Providing a model based on Poisson distribution for malware propagation assessment in peer-to-peer networks

Shadi Haghi and Mahdi Mollamotalebi*

Department of Computer,
Buinzahra Branch,
Islamic Azad University,
Buinzahra, Iran
Email: Haghi@buiniau.ac.ir
Email: motalebi@qiau.ac
*Corresponding author

Abstract: The use of peer-to-peer networks has increased dramatically in recent years in applications infrastructure such as file sharing, gaming, instant messaging and content distribution. These networks suffered from the problem of large-scale incompatibilities, which was organised by the super peer to overcome this problem. So far, several models have been presented to assess the behaviour of worm propagation in peer-to-peer network, but there has not been any effective review of how the worm propagates in super-peer networks. This paper presents a framework for modelling the connectivity of the super-peer network and then examines the behaviour of active worm propagation based on epidemic models on this framework. The results indicate that the worm propagate more rapidly in super-peer networks. Also, the results of the implementation indicate that the model presented in this study has achieved a significant improvement in reducing the rate of propagation of worms compared to previous work.

Keywords: super-peer network; flat peer-to-peer network; malware; active worm; modelling worm propagation behaviour.

Reference to this paper should be made as follows: Haghi, S. and Mollamotalebi, M. (2022) 'Providing a model based on Poisson distribution for malware propagation assessment in peer-to-peer networks', *Int. J. Internet Technology and Secured Transactions*, Vol. 12, No. 1, pp.1–26.

Biographical notes: Shadi Haghi received her BSc degree in Computer Engineering from Tabarestan University, Iran, in 2011; and MSc degree in Computer Engineering from Islamic Azad University of Buinzahra, Iran, in 2016. She is experienced as a specialist with a demonstrated history of working in the computer networking industry. She is skilled in network monitoring tools, Linux server, Solaris, and DevOps engineering. She is now working in FANAP co. as DevOps Engineer.

Mahdi Mollamotalebi received his BSc degree in Computer Engineering from Islamic Azad University of Qazvin (QIAU), Iran, in 1999; MSc degree in Computer Engineering from Islamic Azad University of Arak, Iran, in 2004; and PhD degree in Computer Science from Universiti Teknologi Malaysia (UTM), Malaysia, in 2013. He is currently a Senior Lecturer and Researcher in

the Islamic Azad University of Buinzahra, Iran. His research interests are computer network management, computer security, internet protocols, grid resource discovery, cloud computing, web search engine, and smart home.

1 Introduction

Peer-to-peer (P2P) networks present a distributed environment to share the resources (e.g., files) between different users. Nowadays, such networks are widely used in various services such as voice over Internet, instant messaging, and file sharing. Each member in P2P plays both the role of a server and a client. The members exchange services and information without centralised controls. The users and resources can join or leave the P2P network dynamically.

P2P network is assumed as a fully distributed environment in which all nodes are the same in terms of their hardware/software capabilities and they share resources (e.g., processing, storage, and communication capacities). The shared resources are used to provide the network services without the high management overheads of client-server structure. They are increasingly popular. As an example, the Kazaa P2P software has tens of millions downloads. The eDonkey2000 network software has over 2 million users connected at any given time, and file-transferring tool BitTorrent has more than 10 millions users.

Because of the advantages of a peer-to-peer network, many Internet users use them to distribute information. Today, such networks have become popular with different applications (such as voice over the internet[1], instant messaging, file sharing, etc.). These networks are exposed to many security threats from worms on the Internet. Worms causes stealing of information, consuming bandwidth, unwanted traffic, and occupying host machine resources.

The P2P networks are widely used by end users. However, they face with security risks because they have an ideal venue for new types of worms that rose from vulnerabilities on the P2P hosts. The worms identify new victims by following P2P neighbour information on infected hosts. Some of the worms are different from the well-known worms in behaviour. The worms propagate rapidly in P2P networks because they do not need to probe unused IP addresses. They also have not high rates of failed connections and they can blend into the normal traffic patterns of the P2P network.

According to scan approach, the worms are categorised as scanning worms and non-scanning worms. The notorious Internet worms usually perform a random scanning to find the potential victims. P2P worms tend to use list of neighbours to choose the potential victims and so, they are non-scanning. According to various attack approaches, three types of non-scanning worms can be identified (Feng et al., 2008):

1 passive worm: it hides itself in malicious files and trick users to download and open them

2 reactive worm: it propagates with legitimate network activities

3 active worm: it automatically connects to and infects the known peers using topological information.

The lack of abnormal network behaviour causes that P2P worms become highly risky threat. Most of the defence mechanisms against scanning worms are not effective. There are millions of P2P networks' subscribers and therefore, worms are able to compromise a significant fraction of the internet population.

Peer-to-peer network worms are divided into two categories: peer-to-peer passive worms, and peer-to-peer active worms. Passive worms hide themselves in popular peer-to-peer files and create a number of copies of their own with different names in the infected user's shared folders. When a user executes or downloads a file, the worm spreads out and infects users, and encourages other members to download these copies. The process of spreading and infecting the new victim who has copied the file continues (Thommes and Coates, 2007). These types of worms are able to infect only by exploiting network members.

On the other hand, active worms automatically propagate themselves through common vulnerabilities in members of peer-to-peer network (Chen and Gray, 2007). Such worms, in addition to the common vulnerabilities of peer-to-peer network members also use neighbouring information[2] in infected members for trapping their victims. If one can accurately identify the vulnerabilities of peer-to-peer network members and the way worms propagate in these types of networks, the threats of network worms can be reduced greatly.

Peer-to-peer systems have the some strength as removing overhead from the primary server and distributing it among all clients, although they could potentially caused by flood-messaging. To cope with such limitations, the super-peer network architecture was introduced using the heterogeneity of members, in which heavier responsibilities were assigned to members with more resources (such as bandwidth, processing power, storage space, and communication capacity), which is called super-peer. Super-peer is the most efficient peer-to-peer network and forms the architecture of peer-to-peer systems such as kazaA, gnutella and Skype (Taheri, 2013).

Considering the suitability of the substrate for such networks, for the propagation of the worm and the related attacks, it is essential to provide an efficient model for modelling the propagation of the worm and its behaviour in order to identify its risks, and also to develop new solutions for the discovery of the worm and to deal with it. So far, several models have been presented to assess the behaviour of worms in peer-to-peer networks. However, there has not yet been a comprehensive study of how worms propagate in super-peer networks. In this research, a framework is proposed for modelling the connectivity of super-peer, and then active worm behaviour is presented based on the epidemic model. The proposed propagation model uses the Poisson distribution, which is the prevalent distribution in the modelling of dynamic malicious activities.

The structure of this article in the following sections is as follows: In the second section, an overview of previous work related to the issue of the propagation of the worm in peer-to-peer networks is presented. In the third chapter, a model for the super-peer network is proposed by integration of the Poisson degree distribution. In Section 4, the proposed model is implemented under various scenarios and is evaluated. Finally, Section 5 concludes the paper.

2 Related work

Due to the large volume of information exchanges on peer-to-peer networks, malware can be multiplied rapidly and causes heavy software losses, such as some of Microsoft's software weaknesses, and hardware problem such as sudden and unforeseen filling of memory resources. Therefore, recognising these networks and their structure as well as types of malware and their function can be effective in understanding the behaviour of worms. Malware is a program that is deliberately designed to perform some unauthorised operations. Meanwhile, worms are considered as one of the most important security threats for a peer-to-peer network. Worm is a program that propagates itself at the network level by exploiting security or policy issues in a service that is utilised. Worms use scanning mechanisms to discover new victims for infection.

Before launching an attack, the worm must search for the vulnerabilities of the target hosts. The scanning strategy can accelerate the propagation of the worm and it is divided into the following categories (Rajesh et al., 2015; Kumar and Chen, 2008):

1 Selective random scan: instead of scanning the entire address space, the worm randomly selects some address set as the target address space. The list of selected addresses is obtained randomly from the entire list of routes.

2 Sequential scanning: the worms are randomly selected in the infected host of an IP address and the rest of the addresses are selected from the neighbours of this address. Worms usually select IP addresses in the network to which they belong. The disadvantage of this method is the repetition of the scan, which blocks the network.

3 Scroll through the hit list: creates a hit list which includes those hosts that have the potential to become infected and, after creating the list, begin infecting its members. This list is prepared by scanning a small part of the internet or by obtaining the entire list of databases distributed by the search.

4 Routing scan: is a type of scanning strategy in which network worms selectively scan the IP address space based on route information on the network.

5 DNS scanning: obtains a table of target addresses from DNS servers. This scanning method has problems such as the difficulty of obtaining the total URL table from the domain name server records and the slow release due to the very large address given by the worm to carry the database.

6 Divide and conquer: worms work together to quickly search for susceptible hosts. The worms send part of the address list of susceptible nodes to other infected nodes to scan that list.

Distributed network architecture is called peer-to-peer, if participants share their hardware resources (such as processing power, storage space, and communication capacity). These resources can be used directly by other peers. In these networks, operations such as search have been inefficient in generating large amounts of duplicate packets on the network, and due to the high bandwidth utilisation, the whole network is drowned in query messages, and this issue limits scalability.

Due to the problem of high message load in peer-to-peer network, a specific type of these networks is provided to better manage messages distribution. In this type of peer-to-peer networks known as super-peer networks, members with more capacity,

capability and stability are selected as super peers, and the remainder will be regular members.

The connection of these super peers to each other will form a top layer in the network hierarchy. In these types of networks, each super peer acts as a server in its own set of members, which forms the lower layer of the network, and sends messages across the top layer of the routing network. The super peer answers the queries sent by the regular members (Yang and Garcia-Molina, 2003; Pyun and Reeves, 2005; Meng et al., 2008). Because super peers act as centralised servers for their members, they can manage queries with more efficiency than their members. Also, due to the fact that in super-peer networks, some of the nodes have more connections than other ones, such networks follow the degree of power law distribution.

2.1 Control methods for worm's propagation in peer-to-peer networks

A detailed propagation model describes worm behaviour properly and helps identifying its propagation control pathways. Viral spreading patterns in contagious diseases can be used to model worm diffusion. The first and simplest model provided for the propagation of the worm is the simple epidemic model (Zheng, 2008). However, when the host is infected with the worm, it always remains in the infected state. As regards that this model did not meet the needs for human interaction, such as security, patch, etc. other models such as the Kermack-Mckendrick (Ganguly and Deutsch, 2004), the two-factor model (Zou et al., 2006), and the SEIR Model (Carlyle, 2010) and time-delayed model were presented (Yao et al., 2014). The above models are used on the Internet, and some of them are described below.

The Kermack-Mckendrick (Ganguly and Deutsch, 2004; Kienzle and Elder, 2006) model is developed by a simple epidemiological model considering the repair process. When the host recovers, it becomes immune to infection. This model is not perfect for describing the propagation of worms, since it does not include human interactions, i.e., the immunisation, repair of vulnerable/infected hosts, and filtering worms; only repairing infected hosts is taken. Also, this model assumes that the rate of the infection is constant, while this assumption does not apply to worms that are propagated by large volumes (such as the code red) (Newman, 2004).

The two-factor propagation model (Zou et al., 2006) completes the Kermack-Mckendrick model, so that its repairing measures are not limited to infected stations. It also tries to measures to repair them before infection. On the other hand, the processing load and message exchange in this model increase the potential for monitoring stations and identifying infected nodes.

An improved model (quarantine propagation model) (Carlyle, 2010) was also developed from Kermack-Mckendrick, in which infected hosts are also flagged, although these hosts are only exposed to infection and cannot transmit infection to other hosts. The quarantine propagation model uses the strategy of quarantine infected hosts to control the propagation of worms. This is especially useful for controlling new worms whose propagation behaviour is unknown. Prior to providing this model, if hosts were infected, measures such as interrupting network connections, controlling by antivirus, or imposing some restrictions on the firewall were done to eliminate worms. Due to the delay caused by such control processes, quarantining can eliminate the occurrence of Hopf[3] branches (which leads the number of infected hosts to be unpredictable and the propagation of worms to be out of control).

The propagation model for worms in peer-to-peer networks is similar to their propagation on the Internet. In studying the behaviour of worm propagation, it is necessary to consider the specific features of such networks. The worms are categorised into two active and passive categories in terms of how they are propagating in the network. The propagation of passive worms is related to malicious file execution and is not automated. Therefore, it is not sensitive to network topology, and in modelling, there is no significant difference between peer-to-peer networks and other types of networks structurally. But active worms are propagated automatically and cleverly.

Finding the target and victim of an attack is one of the most important activities of active worms, depending on how the worm implements and the search mechanisms. Hence, awareness of network topology and its structure is essential for active worms and how their propagation behaviour works. In the following, we introduce the active worm propagation models in peer-to-peer networks, such as those found in the worm propagation techniques in the Internet.

Feng et al. (2010) have reviewed the effect of similar peer-to-peer network topology in the propagation of the worm by providing a model called SIS (susceptible, infected, and susceptible). In this model, members can be only susceptible and infected. At each step of time, any susceptible member will be infected with a certain probability (β) if it is attached to an infected member.

Also, infected members with a certain probability (γ) will recover and return to the susceptible state. Therefore, the SIS model does not consider the possibility of repairing members due to their immunisation or deletion due to death. If the effective infection rate reaches a certain threshold, it means that the worm has infected all peer-to-peer network members. Also, when this rate is less than the threshold, it means that peer-to-peer network members are susceptible at some time. The shortcoming of this model is that it does not consider the possibility of repairing members due to immunisation or deletion due to death.

Hua et al. (2010) examined the propagation model of active worm in an unstructured network, a new propagation model of active worm in a peer-to-peer unstructured network. The effect of the parameters of peer-to-peer systems, such as the velocity of propagation of active worms was studied in this model. For this purpose, the distribution degree of unstructured peer-to-peer networks was modelled according to the power law distribution. In this model, each member is exposed to three susceptible, infected, and recovered/deleted states.

The behaviour of worm propagation in unstructured peer-to-peer networks is more appropriate than the base model (Kienzle and Elder, 2006). The prevalence of infection in this model is more than the calculated level of the theory in the basic model, which is due to the high rate of propagation in the proposed model compared to the basic model. On the other hand, the probability of repair in both models is the same. Based on the results of this model, the smaller the size of the system, the faster the release of the worm, and the higher the probability of repair, the speed of the propagation of the worm is lower. Also, the sensitivity of this model to the hit list (the list of vulnerable hosts suitable for attack) has shown that when the size of the hit list is constant, the higher the degree of the primitive infected hosts is larger, the higher the propagation velocity.

In addition, nodes with higher connections are infected earlier. When the size of the hit list is variable and the initial degree is constant, the larger the hit list size is, the higher the propagation speed. Two-factor propagation model in peer-to-peer networks (Zhang et al., 2010) is a generalised two-factor of internet worm model. This model assumes that

all hosts are connected to a peer-to-peer network and that as soon as a host is infected; all its neighbours are immediately attacked.

But the host that was previously infected is not attacked again; in other words, a host cannot be attacked several times. Since the worm's propagation process is very fast, configurations changes for peer-to-peer networks are considered constant. In this model, the number of infected hosts is reduced with the propagation of worms and with the recovering of members. The results of further studies in this model indicate that worm propagation in peer-to-peer networks is much faster than other networks, and therefore, worms in these networks are considered a bigger threat. Also, the infection rate of the Kermack-Mckendrick model is more than this model, since there is no direct transfer from the susceptible to the recovered state.

Luo et al. (2011) modelled the behaviour of worm propagation in peer-to-peer networks, with the dynamical characteristics of the members. They showed that the dynamic property of peer-to-peer networks (such as frequent logging of members and resource requests) reduces the rate of worm propagation. This is because that the duration of the availability of a member is random and is determined by the user's behaviour. Therefore, the member who has been identified for infection is no longer available if he leaves the network. When the worms increase the number of simultaneous attacks, the speed of propagation increases too. Also, increasing the number of member neighbours (degree of member) helps to identify valid network members more easily. This increases the number of attack targets and the rate of worm propagation. When the total number of network's member increases, worm's propagation becomes slower; as worms need more time to identify vulnerable members of the attack.

Tang et al. (2014) modelled worm propagation behaviour in peer-to-peer networks, with dynamical properties of members based on randomised quarantine and regular immunisation and by using a random scan strategy analysed the parameters affecting the rate of active worm propagation. Peer-to-peer network members in this model can at any time be in one of the six states such as susceptible, latent, infected, quarantined, safe, and offline. The host in susceptible state is vulnerable by worm's attack, but it has not yet downloaded a worm file. In the latent state, the attended host downloaded the file containing the worm, but has not yet executed; such host is still not exposed to worms and the worm does not have the ability to spill over to other hosts. If node executes a worm file, it will change to the infected state.

Once the node in the infected state is detected by the monitoring software, it is quarantined. When an online node is modified by security software, it becomes safe against active worms. Nodes that leave the peer-to-peer network will be in offline state. The results of investigating the effect of scanning rate on the propagation speed of active worms in this model indicate that the rate of scan become higher, the infection rate will peak sooner.

Also, the lower the detection rate of the monitoring software, the infection rate become higher, and as more worms are detected and quarantined by monitoring software, there is fewer infection rates, respectively. Meanwhile, the higher the number of offline nodes than online nodes, the infection rates will be higher at the initial site of the worm propagation. If the online immune response is become greater, the infection rate become lower.

Regarding the models listed above, active worms are the most important threat for peer-to-peer networks. The results of all models and the impact of parameters such as simultaneous attacks, scan rate, online immune response, and monitoring rates can be

seen that worms are propagated on such networks, much faster than the Internet. Table 1 summarises the features of the reviewed worm propagation models.

Table 1 The summary of worm propagation models

No.	*Model*	*Features*
1	Simple epidemic model	The simplest propagation model; all hosts are either in susceptible for infection or infected states.
2	SIS	Appropriate for the propagation of worms on homogeneous networks
3	Kermack-Mckendrick	When the host is repaired, it becomes immune; not appropriate for modelling the propagation of Internet worms; assumes that the rate of infection is constant.
4	Two-factor model	Is an expansion of Kermack-Mckendrick model; many of original infected nodes are repaired, restarted, or filtered, and subsequently stop the infection.
5	SEIR	An expansion of SIR model by adding the 'exposed' condition to model the period in which the host is exposed to the worm and is infected but it cannot propagate the worm.
6	Distribution model with quarantine and time delay	It has been suggested to detect new worms that do not behave normally. infected hosts are identified by quarantine, and infected and susceptible nodes are detected.
7	Active P2P worms propagation	Showed that the dynamic characteristics of P2P networks reduce the rate of worm propagation. When a member leaves the network, he will no longer be available for infection.
8	Quarantine-based worm propagation in dynamic P2P networks	It is based on random quarantine and regular immunisation. This model shows that the higher the scan rate, the infection rate reaches sooner to the peak point.

3 Presentation of a model for super-peer network based on Poisson's degree distribution

Different types of networks can be modelled using probabilistic distributions. Assuming that P(k) is probable that the random node chosen has a degree k, the distribution degree P(k) indicates the network connectivity. In this section, we modelled a super-peer network using the Poisson distribution. In this model, the relation between the super peers approximates the ER[4] (Jesi et al., 2007; Albert and Barabási, 2005) graph. The average degree of super peers is larger than normal members. In mathematical terms, if r is the fraction of regular members on the network and the rest of the member are super peers, then the distribution degree of the network will be as follows:

$$p(k) = r^5 P(k_{or}) + (1-r) P(k_{sp}) \tag{1}$$

$p(k_{sp})$ is the distribution degree of the super peer and $p(k_{or})$ is the distribution degree of the regular member, and in (2) and (3) are expressed as follows:

$$P(k_{sp}) = \frac{e^{-k_{sp}} \lambda_{sp}^{k_{sp}}}{k_{sp}!} \tag{2}$$

λ_{sp} which is the average degree of the super peer and k_{sp} is super peer degree.

$$P(k_{or}) = \frac{e^{-\lambda_{or}} \lambda_{or}^{k_{or}}}{k_{or}!} \tag{3}$$

$$\lambda_{or} \ll \lambda_{sp}$$

λ_{or} is the average degree of the normal member and k_{or} is the degree of the normal member. Then the average degree of the entire network is obtained as follows:

$$\lambda^{6}_{t} = r\lambda_{or} + (1-r)\lambda_{sp} \tag{4}$$

The parameters required to analyse the proposed model of this paper are shown in Table 2.

Table 2 Parameters used in super-peer networks modelling

	Parameter	*Description*
1	N	The total number of members in the super-peer network
2	N_{sp}	The number of super peers in the network
3	S(t)	The number of susceptible member at time t
4	I(t)	The number of infected members at time t
5	R(t)	The number of members recovered from the infected population at time t
6	J(t)	The number of infected members at time t, J(t) includes both infected hosts as well as hosts previously infected and found to be immune before t. In other words, J(t) = I(t) + R(t)
7	Q(t)	The number of recovered members from susceptible at time t
8	λ_{sp}	average super peer degree
9	λ_{or}	Average regular member degree
10	S	Worm scan rate per second (the number of members that are simultaneously checked by the worm)
11	ω	The number of members in the list of active super peer, which is exchanged between super peer members of the network.
12	p	The average probability of a worm propagate/the probability a member is susceptible.
13	d_i	Degree member of i
14	Inf(t)	Collection of infected nodes at time t
15	γ	Rate of recover of infected member
16	u	The rate of recovered member of susceptible node
17	β	Attack/transmission probability on the communication link
18	δ	The recovered member who has been infected once again

In this research, a topology has been proposed for the worm propagation of a super-peer network accomplished with Poisson distribution. It is implemented by an unconventional simple graph stored in a neighbourhood matrix. Each node was assigned a degree based on the Poisson distribution with the specified values for each of the parameters of the super peer as well as the average degree of super peer and average degree of regular member. Then the relationship between the members and the neighbouring matrix of the members was formed. The pseudo-code for assignment of degrees to members based on

the Poisson distribution model and the creation of a neighbourhood matrix is shown in Algorithm 1.

Algorithms 1 The pseudo-code for assignment of degrees to members based on the Poisson distribution and the creation of a neighbourhood matrix

```
Input: λsp, Nsp, λor, 'Sup-Sup' and 'Sup-Ord' percent
Output: Super-peer network emerged by joining of nodes, adjacency matrix, Ord_No
Foreach super-peeri do
    degsp = generation of random Poisson degree according to λsp
    degsp-sp = degsp * %Sup_Ord
    degsp-ord = degsp – degsp-sp
    add super-i, degsp, degsp-sp, and degsp-ord to adjacency matrix
End
Foreach ordinary-peeri do
    degord = generation of random Poisson degree according to λor
    add ordinary-peer, degord to adjacency matrix
End
Foreach nodei in adjacency matrix do
    If (nodei is super-peer) then
        find degsp-sp random super-peers as neighbours
        add super-peer neighbours to Super-peeri in adjacency matrix
        find degsp-ord random ordinary-peers as children
        add children to Super-peeri in adjacency matrix
    End
    Else
        find degord random ordinary-peers as neighbours
        add ordinary-peer neighbours to ordinary-peeri in adjacency matrix
    End
End
```

The members of the super-peer network constantly exchange a list of super peers in the network. The permanent exchange of the super peers list causes the nodes in the network always have a new list of active and existing super peers in the system. The super peers that left the system do not appear in this list, and members are notified of their absence. The pseudo-code of assignment for active super peer is shown in Algorithm 2.

Algorithm 2 Assignment of active super-peer network

```
Input: N as total number of nodes, ω: size of Sp Refresh List, adjacency matrix
Output: SP Refresh matrix
Foreach super-peeri in adjacency matrix do
    SpRefList = find ω active super-peers
    While (i <= degsp-sp i) do
        neighboursp i = fetch neighbour of super-peeri from adjacency matrix
```

```
        assign SpRefList to neighbour_sp i
    End
    While (j <=deg_sp-ord i) do
        child_sp i = fetch child of Super-peer_i from adjacency matrix
        assign SpRefList to child_spi
    End
End
```

In this research, the super-peer network is static, and all the super peers remain in the system from the beginning of the simulation to the end, and their number does not change. Therefore, the above list is assigned only once to each member when creating the neighbourhood matrix at the beginning of the simulation. The size of this list will be examined as an effective parameter in the velocity of the worm propagation.

Security measures and closure of security holes often occur slowly due to the lack of awareness of members of the existence of worms in the network and their threats. Hence, the constant immunisation rate has not been used, and as more members become infected, the level of awareness of members is higher and, as a result, the immunisation rate increases. To determine the probability of immunisation, the ratio of infected members to the total number of members in the network is calculated. When the degree of prevalence is less than the constant value of C = 1% (Taheri, 2013), no action will be taken to secure the network. In other words, when the number of infected nodes exceeds 1%, infected members are recovered with probability γ and will never be infected. Algorithm 3 shows the pseudo-control code for controlling the worm propagation in the super-peer network in the proposed model.

Algorithm 3 Pseudo-code for the worm propagation in the super-peer network

```
Input: cycleNo(time unit), S, N, u, c as threshold of immunisation, I0, adjacency matrix
Output: number of infected and immunised nodes
Select I0 nodes as initial infected nodes randomly
Initialise status of all nodes as susceptible except I0 initial infected nodes in status matrix
Foreach cycleNo do
    InfectedNodes = find infected nodes in status matrix
Foreach InfectedNodes_i do
    Neighbours = find all neighbours of infected node in adjacency matrix
    While (j <= S) do
        if (Neighbour_s is susceptible ) then
            change status of Neighbour_s to infected
        End
    End
End
threshold = total number of infected nodes/N
If (threshold >= C) then
    RemovedNodes = Select randomly some of infected nodes proportional to k
    Foreach RemovedNodes_i do
```

```
        Change the status of RemovedNodes_i from infected to removed in statusMatrix
    End
End
If (Number of susceptible nodes > 0)
    Calculate (J(t));
    IsolatedNodes = Select randomly some of susceptible nodes proportional to u and J(t)
        Foreach IsolatedNodes_i do
            Change the status of IsolatedNodesi from susceptible to removed in statusMatrix
        End
End
```

Each copy of the worm on an infected node in its lifetime, depending on the rate of scanning, simultaneously attempts to infect new victims. In the proposed model, the effect of the scanning rate parameter of the worm in the super-peer networks is discussed below. In the simulation of the proposed model, the network includes N_{sp} super peer, N_{or} regular node, and in general N members that communicate with each other based on the topology of the super-peer network. Based on the two-factor propagation model (Zou et al., 2005), hosts are in one of three states as follow:

Susceptible → infected → recovered

The proposed model is based on the two-factor model and when the immunisation is done, the member is placed in a recovered condition. It does not matter that recovered node was in infected state or in susceptible state before. Thus, the transfer of status from any host can be as follows:

State 1: susceptible → infected → recovered

State 2: susceptible → recovered

Since the super-peer network topology graph is considered to be unconventional, λ_{or} (average degree of the normal member) involves connecting a normal member to its super peer and also connecting the regular member with the other regular members. For example, if the degree of a regular member is 4, this member will connect to the super peer and three other regular members in the neighbourhood matrix.

In implementing the proposed model, the following assumptions are considered. The super-peer topology graph and flat peer-to-peer is considered non-directional, (i.e., aij = aji =1). Also, to prevent the loop and to reduce the network load, no node has an edge and no connection to itself (i.e., aii =0).

All members are connected to super-peer network. In other words, each regular member has at least one degree because it has at least one connection with his super peer. This condition ensures that there are no isolated nodes in the network. In addition, every regular member has only one super peer.

When a member is infected, his neighbours are immediately attacked. At the beginning of the simulation, a primary infected member is considered, and the rest of the members are susceptible. Infected member is randomly selected from N network members. The infected member will not change its behaviour if it is again infected by another copy of the worm in other members. This means that duplicate infections do not result in an increase in the number of victims.

In our model, it is assumed that the worm needs only a time unit to complete its infecting process on a susceptible member, and this time is the same for all members of the network, such as super user and the regular ones. In the next section, the simulation results of the proposed model are presented and the worm propagation behaviour is analysed in the super-peer network.

4 Results and evaluation of the proposed model

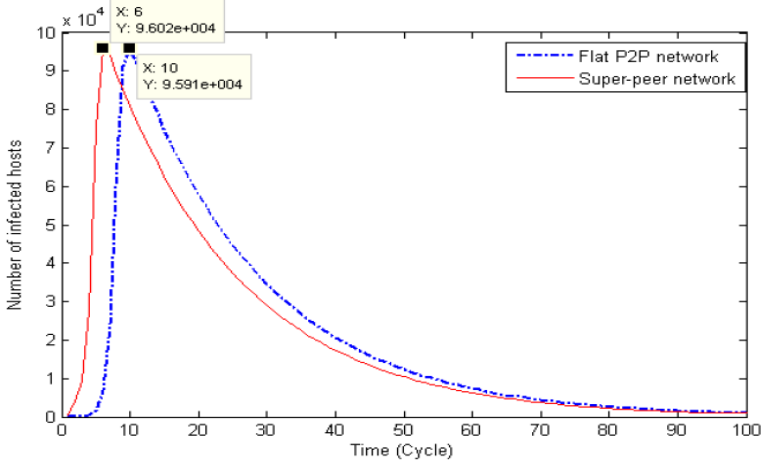
The effect of the parameters such as scanning rate, the average degree of the super peer, the average degree of the regular member, the recovery rate, and the probability of improvement after infection and the probability of transferring on the link on the behaviour of worm propagation has been evaluated. Then, by examining different immunisation policies, suitable policy has been selected for reducing the rate of worm propagation in super-peer network.

According to Table 1, the multiplicity of $\langle N_{sp}, \lambda_{sp}, \lambda_{or}, a, \omega, Q(0), p, I(0), R(0), u, \gamma, C\rangle$ is used to display the different networks parameters. Generally, in the early stages of the worm propagation, no part has been repaired; Therefore,

$$R(0) = 0 \ \& \ Q(0) = 0$$

Initially, the parameters were set to ⟨6,000, 35, 4, 10, 5, 0, 1, 1, 0, 0.06/6,000, 0.05, 1%⟩ and for the results, the average values of 10 simulation load is used.

Figure 1 Comparison worm propagation model of the super peer and flat peer-to-peer network (see online version for colours)



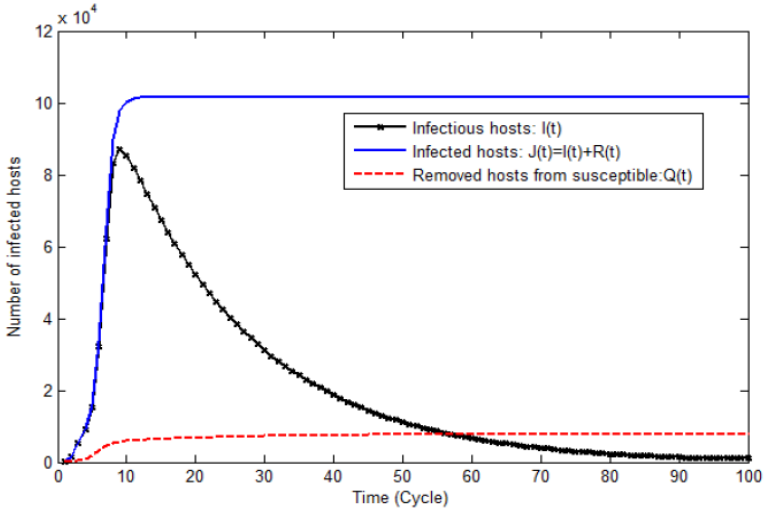
The behaviour of the worm propagation in super-peer networks is compared with its behaviour in a peer-to-peer network (flat). Research in peer-to-peer networks (Adamic et al., 2005; Ripeanu, 2005), has shown that the distribution degree of members in unstructured peer-to-peer networks, follow the power law distribution. In this research, in order to produce a flat peer-to-peer, with the power law distribution degree, the aSHIIP

simulator, which was created by the researchers at the Supelec University, is used. The code for the worm propagation is written in MATLAB with the characteristics and conditions similar to the worm propagation code in super-peer networks. It is implemented by the aSHIIP simulator on a power law distribution topology. Figure 1 compares the results obtained from the implementation of the worm propagation model in a super-peer network and flat peer-to-peer network, in terms of the number of infected nodes in different times. In both charts, a two-factor model has been used.

With this comparison, we find that worm propagation is faster in super-peer networks than flat peer-to-peer ones.

Worm propagation in super-peer networks creates a more serious risk for members of these networks than for members of flat peer-to-peer networks. This is due to the type of topology of these networks and the existence of larger-degree super peer nodes. In Figure 2, the behaviour of the worm propagation in the super-peer network is examined by considering the parameters I(t), J(t) and Q(t) presented in Table 1.

Figure 2 Examination of the behaviour of the worm propagation in super-peer network (see online version for colours)



The number of infected nodes I(t) and also J(t), which contain the previously infected and found to be immune before t, increased rapidly and these two parameters reach their maximum value at the same time. After reaching the peak point, the number of infected nodes I(t) decreases slowly. However, the values of J(t) and Q(t) that are the number of secured nodes transferred from susceptible state remain stable. This is because no more susceptible nodes remain in the system, in other words S(t) = 0.

Figure 3 shows the results of the effect of different scanning rates on the worm propagation in the super-peer network.

Figure 3 Investigating the effect of different scanning rates on the worm propagation in the super-peer network (see online version for colours)

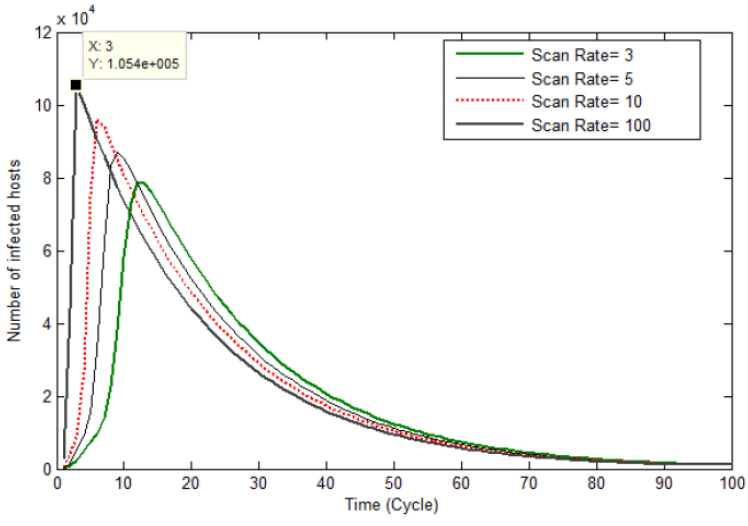
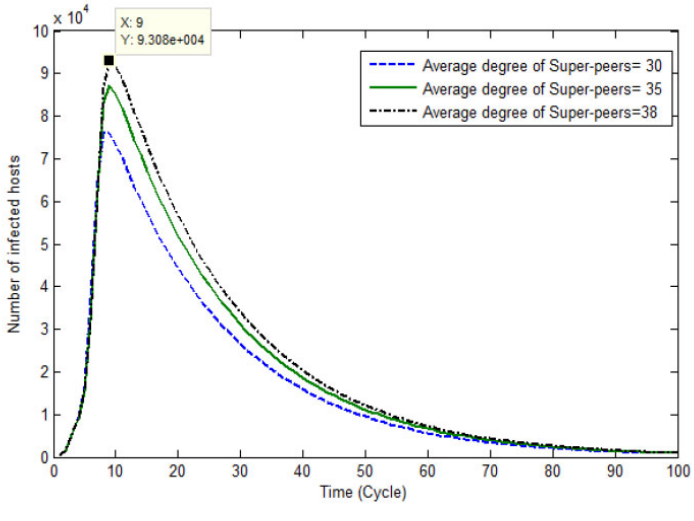


Figure 4 Investigating the effect of different mean values of the super peer degree on worm propagation (λ_{sp}) (see online version for colours)



As shown in Figure 3, when the scanning rate increases, the active worm is propagated more rapidly in the super-peer network. This is because the worm can identify and infect the more number of susceptible hosts simultaneously. Also, the number of infected nodes in the event that the scanning rate is equal to 100, culminated sooner and more than the rest of the scanning rates. The results of the effect of different parameters such as average values of the super peer on the worm propagation (λ_{sp}) are presented in Figure 4. The simulation of the super-peer network is repeated for various sample values of the average

degree of super peer, and the obtained results are shown from the mean values of the 10-times execution in the graph.

The higher the average degree of the super peer, the more adoption of children. As a result, the total number of network members increases. The results indicate that, with the increase in the average degree of super peers, the rate of worm propagation has not increased; just the number of infected nodes has increased. In is because that when the number of network members increases, identifying vulnerable nodes is more time consuming. In Figure 5, the results show the effect of different mean values of the regular members of the super-peer network on the number of infected hosts.

Figure 5 The number of infected nodes per time unit with different values (λ_{or}) in worm propagation (see online version for colours)

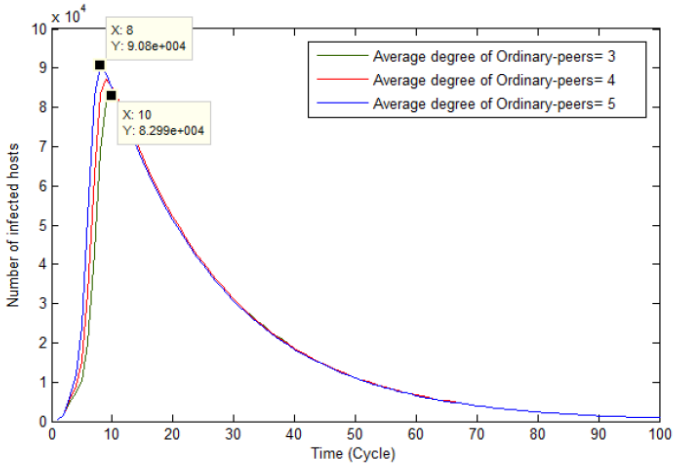
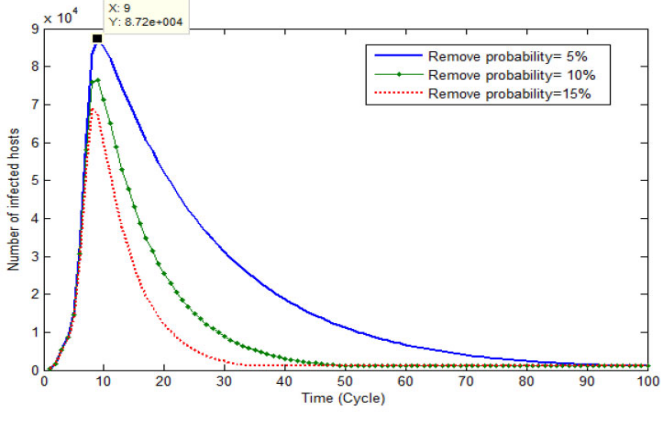


Figure 6 Comparison the number of infected nodes per time unit with different values of recovery probability (see online version for colours)



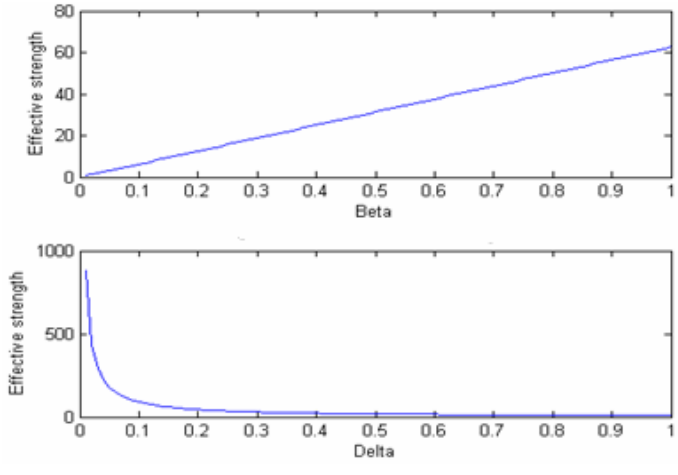
As shown in Figure 5, the worm propagation behaviour is reported in terms of the number of infected nodes by using a two-factor model in the super-peer network with different mean values of regular members. The results indicate that the average degree of

regular members, or the number of connections of regular members with other members, is higher, the rate of worm propagation and the number of infected members increases. The reason is that multiplication in the number of connections of regular members will have less impact on the population growth of the network than multiplication in the number of connections of the super peer.

Figure 6 presents the results of worm propagation behaviour in terms of infected nodes with different values of recovered rate γ. The higher the users pay attention for their system to repair and eliminate infection (in other words, the higher the level of awareness of the people on the network toward the pollution), the Infected hosts are reduced faster. Here we examine two main parameters (the first one is the probability of recovery after the infection, and the second is the probability of the worm attack and propagation), which affects the effective power of the worm (which we represent with S). In this research, in order to standardise the threshold results, the value of the effective strength threshold of normal worm is considered using the propagation function in communication network infrastructure.

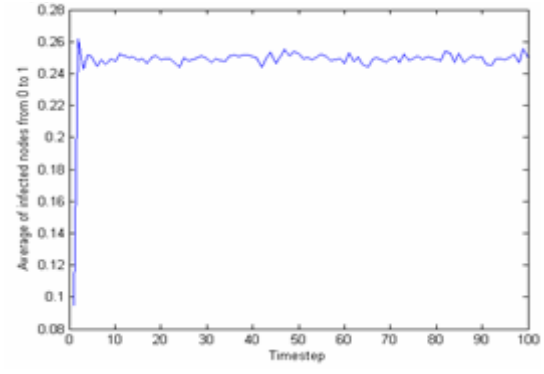
If $S > 1$, it means that the spread of the epidemic in the network is probable. Also, if $S = 1$, it means that the spread of the epidemic is at its peak. First, the effect of the amounts of β, δ is evaluated on the effective strength of the worm independently.

Figure 7 Effective variations in the worm propagation affected by different amounts β, δ (see online version for colours)

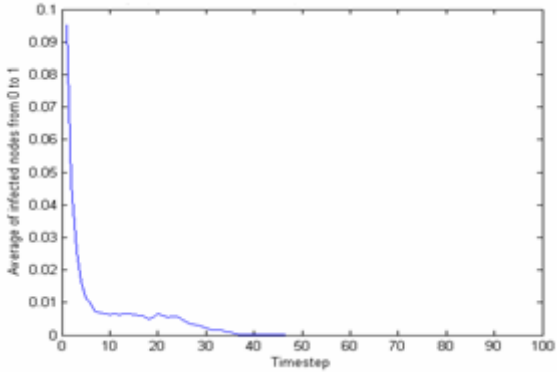


As shown in Figure 7, an increase in the value of β leads to an increase in the effective strength of the worm and more stations in the network become infected. Also, when the amount of δ increases, the immunisation of the infected nodes increases as the effective strength of the worm decreases. The results show that even in the best case, the recovery rate ($\delta = 1$), the effective strength of worm is higher than one (8, 7709), and therefore, the epidemic of the worm propagation in the network is definite and inconsiderable. Figure 8 shows the stability of infection over time in a network for different values of β, δ.

Figure 8 (a) The result of simulating the mean value of the infected nodes for $\beta = 0.2$, $\delta = 0.7$ of the infected nodes for $\beta = 0.01$, $\delta = 0/6$ (b) Results of the simulated average (see online version for colours)



(a)



(b)

The results shown in Figure 8(a) indicate that with increasing of different values of β and δ, if reaching 0.01 and 0.6, the worm propagation decreases and eventually stop. Also, Figure 8(b) shows that if the experiments are repeated, the average number of infected nodes in the network, with the value of $\beta = 0.2$ and $\delta = 0.7$, remain constant and the growth rate of infection is stable.

The proper immunisation policy is to select and secure the nodes with the greatest impact on the graph. In other words, in order to control the worm propagation in the network, the most interactive nodes can be eliminated. In the following, four policies are proposed to assess the immunisation of the super-peer network against the worm's propagation.

4.1 Policy A: select k random nodes for immunisation

In this policy, a number of nodes, including regular node or super peer, are randomly selected. This is done promptly and there is no need for a specific check to determine proper nodes in the graph. This policy is suitable for dynamic graphs where nodes are constantly changing.

This policy responds reliably when a large number of nodes are immunised, otherwise its effectiveness in worm propagation control will be low. Because the effective power (S) is high, more nodes are polluted on the network. Therefore, this policy is suitable for connected graphs where there is a small difference between each node.

The process of implementing this policy is such that, at first k nodes are randomly selected from the graph G. Then these nodes are deleted along with all their edges. Finally, the effective strength of epidemic for new graph is computed. The time complexity of this policy is O(1) as the time required to generate random number is constant. According to the results, the effective strength S is 3.539 in this network. With regard to Figure 9, in which β, δ are constants, logically values of K and effective strength is inversely proportional.

Figure 9 The effect of the random selection of K-node for immunisation on the effective strength of worm (see online version for colours)

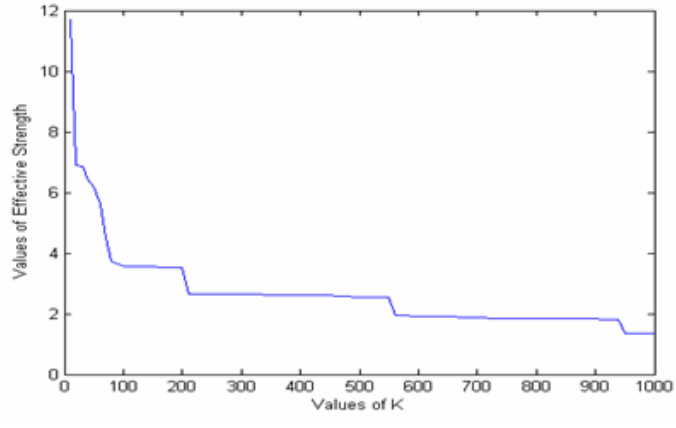
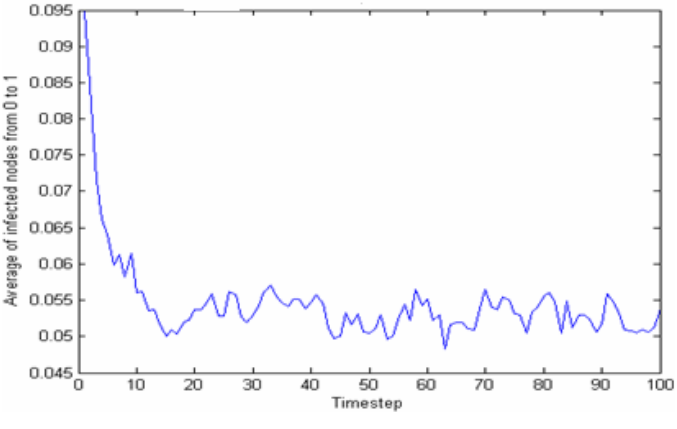


Figure 10 The average of infected nodes after immunisation of k random nodes (see online version for colours)



Due to the high value of S, this policy is weak as can be seen from graph that it requires significant of nodes to be immunised and avoid the network epidemic. Simulation results indicate that a significant number of nodes (about 2,650 nodes out of a total of 6,000 nodes) should be vaccinated to prohibit network-wide epidemic in the network (e.g., worm countermeasures are installed on them).

Figure 10 shows the mean values of normalised infected nodes obtained from the simulation after removing k randomly chosen nodes from the graph. With policy A the results indicate that, the epidemic still persists in the network and it is not properly immunised; hence it is consistent with our empirical calculations.

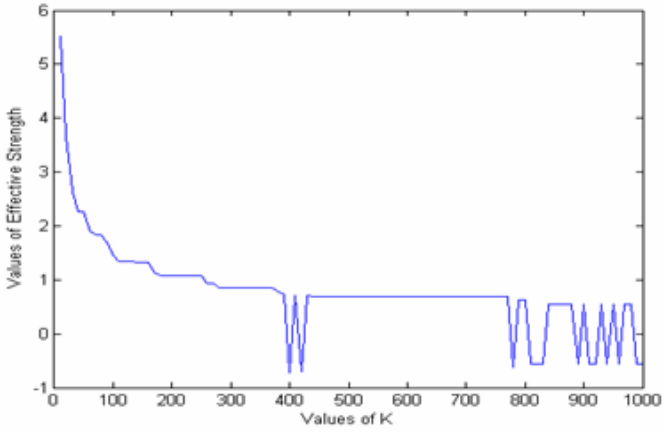
4.2 Policy B: select the k nodes with highest degree for immunisation

In this policy, a list of nodes with the highest degree of communication is provided, and a number of nodes (k nodes) are selected from this list, with all their edges removed. Then the effective strength of the epidemic (worms) is calculated. Hence, the time complexity of the policy can be given as sum of each step which is:

$$O(V+V\text{^}2+kV) \to O\left(V\text{^}2+kV\right) \tag{5}$$

The value of effective strength comes out to be around 1.08 which is very close to threshold of 1 hence, it will prevent network wide epidemic. But since it is above the threshold, the simulation result indicates that epidemic will stay in the network. Figure 11 shows the effective strength based on different values of k.

Figure 11 The effect of choosing K random node from the list of highest levels on the effective strength of the worm (see online version for colours)



The value of effective strength goes below threshold at around k = 250. Hence, this policy could have prevented epidemic if could have immunised few more nodes. This policy provides more optimal results as compared to policy A. The minimum number of vaccinations required to prevent network-wide epidemic comes out to be around 250–260.

This policy is suitable for graphs where there is the most communication between its super-peers such that some randomly selected super-peers are immunised to prevent the worm propagation.

Figure 12 The mean of infected nodes after the immunisation of the selected nodes (see online version for colours)

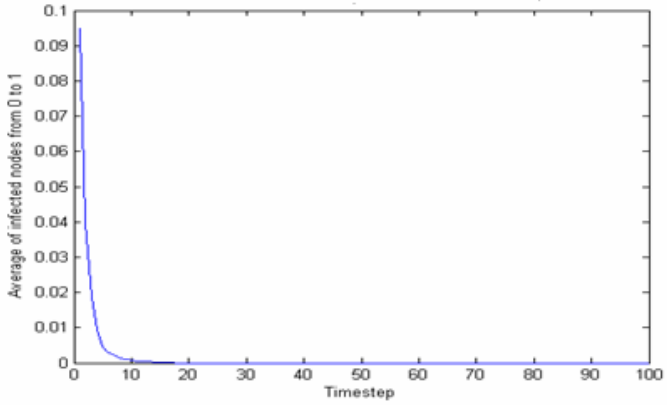


Figure 12 shows the mean of the infected nodes after the immunisation of the selected nodes. These results indicate that the epidemic is stopped after a short time. In other words, policy B's argument for the eradication of the epidemic has been true.

4.3 Policy C: using eigenvector and network's adjacency matrix

In this policy, by analysing the spectrum of graph and analysing its eigenvalue, is determined the activity of each node in the network. For this purpose, first we select absolute largest eigenvalue and then, its vector is analysed to find nodes with high activity in the network. That is selecting k absolute largest value in eigenvector and using the corresponding position in eigenvector as basis of selecting nodes that will be removed from the network. Then we remove these k nodes, along their respective edges from the graph. The immunisation of these stations reduces the transmission of the epidemic.

This policy is appropriate for graphs that their nodes are highly active, and the connectivity between the nodes does not care. The policy is not suitable for high number of nodes and it cannot be considered as a permanent method to limit the worm propagation.

The time complexity of this policy is O (n^3 + kn) where O(n^3) is for calculating eigenvalues and eigenvector, and O(kn) is for selecting k largest node from eigenvector. The effective strength value of virus is 3.07 suggesting that the epidemic will persist continuously.

With regard to the results shown in Figure 13, this policy reduces the network wide epidemic drastically and then behaves linearly. There is no particular value of number of vaccination that can be obtained from this policy. In the following, in Figure 14, the results of simulation show policy C when k selected node removed from the network graph.

Figure 13 The effect of the selection of k nodes with highest value in the eigenvector on the effective strength of the worm (see online version for colours)

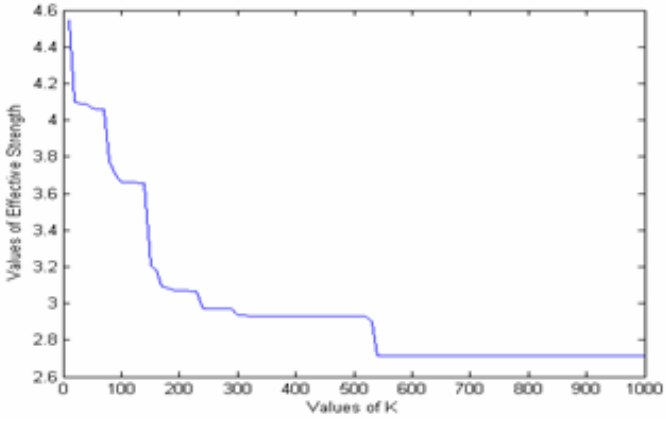
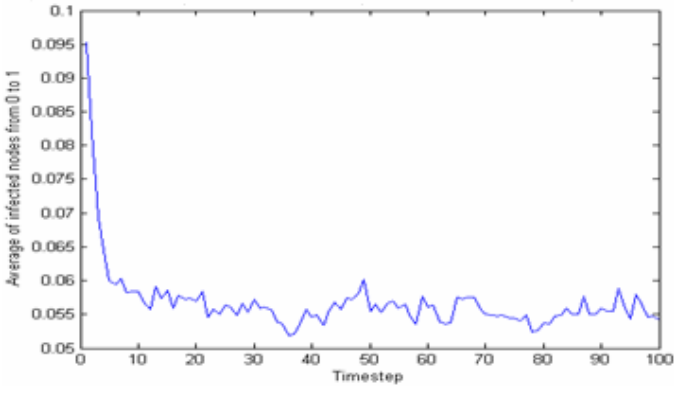


Figure 14 The average of infected nodes based on the immunisation of k-node with the highest value in the eigenvector (see online version for colours)



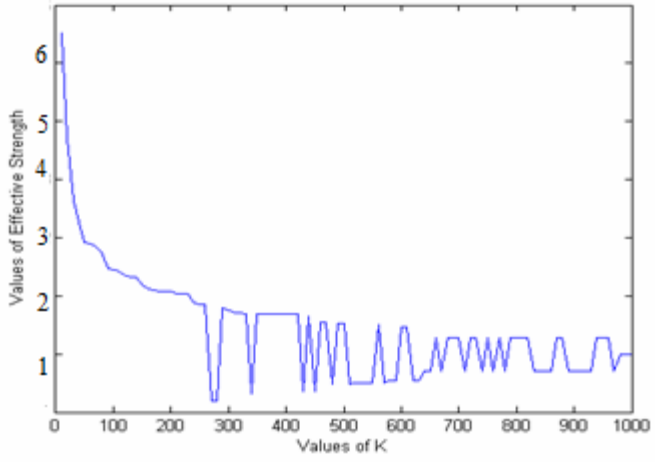
As shown in Figure 14, the policy succeeds in reducing the network-wide epidemic but it never dies out from the network.

4.4 Policy D: select k node with the highest degree of super peer for immunisation

This policy acts better than policy B. Here, the effect of removing each node, and then selecting the next node based on the highest degree of the super peer, is investigated. This method makes it possible to minimise the error in selecting an appropriate super peer when the super peers are connected at the highest levels in the graph. To do this, it is necessary to select the super peer with the highest degree in the network and remove all its edges. This process is repeated k times and the effective strength of the epidemic is evaluated. Therefore, the degree of super peers is selected, from the highest degree to the lowest degree for the removal from the network, respectively. The time complexity of finding degree of each node in a graph is O(V + V^2), and with considering k number of times, the time complexity is O(k(V + V^2)).

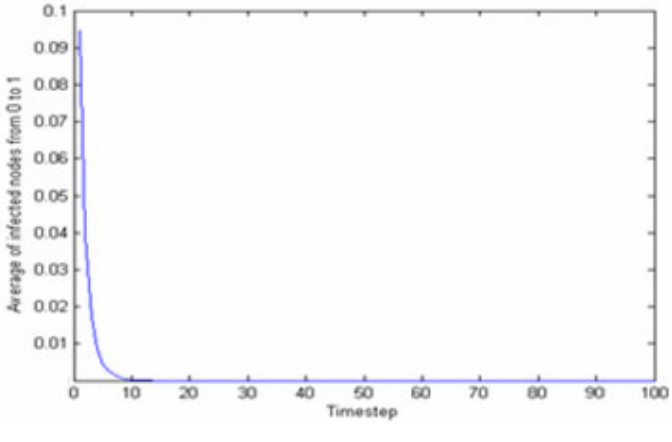
This policy is useful when the number of superpeers is high and they are highly connected. In such condition, if they are immunised from the highest level of connected superpeers, respectively, the worm propagation will be limited more efficiently. Also, if a limited number of nodes should to be immunised, this policy can be used. The value of effective strength in this policy is same as there in policy B as 1.0845. This again means that there will be an epidemic but the low value also suggests that it will prevent network wide epidemic, and again same case happening as policy B. Figure 15 shows the impacts of different values of k on the effective strength of the worm

Figure 15 The effect of k super peer with the highest degree on the effective strength of the worm in policy D, respectively (see online version for colours)



The results shown in Figure 15 indicate that policies B and D have an almost identical impact on the effective strength of worms. The reason is that both of these policies remove highest super peer's degree. Figure 16 shows the results of the simulation of the worm propagation based on policy D and its impact on the average of infected nodes.

Figure 16 The average of infected nodes based on the immunisation of highest degree of super peer (see online version for colours)



The results shown in Figure 16 indicate that the worm propagation of the worm in policy D increase initially and begin to decrease after a shorter time than policy B. Worms also dies out in a more limited number of k (about 200). Thus, it can be concluded that the method used in policy D, the selection of the most suitable super peers for deletion, has the potential for faster eradication of the epidemic in the network.

5 Conclusions

Worm propagation in networks raises problems such as denial of service, disclosure of confidential information, destruction of valuable information, network host disruption and other serious damages. Providing security for peer-to-peer networks in today's applications is essential and inevitable. The most important practical remedy for the worms' destructive effects is limiting the scope or speed of their propagation in the network. This helps to provide the opportunity to repair infected nodes or to protect susceptible nodes and reducing the damage. The sooner network administrators become aware of worm propagation, they can come up with ways to deal with it.

In this research, the parameters affecting of worm propagation in the super-peer network, such as super peer degree, regular peer degree, scan rate, and probability of recovery, were investigated and analysed. Then, an appropriate strategy for controlling the worm's propagation was presented in the super-peer networks. The proposed strategy is based on prioritising super peers with highest degree for immunisation. The simulation results indicate that the above strategy is able to reduce the velocity of worm propagation in super-peer networks. Future research can examine and analyse the behaviour of the active worm propagation considering high dynamics and inactive worms in super-peer networks based on the interest of members in content shared in the network clustering.

References

Adamic, L.A., Lukose, R.M., Puniyani, A.R. and Huberman, B.A. (2005) 'Search in power-law networks', *Physical review E*, Vol. 64, No. 4, p.046135.

Albert, R. and Barabási, A-L. (2005) 'Statistical mechanics of complex networks', *Reviews of Modern Physics*, Vol. 74, No. 1, p.47.

Carlyle, K.R. (2010) *Optimizing Quarantine Regions through Graph Theory and Simulation*, Diss. Kansas State University.

Chen, G. and Gray, R.S. (2007) 'Simulating non-scanning worms on peer-to-peer networks', *Proceedings of the 1st International Conference on Scalable Information Systems*, May, pp.29–41, ACM.

Feng, C., Qin, Z., Cuthbet, L. and Tokarchuk, L. (2010) 'Propagation model of active worms in peer-to-peer networks', *The 9th International Conference Young Computer Scientists 2010, ICYCS 2010*, pp.1908–1912, IEEE.

Feng, C., Zhiguang, Q., Laurence, C. and Laurissa, T. (2008) 'Propagation model of active worms in P2P networks', *The 9th International Conference for Young Computer Scientists*, pp.1908–1912, IEEE.

Ganguly, N. and Deutsch, A. (2004) 'Developing efficient search algorithms for peer-to-peer networks using proliferation and mutation', *Artificial Immune Systems*, pp.357–371, Springer Berlin Heidelberg.

Hua, L., Zheng, Q., Xiaohui, P. and Xiaosong, Z. (2010) 'Propagation model of non-scanning active worm in unstructured peer-to-peer network', *International Conference on Multimedia Information Networking and Security 2010. MINES'10*, Vol. 2, pp.378–381, IEEE.

Jesi, G.P., Montresor, A. and Babaoglu, O. (2007) 'Proximity-aware superpeer overlay topologies', *IEEE Transactions on Network and Service Management*, Vol. 4, No. 2, pp.74–83.

Kienzle, D.M. and Elder, M.C. (2006) 'Recent worms: a survey and trends', *Proceedings of the 2005 ACM*, October, pp.1–10.

Kumar, P. and Chen, M.S. (2008) *Inside the Permutation-Scanning Worms: Propagation Modeling and Analysis*, Department of Computer and Information Science and Engineering University of Florida, Gainesville, Florida 32611, USA.

Luo, W., Liu, J. and Xu, J. (2011) 'An analysis of propagation and capability to attack of active peer-to-peer worms', *2011 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, Vol. 2, pp.506–509.

Meng, S., Shi, C., Han, D., Zhu, X. and Yu, Y. (2008) 'A Statistical study of today's gnutella', *Frontiers of WWW Research and Development-8th Asia-Pacific APWeb*, pp.182–200, Springer Berlin Heidelberg.

Newman, M.E.J. (2004) 'Spread of epidemic disease on networks', *Physical Review E*, Vol. 66, No. 1, p.016128.

Pyun, Y.J. and Reeves, D.S. (2005) 'Constructing a Balanced, log (N)-diameter super-peer topology', *Proceedings of the 4th International Conference on Peer-to-Peer Computing*, Zurich, Switzerland, pp.213–219.

Rajesh, B., Reddy, Y.R.J. and Reddy, B.D.K. (2015) 'A survey paper on malicious computer worms', *International Journal of Advanced Research in Computer Science and Technology*, Vol. 3, No. 2, pp.161–167.

Ripeanu, M. (2005) 'Peer-to-peer architecture case study: Gnutella network', *First International Conference on Peer-to-Peer Computing 2005. Proceedings*, August, pp.99–100, IEEE, DVD.

Taheri, J. (2013) *Collaborative Worm Detection Using Peer-to-Peer Systems*, MSc thesis, Amirkabir University, Tehran.

Tang, H. et al. (2014) 'Propagation of active worms in peer-to-peer networks: modelling and analysis', *Journal of Computers*, Vol. 9, No. 11, pp.2514–2524.

Thommes, R.W. and Coates, M. (2007) 'Epidemiological modelling of peer-to-peer viruses', in *INFOCOM*, Vol. 6, pp.1–12.

Yang, B.B. and Garcia-Molina H., (2003) 'Designing a super-peer network', *Data Engineering. Proceedings. 19th International Conference*, IEEE, pp.49–60.

Yao, Y. et al. (2014) 'Analysis of a delayed internet worm propagation model with impulsive quarantine strategy', *Mathematical Problems in Engineering*, Vol 2014. Article ID 369360, pp.1–18.

Zhang, X.S., Chen, T., Zheng, J. and Li, H. (2010) 'Active worm propagation modelling in unstructured peer-to-peer networks', *Proceedings of the 2nd Symposium International Computer Science and Computational Technology*, pp.35–38.

Zheng, H. (2008) *Internet Worm Research*, For the degree of PhD, pp.12–15, Information Technologies & Science College, Nankai University, Tianjin, China.

Zou, C.C., Gong, W. and Towsley, D. (2005) 'Code red worm propagation modelling and analysis', *Proceedings of the 9th ACM Conference on Computer and Communications Security*, ACM.

Zou, C.C., Gong, W. and Towsley, D. (2006) 'Code red worm propagation modelling and analysis', *Proceedings of the 9th ACM Conference on Computer and Communications Security*, November, pp.138–147, ACM.

Notes

1 VoIP.

2 Neighbourhood Information IP addresses are nodes that are connected to a host, and this information is stored as a table in the host's memory.

3 Hopf bifurcation.

4 Erdos and Renyi graph (ER).

5 If the total number of members in the network is N and the number of ordinary members of N_{or}, then $r = N_{or}/N$.

6 Total.