

Epidemiological Models Applied to Viruses in Computer Networks

¹José Roberto Castilho Piqueira, ¹Betyna Fernández Navarro and
^{1,2}Luiz Henrique Alves Monteiro

¹Telecommunications and Control Engineering Department, Politechnic School,
Sao Paulo University, Avenida Prof. Luciano Gualberto, travessa 3 - 158, 05508-900 São Paulo, Brazil
²Electrical Engineering Graduate Department, Mackenzie Presbyterian University

Abstract: To investigate the use of classical epidemiological models for studying computer virus propagation we described analogies between computer and population disease propagation using SIR (Susceptible-Infected-Removed) epidemiological models. By modifying these models with the introduction of anti-viral individuals we analyzed the stability of the disease free equilibrium points. Consequently, the basal virus reproduction rate gives some theoretical hints about how to avoid infections in a computer network. Numerical simulations show the dynamics of the process for several parameter values giving the number of infected machines as a function of time.

Key words: Basal Reproduction Rate, Computer Virus, Computer Network, Dynamical Systems

INTRODUCTION

Nowadays, computer viruses are an important risk to computational systems endangering either corporation systems of all sizes or personal computers used for simple applications as accessing bank accounting or even consulting entertainment activities schedules. The viruses are being developed simultaneously with the computer systems and the use of INTERNET facilities increases the number of damaging virus incidents.

Since the first trials on studying how to combat viruses, biological analogies were established because biological organisms and computer networks share many characteristics as, for example, large number of connections among large number of simple components creating complex system [1].

Local systems in a computer network can be attacked generating malfunctions that, spreading along the network, produce network-wide disorders following a similar qualitative model of disease spreading for a biological system. This is the main reason for designating attacks against networks by biological terms as worms and viruses.

Using these ideas, it is important to consider that computer viruses have two different levels for being studied: microscopic and macroscopic [2].

The microscopic level has been the subject of several studies. For instance, [3, 4] establishes theoretical principles about how to kill the new viruses created every day. Following the virus development, computer immunology is a new discipline capable of creating efficient anti-virus strategies as programs that are being sold all over the world guaranteeing protection to individual users of a global network [5, 6].

However, the macroscopic approach has not been receiving the same attention in spite of epidemiology analogies being an important tool in order to establish the policies to preventing infections by giving figures about how to update the anti-virus programs.

The interesting but simple model considering exponential variation in the number of computer viruses, proposed by [7], couldn't be considered realistic because the lack of limits for the growth, which is a natural phenomenon either in biological or in computer systems.

There is vast catalog of Mathematical Biology models indicated for epidemiology [8]. One of them, called SIR (Susceptible-Infected-Removed) model, was originally proposed by [9].

Here, we employ a modified version of such a model in order to obtain parameter combinations representing situations with asymptotically stable disease-free solutions.

The relations among network parameters can provide some hints about how to prevent infections in networks. An expression for the maximum infection rate of computers equipped with anti-virus to avoid the propagation of new infections is given. If this number is known, an updating plan for anti-virus programs in a computer network can be elaborated.

The Model: We proposed the model represented in Fig. 1 for the dynamics of the infection propagation in a computer network. The model contains a modification related to the traditional SIR model [8], with an antidotal population compartment (A) representing the nodes of the network equipped with fully effective anti-virus programs.

The total population T is divided into four groups: S of non-infected computers subjected to possible infection; A of non-infected computers equipped with anti-virus; I of infected computers; and R of removed ones due to the infection or not.

The model is called SAIR (Susceptible-Antidotal-Infected-Removed). Its dynamics is described by:

$$dS/dt = N - \alpha \cdot S \cdot A - \beta_{SI} \cdot S \cdot I - \mu \cdot S + \sigma_{IS} \cdot I + \sigma_{RS} \cdot R \quad (1)$$

$$dI/dt = \beta_{SI} \cdot S \cdot I + \beta_{AI} \cdot A \cdot I - \sigma_{IS} \cdot I - \delta \cdot I - \mu \cdot I \quad (2)$$

$$dR/dt = \delta \cdot I - \sigma_{RS} \cdot R - \mu \cdot R \quad (3)$$

$$dA/dt = \alpha \cdot S \cdot A - \mu \cdot A - \beta_{AI} \cdot A \cdot I \quad (4)$$

The parameters of the model are defined as follows:

- * N : influx rate, representing the incorporation of new computers to the network;
- * μ : mortality rate not due to the virus;
- * β_{SI} : infection rate of susceptible computers;
- * $p_{SI} = I / (T-1)$: probability of susceptible computers to establish an effective communication with infected ones;
- * β_{AI} : infection rate of antidotal computers due to the onset of new virus;
- * $p_{AI} = I (1 - \eta) / (T-1)$: probability of antidotal computers to establish an effective communication with infected ones;
- * δ : removing rate of infected computers;
- * k_i / n_i : probability of the execution of an infected file, i.e., probability of conversion of non-infected computers into infected ones;
- * n_i : number of executable files in the i -computer, considering that all the files have the same probability of being executed;
- * k_i^i : number of infected files in the i -computer;
- * k_n^i : number of normal files in the i -computer;
- * σ_{IS} : recovering rate of infected computers;
- * $p_{IS} = (A) / ((T-1)\eta)$: recovering probability of infected computers, i.e., probability of occurring an effective communication between infected computers and antidotal ones;
- * σ_{RS} : recovering rate of removed computers, with an operator intervention;
- * α : conversion of susceptible computers into antidotal ones, occurring when susceptible computers establish effective communication with antidotal ones and the antidotal installs the anti-virus in the susceptible;
- * $p_{SA} = A / (T-1)$: probability of an antidotal computer installing the anti-virus in a susceptible one, when an effective communication is established.

For simplicity, the influx rate is considered to be $N = 0$, representing that there are no incorporation of new computers to the network during the propagation of a virus that is considered to be very fast. The same reason justifies the choice $\mu = 0$.

Under these conditions, the system is modeled by equations:

$$dS/dt = -\alpha \cdot S \cdot A - \beta_{SI} \cdot S \cdot I + \sigma_{IS} \cdot I + \sigma_{RS} \cdot R \quad (5)$$

$$dI/dt = \beta_{SI} \cdot S \cdot I + \beta_{AI} \cdot A \cdot I - \sigma_{IS} \cdot I - \delta \cdot I \quad (6)$$

$$dR/dt = \delta \cdot I - \sigma_{RS} \cdot R \quad (7)$$

$$dA/dt = \alpha \cdot S \cdot A - \beta_{AI} \cdot A \cdot I \quad (8)$$

Since $dS/dt + dA/dt + dI/dt + dR/dt = 0$, then $S + A + I + R = T = \text{constant}$ for any instant t .

Equilibrium Points: In order to investigate the properties of the dynamics of the model, we determine the equilibrium points by considering that all the derivatives of population compartments vanish when this kind of solution holds.

There are disease free equilibrium points, which represent the situations where the infected population is null ($I = 0$). These points are given by:

$$P_1^* = (S = 0, A = T, I = 0, R = 0) \quad (9)$$

$$P_2^* = (S = T, A = 0, I = 0, R = 0) \quad (10)$$

Thus, all computers are susceptible or antidotal when $I = 0$.

Expressions for endemic equilibrium points are given by:

$$P_3^* = (S = (\sigma_{IS} + \delta) / \beta_{SI}, A = 0, I = (\sigma_{RS} \cdot R) / \delta, R = T - S - I) \quad (11)$$

$$P_4^* = (S = (\beta_{AI} \cdot I) / \alpha, A = -(\beta_{SI} \cdot \beta_{AI} \cdot I - \sigma_{IS} \cdot \alpha - \delta \cdot \alpha) / (\alpha \cdot \beta_{AI}), I = T - S - A, R = (\delta \cdot I) / \sigma_{RS}) \quad (12)$$

The expressions for the equilibrium points make possible to obtain the conditions for stability of disease free solutions that are useful to establish the minimum recovering rate that a network is supposed to have in order to avoid the propagation of infections.

Disease Free Stability and Basal Reproduction Rate:

The stability of the equilibrium points determines the viral evolution represented by our SAIR model: if there is asymptotically stable free-disease equilibrium point, then the disease can disappear; if there is not, it becomes endemic.

We obtain the linear approximation of the model around the equilibrium points by calculating the corresponding Jacobian (4x4) matrix [10] as:

$$J = \begin{bmatrix} -A\alpha - I\beta_{SI} & -\beta_{SI}S + \sigma_{IS} & \sigma_{RS} & -\alpha S \\ I\beta_{SI} & \beta_{SI}S + \beta_{AI}A - \sigma_{IS} - \delta & 0 & \beta_{AI}I \\ 0 & \delta & -\sigma_{RS} & 0 \\ A\alpha & -\beta_{AI}A & 0 & \alpha S - \beta_{AI}I \end{bmatrix} \quad (13)$$

Stability of P_1^* : Calculating this Jacobian matrix in P_1^* , we obtain:

$$J = \begin{bmatrix} -A\alpha & \sigma_{IS} & \sigma_{RS} & 0 \\ 0 & \beta_{AI}A - \sigma_{IS} - \delta & 0 & 0 \\ 0 & \delta & -\sigma_{RS} & 0 \\ A\alpha & -\beta_{AI}A & 0 & 0 \end{bmatrix} \quad (14)$$

With eigenvalues given by:

$$\begin{aligned} \lambda_1 &= -T \cdot \alpha \\ \lambda_2 &= \beta_{AI} \cdot T - \sigma_{IS} - \delta \\ \lambda_3 &= -\sigma_{RS} \\ \lambda_4 &= 0 \end{aligned}$$

In spite of one of the eigenvalues being zero the analysis of the stability can be conclusive because the A-axis is a central manifold such that A remains constant for any initial condition [10]. Then, as λ_1 and λ_3 are real and negative the problem is reduced to analyze λ_2 . The condition $\lambda_2 < 0$ is necessary and sufficient for considering P_1^* asymptotically stable. Consequently, the condition for asymptotic stability of this point is:

$$T < (\sigma_{IS} + \delta) / \beta_{AI} \quad (15)$$

Stability of P_2^* : Calculating the Jacobian matrix in P_2^* , we obtain:

$$J = \begin{bmatrix} 0 & -\beta_{SI}S + \sigma_{SI} & \sigma_{RS} & -\alpha S \\ 0 & \beta_{SI}S - \sigma_{IS} - \delta & 0 & 0 \\ 0 & \delta & -\sigma_{RS} & 0 \\ 0 & 0 & 0 & \alpha S \end{bmatrix} \quad (16)$$

With eigenvalues given by:

$$\begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= \beta_{SI} \cdot T - \sigma_{IS} - \delta \\ \lambda_3 &= -\sigma_{RS} \\ \lambda_4 &= \alpha \cdot T \end{aligned}$$

As λ_4 is real and positive, P_2^* is unstable for any combination of parameters.

Basal Reproduction Rate: In epidemiology literature it is well known the concept of basal reproduction rate (R_0). This is a bifurcation parameter meaning that, if $R_0 > 1$, all disease free equilibrium points are unstable and the epidemic process persists. If $R_0 < 1$, there is asymptotically stable disease free equilibrium point; thus, the disease can vanish. In our model, the basal reproduction rate can be determined by analyzing the stability of P_1^* .

From (15), we obtain:

$$R_{01} = (\beta_{AI} \cdot T) / (\sigma_{IS} + \delta) \quad (17)$$

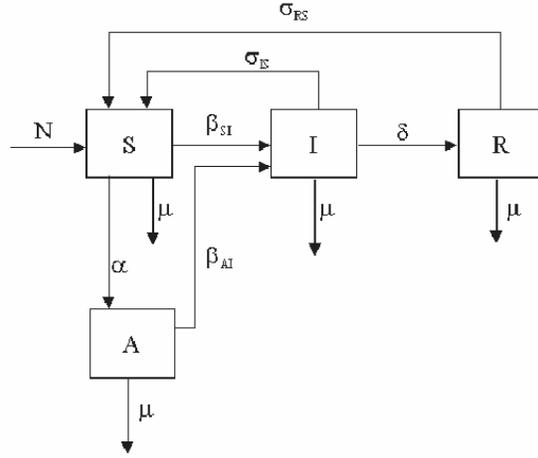


Fig. 1: Susceptible-Antidotal-Infected-Removed Model

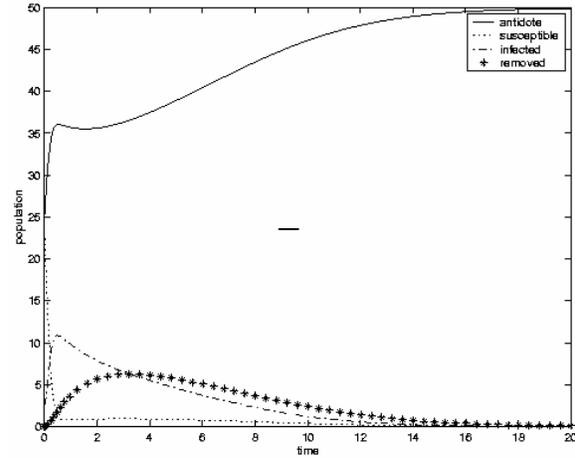


Fig. 2: Dynamics with an Asymptotically Stable Disease-Free Equilibrium Point

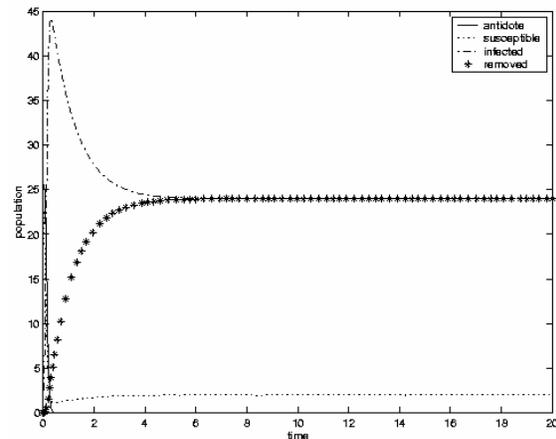


Fig. 3: Unstable Disease-Free Dynamics

Consequently, if $R_{01} < 1$, the virus propagation along the network is avoided. Then, the limit infection rate of antidotal computers β_{AI} is given by:

$$\beta_{AI} = (\sigma_{IS} + \delta)/T \quad (18)$$

Therefore, (18) is the parameter that must be evaluated in a computer network, providing figures about how to maintain the satisfactory operation of a computer network.

Simulations: The condition (18) gives a theoretical prediction about possible ways of avoiding infection in a computer network. We performed numerical simulations of the SAIR model by supposing a local network with 50 computers and half of them equipped with anti-virus programs. The goal is to follow what happens when few infected individuals are introduced in the network. β_{AI} is taken as the control parameter. The values of the other parameters are: $\alpha = 0.1$, $\beta_{SI} = 0.5$, $\sigma_{IS} = 0.5$, $\sigma_{RS} = 0.5$, $\delta = 0.5$.

Considering these values, $\beta_{AI} = 0.02$ represents the limit of the basal reproduction rate. Then we simulate the SAIR model for $\beta_{AI} < 0.02$, and $\beta_{AI} > 0.02$. Figure 2 shows a simulation for $\beta_{AI} = 0.01$. Infected and removed populations vanish and the network, in the long term, is in a good operational state. Figure 3 exhibits a simulation for $\beta_{AI} = 0.5$. All the populations become composed by either infected or removed computer. In the long term, there is a "low" density of operational computers and, consequently, no network.

CONCLUSION

Viral attacks against computer networks are an important research area because the defense strategies need to be able to avoid infection propagation. In this work we presented the SAIR model based on epidemiological studies and conditions for the asymptotically stability of the disease free equilibrium were deduced. Some simulations were performed showing how a parameter, analogous to the epidemic basal reproduction rate, affects the dynamics of the infection propagation.

ACKNOWLEDGEMENT

JRCP and LHAM are supported by CNPq.

REFERENCES

1. Denning, P.J., 1990. Computers under attack. Reading, Mass: Addison-Wesley.
2. Kephart, J.O., S.R. White and D.M. Chess, 1993. Computers and Epidemiology. IEEE Spectrum, pp: 20-26.
3. Cohen, F., 1987. Computer Viruses, Theory and Experiments Therapies. Computer and Security, 6: 22-35.
4. Cohen, F., 1990. A Short Course of Computer Viruses. Computer and Security, 8: 149-160.
5. Nachenberg, C., 1997. Computer Virus-Anti virus coevolution. Communications of the ACM, 40: 40-51.
6. Forrest, S., S.A. Hofmayer and A. Somayaji, 1997. Computer Immunology. Communications of the ACM, 40: 88-96.
7. Tippet, P.S., 1990. Computer Virus Replication. Comput. Syst. Eur., 10: 33-36.
8. Murray, J.D., 2002. Mathematical Biology. New York: Springer-Verlag, 3rd Edn.
9. McKendrick, A.G., 1926. Application of Mathematics to Medical Problems. Proc.Ed. Math. Soc., 44: 98-130.
10. Guckenheimer, J. and P.J. Holmes, 1983. Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields. New York: Springer-Verlag.
11. Castillo-Chavez, C., K. Cooke, W. Huang and S.A. Levin, 1989. On the Role of Long Incubation Periods in the Dynamics of Acquired Immunodeficiency Syndrome (AIDS) - I - Single Population Models. J. Math. Biol., 27: 373-398.