# Epidemic Model for Malware Targeting Telephony Networks

Iosif Androulidakis*, Sergio Huerta†, Vasileios Vlachos‡ and Igor Santos†

*University of Ioannina
Ioannina, Greece

†DeustoTech
University of Deusto
Bilbao, Spain

‡Department of Computer Science and Engineering
Technological Educational Institute (TEI) of Thessaly
Larissa, Greece

Email: *sandro@noc.uoi.gr, †shuerta@deusto.es, ‡vsvlachos@teilar.gr, †isantos@deusto.es

***Abstract*—In this paper, we present a new mathematical epidemiological model of a malware targeting Private Branch eXchanges (PBX). Although the term PBX is old, we argue that traditional PBXs are still extensively used, along the modern VoIP systems, forming part of most nations' critical infrastructure. The proposed model is based on graph theory and generic epidemiological models. Through this model we are able to simulate infections of PBX malware and monitor their evolution. We provide estimations of potential scale and timing characteristics of virus propagation over the PBX network with a mathematical framework to model PBX malware, an epidemic model for PBX infections and an empirical study of the model under a simplified environment. Our results show that 2 days are sufficient for the malware diffusion, highlighting the necessity of immediate countermeasures.**

## I. Introduction

The idea of a special-type malware spreading among PBXs, threatening confidentiality, integrity and availability of these systems has been published in [1]. The same work proposes that PBXs should be considered as part of a nation's critical infrastructure, along with the public telephony network. As far as the authors know, there is no relevant work in the literature, dealing with the modeling of a PBX malware. PBX security itself, has quite a limited academic coverage, with some more references in business and technical journals and publications [2], [3], [4], [5], [6], despite the fact that terrorists have been known to abuse such systems [7], [8].

Although the Private Brand eXchange (PBX) as a term usually refers to the older TDM technology versus the nowadays ever-present VoIP, this work is important for a number of reasons: Traditional PBXs have been exceptionally rigid and enjoy life spans of more than 20 years. As such, infrastructure that was commissioned in early 2000, before the major VoIP "revolution", will stay in operation for (at least another) 10 years from now, reinforcing the need for specific literature. In addition to that, lots of installations were based on hybrid solutions before migrating to full VoIP. Manufacturers in any case, kept large segments of code and functionality from older tested and proven PBX platforms. Therefore older problems are migrated to the new platforms. At the end of the day, given the global crisis, older installations will remain in service for quite long since they have proven their value and rigidness [9].

The adoption of epidemiology to develop models for malware spread was firstly pointed out by Cohen and Murray in 1988 [10] based on the similarities depicted by the self-replicating and the propagation behavior of computer viruses and biological pathogens. Since then, numerous works have extended this approach. Traditionally each host exists only in two discrete states, infected or not infected, but different epidemiological characteristics have been included to the models, as the underlying topology for propagation [11]. Specific attacks have been studied and modeled as Code Red worm [12] as well as generic models for malware [13]. Custom models have been created for specific software applications as the e-mail [14] or on-line social networks [15]. Also, particular types of malware have been created and released in the wild, like cross site scripting worms [16]. Even malware for specific platforms has been developed, as for mobile platforms [17].

Given this background, we present a new mathematical epidemiological model of PBX malware which is able to simulate infections of PBX malware and monitor their evolution.

Summarizing, the goal of the current paper is to provide estimations of potential scale and timing characteristics of

This work has been partially funded by the Ministry of Economy of Spain under the grant TIN2011-29709-C02-02 for the project "MARSOL. Modelos de propagación de malware a través de redes sociales on-line"

virus propagation over the PBX network [1] and therefore the main contributions of this paper are the following:

- A mathematical framework to model PBX malware.
- An epidemic model for PBX infections.
- An empirical study of the model under a simplified environment.

The remainder of this paper is organized as follows. Section II describes the mathematical model and presents the epidemiological model created. Section III details the experiments conducted and describes the obtained results. Section IV discusses the main outcomes of this paper. Finally, Section V concludes this paper and presents possible further work.

## II. Methodology

The proposed model relies on epidemiological theory. A set of differential equations describes the evolution of the diffusion process of the malware.

### A. Epidemic Theory

We describe the diffusion process with a simplified version of the deterministic SEIR model over an age-structured population [18], [19], [20].

Epidemiology uses two different groups of models, deterministic [20] and stochastic [21]. Although real-world epidemics are stochastic processes, deterministic models are preferred when working with large population sizes. As PBX population is up to the range of millions , we are employing deterministic models.

Most popular epidemic models are compartmental, i.e., population is classified (in compartments) according to the present disease phase. Standard compartments are: *Susceptible*, for individuals that might became infected; *Exposed*, for those incubating the disease; *Infectious*, for those who are able to transmit the disease and *Removed*, for those without active role in the diffusion process, after experiencing the disease and have acquired immunity or have been deceased.

Susceptible-Infectious (SI) model is the simplest epidemic model. At the beginning, all the individuals are susceptible and just a few of them are infectious. While susceptible individuals are contacted by infectious ones, they move into the infectious compartment. In this simple model, once an individual became infectious, he remains in this compartment indefinitely. The two most classical models are obtained by considering the consequences to infectious individuals. On the one hand, they can recover and come back to the susceptible state after a period of infection, which leads to the Susceptible-Infectious-Susceptible (SIS) model. On the other hand, if we accept that the infectious individuals die or gain immunity to the disease, they are considered to move into the removed compartment, and play no further role in the diffusion process. This is also called the Susceptible-Infectious-Removed (SIR) model.

These models are governed by differential equation systems. Denoting by $S(t)$ (and resp. $E(t)$, $I(t)$ and $R(t)$) the number of susceptible (and resp. exposed, infectious and removed) individuals at time $t \geq 0$, the equations for the SEIR model are:

$$\begin{aligned} S'(t) &= -\alpha \frac{S(t)I(t)}{N}, \\ E'(t) &= \alpha \frac{S(t)I(t)}{N} - \beta E(t), \\ I'(t) &= \beta E(t) - \gamma I(t), \\ R'(t) &= \gamma I(t); \end{aligned}$$

with the normalizing equation $N = S(t)+E(t)+I(t)+R(t)$, where $N$ is the total size of the population, $\beta$ and $\gamma$ are such that $\beta^{-1}$ and $\gamma^{-1}$ are the average time an individual remains in the exposed and infectious compartments respectively. $\alpha$ is the rate at which contacts between infected and susceptible individuals take place.

Formally, let $J$ be a set of indexes identifying the groups, and denote by $S_i(t)$ (and resp. $E_i(t)$, $I_i(t)$ and $R_i(t)$) the number of susceptible (and exposed, infectious and removed) individuals of age $i \in J$ at time $t \geq 0$; the equations for the SEIR model over an age-structured population are: for each $i \in J$,

$$\begin{aligned} S_i'(t) &= -\frac{S_i(t)}{N} \sum_{j \in J} \alpha_{ij} I_j(t), \\ E_i'(t) &= \frac{S_i(t)}{N} \sum_{j \in J} \alpha_{ij} I_j(t) - \beta E_i(t), \\ I_i'(t) &= \beta E_i(t) - \gamma I_i(t), \\ R_i'(t) &= \gamma I_i(t); \end{aligned} \tag{1}$$

where $\alpha_{ij}$ is the rate at which infectious individuals of age $j$ make contact with susceptible ones of age $i$. As above, there are normalizing equations, for each $i \in J$, the number of individuals in this group is constant, $N_i = S_i(t) + E_i(t) + I_i(t) + R_i(t)$; and the total population, $N = \sum_{i \in J} N_j$, is constant too.

The equations are the same, except when considering a contact between an infectious and a susceptible individual, the rate of infection is specific to the ordered pair of groups that the individuals belong to. The SEIR model can be recovered without groups by considering $\alpha_{ij} = \alpha$ for all possible indexes, and; $S(t) = \sum_{i \in J} S_i(t)$, $E(t) = \sum_{i \in J} E_i(t)$, $I(t) = \sum_{i \in J} I_i(t)$ and $R(t) = \sum_{i \in J} R_i(t)$.

### B. Malware Modeling

The algorithmic blocks of the malware were described in the previous paper [1]; here, we restate them focusing on the diffusion process.

The algorithmic blocks of the malware are presented as white boxes on figure 1. Any PBX is correctly working until an intruder manages to upload the malware into it. This can be the result of an intrusion or, an insider's act. There are two main ways of spreading. Either by direct infection of inter-networked PBXs, or more universal, by war-dialing. War dialing is the technique of consecutive calling telephone numbers in order to discover modems and electronic services to abuse. The term comes from the 1983 classic film "War

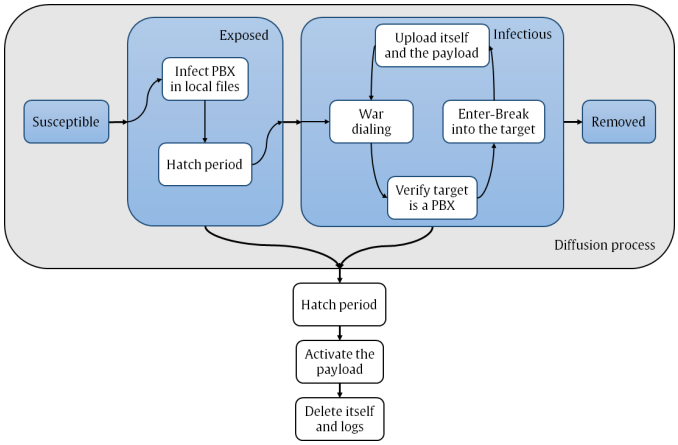


Fig. 1. Flow chart describing the states that a PBX host follows (white boxes), the compartments of the epidemiological model (blue boxes) and the relation between them.

Games" that actually portrayed the technique. According to it, the attacker dials as many as possible numbers in a given range, trying to find modem carriers or other tones that denote the presence of a computer/PBX. Even though modems are no longer in service, most PBXs still utilize them for the remote maintenance procedures.

The malware begins searching in local files (depending on the operating system of the host) for the presence of other PBXs and infects them if possible. Indeed, many larger companies, organizations and institutions employ networks of PBXs with dozens or even hundreds of nodes. After that, there is a hatch period before starting the war dialing procedure in order to maximize the spread. When the war dialing procedure discovers the modem of another PBX (or any other vulnerable service) there is an attempt to infect it. The war dialing procedure continues until the administrator of the PBX discovers and removes the malware; or until a trigger establishes the end of the diffusion process, maybe by a command and control center, and the PBX goes into another hatch period before activating the payload. Finally the malware deletes itself and the traces of its activity.

### C. Epidemic Modelling

In order to include two different diffusion mechanism to the model, the individuals of the epidemic process are not PBXs, but sets of networked PBXs. A networked PBX is usually a PBX, but if one PBX can be discovered in the local files of another one, the two of them are considered as part of one networked PBX.

There are two methods for identifying new targets: The malware may discover new targets searching in the local files or by war dialing. The targets collected by monitoring communication protocols specified in [1] can be directly neglected and therefore we ignore them for the sake of simplicity.

Although finding a new target in the local files is fast, it is quite rare to find a PBX by searching in the local files since this happens only in larger installations, as discussed. Most PBXs are stand-alone systems catering the communication needs of small to medium size entities. Finding a new target in the local files mainly occurs when both PBXs belong to the same company. Infection by local files cannot be considered the primary mechanism of the diffusion process because it is nearly impossible to reach any target using this method by its own. We shall say that 2 vulnerable PBXs are connected if one of them can be found in the local files of the other or if one of them can be found in the local files of a PBX connected with the other. The time needed for the discovery of a new target by the war dialing procedure is so large in comparison with the time needed in order to infect a PBX and the time needed for the discovery of a new target stored in the local files of a infected PBX, that we consider this two last activities to be instant processes. Since connected PBXs would usually be of the same brand and type, the intrusion can exploit the same vulnerability. In any case, the malware has a "library" of techniques and vulnerabilities to exploit, specifically for each PBX brand. So, if a PBX is infected at time $t$ all of the vulnerable PBXs connected with this one are instantly infected. Furthermore, we are going to consider that the PBX connected by local files are all of them vulnerable or none of them because the same configuration is expected in both PBXs. From an epidemiological point of view, a group of connected PBXs has to be considered as a single individual. Therefore, in our model, individuals are going to be either PBX hosts if none other can be found in the local files, or a set of connected PBXs.

The main infection mechanism is the war dialing procedure. All the targets are accessible from any infectious PBX for the war dialing procedure via the public telephony network, therefore we accept the homogeneous mixing assumption.

The relation between the algorithmic blocks of the malware and the compartments of the epidemiological model is the following. Before getting infected a host is working normally, these PBXs hosts are considered the susceptible individuals of the infection. As soon as a host becomes infected, the malware searches for other PBX in local files, infect them if possible, (we are considering this action to be instant) and goes into a hatch period. Hosts in these steps are considered in the exposed compartment. When the hatch period ends, the PBX starts with the war dialing procedure trying to infect a target once it is discovered. PBXs in this loop are considered in the infectious compartment. There are two options to exit this loop; if the malware generates alerts (most probably due to the high utilization of outgoing lines that take part in the war-dialing process), a technician repairs the PBX cleaning the malware, and epidemiologically, we consider this PBX moved into the removed compartment playing no further role in the diffusion process. We consider that the technician is now aware of infections in the PBX host, and does not let the malware reinfecting this host again. This is the reason for working with a variation of the SIR model. If the malware does not generate alerts the war dialing procedure continues until a trigger sets all the infected PBX (exposed or infectious) to a second hatch period. This will end the diffusion process. Subsequently to

TABLE I
NOTATION

| Notation | |
|---|---|
| $N$ | Total population of vulnerable PBX hosts |
| $N_i$ | Vulnerable PBX with $i$ available lines |
| $S_i(t)$ | Susceptible PBXs of group $i$ at time $t$ |
| $E_i(t)$ | Exposed PBXs of group $i$ at time $t$ |
| $I_i(t)$ | Infectious PBXs of group $i$ at time $t$ |
| $R_i(t)$ | Removed PBXs of group $i$ at time $t$ |
| $\alpha_{ij}$ | Rate at which a infectious PBX of group $j$ infects a PBX of group $i$ |
| $\alpha_i$ | Rate at which a susceptible of PBX of group $i$ gets infected |
| $\alpha$ | Rate at which a susceptible of PBX gets infected by only one war dialing line |
| $\beta$ | Inverse average time of the hatch period |
| $\gamma$ | Inverse average time to be cleaned by a technician |

this hatch period the payload is activated and the malware deletes both itself and the logs. Figure 1 shows a schematic representation of the transition between the different steps of the malware and the compartments of the diffusion process.

Not vulnerable PBXs are not considered in the total population of the model. The total population, $N$ is the number of vulnerable PBX hosts by this malware. In any case, this number is maximized by the multitude of vulnerabilities and exploits the malware can take advantage of.

The other main difference between vulnerable individuals is the number of available lines for war dialing. This is the reason for choosing an epidemic model with heterogeneous individuals. The number of available lines is inversely proportional to the time needed for an infectious PBX to discover a new target. Even in cases where there is a single modem available, the total number of out-dialing lines is still of importance. They can be used to perform an initial screening of non-answering or non-existing numbers, therefore limiting the total number of actual telephone numbers that have to further be called by the (single) modem. We divide the population of PBX into groups according to the number of available lines for war dialing, which is the reason for using an age-structured model.

Differences in the time needed to be infected are not relevant. In general, these differences are short (comparing with the time needed for the war dialing procedure).

We need the distribution of the number of vulnerable hosts by the number of available lines for war dialing. This distribution might reflect the local files infection, i. e., some PBX with few war dialing lines will be substituted by another one with more available lines.

At this point, the diffusion process corresponds with the classical SEIR model over an age-structured population (1).

Table II-C summarizes the notation used. $S_i(t)$ (and resp. $E_i(t)$, $I_i(t)$ and $R_i(t)$) are functions of time and we refer to their derivate with respect to $t$.

We can determinate the parameters $\alpha_{ij} = j\alpha_i$, because the rate at which targets are discovered is proportional to the number of available lines to war dialing. We are not considering differences in the time needed to be infected; then, $\alpha_i$ does not depend on $i$, so $\alpha_i = \alpha$.

Note that the distribution of the number of available lines for war dialing is not in the model itself, it is just part of the initial condition to solve the system, $S_i(0) \simeq N_i$.

Finally, the governing equations of the malware diffusion model are the following:

$$\begin{aligned} S_i'(t) &= -\frac{\alpha S_i(t)}{N} \sum_{j\in J} jI_j(t) \\ E_i'(t) &= \frac{\alpha S_i(t)}{N} \sum_{j\in J} jI_j(t) - \beta E_i(t) \\ I_i'(t) &= \beta E_i(t) - \gamma I_i(t) \\ R_i'(t) &= \gamma I_i(t) \end{aligned} \tag{2}$$

with initial conditions $S_i(0) = N_i$, $E_i(0) = 0$, $I_i(0) = 0$ and $R_i(0) = 0$ unless for one $j \in J$ (the initial infected host) that we have $S_j(0) = N_j - 1$, $E_j(0) = 1$, $I_j(0) = 0$ and $R_j(0) = 0$.

## III. Approximate Evolution of the Epidemic

In order to numerically solve the differential equation system we need to approximate the parameters. For the distribution of the PBXs, $N_i$, we will use a power distribution (which is in accordance with the groups we generate in order to reflect the local connectivities) with parameter 2.5. We let from 1 to 500 outlines, having $N_i = \lceil 500^{2.5} i^{-2.5} \rceil$ PBX's with $i$ outlines for $i = 1, \ldots, 500$. This leads to a total population of 7.5 millions.

Taking into consideration the percentage of PBX in the whole telephone numbers, the percentage of PBX vulnerable between all the PBX and the average time needed to infect a host once it has been discovered, we establish 1000 seconds, $\alpha^{-1} = 1000$ for the average time to find and infect a new target war dialing with only one line. 1000 seconds to find a new modem with a single line is plausible when is not used with brute-force dialing, but rather by dialing specific extensions where modems usually "reside" (e.g. 999, 111, 000 etc.).

We set 1 day, $\beta^{-1} = 86400$ seconds, for the average time exposed or duration of the first hatch period (the one before the war dialing procedure). We establish one day assuming that the malware keeps hatched until night and employees are not using the telephone lines.

In order to estimate the time running the malware before it is detected and removed, we assume a routine scan on the hosts once every month. So as an average a malware is detected every 15 days, $\gamma^{-1} = 1296000$ seconds. Some will not have any scan, and some other will have more frequents scans. This matches with the exponential decay of the infectious hosts.

Numerically solving 2 we have the figure 2.

We have to differentiate between the diffusion process and the malicious activity of the malware. As soon as the PBX is exposed, the malware is running on the machine, triggering the diffusion, but the malicious activity will presumably be executed after 36 hours at most. To get to know the number of PBXs being affected by the malware, the sum of exposed and infected is the quantity to read. It can be easily read as the decay of the susceptible hosts because removed hosts are nearly 0%.

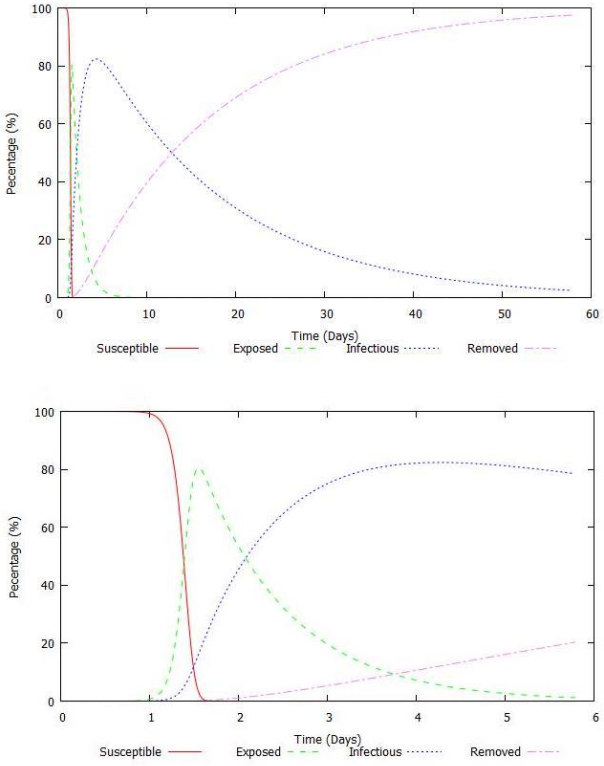


Fig. 2. Top: Evolution of the epidemic. Bottom: Detail of the first six days.

We remark that the whole vulnerable population leaves the susceptible compartment in the first 36 hours. The change occurs between the hour 24 and the hour 36. At that moments we only need 1-5% of the PBX population war dialing. The most worrying about this fact is that the expected overcharge on the telephony network that turns the malware noticeable will not occur. As with most of the malware infections, early detection and education are the main counter-mechanisms.

## IV. Discussion

Although we were familiar with the dangers of a malware targeting the PBX infrastructure, we were not aware of the expected the devastating effects related with this kind of malware. The epidemiological model provided in this paper sheds light on this specific malware attack. Therefore we can better understand the size and the effects of this threats and plan the possible countermeasures. The epidemiological model developed provides a threshold (the ratio between infection and recovery ones), which can help us conclude that the malware would infect within a short timeframe all vulnerable hosts if no countermeasures are put in place.

The difference of the infection ratio and the recovery provide clear conclusions. The reaction time is short, only 24 hours.

## V. Conclusion

PBXs practically serve all vital societal sectors, forming a part of the critical infrastructure, along with the public telephony network. As such, a malware targeting PBXs would have a devastating effect on communications confidentiality, integrity and availability, especially during wartime. Previous paper demonstrated the danger of a malware targeting PBX infrastructure. This work has quantified the effects of this type of malware on PBX's. As main conclusion we have shown that PBX malware needs to be counteracted during the first 24 hours. Early detection tools are fundamental. If more time is required, then increased frequency of routine checks on the hosts is necessary. Further work will be focused on the empirical study of the evolution of PBX malware, rather than the theoretical approximation.

## References

[1] I. I. Androulidakis, V. Vlachos, and Y. Kamphuis, "On a malware targeting private telephony networks during cyber conflict," *Telfor Journal*, vol. 5, no. 1, pp. 2–7, 2013.

[2] D. West, "De-mystifying telecom fraud," *Telecom Business, July*, 2000.

[3] V. Blake, "Pabx security, information security technical report," *vol*, vol. 5, pp. 34–42, 2000.

[4] K. Archer, G. White *et al.*, "Voice and data security," 2001.

[5] C. Pollard, "Telecom fraud: the cost of doing nothing just went up, white paper," *Insight Consulting, Feb*, 2005.

[6] I. Androulidakis, "Pretty (private telephony security)-securing the private telephony infrastructure," *Inf Secur Int J*, vol. 28, no. 1, 2012.

[7] C. Communications Fraud Control Association, "Worldwide telecom fraud survey," 2009.

[8] A. Technica, "How filipino phreakers turned pbx systems into cash machines for terrorists," http://arstechnica.com/tech-policy/news/2011/11/how-filipino-phreakers-turned-pbx-systems-into-cash-machines-for-terrorists.ars, 2011.

[9] I. I. Androulidakis, *PBX Security and Forensics: A Practical Approach*. Springer Publishing Company, Incorporated, 2012.

[10] W. H. Murray, "The application of epidemiology to computer viruses," *Computers & Security*, vol. 7, no. 2, pp. 139–145, 1988.

[11] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on*. IEEE, 1991, pp. 343–359.

[12] C. C. Zou, W. Gong, and D. Towsley, "Code red worm propagation modeling and analysis," in *Proceedings of the 9th ACM conference on Computer and communications security*. ACM, 2002, pp. 138–147.

[13] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.

[14] C. Zou, D. Towsley, and W. Gong, "Modeling and simulation study of the propagation and defense of internet e-mail worms," *Dependable and Secure Computing, IEEE Transactions on*, vol. 4, no. 2, pp. 105–118, April 2007.

[15] G. Yan, G. Chen, S. Eidenbenz, and N. Li, "Malware propagation in online social networks: Nature, dynamics, and defense implications," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '11. New York, NY, USA: ACM, 2011, pp. 196–206. [Online]. Available: http://doi.acm.org/10.1145/1966913.1966939

[16] M. Faghani and H. Saidi, "Social networks' xss worms," in *Computational Science and Engineering, 2009. CSE '09. International Conference on*, vol. 4, Aug 2009, pp. 1137–1141.

[17] S.-M. Cheng, W.-C. Ao, P.-Y. Chen, and K.-C. Chen, "On modeling malware propagation in generalized social networks," *Communications Letters, IEEE*, vol. 15, no. 1, pp. 25–27, January 2011.

[18] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM review*, vol. 42, no. 4, pp. 599–653, 2000.

[19] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.

[20] R. M. Anderson and R. M. May, *Infectious diseases of humans*. Oxford university press Oxford, 1991, vol. 1.

[21] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*. Springer New York, 2000, vol. 151.