Research Article

Towards the Epidemiological Modeling of Computer Viruses

Xiaofan Yang\textsuperscript{1,2} and Lu-Xing Yang\textsuperscript{2}

\textsuperscript{1} School of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
\textsuperscript{2} College of Computer Science, Chongqing University, Chongqing 400044, China

Correspondence should be addressed to Xiaofan Yang, xfyang1964@gmail.com

Received 25 July 2012; Accepted 20 August 2012

Academic Editor: Yanbing Liu

Copyright © 2012 X. Yang and L.-X. Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Epidemic dynamics of computer viruses is an emerging discipline aiming to understand the way that computer viruses spread on networks. This paper is intended to establish a series of rational epidemic models of computer viruses. First, a close inspection of some common characteristics shared by all typical computer viruses clearly reveals the flaws of previous models. Then, a generic epidemic model of viruses, which is named as the SLBS model, is proposed. Finally, diverse generalizations of the SLBS model are suggested. We believe this work opens a door to the full understanding of how computer viruses prevail on the Internet.

1. Introduction

As a technical term coined by Cohen, a computer virus is a malicious program that can replicate itself and spread from computer to computer. Once breaking out, a virus can perform devastating operations such as modifying data, deleting data, deleting files, encrypting files, and formatting disks \cite{1}. In the past, massive outbreaks of computer viruses have brought about huge financial losses. With the advent of the era of cloud computing and the Internet of Things, the threat from viruses would become increasingly serious, even leading to a havoc \cite{2}. As we all know, antivirus software is the major means of defending against viruses. With the continual emergence of new variants of existing viruses as well as new types of virus strains, the struggle waged by human being against viruses is doomed to be endless, arduous, and devious; indeed, the development of new types of antivirus software always lags behind the emergence of new types of viruses. As thus, antivirus technique cannot predict the evolution trend of viruses and, hence, cannot provide global suggestions for their prevention and control.
Inspired by the intriguing analogies between computer viruses and their biological counterparts, Cohen [3] and Murray [4] inventively suggested that the techniques developed in the epidemic dynamics of infectious diseases should be exploited to study the spread of computer viruses. Later, Kephart and White [5] borrowed a biological epidemic model (the SIS model) to investigate the way that computer viruses spread on the Internet. The researches in this field have since been made mainly in the following two different directions.

(i) The finding that the autonomous system level topological structure of the Internet follows diverse power law distributions [6–8] has stimulated the interest in the spreading behavior of viruses on complex networks. Previous work in this direction focused on the existence and estimation of the epidemic threshold under the SI model [9, 10], the SIS model [11–21], and the SIR model [19, 21–24], leading to the most surprising finding that the epidemic threshold vanishes for scale-free networks with infinite size [11]. Due to the extreme diversity of topologies of large-sized complex networks, the global stability of the endemic equilibrium, if present, was examined experimentally rather than theoretically. Although Pastor-Satorras and Vespignani [11] indicated the necessity of studying other types of epidemic models on complex networks, to our knowledge no relevant work has been reported in the literature.

(ii) The strong desire to understand the spread mechanism of computer viruses has motivated the proposal of a variety of epidemic models that are based on fully connected networks, that is, networks where each computer is equally likely to be accessed by any other computer. Previous work in this direction was focused mainly on the theoretical study of complex dynamical properties of the models, such as the global stability of equilibria, the emergence of periodic solutions, and the occurrence of chaotic phenomena [25–34].

The epidemic dynamics of computer viruses is still in its infancy. While previous models lay emphasis on the similarity between computer viruses and infectious diseases, the majority of them more or less neglect the intrinsic difference between them.

This paper is intended to present a series of rational epidemic models of computer viruses. A close inspection of the characteristics of computer viruses reveals the flaws of previous models. On this basis, a generic epidemic model of viruses, which is known as the SLBS model, is proposed. By taking into account the impact of various factors, such as the impulsive emergence of new viruses, the impulsive succeed in the development of new antivirus software, and the fluctuation of the system parameters, a variety of generalizations of the SLBS model are suggested. We believe the proposed models open a door to the macroscopic understanding of the spread of computer viruses on the Internet.

The subsequent materials are organized this way: Section 2 elucidates the defects of previous models. Sections 3 and 4 formulate the SLBS model and some of its generalizations, respectively. Finally, This work is summarized in Section 5.

2. Flaws of Previous Models

2.1. Basic Terminologies

For convenience, let us introduce the following terminologies.

A computer is referred to as internal or external depending on whether it is connected to the Internet or not.

A computer is referred to as infected or uninfected depending on whether there is a virus staying in it or not.
A computer is referred to as the *host computer* of a virus if the virus has entered it and is staying in it. By the *life cycle* of a virus we mean the interval from the time it enters its host computer to the time it is eradicated. By the *lifetime* of a virus, we mean the length of its life cycle. The lifetime of a virus is not fixed. Rather, it is affected by a multiplicity of factors.

### 2.2. Principle of Computer Viruses

The ultimate goal of a clever computer virus is to devastate as many computer systems as possible. To realize that goal, the virus would try to stealthily infect as many computers as possible before it finally breaks out. As thus, a typical virus would undergo two consecutive phases: the *latent period*, that is, the interval from the time the virus enters its host computer to the time exactly before it inflicts damage on the host system, and the *breaking-out period*, that is, the interval from the time the virus begins to inflict damage to the time it is wiped out. In this paper, we will always assume that, in its life cycle, a virus has both latent and breaking-out periods. Furthermore, an infected computer will be referred to as *latent* or *breaking-out* depending on whether all viruses staying in it are in their respective latent periods or at least one virus staying in it is in its breaking-out period.

### 2.3. A Common Flaw of Models with E Compartment

For some biological infectious diseases, an infected individual may experience a particular period, named as the *exposed period*, before having infectivity [35]. So, the corresponding epidemic models must have a separate E compartment, that is, the compartment of all exposed individuals. Some previous epidemic models of computer viruses were established by borrowing biological epidemic models with E compartment, implying the prior assumption that some infected computers possess no infectivity [25, 29–31, 36–39].

The most striking characteristic shared by all computer viruses is their infectivity. On one hand, once infected with a narrowly defined virus, a computer possesses infectivity immediately, because it can infect other computers through sending emails with infected attachments or transmitting infected files. On the other hand, once infected with a worm, a computer also possesses infectivity immediately, because it can infect those computers with specific system vulnerabilities. Therefore, in the real world there exists no infected computer at all that has no infectivity. Equivalently, there exists no exposed computer, implying that a rational epidemic model of computer viruses should have no E compartment.

### 2.4. A Common Flaw of Models with All Infected Computers in a Single I Compartment

Most previous epidemic models of computer viruses have all infected computers in a single I compartment, that is, neither of these models makes a further classification of the infected computers [9–28, 32–34, 40–42].

On one hand, the cure rate of an infected computer, that is, the probability with which it is cured, is a major concern in the modeling process. Indeed, a breaking-out computer can get treated with a higher probability, because it usually suffers from a marked performance degradation or even breaks down, which can be perceived evidently by the user. In contrast, a latent computer can get treatment only with a much lower probability, because it usually
can work normally and hence the user cannot become aware of the presence of any virus at all. In the context of epidemiological modeling, therefore, there is a clear distinction between latent computers and breaking-out computers.

On the other hand, as opposed to a latent internal computer, a breaking-out internal computer has a higher probability to be disconnected from the Internet, because the possible system breakdown caused by the virus outbreak would yield the disconnection automatically.

In conclusion, a sound epidemic model of computer viruses should possess a compartment of all latent computers \( L \) and a compartment of all breaking-out computers \( B \) simultaneously.

2.5. A Common Flaw of Models with Permanent R Compartment

Some previous epidemic models of computer viruses have a permanent \( R \) compartment, that is, the compartment of all uninfected computers having permanent immunity \([19, 21–24, 26–31]\). Such models are especially suitable for a specific computer virus.

When modeling the spread of a large family of existing and future viruses sharing a small number of common features, all currently uninfected computers worldwide will always be confronted with the threat from new variants of existing viruses as well as new virus strains. As thus, it is likely that a computer that has previously been cured be infected by new kinds of viruses, implying that no computer can acquire permanent immunity. In a word, a model that aims to capture the spread of a large family of computer viruses should not possess a permanent \( R \) compartment.

3. The SLBS Model: A Generic Model

This section is intended to propose a generic epidemic model of computer viruses. Based on the previous discussions, all internal computers are classified as three categories: uninfected internal computers \( S \) computers), latent internal computers \( L \) computers), and breaking-out internal computers \( B \) computers). In parallel, all external computers are classified as three categories: uninfected external computers \( S^* \) computers), latent external computers \( L^* \) computers), and breaking-out external computers \( B^* \) computers). Let \( S(t) \), \( L(t) \), and \( B(t) \) denote the numbers of \( S \), \( L \), and \( B \) computers at time \( t \), respectively. Next, let us impose the following assumptions.

(A1) The Internet is fully connected, that is, every internal computer is equally probable to be accessed by any other internal computer.

(A2) \( S^* \) computers are connected to the Internet at constant rate \( \mu_1 \), while \( L^* \) computers are connected to the Internet at constant rate \( \mu_2 \). Let \( \mu = \mu_1 + \mu_2 \).

(A3) In normal case, every internal computer is disconnected from the Internet with constant probability \( \delta_1 \).

(A4) Due to the outbreak of viruses, every \( B \) computer is disconnected from the Internet with constant probability \( \delta_2 \).

(A5) Due to the contact with infected removable storage media, every \( S \) computer is infected with constant probability \( \theta \).
Due to the outbreak of viruses, every $L$ computer becomes a $B$ computer with constant probability $\alpha$.

Due to the contact with $L$ or $B$ computers, at time $t$ every $S$ computer becomes an $L$ computer with probability $f(L(t) + B(t))$, where the function $f$ is continuously differentiable.

Every $B$ computer is cured with constant probability $\gamma_1$, every $L$ computer is cured with constant probability $\gamma_2$, and every $B$ computer is partially cured, that is, becomes an $L$ computer, with constant probability $\gamma_3$.

Based on this collection of assumptions, the corresponding mean-field model, which will be referred to as the SLBS model, is formulated as

$$
\dot{S} = \mu_1 + \gamma_1 B + \gamma_2 L - f(L + B)S - (\delta_1 + \theta)S,
\dot{L} = \mu_2 + f(L + B)S + \theta S + \gamma_3 B - (\alpha + \gamma_2 + \delta_1)L,
B = \alpha L - (\gamma_1 + \gamma_3 + \delta_1 + \delta_2)B,
$$

(3.1)

where $S = S(t)$, $L = L(t)$, and $B = B(t)$.

Based on the following reasons, the SLBS model is well qualified to serve as one of the most fundamental epidemic models of computer viruses.

(i) This model captures the main features of computer viruses.

(ii) Most factors that have conspicuous effect on the diffusion of viruses are incorporated into this model.

(iii) As a generic model, this model includes as special cases a large number of particular models of interest.

(iv) More complicated spread mechanisms of viruses can be characterized by modifying or extending this model properly.

Now, let us give a brief analysis of the SLBS model. First, assume every $L$ or $B$ computer infects any $S$ computer mutually independently and with constant probability $\beta$. A simple calculation gives

$$f(L + B) = 1 - (1 - \beta)^{L + B}.$$

(3.2)

Suppose $\beta \ll 1$, which is consistent with actual conditions. There are three possibilities, which are listed as follows:

(i) $L + B \ll \beta^{-1}$. Then $f(L + B) \approx \beta(L + B)$;

(ii) $L + B \sim \beta^{-1}$. Then $f(L + B) \approx 1 - e^{-\beta(L + B)}$;

(iii) $L + B \gg \beta^{-1}$. Then $f(L + B) \approx 1$.

Second, let $N(t) = S(t) + L(t) + B(t)$. Then

$$\frac{dN(t)}{dt} = \mu - \delta_1 N(t) - \delta_2 B(t) \leq \mu - \delta_1 N(t).$$

(3.3)
If \( N(t) > \mu/\delta_1 \), then \( dN(t)/dt < 0 \), implying \( \limsup_{t \to \infty} N(t) \leq \mu/\delta_1 \). After a moment of reflection, it can be seen that, for arbitrarily small \( \varepsilon > 0 \), the simply connected compact set

\[
\Omega_\varepsilon = \left\{ (S,L,B) \in \mathbb{R}_+^3 : S + L + B \leq \frac{\mu}{\delta_1} + \varepsilon \right\}
\]  

(3.4)

is positively invariant for the SLBS model.

Finally, the SLBS model would have a unique virus-free equilibrium \( E_0 = (\mu/\delta_1, 0, 0) \) if \( \mu_2 = \theta = 0 \). Otherwise, this model would have no virus-free equilibrium. As far as the SLBS model is concerned, the following problems are yet to be studied:

(i) stability of the virus-free equilibrium, if it exists,

(ii) existence and number of endemic equilibria, as well as their respective stabilities,

(iii) more complex dynamic behaviors, such as bifurcations and chaos, of the model.

Very recently, the authors [43–45] proposed three new models, which are formally analogous to special instances of the SLBS model. All of the three models, however, assume that the number of computers connected to the Internet keeps constant, which is not perfectly consistent with actual conditions. The proposed SLBS model removes that unrealistic assumption and, hence, can better describe the epidemics of viruses.

### 4. Some Generalizations of the SLBS Model

#### 4.1. The Impulsive SLBS Model

From the smoothness of the right-hand-sided functions in the SLBS model, it can be concluded that the solutions to the model are all smooth. In reality, however, the emergence of a new type of viruses often leads to a sharp rise in the number of infected computers. Likewise, the appearance of a new type of patches could yield a drastic drop in the number of infected computers. In this context, the SLBS model should be modified by incorporating impulsive terms.

Let \( \{t_k\}_{k \in \mathbb{N}} \), \( t_k \to \infty \), denote the sequence of time instants at each of which the number of infected computers rises rapidly, and let \( \{s_k\}_{k \in \mathbb{N}}, s_k \to \infty \), denote the sequence of time instants at each of which the number of infected computers falls dramatically. Let us adopt the assumptions \((A1)-(A6)\) imposed in the SLBS model, and modify the assumptions \((A7)-(A8)\) in the following fashion.

\[(A7')\] If \( t = t_k \) for some \( k \), exactly \( pS(t_k) \) \( S \) computers are infected simultaneously at time \( t \), where \( p \) is a constant. Otherwise, the assumption is the same as \((A7)\).

\[(A8')\] If \( t = s_k \) for some \( k \), exactly \( q_1B(s_k) \) \( B \) computers are cured simultaneously at time \( t \), exactly \( q_2L(s_k) \) \( L \) computers are cured simultaneously at time \( t \), and exactly \( q_3B(s_k) \) \( B \) computers are partially cured, that is, become \( L \) computers, simultaneously at time \( t \). Otherwise, the assumption is the same as \((A8)\).
Based on this collection of assumptions, the corresponding model, which will be referred to as the *impulsive SLBS model*, is formulated as

\[ S = \mu_1 + \gamma_1 B + \gamma_2 L - f(L + B)S - (\delta_1 + \theta)S, \quad t \neq t_k, s_k, \; k \in N, \]
\[ L = \mu_2 + f(L + B)S + \theta S + \gamma_3 B - (\alpha + \gamma_2 + \delta_1)L, \quad t \neq t_k, s_k, \; k \in N, \]
\[ \dot{B} = aL - (\gamma_1 + \gamma_3 + \delta_1 + \delta_2)B, \quad t \neq s_k, \; k \in N, \]
\[ S(t_k^+) = (1 - p)S(t_k), \quad k \in N, \]
\[ L(t_k^+) = L(t_k) + pS(t_k), \quad k \in N, \]
\[ S(s_k^+) = S(s_k) + q_1B(s_k) + q_2L(s_k), \quad k \in N, \]
\[ L(s_k^+) = (1 - q_2)L(s_k) + q_3B(s_k), \quad k \in N, \]
\[ B(s_k^+) = (1 - q_1 - q_3)B(s_k), \quad k \in N. \]

(4.1)

The impulsive SLBS model is a generic model, which subsumes the following two particular models of interest:

(i) *Impulsive intoxication model*, which is formulated as

\[ S = \mu_1 + \gamma_1 B + \gamma_2 L - f(L + B)S - (\delta_1 + \theta)S, \quad t \neq t_k, \; k \in N, \]
\[ L = \mu_2 + f(L + B)S + \theta S + \gamma_3 B - (\alpha + \gamma_2 + \delta_1)L, \quad t \neq t_k, \; k \in N, \]
\[ \dot{B} = aL - (\gamma_1 + \gamma_3 + \delta_1 + \delta_2)B, \quad t \neq s_k, \; k \in N, \]
\[ S(t_k^+) = (1 - p)S(t_k), \quad k \in N, \]
\[ L(t_k^+) = L(t_k) + pS(t_k), \quad k \in N; \]

(4.2)

(ii) *Impulsive detoxication model*, which is formulated as

\[ S = \mu_1 + \gamma_1 B + \gamma_2 L - f(L + B)S - (\delta_1 + \theta)S, \quad t \neq s_k, \; k \in N, \]
\[ L = \mu_2 + f(L + B)S + \theta S + \gamma_3 B - (\alpha + \gamma_2 + \delta_1)L, \quad t \neq s_k, \; k \in N, \]
\[ \dot{B} = aL - (\gamma_1 + \gamma_3 + \delta_1 + \delta_2)B, \quad t \neq s_k, \; k \in N, \]
\[ S(s_k^+) = S(s_k) + q_1B(s_k) + q_2L(s_k), \quad k \in N, \]
\[ L(s_k^+) = (1 - q_2)L(s_k) + q_3B(s_k), \quad k \in N, \]
\[ B(s_k^+) = (1 - q_1 - q_3)B(s_k), \quad k \in N. \]

(4.3)

### 4.2. A Consideration of the Delay Terms

There are three potential delay factors that have notable influence on the spread of computer viruses.
(i) Due to the time cost needed to develop new viruses, there is a delay from the time a $B$ computer is cured to the time this computer is infected again.

(ii) Due to the intrinsic latent period of viruses, there is a delay from the time an $S$ computer is infected to the time this computer breaks out.

(iii) Due to the time cost needed to develop new patches, there is a delay from the time an $L$ computer breaks out to the time this computer is cured.

A question arises: is it necessary to incorporate delay terms in the standard SLBS model? In order to answer this question, let us make a brief analysis from four aspects.

(i) The SLBS model assumes that an $S$ computer is infected randomly, which implicitly includes a time delay in developing new viruses.

(ii) The SLBS model supposes that an $L$ computer breaks out randomly, which, to a certain extent, implies a latency-related delay.

(iii) The SLBS model postulates that a $B$ computer is cured randomly, which, in some sense, also implies a time delay in developing new antivirus software.

(iv) The incorporation of delay terms in the SLBS model would greatly enhance the hardness in the theoretical study of the resulting models.

Due to these reasons, we do not suggest to study SLBS models incorporated with delay terms.

### 4.3. The Stochastic SLBS Model

All of the above-mentioned models are based on the assumption that all system parameters do not change with time. In reality, however, there are numerous uncertain factors, which are often abstracted as noises, that have significant influence on these parameters. As a result, some or all system parameters are constantly varying with time. Therefore, the predictions made from any deterministic model may have a significant deviation from the actual condition.

An alternative to the deterministic modeling of viruses is to incorporate noises in some or all system parameters so as to form a stochastic model. As an instance, noise terms can be incorporated in the $\mu_1$ and $\mu_2$ parameters of the original SLBS model to produce a particular stochastic SLBS model of the form

\[
\begin{align*}
S &= (\mu_1 + \sigma_1 W_1) + \gamma_1 B + \gamma_2 L - f(L + B)S - (\delta_1 + \theta)S, \\
\dot{L} &= (\mu_2 + \sigma_2 W_2) + f(L + B)S + \theta S + \gamma_3 B - (\alpha + \gamma_2 + \delta_1)L, \\
\dot{B} &= \alpha L - (\gamma_1 + \gamma_3 + \delta_1 + \delta_2)B,
\end{align*}
\]

(4.4)

where $W_1 = W_1(t)$ and $W_2 = W_2(t)$ stand for the standard one-dimensional Wiener processes (i.e., Brownian motions) and $\sigma_1$ and $\sigma_2$ stand for the standard deviations associated with $W_1$ and $W_2$, respectively.
5. Concluding Remarks

By inspecting the characteristics of computer viruses carefully, the flaws of some previous epidemic models of viruses have been indicated. On this basis, a generic epidemic model of viruses (the SLBS model) has been established, and some of its generalizations have been suggested.

Towards this direction, a great diversity of particular models with parameter restrictions are yet to be investigated. Besides, the standard SLBS model is based on fully connected networks and hence cannot capture the effect of the topological structure of the Internet on the spread of computer viruses. It would be highly rewarding to study the qualitative properties of the SLBS model on scale-free networks.

Acknowledgments

The authors are grateful to the anonymous reviewers for their valuable comments. This work is supported by Doctorate Foundation of Educational Ministry of China (Grant no. 20110191110022).

References


Submit your manuscripts at
http://www.hindawi.com