A Novel Stochastic Approach for Modeling Random Scanning Worms

Markos Avlonitis, Emmanouil Magkos, Michalis Stefanidakis and Vassilis Chrissikopoulos
Department of Informatics
Ionian University
Platia Tsirigoti 7, 49100 Corfu, Greece
Email:{avlon,emagos,mistral,vchris}@ionio.gr

***Abstract*—Scanning worms grow with different local velocities in different areas, because of non-uniformities that are present in real networks. The present work introduces a new stochastic model elaborating the classical epidemiological model for random scanning strategies. More specifically, random effects in worm spreading velocity are modeled by means of a stochastic differential equation where an explicit expression quantifying randomness is proposed. Furthermore, we explore whether deterministic or stochastic models are appropriate in order to describe the worm propagation phenomenon. To this end we introduce the scale of observation as a crucial parameter. Simulation results are presented validating the proposed analytical results.**

I. Introduction

Scanning worms search for their targets by scanning target ports of other nodes in order to locate software applications with specific vulnerabilities that allow delivery of the malicious code. They self-propagate very fast and in large scales because of the relatively homogeneous software base and the high bandwidth connectivity between Internet nodes [1]. Depending on their scanning strategy, scanning worms can also be seen as random, local preference, sequential, or topological scanning worms ([2], [3]). The threat of an advanced scanning strategy that may appear in the future has been repeatedly discussed in the literature, under the names of hitlist [3], routing ([4], [5]) and importance scanning worms [6], permutation [3] or divide-conquer worms [5]. At high level these are selective worms that spread faster by carefully selecting their victims instead of "blindly" scanning the universe for possible targets ([5], [4], [3]).

Mathematical models can help the security research community to understand the threat and analyse the propagation pattern during the lifetime of a worm. By analysing a worm's behaviour and the factors that influence its spread, we can have insights into effectively detecting and containing a fast spreading worm ([7], [8], [9]).

Worm propagation is in fact a *stochastic* process, as random effects are present in real networks. This source of randomness roots in the various non-uniform network parameters that influence the propagation of a worm ([3], [7], [4], [10], [11], [6], [12], [13], [14]). Most of these can be categorized as environment-related ($e.g.$, bandwidth, traffic, topology, vulnerable hosts distribution), human-related (preventive and reactive measures, removal tools, disconnecting or isolating hosts, blocking access to a service, operating system updating or restoring, training users etc), policy-related (network or host-level firewall policies, automatic quarantine) or worm-related ($e.g.$, scanning strategy, scan rate, congestion, victim deaths).

Our Contribution: This work introduces a novel approach for modeling fast spreading worms in the Internet. A stochastic model elaborating the classical epidemiological model for random scanning strategies is proposed. Random effects in worm spreading velocity are modeled by means of a stochastic differential equation where an explicit expression quantifying randomness is proposed. Furthermore, we explore whether deterministic or stochastic models are appropriate in order to describe the propagation phenomenon. To this end we introduce the scale of observation as a crucial parameter. Simulation results validate the proposed analytical results.

II. Related work

Epidemiological models for analysing the spread of computer malware are not new [15]. Early attempts capture the mechanism of random scan worms and use the simple epidemic model to study the initial part of worm spreading, where human countermeasures and congestions do not affect the propagation ([7], [3]). In recent years, a number of deterministic models were designed to consider the parameters that affect the worm propagation, for random scanning (*e.g.*, [7], [3], [9], [16], [13]), local preference (*e.g.*, [11], [8], [12]) or other advanced strategies ([5], [8], [4], [14], [6]). The two-factor model in [7] takes into account the congestion caused by the worm scan packets, as well as the reactive (human) countermeasures that turn infected or susceptible nodes into an immune state. Models that consider the preventive measures ($e.g.$, antivirus and patch management [17]), link bandwidth between systems ([9], [16], [13]), network topology [18], the slow down caused by automatic treatment and containment measures ([19], [1], [11], [20]), infection delay and user vigilance [21], have also been proposed in the literature.

Due to the observed randomness affecting worm propagation in the Internet ($e.g.$, [10], [1]) the so called stochastic models have been emerged in the literature ($e.g.$, [22], [23]). These models, contrary to deterministic models that use differential equations to express a mean field behavior, are based on the observation that worm propagation is an inherently random process. In the above models randomness emerges because of the scanning strategy while other sources of randomness, $e.g.$,

bandwidth limitation or network topology, are not covered. The above approaches propose discrete Markov models in order to predict propagation at early stages introducing as an appropriate variable the amount of time for the next infection, as well as estimating its mean value and variance in order to construct robust detection protocols.

III. Determinism versus randomness

Let us assume a random scanning worm that propagates over a network with N unique hosts, where $N_s \leq N$ of these addresses could potentially become infected by the worm. In an arbitrary ensemble of hosts ($i.e.$, a *subnet*) and at arbitrary time the population N is split into infected and susceptible sub-populations, represented by $I(t)$ and $S(t)$, respectively.

If recovery and/or removal of hosts are taken into account, extra population densities or model parameters must be introduced (*e.g.*, [7], [20]). In this article we focus in the simplest case, although our approach can be generalized in a straightforward way.

Infected hosts increase their population by sending infectious packets to other randomly selected hosts. This process proceeds at a constant rate. If a susceptible host receives an infectious packet, then that host becomes infected. In order to extract the propagation rate of a randomly scanning worm, it is assumed that at arbitrary time, any infected host sends an infectious packet. This packet at time t has a probability $S(t)/N$ of being sent to a susceptible host. As a result, if β is the constant scan rate then $\beta S(t)I(t)/N$ is the rate at which the infection is propagated. The classical epidemic model can be expressed with the following ordinary differential equation.

$$\frac{dI(t)}{dt} = \frac{\beta}{N} S(t)I(t) \tag{1}$$

or

$$\frac{dI(t)}{dt} = f(I(t)) \tag{2}$$

where $f(I(t)) = \frac{\beta}{N}(N - I(t))I(t))$ is the spreading "force" over the scale of the entire Internet. While the classical model successfully describes average properties of worm propagation in the Internet, it is not able to treat randomness arising in real networks because of network inhomogeneities or sources of randomness.

In order to decide about the nature of the constitutive variables $(S(t), I(t))$ and parameters (β) that enter into the propagation problem (generalization to more complex models follows the same reasoning) it is noted that the previous relations may be referred to different network scales: within one single subnet i, within a neighbourhood of subnets or within the entire Internet. We distinguish these size scales to *micro*, *meso* and *macroscale*, correspondingly. As a milestone of this work we argue that models that refer to different scales predefine the nature of the above variables and parameters. More specifically, if a model tries to describe the behaviour of the worm propagation in microscale then a probabilistic model is the only choice, and $(S(t), I(t))$ are interpreted as random variables ($e.g.$, [22], [23]). On the other hand if a model is referred to the macroscopic behaviour of a worm, deterministic models are more appropriate and $(S(t), I(t))$ are interpreted as deterministic variables, ($e.g.$, [15], [7], [3]). The link between these models is an approach that is able to describe worm propagation in the mesoscale, which is an appropriate scale for real-world monitoring systems. To this end, a new stochastic model is proposed in the next section.

IV. A stochastic differential model

In mesoscale, the population variables $S(t), I(t)$ are interpreted as stochastic variables. Moreover the infection parameter β is also random and assumed to fluctuate by the amount $\delta\beta$ around a mean value $< \beta >$, *i.e.*,

$$\beta =< \beta > +\delta\beta \tag{3}$$

where the operator $< \cdot >$ interprets average values in time. Substituting the infection parameter β in the corresponding evolution equation (Eq. (1)) a random fluctuating part of the spreading "force" is obtained,

$$\frac{dI}{dt} = f_{<\beta>}(I) + \delta\beta \tag{4}$$

where

$$f_{<\beta>}(I) = \frac{<\beta>}{N}(N - I) \cdot I \tag{5}$$

and

$$\delta f = \frac{1}{N}(N - I) \cdot I \cdot \delta\beta \tag{6}$$

Eq. (4) is our final proposed evolution equation describing the dynamics of worm propagation. It is emphasized that Eq. (4) cannot be directly integrated, because of the random fluctuating term of the second part. It belongs to a general class of stochastic differential equations, being able to describe with success the evolution of dynamical systems in the mesoscale ($e.g.$, [24], [25], [26]). Solutions of the proposed stochastic evolution equation and corresponding estimates of dynamical parameters arising from Eq. (4) will be left to future work.

The stochastic differential model provides a quantitative estimate for the inherent randomness. Indeed, according to Eq. (4) a measure of the resulting noise is,

$$Q = \sqrt{< \delta f^2 >} \tag{7}$$

It is important to emphasize that the value of Q interprets the inherent randomness that emerges in real networks. As was discussed earlier in this article, these sources of randomness may root to the heterogeneity of network infrastructure or randomness in traffic conditions. To have an estimate of this, the white noise limit for the fluctuating part of the infection parameter is adopted,

$$\delta\beta = \dot{w}, < \delta\beta >= 0 \tag{8}$$

and

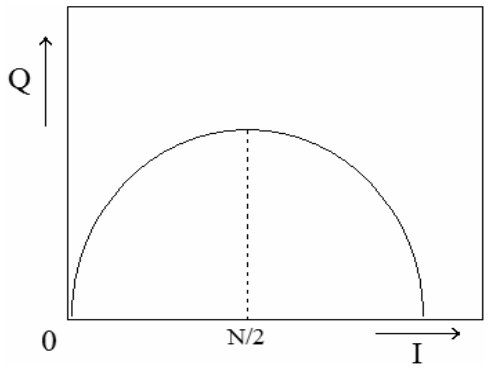

Fig. 1. Fluctuation amplitude of infected hosts for arbitrary model parameters

$$< \delta\beta_i \cdot \delta\beta_j >= \sigma^2 \delta_{ij} \tag{9}$$

where σ^2 is the amplitude of the white noise. Furthermore, substituting $I =< I > +\delta I$, using Eq. (9) and neglecting higher orders, the following estimate for the fluctuation amplitude of the infected population in the Internet is obtained,

$$Q(I) = \frac{\sigma}{N}(N - < I >) \cdot < I > \tag{10}$$

Eq. (10) is one of the main contributions of this paper since it quantifies the inherent randomness of worm propagation. The estimation of the amplitude of emerged randomness is given in terms of the variables and parameters of the problem.

In Figure 1, the shape of fluctuation amplitude $Q(I)$ is drawn for arbitrary parameters. The model predicts that at the beginning and at the end of the propagation the amplitude is closed to zero while it reaches its maximum at $I = N/2$.

V. Simulation results

Next to the development of the theoretical model, we studied the characteristics of malware infection spread via detailed discrete event simulation. In this section we describe our simulator's setup and results.

The code of the simulator models a fast UDP scanning worm with a minimal payload packet. The employment of UDP enables an aggressive behavior of the worm without TCP handshaking delays. In this way, the scanning rate of an infected host is effectively limited only by the available bandwidth of its interconnecting network interface. In our setup the average scanning rate is set to 1 probe per ms, which is also our base simulation timing unit. Following the theoretical model, the simulated worm is assumed to exhibit a uniform scanning strategy, by targeting every node in the setup with equal probability. Our main goal is to study the early stage of rapid infection spread, consequently each simulated computer node is modeled to be in one of two states, either susceptible or infected. That is, no recovery or immunization actions are taken during simulation execution.

During simulation time, we study an Internet portion of 256 C-class networks. Each network is treated as an independent LAN with a network backbone interconnecting all 256 LANs.

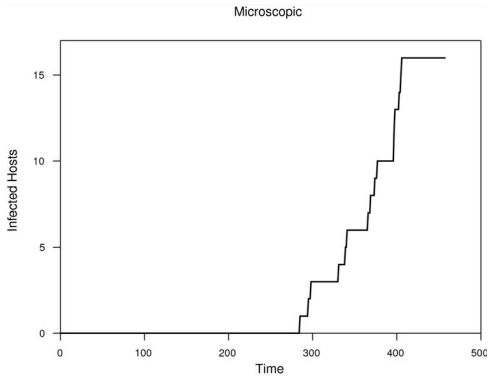

Fig. 2. Infected hosts in microscale

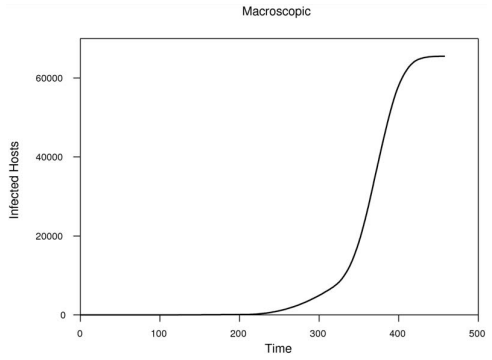

Fig. 3. Infected hosts in macroscale

Each LAN internally is assumed to have a total bandwidth of 100 Mbps, with the same bandwidth available on egress nodes of every LAN towards the interconnecting backbone. Although an arbitrary traffic is expected within and between LANs, the UDP worm generated traffic is studied as the dominant factor of bandwidth limitation within each LAN. According to the previously mentioned worm characteristics we assume 1% bandwidth overhead for each infected node in a LAN.

In Figures 2-4 the time evolution of the infected hosts is shown in different length scales, micro, macro and mesoscale correspondingly. The simulation results confirm the central aspect introduced in this work, $i.e.$, the scale of observation predefines the deterministic or stochastic nature of the monitored variables.

Indeed, in Fig. 2 it can be seen that system evolution in microscale proceeds in discrete steps and as a result a probabilistic treatment is more appropriate for a robust description of worm propagation. On the other hand in Fig. 3 it can be seen that system evolution in macroscale proceeds in a completely deterministic fashion. In this scale the classical epidemiological models can accurately describe worm spreading.

The picture is quite different in mesoscale where interplay between randomness and determinism is present. Indeed, in Fig. 4 a serration-type of system evolution is observed and in this case a stochastic differential model is more appropriate for the description of worm propagation in the Internet.

In order to validate our stochastic differential model, in Fig.

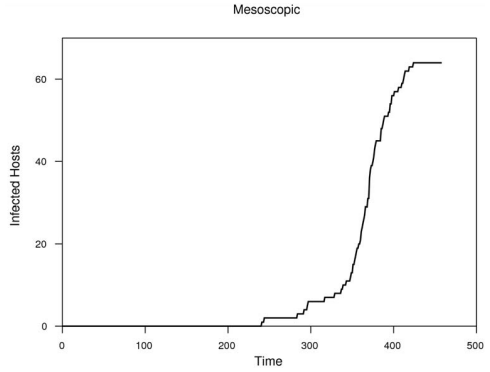

Fig. 4. Infected hosts in mesoscale

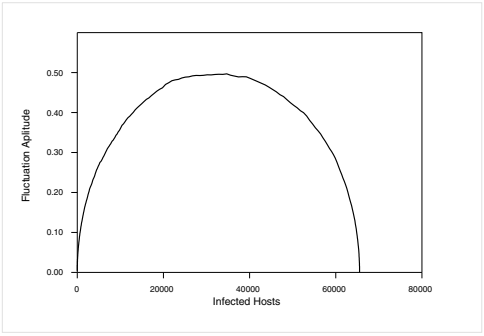

Fig. 5. Fluctuation amplitude of infected hosts

5 the fluctuation amplitude of the population of the infected hosts is plotted. It can be seen that simulation results confirm the analytical expression of Eq. (10).

VI. Conclusions and future work

The present work introduced a novel stochastic model elaborating the classical epidemiological model for random scanning worm strategies. In a future work we intend to elaborate our approach in order to give a theoretical result on the critical size of the network that needs to be monitored (a) in uninfected network segments, for detecting illegitimate traffic, or (b) in infected segments, for studying the worm behaviour. We view this critical size as the least scale size of detection network needed to ensure that a worm is detected within a certain time. Intuitively, a small network size reduces the time for early detection, but increases the false alarms, whereas a large network size leads to more accurate detection but induces a performance cost. Our long-term goal is to balance this inherent tradeoff in real-world monitoring systems.

References

[1] D. Moore, C. Shannon, G. Voelker, and S. Savage, "Internet quarantine: requirements for containing self-propagating code," vol. 3, March-3 April 2003, pp. 1901–1910 vol.3.

[2] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, "A taxonomy of computer worms," in *WORM '03: Proceedings of the 2003 ACM workshop on Rapid malcode*. New York, NY, USA: ACM, 2003, pp. 11–18.

[3] S. Staniford, V. Paxson, and N. Weaver, "How to own the internet in your spare time," in *Proceedings of the 11th USENIX Security Symposium*. Berkeley, CA, USA: USENIX Association, 2002, pp. 149–167.

[4] C. C. Zou, D. Towsley, W. Gong, and S. Cai, "Advanced routing worm and its security challenges," *Simulation*, vol. 82, no. 1, pp. 75–85, 2006.

[5] J. Wu, S. Vangala, L. Gao, and K. Kwiat, "An efficient architecture and algorithm for detecting worms with various scan techniques," in *NDSS"04: Proceedings of the 11th Annual Network and Distributed System Security Symposium*, 2004.

[6] Z. Chen and C. Ji, "Measuring network-aware worm spreading ability," May 2007, pp. 116–124.

[7] C. C. Zou, W. Gong, and D. Towsley, "Code red worm propagation modeling and analysis," in *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*. New York, NY, USA: ACM, 2002, pp. 138–147.

[8] C. C. Zou, D. Towsley, and W. Gong, "On the performance of internet worm scanning strategies," *Perform. Eval.*, vol. 63, no. 7, pp. 700–723, 2006.

[9] G. Serazzi and S. Zanero, "Computer virus propagation models," in *MASCOTS Tutorials*, ser. Lecture Notes in Computer Science, vol. 2965. Springer, 2003, pp. 26–50.

[10] C. Zou, W. Gong, D. Towsley, and L. Gao, "The monitoring and early detection of internet worms," *ACM Transactions on Networking*, vol. 13, no. 5, pp. 961–974, Oct. 2005.

[11] Z. Chen, L. Gao, and K. Kwiat, "Modeling the spread of active worms," vol. 3, March-3 April 2003, pp. 1890–1900 vol.3.

[12] Z. Chen, C. Chen, and C. Ji, "Understanding localized-scanning worms," April 2007, pp. 186–193.

[13] N. Weaver, I. Hamadeh, G. Kesidis, and V. Paxson, "Preliminary results using scale-down to explore worm dynamics," in *WORM '04: Proceedings of the 2004 ACM workshop on Rapid malcode*. New York, NY, USA: ACM, 2004, pp. 65–72.

[14] A. Kamra, H. Feng, V. Misra, and A. Keromytis, "The effect of dns delays on worm propagation in an ipv6 internet," in *Proceedings of IEEE Infocom*. Miami, FL, USA: IEEE, 2005.

[15] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *IEEE Symposium on Security and Privacy*, 1991, pp. 343–361.

[16] G. Kesidis, I. Hamadeh, Y. Jin, S. Jiwasurat, and M. Vojnović, "A model of the spread of randomly scanning internet worms that saturate access links," *ACM Trans. Model. Comput. Simul.*, vol. 18, no. 2, pp. 1–14, 2008.

[17] M. Faghani, H. Saidi, and M. Ataei, "Effects of security solutions on worm propagation," Aug. 2008, pp. 25–29.

[18] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," vol. 2, March 2005, pp. 1455–1466 vol. 2.

[19] C. C. Zou, W. Gong, and D. Towsley, "Worm propagation modeling and analysis under dynamic quarantine defense," in *WORM '03: Proceedings of the 2003 ACM workshop on Rapid malcode*. New York, NY, USA: ACM, 2003, pp. 51–60.

[20] M. Avlonitis, E. Magkos, M. Stefanidakis, and V. Chrissikopoulos, "Treating scalability and modelling human countermeasures against local preference worms via gradient models," *Journal in Computer Virology*.

[21] Y. Wang and C. Wang, "Modeling the effects of timing parameters on virus propagation," in *WORM '03: Proceedings of the 2003 ACM workshop on Rapid malcode*. New York, NY, USA: ACM, 2003, pp. 61–66.

[22] D. M. Nicol, "The impact of stochastic variance on worm propagation and detection," in *WORM '06: Proceedings of the 4th ACM workshop on Recurring malcode*. New York, NY, USA: ACM, 2006, pp. 57–64.

[23] K. R. Rohloff and T. Bacşar, "Deterministic and stochastic models for the detection of random constant scanning worms," *ACM Trans. Model. Comput. Simul.*, vol. 18, no. 2, pp. 1–24, 2008.

[24] R. L. W. Horsthemke, *Noise-induced Transitions*. Springer-Berlin, 1984.

[25] H. Haken, *Synergetics: Introduction and Advanced Topics*. Springer, 2004.

[26] M. A. E. C. Avlonitis, M. Zaiser, "Some exactly solvable models for the statistical evolution of internal variables during plastic deformation," *Probabilistic Engineering Mechanics*, vol. 15, pp. 131–138, 2000.