On Modeling Viral Diffusion in Heterogeneous Wireless Networks

Hoai-Nam Nguyen* and Yoichi Shinoda

Japan Advanced Institute of Science and Technology
#1-518, 1-8 Asahidai, Nomi, Ishikawa, Japan 923-1211
{namnh,shinoda}@jaist.ac.jp

Abstract. Smart phones and computers now are able to co-work in a wireless environment where malware can propagate. Although many investigations have modeled the spread of malware, little has been done to take into account different characteristics of items to see how they affect disease diffusion in an ad hoc network. We have therefore developed a novel framework, consisting of two models, which consider diversity of objects as well as interactions between their different classes. Our framework is able to produce a huge result space thus makes it appropriate to describe many viral proliferating scenarios. Additionally, we have developed a formula to calculate the possible average number of newly infected devices in the considered system. An important contribution of our work is the comprehension of item diversity, which states that a mixture of device types causes a bigger malware spread as the number of device types in the network increases.

1 Introduction

Malware such as viruses, worms, or malicious codes/programs has long posed a significant threat to computers, regardless of fixed or mobile ones, but so far mobile phones have been relatively safe. Recently, however, attacks by various types of viruses, trojans, and spyware on mobile phones, especially smart phones, have been increasing. Reports about worms such as **cabir** [2], [5] and trojans such as **skulls** and **mquito** [5] raise serious questions about mobile phones' vulnerabilities. Such malicious codes[1] are able to infect many users under appropriate conditions. For example, a sports event in Helsinki in August 2005 and a public concert in Germany, Live8, were marked by outbreaks of viruses [10]. Another peril when a phone is under control of a hacker, which can be accomplished using a virus, is described in [16], [17]. They state that it is possible for a malicious user to break down a phone network using SMS messages.

Smart phones, which have been undergoing an explosive growth in sheer numbers, along with the commensurate improvements in computing ability and generality, have become an attractive target for malware. Improvements in hardware and software for phones, especially their communications capabilities, create a myriad of avenues for

* Corresponding author.

[1] We will also call virus, worm, and trojan malicious code, malicious program, malware, or malcode if not specifically mentioned.

A.U. Schmidt and S. Lian (Eds.): MobiSec 2009, LNICST 17, pp. 238–252, 2009.

hackers to exploit and attack. There is malware capable of dwelling in both computers and phones so it is certainly possible for malcode to hit phones from computers and vice versa. In general, then, malicious threats can infect computers or phones in one of the followings ways [3]:

(1) Through communication interfaces such as Bluetooth, Wifi, or infrared. A malware can directly contaminate other devices whenever they are reachable and connectable and they have exploitable weaknesses. We count computers that have wireless interfaces as susceptible objects of contagion as well. Note that, for this propagation relation, we have to consider the bidirectional effect, since a malicious program can move not only from computer to computer but also from computer to phone and vice versa.
(2) Through user download programs infected with the malware.
(3) By using the Multimedia Messaging Service (MMS) to diffuse to other phones.
(4) By a memory card which already contains a malicious program. When the user uses that card in a phone or computer, the compromised program transmits itself to the device.

Of the above ways, only (3) is restricted to phone-phone relations, whereas the others are risky for both computers and phones.

In addition to the improved device technologies, the mobility of users is another factor that makes malware spread wider and faster. Indeed, when a user with an infected device is on the move, he/she is likely to make more contacts, thus giving malicious code an opportunity to transfer to more hosts.

In this paper, we develop a novel analytical framework, by which we are able to explore how viral transmissions spread through a network containing devices that are heterogeneous in terms of various characteristics (such as device types). Here in our framework, each network item experiences states from susceptible to infected. These states with variability and changes are carefully described by mathematical parameters. One of the most advantageous points of our models is their huge result space. Indeed, if we use different combinations of parameters, we consequentially get various outputs that correspond to many spreading scenarios of malware. Our framework also takes into account all the contamination methods mentioned above.

We then use that framework to derive the necessary conditions for the existence and persistence of malware. With the framework, it is able for us to predict the possible average number of newly infected individuals and that may assist network administrators to have a right strategy to combat or mitigate the malware. Investigations of malware free equilibrium and its locally asymptotic stability are additional contributions of our work. Moreover, we check whether diversity of participant characteristics in a network has any influence on virus/worm propagation. It is quite intuitive to see that viral dissemination is wider if items in a network possess more characteristics, but we have gone further to theoretically prove this statement.

The rest of this paper is organized as follows: in Section 2, we review and discuss the related works. This section also differentiates our work from other studies reported in the literature. Next, Section 3 describes the deterministic analytical models in depth. We present a detailed discussion of the proposed models in Section 4. Numerical results

of conducted simulations are given in Section 5. Finally, Section 6 contains concluding remarks and outlines our future work.

2 Related Works and Motivation

Traditional networks with fixed nodes connected through wired media have suffered many serious virus/worm outbreaks such as those created by the Melissa [1] or Codered [7], [20] worms. With improvements in advanced wireless technology, networks with laptops having wireless interfaces have grown rapidly in both quantity and complexity. A group of nodes that communicate via wireless interfaces, forming a network without any pre-defined topological structure, is called a wireless ad hoc network. In this type of network, a node can move freely and contact other nodes through wireless interfaces. Modeling the spread of malware on this type of network has received considerable attention from researchers, [6], [9] to name a few. Nevertheless, these works have not taken smart phones into account, a new source of inspiration for attackers.

Ramachandran and Sikdar [13], [14] have studied the spread of malware in a phone network, assuming that nodes move from one patch to the others at a specific rate. They incorporate movement, or in their words, "the heterogeneity of locality," into their model to make it more precise. However, they do not take into consideration the inner characteristics of each class of objects in the network. They indirectly suppose the same condition for all devices by delivering the same values for parameters. As a result, the model is unable to distinguish individual classes if diverse entities exist concurrently. In addition, these studies do not contain an analysis of the relationship between the number of patches and the spread of the disease.

One of our published papers, [11], proposes a 4-compartment model to describe how viral forces propagate in a wireless network. We used the same reasoning and observation but adding one more state that a susceptible item may experience while participating in a network. The added latent state is understandable as the incubation period of the virus. That paper, however, does not provide an in-depth investigation on the latent state such as calculating the average time a network item has to be in if it is exposed to the spreading malware.

Proposing a model that will depict all characteristics of a set of connectable and reachable nodes, computers and phones, is a difficult task since we have to include as many parameters as possible. In our research, one major input of interest is the fundamental difference between classes of objects in the network[2]. We classify and arrange items in a network into different groups based on their characteristics. It is understandable as item diversity and we aim to fill the gaps of not taking care of device diversity left by previous research. With that goal, our models take the following aspects into consideration:

– **heterogeneity of entity**, i.e., the existence of different types of devices in the network

[2] For ease of reading, we now refer to a network either as a collection of computers and phones, computers only, or phones only, that all reside in a considered territory.

– interactions between different classes of entities. As a matter of fact, there are many types of operating system (OS) for smart phones and each of which has its own particular vulnerabilities for attackers to exploit. Consequently, we count phones with the same OS as items in the same class while those with different ones are put into different classes. Note that computers are recognized as a separate set.
– diversity of entities. We examine whether diversity has an impact on the persistence and stability of a disease as well as its spread dynamics.

3 Modeling Disease with Heterogeneity of Nodes

We built up our model using a typical compartmental epidemic model for two cases, each of which has three entity classes. Under different assumptions, the last class in each model differs while the rest remain the same. Here is the description of the two cases: at an arbitrary point in time, a device is in one of the following three classes: susceptible (S), infected (I), and moved back to susceptible state (S) for the first case or recovered and immune to re-infection (R) for the second case. Every object in the network is initially susceptible to disease if it is not already infected. The object stays in the susceptible phase until it comes into contact with an infectious one or with the malware. As soon as the malware propagates to a susceptible item, the item moves to the infected state. For the first case, when a device is cured of disease, it recovers and therefore is subject to re-infection if no immunity has been introduced. This case is the SIS model. In the second case, after the end of contamination, a device produces resistance to the disease and is removed from the infected condition and does not progress back to susceptible but recovered and immune class. The SIR model represents this case.

3.1 Assumptions

In this section, we state our assumptions. We assumed that multiple items can be infected but one species is the primary infectious vector. Since computers suffer by far the most attacks and they outnumber other devices in terms of the diversity of malware, we made computers the main attack force of offenders while the others are spillover species. At the moment, computer viruses do not usually try to propagate to mobile phones. However, as malware develops, it is clear that this could be a forthcoming problem, possibly in the near future [3]. Because of the higher possibility that malcodes transfer from computers to smart phones compared to that from the reverse direction, we chose computers to be the main disease reservoir. Another assumption we made is that computers in a network are of the same kind. Worms often exploit holes in software and there are many applications that run on both computers and phones under different operating systems. Accordingly, malware is supposedly able to diffuse to nodes using distinct operating systems.

When several phone users coincidentally gather in an area and make calls or send messages, it is likely that those overcrowded occurrences will overwhelm the capacity of the network. It means that as more phones come into the congestion, the less service they receive from the operator, and thus infection vectors for phones such as MMS or dialing are closed off. The same thing may happen to computers as well, but in a dissimilar scenario. For example, in an ad hoc network where data moves from computer

to computer in a hop-by-hop manner, there is a possibility that one computer would serve as the main transporter between two components of the network. The more data sent through this bottleneck, the greater the risk that this computer will become overloaded and break down. At a certain level, the collapse of this computer will result in a disconnected component from the network. All items in that detached component are considered dead. The above scenarios lead us to the following definition: The number of devices that malware can not use for propagation due to the congestion is called *density-dependent death rate*.

3.2 Elaborate Models

In this section, we first derive the SIS epidemic model and then formulate the SIR model for n different types of hosts. As mentioned, one species is considered the main reservoir while the rest, $n-1$ species, act as spillover ones.

Let S_1 and I_1 denote the susceptible and infected reservoir population and S_k and I_k the susceptible and infected spillover populations, $k = 2, \ldots, n$. Therefore, the total population size for each type k is $N_k = S_k + I_k$, $k = 1, \ldots, n$. We then find an equation to describe the changes in the susceptible set of each population. We denote by b_k the rate at which each species has more entities, $k = 1, \ldots, n$. In other words, b_k illustrates the birth rate of the kth type. The total number of new-born units in the kth group is then calculated by $N_k b_k$.

Next we denote by $d_k(N_k)$ the density-dependent death rate of the kth type. Note that this parameter depends on the total population size of the kth group. In the following formulae, we shorten it by using d_k. Consequently, the quantity of out-of-order disease-unrelated units from S_k is $S_k d_k$.

To demonstrate the disease-related interactions between groups through communicating interfaces such as Wifi, Bluetooth, or infrared, we define a factor $\alpha_{ki}(N_i)$, dependent on N_i, for each pair of device types k and i, $k = 1, \ldots, n$ and $i = 1, \ldots, n$. This rate is understood as an infected item in group i tries to infect a susceptible individual in group k. The rate magnitude of each type i is proportional to the ratio of infective pieces over the total population size, which is given by $\frac{I_i}{N_i}$. Scanning through all populations, the S_k group loses total $S_k \sum_{i=1}^{n} \alpha_{ki} \frac{I_i}{N_i}$.

When a user downloads a program containing malware and incautiously executes that program, the device in use (computer or phone) directly moves to the infective target compartment. Let β_k be the rate at which items in S_k download some program from the Internet and let γ be the probability that a program contains malware. For ease of analysis, we assume that when a device is infected by a specific malware, it will not be compromised by other malcodes. Accordingly, malware only affects susceptible items but not already infected ones. This argument gives us the number of entities that move from S_k to I_k is $\beta_k \gamma S_k$.

The third mechanism through which malware can be dispersed is use of the MMS service. A compromised phone may randomly choose a number from the contact list or generate a random number to dial then send the malware enclosed in an MMS message. The rate of infected devices in the kth type that try to dial other numbers is referred to as η_k. Nonetheless, some of the dialed numbers are out-dated (if chosen from contact list)

or do not exist (if randomly generated) so those attempts may fail sometimes. For that reason, we use ϑ as the probability that the malware succeeds in sending messages to other numbers. We also note that when the target phone is off, it belongs to the exposed class since there is a poisoned piece awaiting it in the queue and when it is on, the phone directly moves to the infected class. However, when we count total loss of the susceptible population, this method of dissemination shifts items from susceptible to exposed and infected classes a change of $-\eta_k \vartheta S_k \frac{I_k}{N_k}$.

The last carrier for malware is memory cards. Analogously, we call μ_k the rate at which users in susceptible group S_k use memory cards and τ the probability that a card contains at least one malware. Hence, this carrier subtracts from S_k an amount of $\mu_k \tau S_k$.

Finally, the recovery rate, the rate at which devices are patched, of type k is denoted by φ_k so the total number of recovered pieces is $\varphi_k I_k$.

Taking all into account, changes in S_k, $k = 1, \ldots, n$, can be formulated as follows:

$$\frac{dS_k}{dt} = N_k b_k - S_k d_k + \varphi_k I_k - S_k \left(\sum_{i=1}^{n} \alpha_{ki} \frac{I_i}{N_i} + \beta_k \gamma + \eta_k \vartheta \frac{I_k}{N_k} + \mu_k \tau \right). \quad (1)$$

We use λ_k to demonstrate the disease-related death rate at which infected entities are inoperative by the malware. With the same reasoning the following equation tells us about variations in each I_k, $k = 1, \ldots, n$:

$$\frac{dI_k}{dt} = -I_k d_k - (\varphi_k + \lambda_k) I_k + S_k \left[\sum_{i=1}^{n} \alpha_{ki} \frac{I_i}{N_i} + \beta_k \gamma + \eta_k \vartheta \frac{I_k}{N_k} + \mu_k \tau \right]. \quad (2)$$

Equations (1) and (2) form the SIS epidemic model.

To obtain the SIR model, we add a compartment of entities that produce immunity after disease recovery. For each type k, $k = 1, \ldots, n$, a class of recovered and immune items, R_k, is added to the existing model (1) and (2) . Similarly, we build up the SIR model as follows:

$$\frac{dS_k}{dt} = N_k b_k - S_k d_k - S_k \left(\sum_{i=1}^{n} \alpha_{ki} \frac{I_i}{N_i} + \beta_k \gamma + \eta_k \vartheta \frac{I_k}{N_k} + \mu_k \tau \right), \quad (3)$$

$$\frac{dI_k}{dt} = -I_k d_k - (\varphi_k + \lambda_k) I_k + S_k \left[\sum_{i=1}^{n} \alpha_{ki} \frac{I_i}{N_i} + \beta_k \gamma + \eta_k \vartheta \frac{I_k}{N_k} + \mu_k \tau \right], \quad (4)$$

$$\frac{dR_k}{dt} = -R_k d_k + \varphi_k I_k. \quad (5)$$

All parameters are likewise understood as in the SIS model, except for the φ_k one, which accounts for the rate at which infected entities recover and move to immune compartment R_k. The total population is the sum of the three compartments, that is: $N_k = S_k + I_k + R_k$, $k = 1, \ldots, n$.

Note that all parameters are assumed to be greater than or equal to zero.

4 Discussion on the Models

4.1 Assumption and Parameter Discussion

In this part, we further discuss about the assumptions, parameters, and the models. To distinguish the dynamics of the main reservoir from other species, we assume a higher intrinsic transmission and scanning rate from the reservoir than from the spillover species or between groups of spillover species. Mathematically that is,

$$\alpha_{11}(N_1) \geq \alpha_{k1}(N_1) \geq \begin{cases} \alpha_{ki}(N_i), & k, i = 2, \ldots, n \\ \alpha_{1i}(N_i), & i = 2, \ldots, n. \end{cases} \tag{6}$$

When no infection occurs, we have $\lim_{t\to\infty} N_k(t) = H_k$, where H_k is the carrying capacity for species k. Carrying capacity is the maximum number of individuals that a system can support without degradation. This matches our assumption of the natural density-dependent death rate.

Next we derive the disease-free equilibrium (DFE) for each model. The DFE for (1) and (2) is the unique solution satisfying $S_k^* = H_k$ and $I_k^* = 0$, $k = 1, \ldots, n$. Thus the equilibrium point, Q, is:

$$Q_0 = (S_1^*, I_1^*, \ldots, S_n^*, I_n^*) = (H_1, 0, \ldots, H_n, 0). \tag{7}$$

On the contrary, a nonnegative endemic equilibrium point Q_1 exists with $I_k^* > 0$ which requires that $I_i^* > 0$ when $\alpha_{ik}(N_k) > 0$. That means if there is transmission between two species groups, the persistence of the disease in one can result in persistence of the disease in the other. The DFE for model (3), (4), and (5) is $S_k^* = H_k$ and $I_k^* = R_k^* = 0$, $k = 1, \ldots, n$.

4.2 Analysis of Disease Extent

The malware may have an opportunity to spread throughout the network or it may die out at a certain time. In this section, we study the stability of the disease and derive the necessary conditions for its stable state. We compute the basic reproduction number $\mathcal{R}_0$ to examine the mean number of possible secondary infected cases introduced into a susceptible population by a single infected case [8]. This number describes the possible average number of newly infected devices and it will be proven that this number is identical for the proposed SIS and SIR models.[3]

Theorem 1. *Let $\rho(M)$ denote the spectral radius of matrix M. The possible average number of newly infected items, $\mathcal{R}_0$, for models* (1)*,* (2) *and* (3)*,* (4)*,* (5) *is given by the spectral radius of the matrix $M_n = (\mathcal{R}_{ij})_{n\times n}, i, j = 1, \ldots, n$,*

$$\mathcal{R}_0 = \rho(M_n),$$

[3] "Basic reproduction number" and "average number of newly infected units/item/entities/ individuals" will be used interchangeably.

where

$$\mathcal{R}_{ij} = \begin{cases} \dfrac{\alpha_{ii} + \eta_i \vartheta}{b_i + \varphi_i + \lambda_i} & \text{if } i = j, \\ \dfrac{H_i \alpha_{ij}}{H_j (b_j + \varphi_j + \lambda_j)} & \text{otherwise,} \end{cases} \tag{8}$$

is the ijth entry in the matrix M_n, $i, j = 1, \ldots, n$.

The basic reproduction number functions as a threshold parameter in understanding the state of disease. It is proven in [18] that if $\mathcal{R}_0 < 1$, then the DFE is locally asymptotically stable, meaning that the disease does not spread through the whole network, but if $\mathcal{R}_0 > 1$, the DFE is unstable and the disease may break out.

The following theorem gives us an important view by which we have a more profound understanding of malware spread in a heterogeneous wireless network.

Theorem 2. *Assume that the average number of newly infected units $\mathcal{R}_0$ for a n-type system of models* (1)*,* (2) *and* (3)*,* (4)*,* (5) *is given by the spectral radius of the $n \times n$ matrix M_n defined in Theorem 1. If one more spillover species comes into the existing network and all the parameters in the equations remain unchanged, the new SIS and SIR models with $(n + 1)$ types of items have the basic reproduction number, i.e., the spectral radius of the $(n + 1) \times (n + 1)$ matrix M_{n+1}, $\rho(M_{n+1})$, that satisfies*

$$\rho(M_{n+1}) \geq \rho(M_n) \geq \mathcal{R}_{11}. \tag{9}$$

An expansion for the Theorem 2 is as follows:

Claim. Additionally, if the transmission rate for the kth species satisfies

$$\alpha_{k1}(H_1)\alpha_{1k}(H_k) \neq 0, \qquad k \neq 1, \tag{10}$$

then

$$\rho(M_n) > \mathcal{R}_{11}, \qquad \forall n > 1. \tag{11}$$

Proofs for these theorems and claim are provided in the appendices.

Theorem 2 gives us an important result: If a new population is introduced into an existing system and this new population has disease-related transmissions back and forth with populations that are already in the system, the possible average number of infected devices tends to increase. That means, the more types of individuals appear in the network, the higher the probability that an epidemic infection spreads wider. Although it is straightforward to see this, we have set a theoretically sound base by the above theorem. We will show more evidence in the simulation section where the numerical outcomes totally agree with what stated here.

5 Simulation Results

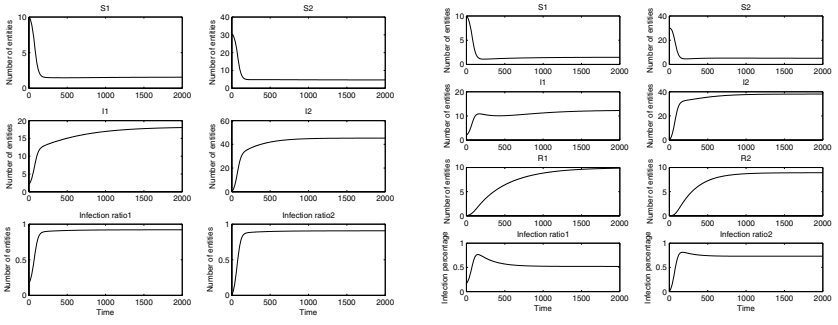
In this section, we evaluate several examples to study the dynamics of the SIS and SIR models. Examples had unlike combinations of parameters to enhance the theoretical findings. According to the first setting, we assumed that a malcode is newly emerged and it is able to massively spread. The same problem of serious spread has been found in

traditional fixed networks [15], [19] and we raise awareness about it in the wireless environment and tested by our models. To reflect the massive diffusion, high contact rates and low recovery rates were employed. In the first simulation, we tested the variations in a network with two device types: computers and phones with a specific operating system, say OS1. Let $b_1 = 0.3$ and $b_2 = 0.5$ for the birth rate. We used a simple form for the density-dependent death rates, $d_1(N_1) = 0.1 + 0.01N_1$ and $d_2(N_2) = 0.1 + 0.008N_2$. Attempts to transfer malware by WLAN or Bluetooth were modeled by $\alpha_{ij}, i, j = 1, 2$ as follows: $\alpha_{11} = 0.15N_1,\ \alpha_{12} = 0.05N_2,\ \alpha_{21} = 0.08N_1,$ and $\alpha_{22} = 0.05N_2$. The download rates for two species are $\beta_1 = 5$ and $\beta_2 = 2$; the probability that a program is malware is $\gamma = 0.005$. Since MMS is only valid for phones, we let $\eta_1 = 0$ for computers and $\eta_2 = 5$ for phones; and the probability a phone is successful in sending the compromised message is $\vartheta = 0.1$. We then describe the last infection mechanism through memory cards. Assuming that computer users and phone users use memory cards at the same rate, let $\mu_1 = \mu_2 = 3$. The probability a memory card contains malware is $\tau = 0.1$.

Antivirus softwares for computers are updated faster than for phones and computer users tend to scan their computers more frequently, so the recovery rates for computers and phones were set at $\varphi_1 = 0.2$ and $\varphi_2 = 0.1$, respectively. Nowadays, less malware, upon usurpation on victims, actually causes the sufferers to break down . Thus, we set the disease-related death rates $\lambda_1 = \lambda_2 = 0.1$.

The possible average number of infected items $\mathcal{R}_0$ for the SIS model therefore had a value of 9.7. As we can see in Fig. 1(a), the infected populations, I_1 and I_2, go up in the initial period while the susceptible populations sharply decrease because of the massive infection. They all stabilize at certain numbers during the simulation. This is in agreement with the value of $\mathcal{R}_0$, which indicates that the disease had obviously prevailed.

The second simulation gives us a view of the SIR model with two types of devices, the same as in the first simulation. All parameters in the first simulation were kept unchanged. Outcomes in Fig. 1(b) showed that all compartments soon reach their stable



(a) (b)

Fig. 1. Numerical results for massive spreads in 2 species of the: (a) SIS model with $S_1^0 = 10,\ S_2^0 = 30,\ I_1^0 = 2$, and $I_2^0 = 0$; (b) SIR model with $S_1^0 = 10,\ S_2^0 = 30,\ I_1^0 = 2, I_2^0 = 0, R_1^0 = 0$, and $R_2^0 = 0$

states. Similar to the SIS model, the infectious population in both computers and phones existed and persisted.

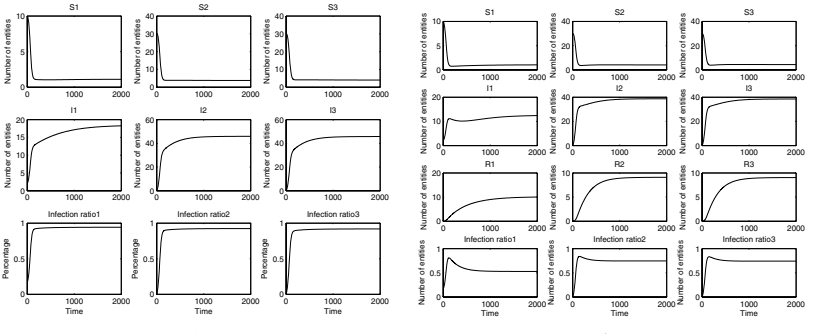
In the third and fourth numerical simulations, we tested the SIS and SIR having one more newly added type of phone with operating system OS2. All parameters in the system were kept unchanged and we used the following extra parameters in need caused by the addition: The density-dependent death rate for the third entity is $d_3 = 0.1 + 0.008N_3$. Attempt rates to transfer malware through wireless interfaces, given in matrix form, was:

$$\alpha = \begin{pmatrix} 0.15N_1 & 0.07N_2 & 0.05N_3 \\ 0.08N_1 & 0.08N_2 & 0.03N_3 \\ 0.08N_1 & 0.03N_2 & 0.07N_3 \end{pmatrix}.$$

Items of this type download with the rate $\beta_3 = 2$. We let the rate at which compromised phones of the new type send MMS messages be 5, i.e., $\eta_3 = 5$. Users using this type of phones were assumed to use memory cards at the rate $\mu_3 = 3$. This kind of phone had the recovery rate of 0.1, denote by $\varphi_3 = 0.1$. The disease-related death was set $\lambda_3 = 0.1$.

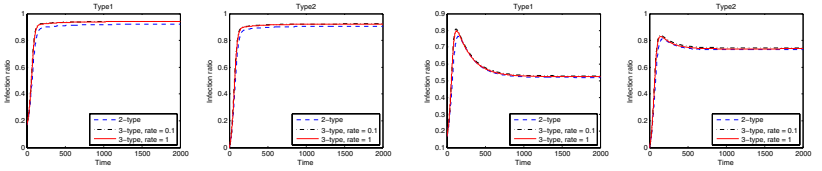
For the two latter simulations, the average number of newly infected items was 12, bigger than that for the system with two entity types. We can see the same variation patterns of populations from Fig. 2(a) and 2(b) as what had been observed in the former simulations. However, if we compare SIS with SIS (Fig. 1(a) vs. Fig. 2(a)) and SIR with SIR (Fig. 1(b) vs. Fig. 2(b)) results obtained from simulations for 2-type and 3-type networks, respectively, there are disparities. The portion of infected entities during time in both species, computers and phones with OS1, from Fig. 1(a) and Fig. 1(b) are smaller than that from Fig. 2(a) and Fig. 2(b), respectively paired as mentioned above.

We also conducted other numerical simulations to assess the influence of device diversity on viral dissemination. As the theoretical result has shown, the new system forming from the existing one with at least one more new device kind has rising average number of newly infected entities, no matter how big the recovery rate of the new network items. For both models having 3 entity types, we changed the recovery rate of



(a) (b)

Fig. 2. Numerical results for massive spreads in 3 species of the: (a) SIS model with $S_1^0 = 10,\ S_2^0 = 30,\ S_3^0 = 30,\ I_1^0 = 2, I_2^0 = 0$, and $I_3^0 = 0$; (b) SIR model with $S_1^0 = 10,\ S_2^0 = 30,\ S_3^0 = 30,\ I_1^0 = 2, I_2^0 = 0, I_3^0 = 0, R_1^0 = 0, R_2^0 = 0$, and $R_3^0 = 0$



(a) (b)

Fig. 3. Analysis of different recovery rates of the third species for the (a) SIS and (b) SIR model with massive spread

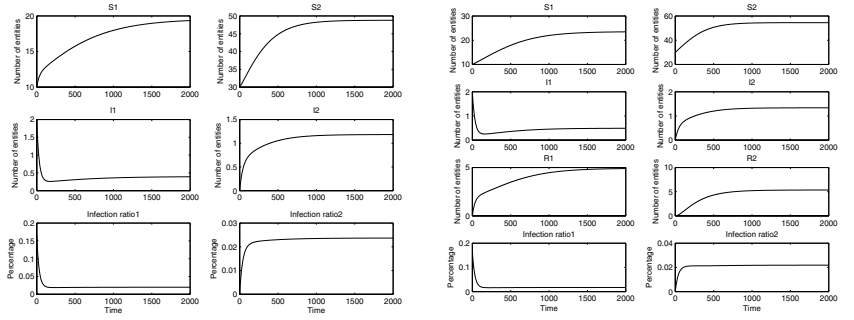
the third population from 0.1 to 1. The portion of infected items over the total number of items for computers and phones with OS1, as theoretically judged, are bigger than that for the network with 2 types of items. We report this difference in Figs. 3(a) and 3(b). The bottommost line shows that the portion of infected entities for two first device types in a 2-type network is smaller than that in a 3-type system. These numerical results strongly agree with the new value of the basic reproduction number that when a new entity kind is added into an existing system and there are transmissions between this new type and existing ones, the disease is likely to span to a wider extent.

In the second scenario, the malware had lower speed of spreading and users were more experienced in mitigating the viral contagion. Therefore, the wireless contact rates had lower values while the recovery rates had higher ones. All parameters are given in matrix or vector forms in Table 1. The outcomes are shown in Figs. 4(a) and 4(b). The resulted average number of newly infected items is calculated, $\mathcal{R}_0 = 0.32$. A small value of $\mathcal{R}_0$ infers that the disease has no opportunity to prevail.

Table 1. Parameters for the second scenario with a 2-type network

Parameters	Values
b	$[0.3, 0.5]$
d	$[0.1 + 0.01N_1, 0.1 + 0.008N_2]$
α	$\begin{pmatrix} 0.015N_1 & 0.005N_2 \\ 0.008N_1 & 0.005N_2 \end{pmatrix}$
β	$[5, 2]$
γ	0.005
η	$[0, 2]$
ϑ	0.1
μ	$[3, 3]$
τ	0.1
φ	$[1.5, 1]$
λ	$[0.01, 0.01]$

We analyzed the same problems when one more type of phone appears with the network and when this new type has variable recovery rates. Additional parameters are:



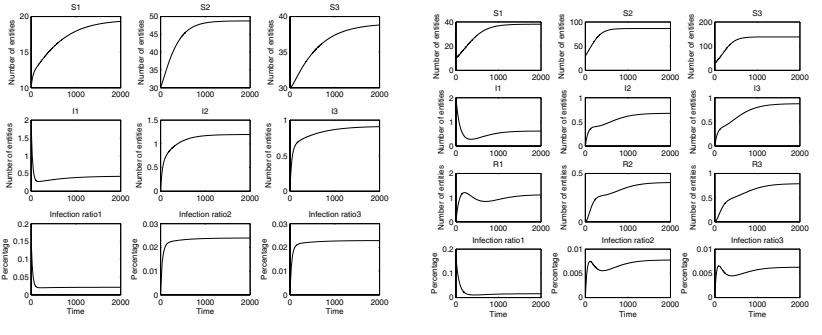
(a) (b)

Fig. 4. Numerical results for mitigated spreads in 2 species of the: (a) SIS model with $S_1^0 = 10,\ S_2^0 = 30,\ I_1^0 = 2$, and $I_2^0 = 0$; (b) SIR model with $S_1^0 = 10,\ S_2^0 = 30,\ I_1^0 = 2, I_2^0 = 0, R_1^0 = 0$, and $R_2^0 = 0$

$d_3 = 0.01 + 0.008N_3, \beta_3 = 2, \eta_3 = 5, \mu_3 = 3, \varphi_3 = 1$, and $\lambda_3 = 0.1$. Wireless contact rates in matrix form is

$$\alpha = \begin{pmatrix} 0.015N_1 & 0.005N_2 & 0.005N_3 \\ 0.008N_1 & 0.005N_2 & 0.003N_3 \\ 0.008N_1 & 0.003N_2 & 0.004N_3 \end{pmatrix}.$$

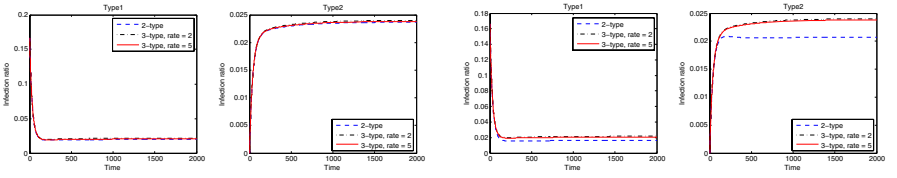
Outcomes for this parameter combination are illustrated in Figs. 5(a) and 5(b). In this case, the reproduction number is $\mathcal{R}_0 = 0.33$, just a little bit bigger than that of a 2-type system. An explanation is that all propagating mechanisms had small rates and probabilities so the number of infected individuals did not grow much.



(a) (b)

Fig. 5. Numerical results for mitigated spreads in 3 species of the: (a) SIS model with $S_1^0 = 10,\ S_2^0 = 30,\ S_3^0 = 30,\ I_1^0 = 2, I_2^0 = 0$, and $I_3^0 = 0$; (b) SIR model with $S_1^0 = 10,\ S_2^0 = 30,\ S_3^0 = 30,\ I_1^0 = 2, I_2^0 = 0, I_3^0 = 0, R_1^0 = 0, R_2^0 = 0$, and $R_3^0 = 0$

We changed the recovery rate of the third item kind to $\varphi_3 = 5$ to see the influence of item heterogeneity. In these conductions, the total portion of infected items for computers and phones with OS1 were higher than that of a 2-type network as easily seen in Figs. 6(a) and 6(b).



(a) (b)

Fig. 6. Analysis of different recovery rates of the third species for the (a) SIS model and (b) SIR model with mitigated spread

6 Conclusions and Future Works

In this paper, we have examined the situation that wireless devices having different characteristics exist in a network with threats of viral infection. Various sorts of items, e.g. computers and phones, are distinguished and taken into account to precisely formulate two novel analytical models which we propose as an aid to understanding the dynamics of the spread malware through an ad hoc network. Additionally, we have derived a method to calculate the basic reproduction number, understandable as the possible average number of newly infected items, and used this number to acquire a grasp of disease diffusion as well as its stability. We also have theoretically contributed to comprehending the influence of entity diversity on malware propagation: The more distinct populations, with vulnerabilities to malware infection, introduced into an existing network, the bigger the viral dissemination is. Moreover, a huge result space is also producible by our 3-compartment framework thus makes it appropriate to describe many viral proliferating scenarios. The correctness of our models is affirmed by several numerical results under various situations.

Our work has some restrictions as follows: The simulations were conducted based on numerical results only and we did not use real network traces, and our models can be more general if we consider more states that each network item may undergo while participating in the network.

We intend to further scrutinize the basic reproduction number so that we can usefully utilize it for an investigation of the spread of malware and add mobility into our model to make it better reflect the real world. Another extension of our paper can be made on considering multiple susceptible and infected reservoirs, this could happen in reality.

Acknowledgements

We gratefully appreciate the support from the 21st century COE program “Verifiable and Evolvable e-Society,” funded by the Japanese Ministry of Education, Culture,

Sports, Science and Technology. We thank Dr. Yasuhiro Ohara (JAIST) for his insightful comments to improve the quality of this paper.

References

1. Chen, T.M., Robert, J.M.: Worm epidemics in high-speed networks. Computer 37(6), 48–53 (2004)
2. Coursen, S.: The future of mobile malware. Network Security 2007(8), 7–11 (2007)
3. Hypponen, M.: Mobile malware. Invited talk, 16th USENIX Security symposium, http://www.usenix.org/events/sec07/tech/hypponen.pdf (accessed, October 2008)
4. Lancaster, P., Tismenetsky, M.: The theory of matrices: with applications, 2nd edn. Academic Press, London (1985)
5. Leavitt, N.: Mobile phones: the next frontiers for hackers? Computer 38(4), 20–23 (2005)
6. Mickens, J.W., Noble, B.D.: Modeling epidemic spreading in mobile environments. In: ACM Workshop on Wireless Security (2005)
7. Moore, D., Shannon, C., Claffy, K.: Code-red: A case study on the spread and victims of an Internet worm. In: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement (2002)
8. Murray, J.D.: Mathematical biology: I. An introduction. Springer, Heidelberg (2002)
9. Nekovee, M.: Worm epidemics in wireless ad hoc networks. New Journal of Physics 9 (2007)
10. BBC news service. Mobiles get anti-virus protection, http://news.bbc.co.uk/2/hi/technology/4207476.stm (accessed, Octorber 2008)
11. Nguyen, H.-N., Shinoda, Y.: A novel analytical framework to model malware diffusion in heterogeneous wireless networks. In: Proceedings of the 10th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (2009)
12. Ortega, J.M.: Matrix theory: A second course. Plenum Press, New York (1987)
13. Ramachandran, K., Sikdar, B.: Modeling malware propagation in networks of smart cell phones with spatial dynamics. In: Proceedings of IEEE INFOCOM 2007 (2007)
14. Ramachandran, K., Sikdar, B.: On the stability of the malware free equilibrium in cell phones networks with spatial dynamics. In: IEEE International Conference on Communications, ICC 2007 (2007)
15. Su, J., Chan, K.K., Miklas, A.G., Po, K., Akhavan, A., Saroiu, S., Lara, E.D., Goel, A.: A preliminary investigation of worm infections in a bluetooth environment. In: Proceedings of the ACM Workshop on Rapid Malcode, WORM (2006)
16. Traynor, P., Enck, W., McDaniel, P., LaPorta, T.: Mitigating attacks on open functionality in SMS-capable cellular networks. In: Proceedings of MobiCom 2006 (2006)
17. Traynor, P., McDaniel, P., LaPorta, T.: On attack causality in Internet-connected cellular networks. In: Proceddings of 16th USENIX Security Symposium (2007)
18. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences 180, 29–48 (2002)
19. Wierman, J.C., Marchette, D.J.: Modeling computer virus prevalence with a susceptible-infected-susceptible model with reintroduction. Computational Statistics and Data Analysis 45(1), 3–23 (2004)
20. Zou, C.C., Gong, W., Towsley, D.: Code red worm propagation modeling and analysis. In: Proceedings of the 9th ACM Conference on Computer and Communications Security (2002)

A Proof for Theorem 1

Result from [18] can be directly applied to get a proof for this Theorem.

B Proof for Theorem 2

We augment the matrix M_n with one row and one column of zeros, the resulted matrix M_n^0 has the following form:

$$M_n^0 = \begin{pmatrix} M_n & \mathbf{0}^T \\ \mathbf{0} & 0 \end{pmatrix},$$

where $\mathbf{0}$ is the $1 \times n$ vector. The next generation matrix for the new system of $(n+1)$ species is built as follows:

$$M_{n+1} = \begin{pmatrix} M_n & M_{*,n+1} \\ M_{n+1,*} & \mathcal{R}_{n+1,n+1} \end{pmatrix},$$

where $M_{*,n+1}$ and $M_{n+1,*}$ are the column and row vector with corresponding elements computable from Theorem 1, respectively. The spectral radius of M_{n+1}, $\rho(M_{n+1})$, is the basic reproduction number of the new system consisting of $(n+1)$ types of entities. With the assumptions made for the SIS and SIR models, it follows that the column and row vectors $M_{*,n+1}$ and $M_{n+1,*}$ of M_{n+1} have nonnegative entries. Consequently, from the theory of nonnegative matrices [4], [12], we have

$$\rho(M_n) = \rho(M_n^0) \leq \rho(M_{n+1}).$$

When $n = 1$, $\rho(M_1) = \mathcal{R}_{11}$. Hence, the inequality (9) holds.

C Proof for Claim

In order to prove the strict inequality (11), we see that, according to the assumption (10), $\mathcal{R}_{1k} > 0$ and $\mathcal{R}_{k1} > 0$ since they can not be equal to zero. Let $k = 2$ and consider the basic reproduction number of a two-species system, represented by the spectral radius of M_2, $\rho(M_2)$, where

$$M_2 = \begin{pmatrix} \mathcal{R}_{11} & \mathcal{R}_{12} \\ \mathcal{R}_{21} & \mathcal{R}_{22} \end{pmatrix}.$$

It is easy to see that $\rho(M_n) \geq \rho(M_2) > \mathcal{R}_{11}, \forall n > 2$. The inequality (11) holds.