Computer Networks 74 (2014) 103–113

Contents lists available at ScienceDirect

Computer Networks

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

Containing smartphone worm propagation with an influence maximization algorithm

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 26 December 2013
Received in revised form 1 September 2014
Accepted 1 September 2014
Available online 22 September 2014

Keywords:
Smartphones
Worm containment
Influence maximization
Social relationship graph
Voting algorithm
Immunization

ABSTRACT

In recent years, wide attention has been drawn to the problem of containing worm propagation in smartphones. Unlike existing containment models for worm propagation, we study how to prevent worm propagation through the immunization of key nodes (e.g., the top k influential nodes). Thus, we propose a novel containment model based on an influence maximization algorithm. In this model, we introduce a social relation graph to evaluate the influence of nodes and an election mechanism to find the most influential nodes. Finally, this model provides a targeted immunization strategy to disable worm propagation by immunizing the top k influential nodes. The experimental results show that the model not only finds the most influential top k nodes quickly, but also effectively restrains and controls worm propagation.

1. Introduction

The number of smartphones in use is steadily increasing, as they are becoming more and more popular. For an increasing percentage of users, smartphones have become an integral part of their everyday lives. Moreover, all smartphones are now being equipped with advanced features, such as SMS (short messaging services), and MMS (multimedia messaging service) [1,2]. SMS and MMS attract more and more people due to their convenience, speed, and economic characteristics, making them a common means of communication in the daily lives of people. However, the significant development and pervasive use of smartphones also has attracted worm writers to pursue their malicious goals by exploiting smartphones' vulnerabilities [3].

According to recent security reports [4–8], the number of malicious exploits and executed attacks has recently surged. In 2010, more than 1 million cell phone users in China were infected by the 'Zombie' virus, which can automatically send text messages, and the attack resulted in a loss of $300,000 per day. The Juniper Networks Mobile Threat Center (MTC) released its *2011 Mobile Threats Report* in February 2012, which reported that mobile malware increased 155% across all platforms over the previous year, and provided evidence of a new level of maturity in the security threats aimed at mobile devices.

The *influence* of a node reflects its importance in a social network. The reason is that a node with a larger influence usually has more connections with others. Therefore, it is extremely important for us to design an effective and efficient mechanism to find the top k influential individuals, something which remains an important yet challenging problem [9]. Most of the existing models focus on greedy algorithms and mainly suffer

* Corresponding author.
E-mail addresses: min@csu.edu.cn (M. Wu), csgjwang@csu.edu.cn (G. Wang).

http://dx.doi.org/10.1016/j.comnet.2014.09.004
1389-1286/

from low computational efficiency, greatly hindering their application to real-world social networks. Although the existing models provide some valuable insights into the problem of influence maximization in social networks, strategies based on greedy algorithms fail to decrease the complexity.

Due to the scale-free characteristics of mobile social networks, a traditional immunization strategy [10], such as a random immunization strategy, a targeted immunization strategy, or an acquaintance immunization strategy, remains a huge challenge: a random immunization strategy needs to immunize 80 percent of the total number of nodes, a targeted immunization strategy requires knowing the global topology of the network, and an acquaintance immunization strategy needs to vaccinate the highly-connected important nodes.

How can we find the important nodes quickly and contain worm dissemination by immunizing these nodes? To answer this question, we need to evaluate which nodes are important without prior knowledge of the global network topology, and decrease the computational complexity for finding the important nodes in large-scale networks. In this paper, we present a novel approach that can significantly reduce the running time, can identify the most influential k nodes effectively and accurately without information about the global topology of the network, and effectively control worm propagation in a network. The contributions of our work are summarized as follows:

- We establish a social relationship graph based on the theory of complex networks. This graph is constructed using the actual SMS/MMS communication data from people's daily lives for social interactions. It is built to reveal the connections of social interaction and the spreading of SMS/MMS.
- We design a method of analysis of the behavior of a mobile social network based on the social relationship graph. The related factors of behavior analysis and their computing model are provided in this method, and they are used to count and analyze the characteristics of the mobile social network, such as in-degree, out-degree, the number of friends, activity degree, and intimacy degree.
- We design a mechanism to mine the top k influential nodes based on a voting algorithm. In this mechanism, each node for votes the most influential node among its set of friend nodes, according to the intimacy degree with these friends. Then, the heap sorting algorithm is employed to sort the results of the voting to discover the most influential former k nodes.
- We find that the immunization of the top k influential nodes can disable worm propagation more effectively than the traditional random immunization strategy or acquaintance immunization strategy, based on the worm propagation model presented in [1,2], and by improving the targeted immunization strategy.
- Extensive simulations using a real-world SMS/MMS-based communication data set demonstrate that the proposed algorithm is more effective and efficient than the existing models.

The remainder of this paper is structured as follows: In Section 2, we provide an overview of related work. We discuss the construction of the social relationship graph in Section 3. In Section 4, we perform an analysis seeking the factors of node influence and their computing models and present a mechanism to mine the influential top k nodes in Section 5. In Section 6, we design a containment scheme and provide the results of a model validation in Section 7. Finally, we conclude this paper in Section 8.

2. Related work

In this section, we review related work in terms of three dimensions. The first dimension is the worm propagation modeling; the second is related to the influence maximization algorithm; and the last one is related to worm containment models.

2.1. Worm propagation modeling

Zheng et al. [11] 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 poisoning. But the authors did not consider the impact of individual differences 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 [12] 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 some input parameters, such as the average node degree, average node meeting rate, and the link duration distribution. However, it is difficult to apply this model to analyze the propagation of SMS/MMS worms.

Peng and Wang [13] proposed a worm propagation modeling scheme (WPM) that used two-dimensional (2D) cellular automata to simulate the dynamics of the worm propagation process from a single node to the entire Bluetooth network. Although the WPM scheme integrates an infection factor and a resistance factor, it fails to provide specific expressions to compute these two factors.

Ruitenbeek et al. [14] proposed response mechanisms to analyze the effects of multimedia messaging system (MMS) viruses that spread by sending infected messages to other phones. Fleizach et al. [15] developed an event-based simulator to evaluate the effects of malware propagation using communication services like VOIP and MMS in mobile phone networks. However, they did not use real traffic data in their worm propagation model.

Peng et al. [16] proposed an approach to characterize the propagation dynamics of SMS/MMS-based worms. The authors introduced social network theory to characterize mobile worms that spread using MMS or SMS and typically exploit the social network of users to propagate from one mobile device to another. Moreover, the impact on the malware propagation of individual differences was also taken into account.

2.2. Influence maximization algorithm

Domingos and Riehardson [17] were the first to study the problem of influence maximization, and proposed a probabilistic model to solve this problem. In recent years, evaluating the influence of nodes in social networks and finding the top k influential nodes has drawn wide attention and has become quite an active research area.

Kempe et al. [18] proved that the problem of influence maximization is NP-hard, and then proposed a climbing-up greedy algorithm based on the independent cascading model and linear threshold model and provided $(1-1/e)$ approximation of the optimal solution.

Kimura et al. [19] examined the influence maximization problem (top k nodes problem) using SIR models (namely the IC and LT models) in a directed graph. They solved the problem by means of a greedy hill climbing algorithm on the basis of bond percolation and demonstrated a higher performance and a large reduction in computational cost compared with the conventional methods that simulated the random process many times.

Saito et al. [20] studied the influence maximization problem (top k nodes problem) using SIS models as final-time and integral-time maximization problem in a directed graph. They tried to solve the problems with a greedy algorithm on the basis of bond percolation, pruning, and burnout. They found that more influential users can be discovered than by approaches based on centrality measures and that the identified influential users differ remarkably depending on the chosen influence maximization problem.

Ma et al. [21] examined the influence maximization problem (top k nodes problem) using a heat diffusion process in a directed and an undirected graph. They solved the problem with a top k, k-step greedy, and enhanced k-step greedy algorithm.

Zhang et al. [22] examined the influence maximization problem (top k nodes problem) using an SIR model (namely IC). They adapted the IC model by weighting the edges that account for users' preferences for specific topics, and solved the problem with an optimized greedy algorithm including Monte Carlo simulation. Experimental results show that the approach significantly outperforms the traditional greedy algorithm in terms of information diffusion on specific topics.

Zhou et al. [23] proposed a two-stage algorithm, called Greedy Algorithm based on Users' Preferences (GAUP), to mine the top k influential nodes in social networks based on user preferences. The GAUP algorithm works in two stages: in the first stage, a vector space model is designed to compute user preferences; in the second stage, a greedy algorithm is adopted to find the most influential nodes.

Estevez et al. [24] proposed a set covering greedy algorithm to solve the problem of influence maximization. This algorithm repeatedly chooses the node with the highest "uncovered degrees". Once a node is chosen, all its neighbors and itself are labeled as "covered". This procedure is continued until k nodes are chosen.

Wang et al. [25] proposed a community-based greedy solution to the problem of influence maximization. In order to reduce the running time, they first detected communities based on the IC model and then mined the top k nodes across communities. They developed a cost function that optimized the community assignment in mobile networks.

Chen et al. [26] proposed a MixGreedy algorithm to reduce the computational complexity. This algorithm improved the performance by removing those unreachable nodes in the graph G to form a smaller new graph G^*.

Liu et al. [27] presented a framework, called IMGPU, to accelerate the influence maximization by leveraging the parallel processing capability of a graphics processing unit (GPU). It firstly converts the social graph into a Directed Acyclic Graph (DAG) to avoid redundant calculations. Then a Bottom-Up Traversal Algorithm (BUTA) was designed and mapped to the GPU using the CUDA programming model.

Narayanam and Narahari [28] examined the influence maximization problem (top k nodes problem) and the λ-coverage problem (finding a minimum set of influential nodes that influences a given percentage $\lambda$of nodes in the network) using an SIR model (namely LT) in a directed graph. They solved both problems by means of a Shapely-value based influential nodes (SPIN) algorithm based on a cooperative game, indicating that the SPIN algorithm is more powerful and computationally efficient than existing algorithms.

Leskovec et al. [29] proposed an algorithm called CELF (cost-effective lazy forward), which was reported to be 700 times faster than the algorithm proposed by Kempe et al. It was also based on the submodular property of the cascade influence function.

Wang and Feng [30] proposed a potential-based node selection algorithm to select some inactive nodes. This algorithm might not be optimal in the starting phase, yet can trigger more nodes in a later stage of diffusion.

2.3. Worm containment models

Bose and Shin [31] proposed a framework to contain malicious software spreading in messaging networks such as IM and SMS/MMS, called proactive group behavior containment (PGBC for short).

Xie et al. [32] proposed a systematic countermeasure against the propagation of cell-phone worms that exploit multimedia messaging service (MMS) for spreading. In this containment strategy, a graphic Turing test (GTT) and an identity-based signature were adopted to block unauthorized messages from leaving compromised phones; at the network level, a push-based automated patching scheme was presented for cleansing compromised phones. However, the defense scheme is more like intrusion prevention.

Xie et al. [33] proposed a two-level defense, which consists of an access-control based scheme and a GTT-based scheme to contain the propagation of MMS-based worms.

Zhu et al. [34] proposed a systematic approach to contain MMS worm propagation based on social networks. First, the authors built an undirected weighted graph G to represent the mobiles' social relationships in the cellular network from a real trace. Second, they divided the mobiles in cellular networks into multiple partitions applying either a balanced partitioning or a clustered partitioning algorithm based on the graph G. Third, a

minimum vertex separator from the set of cut edges was computed by designing a minimum vertex separator algorithm.

Gao and Liu [35] examined two strategies for restraining SMS-based virus propagation that are based on the methodology of autonomy-oriented computing (AOC). One of the two strategies is a pre-immunization strategy that can effectively reduce the number of infected users. The other is an adaptive dissemination strategy that adjusts the search behavior of the autonomous entities in the pre-immunization strategy, and can disseminate security notifications or patches to as many phones as possible in the mobile network, with a lower communication redundancy.

Zyba et al. [36] considered the dynamics of mobile phone malware that propagates by proximity contact, and evaluated potential defenses against it. The authors explored three strategies to defend mobile phones against proximity malware: local detection, proximity signature dissemination, and broadcast signature dissemination. However, its limitations are that signature flooding costs too much and that the local view of each node constrains the global optimal solution.

Li et al. [37] proposed an a community-based proximity malware coping scheme (CPMC for short). CPMC explores the social relationships and community structure of smartphone-based mobile networks. The scheme integrates short-term coping components, which deal with individual malware, and long-term evaluation components, which offer vulnerability evaluation towards individual nodes.

Tang et al. [38] researched how to block the propagation of Bluetooth-based malware through the immunization of key nodes. The authors introduced the notion of temporal closeness centrality to rank nodes by speed, and designed a time-aware containment strategy that spreads a patch message starting from nodes with high temporal closeness centrality. However, this strategy requires prior knowledge of future contacts, which is not available in practice.

In 2012 to overcome the drawbacks of [38], Tang et al. [39] proposed a predictive socio-temporal aware central node identification method that only requires the past history of device contacts. The authors presented socio-temporal opportunistic patching (STOP for short), a two-tier predictive mobile malware containment system. In STOP, a top k prediction model and a prediction function are designed to identify the top k ranking temporal centrality nodes based on past observations.

3. Construction of the social relationship graph

By analyzing the characteristics of the degree distribution of a mobile social network, we know that this kind of network possesses the related characteristics of a complex network. Thus, to characterize social interactions between smartphone users, we introduce a social relationship graph [16] in this paper. A social relationship graph is represented by a directed weighted graph, $G(V, E, W)$, where the set V of vertices corresponds to the smartphones in the cellular network, the set E of directed edges corresponds to the traffic flow between the smartphones, and the set W is the set of weights indicating the total number of SMS/MMS messages.

From a real-world data set which we 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 a smartphone social network, we take ten users from the data set and use them as an example. The data of this sample social network is listed in Table 1.

According to Table 1, we treat each smartphone as a vertex, so a weighted social relationship graph can be obtained: it is shown in Fig. 1.

In Fig. 1, we find that the social-interactions based SMS/MMS are in our everyday lives. If only the one-way behavior is presented, even if W is large, it is difficult to show that the relationship between these two users is very close. For example, if a smartphone user always sends advertising information to other people, it cannot be said that their relationships are close. Therefore, in this paper, we take the smaller of the weight values of the two directed edges between two nodes to measure the social relation between

Table 1
The number of interactions between any two cellular phone users in a week.

Between two smartphones	The number interactions
$A \to B$	2
$A \to C$	6
$A \to D$	4
$B \to A$	5
$B \to D$	15
$B \to E$	2
$B \to F$	7
$C \to A$	4
$C \to D$	8
$C \to G$	12
$C \to I$	3
$D \to A$	9
$D \to B$	6
$D \to C$	3
$D \to F$	4
$D \to G$	13
$E \to B$	3
$E \to F$	8
$E \to H$	8
$F \to B$	9
$F \to D$	8
$F \to E$	1
$F \to G$	14
$F \to H$	0
$F \to J$	5
$G \to C$	8
$G \to D$	15
$G \to F$	22
$G \to I$	6
$G \to J$	6
$H \to E$	6
$H \to F$	10
$H \to J$	12
$I \to C$	7
$I \to G$	0
$I \to J$	6
$J \to F$	9
$J \to G$	20
$J \to H$	13
$J \to I$	0

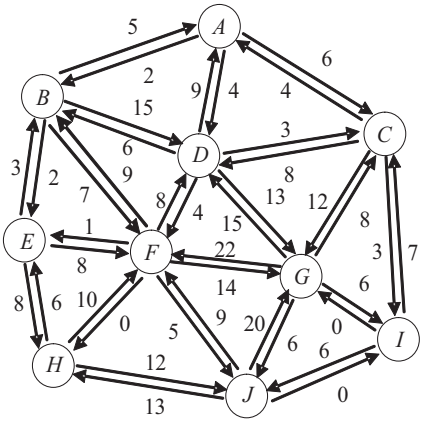


Fig. 1. A weighted social relationships graph (directed) based on the total number of SMS messages between any two nodes in a week.

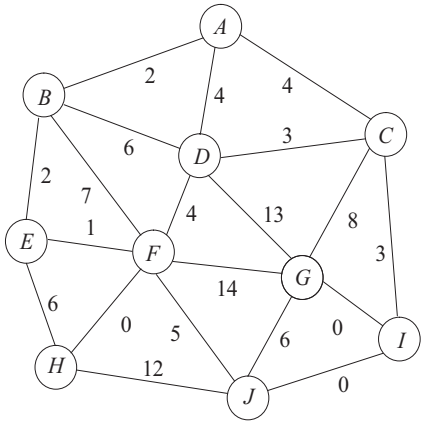


Fig. 2. A weighted social relationships graph (undirected) based on the total number of SMS messages between any two nodes in a week.

those nodes. That is, the social graph weight, denoted by W, is given by

$$W = \min\{C_{ij}, C_{ji}\} \tag{1}$$

where C_{ij} denotes the number of SMS/MMS messages sent from node i to j.

Thus, according to Eq. (1), we can obtain the effective interactions between any two nodes in a week, as shown in Table 2.

In addition, the transition result for Fig. 1 is shown in Fig. 2. The behaviors of SMS/MMS-based social interactions between two smartphone users are characterized accurately by changing the directed weighted graph $G(V, E, W)$ into the undirected weighted graph $G'(V, E, W)$.

4. Behavior analysis of mobile social networks

In this section, we firstly provide an analysis of the factors of node influence, then present the behavior analysis of the computing model.

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

Between two smartphones	The number of effective interactions
A, B	2
A, C	4
A, D	4
B, D	6
B, E	2
B, F	7
C, D	3
C, G	8
C, I	3
D, F	4
D, G	13
E, F	1
E, H	6
F, G	14
F, H	0
F, J	5
G, I	0
G, J	6
H, J	12
I, J	0

4.1. Relevant factors

Sending SMS and MMS to perform a social interaction reflects, to a certain extent, the real-life social relationships between people. It is also associated with the features of social networks, constituted by a particular social network, the mobile social network.

A mobile social network is a social network extracted from a data set of mobile communications, where each smartphone is represented by a node, and the relations between any two smartphones are represented by the set of edges. Based on complex network theory, we can scientifically and accurately identify and understand the social relations of mobile users, which will help us to contain worm propagation through the immunization of the influential users. In general, there are several factors to be considered, such as the number of friend nodes, the number of interactions between nodes, and node activity.

(1) The number of friend nodes.
If user A often sends an SMS/MMS to many users, this is a large contribution to the formation of mobile social networks. The larger the number of people receiving an SMS or MMS from A, the better social relations A has, and the greater the impact A on the network.
(2) The number of interactions between any two users.
If the number of interactions between user A and B is high, this means that the relationship between them is closer, and the influence of one user on another is greater. If there is a much closer relationship between a user with the other users, the influence of the user is much greater.
(3) Node activity.

In a unit period of time, if user A has many friends, and sends many SMS/MMS messages to them, this indicates that A is more active and has a much greater influence on the mobile social network in this period of time.

From the above analysis, it is seen that the number of friend nodes, the number of interactions between nodes, the node activity, and other factors are to be considered,

for the calculation of the influence of nodes in a mobile social network. Combining the above factors, we propose a computation model of node influence based on the social relationship graph.

4.2. Computing model

We wish to give a definition of the *activity degree* for a node in a mobile social network. It will be denoted by *AD*, and could be represented by measures such as the number of friend nodes and the number of SMS/MMS messages sent to other nodes.

First, we introduce a function $f_{ij}(t)$ to characterize the friendship at time t between node i and j.

$$f_{ij}(t) = \begin{cases} 1, & \min\{C_{ij}, C_{ji}\} > 0 \\ 0, & \min\{C_{ij}, C_{ji}\} = 0 \end{cases} \tag{2}$$

Now, let N be the total number of nodes in a mobile social network. Let N_i be the number of friend nodes of node i. Thus, N_i is described as follows.

$$N_i = \sum_{j=1}^{N} f_{ij}(t). \tag{3}$$

As a preliminary to the definition of *AD*, we must introduce the following concept.

Definition 1 (*Intimacy degree*). Social activities, such as making invitations, leaving words, sharing, are performed by sending SMS/MMS message between smartphone users. The closeness between users, their *intimacy degree*, can be reflected by the frequency of these activities, and is denoted by *ID*. More precisely, the *ID* between nodes i and j is denoted by ID_{ij} and is given by the following formula.

$$ID_{ij} = \frac{\min\{C_{ij}, C_{ji}\}}{\max\{\min\{C_{uv}, C_{vu}\}\}} \tag{4}$$

Let S_i denote the total number of effective interactions of node i with its friend nodes, i.e., S_i is defined as follows.

$$S_i = \sum_{k \in N_i} \min\{C_{ik}, C_{ki}\} \tag{5}$$

Definition 2 (*Activity degree*). If a node has many friend nodes, but the node fails to send SMS/MMS messages to its friend nodes, the influence of the node can be considered relatively small in a mobile social network. Thus, we introduce *AD* to measure which node is more active.

$$AD_i = \omega_1 \frac{S_i}{\max\{S_u\}} + \omega_2 \frac{N_i}{\max\{N_u\}} \tag{6}$$

where $i, j, u, v \in N$, $\max\{S_u\}$ denotes the maximum total number of sent SMS/MMS messages for all the nodes, $\max\{N_u\}$ denotes the maximum number of friend nodes for all the nodes, ω_1 and ω_2 denote the factors of weight, and $\omega_1 + \omega_2 = 1$.

5. Influence maximization model based on a voting algorithm

It is well-known that the influence maximization problem is NP-hard. Thus, it is challenging to mine the most influential top k nodes in a large scale mobile social network. Fortunately, during the course of an ordinary election process, such as the election of a mayor or governor, the voters tend to cast their ballots for the more trusted, the more reliable, or the more familiar candidates. In this paper, we exploit the voting algorithm of [40] to find the most influential top k nodes.

5.1. Voting algorithm

Since each node knows the related information (e.g., phone number) of its friend nodes in advance in the SMS/MMS-based communication network, we can assume, in the initialization process of a mobile social network, that each node also knows the ID number of its friends.

The voting rules are described as follows:

(1) Each node only votes for friend nodes with whom they communicate frequently (i.e., the highest intimacy degree) with SMS/MMS messages in the time unit.
(2) Each node only has one ballot to vote for a friend node each time.

After starting the voting process, each node performs its voting according to the voting rules. According to Eqs. (4) and (6) and the voting rules, we observe that the larger number of friend nodes and the greater amount of SMS/MMS messages are sent, the larger number of votes there are. To see the greatest influence in the selection of nodes, we need to consider not only the number of friends, but also consider the frequency of node activity. If a node has more friends, it is also possible to get more votes. On the other hand, if a node has few friends, even if its activity is frequent, it is impossible to get many votes. For this reason, we will employ *AD* to further refine the search for the most influential nodes.

5.2. Mining algorithm for top k influential nodes

During the course of voting, each node votes for another node only if that node is the most intimate one among its neighbor nodes, according to the size of *ID*. When the voting process is completed, it is possible that multiple nodes may share the same number of votes. It would be difficult to determine which nodes have a greater influence. In order to mine the influential nodes accurately and effectively, we exploit *AD* to determine which nodes have a greater influence.

Let Δt be the time unit. The five step algorithm is described as follows.

Step 1: Each node calculates the closeness degree between itself and its friend nodes according to Eq. (4);

Step 2: Each node votes for its friend nodes according to the voting algorithm;
Step 3: Each node calculates *AD* according to Eq. (6);
Step 4: When the voting procedure is completed, the number of votes is obtained for each node. The top k influential nodes can be mined by sorting the voting results with the heap sort algorithm. During the course of the sorting process, if the number of votes is the same for some nodes, choose the one with the largest *AD* as the most influential node.
Step 5: $t = t + \Delta t$.

The pseudocode of the mining algorithm for the top k influential nodes is shown in Algorithm 1.

Algorithm 1. Mining algorithm for top k influential nodes

Input: $G(V, E, W), k, S = \emptyset$;
Output: The set S for top k influential nodes;
1: **for** i = 1 to E **do**
2: Constructs friend relationship network;
3: **end for**
4: **for** i to N **do**
5: Computes *ID* according to Eq. (4);
6: Computes *AD* according to Eq. (6);
7: Node i casts its ballot;
8: **end for**
9: Computes the number of votes for each node;
10: Calls the function of heap sorting algorithm;
11: S = the sorting results of former k nodes,
12: return S.

5.3. An example for behavior analysis

According to Fig. 2, the number of friend nodes for each node is listed in Table 3.

In the light of Eq. (6), the result of *AD* for each node is presented in Table 4.

We set particular values for two parameters: $\omega_1 = 0.7$ and $\omega_2 = 0.3$. On the basis of the voting algorithm and computing model, the results of the voting for each node are listed in Table 5.

Combining the number of votes and the size of *AD*, the results of the sorting are listed in Table 6.

Table 3
The number of friend nodes for each node.

Terminal	The number of friend nodes
A	3
B	4
C	4
D	5
E	3
F	5
G	4
H	2
I	1
J	3

Table 4
The results of *AD* for each node.

Terminal	Activity degree
A	0.35
B	0.53
C	0.54
D	0.81
E	0.33
F	0.82
G	0.94
H	0.43
I	0.11
J	0.57

Table 5
The results of voting for each node.

Node	The number of votes
A	0
B	0
C	1
D	1
E	0
F	2
G	3
H	2
I	0
J	1

Table 6
The results of sorting.

Node	*No.*
G	1
F	2
H	3
C	4
D	5
J	6
A	7
B	8
E	9
I	10

5.4. Analysis of the algorithm's complexity

To compare the performance of our proposed algorithm with those of the existing algorithms, some notation is introduced as follows: N denotes the total number of nodes in the network; k denotes the number of the most influential nodes; c denotes the average number of friend nodes; and E denotes the total number of edges. The comparison results for the time complexity between the voting algorithm, the climbing-up greedy algorithm, and the set covering greedy algorithm, are shown in Table 7.

Table 7
Comparison of the complexity of each algorithm.

Algorithm	Complexity
Climbing-up greedy	$O(k \times N \times c^2 + E)$
Set covering greedy	$O(k \times N \times c + E)$
Voting	$O(N \times c + N \log(k) + E)$

It is shown in Table 7 that the complexity of the voting algorithm is smaller than that of either the climbing-up greedy algorithm or the set covering greedy algorithm.

6. Containment scheme for worm propagation

In order to prevent a worm from rapidly becoming an epidemic in a network, it is extremely important to design an effective and efficient mechanism to restrain the worm's propagation. At present, one of the most popular methods is network immunization. Immunization means that some nodes in the network are immunized and hence will not be infected by any worms. Considering the cost associated with immunization in large scale networks, the real point of immunization becomes how to immunize a small or the minimum number of important nodes.

In this paper, we improve the targeted immunization strategy based on the identification of the most influential nodes by using an influence maximization model. The pseudocode of the containment scheme for worm propagation is shown in Algorithm 2.

Algorithm 2. Containment strategy based on target immunization

Input:
The number of immunized nodes k;
The cycle time of immunization Δt;
The total time of worm propagation T;
Output:
k nodes been immunized;
1: **while** $t \leqslant T$ **do**
2: **if** $t \neq \Delta t$ **then**
3: Calls the mining algorithm for top k influential nodes;
4: Immunizes those k nodes;
5: **end if**
6: **end while**
7: return k.

7. Performance evaluation

In this section, we investigate how a social relationship graph is constructed by using the message records collected by one of the largest cellular networks in China. The data set of 0.4 million users in this network exchanged about 20 million SMS/MMS messages over a three-week period in October 2012. The content of the message records has been deleted, while the uniqueness of the identifiers of the phone numbers involved are replaced by pseudocodes.

In order to conduct the performance evaluation, we designed and developed a C++ simulator to implement our proposed mechanism, 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 took 5114 users for our experiments, rather than including all the users.

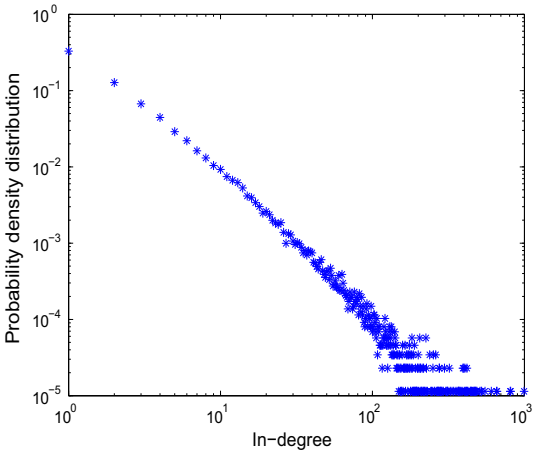


Fig. 3. The probability density distribution of in-degree.

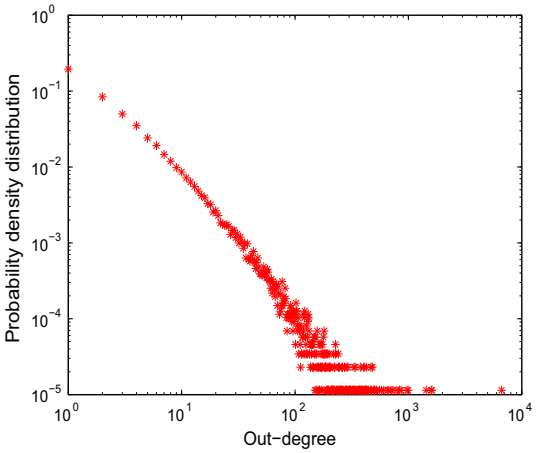


Fig. 4. The probability density distribution of out-degree.

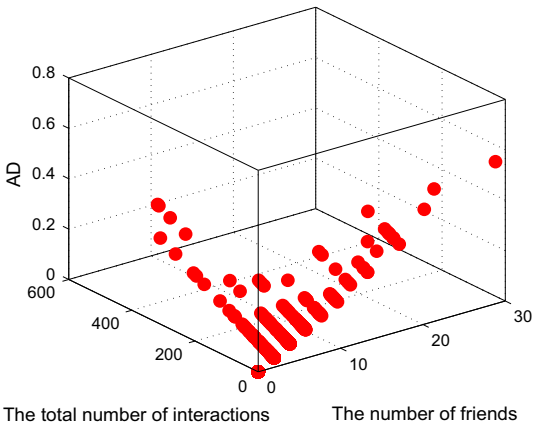


Fig. 5. The relation between the total number of interactions with its friends, the number of friends, and *AD*.

Firstly, we analyzed the characteristics of the complex network for the real-world data set. Figs. 3 and 4 show the degrees of the nodes for the probability density

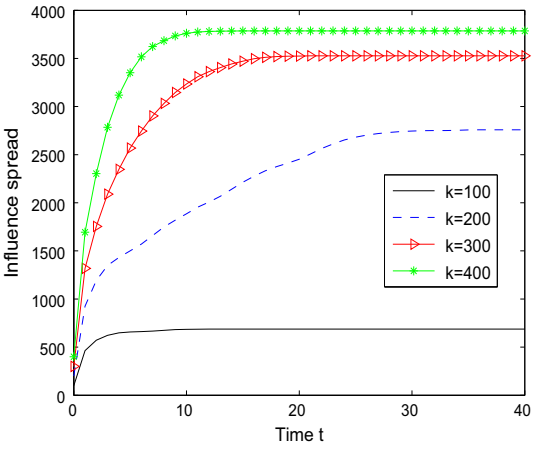


Fig. 6. A comparison of influence spread of voting algorithm with different k.

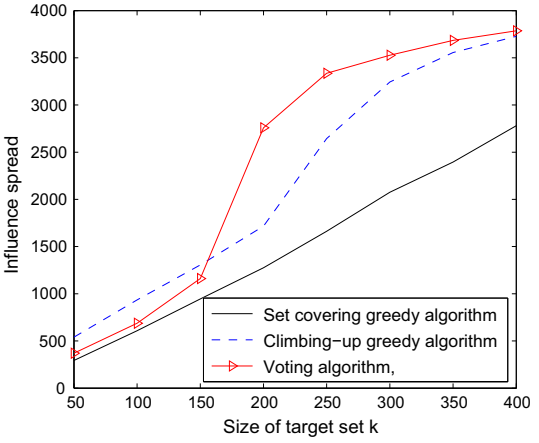


Fig. 7. A comparison of the influence spread of different algorithms with different k.

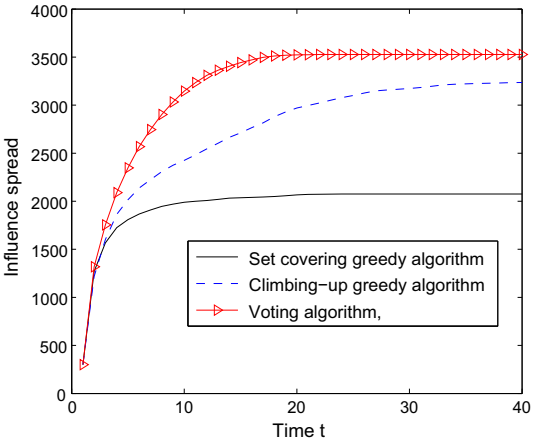


Fig. 8. A comparison of influence spread of different algorithms with different times ($k = 300$).

distribution of in-degree and out-degree, respectively. From Figs. 3 and 4, we observe that the degrees of the nodes obeys a power law distribution, which indicates that the mobile social network is a scale-free network.

In addition, we carried out an experiment to verify Eq. (6). Fig. 5 shows the relation between the total number of interactions with friends, the number of friends, and the *AD*. We know that for a node, the larger the number of its friends and the larger the total number of its interactions with its friends, the larger its *AD* will be, and so the greater will be its influence.

Fig. 6 shows the influence spread of the voting algorithm with different k at time t. As can be seen from the results, the number of influence spread for the voting algorithm increases as the value of k increases.

Fig. 7 shows the influence spreads of different algorithms with different k. From the results, we find that the influence spread of the voting algorithm is better than for the climbing-up greedy algorithm and the set covering greedy algorithm. The influence spread increases slowly, as k changes

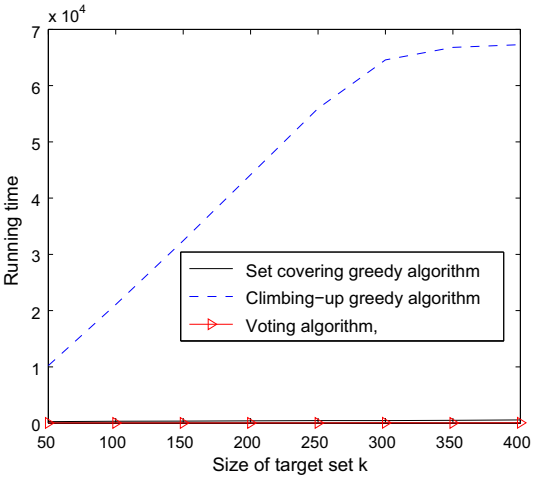


Fig. 9. A comparison of running times of different algorithms.

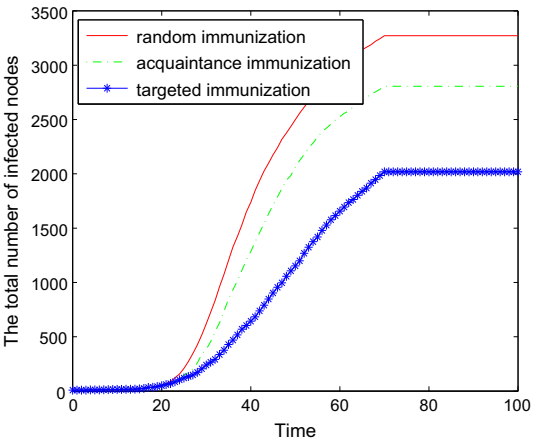


Fig. 10. A comparison of the total number of infected nodes between the random immunization strategy, the acquaintance immunization strategy, and the targeted immunization strategy.

from 1 to 150. Then as the value of k increases, the influence spread increases quickly. The reason is that only the top 1–150 nodes are influential nodes and the succeeding nodes do not contribute to increasing the influence spread.

Fig. 8 shows the influence spread of different algorithms with $k = 300$ at time t. We observe that the influence spread of the voting algorithm is larger than that for the climbing-up greedy algorithm or for the set covering greedy algorithm as time t increases.

Fig. 9 shows the running times of the different algorithms with different k. We see that the running time of the climbing-up greedy algorithm is proportional to the value of k. However, the running time of the voting algorithm and set covering greedy algorithm is very low, almost negligible. This is because the voting algorithm and the set covering greedy algorithm only have to spend a little time for adjustment as k increases.

In order to verify the efficiency of the proposed containment scheme, we compare our strategy with the random immunization strategy and the acquaintance immunization strategy using the model of worm propagation presented in [16]. In Fig. 10, we find that the improved targeted immunization strategy is more effective for worm containment than the immunization strategy or the acquaintance immunization strategy.

8. Conclusion and future research

In this paper, we have presented a model for worm containment based on the idea of an influence maximization algorithm. At first, in order to investigate the efficiency of the mining model for the top k nodes based on a voting algorithm, we compared this mining model with the climbing-up greedy algorithm and the set covering greedy algorithm using a real-world data set. The experimental results show that the proposed algorithm is superior to the climbing-up greedy algorithm and the set covering greedy algorithm in time complexity.

In addition, we have also verified our worm containment scheme based on an influence maximization model by a comparison with some existing strategies. The experimental results show that the presented worm containment scheme is effective for large scale mobile social networks.

Due to the complexity of the social relations between people, there is a diversity, randomness, and uncertainty reflected in the SMS/MMS-based interactions between mobile phone users; to a certain extent, this partly reflects the social interactions between mobile phone users and their relationships, but does not reflect or summarize the full social relations between them. Therefore, how to characterize the real and stable social relations, such as work relationships and family relationships, on the basis of the social behavior between mobile phone users, is worthy of further study.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61272151, 61472451 and 61379041, the Postdoctoral Science Foundation of China under Grant No. 2012M511757, the Ministry of Education Fund for Doctoral Disciplines in Higher Education under Grant No. 20110162110043, the Natural Science Foundation of Guangdong Province under Grant No. S2011040002356, and the Postdoctoral Program of Central South University.

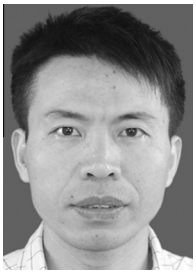
Appendix A. Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.comnet.2014.09.004.

References

[1] S. Peng, G. Wang, S. Yu, Modeling the dynamics of worm propagation using two-dimensional cellular automata in smartphones, J. Comput. Syst. Sci. 79 (5) (2013) 586–595.
[2] S. Peng, M. Wu, G. Wang, S. Yu, Propagation model of smartphone worms based on semi-Markov process and social relationship graph, Comput. Secur. 44 (2014) 92–103.
[3] J. Jamaluddin, N. Zotou, P. Coulton, Mobile phone vulnerabilities: a new generation of malware, IEEE Int. Symp. Consum. Electron. (2004) 199–202.
[4] C. Gao, J. Liu, Modeling and predicting the dynamics of mobile virus spread affected by human behavior, in: Proceedings of the 12th IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM 2011), Lucca, Italy, 2011, pp. 1–9.
[5] S. Peng, S. Yu, A. Yang, Smartphone malware and its propagation modeling: a survey, IEEE Commun. Surv. Tutorials 16 (2) (2014) 925–941.
[6] A.P. Felt, M. Finifter, E. Chin, S. Hanna, D. Wagner, A survey of mobile malware in the wild, in: Proceedings of the 1st ACM workshop on Security and Privacy in Smartphones and Mobile Devices (SPSM 2011), Chicago, Illinois, USA, 2011, pp. 3–14.
[7] D. Shih, B. Lin, H. Chiang, M. Shih, Security aspects of mobile phone virus: a critical survey, Ind. Manage. Data Syst. 108 (4) (2008) 478–494.
[8] S. Peng, A survey on malware containment models in smartphones, Appl. Mech. Mater. 263–266 (2013) 3005–3011.
[9] F. Probst, D.L. Grosswiele, D.R. Pfleger, Who will lead and who will follow: identifying influential users in online social networks, Bus. Inf. Syst. Eng. 5 (3) (2013) 179–193.
[10] C. Gao, J. Liu, N. Zhong, Network immunization with distributed autonomy-oriented entities, IEEE Trans. Parallel Distrib. Syst. 22 (7) (2011) 1222–1229.
[11] H. Zheng, D. Li, Z. Gao, An epidemic model of mobile phone virus, in: Proceedings of the 1st IEEE International Symposium on Pervasive Computing and Applications (SPCA 2006), Urumqi, China, 2006, pp. 1–5.
[12] G. Yan, S. Eidenbenz, Modeling propagation dynamics of bluetooth worms (extended version), IEEE Trans. Mob. Comput. 8 (3) (2009) 353–367.
[13] S. Peng, G. Wang, Worm propagation modeling using 2D cellular automata in bluetooth networks, in: Proceedings of the 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2011), Changsha, China, 2011, pp. 282–287.
[14] E.V. Ruitenbeek, T. Courtney, W.H. Sanders, F. Stevens, Quantifying the effectiveness of mobile phone virus response mechanisms, in: Proceedings of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2007), Edinburgh, UK, 2007, pp. 791–800.
[15] C. Fleizach, M. Liljenstam, P. Johansson, G.M. Voelkery, A. Mhes, Can you infect me now? malware propagation in mobile phone networks, in: Proceedings of the 5th ACM Workshop on Recurring Malcode (WORM 2007), Alexandria, VA, USA, 2007, pp. 61–68.
[16] S. Peng, G. Wang, S. Yu, Modeling malware propagation in smartphone social networks, in: Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2013), Australia, 2013, pp. 196–201.
[17] P. Domingos, M. Riehardson, Mining the network value of customers, in: Proceedings of the 7th ACM Conference on Knowledge Discovery and Data Mining, New York, USA, 2001, pp. 57–66.

[18] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence in a social network, in: Proceedings of the 9th ACM Conference on Knowledge Discovery and Data Mining, Washington, USA, 2003, pp. 137–146.
[19] M. Kimura, K. Saito, R. Nakano, Extracting influential nodes for information diffusion on a social network, in: Proceedings of the 22nd National Conference on Artificial Intelligence, Vancouver, 2007, pp. 1371–1376.
[20] K. Saito, M. Kimura, K. Ohara, H. Motoda, Efficient discovery of influential nodes for SIS models in social networks, Knowl. Inf. Syst. 30 (3) (2012) 613–635.
[21] H. Ma, H. Yang, M. Lyu, I. King, Mining social networks using heat diffusion processes for marketing candidates selection, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, 2008, pp. 233–242.
[22] Y. Zhang, J. Zhou, J. Cheng, Preference-based top-k influential nodes mining in social networks, in: Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications, Changsha, 2011, pp. 1512–1518.
[23] J. Zhou, Y. Zhang, J. Cheng, Preference-based mining of top-influential nodes in social networks, Future Gener. Comput. Syst. 31 (2014) 40–47.
[24] P.A. Estevez, P. Vera, K. Saito, Selecting the most influential nodes in social networks, in: Proceedings of the International Joint Conference on Neural Networks, Orlando, Florida, USA, 2007, pp. 2397–2402.
[25] Y. Wang, G. Cong, G. Song, K. Xie, Community-based greedy algorithm for mining top-k influential nodes in mobile social networks, in: Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 2010, pp. 1039–1048.
[26] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social network, in: Proceedings of the 15th ACM Conference on Knowledge Discovery and Data Mining, Paris, France, 2009, pp. 199–208.
[27] X. Liu, M. Li, S. Li, S. Peng, X. Liao, X. Lu, IMGPU: GPU accelerated influence maximization in large-scale social networks, IEEE Trans. Parallel Distrib. Syst. 25 (1) (2014) 136–145.
[28] R. Narayanam, Y. Narahari, A Shapley value-based approach to discover influential nodes in social networks, IEEE Trans. Autom. Sci. Eng. 8 (1) (2011) 130–147.
[29] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J.M. VanBriesen, N.S. Glance, Cost-effective outbreak detection in networks, in: Proceedings of the 13th ACM Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 2007, pp. 420–429.
[30] Y. Wang, X. Feng, A potential-based node selection strategy for influence maximization in a social network, in: Proceedings of the 5th International Conference on Advanced Data Mining and Applications, 2009, pp. 350–361.
[31] A. Bose, K.G. Shin, Proactive Security For Mobile Messaging Networks, in: Proceedings of the 5th ACM Workshop on Wireless Security (WiSe 2006), Los Angeles, California, USA, 2006, pp. 95–104.
[32] L. Xie, H. Song, T. Jaeger, S. Zhu, Towards a systematic approach for cell-phone worm containment, in: Proceeding of the 17th International Conference on World Wide Web (WWW 2008), Beijing, China, 2008, pp. 1083–1084.
[33] L. Xie, X. Zhang, A. Chaugule, T. Jaeger, S. Zhu, Designing System-level Defenses against Cellphone Malware, in: Proceeding of the 28th IEEE International Symposium on Reliable Distributed Systems (SRDS 2009), 2009, pp. 83–90.
[34] Z. Zhu, G. Cao, S. Zhu, S. Ranjany, A. Nucciy, A social network based patching scheme for worm containment in cellular networks, in: Proceedings of the 28th IEEE Conference on Computer Communications (INFOCOM 2009), Rio de Janeiro, Brazil, 2009, pp. 1476–1484.
[35] C. Gao, J. Liu, Modeling and restraining mobile virus propagation, IEEE Trans. Mob. Comput. 12 (3) (2013) 529–541.
[36] G. Zyba, G.M. Voelker, M. Liljenstam, A. Mehes,, P. Johansson, Defending mobile phones from proximity malware, in: Proceedings of the 28th IEEE Conference on Computer Communications (INFOCOM 2009), Rio de Janeiro, Brazil, 2009, pp. 1503–1511.
[37] F. Li, Y. Yang, J. Wu, CPMC: an efficient proximity malware coping scheme in smartphone-based mobile networks, in: Proceedings of the 29th IEEE Conference on Computer Communications (INFOCOM 2010), San Diego, CA, USA, 2010, pp. 1–9.
[38] J. Tang, C. Mascolo, M. Musolesi, V. Latora, Exploiting temporal complex network metrics in mobile malware containment, in: Proceedings of the IEEE 12th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM2011), Lucca, Italy, 2011.
[39] J. Tang, H. Kim, C. Mascolo, M. Musolesi, STOP: Socio-Temporal Opportunistic Patching of Short Range Mobile Malware, in: Proceedings of the 13th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2012), San Francisco, USA, 2012.
[40] S. Peng, W. Jia, G. Wang, Voting-Based clustering algorithm with subjective trust and stability in mobile ad-hoc networks, in: Proceedings of the 2008 IEEE/IFIP International Conference On Embedded and Ubiquitous Computing (EUC 2008), Shanghai, China, 2008, pp. 3–9.



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

Min Wu received a B.S. in 1983 and an M.S. in 1986 in Engineering from Central South University, Changsha, China, and a Ph.D. 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 at Central South University and his research interests are process control, robust control, and intelligent system.

Guojun Wang received a B.Sc. in Geophysics, and an M.Sc. and Ph.D. in Computer Science from Central South University, China. He is Head and Professor of the Department of Computer Science and Technology in Central South University. He is also Director of the 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 Hong Kong Polytechnic University. His research interests include network and information security, the Internet of Things, and cloud computing. He is a distinguished member of the CCF, and a member of the IEEE, ACM, and IEICE.

Shui Yu received the B.Eng. in 1993 and the M.Eng. in 1999 from the University of Electronic Science and Technology of China, Chengdu, China. He received the 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 the IEEE.