IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 15, 2019, accepted September 17, 2019, date of publication September 23, 2019, date of current version October 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2942809

Security Countermeasures of a SCIRAS Model for Advanced Malware Propagation

J. D. HERNÁNDEZ GUILLÉN[1], A. MARTíN DEL REY[2], (Member, IEEE), AND ROBERTO CASADO-VARA[3], (Member, IEEE)

[1]Department of Applied Mathematics, University of Salamanca, 37008 Salamanca, Spain
[2]Department of Applied Mathematics, Institute of Fundamental Physics and Mathematics, University of Salamanca, 37008 Salamanca, Spain
[3]BISITE Digital Innovation Hub, University of Salamanca, 37007 Salamanca, Spain

Corresponding author: A. Martín del Rey (delrey@usal.es)

This work was supported in part by the Ministerio de Ciencia, Innovación y Universidades (MCIU, Spain), in part by the Agencia Estatal de Investigación (AEI, Spain), and in part by the Fondo Europeo de Desarrollo Regional (FEDER, UE) under Project TIN2017-84844-C2-2-R (MAGERAN) and Project SA054G18 (MASEDECID), supported by the Consejería de Educación (Junta de Castilla y León, Spain). The work of J. D. Hernández Guillén was supported in part by the University of Salamanca, Spain, and in part by the Banco Santander under a doctoral grant.

ABSTRACT In the new and sophisticated cyber attacks (mainly, advanced persistent threats) the advanced specimens of malware such that zero-day malware play a crucial role. Due to its stealthy behavior it is very important to study and analyze its propagation process by designing mathematical models that could predict in an efficient way its spread on a network. With no doubt the computational implementation of these theoretical models leads to the develop of solutions to be used in the Security Operation Centers (SOC) with forensic purposes. The main goal of this work is to introduce a novel mathematical model to simulate advanced malware. Specifically, it is a compartmental and global SCIRAS (Susceptible-Carrier-Infectious-Recovered-Attacked-Susceptible) model where susceptible, carrier, infectious, recovered and attacked devices are considered. The local and global stability of its equilibrium points are studied and the basic reproductive number is computed. From the analysis of this epidemiological threshold, the most efficient security countermeasures are derived.

INDEX TERMS Basic reproductive number, malware spread, mathematical model, advanced persistent threats, zero-day malware.

I. INTRODUCTION

Advanced persistent threats (APTs for short) are sophisticated and complex cyber-attacks combining not only different and advanced technologies and methodologies, but also detailed information and data of the targeted network obtained from (usually) intelligence resources [1], [2]. These cyber-attacks exhibit the following main characteristics [3]: (1) they are targeted attacks, that is, the principal goal of an APT is to achieve a specifically targeted and highly valuable objective; (2) they are persistent attacks in the sense that they are constituted by several phases to perform a long-time campaign with repeated attempts; (3) They exhibit an stealthy and evasive behavior with a high level of adaptation to defenders' efforts; and, finally, (4) they are well-resourced and highly organized attacks. These types of attacks are organized and/or sponsored by large organizations or government agencies [4].

The attack methods used in APTs are diverse and sophisticated, and its choice depends on the characteristics of the targeted environment [5]. These tools include, among others, social engineering, custom encryption technology, binary command-and-control code, rootkits, and advanced malware that exploits (zero-day vulnerabilities): zero-day malware.

Zero-day malware can be defined as a specimen of malicious code that exploits an unknown (and, consequently, non-patched) vulnerability. As a consequence this type of malware exhibits an evasive and stealthy behavior to propagate as undetected as possible [6].

Most of efforts of scientific and technological community are devoted to the design of defense mechanisms against APTs ([7], [8]) and to implement efficient methods to detect this type of cyberattacks (see, for example, [9]–[11]). Apart from this approach it is also of interest to propose and

The associate editor coordinating the review of this manuscript and approving it for publication was Aniruddha Datta.

analyze models that simulate the temporal evolution of these cyber-attacks, specially the spread of zero-day malware. In this sense, although several models for (standard) malware propagation have been proposed in the scientific literature (see [12]–[14] and references therein), very few have appeared dealing with zero-day malware spreading. In fact, as far as we know there is only four works dealing with the use of Mathematical Epidemiology to [15]–[18].

In [15] a computer engineering approach to this phenomenon is done. In this work the authors designed a novel simulator, based on the finite state machine paradigm, to simulate the spreading of zero-day worms on a full IPv4-sized network. On the other hand, an epidemiological model to combat a phishing attack containing zero-day malware was introduced in [16]. Specifically it is a deterministic and global model where susceptible, infectious, quarantined and recovered devices are considered and, in addition, a cyber resilience recovery model was proposed. In [17] a global and deterministic model was introduced and its stability analysis was studied; in this case the compartments involved in the dynamics were weak-defensive nodes, attacked nodes, strong-defensive nodes and compromised nodes. Finally, in [18] a theoretical model to simulate an advanced persistent distributed denial-of-service attack was presented. It is a compartmental and stochastic model where the population of devices is divided into four classes: susceptible, infected, tolerant and missed nodes. The equilibrium points are computed and its main qualitative characteristics are studied.

The model proposed in this work is also a compartmental, global and deterministic. The novelty of this model, that makes it different from those mentioned above, is that there are two main characteristics of the APTs involved in the dynamics: the stealthy and the use of intelligence resources to decide whether a compromised device should be successfully attacked or not. Consequently, in our proposal we will consider two "infected" compartments: infectious devices (those susceptible ones reached by malware) and attacked devices (the reached devices that are classified by advanced malware as targeted devices). Moreover, also carrier devices play an important role in our model since they can be considered as efficient transmission vectors although they cannot be effectively damaged.

The rest of the paper is structured as follows: In section II the general description of the new theoretical model is presented; its mathematical formulation is developed in section III, and its qualitative analysis is introduced in section IV. In section V some illustrative simulations showing the steady states are presented; the analysis of the basic reproductive number to obtain efficient security countermeasures is detailed in section VI. Finally, the conclusions are presented in section VII.

II. GENERAL DESCRIPTION OF THE SCIRAS MODEL

The main purpose of the model proposed in this work is to simulate the propagation of advanced malware on a computer network. In this work we will suppose that malware presents the following main characteristics:

(i) Using previously collected information, the specimen of malware is able to determine if a device could be considered as a potential target or not.
(ii) Advanced malware has the ability to decide if the reached device must be effectively attacked or not.
(iii) It exhibits a stealthy and evasive behavior.

Taking into account these considerations, it is assumed that a susceptible device that has been reached by the advanced malware becomes infectious or carrier depending on the decision taken by malware after the analysis of such device. If malware considers that the device lacks the basic specifications of a potential target, then the host becomes carrier; otherwise it happens to be infectious. Note that both carrier and infectious devices are considered as transmission vectors for malware but the malicious activity could be carried out only on infectious devices.

Moreover, an infectious device becomes attacked when the malware catalogs it as an objective. This decision process is based on the gathering information on the host. On the other hand, if malware does not consider the infected device as a target then it removes itself and the device becomes recovered.

Due to the stealthy behavior of the specimen of malware, it removes itself from the host once its activity is finished. In this sense, infectious, carrier and attacked devices become recovered at a certain rate. As this type of malware can be adapted to certain security countermeasures, permanent immunity is not guaranteed; consequently, a reinfection process must be considered in the model.

Finally, a vaccination process through security countermeasures (upgrade and security patches, etc.) is considered. Note that it is reasonable to suppose that the effectiveness of these measures is very limited due to the nature of the cyber-attack.

III. MATHEMATICAL FORMULATION OF THE SCIRAS MODEL

As is previously mentioned, the epidemiological model proposed in this work is a compartmental and global model where each device can belong to different five classes at each step of time t: susceptible $S(t)$, carrier $C(t)$, infectious $I(t)$, attacked $A(t)$, or recovered $R(t)$. Specifically, it is a SCIRAS model where both reinfection and vaccination processes are considered. Moreover, it assumed that there is not population dynamics, hence

$$S(t) + I(t) + C(t) + A(t) + R(t) = N > 0, \qquad (1)$$

for every t. The main specifications of advanced malware stated in the previous section are reflected in the model as follows (see Fig. 1):

- The infection can be caused by both carriers and infectious devices, and this process depends on the transmission rate $0 \leq a \leq 1$, which is the same for these two compartments. As a consequence, the incidence (that is,

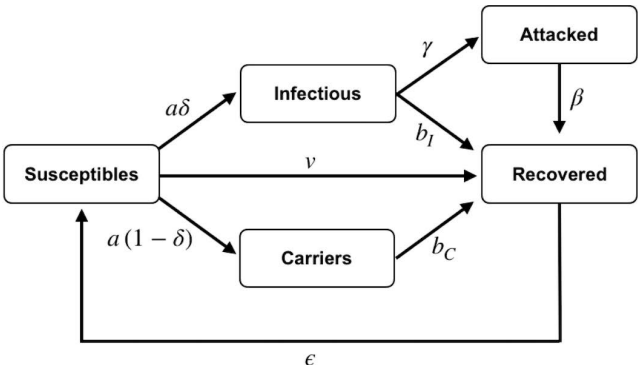


FIGURE 1. Flow diagram representing the dynamics of the model.

the new infected -carrier and infectious- devices) at step of time t is given by $aS(t)(C(t)+I(t))$. Furthermore, if δ stands for the fraction of susceptible devices which are potential targets for the cyber attack then the total incidence can be rewritten as follows:

$$\begin{aligned}\text{incidence} &= \delta aS(t)(C(t)+I(t)) \\ &\quad +(1-\delta)aS(t)(C(t)+I(t)), \end{aligned} \tag{2}$$

where $\delta aS(t)(C(t)+I(t))$ represents the new infectious devices at t, and $(1-\delta)aS(t)(C(t)+I(t))$ is the number of new carrier devices at step of time t.

- If security patches are installed, a fraction of non-infected devices, $vS(t)$, can acquire temporal immunity to cyber-attack. Due to the characteristics of advanced malware (it can exploit zero-days) it is possible assume that $0 \leq v << 1$.
- If the security software installed in the devices and/or network successfully detects and removes the malware, also carrier and infectious devices acquire temporal immunity at rates b_C and b_I, respectively. As in the previous case, $0 \leq b_C, b_I << 1$. As a consequence, $b_C C(t)$ and $b_I I(t)$ represent the new recovered devices from carrier and infectious compartments respectively.
- A fraction of infectious devices, $\gamma I(t)$, are classified as targets by malware and, consequently, they are effectively attacked. Once malware finishes it malicious activity, the host becomes recovered at rate $0 \leq \beta \leq 1$. That is, $\beta A(t)$ represents the number of new recovered devices from attacked compartment at step of time t.
- Finally, recovered devices lose their temporal immunity and turn back to be susceptible at recovery rate $0 \leq \epsilon \leq 1$.

Taking into account all these assumptions, the following SODE determines the dynamics of the system:

$$S'(t) = \epsilon R(t) - aS(t)[I(t)+C(t)] - vS(t), \tag{3}$$
$$C'(t) = a(1-\delta)S(t)[I(t)+C(t)] - b_C C(t), \tag{4}$$
$$I'(t) = a\delta S(t)[I(t)+C(t)] - b_I I(t) - \gamma I(t), \tag{5}$$
$$A'(t) = \gamma I(t) - \beta A(t), \tag{6}$$
$$\begin{aligned}R'(t) &= b_C C(t) + b_I I(t) + \beta A(t) + vS(t) \\ &\quad -\epsilon R(t). \end{aligned} \tag{7}$$

Note that from (1), this SODE can be rewritten as follows:

$$\begin{aligned}S'(t) &= -aS(t)[I(t)+C(t)] - vS(t) \\ &\quad +\epsilon(N - S(t) - C(t) - I(t) - A(t)), \end{aligned} \tag{8}$$
$$C'(t) = a(1-\delta)S(t)[I(t)+C(t)] - b_C C(t), \tag{9}$$
$$\begin{aligned}I'(t) &= a\delta S(t)[I(t)+C(t)] - b_I I(t) \\ &\quad -\gamma I(t), \end{aligned} \tag{10}$$
$$A'(t) = \gamma I(t) - \beta A(t). \tag{11}$$

IV. QUALITATIVE ANALYSIS

A. STEADY STATES

As is well known, the steady states of the SODE (8)-(11) are the solutions of the following system of non-linear equations:

$$\begin{aligned}0 &= -aS(t)[I(t)+C(t)] - vS(t) \\ &\quad +\epsilon[(N - S(t) - C(t) - I(t) - Q(t)], \end{aligned} \tag{12}$$
$$0 = a(1-\delta)S(t)[I(t)+C(t)] - b_C C(t), \tag{13}$$
$$0 = a\delta S(t)[I(t)+C(t)] - b_I I(t) - \gamma I(t), \tag{14}$$
$$0 = \gamma I(t) - \beta Q(t). \tag{15}$$

A simple computation shows that this system has two solutions: the disease-free equilibrium point given by

$$E_0 = (S_0, C_0, I_0, Q_0) = \left(\frac{\epsilon N}{v+\epsilon}, 0, 0, 0\right), \tag{16}$$

and the endemic equilibrium point:

$$E^* = \left(S^*, C^*, I^*, Q^*\right), \tag{17}$$

where

$$S^* = \frac{N\epsilon(v+\epsilon)}{B}, \tag{18}$$
$$C^* = -\frac{\beta(\delta-1)N(B-1)\epsilon(b_I+\gamma)}{AB} \tag{19}$$
$$I^* = \frac{\beta b_C \delta N(B-1)\epsilon}{AB}, \tag{20}$$
$$Q^* = \frac{b_C \gamma \delta N(B-1)\epsilon}{AB}. \tag{21}$$

with

$$\begin{aligned}A &= b_c\beta(b_I+\gamma) - \beta(b_I+\gamma)(-1+\delta)\epsilon \\ &\quad +b_C(\beta+\gamma)\delta\epsilon > 0, \end{aligned} \tag{22}$$
$$B = \frac{aN\epsilon[b_I+\gamma+b_C\delta-(b_I+\gamma)\delta]}{b_C(b_I+\gamma)(v+\epsilon)}. \tag{23}$$

Note that the endemic solution only exists if $B > 1$ (moreover, $AB \neq 0$).

B. BASIC REPRODUCTIVE NUMBER

Applying the next-generation method [20], we obtain that the basic reproductive number associated to the proposed model is the spectral radius of the following matrix (next-generation

matrix):

$$G = \begin{pmatrix} \frac{aN\,(1-\delta)\,\epsilon b_C}{v+\epsilon} & \frac{aN\,(1-\delta)\,\epsilon\,(b_I+\gamma)}{v+\epsilon} & 0 \\ \frac{aN\delta\epsilon b_C}{v+\epsilon} & \frac{aN\delta\epsilon\,(b_I+\gamma)}{v+\epsilon} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (24)$$

that is:

$$R_0 = \frac{aN\,(b_I+\gamma+b_C\delta-(b_I+\gamma)\delta)\,\epsilon}{b_C(b_I+\gamma)\,(v+\epsilon)}. \quad (25)$$

Note that the condition for the existence of the endemic equilibrium point is, precisely, that $R_0 = B > 1$.

C. STABILITY OF THE EQUILIBRIUM POINTS

Considering the qualitative theory of ordinary differential equations, a rather long calculus leads to the the following results related to the local stability of the equilibrium points:

Theorem 1: The disease-free equilibrium point E_0 is locally and globally asymptotically stable if $R_0 < 1$.

Theorem 2: The endemic equilibrium point E^* is locally asymptotically stable if $R_0 > 1$.

Theorem 3: the endemic equilibrium point E^* is globally asymptotically stable if $R_0 > 1$ under the following assumptions:

$$-(1-\delta)a\frac{c^2}{N} + aN\delta - v - 2ac - \epsilon < 0, \quad (26)$$

$$-a(1-\delta)\frac{c^2}{N} + \delta aN - b_I - 2\gamma + 2aN\tilde{\delta} < 0, \quad (27)$$

where $\tilde{\delta} = \max\{\delta, (1-\delta)\}$ and c is the persistence constant.

V. ILLUSTRATIVE SIMULATIONS OF THE SCIRAS MODEL

In what follows two simulations to illustrate the different behaviors of the system are shown. It is assumed that $N = 100$ with $S(0) = 95$ and $I(0) = 5$ and the evolution of each compartment is computed during the first week after the start of the outbreak (168 hours). In the first one (see Fig. 2) the disease-free equilibrium point is reached. In this case, the numerical values of the epidemiological coefficients are the following:

$$\begin{aligned} a &= 5\times10^{-4}, \quad \delta = 0.9, \\ v &= 0.05, \quad \gamma = 5\times10^{-3}, \\ b_C &= 4\times10^{-3}, \quad b_I = 0.03, \\ \beta &= 5\times10^{-6}, \quad \epsilon = 5.5\times10^{-3}. \end{aligned} \quad (28)$$

As a consequence $R_0 \approx 0.2513 < 1$, and the disease-free equilibrium point is

$$E_0 \approx (10.01, 0, 0, 0, 90.99)\,. \quad (29)$$

On the other hand, if the value of the transmission coefficient is changed and $a = 2\times10^{-3}$ is considered, then the system tends to the endemic equilibrium point (see Fig. 3):

$$E^* \approx (9.86, 0.00049, 0.00051, 0.51, 89.63)\,, \quad (30)$$

where $R_0 \approx 1.005 > 1$.

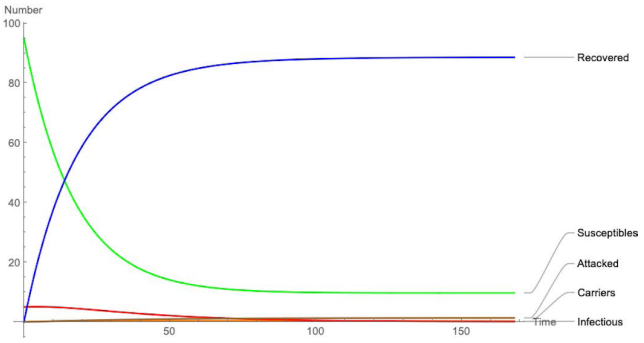


FIGURE 2. Disease-free behavior of the model.

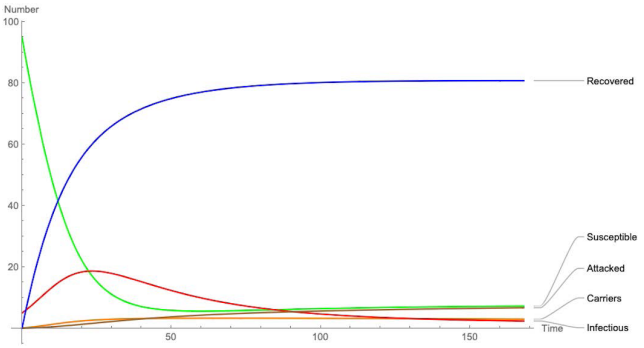


FIGURE 3. Endemic behavior of the model.

VI. DESIGN OF EFFICIENT CONTROL MEASURES

There are mainly three threshold parameters related to mathematical models to simulate malware spreading: the basic reproductive number R_0, the replacement number R, and the contact number σ (see [19]). Roughly speaking, the basic reproductive number R_0 can be defined as the average number of secondary infections caused by an only one infectious device in an entire susceptible population during its entire infectious period. The replacement number R stands for the average number of secondary infections caused by an infectious device during its entire infectious period. Finally, the contact number σ has been defined as the average number of adequate contacts of an infectious device during its entire infectious period.

The most important is the basic reproductive number (also known as the basic reproduction ratio or the basic reproductive rate) since it plays a central role in the study of the behavior of the solutions of the system [20], [21] (as is illustrated in Sect. IV).

Consequently the basic reproductive number plays a very important role in the design of efficient control measures. Specifically, if $R_0 < 1$ the malware outbreak dies out and, consequently, the reduction of the numeric value of the R_0 will be the main goal of all security countermeasures.

In what follows, we will analyze the basic reproductive number in order to provide explicit expressions for the control of the malware epidemic. Specifically, in the next two subsections we will describe the most important control measures that consider the modification of one or two epidemiological coefficients.

A. ONE-PARAMETER ANALYSIS

For the sake of simplicity assume that $\alpha = b_I + \gamma$. Then, from the expression of the basic reproductive number (25) we obtain:

$$R_0 = \frac{aN\epsilon[(1-\delta)\alpha + \delta b_C]}{b_C\alpha(\nu+\epsilon)}. \tag{31}$$

As a consequence the basic reproductive number depends on 7 coefficients: a (the transmission rate), N (the total number of devices), ϵ (the recovery rate), δ (the fraction of targeted devices), $\alpha = b_I + \gamma$, b_C (the recovery rate from carrier), and ν (the rate at which susceptible devices acquire temporal immunity).

If it is supposed that six of these seven coefficients remains constant over time, then R_0 can be considered as a function of only one variable x (the remaining non-constant coefficient). As a consequence, the study of $\frac{\partial R_0}{\partial x}$ will give us information about the monotony of the function $R_0(x)$ and we can draw conclusions about the behavior of the basic reproductive number when only one coefficient varies.

Consequently, and supposing $0 < a, \epsilon, \alpha, b_C, \nu, N, \delta \leq 1$, the following holds:

(1) If the transmission rate a varies, $R_0 = R_0(a)$, then:

$$\frac{\partial R_0}{\partial a} = \frac{N\epsilon[(1-\delta)\alpha + \delta b_C]}{b_C\alpha(\nu+\epsilon)} > 0. \tag{32}$$

As a consequence R_0 decreases as a decreases.

(2) If the total number of devices N is non-constant, then:

$$\frac{\partial R_0}{\partial N} = \frac{a\epsilon[(1-\delta)\alpha + \delta b_C]}{b_C\alpha(\nu+\epsilon)} > 0, \tag{33}$$

and, as the previous case, R_0 decreases when N decreases.

(3) Suppose that $R_0 = R_0(\delta)$, then:

$$\frac{\partial R_0}{\partial \delta} = \frac{aN\epsilon(b_C - \alpha)}{b_C\alpha(\nu+\epsilon)}. \tag{34}$$

If we assume that $b_C < b_I$ (which is a realistic assumption) then $b_C - \alpha < 0$ and, consequently, $\frac{\partial R_0}{\partial \delta} < 0$. Thus R_0 decreases when δ increases and $b_C < b_I$.

(4) If the coefficient ν is variable, then $R_0 = R_0(\nu)$ and simple calculus shows that:

$$\frac{\partial R_0}{\partial \nu} = -\frac{aN\epsilon[(1-\delta)\alpha + \delta b_C]}{b_C\alpha(\nu+\epsilon)^2} < 0. \tag{35}$$

As a consequence if ν increases then R_0 decreases.

(5) Now, suppose that the non-constant coefficient is α, then

$$\frac{\partial R_0}{\partial \alpha} = -\frac{aN\epsilon\delta}{\alpha^2(\nu+\epsilon)} < 0. \tag{36}$$

Then R_0 decreases when $\alpha = b_I + \gamma$ increases.

(6) If $R_0 = R_0(b_C)$ then:

$$\frac{\partial R_0}{\partial b_C} = -\frac{aN\epsilon(1-\delta)}{b_C^2(\nu+\epsilon)} < 0. \tag{37}$$

Consequently R_0 decreases when b_C increases.

(7) Finally, set $R_0 = R_0(\epsilon)$. Then:

$$\frac{\partial R_0}{\partial \epsilon} = \frac{aN\alpha b_C\nu[(1-\delta)\alpha + b_C\delta]}{b_C^2\alpha^2(\nu+\epsilon)^2} > 0, \tag{38}$$

and R_0 decreases when ϵ decreases.

Taking into account all these results, we can derive that when only one coefficient varies the basic reproductive number decreases when:

- The parameters a, N, δ (if $b_C > b_I$), ϵ decrease.
- The parameters δ (if $b_C < b_I$), $\nu, \alpha = b_I + \gamma$ increase.

Consequently the following security measures reduce the impact of the malware epidemic:

- Decreasing the transmission rate, total number of devices (particularly, the number of devices endowed with the targeted operative system when the recovery rate of carriers is greater than the recovery rate of infectious), or the rate of lose of immunity.
- Increasing the infectious recovery rate and/or the vaccination rate.

B. TWO-PARAMETER ANALYSIS

Now, we will define efficient security strategies that imply the jointly use of two coefficients. In this case the basic reproductive number can be considered as a function of two variables x and y, $R_0(x, y)$, which stand for the epidemiological coefficients that can vary; the other five parameters remain constant.

Suppose that a particular step of time t_0, the values of the variable coefficients are x_0 and y_0 respectively, such that $R_0(x_0, y_0) > 1$ (that is, the system is in the endemic region -the number of infectious devices is increasing-). Set $p_0 = (x_0, y_0)$ the initial point in the xy-plane such that it is placed in the endemic region defined by $R_0(x, y) - 1 > 0$. As a consequence, the challenge is to find the fastest way to get the threshold curve $R_0(x, y) - 1 = 0$ from the initial point $p_0 = (x_0, y_0)$. Taking into account the expression of the basic reproductive number (31), the threshold curve $R_0(x, y) = 1$ can be described by different rational expressions of the form

$$y = r_0(x) = \frac{c_1x + c_2}{c_3x + c_4} \tag{39}$$

where $0 < c_1, c_2, c_3, c_4 \leq 1$ (see Table 1).

The most efficient strategy to get $R_0(x, y) - 1 = 0$ is given by the trajectory defined by the segment $\overline{p_0p_1}$ where $p_1 = (x_1, y_1)$ is the nearest point to p_0 such that $R_0(x_1, y_1) = 1$ (see Fig. 4). Note that the parametric equations of this segment are the following:

$$x = \lambda x_1 + (1-\lambda)x_0, \tag{40}$$

$$y = \lambda r_0(x_1) + (1-\lambda)y_0, \quad 0 \leq \lambda \leq 1, \tag{41}$$

where $x = x_1$ is the minimum of the function:

$$d(x) = \sqrt{(x-x_0)^2 + (r_0(x) - y_0)^2}. \tag{42}$$

TABLE 1. Rational expression of $R_0(x, y) = 1$.

(x, y)	$R_0(x, y) = 1$
(a, N)	$y = r_0(x) = \dfrac{1}{c_3 x}$
$(a, \epsilon), (N, \epsilon)$	$y = r_0(x) = \dfrac{1}{c_3 x + c_4}$
$(a, \delta), (N, \delta)$	$y = r_0(x) = \dfrac{c_1 x + 1}{c_3 x}$
$(a, \alpha), (N, \alpha), (N, b_C), (\delta, \alpha), (a, b_C)$	$y = r_0(x) = \dfrac{c_1 x}{c_3 x + 1}$
$(a, v), (N, v)$	$y = r_0(x) = c_1 x + c_2$
$(\epsilon, \delta), (\alpha, v), (b_C, v)$	$y = r_0(x) = \dfrac{c_1 x + c_2}{c_3 x}$
$(\epsilon, \alpha), (\epsilon, b_C), (\alpha, b_C)$	$y = r_0(x) = \dfrac{c_1 x}{c_3 x + c_4}$
(ϵ, v)	$y = r_0(x) = c_1 x$
(δ, b_C)	$y = r_0(x) = \dfrac{c_1 x + c_2}{c_3 x + c_4}$
(δ, v)	$y = r_0(x) = c_1 x + c_2$

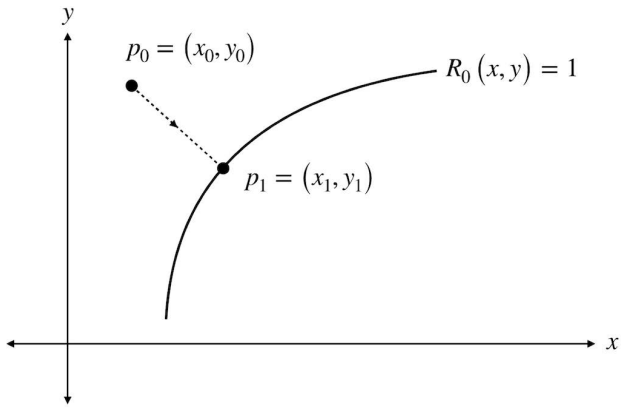


FIGURE 4. Illustrative representation of the fastest way to get the threshold curve.

Consequently the optimal strategy is to increase (resp. decrease) x and y from x_0 and y_0 to x_1 and y_1 respectively, and following Eqs. (40)-(41) (see Fig. 5). That is, the parameter λ must be increased to 1 and the non-constant epidemiological coefficients x and y must be computed according to Eqs. (40)-(41).

As an illustrative example of this procedure assume that the initial values of the system are the following:

$$\begin{aligned} a &= 2 \times 10^{-2}, \quad \delta = 0.9, \\ v &= 0.05, \quad \gamma = 0.5, \\ b_C &= 0.01, \quad b_I = 0.05, \\ \epsilon &= 5.5 \times 10^{-3}, \end{aligned} \tag{43}$$

then $R_0 \approx 2.30631 > 1$. Suppose that the non-constant coefficients are $x = \alpha$ and $y = b_C$, then the explicit expression of the threshold curve is:

$$y = r_0(x) = \frac{0.0011x}{0.055x - 0.0099}. \tag{44}$$

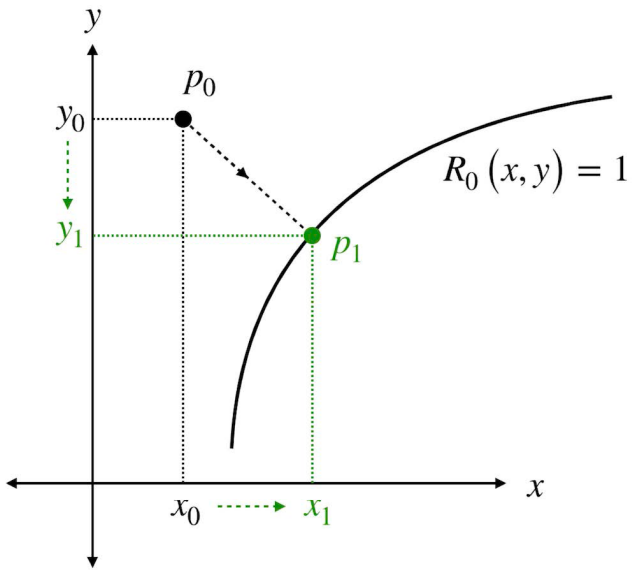


FIGURE 5. Optimal variation of non-constant epidemiological coefficients to control the malware outbreak.

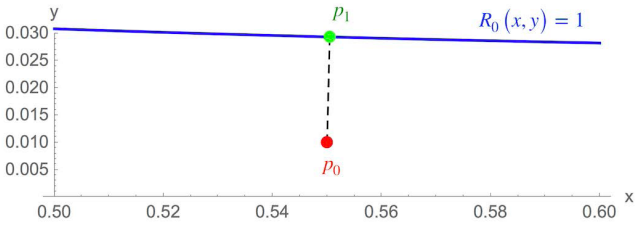


FIGURE 6. Illustrative example when $x = \alpha$ and $y = b_C$.

The initial point is $p_0 = (x_0, y_0) = (0.55, 0.01)$ and a simple calculus shows that $p_1 \approx (0.55, 0.029)$. As a consequence the optimal strategy to reduce the basic reproductive number modifying $x = \alpha$ and $y = b_C$ is increasing the parameter λ such that (see Fig. 6):

$$x = 0.00049\lambda + 0.55, \tag{45}$$

$$y = 0.019\lambda + 0.01, \quad 0 \le \lambda \le 1. \tag{46}$$

VII. CONCLUSION

In this work a novel mathematical model for simulating the behaviour of an advanced malware outbreak has been introduced. This is a compartmental, deterministic and global model whose dynamics is based on a system of ordinary differential equations. As a consequence the qualitative theory of differential equations can be applied to study the behaviour of the solutions. In this sense, two types of steady states can be reached: the disease-free steady state where malware disappears from the network, and the endemic steady state where there will be infectious devices at every step of time.

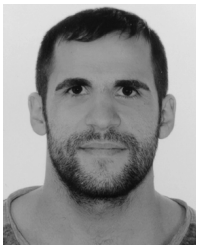
The basic reproductive number is computed and it is shown that this threshold parameter determines the behaviour of the system depending on whether its numerical value is greater

than or less than 1. An analysis of this coefficient has been done determining the most efficient control measures when one or two epidemiological coefficients are varied.

Future work aimed at designing individual-based models to simulate advanced malware behavior considering the individual characteristics of the devices. Moreover, different network topologies must be analyzed over both stochastic and deterministic local transition rules. In this case the paradigm of multi agent systems or computational intelligence must be used to design such models.

REFERENCES

[1] A. K. Sood and R. J. Enbody, "Targeted Cyberattacks: A superset of advanced persistent threats," *IEEE Security Privacy*, vol. 11, no. 1, pp. 54–61, Jan. 2013.
[2] S. Singh, P. K. Sharma, S. Y. Moon, D. Moon, and J. H. Park, "A comprehensive study on APT attacks and countermeasures for future networks and communications: Challenges and solutions," *J. Supercomput.*, vol. 75, pp. 4543–4574, Aug. 2019.
[3] C. Tankard, "Advanced persistent threats and how to monitor and deter them," *Netw. Secur.*, vol. 2011, no. 8, pp. 16–19, 2011.
[4] A. Ahmad, J. Webb, K. C. Desouza, and J. Boorman, "Strategically-motivated advanced persistent threat: Definition, process, tactics and a disinformation model of counterattack," *Comput. Secur.*, vol. 86, pp. 402–418, Sep. 2019.
[5] P. Chen, L. Desmet, and C. Huygens, "A study on advanced persistent threats," in *Communications and Multimedia Security* (Lecture Notes in Computer Science), vol. 8735, B. De Decker, and A. Zúquete, Eds., Berlin, Germany: Springer, 2014, pp. 63–72.
[6] J.-Y. Kim, S.-J. Bu, and S.-B. Cho, "Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders," *Inf. Sci.*, vols. 460–461, pp. 83–102, Sep. 2018.
[7] Y. Li, W. Dai, J. Bai, X. Gan, J. Wang, and X. Wang, "An intelligence-driven security-aware defense mechanism for advanced persistent threats," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 646–661, Mar. 2019.
[8] K. Lv, Y. Chen, and C. Hu, "Dynamic defense strategy against advanced persistent threat under heterogeneous networks," *Inf. Fusion*, vol. 49, pp. 216–225, Sep. 2019.
[9] I. Ghafir, M. Hammoudeh, V. Prenosil, L. Han, R. Hegarty, K. Rabie, and F. J. Aparicio-Navarro, "Detection of advanced persistent threat using machine-learning correlation analysis," *Future Gener. Comput. Syst.*, vol. 89, pp. 349–359, Dec. 2018.
[10] G. Tecuci, D. Marcu, S. Meckl, and M. Boicu, "Evidence-based detection of advanced persistent threats," *Comput. Sci. Eng.*, vol. 20, no. 6, pp. 54–65, Nov./Dec. 2018.
[11] S. S. Chakkaravarthy, V. Vaidehi, and P. Rajesh, "Hybrid analysis technique to detect advanced persistent threats," *Int. J. Intell. Inf. Technol.*, vol. 14, no. 2, pp. 59–76, Apr. 2018.
[12] A. M. del Rey, "Mathematical modeling of the propagation of malware: A review," *Secur. Commun. Netw.*, vol. 8, no. 15, pp. 2561–2579, Oct. 2015.
[13] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 925–941, 2nd Quart., 2014.
[14] V. Karyotis and M. H. R. Khouzani, *Malware Diffusion Models for Modern Complex Networks: Theory and Applications*. Cambridge, MA, USA: Morgan Kaufmann, 2015.
[15] L. Tidy, S. Woodhead, and J. Wetherall, "Simulation of zero-day worm epidemiology in the dynamic, heterogeneous Internet," *J. Defense Model. Simul. Appl. Methodol. Technol.*, vol. 12, no. 2, pp. 123–138, Oct. 2015.
[16] H. Tran, E. Campos-Nanez, P. Fomin, and J. Wasek, "Cyber resilience recovery model to combat zero-day malware attacks," *Comput. Secur.*, vol. 61, pp. 19–31, Aug. 2016.
[17] C. Zhang and J. Xiao, "Stability analysis of an advanced persistent distributed denial-of-service attack dynamical model," *Secur. Commun. Netw.*, vol. 2018, May 2018, Art. no. 5353060.
[18] C. Zhang, J. Peng, and J. Xiao, "An advanced persistent distributed denial-of-service attacked dynamical model on networks," *Discrete Dyn. Nature Soc.*, vol. 2019, Feb. 2019, Art. no. 2051489.
[19] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, 2000.
[20] O. Diekmann, H. Heesterbeek, and T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton, NJ, USA: Princeton Univ. Press, 2013.
[21] P. van den Driessche and J. Watmough, "Further notes on the basic reproduction number," in *Mathematical Epidemiology* (Lecture Notes in Mathematics), F. Brauer, P. van den Driessche, and J. Wu, Eds., Berlin, Germany: Springer-Verlag, 2008, pp. 159–178.



J. D. HERNÁNDEZ GUILLÉN was born in Tenebrón, Salamanca, Spain. He received the B.S. and M.S. degrees in mathematics from the University of Salamanca, Spain, in 2015. He is currently pursuing the Ph.D. degree in computer engineering and mathematical modeling.

He has published five articles in journals indexed in JCR-WoS. His research interests include the design and computational implementation of mathematical models to simulate malware propagation.

Dr. Guillén has received a prize for being the student with higher grades in mathematics degree.

A. MARTÍN DEL REY (M'19) was born in Salamanca, Spain, in 1972. He received the B.S. and M.S. degrees in mathematics from the University of Salamanca, in 1996, and the Ph.D. degree in mathematics from UNED/CSIC, in 2000.

Since 2008, he has been an Assistant Professor with the Department of Applied Mathematics, Universidad de Salamanca, Spain. He is the author of more than 50 articles published in journals indexed in JCR-WoS, and 34 conference proceedings indexed in CPCI-S (WoS). His research interests include mathematical models for security and cyber-security, cryptography, complex network analysis, and cellular automata. He is an Academic Editor of the journal *Security and Communication Networks*.

ROBERTO CASADO-VARA received the degree in mathematics from the University of Salamanca, the master's degree in big data and visual analytics from the International University of la Rioja, and the Ph.D. degree in computer science from USAL, in 2019. He has been with Viewnext as a Data Scientist and Powercenter Consultant for important clients in the pharmaceutical and public administration sectors. He is currently researching in computer engineering as a member of the BISITE Research Group. As a researcher, his interests are focused on deep learning, advanced mathematical models for intelligent robust and non-linear control and monitoring, blockchain and knowledge discovery data, as well as other fields.