J Comput Virol (2007) 3:87–92
DOI 10.1007/s11416-007-0048-y

EICAR 2007 BEST ACADEMIC PAPERS

A spatial stochastic model for worm propagation: scale effects

Markos Avlonitis · Emmanouil Magkos · Michalis Stefanidakis · Vassilis Chrissikopoulos

Received: 12 January 2007 / Accepted: 22 March 2007 / Published online: 13 April 2007

Abstract Realistic models for worm propagation in the Internet have become one of the major topics in the academic literature concerning network security. In this paper, we propose an evolution equation for worm propagation in a very small number of Internet hosts, hereinafter called a subnet and introduce a generalization of the classical epidemic model by including a second order spatial term which models subnet interactions. The corresponding gradient coefficient is a measure of the characteristic scale of interactions and as a result a novel scale approach for understanding the evolution of worm population in different scales, is considered. Results concerning random scan strategies and local preference scan worms are presented. A comparison of the proposed model with simulation results is also presented. Based on our model, more efficient monitoring strategies could be deployed.

1 Introduction

Computer worms are autonomous programs that self-propagate across computers and networks, by exploiting security flaws in widely used services. Early worms such as Code Red [10], Slammer (or Sapphire) [9], Blaster [2], Nimda [12] and their variants have caused significant damage in the Internet infrastructure. While early worms aimed mostly at denial of service (DOS) attacks and congesting network lines, future worms are expected to bear catastrophic payloads such as access to confidential information, data erasure and corruption, hardware damage, etc [19]. Future worms are also expected to employ fast spreading strategies, cooperative scanning, and distributed control techniques [17]. Such worms, exploiting a relatively homogenous software base, high-bandwidth connectivity and poor security policies, may result in significant congestion within the Internet core and may even bring down a large part of the information infrastructure within a short time interval.

After being deployed, each worm instance applies a scanning strategy in order to propagate. Three main scanning strategies are considered in the literature [17,24]. Random scanning worms (e.g. Code Red I) uniformly scan the entire address space to find vulnerable targets. In local preference scanning (e.g. Blaster, Code Red II, Nimda) the worm scans preferably the IP addresses that are close to its address (e.g. within the same /8, /16, or /24 network). In sequential scanning, the infected host sequentially scans from a starting IP address, selected by the worm, and then increments the address by one.

Worm propagation models are mathematical models that attempt to capture the propagation dynamics of scanning worms in order to understand better the various worm types and behaviors, as well as to design and evaluate strategies for monitoring and early detection of the worm. While it seems hard to create realistic models mainly due to the heterogeneity of the Internet networks, recent research has come up with analytical models and simulation results that approximate well the behavior of real random scanning worms such as the Code Red and Slammer worms, for which real measurements are disposable on the Internet. Admittedly, the difficulty in evaluating analytical models that describe scanning strategies

M. Avlonitis (✉) · E. Magkos · M. Stefanidakis · V. Chrissikopoulos
Department of Informatics, Ionian University,
Platia Tsirigoti 7, 49100 Corfu, Greece
e-mail: avlon@ionio.gr

E. Magkos
e-mail: emagos@ionio.gr

M. Stefanidakis
e-mail: mistral@ionio.gr

V. Chrissikopoulos
e-mail: vchris@ionio.gr

other than random scanning, such as local preference scanning, cannot be evaluated with real world data. In such a case, a simulation model with realistic network characteristics has to be designed [14].

Our contribution The classical epidemic model is able to describe random scanning strategies assuming the Internet as a uniform underlying infrastructure. Recent models have extended the classical model to describe random scanning strategies assuming a non-uniform network infrastructure (e.g. [15]). Others [24], have modeled a local preference strategy, however assuming a uniform underlying infrastructure. As a result, most worm propagation models in the literature introduce their evolution equations by considering the population of infected hosts in the whole Internet (classical epidemic) or in large scale networks (e.g. [15,24]) in order to describe the behavior of a worm. In this paper we extend the classical model by introducing the notion of an elementary network quantity, i.e. a very small number of Internet hosts, hereinafter called a subnet, and propose an evolution equation for worm propagation into an arbitrary subnet. The formalism can take into account non-uniformities that are due either to local interactions between neighboring subnets (e.g. as a result of a local preference strategy) or to the heterogeneity of the underlying infrastructure, (e.g. bandwidth variations, different topologies, human countermeasures etc.). In particular, we present an application of the proposed formalism for local preference worms. As a possible application, the proposed approach could be used to construct monitoring strategies, where the propagation of the worm in the whole Internet may be predicted by examining worm propagation within a representative neighborhood of local subnets.

This paper is organized as follows. In Sect. 2 we present related work on worm propagation models and monitoring strategies. In Sect. 5 we extend the classical epidemic model by introducing a method for studying the behaviour of random scanning and local preference worms. In Sect. 6 we validate the presented model with simulation results. Section 7 concludes the paper.

2 Related work

Current worm propagation models follow the line of work that begun from classical epidemiological models (e.g. [1,7]). The first complete application of mathematical models to computer virus propagation was proposed in [6]. Epidemiological models are given names according to the initial letters of the possible states an entity can be at. For example, the Kermack–McKendrick model with the states {Susceptible, Infected and Recovered} is referred to as the SIR model, where a node cannot be infected more than once. The model in [6] was a SIS model, where a node is infected and cured repeatedly. In the simple SI epidemic model [5], a host is infected by the virus and will stay infected forever.

2.1 Scanning strategies

The RCS model, described in [17] attempted to model random scanning worms which peak before a remedy is deployed (i.e. it follows the SI model). The authors used empirical data from the Code Red v.2.0 worm and considered the Internet topology as an undirected completed graph. Later, in [22] the RCS model was extended to include host disinfection (the "R" in the SIR model). In the model of [13] for random scanning worms, the states of the SIR model are extended to include susceptible nodes that are removed from the network or simply quarantined. In [24], mathematical models and simulations for current and future scanning strategies are presented without considering human interaction (i.e. a SI model) or network congestion. In [24] it is also said that local preference strategy improves worm propagation when the vulnerable hosts are more densely distributed within a given network. They also model local preference worms following the SI model and present simulation results.

3 Non-homogeneous aspects

Besides human intervention, analytical models attempt to reflect how worm propagation can be influenced by other factors, such as the underlying network topology [18], the differences of bandwidth within and between networks, the infection delay [20] or even the congestion caused by the worm itself [22]. For example, in [15] the RCS model was further extended to describe bandwidth limited (UDP scanning) worms such as the Slammer worm. The model takes into account the possible variations in the capacity of connections between different autonomous systems. In another proposal, the work in [8] for bandwidth-limited worms models the spread of random scanning worms given bandwidth limitations within heterogeneous networks. In [21], modelling worms with varying scan rate was presented.

4 Monitoring strategies

Traditional techniques and strategies for protecting against worm intrusions in a proactive and reactive manner are human based [11]. Admittedly though, early detection and response for fast spreading worms cannot involve human reaction as the worm may infect the majority of vulnerable nodes before human countermeasures can be taken. As a result, much attention has been shed on automated real-time monitoring. Anomaly detection techniques [21,23] with distributed monitoring within the local network, may trigger automated

response (e.g. automated alarms, worm signatures, quarantine of nodes, automated patch dissemination [16] etc.). In recent works, local host information is exploited for early detection of worms within local networks [3,4,16]. Staniford et al. [17] proposed the establishment of a Cyber Decease Center to collect global security-related information for early worm detection. Similar proposals for global strategies include the use of distributed network telescopes [11], or the employment of a network of honeypots (i.e., hosts that pretend to own a number of IP addresses) to attract and identify attackers [15]. Special filters that capture the dynamics of worm propagation could also be inserted as part of the filtering in network routers. For example, the Kalman filter for random scanning worms is based on the simple SI epidemic model [23]. All observation data collected by the distributed monitors, are being sent in real time to a central server which, in real time compares the results with the figures expected by the propagation model and acts accordingly.

5 A spatial stochastic propagation model

In this section we are interested to model the evolution of the worm population in the whole internet by focusing in the behavior of local subnets instead of considering large networks or Autonomous Systems. This point of view has the advantage that interactions between subnets may be taken into account for worm propagation. As a result the Internet can be macroscopically thought of as the interconnection of many interacting subnets. In each subnet an epidemic model for worm propagation may be considered.

Following the line of work by Serazzi et al. [15], let N_i be the number of susceptible hosts in the ith subnet and I_i the infected hosts in the same subnet. Suppose that K is the average propagation speed of the worm and in a first approximation let us say that it is constant in every single subnet. Assuming a random scanning strategy, there is a probability P_{IN} that a host inside the subnet targets a host inside the same subnet and a probability P_{OUT} that instead it attacks another subnet. Without lost of generality we assume that all subnets have the same size $N_i = N_s$. Under the assumption that the worm randomly generates the target IP address, the following evolution equation for the density of infected hosts $a_i = \frac{I_i}{N_S}$ (I_i is the number of infected hosts) in an arbitrary subnet holds (taking into account both the internal and external worm infections attempts [15])

$$\frac{da_i}{dt} = \left[a_i K \frac{N_S}{N} + \sum_{j=1, j\neq i}^{n} a_j K \frac{N_S}{N} \right] (1 - a_i) \tag{1}$$

where n is the number of subnets in the Internet which has a total of N susceptible hosts. Simplifying further,

$$\frac{da_i}{dt} = \left[\sum_{j=1}^{n} a_j K \frac{N_S}{N} \right] (1 - a_i) \tag{2}$$

In a continuum limit a single subnet may be viewed as a volume element at arbitrary position x. Here the space coincides with the space of susceptible IP addresses and the above evolution equation may be written as

$$\frac{da(x,t)}{dt} = K \frac{N_S}{N} (1 - a(x,t)) \left[\int_n a(y,t) dy \right] \tag{3}$$

where integration is performed over all subnets $n = \frac{N}{N_s}$.

In the more general case, assuming quite smooth variations of the number of infected hosts in neighborhood sites, in a first approximation using a Taylor expansion around x ($y = x + r$) we may write

$$\frac{da_X}{dt} = K \frac{N_S}{N} (1 - a_X) \int_n \left(a_X + r \frac{\partial a_X}{\partial x} + \frac{1}{2} r^2 \frac{\partial^2 a_X}{\partial x^2} \right) dr \tag{4}$$

where we have used the abbreviation $a_X = a(x,t)$.

Equation (4) is our final result. The result is a spatial generalization of the simple epidemic model in order to capture non-uniformities between subnets. In this form the proposed model is able to describe different patterns of infected hosts, either because of non-uniform scanning strategies (e.g. local preference), or because of the inherent heterogenoneity of the Internet networks. As a result, a more realistic model of worm propagation can be constructed.

5.1 Case studies

(A) **Random scanning worm**

In the simplest case (also followed by most analytical models in the literature), we consider a uniform scanning strategy and the Internet as homogeneous. In this case, the number of infected hosts uniformly increases within the Internet. As a result a uniform spatial distribution emerges and the spatial partial derivatives in Equation (4) vanishes,

$$\frac{da_X}{dt} = K \frac{N_S}{N} (1 - a_X) \int_n a_X dr \tag{5}$$

or, using that $a = \frac{1}{N} \int_n N_S a_X dr$ (in accordance with the discrete definition $a = \frac{1}{N} \sum_i N_i a_i$ where a is the total or average density of infected hosts in the Internet),

$$\frac{da_X}{dt} = K a (1 - a_X) \tag{6}$$

In this scenario our model coincides with the well known epidemic model. As a result the above analysis shows that the inherent weakness of the classical epidemic model lies in the assumption of homogenous Internet structure. Since uniform infrastructure was assumed, the behavior of an arbitrary subnet (theoretically of any size but in practice there is a lower limit below of which the phenomenon is discrete in nature) coincides with the behavior of the Internet as a whole. Indeed, Equation (5) may be rewritten as follows,

$$\frac{da_X}{dt} = K\frac{N_S}{N}(1-a_X)a_X \int_n dr \tag{7}$$

and using that $\int_n dr = n$ we end up with (recall also that $N = N_S n$),

$$\frac{da_X}{dt} = Ka_X(1-a_X). \tag{8}$$

On the other hand using Equation (6) (recall that $a = \frac{1}{N}\int_n N_S a_X dr$),

$$\frac{da}{dt} = \frac{1}{N}\int_n N_S \frac{da_X}{dt}dr = \frac{N_S}{N}aK\left(n - \frac{aN}{N_S}\right) \tag{9}$$

or finally

$$\frac{da}{dt} = Ka(1-a). \tag{10}$$

Comparing Equation (8) and (10) we confirm what was intuitively expected (as stated before), that when no non-uniformities are present, the propagation of a worm in arbitrary subnet proceeds independently and as a result the average behavior of a worm population in the Internet coincides with its propagation behavior within any single subnet.

(B) **Local preference scanning**

In this section we model a local preference scan worm with a uniform probability to scan addresses in its own "/m" prefix network. As a result a non-uniform distribution of infected hosts emerges and the spatial derivatives in Equation (4) are no longer negligible. Following Zou et al. [24] we rewrite Equation (4) as follows,

$$\frac{da_X}{dt} = \beta N_S(1-a_X)\int_n \left(a_X + r\frac{\partial a_X}{\partial x} + \frac{1}{2}r^2\frac{\partial^2 a_X}{\partial x^2}\right)dr \tag{11}$$

where $\beta = \frac{K}{N}$ is called the pairwise rate of infection. In [24] this is expressed as,

$$\beta = \frac{\eta}{\Omega} \tag{12}$$

where η is an average scan rate and Ω is the total number of IP addresses, while for a local preference worm they introduce the following β' and β'' pairwise rates of infection in local and remote scan respectively,

$$\beta' = \frac{p\eta}{2^{32-m}}, \qquad \beta'' = \frac{(1-p)\eta}{(Q-1)2^{32-m}} \tag{13}$$

where Q is the number of "/m" prefix networks in Ω. For the arbitrary subnet at x, the integration in Equation (11) must be split into two domains, first into the same network and a second one over the remaining networks. As a result we write (using the corresponding rates β' and β''),

$$\frac{da_X}{dt} = N_S(1-a_X)\left[\int_{Qx}\beta'\left(a_X + r\frac{\partial a_X}{\partial x} + \frac{1}{2}r^2\frac{\partial^2 a_X}{\partial x^2}\right)dr + \int_{\Omega - Qx}\beta''\left(a_X + r\frac{\partial a_X}{\partial x} + \frac{1}{2}r^2\frac{\partial^2 a_X}{\partial x^2}\right)dr\right] \tag{14}$$

where Q_x is the "/m" prefix network which contains the arbitrary subnet. To proceed further we make the following assumption: for quite smooth heterogeneities the second order spatial derivative vanishes in the $\Omega - Q_x$ domain. Performing the integration and noting that the first spatial derivative vanishes because of symmetry we end up with,

$$\frac{da_X}{dt} = N_S(1-a_X)\left[\beta' a_X + \beta' c\frac{\partial^2 a_X}{\partial x^2} + (Q-1)\beta'' a_X\right] \tag{15}$$

or

$$\frac{da_X}{dt} = N_S(1-a_X)\left\{\left[\beta' + (Q-1)\beta''\right]a_X + \beta' c\frac{\partial^2 a_X}{\partial x^2}\right\} \tag{16}$$

where

$$c = \frac{1}{2}\int_{Qx} r^2 dr \tag{17}$$

This is the final evolution equation for local preference scanning worms in an arbitrary subnet. The contribution of the formalism is a specific law for the local evolution of worm propagation taking into account the heterogeneities arising from a local preference strategy. Equation (16) can be analytically solved in order to estimate the evolution of the worm population in the whole Internet. The formalism introduces as a crucial model parameter, the gradient coefficient c in Equation (17) which represents a critical scale of interactions between subnets. This means that in a neighborhood of this scale the worm population proceeds independently. As a result, the evolution of the worm population within this neighborhood of subnets coincides with the evolution of the population in the Internet as a whole.

6 Simulation results

In order to verify the analytical model outcomes, we have built a simple discrete event simulator for a setup equivalent to a /16 network. Under this setup, a total number of 256 networks (LANs) has been simulated. Each LAN has 256 hosts, all susceptible to worm infection. Initially, a single computer node in an arbitrary LAN is in infected state.

The general characteristics of the simulated worm are: 1 infection probe per time unit (which in our scenario roughly corresponds to 1 ms), UDP-like scanning with no connection establishment delays, 10 time units for intra-LAN and 100 time units for inter-LAN infection propagation. The simulation executions follow the evolution of worm infection in two distinct cases:

- In the first case, the addresses generated for infection scanning have a *uniform* (random) distribution, disregarding any information about locality. Each infection probe can target any other host in simulated setup with equal probability.
- In the second case, the worm exhibits a *local preference* in the probe addresses it generates. Under our scenario (and following the characteristics of a Blaster-like worm) 40% of the generated addresses target other hosts in the same LAN, while the remaining 60% targets hosts in random LANs.

In both cases, we study the density of infected hosts within the total susceptible hosts, either in the whole setup (total infection density), or in each subnet LAN (local infection density). The goal of this study is to investigate and validate the proposed model in relation to the predicted scale behavior of worm propagation. To this end, the evolution of densities of infected nodes in global (over total setup) and local (per subnet LAN) ranges, are examined. Simulation results for the random scanning case a redepicted in Fig. 1. Analysis of the average error in estimating the global infection density, by using values of local densities into subnets of 256, 128 or 64 hosts, was performed. The results show a negligible average error of the order of 1% in accordance with the proposed model until a certain limit of analysis (∼50 hosts) is applied. Below this limit the phenomenon tends to be discrete. The same analysis for local preference probing was performed. Simulation results are depicted in Fig. 2. The departure of the local evolution of the worm population from the global evolution is evident below a critical subnet size. Indeed in the case of a subnet size of 64, 128, 256 there is an average error of 40% in the estimation of the global infection density. On the other hand, when a critical size of 512 hosts is considered, the corresponding estimation error is of the order of 15%. This preliminary result shows in a clear way the significance of accurately estimating the critical size of the subnet over which monitoring must be performed in order to predict the global density of infection.

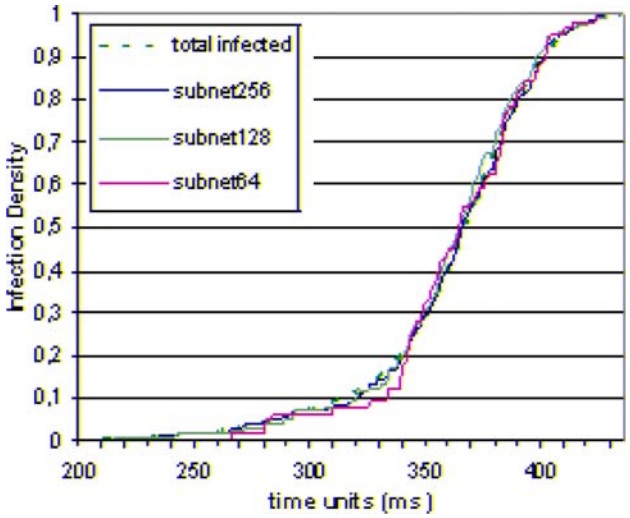

Fig. 1 Infection in an arbitrary LAN—random scanning

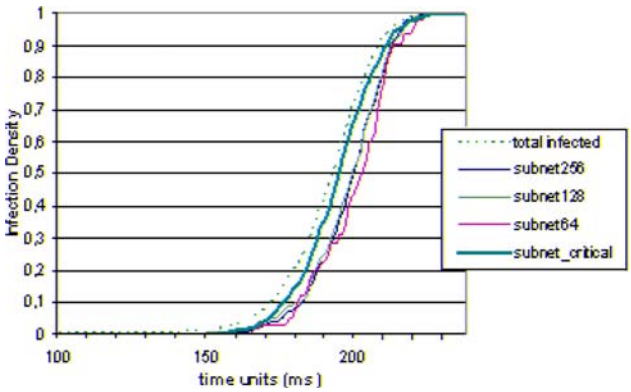

Fig. 2 Infection in an arbitrary LAN—local preference

7 Discussion and future work

In this paper we presented an evolution equation for describing a worm's propagation within a very small scale network of hosts, i.e. the subnet. At this scale it is possible to capture more accurately non-uniformities that are due either to local interactions between neighboring subnets or to the heterogeneity of the underlying infrastructure within a neighborhood of subnets. This was done by including a second order spatial term which models the aforementioned subnet interactions. The corresponding gradient coefficient is a measure of the characteristic scale of interactions and as a result a novel scale approach for understanding the evolution of worm population in different scales, was considered. In particular, we presented an application of the proposed

formalism for random scanning and local preference worms, and presented simulation results.

A practical implication of the proposed approach could be the construction of more effective monitoring strategies. Indeed, global strategies for monitoring worm activity require a large monitored network to become effective. In view of fast spreading worms, it seems difficult to setup nation-wide or global monitoring infrastructures for early warning [23]. On the other hand, local strategies can be more effective in early detecting and setting up threshold alarms. Furthermore, monitoring traffic within local networks could turn out to be more effective against fast worms with non-uniform behavior, such as local preference scanning. However, as it was shown in Fig. 2 this can be misleading because of interactions between subnets. Our analysis demonstrates that we may maintain local monitoring strategies, by introducing a neighborhood of subnets of appropriate size over which the evolution of worm population follows correctly the evolution of the population in the global Internet. In such a monitoring architecture, special filters could be run either at the router of each subnet or at distinct hosts among subnets, in order to capture the evolution of the worm population. A centrally controlled entity, similar to [23], would gather the distributed log data and compare them in real time against the epidemic model for counting the alarm threshold.

In a future work we intend to elaborate on the details of an effective monitoring architecture for early detection of fast spreading worms. Furthermore, observe that the generalization of the classical epidemic model, proposed in Sect. 5, considered local preference scanning worms and a homogeneous infrastructure. The model could be extended to take into account other non-uniformities such as bandwidth variations, different topologies, varying scan rates etc. Finally, the gradient coefficient c of Sect. 5, was introduced in a more or less phenomenological way: although simulation results confirmed its validity, we have not provided a rigorous estimation of its value, derived from local preference scanning features. This should also be addressed in a future work.

References

1. Anderson, R.M., May, R.M.: Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford (1991)
2. Blaster (2003) eEye Digital Security, Blaster worm analysis. http://www.eeye.com/html/Research/Advisories/AL20030811.html
3. Costa, M., Crowcroft, J., Castro, M., Rowstron, A., Zhou, L., Zhang, L., Barham, P.: Vigilante: End-to-End Containment of Internet Works. In: 20th ACM symposium on Operating systems principles, pp. 133–147 (2005)
4. Gu, G., Sharif, M., Qin, X., Dagon, D., Lee, W., Riley, G.: Worm detection, early warning and response based on local victim information. In: 20th Annual Computer Security Applications Conference, pp. 136–145 (2004)
5. Heathcote The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
6. Kephart, White: Directed graph epidemiological models of computer viruses. IEEE Symposium on security and privacy (1991)
7. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. In: Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, vol. 115, No. 772, pp. 700–721 (1927)
8. Kesidis, G., Hamadeh, I.: Jiwasurat Soranun.: coupled Kermack–McKendrick models for randomly scanning and bandwidth-saturating internet worms. In: QoS-IP 2005, LNCS 3375, pp. 101–109 (2005)
9. Moore D., Paxson V., Savage S. Inside the slammer worm. Slammer **1**, 33–39 (2003)
10. Moore D., Shannon C., Brown J.: Code-red: a case study on the spread and victims of an internet worm. In: Proceedings of 2-th Internet Measurement Workshop (IMW) (2002)
11. Moore, D., Shannon, C., Voelker, G.M., Savage, S.: Internet Quarantine: Requirements for Containing Self-Propagating Code. INFOCOM (2003)
12. Nimda (2001) Symantec, "W32.nimda.a@mm," http://www.symantec.com/avcenter/venc/data/w32.nimda.a@mm.html
13. Onwubiko, et al.: An improved worm mitigation model for evaluating the spread of aggressive network worms (2005)
14. Sharif, M.I., Riley, G.F., Lee, W.: Comparative study between analytical models and packet-level worm simulations. In: 19th Workshop on Principles of Advanced and Distributed Simulation. IEEE, pp. 88–98 (2005)
15. Serazzi, G., Zanero, S.: Computer virus propagation models. In: Calzarossa, M.C., Gelenbe, E. (eds.) Tutorials of the 11th IEEE/ACM Int'l symp. on modeling, analysis and simulation of computer and telecom—systems—MASCOTS 2003. Springer, New York (2003)
16. Sidiroglou, S., Keromytis, A.D.: (2005) Countering network worms through automatic patch generation. IEEE Security and Privacy
17. Staniford S., Paxson V., Weaver N.: How to own the internet in your spare time. In: Proceedings of the 11th USENIX security symposium (Security '02) (2002)
18. Wang Y., Chakrabarti D., Wang C., Faloutsos C.: Epidemic spreading in real networks: an eigenvalue viewpoint. In: Proceedings of the 22nd International Symposium on Reliable Distributed Systems (2003)
19. Weaver N., Paxson V., Staniford S., Cunningham R.: A taxonomy of computer worms. In: ACM Workshop on Rapid Malcode (WORM) (2003)
20. Yang, W., Chenxi, W.: Modeling the effects of timing parameters on virus propagation. In: Proceedings of the 2003 ACM workshop on Rapid Malcode, pp. 61–66. ACM Press, New York (2003)
21. Yu, W., Wang, X., Xuan, D., Lee, D.: Effective detection of active worms with varying scan rate. In: Proceedings of IEEE International Conference on Security and Privacy in Communication Networks (SecureComm), to appear (2006)
22. Zou, C.C., Gong, W., Towsley, D.: Code red worm propagation modeling and analysis. In: Proceedings of the 9th ACM conference on computer and communications security, pp 138-147. ACM Press, New York (2002)
23. Zou, C., Gong, W.B., Towsley, D., Gao, L.X.: Monitoring and early detection for internet worms. In: Proceedings of the 10th ACM Conference on Computer and Communication Security (CCS), October (2003)
24. Zou Cliff, C., Towsley, D., Gong, W.: On the performance of Internet worm scanning strategies. Performance Evaluation 63 (2006), Science Direct, pp. 700–723 (2006)