Modeling the propagation of Peer-to-Peer worms

Xiang Fan\textsuperscript{a}, Yang Xiang\textsuperscript{b,∗}

\textsuperscript{a} School of Management and Information Systems, Central Queensland University, Rockhampton, Queensland 4702, Australia

\textsuperscript{b} School of Information Technology, Deakin University, Burwood, Victoria 3125, Australia

\textbf{A B S T R A C T}

Propagation of Peer-to-Peer (P2P) worms in the Internet is posing a serious challenge to network security research because of P2P worms’ increasing complexity and sophistication. Due to the complexity of the problem, no existing work has solved the problem of modeling the propagation of P2P worms, especially when quarantine of peers is enforced. This paper presents a study on modeling the propagation of P2P worms. It also presents our applications of the proposed approach in worm propagation research.

Motivated by our aspiration to invent an easy-to-employ instrument for worm propagation research, the proposed approach models the propagation processes of P2P worms by difference equations of a logic matrix, which are essentially discrete-time deterministic propagation models of P2P worms. To the best of our knowledge, we are the first using a logic matrix in network security research in general and worm propagation modeling in particular.

Our major contributions in this paper are firstly, we propose a novel logic matrix approach to modeling the propagation of P2P worms under three different conditions; secondly, we find the impacts of two different topologies on a P2P worm’s attack performance; thirdly, we find the impacts of the network-related characteristics on a P2P worm’s attack performance in structured P2P networks; and fourthly, we find the impacts of the two different quarantine tactics on the propagation characteristics of P2P worms in unstructured P2P networks. The approach’s ease of employment, which is demonstrated by its applications in our simulation experiments, makes it an attractive instrument to conduct worm propagation research.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Worms and their variants have been a serious challenge to network security research for many years. Worms can be classified according to the techniques by which they discover new targets to infect. Scanning, which entails probing a set of addresses to identify vulnerable hosts, is the technique most widely employed by worms [1]. Scanning can be implemented differently, which leads to several different types of scanning such as random scanning, localized scanning [2], sequential scanning [3], routable scanning [4], selective scanning [4], importance scanning [5,6], and topological scanning. Topological scanning was employed by the Morris Internet Worm of 1988 as its target discovery technique [7].

Worms employing all other types of scanning, except topological scanning, among the above types do not need to have any knowledge on the topology of the network they intend to propagate across. On the contrary, worms employing topological scanning must have more information on the network they intend to propagate over, or have the capability to discover that information if they do not have it in advance. Therefore, worms employing topological scanning are also called topology-aware worms.

Typical examples of topology-aware worms are worms attacking a flaw in a Peer-to-Peer (P2P) application and propagating across the P2P network by getting lists of peers from their victims and directing their subsequent attacks to those peers. This sort of topology-aware worms are called P2P worms. The Slapper worm of 2003 was a typical example of P2P worms [8]. The subsequent appearance of variations of the Slapper worm (the Slapper.B worm a.k.a. Cinik and the Slapper.C worm a.k.a. Unlock) indicates that P2P worms are becoming increasingly complex and sophisticated [8].

Due to the recent popularity of P2P systems with their increasing number of users, they have become the most effective vehicles for topology-aware worms to achieve the fastest propagation across the Internet. Propagation of P2P worms on top of P2P systems can result in devastating damage, as illustrated by [9]. P2P worms are posing a serious challenge to the Internet.

In order to find an effective and efficient countermeasure against the propagation of P2P worms, we must fully understand their propagation mechanisms. This paper presents a study on modeling the propagation of P2P worms under three different conditions. Our major contributions in this paper are firstly, we
propose a novel logic matrix approach to modeling the propagation of P2P worms under three different conditions; secondly, we find the impacts of two different topologies on a P2P worm’s attack performance; thirdly, we find the impacts of the network-related characteristics on a P2P worm’s attack performance in structured P2P networks; and fourthly, we find the impacts of two different quarantine tactics on the propagation characteristics of P2P worms in unstructured P2P networks.

The rest of the paper is organized as follows. We survey related work in Section 2. We present the proposed innovative logic matrix approach in Section 3. Then, in Section 4, we use the logic matrix approach to investigate the impacts of two different topologies on a P2P worm’s attack performance, the impacts of the network-related characteristics on a P2P worm’s attack performance in structured P2P networks, and the impacts of two different quarantine tactics on the propagation characteristics of P2P worms in unstructured P2P networks. Finally, Section 5 concludes this paper, and points out future research directions.

2. Related work

Mathematical models developed to model the propagation of infectious diseases have been adapted to model the propagation of worms [10]. In the area of epidemiology, both deterministic and stochastic models exist for modeling the spreading of infectious diseases [11–14]. In the network security, both deterministic and stochastic propagation models of worms based on their respective counterparts in epidemiology have emerged.

Deterministic propagation models of worms can be further divided into two categories: continuous-time and discrete-time. Since the propagation of worms is a discrete event process, discrete-time propagation models of worms are more accurate than their continuous-time counterparts in the deterministic regime.

Some typical examples of deterministic propagation models of worms are as follows:

- In the classical simple epidemic model [11–14], all hosts stay in one of only two states at any time: ‘susceptible’ (denoted by ‘S’) or ‘infectious’ (denoted by ‘I’), and thus it is also called the SI model. Stanford et al. [15] presented a propagation model for the Code-Red v2 worm, which is essentially the above classical simple epidemic model.
- The classical general epidemic model (Kermack–McKendrick model) [11–14] improves the classical simple epidemic model by considering removal of infectious hosts due to patching (installing software designed to fix security vulnerabilities).
- The two-factor worm model [10] extends the classical general epidemic model by accounting for removal of susceptible hosts due to patching and considering the pair-wise rate of infection as a variable rather than a constant.
- The discrete-time Analytical Active Worm Propagation (AAWP) model [16] takes into account the time an infectious host takes to infect other hosts, which is an important factor for the spread of worms [17].

Among the above models, all others are continuous-time except the last one, which is discrete-time.

Stochastic propagation models of active worms are based on the theory of stochastic processes. All of them are discrete-time in nature.

Two typical examples of stochastic propagation models of worms are as follows:

- Rohloff and Basar presented a stochastic density-dependent Markov jump process propagation model [18], for worms employing the random scanning approach, drawn from the field of epidemiology [12,19].
- Sellke et al. presented a stochastic Galton–Watson Markov branching process model [20] to characterize the propagation of worms employing the random scanning approach.

A more detailed survey on modeling the propagation process of worms can be found in our previous work [21].

The common limitation is that all of the existing models are not applicable to worms employing topological scanning. No existing model can describe the propagation of P2P worms.

Our novel logic matrix approach proposed in this paper models the propagation processes of P2P worms by difference equations of a logic matrix, which are essentially discrete-time deterministic propagation models of P2P worms. The proposed models are suitable for modeling P2P worms because these models take into account the topology of a P2P network. Existing models do not consider topology issues, which is the root cause of their common limitation mentioned above. Our work in this paper is motivated by the aspiration to invent an easy-to-employ tool to conduct network security research in general and worm propagation modeling research in particular, there being a current absence of such research instruments. Using a logic matrix in worm propagation modeling forms the major difference between this work and existing work.

In our models, temporal issues, such as the time lag for worms to infect peers and the time spent in quarantining peers, have intentionally not been considered. We acknowledge these issues and leave them as our future work. However, the paramount objective of these models is to facilitate determining the maximum number of peers in a P2P system that can be infected, which is the key element to lead effective defense mechanisms. Moreover, these temporal issues are normally strongly affected by human factors during the propagation process; these can be difficult to decide.

3. The logic matrix approach to propagation modeling of Peer-to-Peer worms

At the beginning of this section, we extend the definition of a matrix to allow its elements to be variables or constants of logic type; and term such kind of matrices logic matrices. Several operations of logic matrices are defined. Next, the topology, state, vulnerability status and quarantine status of a network are represented by its topology logic matrix, state logic matrix, vulnerability logic matrix, and quarantine logic matrix, respectively. Finally, an innovative logic matrix formulation of the propagation process of P2P worms under three different conditions is derived from first principles.

3.1. Logic matrix and its operations

We extend the definition of matrix to allow variables or constants of logic type as its elements and term such a kind of matrix a logic matrix. The values of variables of logic type can only be one of the two constants of logic type: True (denoted by ‘T’) or False (denoted by ‘F’). If a logic matrix has only one row or one column, we can also term it a row logic vector or a column logic vector, respectively.

We define the absolute value of a variable $l$ of logic type (denoted by $|l|$) as 1 when its value is ‘T’, and 0 when ‘F’; and define the absolute value of a logic matrix $L$ (denoted by $|L|$) as the total number of its elements with value ‘T’. According to the above definitions, the absolute value of a logic matrix $L$ can be worked out by summing the absolute value of each of its elements $l$, i.e.,

$$|L| = \sum |l|.$$  (1)
A logic matrix $L$ can be inverted. The resultant $\overline{L}$ is a logic matrix of the same dimension with its elements $l_{inv}$ being the result of the logical NOT operation of the corresponding element $l$ of the logic matrix to be inverted. It can be defined mathematically as follows:

$$ l_{inv} = \overline{l}, \tag{2} $$

where the bar over $l$ indicates logical NOT operation.

Two logic matrices $A$ and $B$ can be added together if and only if their dimensions are the same, i.e., they have the same number of rows and the same number of columns. The resultant $S = A + B$ is a logic matrix of the same dimension with its element $s_{ij}$ (in the $i$-th row and the $j$-th column) being the result of the logical OR operation of the corresponding elements $a_{ij}$ and $b_{ij}$ of the two logic matrices to be added together. It can be defined mathematically as follows:

$$ s_{ij} = a_{ij} \lor b_{ij}, \tag{3} $$

where the $\lor$ sign between $a_{ij}$ and $b_{ij}$ indicates the logical OR operation.

A mutation law applies to the logic matrix addition defined above.

Two logic matrices $A$ and $B$ can be multiplied element-by-element if and only if their dimensions are the same, i.e., they have the same number of rows and the same number of columns. The resultant $P = AB$ is a logic matrix of the same dimension with its element $p_{ij}$ (in the $i$-th row and the $j$-th column) being the result of the logical AND operation of the corresponding elements $a_{ij}$ and $b_{ij}$ of the two logic matrices to be multiplied element-by-element. It can be defined mathematically as follows:

$$ p_{ij} = a_{ij} \land b_{ij}, \tag{4} $$

where $a_{ij} \land b_{ij}$ indicates the logical AND operation of $a_{ij}$ and $b_{ij}$.

A mutation law applies to the logic matrix element-by-element multiplication defined above.

A logic matrix $A$ can be multiplied by another logic matrix $B$ in the manner of traditional matrix multiplication if and only if their inner dimensions are the same, i.e., the number of columns of the multiplicand logic matrix (the left one) is equal to the number of rows of the multiplier logic matrix (the right one). The resultant $P = AB$ is a logic matrix with the same number of rows as $A$ and the same number of columns as $B$. We define the value of element $p_{ij}$ (in the $i$-th row and the $j$-th column) of the product as determined by the following equation:

$$ p_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \tag{5} $$

where $a_{ik} b_{kj}$ indicates the logic AND operation of $a_{ik}$ and $b_{kj}$, $n$ denotes the inner dimensions of the multiplicand and the multiplier logic matrices, and $\sum$ denotes the logical OR operation of all resultants of those logical AND operations.

Contrary to logic matrix addition and logic matrix element-by-element multiplication, a mutation law does not apply to the logic matrix multiplication in the manner of traditional matrix multiplication defined above.

Now the stage for later discussion has been set. In the next two sub-sections, we will introduce the concepts of a P2P network's topology logic matrix, state logic matrix, vulnerability logic matrix, and quarantine logic matrix, respectively; and derive our innovative logic matrix formulation of the propagation process of P2P worms under three different conditions from first principles.

### 3.2. The Logic Matrix Representations

According to the traditional directed graph theory, a P2P overlay network can be represented by a directed graph $G$, with its set of vertices $V$ representing all peers connected to form the network, and its set of directed edges $E$ representing all directed links among these peers. A directed link from peer $i$ to peer $j$ means peer $j$ is a neighbor of peer $i$, but peer $i$ is not a neighbor of peer $j$ if there does not exist a directed link from peer $j$ to peer $i$ at the same time. A peer is only able to send messages to its neighbors directly.

Topology of a P2P overlay network consisting of $n$ peers can be represented by an $n \times n$ square matrix $T$ with its element $t_{ij}$ (in the $i$-th row and the $j$-th column) indicating whether there is a directed link from peer $i$ to peer $j$.

In this paper, we propose a different approach from that used under the traditional directed graph theory to indicating the existence or not of a directed link. The logic constant ‘$T$’ is used to indicate there is a directed link, and the logic constant ‘$\overline{T}$’ to indicate there is not. Therefore, topology of a P2P overlay network consisting of $n$ peers can be represented by an $n \times n$ logic square matrix. We term it the topology logic matrix of the P2P overlay network.

Each row of the topology logic matrix of a P2P overlay network forms a row logic vector, which is a logic vector representation of outbound links (neighbors) of a particular peer belonging to the network. We call this row logic vector the peer’s topology out-degree logic vector. Each column of the topology logic matrix of a P2P overlay network forms a column logic vector, which is a logic vector representation of inbound links of a particular peer belonging to the network. We call this logic column vector the peer’s topology in-degree logic vector. For example, the $i$-th row of a topology logic matrix represents all outbound links (neighbors) of peer $i$; and the $j$-th column of the topology logic matrix represents all inbound links of peer $j$.

It can be easily derived that the values of topology in-degree and topology out-degree of each peer belonging to a P2P overlay network equate to the absolute values of the peer’s topology in-degree logic vector and topology out-degree logic vector, respectively, which can be worked out by using (1).

Similarly, we represent states of all the $n$ peers belonging to the P2P overlay network by a row logic vector $S$ of length $n$ with its element $s_j$ (the $j$-th element) indicating whether peer $j$ has been infected by the worm and become infectious. The logic constant ‘$T$’ is used to indicate a peer has been infected and become infectious, and the logic constant ‘$\overline{T}$’ to indicate it has not. We term the above row logic vector the P2P overlay network’s state logic vector. It can be derived that the total number of infected and infectious peers in a P2P overlay network equates to the absolute value of the network’s state logic vector, which can be worked out by using (1).

A healthy peer vulnerable to the worm can be infected by the worm and become infectious. However, a peer which is not vulnerable to the worm will not be infected by the worm and become infectious. We represent the vulnerability status of all the $n$ peers in the P2P overlay network by a row logic vector $V$ of length $n$ with its element $v_j$ (the $j$-th element) indicating whether peer $j$ is vulnerable to the worm. The logic constant ‘$T$’ is used to indicate a peer is vulnerable to the worm, and the logic constant ‘$\overline{T}$’ to indicate it is not. We term the above logic row vector the P2P overlay network’s vulnerability logic vector. It can be derived that the total number of peers vulnerable to the worm in a P2P overlay network equates to the absolute value of the network’s vulnerability logic vector by using (1).

We represent the quarantine status of all the $n$ peers belonging to the P2P overlay network by a row logic vector $Q$ of length $n$ with its element $q_j$ (the $j$-th element) indicating whether peer $j$ has been...
quarantined for the worm. A quarantined healthy peer will not be infected by the worm; and a quarantined infected and infectious peer will be cured and will not be infected again by the worm. The logic constant 'I' is used to indicate a peer has been quarantined, and the logic constant 'Q' to indicate it has not. We term the above row logic vector the P2P overlay network's quarantine logic vector. It can be derived that the total number of quarantined peers in a P2P overlay network equals to the absolute value of the network's quarantine logic vector, which can be worked out by using (1).

3.3. The logic matrix formulation

Based on the above extensions to the matrix and its operations, and extensions to the matrix representation of a network in the traditional directed graph theory, we are now ready to derive our innovative logic matrix formulation of the propagation process of P2P worms. The derivation is based on the following assumptions:

An infected and infectious peer will send the worm packets to all other peers belonging to the same P2P overlay network to which it has an outbound link, regardless of the state (infected by the worm and infectious or not) and the quarantine status (quarantined for the worm or not) of those peers. A healthy (not infected by the worm and not infectious) peer will be infected by the worm and become infectious once it receives the worm packets from an infectious peer, provided the healthy peer is not quarantined for the worm. An infected and infectious peer will remain in that state once it receives the worm packets from an infectious peer, provided the infected and infectious peer is not quarantined for the worm. A healthy peer quarantined for the worm will not be infected by the worm; and an infected and infectious peer quarantined for the worm will be cured and will not be infected again by the worm.

The time lags from sending the worm packets, to receiving the worm packets, to having the recipient peers infected by the worm, to the peers infected by the worm becoming infectious will not be considered, nor will the time spent in quarantining peers.

There are a total of \( I_0 \) peers which are infected by the worm and infectious. According to the above assumptions, the P2P overlay network's initial state (State 0) can be represented by its initial state logic vector \( S_0 \) of length \( n \); and the absolute value of \( S_g \) equates to the total number of peers which are initially infected by the worm and infectious \( (I_g)\), i.e.,

\[
|S_0| = I_0. \tag{6}
\]

Generally, state \( g \) of the logical P2P overlay network can be represented by its state logic vector \( S_g \), and the absolute value of \( S_g \) equates to the total number of peers which are infected by the worm and infectious at that state \( (I_g)\), i.e.,

\[
|S_g| = I_g. \tag{7}
\]

The next state \((State g + 1)\) of the logical P2P overlay network can be represented by its state logic vector \( S_{g+1} \), and the absolute value of \( S_{g+1} \) equates to the total number of peers which are infected by the worm and infectious at that state \( (I_{g+1})\), i.e.,

\[
|S_{g+1}| = I_{g+1}. \tag{8}
\]

We notice that the logical P2P overlay network's next state represented by its state logic vector \( S_{g+1} \) is fully determined by the network's current state represented by its state logic vector \( S_g \), the network's topology represented by its topology logic matrix \( T \), the network's vulnerability status represented by its vulnerability logic vector \( V \), and the network's quarantine status represented by its quarantine logic vector \( Q \).

If all peers are vulnerable to the P2P worm, we find the relationship among \( S_{g+1}, S_g, T, V, \) and \( Q \) can be described mathematically as follows:

\[
S_{g+1} = S_g + S_g T Q. \tag{9}
\]

Let \( S_{g+1}^{new} \) stand for the second term in the above equation (after the + sign), the above equation can now be simplified to

\[
S_{g+1} = S_g + S_{g}^{new}. \tag{10}
\]

The term represented by \( S_{g+1}^{new} \) actually says if at State \( g + 1 \) at least one peer among those peers from which peer \( j \) has inbounds links is infectious, peer \( j \) will be infected by the worm and become infectious at State \( g + 1 \) provided peer \( j \) is not quarantined.

Since both \( S_g \) and \( Q \) are row logic vectors of length \( n \) and \( T \) is an \( n \times n \) square logic matrix, \( S_{g+1}^{new} \) will be a row logic vector of length \( n \). It can be derived that \( S_{g+1}^{new} \) is a logic vector representation of all those peers that can be infected by the worm at State \( g + 1 \), given the network's state at State \( g \) represented by its state logic vector \( S_g \), the network's topology represented by its topology logic matrix \( T \), and the network's quarantine status represented by its quarantine logic vector \( Q \). \( S_{g+1}^{new} \) may or may not include peer or peers infected by the worm at states prior to State \( g + 1 \). Then, (9) and (10) can be easily derived.

If quarantine is not enforced at all and not all peers are vulnerable to the P2P worm, (9) will be changed to

\[
S_{g+1} = S_g + S_g T V. \tag{11}
\]

The term represented by the second term in the above equation (after the + sign) actually says if at State \( g \) at least one peer among those peers from which peer \( j \) has inbounds links is infectious, peer \( j \) will be infected by the worm and become infectious at State \( g + 1 \) provided peer \( j \) is vulnerable to the worm.

If quarantined is not enforced at all and all peers are vulnerable to the P2P worm, (11) will be simplified to

\[
S_{g+1} = S_g + S_g T. \tag{12}
\]

Eq. (12) is also a special case of (9) when \( Q \) is a row logic vector with all its elements being 'I'.

Eqs. (9), (11) and (12) are actually discrete-time deterministic propagation models of P2P worms under three different condition, respectively, written in the form of difference equations of the logic matrix.

Starting from some certain state, there will be no newly infected peer to occur and thus actually, the propagation will stop. The state from which the propagation will cease is the earliest state whose state logic vector \( S_c \) satisfies the following equation:

\[
|S_{c+1}| = |S_c|. \tag{13}
\]

where \( S_{c+1} \) stands for the state logic vector of the state immediately after the state with state logic vector \( S_c \). We call the earliest state whose state logic matrix \( S_c \) satisfies (13) the final state of the P2P overlay network.

The proposed logic matrix approach essentially translates the propagation processes of P2P worms into a sequence of logic matrix operations.

4. Simulation experiments: applications of the logic matrix approach

Our evaluation metric for attack performance in this paper is a P2P worm's coverage rate (denoted by \( c \)) in a logical P2P overlay network. It is defined as the ratio of number of peers in the network that can be infected by the worm to number of peers in the network.
that are vulnerable to the worm. It can be worked out by using the following equation:

$$c = \frac{|S_c|}{|V|}.$$  \hspace{1cm} (14)

where $S_c$ is the state logic vector of the network when the propagation process has just stopped, and $V$ is the vulnerability logic vector to the worm of the network.

One of our evaluation metrics for network-related characteristics of P2P networks in this paper is vulnerability rate (denoted by $v$) to a P2P worm of a logical P2P overlay network, which is defined as the ratio of number of peers in the network that are vulnerable to the worm to total number of peers in the network. It can be worked out by using the following equation:

$$v = \frac{|V|}{n},$$ \hspace{1cm} (15)

where $V$ is the vulnerability logic vector to the worm of the network, and $n$ is total number of peers in the network.

The other two of our evaluation metrics for network-related characteristics of P2P networks in this paper are topology out-degree, which refers to the number of logical neighbors maintained by each peer locally; and network size, which refers to the total number of peers in a P2P network.

Our defense-related evaluation metric in this paper is quarantine rate (denoted by $q$) for a P2P worm of a logical P2P overlay network. It is defined as the ratio of number of peers belonging to the network that are quarantined for the worm to total number of peers belonging to the network; and can be worked out by using the following equation:

$$q = \frac{|Q|}{n},$$ \hspace{1cm} (16)

where $Q$ is the quarantine logic vector for the worm of the network and $n$ is total number of peers belonging to the network.

We apply the proposed logic matrix approach in our simulation experiments under the following three different conditions using MathWorks’ MATLAB.

## 4.1. All peers being vulnerable to the P2P worm and no quarantine at all

In this case, we investigate the impacts of the two different topologies, namely the simple random graph topology and the pseudo power law topology on the coverage rate of P2P worms.

### 4.1.1. The simple random graph topology

We investigate the impacts of the two parameters, namely the number of initially infected computers belonging to a P2P network and the mean value of topology out-degree of the network, on the coverage rate of P2P worms in the network.

Our implementation in MATLAB assumes there are a total of 10,000 peers (computers) belonging to the logical P2P overlay network under consideration. Therefore, the topology of the overlay network is represented by its topology logic matrix, which is a 10,000 by 10,000 square logic matrix; and its initial state is represented by its initial state logic vector, which is a 1 by 10,000 logic matrix (row logic vector). In the experiments conducted for this sub-section, we assume each peer has the same value of topology out-degree. Peers to which each peer has outbound links are randomly selected from all peers except the peer itself belonging to the overlay network, which means we do not allow loop, that is, no peer has an outbound link to itself. Therefore, we call the topology of the overlay network in the experiments conducted for this sub-section the simple random graph topology.

### 4.1.2. The pseudo power law topology

We investigate the impacts of the two parameters, namely the number of initially infected computers belonging to a P2P network and the maximum value of topology out-degree of the network, on the coverage rate of P2P worms in the network.

In the experiments conducted for this sub-section, we assume only a very small number (10 in our experiments) of peers have the maximum value of topology out-degree, and all other peers have the minimum value (1 in our experiments) of topology out-degree. Although the distribution of topology out-degree in our experiments does not strictly follow a power law, it does have the most important features of power law distribution, namely peers with the maximum value of topology out-degree are rare and most peers have the minimum value of topology out-degree. Therefore, we call the topology of the overlay network in the experiments conducted for this sub-section the pseudo power law topology.

We conduct our simulation under different combinations of values of the number of initially infected computers and the maximum value of topology out-degree.

Firstly, we fix the number of initially infected peers (computers) belonging to the overlay network to be 1, and try to find out the impact of mean value of topology out-degree on the coverage rate of P2P worms in the overlay network. The initially infected peer is randomly select from all peers belonging to the overlay network. A total of 5 scenarios listed in Table 1 are investigated. The experiment for each scenario is repeated 100 times. Next, the mean value of coverage rate and coefficient of variation of coverage rate are worked out. Results from the experiments are listed in Table 1.

As shown by Table 1, the mean value of topology out-degree has great impact on both the mean value and coefficient of variation of coverage rate of P2P worms in the overlay network featuring the simple random graph topology. An increase in the mean value of topology out-degree results in an increase in the mean value of coverage rate but a decrease in the coefficient of variation of coverage rate. When the mean value of topology out-degree is increased to 3, the mean value of coverage rate is increased to over 90% and its coefficient of variation becomes very small, which indicates 3 is the minimum mean value of topology out-degree which can make a P2P worm able to infect most peers with very high certainty.

Next, we fix the number of initially infected peers (computers) belonging to the overlay network to be 10/100, and repeat the above experiments. Results from the experiments are listed in Table 2.

Table 2 shows similar trends to those shown by Table 1, which indicates the impact of number of initially infected peers on the coverage rate of a P2P worm in the overlay network featuring the simple random graph topology is insignificant.

### 4.1.2. The pseudo power law topology

We investigate the impacts of the two parameters, namely the number of initially infected computers belonging to a P2P network and the maximum value of topology out-degree of the network, on the coverage rate of P2P worms in the network.

In the experiments conducted for this sub-section, we assume only a very small number (10 in our experiments) of peers have the maximum value of topology out-degree, and all other peers have the minimum value (1 in our experiments) of topology out-degree. Although the distribution of topology out-degree in our experiments does not strictly follow a power law, it does have the most important features of power law distribution, namely peers with the maximum value of topology out-degree are rare and most peers have the minimum value of topology out-degree. Therefore, we call the topology of the overlay network in the experiments conducted for this sub-section the pseudo power law topology.

We conduct our simulation under different combinations of values of the number of initially infected computers and the maximum value of topology out-degree.

Firstly, we fix the number of initially infected peers (computers) belonging to the overlay network to be 1, and try to find out the impact of the maximum value of topology out-degree on the
coverage rate in the overlay network. The initially infected peer is randomly select from all peers belonging to the overlay network. A total of 5 scenarios are investigated. In the experiments conducted for this sub-section, we assume each peer has either the maximum value of topology out-degree or the minimum value of topology out-degree. Peers to which each peer has outbound links are randomly selected from all peers except the peer itself belonging to the overlay network. The experiment for each scenario is repeated 100 times. Next, the mean value of coverage rate and coefficient of variation of coverage rate are worked out. Results from the experiments are listed in Table 3.

As shown by Table 3, when all initially infected peers are randomly selected from all peers, the maximum value of topology out-degree has a little impact on the mean value and coefficient of variation of coverage rate of P2P worms in the overlay network featuring the pseudo power law topology. An increase in the maximum value of topology out-degree results in a little increase in the mean value of coverage rate and a little increase in coefficient of variation of coverage rate as well, which indicates the small gain in coverage rate could be offset by the small loss in certainty. The worm is not able to infect most peers with very high certainty.

Then, we fix the number of initially infected peers (computers) belonging to the overlay network to be 10, and repeat the above experiments. Results from the experiments are listed in Table 4.

Table 4 shows similar trends (just an insignificantly higher coverage rate and an insignificantly lower coefficient of variation of coverage rate) to those shown by Table 3, which indicates, when all initially infected peers are randomly selected from all peers, the impact of number of initially infected peers on the coverage rate of a P2P worm in the overlay network featuring the pseudo power law topology is insignificant.

Finally, initially infected peers are randomly select from only those peers with maximum topology out-degree and we repeat all of the above experiments described in this sub-section. Results from the experiments are listed in Tables 5 and 6.

As shown by Tables 5 and 6, when all initially infected peers are randomly selected from only those peers with maximum topology out-degree, the maximum value of topology out-degree has a great impact on both the mean value and coefficient of variation of coverage rate of P2P worms in the overlay network featuring the pseudo power law topology. An increase in the maximum value of topology out-degree results in an increase in the mean value of coverage rate but a decrease in the coefficient of variation of coverage rate. However, the impact of the number of initially infected peers is insignificant. When the maximum value of topology out-degree reaches 2000, the worm is able to infect most peers with very high certainty, regardless of the number of initially infected peers.

4.2. Not all peers being vulnerable to the P2P worm and no quarantine at all

In a structured P2P network, the topology out-degree $d$ of each peer is a constant. It is characterized by the following probability distribution:

$$\begin{align*}
P(d = k) &= 1 \\
P(d \neq k) &= 0,
\end{align*}$$

where $k$ is a constant.

In this sub-section, we only consider structured P2P networks. Therefore, all peers in the network have the same topology out-degree.

Our objective is to investigate the impacts of the network-related characteristics (measured by the evaluation metrics: vulnerability rate $v$, topology out-degree $d$, and network size $n$) on a P2P worm’s attack performance in structured P2P networks (measured by the evaluation metric: coverage rate $c$).

Our simulation experiments include scenarios with vulnerability rate of 1.0. Experimental results from them set the benchmark to compare to. When all peers are vulnerable to the worm, (12) instead of (11) forms the foundation of our implementation of the proposed logic matrix approach. Otherwise, our implementation is based on (11).
Our simulation experiments are based on the following assumptions:

- Topology out-degree \((d)\) of each peer in the structured P2P network under consideration strictly follows the probability distribution [17]. Neighbors of a peer are randomly selected from all other peers except the peer itself.
- Peers vulnerable to the worm are selected randomly from all peers in the network.
- There is only 1 initially infected peer, which is selected randomly from all peers in the network that are vulnerable to the worm.

Based on the above assumptions, we first populate the topology logic matrix of the structured P2P network under consideration by letting the probability that a randomly selected peer has \(k\) neighbors follow [17]. Then, we populate the vulnerability logic vector of the network, before populating the initial state logic vector of the network.

We conduct our simulation experiments for the three different sets of scenarios. Each of our simulation experiment is repeated 100 times, and then average values of coverage rate are reported as final results.

For our first set of scenarios, we fix topology out-degree at 3. We let vulnerability rate vary from 1.0 to 0.2 with step size 0.2; and let network size vary from 1000 to 10,000 with step size 1000. A vulnerability rate of 1.0 actually means all peers in the network are vulnerable to the worm. The experimental results from the above set of scenarios are illustrated by Fig. 1.

Fig. 1 reveals that under the set conditions, the coverage rate of a P2P worm in a logical P2P overlay network will decrease if vulnerability rate is decreased. This is sensible since more vulnerable peers in the network naturally lead to higher attack performance measured by the coverage rate. The upper bound of the coverage rate is approximately 0.95. It is achieved when all peers are vulnerable, i.e., \(v = 1.0\) (the top curve in Fig. 1). The lower bound of the coverage rate is close to 0. It is achieved when 20% peers are vulnerable, i.e., \(v = 0.2\) (the bottom curve in Fig. 1).

The above findings imply that both attackers and defenders can manipulate vulnerability rate \(v\) to improve or worsen attack performance, respectively, and more importantly that limiting vulnerability rate to be below 0.4 is critical to defenders.

Fig. 1 also shows that, when network size is in the range 1000–10,000 inclusive, it has no significant impact on attack performance measured by the coverage rate if both topology out-degree and vulnerability rate are fixed. This finding implies that we can choose a smaller value in the range 1000–10,000 for network size \(n\) in our later experiments to shorten simulation time, and that neither attackers nor defenders can manipulate network size \(n\) to improve or worsen attack performance, respectively.

For our second set of scenarios, we fix vulnerability rate at 0.5. We let topology out-degree vary from 1 to 5 with step size 1; and let network size vary from 1000 to 10,000 with step size 1000. The experimental results from the above set of scenarios are illustrated by Fig. 2.

Fig. 2 reveals that under the set conditions, the coverage rate of a P2P worm in a logical P2P overlay network will increase if topology out-degree is increased. This is sensible since the more neighbors a peer in the network has naturally leads to higher attack performance measured by the coverage rate. The upper bound of the coverage rate is approximately 0.85. It is achieved when all peers have 5 neighbors, i.e., \(d = 5\) (the top curve in Fig. 2). The lower bound of the coverage rate is 0. It is achieved when all peers have only 1 neighbor, i.e., \(d = 1\) (the bottom curve in Fig. 1). The coverage rate drops significantly from above approximately 0.4 to below 0.05 when topology out-degree decreased from 3 to 2. The above findings imply that both attackers and defenders can manipulate topology out-degree \(d\) to improve or worsen attack performance, respectively, and more importantly that limiting topology out-degree to be below or equal to 2 is critical to defenders.

Based on the common finding from our first 2 sets of simulation experiments that when network size is in the range 1000–10,000 inclusive, it has no significant impact on attack performance, for our third set of scenarios, we fix network size at 5000 to shorten simulation time. We investigate the two cases given below:

Case 1—In this case, we let vulnerability rate vary from 0.1 to 1.0 with step size 0.1; and let topology out-degree vary from 1 to 5 with step size 1. Here, our focus is on the impact of vulnerability rate rather than topology out-degree on attack performance measured by the coverage rate. Therefore, we choose a smaller step size for vulnerability rate, but only a few topology out-degree values are investigated.

The experimental results from the above case are illustrated by Fig. 3. Fig. 3 reveals that generally the coverage rate of a P2P worm in a logical P2P overlay network will increase if vulnerability rate is increased. This is sensible since more vulnerable peers in the network naturally lead to higher attack performance measured by the coverage rate.

More importantly, Fig. 3 also shows that the takeoff points on the curves do not correspond to the same value of vulnerability rate. Here, takeoff point refers to the point on a curve in Fig. 3 immediately to the right of which the slope of the curve increases.
and that the Fig. 3 reveals that generally the coverage rate of a P2P worm in a logical P2P overlay network will increase if topology out-degree is reduced. This is sensible since fewer vulnerable peers demand more neighbors a peer in the network has, to achieve the same attack performance.

4.3. All peers being vulnerable to the P2P worm and quarantine being nonexistent

In an unstructured P2P network, the topology out-degree ($d$) of each peer is a variable. It is characterized by the following power law distribution:

$$
\begin{align*}
\text{if } D_{\min} \leq k \leq D_{\max} &:\quad P(d = k) = \frac{C}{k^A}, \\
\text{or } k \neq D_{\max} &:\quad P(d = k) = 0,
\end{align*}
$$

where $D_{\min}$ and $D_{\max}$ stands for the minimum topology out-degree and maximum topology out-degree, respectively. $A$ represents the power law degree, and $C$ is a constant. The set of equations (18) gives the probability that a randomly selected peer has $k$ neighbors.

In this subsection, we only consider unstructured P2P networks. Therefore, not all peers in the network have the same topology out-degree.

Our paramount objective is to find a quarantine tactic whose enforcement will lead to a lower attack performance (measured by the attach-related evaluation metric: coverage rate $c$) at a lower cost of defense effort (measured by the defense-related evaluation metric: quarantine rate $q$).

According to probability theory, the following equations must hold:

$$
1 = \sum_{k=D_{\min}}^{D_{\max}} P(d = k) = C \sum_{k=D_{\min}}^{D_{\max}} \frac{1}{k^A}, \\
$$

where $E(d)$ stands for expected value of topology out-degree.

Then, it can be easily derived from (19) and (20) that the power law degree $A$ is a function of $D_{\min}, D_{\max}$, and $E(d)$ described implicitly by the following equation:

$$
E(d) = \frac{D_{\max}}{\sum_{k=D_{\min}}^{D_{\max}} \frac{1}{k^A}}.
$$

Finally, once the power law degree $A$ is determined according to (20), given $D_{\min}, D_{\max}$, and $E(d)$, the constant $C$ can be worked out according to (19) or (20).

The most important feature of the above power law distribution of topology out-degree in the unstructured P2P system is that there are fewer peers with larger topology out-degree than those with smaller topology out-degree.

Let $D_{\min} = 1, D_{\max}$ vary from 100 to 1000 and the expected value of topology out-degree $E(d)$ vary from 2 to 32, we numerically determine power law degree $A$. The results are shown in Fig. 5.

Fig. 5 shows that a larger maximum topology out-degree requires a larger power law degree, and that a larger expected value of topology out-degree demands a smaller power law degree.

<table>
<thead>
<tr>
<th>Number of Peers: 5,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = 1$</td>
</tr>
<tr>
<td>$v = 0.2$</td>
</tr>
</tbody>
</table>

Fig. 3. Coverage rate as a function of topology out-degree and vulnerability rate when network size is fixed at 5000 (Case 1).

Fig. 4. Coverage rate as a function of vulnerability rate and topology out-degree when network size is fixed at 5000 (Case 2).
Our simulation experiments are based on the following assumptions:

- **Topology out-degree** ($d$) of each peer belonging to the unstructured P2P network under consideration strictly follows the power law distribution \( (18) \). $E(d) = 3$, $D_{\text{min}} = 1$, and $D_{\text{max}}$ varies from 100 to 1000 with step size 100. Neighbors of a peer are randomly selected from all other peers except the peer itself.
- Peers quarantined are selected accordingly, based on the quarantine tactics enforced, which are detailed in the next two subsections.
- There is only 1 initially infected peer, which is selected randomly from all peers not quarantined.
- We conduct our simulation experiments for two different values of \( n \) (total number of peers belonging to the system). We first assume \( n \) to be 5000 and then double it, i.e., assume \( n \) to be 10,000. We believe 10,000 peers are sufficient for our simulation experiments, and intend to investigate whether 5000 peers will generate significantly different results.

Based on the above assumptions, we populate the topology logic matrix of the unstructured P2P network under consideration by letting the probability that a randomly selected peer has \( k \) neighbors follow \( (18) \). How to populate the quarantine logic vector of the network is detailed later. Once it is populated, we can populate the initial state logic vector of the network.

Our simulation experiments include scenarios with no quarantine at all. The experimental results from these set the benchmark to compare to. When there is no quarantine, \( (12) \) instead of \( (9) \) forms the foundation of our implementation of the proposed logic matrix approach. When quarantine is enforced, our implementation is based on \( (9) \).

We conduct our simulation experiments for two different quarantine tactics, namely random quarantine and larger topology out-degree priority quarantine. Each of our simulation experiment is repeated 100 times, and then the average values of coverage rate are reported as final results.

### 4.3.1. Random Quarantine

Random quarantine means peers quarantined are randomly selected from all peers. We populate the quarantine logic vector of the unstructured P2P network under consideration by letting each peer have the same probability of being quarantined when this quarantine tactic is enforced. Then, we populate the initial state logic vector of the network.

We conduct our experiments for the 5 sets of scenarios with quarantine rate \( q \) varying from 0 to 0.4 with step size 0.1. A quarantine rate of 0 actually means no quarantine at all. We include no quarantine as a special case of random quarantine, which facilitates comparison of experimental results.

The experimental results from random quarantine are illustrated by Figs. 6 and 7 for the two cases: \( n \) (total number of peers belonging to the P2P network) = 5000 and \( n = 10,000 \), respectively.

Figs. 6 and 7 reveal that generally, coverage rate of a P2P worm in a logical P2P overlay network will decrease if quarantine rate is increased. This is sensible because a higher defense effort will naturally lead to a lower attack performance. However, as mentioned previously, our paramount objective is to find a quarantine tactic whose enforcement will lead to a lower attack performance at a lower cost of defense effort. Therefore, the above finding cannot serve our paramount objective.

Figs. 6 and 7 also show that maximum topology out-degree has no significant impact on attack performance and defense effort when it is in the range 100–1000 inclusive, and that 5000 peers will not generate significantly different results. The implications of the above findings include that we can choose the smallest value of $D_{\text{max}}$ (100) and the smaller value of $n$ (5000) in our future experiments to shorten simulation time, and that neither attackers nor defenders can manipulate $n$ or $D_{\text{max}}$ to improve attack performance or reduce defense effort, respectively.
4.3.2. Larger topology out-degree priority quarantine

Larger topology out-degree priority quarantine means peers with larger topology out-degree are quarantined prior to peers with smaller topology out-degree.

When this quarantine tactic is enforced, we populate the quarantine logic vector of the unstructured P2P network under consideration by following the procedure given below:

Firstly, we work out the absolute value of each peer's topology out-degree logic vector. Secondly, all peers are sorted in descending order of the absolute value calculated above. By doing this, we actually sort all peers into a list in descending order of number of neighbors since, as mentioned previously, each peer’s topology out-degree logic vector is a logic vector representation of its outbound links (neighbors). Thirdly, we quarantine peers in the same order as their order in the sorted list of peers. Then, we populate the initial state logic vector of the network.

We conduct our experiments for the 5 sets of scenarios with quarantine rate $q$ varying from 0 to 0.16 (40% of 0.4, which is the maximum quarantine rate investigated under random quarantine) with step size 0.04.

The experimental results from larger topology out-degree priority quarantine are illustrated by Figs. 8 and 9 for the two cases: $n$ (total number of peers belonging to the P2P network) = 5000 and $n = 10,000$, respectively.

Figs. 8 and 9 reveal that generally, coverage rate of a P2P worm in a logical P2P overlay network will decrease if quarantine rate is increased. Figs. 8 and 9 also show that maximum topology out-degree has no significant impact on attack performance and defense effort when it is in the range 100–1000 inclusive, and that 5000 peers will not generate significantly different results. The above findings are the same as those from random quarantine.

If we compare the bottom curve in Fig. 6 to the bottom curve in Fig. 8, it can be found that larger topology out-degree priority quarantine demands a lower defense effort (quarantine rate $q = 0.16$) to achieve a lower attack performance (coverage rate $c < 0.1$), and that random quarantine demands a higher defense effort (quarantine rate $q = 0.4$) to achieve a higher attack performance (coverage rate $c < 0.2$). The same result as above can be found if we compare the bottom curve in Fig. 7 to the bottom curve in Fig. 9.

The above finding exactly serves our paramount objective, which is to find a quarantine tactic whose enforcement will lead to a lower attack performance at a lower cost of defense effort.

Therefore, according to our experimental results, larger topology out-degree priority quarantine outperforms random quarantine. Larger topology out-degree priority quarantine is exactly the quarantine tactic we are looking for since it demands only 40% (0.16/0.4) defense effort to achieve 50% (0.1/0.2) attack performance, compared to random quarantine. In other words, larger topology out-degree priority quarantine is much more efficient than random quarantine.

5. Conclusion

This paper presents a study on modeling the propagation processes of P2P worms. In this paper, based on our definitions of logic matrix and its operations, we have proposed the logic matrix representation of a P2P overlay network's topology, topology out-degree, topology in-degree, state, vulnerability status, and quarantine status; and derived our unique logic matrix approach to modeling the propagation of P2P worms. Based on this model, we find the impacts of the two different topologies on a P2P worm's attack performance, the impacts of the network-related characteristics on a P2P worm's attack performance in structured P2P networks, and the impacts of the two different quarantine tactics on the propagation characteristics of P2P worms in unstructured P2P networks.

To the best of our knowledge, we are the first using logic matrix in network security research in general and worm propagation research in particular. The proposed approach's ease of employment makes it an attractive instrument to conduct worm propagation research. We have demonstrated the innovative logic matrix formulation proposed in this paper, which are discrete time deterministic propagation models of P2P worms described by difference equations of logic matrix, is a highly effective and efficient tool for investigating the propagation processes of P2P worms.

In the future, we plan to extend P2P worm models by considering temporal issues, such as the time lag for worms to infect peers and the time spent in quarantining peers. Time series-based matrices can be potentially used in the extended model. We will also look for a more effective and efficient quarantine tactic if temporal issues are considered.

To make it more practical to accommodate the dynamic P2P network where peers can join and leave a network, a P2P network's topology logic matrix needs to include peers that will join the network in the future. It also needs to be updated once a peer joins or leaves the network. This could be simulated by randomly selecting peers joining and leaving the network, which means the topology logic matrix of the P2P network is constantly changing. We are going to incorporate the above idea into the simulation conducted previously.
References


Xiang Fan is currently with School of Management and Information Systems, Central Queensland University. His research interests include network security in general and propagation of active worms in particular. He is currently working in a research group developing active defense systems against large-scale network attacks and new Internet security countermeasures.

Dr. Yang Xiang is currently with School of Information Technology, Deakin University, Australia. His research interests include network and system security, and wireless systems. In particular, he is currently working in a research group developing active defense systems against large-scale network attacks and new Internet security countermeasures. He has served as PC Chair for the 2009 3rd International Conference on Network and System Security (NSS 2009), the 11th IEEE International Conference on High Performance Computing and Communications (HPCC 2009), and the 14th IEEE International Conference on Parallel and Distributed Systems (ICPADS 2008). He has been PC member for many international conferences such as IEEE ICC, IEEE GLOBECOM and IEEE ICPADS. He is on the editorial board of Journal of Network and Computer Applications.