A Worm Propagation Model based on Scale Free Network Structures and People's Email Acquaintance Profiles

T. Komninos[1,3] P. Spirakis[1,3] Y.C. Stamatiou[2,3] G. Vavitsas[1,3]

[1]*University of Patras, Department of Computer Engineering, 26500, Rio, Patras, Greece*
[2]*University of Ioannina, Department of Mathematics, 451 10, Ioannina, Greece*
[3]*Research and Academic Computer Technology Institute, N. Kazantzaki, University of Patras, 26500, Rio, Patras, Greece.*

Abstract
In light of the rise of malicious attacks on the internet and the various networks and applications attached to it, new approaches propagation exploited by worms is through the victim's contact book. The contact book, which reflects the acquaintance profiles of people, is used as a ``hit-list'', to which the worm can try and send itself in order to spread fast. In this paper we propose a discrete worm propagation model that relies upon a combined email and Instant Messaging (IM) communication behavior of users in a scale free environment. We also model the effect in propagation based on user reaction when a threat is recognized, the installation and update of antivirus software as well as the network connectivity, arising conclusions about the behavior of the network infrastructure in presence of a worm. Our analysis is based on Wormald's differential equations method for approximating ``well-behaving'' random processes with deterministic functions.

Key words:
Worm propagation modelling, e-mail and Instant Messaging Scale free network, intrusion detection

1. Introduction

A *worm* is a self-contained malicious code that is able to spread itself in computer networks. Propagation, usually, occurs through the exploitation of network connections, shared storage, email, Instant Messengers or Peer to Peer (P2P) file sharing networks. Recent research has found that several critical technological networks are scale free structures with power law distributions, such as the Internet at the autonomous system level and the router level [3], the Web [1] [4], and physical SDH telecommunications networks [5]. Studies have examined the vulnerability of scale free networks finding that they are resilient to random attacks, but highly susceptible to targeted attacks [2]. These studies have analyzed scale free networks assuming that all nodes are homogeneously susceptible to attack or infection. Usually in real world networks only subsets of nodes are susceptible to attack or infection. One example of such a scenario is Internet worms which are designed to attack only in specific operating systems or platforms. Simple Mail Transfer Protocol (SMTP), for instance, is one of the most common malicious code propagation vehicles. To spread by email, a worm can propagate towards modeling worm activity in networks is called for. One frequently utilized method for W32/Novarg [7], Sober X, Netsky P and Mytob ED [18]

as an email attachment or embed itself into html code within the email body. Then it obtains email addresses from the victim's computer in order to propagate. Worm propagation modelling has attracted the attention through a series of incidents such as the CodeRed [21] worm, Nimda [8] worm, Slammer worm [16], Sobig [9], W32/Bagle and

Recently, worms have appeared that are able to propagate using another social-like popular communication method such as Instant Messengers (IM) or Peer-to-Peer (P2P) file sharing networks [11]. IM networks provide the ability not only to transfer text messages, but also files supporting peer-to-peer file sharing, leading to the immediate spread of files that are infected. Worms use social engineering to trick people into downloading and execute malicious code [10]. Using IM, worms spread faster as locating potential victims does not require scanning attempts to possibly unknown or unused IP addresses. What they need is simply online users' contact list. However, there were some IM worms which have exploited the processing vulnerabilities described in [15] to allow automatic execution of code. These worms are much faster than any other that requires user intervention and, thus, causes significant devastation. As more users adopt IM services, new worms will spread combining different propagation vectors, not only using email but also IM and P2P links.

While many researchers deal with the development of new techniques for the detection and elimination of worms, there seems to be, relatively, little activity in the theoretical modelling of viral code replication and propagation. Less research effort has been expended on modelling worms that use IM and email simultaneously. Creating reliable models of virus and worm propagation is beneficial for many reasons. First, it allows researchers to better understand the threat posed by new attack vector and new propagation techniques. For instance, the use of conceptual models of worm propagation allowed researchers to predict the behavior of future malware, and later to verify that their predictions were substantially correct [13]. In second place, using such models, researchers can develop and test new and improved models for containment and disinfection of worms without resorting to risky "in vitro" experimentation of zoo worm release and cleanup on test bed networks. If these models are combined with good load modelling techniques such as the

Manuscript received February 5, 2007
Manuscript revised February 25, 2007

queueing networks, we can use them to predict failures of the global network infrastructure when exposed to worm attacks.

In [19] Wang *et al.* study a worm propagation model based on a clustered and a tree-like hierarchic topologies. In their model, copies of the worm propagate at a constant rate without needing user interactions. The lack of a user model coupled with the clustered and tree-like topologies make it unsuitable for modelling the propagation of email and IM worms/viruses over the Internet. Zou et al. studied Code Red worm propagation based on the classical epidemic Kermack-Mckendrick model [21]. Newman et al. derived the analytical solution of the percolation threshold of small world topology [12, 19]. Albert et al. were the first to explain the vulnerability of power law networks under attacks [16]. The authors conclude that the power law topology is vulnerable under deliberate attack. Wang, Knight *et al.* study the effect of immunization on worm propagation [19]. They compare the effect of random immunization and selective immunization. They show that immunizing nodes with highest degrees has better effect than random immunization. This is different from reality where the immunization is randomly applied to hosts by users or administrators.

In [21] Zou *et al.* an email model is given as an undirected graph of relationships between people. It is assumed that each user opens an incoming worm attachment with a certain probability, depending on the user and not on time. This, however, does not describe well the typical user behaviour. Indeed, as the new worm starts spreading there is no user alertness, who tend to open the contaminated attached file. As news about the worm are circulated, users become more cautious. Thus users' behaviour should depend on time. The authors consider a ``reinfection" model, where a user sends out copies of the worm each time an infected attachment is reopened, but this does not add to the infected population as long as the host is either already infected or it has been immunized by an antivirus. An interesting conclusion can be drawn from this study: the overall spread rate of worms increases as the variability of users' email checking times increases. Thus a worm is more vicious as a better social engineering technique is applied. Mannan and van Oorschot [14] review selected IM worms and summarize their main characteristics, motivating a brief overview of the network formed by IM contact lists, and a discussion of theoretical consequences of worms in such networks.

In [6], we proposed email, IM and P2P networks as forming a kind of ``social" network. These networks can be macroscopically considered as an interconnection of a number of Autonomous Systems (AS). In this paper we propose a new model that adapts the previous model to a more realistic, scale free network infrastructure. Using this model we can determine the impact of a worm spreading without having proper antivirus or informed users in a scale free environment.

2 Acquaintance Networks: motivation and formalism

An *acquaintance network* consists of several hypernodes, where each hypernode represents a specific *domain* or LAN (e.g. a university or a company network). Each hypernode contains several nodes which represent personal computers or users' contact information (e.g. email or IM addresses). We assume that with probability p_{intacq} a node of a hypernode contains in its contact book the contact address of another node of the same hypernode. Also, with p_{extacq} a node of a hypernode is associated with a node (user) in a different hypernode. Finally, with probability p_{hyper} we consider that there is a connection between two hypernodes (which means that at least one user of one hypernode is associated with at least one user of the other hypernode). The connections between hypernodes forms the *network acquaintance graph* while the connections between nodes forms the *person acquaintance graph*. Our focus is on modelling a worm outbreak which starts at some random set of nodes and propagates along the acquaintance links.

More formally, an *acquaintance network* consists of a set of hypernodes $X_1,\ldots,X_n$ containing node sets $D_1,\ldots,D_n$ respectively, and a set of acquaintance relations C. An edge $R_{i_1,i_2} \in C$ is a subset of $D_{i_1} \times D_{i_2}$, with i_1, i_2 distinct. We say that R_{i_1,i_2} bounds hypernodes nodes X_{i_1}, X_{i_2} to mutual acquaintance, because of mutual acquaintances stemming from the nodes they contain. The *person acquaintance hypergraph* of a network acquaintance graph is an n-partite graph. Its i th part corresponds to hypernode X_i and it has exactly $|D_i|$ vertices, one for each node in D_i. There exists an eedge $\{v_{i_1}, v_{i_2}\}$ if and only if the corresponding nodes $d_{i_1} \in D_{i_1}$ and $d_{i_2} \in D_{i_2}$ belong to some acquaintance relation that bounds the corresponding variables.

Email, IM and P2P contacts form a kind of social network. Modelling these networks as graphs, with each node representing a host, is clearly unfeasible. On the other hand, they can be macroscopically considered as an interconnection of a number of Autonomous Systems (AS). An AS is a subnetwork usually a Domain Network which is administered by a single authority. For this reason we propose a hypernode based model with a hyper-node representing a Domain Network.

According to the above formalism (which, actually, stems from the formalism of the Constraint Satisfaction Problem (CSP)), the network acquaintance graph represents the structure of email/IM acquaintances across network domains (or LANs). The set of hypernodes $X_1,\ldots,X_n$ represent *Autonomous Systems (AS) or Domains* (e.g. universities) of the acquaintance network, while the node sets $D_1,\ldots,D_n$ represent distinct contact addresses of each domain. These addresses comprise an email or IM acquaintances network. We can safely assume, without loss of generality, that every distinct email or IM address is associated with one host computer which is associated with a single user. The connections between the hypernodes form the *network acquaintance graph*, while the connections among the nodes of each distinct email and IM contact address, form the *person acquaintance graph*. Also, in our model the quantity $B(i)$ represents the number of the infected nodes at step i while

$W(i)$ represents the number of immunized nodes, i.e. nodes on which updated antivirus software is already installed. The quantity $nd - B(i) - W(i)$ represents the number of *susceptible* nodes, that is the number of nodes that have no defence against the new worm.

We will now define some probabilities related to our model: p_{hyper} is the probability that two hypernodes (AS-Domains) are connected. We set this probability as $p_{hyper} = (g-1)m^{(g-1)}k^{(-g)}$ where $2 < g < 3$ is the degree exponent, m is the minimum degree in the network and we are approximating k as a continuous variable. This probability declares how many neighbors on average a hypernode has. Then p_{extacq} is the probability with which a user (or a node) has a contact with a specific user of another hypernode. Also, p_{intacq} is the probability that a node has a contact with a specific node belonging in same hypernode. Moreover, p_{antv} represents the percentage of the nodes protected by updated antivirus software. We can model this probability using a function of time and network size that is gradually increasing with time and decreasing with the size of the network. In particular we set $p_{antiv}(n,t) = \frac{g(t)}{n}$ where $g(t)$ is a monotonicaly increasing function of t and n is the network size which is the number of hypernodes in our model, and p_{openm} is the probability that a user opens his/her email or IM message. We also model this probability as a function of time, $p_{openm}(t) = f(t)$, which is monotonicaly decreasing with time, since as time passes and information of the worm outbreak circulates people are more cautious in opening suspicious email messages. The pair $I = (p_{antv}, p_{openm})$ is called an *attack reaction pair* since it characterizes a new worm that has started propagating within a network.

3 Scale-free random acquaintance graphs

As we saw in the previous section, an acquaintance graph is defined by the following four parameters: (a) The number of hypernodes n, (b) the size of each hypernode d, (c) The network acquaintance graph, and (d) The person acquaintance graph.

In this paper, the random network acquaintance graph will be based on a scale-free random graph model, the model $G_{n,m,g}$, which is defined as follows. Each of n available network nodes selects uniformly and independently of the others to have s neigbours, where s is either 0 (i.e. the node chooses to be disconnected from the rest of the network) or it ranges from m, which is the minimum degree other than 0 that is allowed in the network, up to $n-1$. The probabilities with which these choices are made are given by the following probability function, where $2 < g < 3$:

$$Pr[s = k] = \begin{cases} (g-1)m^{(g-1)}k^{(-g)}, & m \le k \le n-1 \\ m^{g-1}n^{1-g}, & k = 0. \end{cases} \quad (1)$$

Between two hypernodes we may have a double connection (from hypernode A to hypernode B and from B to A) but this probability is very small and we do not take it into account. We will now compute the probability that two nodes of degrees s,t respectively are joined by an edge. This probability will be used in the computation of p_{hyper}.

For a randomly constructed random graph according to model $G_{n,m,g}$, the probability of an edge existing between two nodes of degrees s,t ($0 \le s,t \le n-1$), respectively is given by the following (note that $2 < g < 3$):

$$p_{s,t} = \begin{cases} \frac{s+t}{n-1} + O(\frac{1}{n^{g-1}}), & st = o(n^{3-g}) \\ \frac{s+t}{n-1} - \frac{st}{(n-1)^2} + O\left(\frac{1}{n^{g-1}}\right), & st = \Omega(n^{3-g}). \end{cases} \quad (2)$$

Proof. For any two nodes with randomly chosen degrees s,t respectively, there is an edge between them if their neigbour sets intersect. The probability of this event can be written as follows

$$\begin{aligned} p_{s,t} &= Pr[neigbboursetsintersect] \\ &= Pr[neigbboursetsintersect \mid s+t \ge n-1] \cdot Pr[s+t \ge n-1] \\ &+ Pr[neigbboursetsintersect \mid s+t < n-1] \cdot Pr[s+t < n-1] \\ &= 1 \cdot Pr[s+t \ge n-1] \\ &+ Pr[neigbboursetsintersect \mid s+t < n-1] \cdot (1 - Pr[s+t \ge n-1]). \end{aligned} \quad (3)$$

Also, the following holds:

$$Pr[s+t \ge n-1] \le Pr[s \ge (n-1)/2 \vee t \ge (n-1)/2].$$

Continuing, we obtain the following:

$$\begin{aligned} Pr[s+t \ge n-1] &\le Pr[s \ge (n-1)/2 \vee t \ge (n-1)/2] \\ &\le Pr[s \ge (n-1)/2] + Pr[t \ge (n-1)/2] \\ &\le 2(g-1)m^{(g-1)} \sum_{k=\frac{n-1}{2}}^{n-1} k^{(-g)} \\ &\le 2(g-1)m^{(g-1)} \int_{u=\frac{n-1}{2}-1}^{n-1} u^{-g} du \\ &= 2(g-1)m^{(g-1)} \cdot \frac{2^g n(n-3)^{-g} - 3 \cdot 2^3 (n-3)^{-g} - 2n(n-1)^{-g} + 2(n-1)^{-g}}{2(g-1)} \\ &\le m^{g-1} 2^g n^{-g+1}. \end{aligned} \quad (4)$$

Also, it is easy to see that the following holds:

$$Pr[neigbboursetsintersect \mid s+t < n-1] = \frac{s}{n-1} + \frac{t}{n-1} - \frac{st}{(n-1)^2}. \quad (5)$$

Using (4) and (5) we can rewrite (3) as follows:

$$\begin{aligned} p_{s,t} &= Pr[neigbboursetsintersect \mid s+t<n-1] \\ &- Pr[neigbboursetsintersect \mid s+t<n-1]Pr[s+t\geq n-1] \\ &+ Pr[s+t\geq n-1] \\ &= \frac{s}{n-1}+\frac{t}{n-1}-\frac{st}{(n-1)^2} \\ &- \left[\frac{s}{n-1}+\frac{t}{n-1}-\frac{st}{(n-1)^2}\right]\cdot O(\frac{1}{n^{g-1}}) \\ &+ O(\frac{1}{n^{g-1}}) \\ &= \frac{s}{n-1}+\frac{t}{n-1}-\frac{st}{(n-1)^2}+O(\frac{1}{n^{g-1}})-o\left(\frac{1}{n^2}\right) \end{aligned} \quad (6)$$

from which (2) follows.

We will now compute p_{hyper}.

Theorem 1. For a randomly constructed random graph according to model $G_{n,m,g}$, p_{hyper} is given by the following expression:

$$p_{hyper} = \frac{2m(g-1)}{(n-1)(g-2)} + o(\frac{1}{n}). \quad (7)$$

Proof. Take any two vertices v, w of the graph. Then the probability of being connected can be written, using the law of total probability, as follows (using, also, (2)):

$$\begin{aligned} p_{hyper} &= \sum_{s=m}^{n-1}\sum_{t=m}^{n-1} Pr[v, wadjacent \mid deg(v)=s \wedge deg(w)=t] \\ &\cdot Pr[deg(v)=s \wedge deg(w)=t] \\ &= \sum_{s=m}^{n-1}\sum_{t=m}^{n-1}\left(\frac{s+t}{n-1}-o(\frac{1}{n})\right)(g-1)m^{(g-1)}s^{(-g)}(g-1)m^{(g-1)}t^{(-g)} \\ &= (g-1)^2 m^{2(g-1)}\frac{1}{n-1}\sum_{s=m}^{n-1}\sum_{t=m}^{n-1}(s+t)s^{-g}t^{-g} \\ &- (g-1)^2 m^{2(g-1)}o(\frac{1}{n})\sum_{s=m}^{n-1}\sum_{t=m}^{n-1}s^{-g}t^{-g}. \end{aligned} \quad (8)$$

Using integral approximations, we can see that the second term in the final expression of (8) is $o(\frac{1}{n})$ while the first term is equal to $\frac{2m(g-1)}{(n-1)(g-2)} + o(\frac{1}{n})$, completing the proof.

Given a $G_{n,m,g}$ random network, we can construct the person acquaintance graph by having each of the possible d^2 edges that may exist between two hypernodes that are adjacent in the network acquaintance graph selected uniformly and independently with probability p_{extacq} and by having each of the d^2 edges that may exist between two nodes of the same hypernode selected uniformly and independently with probability p_{intacq}. If no edges are introduced we repeat the edge formation process. We will denote by $G(p_{hyper}, p_{extacq}, p_{intacq}, d, n)$ the generated acquaintance network.

4 Virus Propagation Model

We assume that a worm spreads itself by attaching its malicious code to an email, a file transferred or a URL to an infected link and sending it to all contact addresses it finds on a users' computer. IM contact lists enable users to track the presence status of their contacts. To a worm, an online contact list provides an instant hit-list. Note that most email clients provide an address book which does not reveal any online status of the users thus the propagation is slowed down by the time the user interacts opening his email. A host is infected when the user opens the attached or transferred file or when the client previews it or the exploited vulnerability makes it to execute automatically. According to the theory above we will now refer to the model that the worm uses in order to propagate between the hypernodes and the nodes accordingly to the existing connections. We assume that a worm randomly infects a node v of a hypernode. By exploiting the address book of the node the worm starts to propagate. Initially, all the nodes are susceptible to infection. At step zero a randomly chosen set of nodes becomes infected. Then the infection spreads as follows: at every infection cycle i the nodes that are infected turn into black and start to infect other susceptible nodes by sending infected messages to the hit-list they have. The messages that are been sent by the infected nodes follow the edges of the person acquaintance graph. We assume that all generated messages, infectious or not, are send sequentially, as IM and email servers are receiving and sending messages sequentially. Step i is completed after the i th message is ready to be dispatched. Every user that receives an infected message opens this message with probability p_{openm}. The user's computer becomes infected if there is no updated antivirus program installed at the computer. A susceptible or infected (black) node becomes white (immunized) with probability p_{antv} if an updated antivirus program is installed at the node.

5 Theoretical analysis and model evaluation

We will now analyze theoretically the proposed worm propagation model by applying the differential equations method.

Definition 1 *A function f satisfies a Lipschitz condition on $D \subset \Re^j$ if there exists some constant $L>0$ such that*

$$|f(u_1,\ldots,u_j)-f(v_1,\ldots,v_j)| \leq L\sum_{i=1}^{j}|u_i-v_i|$$

for all $(u_1,\ldots,u_j)$ and $(v_1,\ldots,v_j)$ in D.

Definition 2. *Given a random variable X depending on n, denoted by $X^{(n)}$, we say that $X^{(n)} = o(f(n))$ always if*

$$\max\{x \mid Pr[X^{(n)}=x] \neq 0\} = o(f(n)).$$

Theorem 2. *Let $Y_i^{(n)}(t)$, $n \geq 1$, be a sequence of real-valued*

random variables, $1 \le i \le k$ *for some fixed* k *, such that for all* i *, all* t *and all* n *,* $|Y_i^{(n)}(t)| \le Bn$ *(* $n > 0$ *) for some constant* B *. Let* $H(t)$ *be the history of the sequence, i.e. the matrix* $\langle \bar{Y}(0), \ldots, \bar{Y}(t) \rangle$ *,*

where $\bar{Y}(t) = (Y_1^{(n)}(t), \ldots, Y_k^{(n)}(t))$ *.*

Let $I = \{(y_1, \ldots, y_k) : Pr[\bar{Y}(0) = (y_1 n, \ldots, y_k n)] \neq 0 \text{ for some } n\}$ *. Let D be some bounded connected open set containing the intersection of* $\{(s, y_1, \ldots, y_k) : s \ge 0\}$ *with a neighborhood of* $\{(t/n, y_1, \ldots, y_k) : (y_1, \ldots, y_k) \in I\}$ *. (That is, after taking a ball around the set I, D is required to contain the part of the ball in the half-space corresponding to* $s = t/n$ *,* $s \ge 0$ *.)*

Let $f_i : \Re^{k+1} \to \Re, 1 \le i \le k$ *,and suppose that for some* $m = m(n)$ *,*

(i) for all i *and uniformly over all* $t < m$ *, always*

$$E[Y_i^{(n)}(t+1) - Y_i^{(n)}(t) \mid H(t)] = f_i(t/n, Y_0^{(n)}(t)/n, \ldots, Y_k^{(n)}(t)/n) + o(1),$$

(ii) for all i *and uniformly over all* $t < m$ *,*

$$Pr[|Y_i^{(n)}(t+1) - Y_i^{(n)}(t)| > n^{1/5}] = o(n^{-3}), always,$$

(iii) for each i *, the function* f_i *is continuous and satisfies a Lipschitz condition on D.*

Then

(a) for $(0, z^{(0)}, \ldots, z^{(k)}) \in D$ *the system of differential equations*

$$\frac{dz_i}{ds} = f_i(s, z_0, \ldots, z_k), 1 \le i \le k$$

has a unique solution in D for $z_i : \Re \to \Re$ *passing through* $z_i(0) = \hat{z}^{(i)}, 1 \le i \le k$ *, and which extends to points arbitrarily close to the boundary of* D *;*

(b) almost surely $Y_i^{(n)}(t) = z_i(t/n) \cdot n + o(n)$*, uniformly for* $0 \le t \le \min\{\sigma, m\}$ *and for each* i *, where* $z_i(s)$ *is the solution in (a) with* $\hat{z}^{(i)} = Y_i^{(n)}(0)/n$ *, and* $\sigma = \sigma(n)$ *is the supremum of those* s *to which the solution can be extended.*

This theorem says is that if we have a number of co-evolving discrete random variables (associated with some discrete random process) that satisfy a Lipschitz condition and their expected fluctuation at each time step is known, then the value of these variables at each time step can be approximated using the solution of a system of differential equations. Furthermore, the system of differential equations results directly from the expressions for the expected fluctuation of the random variables describing the random process.

In [6] we have analyzed the process that involves two jointly evolving random variables: $B(i)$, the number of black nodes at step i of the worm spread process, and $W(i)$, the number of immune nodes at step i of the process. According to our discussion in Section, step i is completed after the ith message (according to some global ordering) is ready to be dispatched from one node (i.e. personal computer) belonging to some hypernode (a domain). For our model, the following holds:

Theorem 3. *Let* $p_{hyper} = \frac{c}{n}$*, with* $c = \frac{2m(g-1)}{g-2}$*,* p_{extacq} *a constant independent of* n *and* t*,* $p_{antiv}(n,t) = \frac{g(t)}{n}$*,* p_{intacq} *a constant independent of* n, t *, and* $p_{openm}(t) = f(t)$*. Let, also,*

$$h = [c\, p_{extacq}\, w(0) - p_{extacq}\, c + p_{intacq}\, w(0) - p_{intacq}] \qquad (9)$$

$$u(x) = e^{\left[-hd \int_0^x f(y) e^{\left(-d \int_0^y g(z)\, dz \right)} dx \right]}.$$

Then the solution to the system of differential equations given in [6] is the following:

$$w(t) = \exp\left[d \int_0^t g(z)\, dz + w(0) - 1 \right] \exp\left[-d \int_0^t g(z)\, dz \right]$$

$$b(t) = b(0) \frac{u(t)}{b(0) d (p_{extacq}\, c + p_{intacq}) \int_0^t u(s) f(s)\, ds + 1}.$$

We will now plot these solutions, for various values of the parameters, in order to see the interaction between the numbers of black and white nodes. In our model, the main parameters that affect this interaction are p_{antv} , p_{openm} and p_{hyper} . In addition, as we have already argued above, the probabilities p_{antv} and p_{openm} should depend on the time parameter. In particular, we have set $p_{antv} = \frac{1}{1 + a e^{(-\beta t + \gamma)}}$, $p_{openm} = \frac{0.9}{\delta + e^{(+\zeta t - \theta)}}$, where $a, \beta, \gamma, \delta, \zeta$ and θ are constants. We also have $p_{hyper} = \frac{c}{n}$, with $c = 2m(g - 1/(g-2)$ where $2 < g < 3$ is the degree exponent and m is the minimum degree in the network. From p_{hyper} we can observe that when g is decreasing, the average degree (how many neighbors on average a hypernode has) is increasing and is equal to $m(g-1)/(g-2)$, and vice versa. When the average degree is large then the worm outbreak's much faster, and this is depicted to the following figures. The chosen function for p_{antv} is, initially, monotonicaly increasing with a small rate while afterwards it increases at a faster rate. This has

the interpretation that after a new worm has been analyzed, as time goes by, more people start downloading and installing defense software against it while at first only few antivirus installations take place. The chosen function for p_{openm} has the opposite behaviour. At first, this function is monotonicaly decreasing at a slow rate, reflecting the fact that people tend to open emails without a second thought. Then the function is decreasing with a faster rate reflecting the fact that information about the worm becomes available and people become more cautious with opening their email.

In the figures that follow, we plot the percentage of black and white nodes, as a function of time, for $d = 20$ (hypernode or local network size). We can see the effect that p_{antv}, p_{openm} and p_{hyper} have to the relative size of the populations of white and black nodes for the value of d, all other parameters being fixed. First we observe that the effect of the worm in Figure 3 is more severe than in the Figures 1, 2 and 4 This is due to the fact that in Figure 3 we have a large p_{openm} and g and a small p_{antv}, which makes the worm's outbreak much faster. With regard to the effect of the antivirus installation rate as well as the users' opening mail easiness, in Figures 1 and 2 we have higher installation rate and less easiness, in comparison with Figures 3 and 4. One can also tune the other parameters of the model and, thus, construct ``what-if'' scenarios for various worm propagation patterns and and acquaintance network structures.

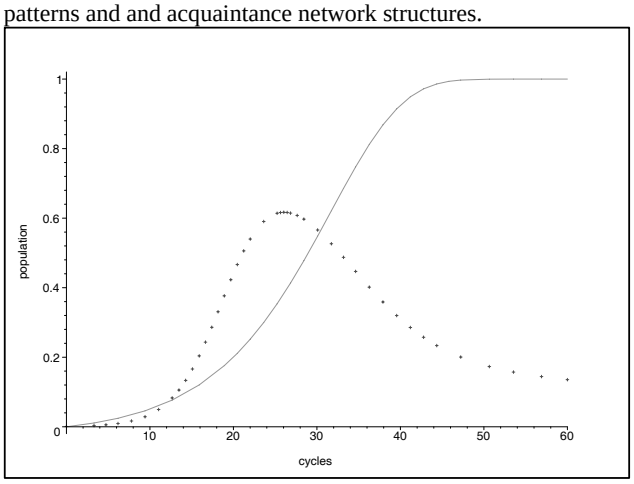

Fig. 1 d=20, large p_{antv}, small p_{openm} and large g,(black nodes dotted, white nodes continuous)

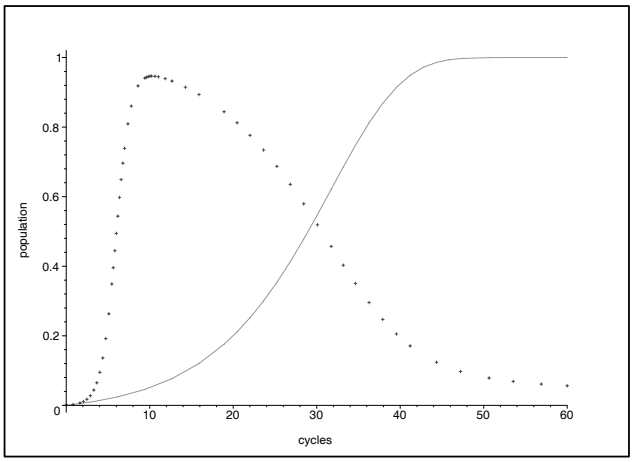

Fig. 2 d=20, large p_{antv}, small p_{openm} and small g,(black nodes dotted, white nodes continuous)

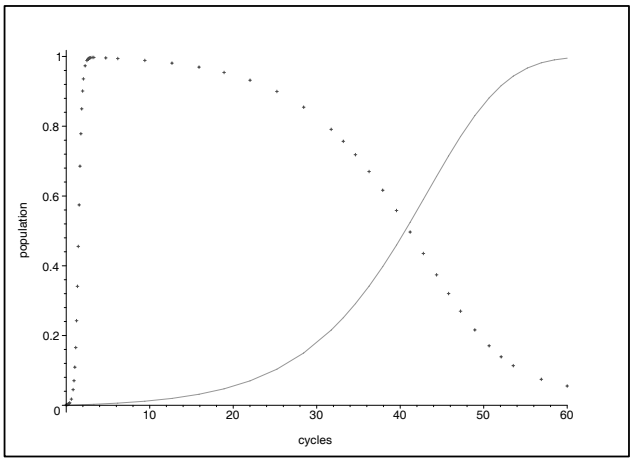

Fig 3 d=20, small p_{antv}, large p_{openm} and small g (black nodes dotted, white nodes continuous)

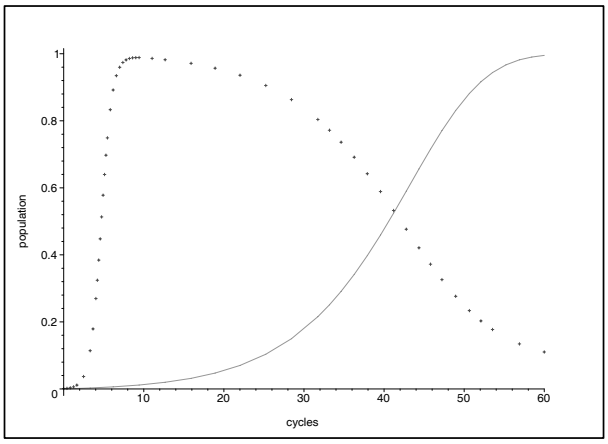

Fig 4 d=20, small p_{antv}, large p_{openm} and large g (black nodes dotted, white nodes continuous)

6 Conclusions

The availability of reliable models of computer worms propagation would prove useful in a number of ways, in order both to predict future threats and to develop new containment measures. In this paper we have proposed a model for users' acquaintance profiles based on a scale free network infrastructure, as they result from their email address books or their IM communication habits. This model can be used for the study of worm propagation as a function of the antivirus installation rate the users' easiness in opening their email attachments as well as the probability in which the hypernodes are connected. We showed that the theoretical analysis of this model leads to a system of differential equations that result from the application of Wormald's theorem to the analysis of the expected fluctuations of infected as well as immunized nodes. These equations can be analytically solved, offering a practical means of conducting ``what-if'' scenarios by tuning the parameters of the model. We believe that our model can be used as a basis for extensions by including other factors which may affect virus propagation (e.g. network link speed) having, at the same time, a straightforward theoretical analysis with the aid of Wormald's theorem.

7 References

[1] Albert R, Jeong H, Barabasi A, 1999, ``The diameter of the World Wide Web,'' Nature 401:130-131.

[2] Albert R, Jeong H, and Barabasi A.L., 2000, ``Attack and error tolerance in complex networks,'' Nature 406:378.

[3] Faloutsos C, Faloutsos P, Faloutsos M, 1999, ``On power-law relationships of the Internet Topology,'' Computer Communication Review 29:251-260

[4] Huberman B, Adamic L, 1999, ``Growth dynamics of the World Wide Web,'' Nature 401:131-134

[5] Spencer J. and Sacks L., ``On power-Laws in SDH Transport Networks,'' in *IEEE ICC 2003*, May 2003, Anchorage, Alaska, USA.

[6] T.Komninos, Y.C. Stamatiou and G.Vavitsas, ``A worm propagation model based on people's email acquaitance profiles,'' in *Wine 2006*, Patras, Greece.

[7] CERT advisory CA-2004-02.

[8] CERT advisory CA-2001-26 Nimda Worm.

[9] CERT incident note IN-2003-03.

[10] A. Gostev, ``Malware Evolution: January - March 2005,'' Kaspersky Lab Report 4.

[11] IMlogic Threat Center, Symantec Corporation. http://www.imlogic.com/im_threat_center/index.asp

[12] J.O. Kephart and S.R. White, ``Measuring and Modeling Computer Virus Prevalence,'' in *Proc. 1993 IEEE Computer Society Symposium on Research in Security and Privacy*, Oakland, California, 1993.

[13] Staniford S., Paxson V., Weaver N., ``How to own the internet in your spare time,'' in *Proc. of the 11th USENIX Security Symposium (Security '02)*, (2002).

[14] M. Mannan and P. Oorschot,``On Instant Messaging Worms, Analysis and Countermeasures,'' in *Proc. of the 2005 ACM workshop on Rapid malcode (WORM'05)*.

[15] Microsoft, ``How to update your computer with the JPEG processing (GDI+) security update''. http://www.microsoft.com/athome/security/update/bulletins/200409_jpeg_tool.mspx

[16] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, ``Inside the slammer worm,'' *IEEE security and privacy*, **1(4)**,33--39, July 2003.

[17] G.M. Murphy, *Ordinary Differential Equations and their Solutions*, D. Van Nostrand Company Inc., 1960.

[18] Symantec Internet Security Threat Report Trends for January 05-Dec. 05, Vol. VIII and IX, 2005.

[19] C. Wang, J. Knight, and M. Elder, ``On computer viral infection and the effect of immunization,'' in *Proc. of the 16th annual computer security applications conference (ACSAC r00)* , New Orleans, LA, Dec. 2000.

[20] N.C. Wormald, ``The differential equation method for random graph processes and greedy algorithms,'' *Ann. Appl. Probab.* **5**, 1217--1235, 1995.

[21] C.C Zou, W. Gong, and D. Towsley,``Code-red worm propagation modeling and analysis.,'' in *Proc. of the 9th ACM conference on Computer and Communications Security*, ACM Press, pp. 138--147, 2002.

Theodore Komninos received his MSc in Computer and Networks Security (2000) from Department of Computer Engineering and Informatics, School of Engineering, University of Patras, Greece, his MBA (1993) from the Hellenic Management Association. He owns a diploma in Computer Engineering and Informatics (1989) from the University of Patras and a diploma in Civil Engineering (1986) from the

same University. In 1989 he joined RACTI and he is now member of the BoD and Director of Educational Technology Sector, Director of Systems & Networks Support Sector and Director of Networking and Information Systems Security Sector. He is lead auditor and Special Advisor for Information, Systems and Network Security for the Greek Ministry of Education, a member of Network Specialists of the Greek Research Network (GRNet-member of GEANT) and has extensive experience of CSF programs. Mr. Komninos is supervising Postgraduate Diploma Thesis and Master Thesis in the area of Information, Systems & Network Security, Distributed Networking Intrusion Detection Systems and Information Warfare. His research interests include Information, Systems & Network Security, Distributed Networking Intrusion Detection Systems, Information Warfare, Design of innovative environments for Intrusion Detection Systems, and New Technologies in Education. He is also co-author of the book "Strengthening Security of Systems and Networks. Dare the intruders", Greek Letters-CTI Press, 2003 (in Greek) and author of the 4th Chapter titled "Choosing the right computer equipment for secondary schools" in the book "Time is the Judge" (pony translation of Greek title).

Prof. Paul Spirakis (google: Paul Spirakis) born in 1955, obtained his PhD from Harvard University, USA, in 1982. He served as a postdoctoral researcher at Harvard University and as an assistant professor at New York University, (the Courant Institute)**.** He was appointed as a Full Professor in the Department of Computer Science and Engineering of Patras University (Greece) in 1990.

Paul Spirakis was honored several times with international prizes and grants (e.g. NSF), also the top prize of the Greek Mathematics Society. He was acknowledged between the top 50 scientists worldwide in Computer Science with respect to "The best Nurturers in Computer Science Research", published by B. Kumar and Y.N. Srikant, ACM Data Mining, 05. He was appointed as a Distinguished Visiting Scientist of Max Planck Informatik. Paul Spirakis is the Director of the Research Academic Computer Technology Institute (RA.CTI). His research interests include Algorithms and Complexity and interaction of Complexity and Game Theory. He has extensively published in most of the important Computer Science Journals and most of the significant refereed conferences. He has edited various conference proceedings and is currently an Editor of Several Prestigious Journals. He has published two books through Cambridge University Press, and eight books in Greek. Paul Spirakis was the Greek National Representative in the Information Society Research Programme (IST) from January 1999 till June 2002. He was elected unanimously as one of the two vice-Presidents of the Council of the European Association for Theoretical Computer Science (EATCS). He has been a member of ISTAG (Information Society Technologies Advisory Group) a prestigious body of about 40 individuals advising EU for research policy, form January 2003 to January 2005. He consults for the Greek State, the European Union and several major Greek Computing Industries.

Prof. Yannis Stamatiou was born in Volos in 1968. He holds a degree of Computer Engineering & Informatics, from the University of Patras and a Ph.D. from the same department. He is currently an Assistant Professor at the University of Ioannina, Mathematics Department, Greece, and a scientific consultant on security and cryptography issues of the Research and Academic Computer Technology Institute (RACTI). His scientific interests lie in the fields of security and cryptography as well as the study of threshold phenomena arising in computationally intractable problems. He is a member of ACM and IEEE.

George Vavitsas was born in Trikala in 1981. He holds a degree of Computer Engineering & Informatics, from the University of Patras. He is cullently obtaining his MSc in Computer and Networks Security (2005) at the Department of Computer Engineering and Informatics, School of Engineering, University of Patras. His research interests include Information Systems & Network Security, worm propagation models, algorithms, Game Theory, Information Warfare. He is now member of Networking and Information Systems Security Sector of the Research and Academic Computer Technology Institute (RACTI).