Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Propagation model of smartphone worms based on semi-Markov process and social relationship graph

Sancheng Peng [a,b], Min Wu [a], Guojun Wang [a,*], Shui Yu [c]

[a] School of Information Science and Engineering, Central South University, Changsha 410083, China
[b] School of Computer Science, Zhaoqing University, Zhaoqing 526061, China
[c] School of Information Technology, Deakin University, 221 Burwood HWY, Burwood, VIC 3125, Australia

ARTICLE INFO

Article history:
Received 22 August 2013
Received in revised form
11 March 2014
Accepted 15 April 2014

Keywords:
Smartphone
Worm
Propagation model
Semi-Markov process
Social relationship graph

ABSTRACT

Smartphone applications are getting more and more popular and pervasive in our daily life, and are also attractive to malware writers due to their limited computing source and vulnerabilities. At the same time, we possess limited understanding of our opponents in cyberspace. In this paper, we investigate the propagation model of SMS/MMS-based worms through integrating semi-Markov process and social relationship graph. In our modeling, we use semi-Markov process to characterize state transition among mobile nodes, and hire social network theory, a missing element in many previous works, to enhance the proposed mobile malware propagation model. In order to evaluate the proposed models, we have developed a specific software, and collected a large scale real-world data for this purpose. The extensive experiments indicate that the proposed models and algorithms are effective and practical.

1. Introduction

Smartphones integrate the communication capabilities of cell-phones with personal digital assistant (PDA) functions, which enables them running more and more applications, such as emails, web services, simple messaging service/multimedia messaging service (SMS/MMS), and online shopping (Peng et al., 2013a, 2013b; Peng, 2013). In addition, an application-based interface is used in most smartphones. It enables users to download individual programs that can perform a variety of tasks. However, the significant development and pervasive usage of smartphones also attract worm writers to target on their vulnerabilities to commit their malicious goals (Jamaluddin et al., 2004).

According to the recent security reports (Gao and Liu, 2011; Felt et al., 2011; Shih et al., 2008; Peng et al., 2014), the number of malicious exploits and executed attacks have gone through a surge in recent years. In 2010, more than 1 million cell phones in China were infected by the 'Zombie' virus, which can automatically send text messages, and the attack resulted in a loss of combined $300,000 per day. Juniper Networks Mobile Threat Center (MTC) released its 2011 Mobile Threats Report in February 2012, which showed that mobile malware increased 155% across all platforms compared to the previous year, and provided the evidence of a new level of maturity in security threats targeting on mobile devices.

* Corresponding author.
E-mail address: csgjwang@csu.edu.cn (G. Wang).
http://dx.doi.org/10.1016/j.cose.2014.04.006

However, few smartphones have been designed to guard against worm attacks, making them enticing targets for worm writers and hackers. If smartphones have been compromised by worms, it could cause their services interrupted, such as data loss, system damage, privacy leakage, and financial loss.

Mobile malware can spread through both short-range radio (i.e., Bluetooth and WiFi) and long-range communication (i.e., SMS, MMS and email). The malware of Bluetooth-based and WiFi-based is also called proximity malware (Li et al., 2010). The first Bluetooth-based worm is Cabir (Polla et al., 2013), which propagates through Bluetooth connections to other Bluetooth-enabled devices that it can find. The first SMS/MMS-based worm is Commwarrior (Xia et al., 2008), which searches through a user's local address book for phone numbers and then sends SMS/MMS messages containing infected files to the users in the address book.

Many efforts have been recently made to model worm propagation among smartphones, such as Bluetooth-based (Peng and Wang, 2011), and SMS/MMS-based (Van Ruitenbeek et al., 2007; Fleizach et al., 2007). However, most models depend almost entirely on the technology of differential equations, and fail to take into account the perspectives of individual worm difference on propagation dynamics, and seldom consider the impact from social relationships (Zhu et al., 2009) of smartphone users either. It is complex and challenging to establish practical and appropriate models on worm propagation dynamics.

In our previous work (Peng and Wang, 2011; Peng et al., 2013b), we have proposed two models to characterize the propagation of Bluetooth-based worms. In these two models, we have exploited cellular automata (CA) to describe the spatial-temporal process of worm spreading. Due to the difference between Bluetooth and SMS/MMS, our previous models are difficult to characterize the propagation dynamics of SMS/MMS-based worms. Moreover, because of the advantages of SMS/MMS, such as convenience, shortcut, inexpensiveness, and privateness, SMS/MMS have become a common means of communication in people's everyday lives. Many people often make their social interactions by sending SMS/MMS messages, which increases chances to worms. Therefore, it is further motivated to deal with the modeling and analysis for SMS/MMS-based worm propagation in this paper.

In this paper, based on the models presented in references (Zhu et al., 2009) and (Peng et al., 2013a), we propose a new mechanism to characterize the propagation dynamics of SMS/MMS-based worms. In our mechanism, we combine two techniques to tackle the problem. First of all, we employ the semi-Markov process (Hou et al., 2005; Peng et al., 2011; Xing and Wang, 2008; Hao et al., 2011; Chen et al., 2011) to model node behaviors, and then exploit smartphone social networks to characterize mobile worms that spread through SMS or MMS. We believe that social network theory (Zhu et al., 2009; Peng et al., 2013a) can be useful to simulate this kind of networks, because the relationships between any two nodes could be capable of modifying all network relationships, and worm propagation depends heavily on the social contact of the responding smartphones. This characteristic is similar to what have been found in many dynamic systems, which are generally simulated through social network graph. The contributions of this paper are summarized as follows.

- We present a node-state transition model to describe the complexity of worm and uncertainty of worm propagation according to the semi-Markov process, and provide an effective empirical derivation to further show how to combine social interaction and semi-Markov process, by using the theoretical analysis of limiting probability.
- We construct a social relationship graph of mobile devices by extracting their communication patterns. This graph describes the social relationships between any two smartphones, which are usually exploited by SMS/MMS worms for spreading.
- We propose a detailed analytical model to characterize the propagation dynamics of SMS/MMS-based worms using the social relationship graph. Moreover, we consider the following realistic modeling assumptions: 1) the infected factor of infectious devices is different, and 2) the resisted factor of each device for worm spread is different.
- We evaluate the performance of our solution using a customized program based on messaging records collected from real cellular networks. Through extensive numerical simulations and analysis, we confirm that our strategies can characterize the propagation of mobile worms more effectively than the traditional propagation models.

The remainder of this paper is structured as follows. In Section 2, we provide an overview of related work, and present the construction of social relationship graph in Section 3. In Section 4, we discuss the modeling issue of node-state transition, and establish a worm modeling scheme with social relationship graph in Section 5, and provide the results of the model validation in Section 6. Finally, we conclude this paper in Section 7.

2. Related work

There are multiple infection vectors that are able to deliver malicious content to Smartphones, such as SMS/MMS, Bluetooth, Internet access, and file duplication with USB. However, the existing models focus on SMS/MMS and Bluetooth. As a result, we present the related work in three dimensions. The first dimension is the worm propagation models, the second one is the social relationship graph and semi-Markov process, and the last part is the discussion and comparison on the existing worm propagation models.

2.1. *Worm propagation model*

Su et al. (2006) investigated whether a large-scale Bluetooth worm outbreak was viable or not in practice, and used trace-driven simulations to examine the propagation dynamics of Bluetooth worms. They found that Bluetooth worms could infect a large population in just a few days. However, the authors did not consider the effect of the real-world social interactions on the propagation dynamics of Bluetooth worms.

Zheng et al. (2006) focused on modeling population distribution density, Bluetooth radius, and node velocity. They

pointed out a variety of quarantine methods that could greatly reduce the potential poisonousness. But the authors did not consider the impact of individual difference on the propagation dynamics of different worms, and did not characterize the effect of the real-world social interactions on the propagation dynamics of Bluetooth worms.

Yan and Eidenbenz (2009) presented a model to study the spread of Bluetooth worms and investigated the impact of mobility patterns on Bluetooth worm propagation. In their proposed model, the impact of mobility patterns on Bluetooth worm propagation can be investigated by introducing the input parameters, such as average node degree, average node meeting rate, and the link duration distribution. However, it is difficult to be applied this model to analyze the propagation of SMS/MMS worms.

Rhodes and Nekovee (2008) provided the effect of population characteristics and device behavior on the outbreak dynamics of Bluetooth worms. However, they did not consider the impact of individual difference on propagation dynamics of different worms.

Martin et al. (2010) predicted the spreading of cell phone viruses using the SIS model from mathematical epidemiology. However, the authors did not take into account the impact of individual difference on the propagation dynamics of proximity-based viruses.

Peng and Wang (2011) proposed a worm propagation modeling scheme (WPM), which utilized the two-dimensional (2D) cellular automata to simulate the dynamics of worm propagation process from a single node to the entire Bluetooth network. Although the WPM scheme integrates infection factor and resistance factor, it fails to provide specific expressions to compute these two factors.

Van Ruitenbeek et al. (2007) presented a mechanisms to analyze the effects of multimedia messaging system (MMS) viruses that spread by sending infected messages to other phones. However, the authors did not provide the impact of social interactions on the propagation dynamics of MMS-based viruses.

Fleizach et al. (2007) designed an event-based simulator to evaluate the effects of malware propagating using communication services like VOIP or MMS in mobile phone networks. The authors used the US census data and estimated address book degree distribution, but did not use real traffic data in their worm propagation modeling examinations.

2.2. Social relationship graph and semi-Markov process

Zhu et al. (2009) constructed a social relationship graph of mobile devices by extracting their communication patterns based on a network trace. This graph characterizes the social relationships among mobile phones that are usually utilized by mobile worms for propagation.

Xing and Wang (2008) proposed a model to characterize the evolution of node misbehavior using semi-Markov process (SMP) in 2D mobile ad hoc networks. Peng et al. (2011) presented another model to describe the evolution of node misbehavior using SMP in 3D mobile ad hoc networks. However, it is difficult to use these two models to analyze the node-state transition for mobile worms.

2.3. Discussion and comparison

A comparison of worm propagation models is listed in Table 1. From this table, it is known that differential equations are widely used to model worm propagation in existing work. That is, most worm propagation models are based on deterministic models, and only a small number are based on stochastic and spatial-temporal models.

Although most previous work can provide some valuable insight into the characteristics and dynamics of worm propagation, the models based on differential equations fail to capture the local characteristics of spreading processes, nor do they include interaction behaviors among individuals. Furthermore, in our previous work, the individual difference has been introduced in our worm propagation models. However, we have not provided the related empirical derivation.

Moreover, as to the existing works on social relationship graph, these works do not use this graph to analyze the propagation dynamics of MMS-based worms, but to prevent the spreading of MMS-based worms; as to the existing works on modeling of node misbehaviors, these works do not either use semi-Markov process to model node behaviors caused by worms.

Therefore, in this paper, we introduce semi-Markov process to model node behaviors in mobile social network, and then provide a theoretical analysis of limiting probability for our node-state transition model. In addition, we construct a social relationship graph to characterize propagation dynamics of SMS/MMS-based worms.

3. Construction of social relationship graph

Cellular services, such as SMS and MMS (Peng et al., 2013a; Zhu et al., 2009), can be used as attack vectors for smartphones. For example, SMS and MMS messages can be used to deliver malicious content, or to maintain communication

Table 1 – Comparison of worm propagation models.

Worm propagation model	*Modeling theory*	*Malware type*	*Individual difference*	*Modeling on social interaction*
Su et al., 2006	NA	Bluetooth worm	No	No
Zheng et al., 2006	Differential equations	Bluetooth worm	No	No
Yan and Eidenbenz, 2009	Differential equations	Bluetooth worm	No	No
Rhodes and Nekovee, 2008	Differential equations	Bluetooth worm	No	No
Martin et al., 2010	Differential equations	Bluetooth worm	No	No
Peng and Wang, 2011	Cellular automata	Bluetooth worm	Yes	No
Van Ruitenbeek et al., 2007	NA	SMS/MMS worm	No	No
Fleizach et al., 2007	Stochastic process	SMS/MMS worm	No	No

with the attacker. In regard to SMS, the attacker uses SMS to send URL links, then the user could be lured to open a browser using these URLs. In term of MMS, the message itself could be the malicious payload. A victim receiving this message will most likely open and download the message, since he/she believes it comes from someone he/she knows and trusts. Thus, the social relationship between any two mobile users in a cellular network should be considered for an effective SMS/MMS-based worm propagation model. By analyzing the social interactions among smartphones, i.e., which devices are more likely to exchange messages with others, the propagation path of such worm can be predicted.

If Alice and Bob exchange messages with each other on a regular basis, Alice would in a high probability to download and open the messages from Bob, because they trust each other and wouldn't doubt the credibility of the messages. However, if Alice's smartphone has been infected by SMS/MMS-based worms, and she accidentally sends this message to Bob, Bob's smartphone is prone to be infected by these worms. The fact is that if they have never sent messages to each other, Bob's smartphone may not get infected despite the fact that Alice's smartphone has been infected. In this paper, we therefore exploit SMS and MMS records among smartphones to predict the propagation path of worm.

Social relationship graph is represented by a directed weighted graph, $G(V, E, W)$, where set V of vertices corresponds to the cellular phones in cellular networks, set E of directed edges corresponds to the traffic flow among cellular phones i to j, and set W of weight values corresponds to the total number of SMS messages sent from cellular phone i to j. The degree of vertex i, d_i, is the number of smartphones that its owner communicates with. The amount of messages initiated from i to j is denoted by C_{ij}.

We introduce two functions $f(i)$ and $g(i,j)$ to map each vertex i ($i \in V$) and each edge (i, j) ($(i, j) \in E$), respectively. Thus, the graph can be weighted with $f(i)$ and $g(i, j)$ determining the weights of vertex and edge, respectively. The mapping functions of the weights of vertex and edge are listed as follows.

$$f(i) = d_i \tag{1}$$

$$g(i,j) = C_{ij} + C_{ji} \tag{2}$$

Both of the weights of vertices and edges contribute to a significance parameter, which represents the probability of being infected by worms. Based on Equation (1), we know the weights of vertices depend on their degrees. For SMS/MMS-based worm, a smartphone with a high in-degree means that it is more likely to be infected, while a smartphone with a high out-degree is more likely to infect other smartphones. Therefore, those smartphones, either with high in-degree or out-degree, should be assigned a high vertex weight.

From the real-world data set, which is collected from one of the largest cellular networks in China. We can extract a smartphone social network. The network is huge and complex. In order to explain the idea of smartphone social network, we take ten users from the data set and use them as an example, and the data of this sample social network is listed in Table 2.

Table 2 – The number of interactions in a week between any two cellular phone users.

Between two smartphones	*The number of interactions*
A → B	2
A → D	5
A → E	15
A → G	7
B → A	3
B → C	8
B → G	8
C → B	6
C → G	10
C → J	12
D → A	2
D → E	4
D → F	6
E → A	6
E → D	9
E → F	3
E → G	4
E → H	13
F → D	4
F → E	8
F → H	12
F → I	3
G → A	9
G → B	1
G → C	0
G → E	8
G → H	14
G → J	5
H → E	5
H → F	8
H → G	9
H → I	6
H → J	6
I → F	7
I → H	0
I → J	6
J → C	13
J → G	9
J → H	20
J → I	0

According to Table 2, we abstract each cellular phone as a vertex, a weighted social relationship graph can be obtained accordingly, which is shown in Fig. 1.

4. Model of node-state transition

In this section, based on the classification of node states, modeling on node-state transition is proposed using the semi-Markov process.

4.1. *Classification of node states*

According to the spread property of SMS/MMS-based worms, the epidemic states of a node are as follows (Fan et al., 2010): susceptible (S), exposed (E), infected (I), and recovered (R).

1) Susceptible state (S): nodes have not been infected by any worm in the network but are prone to infection.

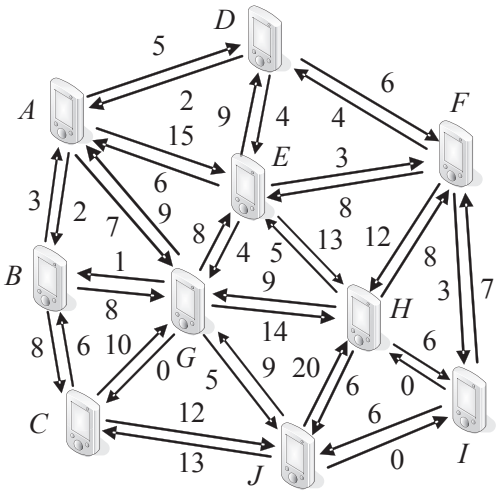

Fig. 1 – Social relationship graph of ten smartphone users from the real-world data set in a week.

2) Exposed state (E): nodes have been infected by worm but have not spread the worm to the susceptible smartphones while transmitting data or controlling the messages sent to the phones for the time being.
3) Infectious state (I): nodes have been infected by worms in the network and they may infect nodes in state S.
4) Recovered state (R): nodes that used to be infected by worms and then recovered from the infection. Those nodes are cleaned and immune to the same type of cleaned worms.

We note the number of susceptible, exposed, infectious, and recovered nodes at time t as $S(t)$, $E(t)$, $I(t)$, and $R(t)$, respectively. Details of some other terms of the model are further explained as follows.

- N denotes the total number of susceptible, exposed, infectious, and recovered individuals. Let the birth rate not be equal to the death rate, the total population size is a variable. Thus, $S(t) + E(t) + I(t) + R(t) = N$.
- p_{XY} denotes the probability of a node transform from state X to state Y, where X, Y are in set $\{S, E, I, R\}$. For example, p_{SE} represents the probability of a node transform from state S to state E.

4.2. *Modeling on node-state transition using semi-Markov process*

Simple Markov chain is difficult to evaluate node behavior in social networks. The reasons are as follows.

1) The transition time from one state to another is a random variable.
2) It is very difficult to characterize the transition time using exponential distributions. For instance, a node is more inclined to exposed due to weak safety awareness as time passes, and the less proper secure measure adopted, the more likely a node changes its state to infected.
3) System responses to the attacks, which causes the sojourn time of some states to be non-exponential. The further action of a node may depend on how long it has been in the current state and transition intervals may have arbitrary distributions.

Thus, the simple Markov chain may not be used to describe node behaviors, especially in the presence of node misbehaviors. In this paper, the semi-Markov process (Peng et al., 2011; Xing and Wang, 2008) is used to characterize the node-state transition.

A node may change its states as follows.

1) A susceptible node becomes recovered for the sake of installing anti-worm software. It is also prone to be an exposed node due to many reasons, e.g., fail to install anti-worm software, download application software from the Internet.
2) An infected node could become susceptible or recovered by means of installing anti-worm software.
3) An exposed node could become susceptible or recovered through proper secure measure, and could also become infected due to outbreak of worms.
4) A recovered node could become susceptible again if a new worm presents.

However, all these events occur randomly and do not follow exponential distributions. Thus, the above assumptions for the node-state transition are difficult to model node behaviors using simple Markov chain, but are simple enough to exploit using semi-Markov process, which is provided as follows.

A stochastic process $\{X(t), t \geq 0\}$ is a semi-Markov process (SMP), which is described by:

$$X(t) = Z_n, \forall t_n \leq t \leq t_{n+1}, \quad n = 0, 1, \cdots, L \tag{3}$$

where $\{X(t), t \geq 0\}$ denotes the transitions of the process, occur at epochs (or instants of times) $t_0, t_1, \ldots, t_n$ ($t_0 < t_1 < \ldots < t_n$); $\{Z_n, n = 0, 1 \ldots L\}$ denotes the state of system at epoch t_n, so that the Markov property is satisfied at each time epoch t_n; The pair $\{X_n, t_n\}$ forms a Markov renewal sequence on state space $W = \{S, E, I, R\}$; and $\{Z_n, n = 0, 1 \ldots L\}$ constitutes a discrete time Markov chain (DTMC) with state space W.

Here, the Markov chain $\{Z_n, n = 0, 1, \ldots L\}$ is said to be an embedded discrete time Markov chain (DTMC) of the SMP $\{X(t), t \geq 0\}$. Thus, the transition probability matrix of $\{Z_n\}$ is described by

$$P = \begin{pmatrix} 0 & p_{SE} & p_{SI} & p_{SR} \\ p_{ES} & 0 & p_{EI} & p_{ER} \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{4}$$

where $p_{ii} = 0$ represents that $\{Z_n\}$ only has transitions from a state to another different state. In Equation (4), if a transition probability equals to zero, e.g., $p_{IE} = 0$, $p_{RE} = 0$, that means infectious node will not become exposed, and recovered node will not become exposed. Moreover, since the summation of transition probabilities of a state must be equal to 1 in a stochastic matrix, then $p_{IR} = 1$, $p_{RS} = 1$.

The transition from one state to another in an SMP can be denoted by two matrices, $P = (p_{ij})$ and $F(t) = (F_{ij}(t))$, where p_{ij} is the transition probability of node behavior from state i to state j, $F_{ij}(t)$ is the distribution function of time spent from state i to

j. The transitions among states of worm propagation are illustrated in Fig. 2.

The SMP model defined above enables us to consider the evolution process of node state without the assumption of memoryless property. In addition, this model can be used to describe a wide variety of different threats caused by node misbehavior. Let $T_n = t_{n+1} - t_n$ be the sojourn time between n-th and $(n+1)$-th transition, the associated (time-homogeneous) semi-Markov kernel $Q = (Q_{ij}(t))$ can be defined by

$$Q_{ij}(t) = P\{Z_{n+1} = j | Z_n = i\} = p_{ij}F_{ij}(t) \tag{5}$$

where $p_{ij} = \lim_{t\to\infty} Q_{ij}(t) = P\{Z_{n+1} = j | Z_n = i\}$ is the transition probability between states i and j, and $F_{ij}(t) = P\{Z_{n+1} = j, Z_n = i | T_n \le t\}$ is the transition probability matrix of $\{Z_n\}$.

To analyze an SMP, the parameters are to be dealt with as follows.

- $M(i)$ represents mean sojourn time in state $i (i \in W)$.
- p_{ij} represents the transition probabilities between different states i and j.
- $M(i,j)$ represents mean transition time from state i to j.
- π_i represents the probability of each state in steady-state system.
- $E[T_i]$ and $E[T_{ij}]$ represent the conventional notation for exception.

Let T_i be the sojourn time in state i, and T_{ij} be the transition time from states i to j, then the mean sojourn time is expressed as follows.

$$M(i) = E[T_i | Z_n = i] = E[T_i] \tag{6}$$

$$M(i,j) = E[T_{ij} | Z_n = i, Z_{n+1} = j] = E[T_{ij}] \tag{7}$$

$$E[T_i] = \int_0^\infty (1 - F_i(t))dt \tag{8}$$

$$E[T_{ij}] = \int_0^\infty (1 - F_{ij}(t))dt \tag{9}$$

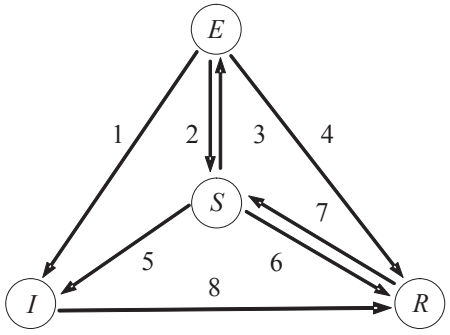

Fig. 2 – State transition relationship for worm propagation.

$$M(i) = \sum_{j\in W} p_{ij}M(i,j) \tag{10}$$

Theorem 1. Given SMP $\{X(t), t \ge 0\}$ associate with state space W and DTMC defined by Equation (5), the transient distribution $P_{ij}(t)$ converges to P_j as $t \to \infty$, where P_j can be calculated by

$$P_j = \lim_{t\to\infty} P_{ij}(t) = \frac{\pi_j M(j)}{\sum_{k\in W}\pi_k M(k)} \tag{11}$$

Proof. Let $T_j(i)$ be the sojourn time in state Z_j for the i-th time ($i, j \ge 0$), and $N_j(k)$ be the arrival time points at state Z_j among the previous k times of transition in a semi-Markov process (Hou et al., 2005; Corradi et al., 2004) . Thus, the time ratio of the arrival times can be obtained at state Z_j among the previous k times of transition as follows.

$$P_{j=k} = \frac{\sum_{i=1}^{N_j(k)} T_j(i)}{\sum_{j=0}^{\infty}\sum_{i=1}^{N_j(k)} T_j(i)} = \frac{\frac{N_j(k)}{k}\sum_{i=1}^{N_j(k)}\frac{T_j(i)}{N_j(k)}}{\sum_{j=0}^{\infty}\frac{N_j(k)}{k}\sum_{i=1}^{N_j(k)}\frac{T_j(i)}{N_j(k)}} \tag{12}$$

When $k \to \infty$, $N_j(k) \to \infty$, by the strong law of large numbers, we have

$$\sum_{i=1}^{N_j(k)} \frac{T_j(i)}{N_j(k)} \to M(j) \tag{13}$$

In addition, let u be the number of transitions between two arrivals at state j. By the central limit theorem of a renewal process, we also have

$$\frac{N_j(k)}{k} \to (E[u])^{-1} = \pi_j \tag{14}$$

Thus, let $k \to \infty$ in Equation (12), we obtain Equation (11). This proves the theorem.

According to the theory of SMP, the probabilities of steady-state are obtained based on the following Equations.

$$\pi \cdot P = \pi \tag{15}$$

$$\sum_{i\in W} \pi_i = 1 \tag{16}$$

To compute P_j, we need to evaluate p_{ij} and $M(i,j)$. Corradi et al. have provided a specific case in Corradi et al. (2004) to show how to evaluate them by using the statistics from measurements.

□

4.3. *Theoretical analysis of limiting probability*

In Subsection 4.2, we have proposed a semi-Markov node behavior model and provided Equation (11) as the solution for the limiting state probability (P_j for $j \in W = \{S, E, I, R\}$). Even with this equation, calculating P_j is a non-trivial task due to the difficulty in determining the transition probabilities p_{ij} and expected sojourn times $M(i)$. Due to the dependency on specific application scenarios of these parameters, we establish an example network and incorporate the following policies in our theoretical analysis.

- Each node has the same initial healthy index H, and may infect a worm once its residual healthy index is below a threshold ξ.
- Each node changes its state from susceptible to exposed when its health index below a threshold ξ_E; each node changes its state from exposed to infected when its health index below a threshold ξ_I; and each node changes its state from infected to recovered when its health index above a threshold ξ_R ($0 < \xi_I < \xi_E < \xi_R < 1$).
- Each node encounters a damage index δ ($0 < \delta < 1$) when they communicate with each other by SMS/MMS in a time unit; and each node has a repairing index θ ($0 < \theta < 1$) in a time unit.
- The time that any node resides in the network (called residence time) is random, depending on the behavior of smartphone users, but with a finite expected value T_{in}.
- At last, we suppose an average recovery time T_{re}, so that recovered nodes can become susceptible again (e.g., if a new worm presents when they rejoin the network).

It is known that the larger the number of interactions between any two nodes is, the larger the damage index is. Thus, δ is described by

$$\delta = \frac{1}{e^{\ln(1/TM+1)}} \tag{17}$$

where TM represents the total number of messages records sent from its friend nodes to node i, and it denoted by $\sum_{j=1}^{N_i} C_{ji}$.

Since no heuristic is known so far to provide an analytical solution of transition probabilities on worm propagation, in this method, we estimate expected transition times $M(i, j)$ using following heuristics in the above scenario. Thus, to estimate $M(S, E)$, it is noticed that a susceptible node become exposed if its residual healthy index is below $(1 - \xi_E)H$ according to the network scenario settings aforementioned. We can have $M(S, E)$ approximated by

$$M(S,E) = \frac{(1-\xi_E)H}{\delta} \tag{18}$$

Similarly, $M(S, I)$, $M(E, S)$ and $M(E, I)$ can be bounded by

$$M(S,I) = \frac{(1-\xi_I)H}{\delta} \tag{19}$$

$$M(E,S) = \frac{(1-\xi_E)H}{\theta} \tag{20}$$

$$M(E,I) = \frac{(\xi_E-\xi_I)H}{\delta} \tag{21}$$

$$M(I,R) = \frac{(\xi_R-\xi_I)H}{\theta} \tag{22}$$

In general, any node changes its state from other states to recovered after an average period $\overline{T_{in}}$. Thus, $M(S, R)$ and $M(E, R)$ can be bounded by

$$M(S,R) = \overline{T_{in}} \tag{23}$$

$$M(E,R) = \overline{T_{in}} \tag{24}$$

At last, any node changes its state from recovered states to susceptible after an average period $\overline{T_{re}}$. Thus, $M(R, S)$ can be bounded by

$$M(R,S) = \overline{T_{re}} \tag{25}$$

Thus, the transition expected time matrix of Z_n is described by

$$Q = \begin{Bmatrix} 0 & M(S,E) & M(S,I) & M(S,R) \\ M(E,S) & 0 & M(E,I) & M(E,R) \\ 0 & 0 & 0 & M(I,R) \\ M(R,S) & 0 & 0 & 0 \end{Bmatrix} = \begin{Bmatrix} 0 & \frac{(1-\xi_E)H}{\delta} & \frac{(1-\xi_I)H}{\delta} & \overline{T_{in}} \\ \frac{(1-\xi_E)H}{\theta} & 0 & \frac{(\xi_E-\xi_I)H}{\delta} & \overline{T_{in}} \\ 0 & 0 & 0 & \frac{(\xi_R-\xi_I)H}{\theta} \\ \overline{T_{re}} & 0 & 0 & 0 \end{Bmatrix} \tag{26}$$

Moreover, we use the method proposed in Corradi et al. (2004) to determine p_{ij} as follows: given the time period [0, t], record the total number of transitions from state i to all other states k and denote it by N_{ik} ($i, k \in W$, $i \neq k$), then, p_{ij} is approximated by

$$p_{ij} = N_{ij} \Big/ \sum_{k \in W} N_{ik} \tag{27}$$

Remark: Although the accuracy of using the heuristic method depends on the soundness of Equations (18)–(27), the proposed approach provides us a method to analyze the effect of a specific dynamic factor, such as individual difference, social relationship, stochastic property of node behaviors. Since we have known that the worm propagation also depends on the state distribution of node behavior, the limiting probability plays an important role in bridging the gap between the worm propagation and any specific dynamic that directly affects the limiting probability. Thus, the node-state transition model based on semi-Markov process proposed in this paper, not only provides us a general mathematical framework to characterize node behavior, but also provides a theoretical basis on the setting of relevant parameters during our simulation, via the node-state distribution.

5. Modeling on worm propagation

In this section, we discuss how a SMS/MMS-based worm propagation model can be constructed with social relationship graph.

5.1. *Description on difference of individuals*

In this model, three parameters are introduced to reasonably describe the infection ability of worms in smartphone based social relationship graphs. Moreover, some variables and terms are made as follows.

- C_{ji} denotes the number of message records sent from j to i in a week.
- $\max\{N_u\}$ ($u \in N$) denotes the maximum total number of friends for all of nodes.

- α_i ($0 \le \alpha_i \le 1$) denotes the resistance intensity of node i.
- β_i ($0 \le \beta_i \le 1$) denotes the safety awareness of node i.
- N_i denotes the number of neighbor nodes for node i.
- SA_j denotes the social ability of node j, and $SA_j = \ln(N_j + 1)/\ln((\max\{N_u\}) + 1)$.

Three factors ID_i, IF_{ji} and RF_{ij} are described as follows.

The first one is infected factor, denoted by IF_{ji}, which represents infection degree from node j to node i ($0 \le IF_{ji} \le 1$). If IF_{ji} equals to 0, it means that the node has no infection to node i. If IF_{ji} equals to 1, it indicates that node j has a strong infection to node i. For example, if the resistance intensity of node i is strong and the social ability of node j is weak, IF_{ji} is small. Otherwise, IF_{ji} is large. Thus, IF_{ji} is given by.

$$IF_{ji} = 1 - \frac{\alpha_i}{e^{SA_j \times C_{ji}}} \tag{28}$$

The second one is resisted factor, denoted by RF_{ij}, which represents resistance degree of node i on infection from node j ($0 \le RF_{ij} \le 1$). If RF_{ij} equals to 1, it means that node i has a strong ability to resist infection. For example, if the safety awareness of node i is strong and the number of interactions between node i and j is small, RF_{ij} is large. Otherwise, RF_{ij} is small. Thus, RF_{ij} is given by.

$$RF_{ij} = \frac{\beta_i}{1 + \ln(C_{ij} + 1)} \tag{29}$$

The last one is infection degree, denoted by ID_i, which is used to measure the danger level from an infected smartphone to susceptible smartphones. ID_i is given by.

$$ID_i = \sum_{j=1}^{N_i} \frac{IF_{ji}}{IF_{ji} + RF_{ij}} \tag{30}$$

To assure that the value of ID_i is between 0 and 1, ID_i is normalized as follows.

$$ID_i^* = \frac{ID_i - \min\limits_{k \in N}\{ID_k\}}{\max\limits_{k \in N}\{ID_k\} - \min\limits_{k \in N}\{ID_k\}} \tag{31}$$

where $\max\{ID_k\}$ and $\min\{ID_k\}$ represent the maximum and minimum of ID among all nodes, respectively.

The social interactions between any two smartphones can be described by Equation (26). If two smartphones have communicated through SMS or MMS, they are more likely to open and activate a worm-infected message from each other. This social relationship graph provides us an overview of how smartphones connect with each other, and how worms propagate themselves by exploiting the relationship. The relation between the number of friends, the total number of interactions and ID is analyzed in Section 6.

5.2. Worm propagation model

According to the characteristics of SMS/MMS worm propagation, and the social relationship graph, we design an algorithm corresponding to state transitions. It is used to characterize the process of SMS/MMS-based worm propagation. Let T be the transmission threshold through which a node i transforms from state S to the other states. The state transition algorithm is presented as Algorithm 1.

Algorithm 1. State transition of node.

```
Network initialization. All nodes N communicate with each other using SMS/MMS;
Node state initialization. A node or some nodes is/are randomly selected, and its sate or their states
  is/are set to be state I, and the states of other nodes are set to be state S;
The information of its/theirs friends can be collected by counting the messaging records;
Node i is accessed at time t, thus
while i ≤ N do
  if (Its state is I) then
    Its neighbor nodes are accessed;
    while j ≤ N_i do
      if (The state of its friend node j is S) and (ID_j is not smaller than T) then
        Node j changes its state from S to E with probability p_SE, or from S to I with probability
        p_SI;
      else
        Node j remains in its previous state;
      end if
      if (IF_ij equals to 0) or (RF_ji equals to 1) then
        Node j changes its state from S to R with probability p_SR;
      end if
    end while
    Node i changes its state from I to R with probability p_IR;
  end if
  if (Its state is E) then
    Node i changes its state from E to R with probability p_ER, or node i changes its state from E to I
    with probability p_EI;
  end if
end while
t equals to t plus Δt;
```

Table 3 – Related parameters for the evaluation of limiting probabilities.

Parameter	H	ξ_E	ξ_I	ξ_R	δ	$\overline{T_{in}}$	$\overline{T_{re}}$
Value	120	0.3	0.15	0.3	0.3	25	20

Table 4 – Related parameters for the evaluation of propagation model.

Parameter	p_{SE}	p_{EI}	p_{ER}	p_{SI}	p_{SR}	p_{IR}	T
Value	0.54	0.54	0.06	0.05	0.12	0.06	0.21

6. Performance evaluation

To demonstrate the existence of limiting probabilities, related parameters are set as the values in Table 3. The limiting probability with respect to the total number of messages records sent from its friend nodes to a node is illustrated in Fig. 3. From this figure, we can see clearly that P_S keeps attenuating, P_I, P_R, and P_E keep mounting, as TM increases; and finally they bound to the limiting values after $TM = 10$. This observation shows the existence of the limiting probabilities and the soundness of the heuristics analysis.

Moreover, to evaluate the feasibility of using social relationship graph to simulate SMS/MMS-based human contact network, and to verify the rationality and effectiveness of the proposed model of worm propagation in smartphones, we collected a large scale real-world data set of smartphone communication from a main mobile service company of China. The data set includes 400,000 mobile users, and 20,000,000 SMS/MMS messages of the user space for three weeks. For security and privacy protection purposes, the content of the messages was deleted, while the uniqueness of the identifiers of involved phone numbers are replaced by pseudo code. In addition, we investigated how a social relationship graph was constructed using the real-world data set.

To conduct the performance evaluation, we developed a specific C++ program to implement our proposed algorithm, which is an extension of the proposed models based on the two categories. Due to the huge scale of the real-world data set, we take only 5114 users in our experiments, rather than including all the users. In the following experiments, we set the parameters as shown in Table 4.

Firstly, we carry out experiments to verify Equations (28)–(31). Fig. 4 shows the relation between the number of friends, the total number of interactions with its friends and ID. We know that the larger the number of friends and the larger the total number of interactions with its friends there are, the larger ID will be. That is, the larger the number of friends and the larger the total number of interactions with its friends can encourage the potential scope and speed of worm propagation.

Moreover, we conduct some other experiments to observe the propagation dynamics process of SMS/MMS-based worm. Fig. 5 shows the continuous response on the total number of infected nodes with different infected rates. As time passes, the total number of infected nodes first increases gradually from $t = 0$ to $t = 20$, and at last reaches the maximum point about $t = 70$. We find that as the probability p_{SE} increases, the total number of infected nodes increases. That is to say, if a worm has the stronger infection, its propagation scope becomes larger.

Fig. 6 shows the transient response on the number of infected nodes $I(t)$. We find that as the probability p_{SE} increases, $I(t)$ increases quicker, and more susceptible nodes will be infected. It is seen that as the probability p_{SE} increases, the number of infected nodes increases, and the outbreak point is achieved ahead of time. $I(t)$ decreases gradually to zero as time passes.

The above simulation analysis shown in Figs. 5 and 6, which is based on one infection resource node. To analyze the impact of multiple infection resource nodes (IRN) on worm propagation, the following cases, such as $N_{IRN} = 3$, $N_{IRN} = 5$, $N_{IRN} = 7$, are provided in Fig. 7 and Fig. 8.

Fig. 7 shows the total number of infected nodes with different infection resource nodes. The total number of

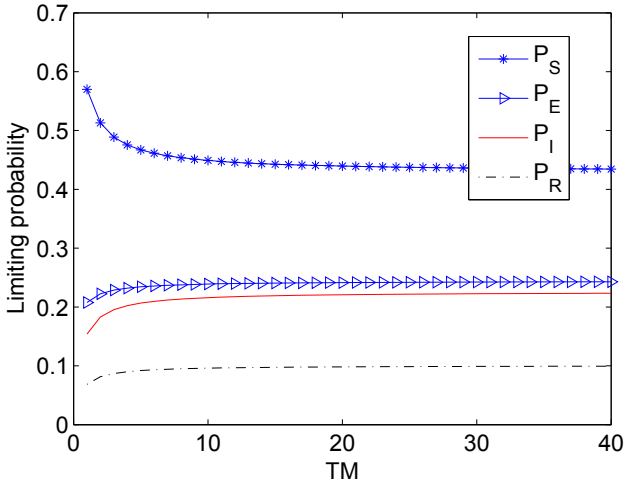

Fig. 3 – Limiting probabilities against the total number of messages records sent from its friend nodes to a node.

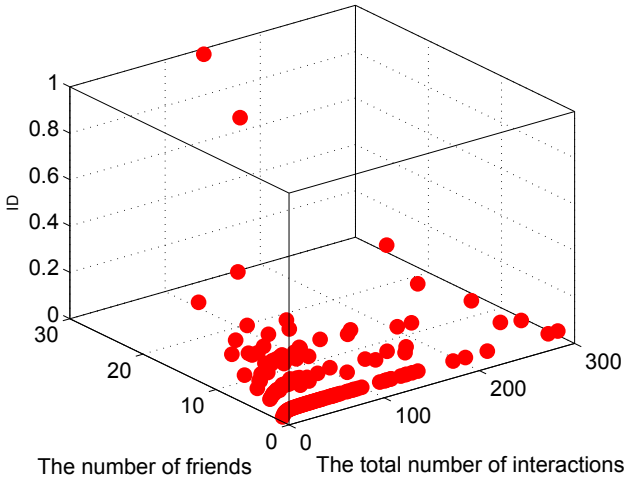

Fig. 4 – The relation between the number of friends, the total number of interactions and ID.

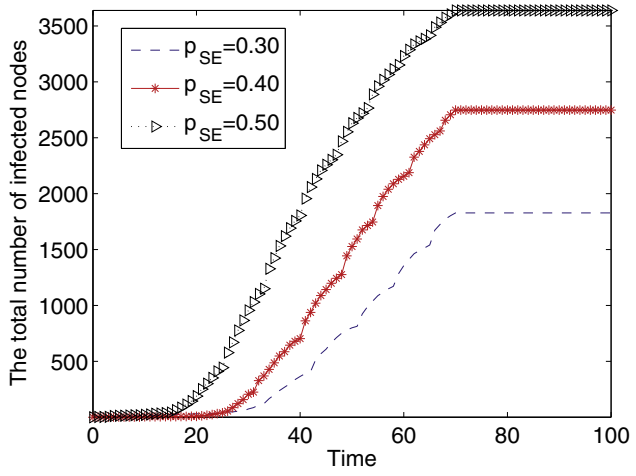

Fig. 5 – The total number of infected nodes against time under different infection rates.

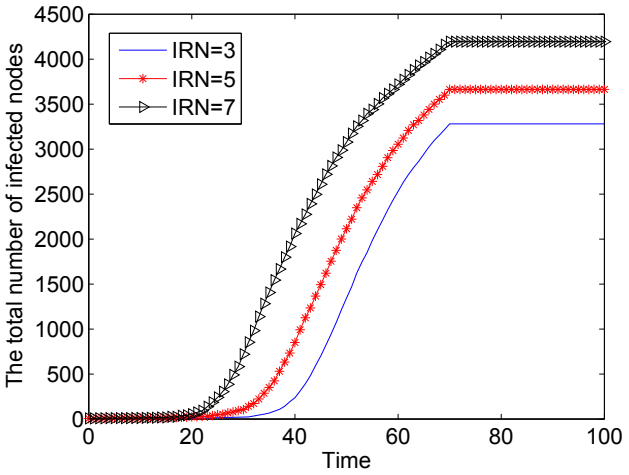

Fig. 7 – The total number of infected nodes with different infection resource nodes.

infected nodes increases slowly, as t changes from 1 to 20. Then as the value of t increases, the total number of infected nodes increases quickly. In addition, it is seen that the more infection resource nodes there are, the larger the total number of infected nodes is.

Fig. 8 shows the number of infected nodes with different infection resource nodes at time t. As can be seen from the results, as the value of N_{IRN} increases, the number of infected nodes increase. From these results, we find that it is important and valuable to contain worm spreading in smartphones through decreasing the number of N_{IRN}.

Comparing Fig. 5 with Fig. 7, we can see that the more N_{IRN} there are, the larger the total number of infected nodes is. Comparing Fig. 6 with Fig. 8, we can also see that the more N_{IRN} there are, the earlier the presence of the outbreak point will be.

Fig. 9 shows the transient response on the number of infected nodes $I(t)$ between SEIR model and our proposed model. Our model can characterize the dynamics of worm propagation more effectively than the general SEIR model. The difference of individual is introduced in our proposed model, which can distinguish the nodes that are prone to change their states from susceptible to exposed, and then can predict effectively the spreading principle of worms. Thus, the outbreak point of our proposed model comes much earlier than that of the SEIR model.

7. Summary and future work

In this paper, we have proposed a method to effectively characterize the propagation of SMS/MMS-based worms using the semi-Markov process and the social relationship graph. In our solution, the theoretical analysis for the limiting probability of state transition was provided. Moreover, the theoretical estimation and experimental analysis were also provided for the relation between smartphone users' social interactions and worm propagation. Its performance was evaluated using a customized program based on messaging records collected from real cellular networks. Through extensive evaluations, we demonstrate that our strategies can

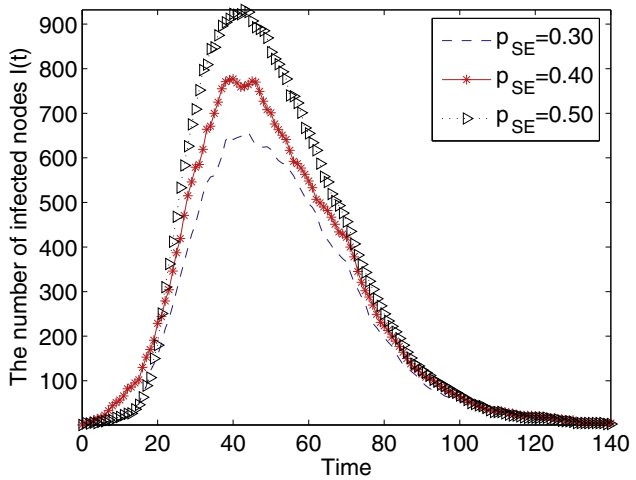

Fig. 6 – The number of infected nodes against time with different infection rates.

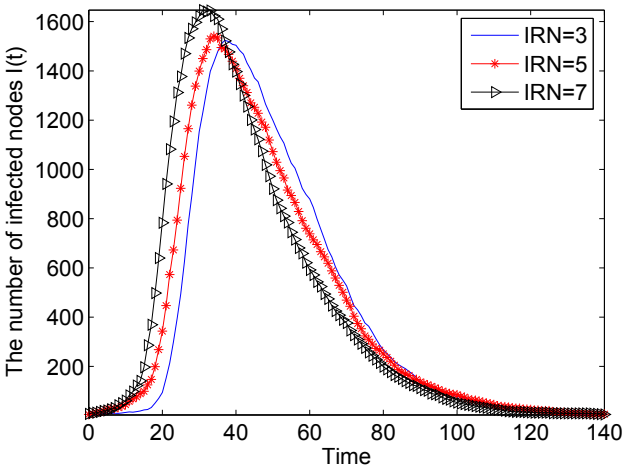

Fig. 8 – The number of infected nodes with different infection resource nodes.

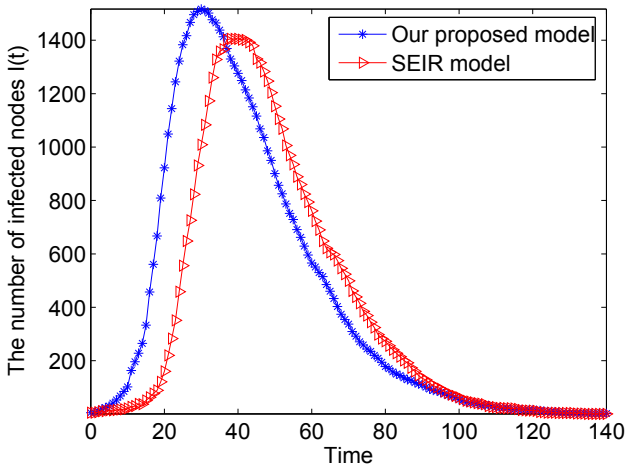

Fig. 9 – The number of infected nodes between SEIR model and our proposed model.

characterize the propagation of mobile worms more effectively than the traditional propagation models.

Recall that a semi-Markov node-state transition model has been proposed and Equation (11) has been provided as the solution for the limiting state probability P_j ($j \in W = \{S, E, I, R\}$) in Section 3. However, it is difficulty in determining the transition probabilities p_{ij} and expected sojourn times $M(j)$, even with this equation, we also have difficulty in calculating P_j. In addition, since these parameters are dependent on specific application scenario, and the estimation for computing of specific probabilities largely depends on the statistics of the data collected from real measurements. In this work, we have not provided details on that point supported by a proper example. In our future work, we will try to collect a large scale data set from practical measurements to provide an effective analysis solution of transition probabilities. Moreover, we will explore the propagation characteristics of hybrid worms and take the effect of vaccination process on the evolution of infected individuals into account.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant Nos. 61379041 and 61073037, the Postdoctoral Science Foundation of China under Grant No. 2012M511757, the Ministry of Education Fund for Doctoral Disciplines in Higher Education of China under Grant No. 20110162110043, the Natural Science Foundation of Guangdong Province of China under Grant No. S2011040002356, and the Postdoctoral Program of Central South University of China.

REFERENCES

Chen C, Pei X, Zhou H, Liu Y. A semi-Markov evaluation model for the survivability of real time database with intrusion tolerance. Chin J Comput 2011;34(10):1907–16 [in Chinese].

Corradi G, Janssen J, Manca R. Numerical treatment of homogeneous semi-Markov processes in transient case-a straightforward approach. Methodol Comput Appl Probab 2004;6:233–46.

Fan Y, Zheng K, Yang Y. Epidemic model of mobile phone virus for hybrid spread mode with preventive immunity and mutation. In: 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM 2010). Chengdu, China; September 2010. pp. 1–5.

Felt AP, Finifter M, Chin E, Hanna S, Wagner D. A survey of mobile malware in the wild. In: 1st ACM workshop on Security and privacy in smartphones and mobile devices (SPSM 2011). Chicago, Illinois, USA; 2011. pp. 3–14.

Fleizach C, Liljenstam M, Johansson P, Voelkery GM, Mhes A. Can you infect me now? Malware propagation in mobile phone networks. In: 4th ACM workshop on Recurring Malcode. Alexandria, VA, USA; 2007. pp. 61–8.

Gao C, Liu J. Modeling and predicting the dynamics of mobile virus spread affected by human behavior. In: 12th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2011). Lucca, Italy; 2011. pp. 1–9.

Hao J, Liu W, Dai Y. An anonymous node state transition model based on semi-Markov process. Acta Electron Sin 2011;39(5): 1082–6 [in Chinese].

Hou Z, Luo J, Shi P. Stochastic stability of linear systems with semi-Markovian jump parameters. ANZIAM J 2005;46:331–40.

Jamaluddin J, Zotou N, Coulton P. Mobile phone vulnerabilities: a new generation of malware. In: IEEE International Symposium on Consumer Electronics; 2004. pp. 199–202.

Li F, Yang Y, Wu J. CPMC: an efficient proximity malware coping scheme in smartphone-based mobile networks. In: 29th IEEE Conference on Computer Communications (INFOCOM 2010). San Diego, CA, USA; 2010. pp. 1–9.

Martin JC, Burge III LL, Gill JI, Washington AN, Alfred M. Modelling the spread of mobile malware. Int J Comput Aided Eng Technol 2010;2(1):3–14.

Peng S, Wang G. Worm propagation modeling using 2D cellular automata in bluetooth networks. In: 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2011). Changsha, China; 2011. pp. 282–7.

Peng S, Wang G, Hu Z, Chen J. Survivability modeling and analysis on 3D mobile ad-hoc networks. J Central South Univ Technol 2011;18(4):1144–52.

Peng S. A survey on malware containment models in smartphones. Appl Mech Mater 2013;263–266:3005–11.

Peng S, Wang G, Yu S. Modeling malware propagation in smartphone social networks. In: 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2013), Australia; 2013. pp. 196–201.

Peng S, Wang G, Yu S. Modeling the dynamics of worm propagation using two-dimensional cellular automata in smartphones. J Comput Syst Sci 2013b;79(5):586–95.

Peng S, Yu S, Yang A. Smartphone malware and its propagation modeling: a survey. IEEE Commun Surveys and Tut 2014;16(2):925–41.

Polla ML, Martinelli F, Sgandurra D. A survey on security for mobile devices. IEEE Commun Survey Tut 2013;15(1):446–71.

Rhodes CJ, Nekovee M. The opportunistic transmission of wireless worms between mobile devices. Phys Stat Mech Its Appl 2008;387(27):6837–44.

Shih D, Lin B, Chiang H, Shih M. Security aspects of mobile phone virus: a critical survey. Ind Manag Data Syst 2008;108(4):478–94.

Su J, Chan K, Miklas A, Po K, Akhavan A, Saroiu S, et al. A preliminary investigation of worm infections in a bluetooth environment. In: 4th ACM Workshop on Recurring Malcode (WORM 2006); 2006. pp. 9–16.

Van Ruitenbeek E, Courtney T, Sanders WH, Stevens F. Quantifying the effectiveness of mobile phone virus response mechanisms. In: 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2007). Edinburgh, UK; June 2007. pp. 791–800.

Xia W, Li Z, Chen Z, Yuan Z. Commwarrior worm propagation model for smart phone networks. J China Univ Posts Telecommun 2008;15(2):60–6.

Xing F, Wang W. On the survivability of wireless ad hoc networks with node misbehaviors and failures. IEEE Trans Dependable Secure Comput 2008;7(3):284–99.

Yan G, Eidenbenz S. Modeling propagation dynamics of bluetooth worms (extended version). IEEE Trans Mob Comput 2009;8(3):353–67.

Zheng H, Li D, Gao Z. An epidemic model of mobile phone virus. In: 1st IEEE International Symposium on Pervasive Computing and Applications (SPCA 2006). Urumqi, China; August 2006. pp. 1–5.

Zhu Z, Cao G, Zhu S, Ranjany S, Nucciy A. A social network based patching scheme for worm containment in cellular networks. In: 28th IEEE International Conference on Computer Communications (INFOCOM 2009). Rio de Janeiro, Brazil; 2009. pp. 1476–84.

Dr. Sancheng Peng received his Ph.D. degree in computer science from Central South University, Changsha, China, in 2010. Currently, he is a Professor with the School of Computer Science, Zhaoqing University, Zhaoqing, China. He was a Research Associate of City University of Hong Kong from 2008 to 2009. His research interests include network and information security, trusted computing, and mobile computing.

Dr. Min Wu received his B.S. and M.S. degrees in Engineering from Central South University, Changsha, China, in 1983 and 1986, respectively, and the Ph.D. degree in Engineering from Tokyo Institute of Technology, Tokyo, Japan, in 1999. He received the IFAC Control Engineering Practice Prize Paper Award in 1999 (together with M. Nakano and J. She). Currently, he is a Professor in Central South University and his research interests are process control, robust control, and intelligent system.

Dr. Guojun Wang received B.Sc. in Geophysics, M.Sc. in Computer Science, and Ph.D. in Computer Science, from Central South University, China. He is Head and Professor of Department of Computer Science at Central South University. He is also Director of Trusted Computing Institute at Central South University. He has been an Adjunct Professor at Temple University, USA; a Visiting Scholar at Florida Atlantic University, USA; a Visiting Researcher at the University of Aizu, Japan; and a Research Fellow at the Hong Kong Polytechnic University. His research interests include network and information security, Internet of Things, and cloud computing. He is a senior member of CCF, and a member of IEEE, ACM, and IEICE.

Dr. Shui Yu received his B.Eng. and M.Eng. degrees from University of Electronic Science and Technology of China, Chengdu, China, in 1993 and 1999, respectively. He received his Ph.D. degree from Deakin University, Victoria, Australia, in 2004. He is currently a Lecturer with the School of Information Technology, Deakin University, Victoria, Australia. His research interests include networking theory, network security, and mathematical modeling. He is a member of IEEE.