Applied Mathematics and Computation 320 (2018) 282–291

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Edge-based modeling of computer virus contagion on a tripartite graph

Wei Pan[a,b,c,*], Zhen Jin[b,c,*]

[a] School of Information and Communication Engineering, North University of China, Taiyuan 030051, China
[b] Complex Systems Research Center, Shanxi University, Taiyuan 030006, China
[c] Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Shanxi University, Taiyuan 030006, China

ARTICLE INFO

Keywords:
File virus
Edge-based contagion
Tripartite graph

ABSTRACT

As a typical computer virus, a file virus can parasitize in executable files and infect other files when the host files are executed. Due to the strong similarity between computer viruses and their biological counterparts, in this paper we adapt the epidemiologically compartmental models to study the computer virus contagion. To trace the transmission process of file viruses and determine effective control measures, we derive a pairwise mathematical model by taking account of edge-based contagions. By constructing a tripartite graph, we can determine the potential edges on which contagions take place. The sensitivity analysis for some parameters is performed, indicating that the contagion of file viruses can be effectively restrained by reducing the use of portable storage devices with computers which have not installed antivirus softwares or by reducing the transmission rate from infected web pages to susceptible computers. It is also found that the final number of infected computers is much lower in scale-free networks than in Poisson degree distributed networks.

1. Introduction

A typical unpleasant situation may occur that a computer crashes with a blue screen when we visit web pages to download documents or when we transfer documents from/to the computer via removable disks. For computers that have installed anti-virus softwares, they can remind users to detect viruses from web pages or portable disks, while this is not the case for computers without antivirus softwares.

In 1983, the term “computer virus” was coined by Cohen [1] after he wrote a program that could “infect” computers, replicate itself and spread from one computer to another. Typically, a computer virus is a program that can harm computers and their files by modifying them to include a possible evolved copy of itself [1]. Since then, the number of computer viruses has been enormously increasing. In the past decades, computer viruses have been recognized as one of serious safety issues due to their threat to computer systems.

In general, computer viruses mentioned above are called file viruses, which infect files mostly by parasitizing. File viruses insert malicious codes into some executable files or data files containing executable codes, so that the malicious

* Corresponding author at: Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Shanxi University, Taiyuan 030006, China.
E-mail addresses: onlyonebetty@163.com (W. Pan), jinzhen@sxu.edu.cn (Z. Jin).

https://doi.org/10.1016/j.amc.2017.09.044
0096-3003/

codes are executed once the files are accessed. It has been several decades since the executable file virus known as "Black Friday" emerged in 1987. The file viruses have some similar traits of biological pathogens, such as parasitism and infectiousness. Cohen [1] and Murray [2] pointed out the connection between the computer virus contagion and biological epidemiology. Kephart and White [3] took the first step toward modeling the spread behavior of computer viruses by considering a susceptible-infected-susceptible (SIS) model on a directed random graph. Afterward many attempts have been made to describe computer virus propagation using feasible explicit models [4–14] based on mathematical epidemiology [15–22]. For instance, Piqueira and Araujo [4] studied a modified version of the susceptible-infected-removed (SIR) model by including an antidotal compartment and derived the bifurcation condition that distinguishes the virus-free and virose equilibria. After that, Han and Tan [5] and Mishra and Pandey [6] proposed a susceptible-infected-recovered-susceptible (SIRS) model and a susceptible-exposed-infected-recovered-susceptible (SEIRS) model, respectively, based on the homogeneous mixing assumption, where the basic reproduction number R_0 [15] and the stability analysis have been provided. Yang and Yang [9] investigated the influences of removable storage devices on the computer virus spread with a node-based compartmental model. Moreover, Ren et al. [10] explored the effects of antivirus ability on the spreading dynamics of computer viruses, particularly on the backward bifurcation and Hopf bifurcation. Muroya and Kuniya [11] presented a nonresident computer virus model and established the global dynamics depending on the value of the basic reproduction number R_0. Recently, Yang et al. [12] developed a mean-field computer virus spreading model to incorporated the effects of the heterogeneous connectivity of the underlying network [17,23], where the authors argued that the strong heterogeneity leads to a smaller transmission threshold, above which the virus will break out through the computer network. In addition, Wang and Cao [13] mathematically proved the globally asymptotical stability of the virus equilibrium for the network-based computer virus propagation model. All of the above mentioned models [4–14] are in essence node-based compartment models developed under the assumption of mass action law [15]. Although the node-based compartment models are beneficial to the determination of epidemic threshold and the prediction of final infected density, they are insufficient to resolve the evolution or dynamical behavior of infectious contacts (namely, contagious edges on which the virus or disease is transmitted). To address this issue, in this paper we propose a pairwise computer virus contagion model on a tripartite graph by taking account of edge-based transmission. It is worth noting that our study incorporates all the effects of removable storage devices [9], antivirus software [4,10] and the heterogenous connectivity [12,13] on the dynamics of computer virus spread. In what follows, we will clarify the basic motivations and assumptions towards edge-based modeling of computer virus spread on a tripartite network.

Similar to biological pathogens, each type of computer virus has its own characteristics and patterns of infection. The file viruses are in general not considered to run automatically. The computer virus in the infected host files can be activated when the host files are executed, thereby infecting noninfected files in the same system [24]. The infected files are not able to recover automatically until the viruses hidden in them have been killed. The virus in an infected file can be killed only if the file is detected by an antivirus software. Actually, the file virus itself as well as the file or the document parasitized does not have any direct harm. However, once the infected file is activated, it leads to the propagation of file viruses like epidemic outbreaks, disabling the normal functions of computers.

Based on both of the above features of computer virus propagation and the complex network theory [25], we consider each carrier (for example, a web page, a computer or a removable storage disk) as a node in a network. The propagation process of the computer virus depends on the carrier type. Using the idea of the classic compartment model firstly proposed by Bernoulli [26], we divide carriers into several compartments depending on whether they have infected host documents. The entire network is regarded as a tripartite graph consisting of three types of nodes, and the two endpoints of each edge in the network are of different types. As we know from Ref [27], contagions occur along infectious edges. Nevertheless, these can result in tremendous difficulty in solving practical problems for portraying the transmission process.

In this work, we employ a mathematical model to describe the contagion of file viruses by taking account of edge-based contagions between computers and web pages or removable storage disks. We address this issue by deriving a system of 12 nonlinear ordinary differential equations (ODEs). We also carry out sensitivity analysis to identify effective measures for controlling the spread of file viruses.

2. Model formulation

In the network investigated, there are three types of nodes, which represent web pages, computers and removable disks, respectively. They are distinguished by different superscripts p, c and r, which are shown in Fig. 1. We further divide web pages into two types: healthy web pages and infected web pages, according to whether they have been embedded in viruses. The computers are accordingly divided into the following three compartments: the susceptible, the infected and the recovered with anti-virus softwares. We also divide removable disks into two compartments, namely, healthy disks and infected ones. In Fig. 1, a connection between a web page and a computer indicates that the computer visits the web page, and an edge connecting a computer with a removable disk means that the removable disk is used in the computer.

We consider the links connecting different nodes as arcs, as shown in Figs. 1 and 2. A healthy web page is infected by being embedded viruses at a rate b and an infected web page recovers at a rate d by clearing viruses owing to human causes. For web pages, their states are independent of others. The susceptible computers can be infected by visiting infected web pages as well as using infected removable disks. An infected computer recovers at rate α by clearing viruses with antivirus softwares. Recovered computers gain immunity after being installed antivirus softwares. The virus in an infected