The Temporal Characteristic of Human Mobility: Modeling and Analysis of Social Worm Propagation

Tianbo Wang, Chunhe Xia, and Qiong Jia

***Abstract*—A common view of modeling social worm propagation is topology construction for social networking layer and human behavior modeling for message checking activity. However, due to the existence of human mobility behavior, traditional analytical models have not been suitable for the propagation dynamics of social worms nowadays, which results in underestimating the scale of infected network and lacking a comprehensive and accurate abstraction for human behaviors. This letter proposes a novel model to understand the propagation dynamics of social worms in hierarchical networks based on the temporal characteristic of human mobility. We characterize the infection scale and the spread speed for social worms adopting discrete model and difference-equation method. Our simulations and analysis show our model to approximate the complicated propagation behaviors of social worms more accurately, and we give two quantitative conclusions on the range of final infection scale and the fastest spread speed. In addition, the results indicate the rationality and validity of our conclusions.**

***Index Terms*—Network security, social worm, human mobility, propagation.**

I. Introduction

SOCIAL worms constitute one of the major network security problems. According to Symantec Corporations report on official Internet security threats [1], the frequency and virulence of their propagation outbreaks have increased dramatically in the last few years. Firstly, they only rely on knowledge of network topology, which do not require any vulnerable information in a computer system or software. They have less network traffic than scanning worms and better **imperceptibility**. Secondly, by exploiting trust between friends, many users fail to recognize malware or malicious codes that are sent by their friends and subsequently users become infected. This makes worm propagation more **effective**. Finally, the carrier of social worms is human mobility, which is the connection of a hierarchical network and the addition of social engineering techniques. The propagation of social worms has the characteristics of **interdisciplinarity** and **multidimension**.

In order to understand how social worms propagate in Internet, worm propagation dynamics is discussed in [2]–[5]. These simulation models can describe worm propagation behavior in social networks very well, and we have a qualitative understanding of how social worms spread. However, they cannot provide analytical study on the nature of the propagation. There are many improved works on the analytic model of propagation dynamics. [6] proposes that the Markov model can incorporate both detailed topology information and simple spatial dependence into achieving a greater accuracy than previous models. However, there are two problems: temporal dynamics and spatial dependence. In order to solve them, the SII model is presented in [7], and the results show that the SII model is more suitable for modeling the propagation of social worms. Meanwhile, The paper of [8] presents a modern email malware model that accounts for two new features of reinfection and self-start. They all have a "*social layer topology*" assumption which does not accord with the topology struction in the real world.

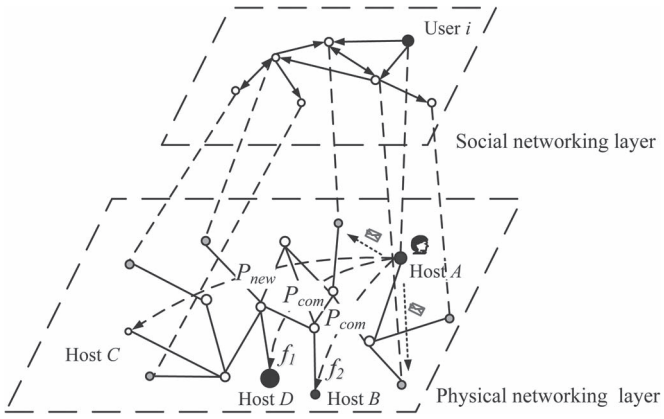

Fig. 1. The propagation scenario of social worms in hierarchical networks. At time t user i in the social networking layer operates host A in location a of the physical networking layer. At the next time $t + \Delta t$ (i.e., T^i_{rest} denotes the time user i spent at one location.), the user i can either (i) Preferential return, $P_{com} = 1 - P_{new}$, the user i visited previously locations with frequency f_i that is proportional to the size of circles drawn at each location, or (ii) Exploration, $P_{new} = \rho S^{-\gamma}$, user i operates host C in location c, where S is the total number of visited locations [9].

However, the above works all neglect the influence of human mobility and hierarchical networks on the spread of social worms. The **hierarchical network** is a network that describes the interdependency between human behaviors and network devices from the perspective of both social networking layer and physical networking layer, as shown in Fig. 1. In the social networking layer, nodes and edges denote users and friend relationships between users respectively. In the physical networking layer, leaf nodes and non-leaf nodes denote hosts and interconnect devices respectively, and edges denote physical connections. A reason for underestimation of infection scale in previous works should be rooted in the failure of considering the dynamic spreading procedure in the physical networking layer. Because different social roles lead to human mobility among different locations, they use different computers to address related work. This results in an one-to-many relationship between a user and hosts. At the same time, they do not pay

Manuscript received January 4, 2015; accepted April 30, 2015. Date of publication May 7, 2015; date of current version July 8, 2015. This work was supported by the National Natural Science Foundation of China under Grant No. 61170295. The associate editor coordinating the review of this paper and approving it for publication was P. D. Yoo.

The authors are with the Beijing Key Laboratory of Network Technology, Beihang University, Beijing 100191, China, and also with the Research Institute in Shenzhen, Beihang University, Beijing 100191, China (e-mail: wangtb@buaa.edu.cn; xch@buaa.edu.cn; jiaqiong1219@buaa.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2015.2430972

attention to analyze the influence of the temporal characteristic of human mobility, which impacts on the infection scale and the spread speed.

This letter proposes a novel analytical model to theoretically analyze the upper and lower bounds of the scale of infected nodes and the spread speed of social worms. Validation against conducted simulation experiments reveals that our analysis helps understand the realistic propagation dynamics of social worms.

II. Propagation Model of Social Worms

To quantitatively analyze spread ability of social worms, we model social worm propagation based on human mobility. Firstly, let random variable $X_{i,n}^t$ denotes the state of a host n used by user i at time t, which is in either susceptible (Sus.) or infected (Inf.). If host n is infected, $X_{i,n}^t = Inf.$, otherwise $X_{i,n}^t = Sus.$. Secondly, we introduce an M by M square matrix to describe the social networking layer, as in

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ \vdots & \vdots & p_{ij} & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{pmatrix} p_{ij} \in [0,1] \tag{1}$$

where p_{ij} represents the propagation probability from user i to user j. $p_{ij} = 0$ means user j is not in the contact list of user i, and M is the size of the social networking layer. We propose an M by N_{max} matrix to denote the number of locations visited by users at time t, as follows:

$$\begin{pmatrix} N_{11}^t & \cdots & N_{1N_{max}}^t \\ \vdots & N_{ij}^t & \vdots \\ N_{M1}^t & \cdots & N_{MN_{max}}^t \end{pmatrix} \tag{2}$$

where N_{ij}^t represents the number of user i visiting location j by time t, and the time user i spent at location j is chosen from the $P(T_{rest}^i)$ distribution. N_{max} denotes the maximum number of visited locations by every user, we can obtain its solution according to values of ρ and γ in [9]

$$F(S \le x) = \lim_{x \to N_{max}} \int_1^x \rho S^{-\gamma} dS = 1 \tag{3}$$

Thirdly, we use $n(t)$ to represent the number of infected nodes in the hierarchical network at time t, as in

$$\begin{aligned} n(t) &= E\left(\sum_{i=1}^{M}\sum_{n=1}^{N_{max}} X_{i,n}^t = Inf.\right) \\ &= \sum_{i=1}^{M}\sum_{n=1}^{N_{max}} E\left(X_{i,n}^t = Inf.\right) \\ &= \sum_{i=1}^{M}\sum_{n=1}^{N_{max}} P\left(X_{i,n}^t = 1\right) \\ &= \sum_{i=1}^{M}\sum_{n=1}^{N_{max}} P\left(X_{i,n}^t = Inf.\right). \end{aligned} \tag{4}$$

We derive the computation of $P\left(X_{i,n}^t = Inf.\right)$ by different equations as follows:

$$\begin{aligned} P\left(X_{i,1}^t = Inf.\right) &= v'(i,1,t) \cdot P\left(X_{i,1}^{t-1} = Sus.\right) \\ &\quad + P\left(X_{i,1}^{t-1} = Inf.\right) \end{aligned} \tag{5}$$

$$\begin{aligned} P\left(X_{i,n}^t = Inf.\right) &= v(i,n,t) \cdot P\left(X_{i,n}^{t-1} = Sus.\right) \\ &\quad + P\left(X_{i,n}^{t-1} = Inf.\right), n > 1 \end{aligned} \tag{6}$$

Then, $v'(i,1,t)$ and $v(i,n,t)$ represent the infected probability that user i operates the same or different host in different locations at time t, respectively. There are three preconditions: 1) the user is checking the social account for new messages; 2) the susceptible user reads those malicious messages, which are represented by $s(i,n,t)$; 3) the user is in location n.

$$\begin{aligned} v'(i,1,t) &= s(i,1,t) \cdot P\left(open_i(t) = 1\right) \\ &\quad \cdot \left[P_{i\Rightarrow 1}(t) + \left(1 - \prod_{n=2}^{N_{max}} P_{dev}^n(t)\right)\right] \end{aligned} \tag{7}$$

$$v(i,n,t) = s(i,n,t) \cdot P\left(open_i(t) = 1\right) \cdot P_{i\Rightarrow n}(t) \cdot P_{dev}^n(t), n > 1 \tag{8}$$

$$P_{i\Rightarrow n}(t) = \frac{N_{in}^t}{\sum_{j=1}^{N_{max}} N_{ij}^t} \tag{9}$$

where $P_{i\Rightarrow n}(t)$ denotes the probability of user i staying in the location n at time t. $P_{dev}^n(t)$ denotes the probability that a user operates a new host in location n at time t, which is proportional to the visiting probability $P_{i\Rightarrow n}(t)$ and the resting time proportion $T_i(n)$ in location n.

$$P_{dev}^n(t) = \alpha \cdot P_{i\Rightarrow n}(t) \cdot T_i(n) \tag{10}$$

where $T_i(n) = T_{rest}^i(n) / \sum_{n=1}^{N_{max}} T_{rest}^i(n)$ and $\alpha = 1 / \sum_{n=1}^{N_{max}} [P_{i\Rightarrow n}(t) \cdot T_i(n)]$. T_{check}^i denotes message checking time of user i. $open_i(t) = 1$ denotes that user i is checking new messages at time t, which means $t \bmod T_{check}^i = 0$, otherwise $open_i(t) = 0$.

Finally, we can compute $s(i,n,t)$, as in

$$\begin{aligned} s(i,n,t) &= 1 - \left(1 - s(i,n,t-1)\left[1 - P\left(open_i(t-1) = 1\right)\right]\right) \\ &\quad \cdot \prod_{j \in N_i} \prod_{m=1}^{N_{max}} \left[1 - p_{ji} \cdot P\left(X_{j,m}^{t-1} = Inf. | X_{i,n}^{t-1} = Sus.\right)\right] \end{aligned} \tag{11}$$

where N_i denotes the set of neighboring users of user i.

III. Theoretical Analysis of the Spread Ability of Social Worms

In this section, we characterize the spread ability of social worms (i.e., the infection scale and the spread speed).

A. Threshold Analysis of the Scale of Infected Nodes

Recall that social worms has two characteristics: the hierarchical topology and the social engineering caused message checking and human mobility.

Conclusion 1: Suppose that $n_s(t)$ is the number of infected nodes at time t in the social networking layer. The range of infection scale of social worms in hierarchical networks is that $[n_s(t), M \cdot N_{max}]$.

Upper-Bound of the Scale of Infected Nodes: In the worst case, user i operating different hosts in different locations can receive and check malicious messages from neighbors. Thus, this means all hosts in different locations should be infected.

Lower-Bound of the Scale of Infected Nodes: No matter where user i operates the corresponding host, the infection state of one host at least should be consistent with the state of nodes in the social networking layer. Thus, the peak value of infected nodes in the social networking layer is lower-bound. For the social networking layer, let random variable X_i^t denotes the state

of node i at time t, $v(i,t)$ represents the infection probability of node i at time t, and $s(i,t)$ represents the probability of user i reading malicious messages from neighboring nodes at time t. Therefore, we revise our presented model in Section II, and obtain the propagation model in the social networking layer, as follow:

$$n_s = n(t) = \sum_{i=1}^{M} P\left(X_i^t = 1\right) = \sum_{i=1}^{M} P\left(X_i^t = Inf.\right) \tag{12}$$

$$P\left(X_i^t = Inf.\right) = v(i,t) \cdot P\left(X_i^{t-1} = Sus.\right) + P\left(X_i^{t-1} = Inf.\right) \tag{13}$$

$$v(i,t) = s(i,t) \cdot P\left(open_i(t) = 1\right) \tag{14}$$

$$s(i,t) = 1 - \left(1 - s(i,t-1)\left[1 - P\left(open_i(t-1) = 1\right)\right]\right) \cdot \prod_{j \in N_i} \left[1 - p_{ji} \cdot P\left(X_j^{t-1} = Inf. | X_i^{t-1} = Sus.\right)\right] \tag{15}$$

On the basis of above analysis, upper-bound and lower-bound the scale of infected nodes in hierarchical networks are given by

$$n_s(t) \le n(t) \le M \cdot N_{max}. \tag{16}$$

B. Quantitative Analysis of the Spread Speed of Social Worms

Message checking and the temporal characteristic of human mobility are two core elements of both information dissemination and hierarchical network formation.

Theorem 1: Suppose that $V(t)$ denotes the spread speed of social worms. The condition that $V(t)$ reaches the maximum is that $T_{check}^i = T_{rest}^i, i = 1, 2, \cdots, M$.

$$\max V(t) = \sum_{i=1}^{M} \max\left[v_i(t)\right] \tag{17}$$

Proof: Suppose that $v_i(t)$ is the infection speed of user i at time t. $n_i(t)$ is the number of infected nodes for user i at time t. η is the average probability that a node is infected.

$$v_i(t) = n_i(t) \cdot \theta\left(T_{check}^i, T_{rest}^i\right) \tag{18}$$

where $\theta(T_{check}^i, T_{rest}^i) \in (0,1]$ is the speed factor. If $v_i(t)$ can reach the maximum, $n_i(t)$ and $\theta(T_{check}^i, T_{rest}^i)$ should both reach their maximum. We discuss them as follows: (1) The condition that $n_i(t)$ reaches the maximum

$$n_i(t) = E\left(\sum_{j=1}^{N_{max}} X_{i,j}^t = Inf.\right) = \sum_{j=1}^{N_{max}} P\left(X_{i,j}^t = 1\right) = \sum_{j=1}^{N_{max}} P\left(X_{i,j}^t = Inf.\right). \tag{19}$$

Because

$$P\left(X_{i,j}^t = Inf.\right) = P_{com} \cdot P_{i \Rightarrow j} \cdot P\left(t\%T_{check}^i = 0, t\%T_{rest}^i \neq 0\right) \cdot \eta + P_{new} \cdot P\left(t\%T_{check}^i = 0, t\%T_{rest}^i \neq 0\right) \cdot \eta \tag{20}$$

According to (19) and (20), thus

$$n_i(t) = P\left(t\%T_{check}^i = 0, t\%T_{rest}^i \neq 0\right) \cdot \eta \cdot \sum_{j=1}^{N_{max}} \left(P_{com} \cdot P_{i \Rightarrow j} + P_{new}\right) = P\left(t\%T_{check}^i = 0, t\%T_{rest}^i \neq 0\right) \cdot \eta \left(P_{com} + N_{max} \cdot P_{new}\right) \tag{21}$$

We can find that $n_i(t)$ depends on $P(t\%T_{check}^i = 0, t\%T_{rest}^i \neq 0)$. If $n_i(t)$ can reach the maximum, $P(t\%T_{check}^i = 0, t\%T_{rest}^i \neq 0)$ must be 1. Because $\lim_{t \to \infty} P_{new} = 0$,

$$T_{rest}^{i,n-1} \le T_{check}^{i,n} \le T_{rest}^{i,n}, n = 1, 2, \cdots, \infty \tag{22}$$

where $T_{rest}^{i,n}$ and $T_{check}^{i,n}$ denote at the nth time user i has the time interval of message checking and the time interval of spending at one location, respectively. (2) The condition that $\theta(T_{check}^i, T_{rest}^i)$ reaches the maximum $\theta(T_{check}^i, T_{rest}^i)$ decides that the cost time t_{cost} when $n_i(t)$ reaches the maximum, thus

$$\theta\left(T_{check}^i, T_{rest}^i\right) \propto \frac{1}{t_{cost}}. \tag{23}$$

According to (23), we know that

$$T_{check}^{i,n+1} > T_{rest}^{i,n} \Rightarrow (n+1) T_{check}^i > n T_{rest}^i. \tag{24}$$

Thus

$$\lim_{n \to \infty} T_{check}^i > \frac{n}{n+1} T_{rest}^i \Rightarrow T_{check}^i > T_{rest}^i. \tag{25}$$

According to (22) and (25), The condition that $v_i(t)$ reaches the maximum is that $T_{check}^i = T_{rest}^i, i = 1, 2, \cdots, M$. □

IV. Simulation Results

In this section, we focus on the scale of infected computers and the temporal characteristic of human mobility in physical networking layer. The spread carrier of social worms is the social network. Thus, the topology of social networking layer is a key component of simulation. We build the topology according to the previous analysis of real social networks [5], [7], which exhibit 'semi-directed", scale-free, assortative, and small-world properties. The topology has 10,000 nodes. We reproduce the degree for each node by the Power-law distribution [7]. Moreover, the infection probability p_{ij} follows the Gaussian distribution [7], the checking time T_{check}^i follows the Exponential distribution [7], and the resting time T_{rest}^i follows the Exponential and Power-law distribution [10], [11]. We draw a compatible propagation simulator from existing simulation models [2], [3], [5], [7]. The implementation is in C++ and Matlab R2014a. The random numbers in the experiments are produced by the C++ TR1 library extensions. The experimental results are averaged by 100 runs. Each run of the spread has two infected nodes at the beginning, which are randomly chosen from the network. Moreover, we set the two nodes with a distance of 6 (the number of edges between them) in the topology, which reflects the impact of the cluster-coefficient [7].

As shown in Fig. 2, other models except our model result in underestimating the infection scale. The main reason is that structural imperfection of the network topology, and previous researchers do not consider the propagation scenario in hierarchical networks. We exhibit the differences Δ in the first inset, and can see the results of previous models deviate from the simulation by two thousands less infections at maximum. There is also a minor divergence between our model and the simulation. Our model is far more accurate than other models. In the second inset, the black solid line denotes the number of infected nodes in the social networking layer at time t, and the

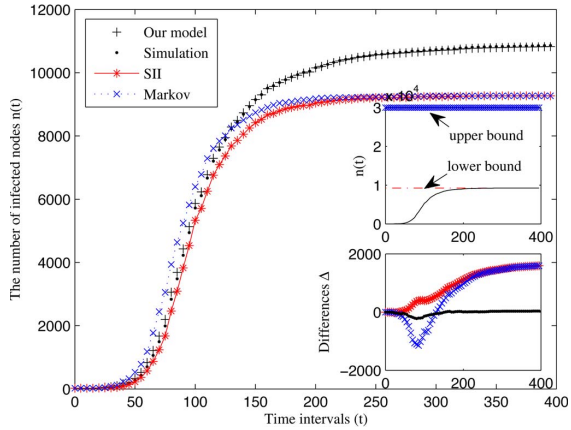

Fig. 2. The comparison of different models and illustration of the infection scale. Topology: $p_{ij} \sim N(0.5, 0.2^2), \alpha = 2.5, < k >= 5.6, \lambda = 0.23$. $T^i_{check} \sim Exp(1/40), P(T^i_{rest}) \propto (T^i_{rest})^{-1.8}$. Δ denotes the differences between the results of our model, previous models and the simulations.

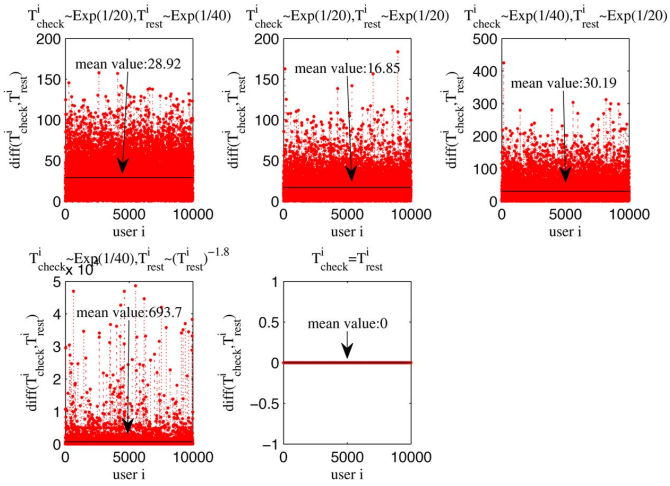

Fig. 3. The difference analysis between checking time and resting time. $diff(T^i_{check}, T^i_{rest})$ denotes the differences of user i's checking time and resting time, which is the non-negative function. Mean value denotes the synchronicity degree of users checking and resting behaviors.

red dotted line denotes the lower-bound of the infection scale, which is the maximum number of infected nodes in the social networking layer. The blue dotted line denotes the upper-bound of the infection scale.

As shown in Fig. 4, according to the different situations of two behaviors in Fig. 3, we can see that the spread speed of social worms mainly depends on the difference degree between their sample distributions. The experimental results show the fastest spread speed satisfy the condition in Theorem 1. This implies that synchronicity of different human behaviors can promote the spread of social worms.

The discrete-quantitative analysis and simulation results imply:

Remark 1: The presented model is able to address the problem of infection scale underestimation, which results from modeling imperfectly of network topology structure and human behavior. At the same time, we give the quantitative range of the infection scale of social worms.

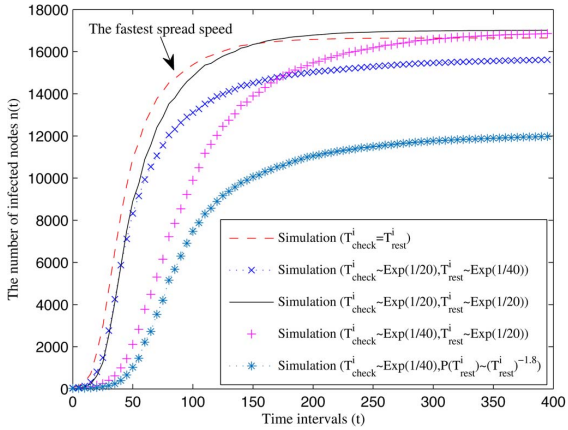

Fig. 4. The comparison with different relationships between checking time and resting time. Topology: $p_{ij} \sim N(0.5, 0.2^2), \alpha = 2.5, < k >= 5.6, \lambda = 0.23$.

Remark 2: Synchronicity of two human behaviors can promote the spread of social worms. Under the certain condition, which is the same as the temporal periods of checking and resting behaviors of every user, the spread speed of social worms is the fastest.

V. Conclusion

This letter presents a novel model to capture the propagation dynamics of realistic social worms more accurately. From two aspects of the infection scale and the spread speed, we conduct further research on the spread characteristics of social worms theoretically and practically. Simulation results indicate the correctness of our work, and researchers could adopt our conclusions to develop network defense strategies.

References

[1] M. Fossi and J. Blackbird, "Symantec Internet security threat report 2013," Symantec Corp., Mountain View, CA, USA, Tech. Rep., Apr., 2014.

[2] G. Yan, G. Chen, S. Eidenbenz, and N. Li, "Malware propagation in online social networks: nature, dynamics, and defense implications," in *Proc. 6th ACM Symp. Inf., Comput. Commun. Security*, 2011, pp. 196–206.

[3] C. Gao, J. Liu, and N. Zhong, "Network immunization and virus propagation in email networks: Experimental evaluation and analysis," *Knowl. Inf. Syst.*, vol. 27, no. 2, pp. 253–279, 2011.

[4] W. Fan, K. Yeung, and K. Wong, "Assembly effect of groups in online social networks," *Phys. A, Statist. Mech. Appl.*, vol. 392, no. 5, pp. 1090–1099, Mar. 2013.

[5] C. C. Zou, D. Towsley, and W. Gong, "Modeling and simulation study of the propagation and defense of Internet e-mail worms," *IEEE Trans. Dependable Secure Comput.*, vol. 4, no. 2, pp. 105–118, Apr.–Jun. 2007.

[6] Z. Chen and C. Ji, "Spatial-temporal modeling of malware propagation in networks," *IEEE Trans. Neural Netw.*, vol. 16, no. 5, pp. 1291–1303, Sep. 2005.

[7] S. Wen *et al.*, "Modeling propagation dynamics of social network worms," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 8, pp. 1633–1643, 2013.

[8] S. Wen *et al.*, "Modeling and analysis on the propagation dynamics of modern email malware," *IEEE Trans. Dependable Secure Comput.*, vol. 11, no. 4, pp. 361–374, Jul./Aug. 2014.

[9] C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," *Nature Phys.*, vol. 6, no. 10, pp. 818–823, 2010.

[10] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang, "Understanding metropolitan patterns of daily encounters," *Proc. Nat. Acad. Sci.*, vol. 110, no. 34, pp. 13774–13779, 2013.

[11] X. Liang, J. Zhao, L. Dong, and K. Xu, "Unraveling the origin of exponential law in intra-urban human mobility," *Sci. Rep.*, vol. 3, pp. 1–7, 2013.