Modeling and Analysis Malware Spread in Short-range Wireless Networks

Shuai FU[1], Chang-guang WANG[1,2*], Li-jing BAI[1],Qing-yang HU[1]
1College of Physics Science and Information Engineering, Hebei Normal University, Shijiazhuang 050016, China
2 Hebei Advanced Thin Films Laboratory
Shijiazhuang 050016, China
wangcg@mail.hebtu.edu.cn

Jian-feng MA
School of Computer
Xidian University
Xi'an 710071, China
jfma@mail.xidian.edu.cn

***Abstract*—The growing popularity of short-range wireless networks makes them increasingly attractive to malware writers, and malware targeting portable wireless-enabled devices such as Bluetooth-enabled smart phones has already begun to appear. In this paper, using the theories of epidemics, we present a propagation model of malware in short-range wireless networks and investigate the propagating behaviors of the malware in such networks by means of simulations. This model can better reflect the specific characteristics of devices mobility, patterns of nodes aggregation and the wireless nature of malware transmission. The simulation results show that the agreement between malware spread in mobile devices and this model is good.**

Keywords- *malware; propagation model; Bluetooth; short-range wireless networks; portable devices*

I. INTRODUCTION

Recent years have witnessed the widespread adoption of portable computing devices which are equipped with a short-range wireless technology such as WiFi[1] or Bluetooth[2]. Wireless connectivity is greatly convenient for its little limitation on users' mobility and allows a great deal of flexibility.

The growing popularity of short-range wireless networks and mobile devices is starting to attract unwanted attention especially as potential targets for malicious activities reach critical mass. Malware is a generic term to denote all kinds of unwanted software (e.g., viruses, worms[3], or Trojan horses). Such software poses a major security threat to computer users. These wireless malware exploit such wireless capabilities in order to propagate themselves between nearby devices, usually without any active user involvement. One notable characteristic of such malware is that they do not require Internet connectivity for their propagation, thus can spread without being detected by existing security systems. Another characteristic is that, since this kind of wireless malware mainly targets portable devices and can exploit the mobility of users for its propagation, its propagation behavior is similar to the spread of human diseases in the population.

Malware attacks on the Internet have been the subject of extensive empirical, theoretical and simulation studies[3-6]. Such studies are of great importance to the design of more effective immunization strategies to prevent and combat Internet epidemics. However, due to the properties of the short-range wireless networks, the propagation model of malware on Internet can not be directly applied to such networks. There have been very limited studies of the malware propagating in short-range wireless networks[2,7].

In this paper, we present a model for the propagating of malware in short-range wireless networks and investigate the properties of malware in such networks through simulations. The rest of this paper is organized as follows. In section II, we characterize the model of short-range wireless networks. In section III, we develop the model of malware propagating in the short-range wireless networks. In section IV, we implement the simulation experimentation and analyze the simulation results. In section V, we close this paper with our conclusions.

II. THE MODEL OF SHORT-RANGE WIRELESS NETWORKS

Short-range wireless networks are constituted by mobile devices such as laptops and smart phones which are equipped with short range radio transmission. Communication is possible between devices within each other's radio range.

A. Modeling the short-range wireless networks

From the point of view of the complex network theory, the topology of the short-range wireless networks is a spatial network. The interactions between the nodes in such networks is a function of their spatial distance[8]. We consider a set of nodes distributed in a two-dimensional plane which communicate using short-range radio transmissions. The strength of the radio signal at a device *r* received from another device *s* decays with the distance between the sender and the receiver. Therefore, we use the path-loss model[10] to describe the network. In the path-loss model, the mean value of the signal power at the receiving device *r* is related to the signal power of the sending device *s* as follows:

$$P_r = \frac{P_s}{c d_{sr}^{\alpha}}, \quad (1)$$

Supported by National High Technology Research and Development Program of China(2007AA01Z429,2007AA01Z405), the Major Program of National Natural Science Foundation of China (60633020),the Science Foundation of Hebei Normal University.

*Correspondence to: Chang-guang WANG

978-1-4244-3693-4/09/$25.00 ©2009 IEEE

where d_{sr} is the Euclidean distance between node s and node r; P_s and P_r are the sending and received power, respectively; c is a constant that depends on a number of factors including the transmission frequency. Parameter α varies between 2 and 5 depending on the specific indoor/outdoor propagation scenario, and for free space propagation its value is 2[1]. The data sended by node s is correctly received at node r on condition that:

$$\frac{P_{sr}}{v} = \frac{P_s/(cd_{sr}^{\alpha})}{v} \geq \beta_0 , \quad (2)$$

where β_0 is an attenuation threshold and v is the noise level at node s. If equation (2) holds, we say s can establish a communication link with r. Equation (2) can be converted into a maximum transmission range for node s:

$$R = (\frac{P_s}{c\beta_0 v})^{\frac{1}{\alpha}} , \quad (3)$$

so each node can establish wireless links with only those nodes within a circle of radius R.

We can, however, assume that all devices use the same transmit power. Consequently, they have the same maximum transmission range R.

B. Modeling the interaction between mobile devices

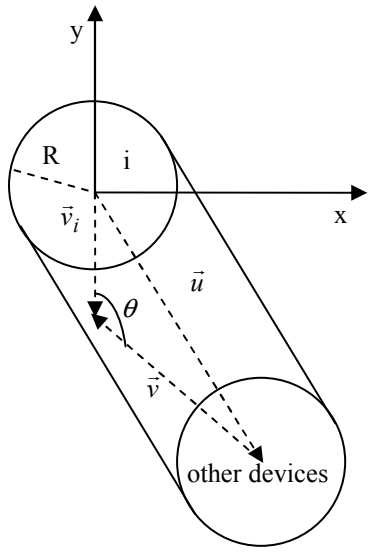

Fig.1 The interaction between a specified device and other devices in the network

Assuming that each device moves independently of the others in the plane, as the communication range of each device is R, each specified device can form a circular contact domain with radius R, taking itself as the center. Again we assume that each device moves with a constant straight-line velocity $\vec{v}$, with their direction vectors distributed uniformly in azimuth. If one of the devices passes within the contact domain of another's, we call a contact being made. Given a specified node, i, which is situated at the coordinate origin, it moves with velocity $\vec{v}_i$ along the negative y axis(dashed line) while another node q moving close to it with velocity vector $\vec{v}$.

As illustrated in Fig.1, these two vectors are at an angle θ, with a relative velocity vector $\vec{u}$. Firstly, it is necessary to find what conditions it should satisfy to make a contact with i during a time period dt. For this purpose, we first calculate the domain area swept by i during its motion. The relative velocity of them is $\vec{u} = \vec{v}_i - \vec{v}$. Thus we have $u = \sqrt{v_i^2 + v^2 - 2v_i v\cos\theta}$. Since the contact domain of each node is equal to the circular area with radius R, the domain that node i sweeps during dt is a lozenge-shaped area. It is formed by the circular starting from the coordinate origin and then moving along the direction of $\vec{u}$, which can be seen clearly from the dashed line in Fig.1. Its area is:

$$dA = 2Rudt \quad (4)$$

so we can conclude that, if node q is in the domain of dA during dt, the contact will be made between node i and q.

III. THE PROPAGATION MODEL OF MALWARE

Several previous studies have analyzed and modeled the propagation of malware on the Internet and wireless networks [2,4,5,6,7]. We assume that malware use unicast technology to attack the short-range wireless networks and use the SIS model which is used to describe the epidemics spreading in the population[8] to characterize the malware spreading in such networks.

According to the SIS model, we assume the nodes in the short-range wireless networks to be in one of the following two states: S(Susceptible) and I(Infected). Then we use $S(t)$ and $I(t)$ to denote the susceptible nodes and infected nodes at the time of t, respectively. Thus the densities of susceptible and infected nodes can be denoted as $s_t = S(t)/N$, $i_t = I(t)/N$, respectively.

A. The contact rate of mobile devices

In the above section, we have provided the conditions under which any two nodes will make contact by describing the motion of two specified nodes. We assume that the probability a single infected node transmitting malwares to any other susceptible nodes is p. If a susceptible node passing within the contact domain of any infected node, it may be infected. In order to obtain the propagation characteristics of malwares, we need to calculate the contact rate between an infected node and all the other nodes in this network.

Consider a specific infected node z. As all the nodes move independently of the others with their direction vectors distributed uniformly in azimuth, the density of the nodes from θ to $\theta + d\theta$ is $\rho_s d\theta/2\pi$. According to Eq.(4), the area swept during dt is: $dA = 2Rudt$. Therefore, the number of nodes entering the swept area in dt is:

$$dN = \rho_s Rud\theta dt/\pi \quad (5)$$

As the contact domain of node z is a circular with radius R, using itself as the center, the number of nodes making contact with z is:

$$dN_R = \int_0^{2\pi} \rho_s R u d\theta dt / \pi \quad (6)$$

Therefore, the average contact rate of node z is:

$$\eta = \int_0^{2\pi} \frac{dN_R}{dt} = \frac{\rho_s R}{\pi} \int_0^{2\pi} u d\theta \quad (7)$$

Substituting for u gives:

$$\eta = \frac{\rho_s R}{\pi} \int_0^{2\pi} (v_i^2 + v^2 - 2 v_i v \cos\theta)^{1/2} d\theta \quad (8)$$

Eq.(8) reduces to

$$\eta = \frac{4\rho_S R}{\pi} (v + v_i) \int_0^{\pi/2} (1 - h \cos^2 2r)^{1/2} dr \quad (9)$$

where $h = 4vv_i / (v + v_i)^2$. For simplicity, it can be assumed that all the nodes in this network have the same speed $\vec{v}$. Eq.(9) can be changed to

$$\eta = \frac{8}{\pi} R \rho_s v \quad (10)$$

It is generally the case that nodes in a network move with a range of different speeds. It has been shown in Ref.[9] that when the node speeds are distributed according to a Maxwell-Boltzmann distribution, for example, Eq.(10) becomes

$$\eta = \frac{8}{\pi} R \rho_s \bar{v} \quad (11)$$

where $\bar{v}$ is the mean node speeds. So we can use this framework to accommodate a distribution of nodes' speeds.

B. *Modeling the propagation of malware*

As is widely used in epidemic modeling, we have the standard frequency dependent mass-action assumption [10]:

$$\frac{1}{A} \frac{dI}{dt} = \beta \frac{SI}{N} \quad (12)$$

where β is the transmission rate and N is the total population size. S is the number of susceptible notes while I represents the infective devices.

According to the model of $\beta SI / N$, $p = I / N$, Eq.(12) can be changed to:

$$\frac{1}{A} \frac{dI}{dt} = \beta S p \quad (13)$$

As the contact rate of any infected node in this network with others is $\eta = \frac{8}{\pi} R \rho_s \bar{v}$, and the infection probability is p, the infection behavior of each node is:

$$\frac{dI_0}{dt} = \frac{8}{\pi} R \rho_s \bar{v} p \quad (14)$$

Therefore, the total infection behavior is:

$$\frac{dI}{dt} = \frac{8}{\pi} R \rho_s \bar{v} p I \quad (15)$$

So the infection behavior of infective nodes per unit area is:

$$\frac{1}{A} \frac{dI}{dt} = \frac{8}{\pi} R \rho_s \bar{v} p \rho_i \quad (16)$$

Combining Eq.(13) with Eq.(16), we have

$$\beta = \frac{8 R \bar{v} \rho_i}{\pi N} \quad (17)$$

As proposed above, $p = I / N$, so Eq.(17) can be converted to:

$$\beta = \frac{8 R \bar{v} p}{\pi N} \quad (18)$$

IV. SIMULATION EXPERIMENTATION

We use the above model to simulate malware propagating in the short-range wireless networks. Let the network consist of 3000 nodes distributing an A=1000m*1000m area. The transmission range of all devices was set at 5m, which is somewhere inside the typical (10m) range of the Bluetooth systems. All the nodes have a mean speed of about 2km/day. The mean infectious duration is 5 days. It is assumed that channel access protocols such as medium access control (MAC) [5] have minimal impact, as there are rarely any other competing devices within the transmission range of an infected device.

A. *The simulation parameters*

In this simulation we consider a typical node density of 3000 nodes/km^2. All the nodes have a mean speed of about 2km/day, and each node's transmission radius is 5m, which corresponds to that seen in a Class 2 Bluetooth device. The transmission probability p=0.1 and a mean infectious duration of 5 days. It is assumed that channel access protocols such as medium access control (MAC) [11] have minimal impact, as there are rarely any other competing devices within the transmission range of an infected device.

B. *Analysis of the simulation results*

1) *The time relation of infective nodes*

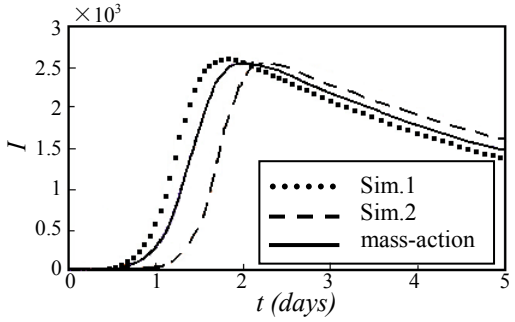

Fig.2 Time relation of the number of infected devices

First, we perform 3 separate simulations. Fig.2 shows a comparison of the time-dependence of the number of infected devices based on the mass-action model with those obtained from individual simulations of the process, and the equivalent mass-action model result is also shown in it. From Fig.2 we can see that each simulation is well described by the mass-action model, though it produces more or less fluctuations. However, the initial dispositions of the susceptible and infective nodes determine when the propagation start to take

off, and how long it will take the index node to complete the transmission to the first susceptible. So there exists some deviation from the mass-action model.

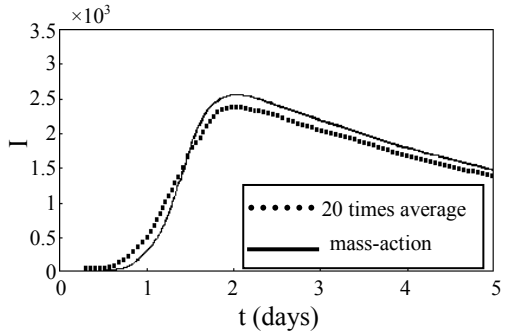

*Fig.*3 Comparison of the mean of 20 simulations and the mass-action model

Fig.3 shows the ensemble average of twenty realizations of the simulation. Sufficient simulations were performed to produce a statistically stable result. As we have discussed above, the difference among the initial dispositions results in the fluctuations of each realization. Averaging them captures the fluctuations that this produces. So it can be seen in Fig.3, the ensemble average is less similar to the mass-action result.

2) The relationship between R and the number of infected nodes

In this set of simulation experiments, we increase the node's transmission range R from 10m to 40m and investigate what will happen. Now it is more typical for class 1 Bluetooth devices. Other simulation parameters stay the same as previous. Fig.4 shows a series of propagation curves for increasing wireless transmission range R . We can observe several changes as follows when R increases from 10m to 40m.

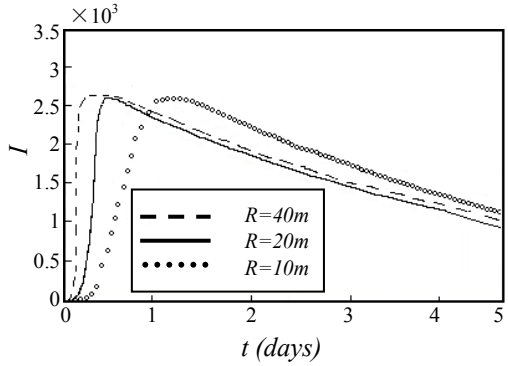

Fig.4 The relationship between R and the number of infected nodes

Firstly, if the total number of nodes keeps the same, as R increases, the time needed to reach the maximum number of infected nodes diminishes. It is clear that the contact domain of any node will become larger as R increases, and then more susceptible nodes will pass within this area. Thus, less time will be taken to infect the same number of nodes. That is to say, the greater the value of R is, the shorter the time needed to infect the maximum number of nodes.

Secondly, as R increases, the less time the malicious will take to complete the first infection case. This is because the larger value of R corresponds to a greater number of susceptible nodes, and the number of susceptible nodes passing within a specified infective node's contact domain becomes greater. So for this infected node, it has more chances to contact a susceptible one. From this we can see that, an infective node will spend less time to complete its first infection when the value of R increases.

Thirdly, when the number of infective nodes reaches the peak, the infection tendency will keep the same as each other. The changes of R has no impact on malicious behavior.

V CONCLUSION

In this paper, we have developed a propagation model of malware in short-range wireless networks. With help of the interaction rate of mobile devices, this model can reflect how the wireless nodes are influenced by factors such as host speed, malware transmission probability and device transmission range. It can also suggest that, as device transmission range increases, the influence of channel allocation algorithms plays an increasing role in affecting the probability of worm transmission. Simulation results show that, the agreement between malware spread in mobile devices and this model is good. An understanding of the propagation properties of the malware in short-range wireless networks is very important for the design of effective detection and prevention strategies for such networks.

REFERENCES

[1] P. Santi, D. Blough. "The critical transmitting range for connectivity in sparse wireless ad hoc networks," IEEE Trans. on Mobile Computing, vol.2, pp.25-29, 2003

[2] C. G. WANG and J. F. MA, "Malware propagation model in wireless Bluetooth networks" Journal of Xidian University. vol.36, no.1, pp. 94–98, 2009.

[3] Jose Nazario. Worm blog. http://www.wormblog.com/im_worms/.

[4] S. Staniford, V. Paxson and N.Weaver. "How to own the Internet in you spare time," In 11th USENIX Security Symposium, San Francisco: USENIX Press, 2002, pp.149-167.

[5] H.Hu, S.Myers, V.Colizza, A.Vespignani, WiFi Epidemiology: Can Your Neighbors'Router Make Your Sick?(2007).ArXiv:0706.3146.

[6] C. G. WANG and J. F. MA, "Malicious code modeling and analysis in weighted scale-free networks," Wuhan University Journal of Natural Sciences. vol.2, pp. 52–54, 2007..

[7] J. W. Mickens, B. D. Noble."Modeling epidemic spreading in mobile environments," In Proc. of the 4th ACM workshop on Wireless security. Cologne, Germany: ACM Press, 2005, pp.77-86.

[8] X. F. WANG, X. LI and G. R. CHEN. "The Theories and Applications of Complex Networks," Beijing,China: Tsinghua University Press, 2006, pp.72-98.

[9] CJ.Rhodes,R.M. Anderson,Contact rate calculation for a basic epidemic model,Mathematical Biosciences(2007).

[10] H.McCallum,N.Barlow,J.Hone,How should pathogen transmission be modelled?Tr.In Ecol and Evol.16(2001)295-300.

[11] M.Nekovee, Worm epidemics in ad-hoc network, New J.of Phys 9 (2007a) 189-202.