mathematics

MDPI

Article

Propagation of the Malware Used in APTs Based on Dynamic Bayesian Networks

Jose D. Hernandez Guillen [1,†], Angel Martin del Rey [2,*,†] and Roberto Casado-Vara [3,†]

[1] Department of Applied Mathematics, University of Salamanca, 37008 Salamanca, Spain; diaman@usal.es
[2] Institute of Fundamental Physics and Mathematics, Department of Applied Mathematics, University of Salamanca, 37008 Salamanca, Spain
[3] Department of Mathematics and Computation, University of Burgos, 09007 Burgos, Spain; rccasado@ubu.es
* Correspondence: delrey@usal.es
† These authors contributed equally to this work.

Abstract: Malware is becoming more and more sophisticated these days. Currently, the aim of some special specimens of malware is not to infect the largest number of devices as possible, but to reach a set of concrete devices (target devices). This type of malware is usually employed in association with advanced persistent threat (APT) campaigns. Although the great majority of scientific studies are devoted to the design of efficient algorithms to detect this kind of threat, the knowledge about its propagation is also interesting. In this article, a new stochastic computational model to simulate its propagation is proposed based on Bayesian networks. This model considers two characteristics of the devices: having efficient countermeasures, and the number of infectious devices in the neighborhood. Moreover, it takes into account four states: susceptible devices, damaged devices, infectious devices and recovered devices. In this way, the dynamic of the model is $SIDR$ (susceptible–infectious–damaged–recovered). Contrary to what happens with global models, the proposed model takes into account both the individual characteristics of devices and the contact topology. Furthermore, the dynamics is governed by means of a (practically) unexplored technique in this field: Bayesian networks.

Keywords: malware propagation; epidemic model; Bayesian network; advanced persistent threat; stochastic model

Citation: Hernandez Guillen, J.D.; Martin del Rey, A.; Casado-Vara, R. Propagation of the Malware Used in APTs Based on Dynamic Bayesian Networks. *Mathematics* **2021**, *9*, 3097. https://doi.org/10.3390/math9233097

Academic Editor: Davide Valenti

Received: 22 October 2021
Accepted: 28 November 2021
Published: 30 November 2021

1. Introduction

Nowadays, security threats to computer systems pose a huge risk to our society. Especially dangerous are those sophisticated cyber attacks called advanced persistent threats since their basic targets are critical infrastructures and other systems that control essential services, such as transport, communications, etc. [1,2].

The National Institute of Standards and Technology (NIST) of the United States defines an APT as follows [3]: "an adversary with sophisticated levels of expertise and significant resources, allowing it through the use of multiple different attack vectors (e.g., cyber, physical, and deception), to generate opportunities to achieve its objectives which are typically to establish and extend its presence within the information technology infrastructure of organizations for purposes of continually exfiltrating information and/or to undermine or impede critical aspects of a mission, program, or organization, or place itself in a position to do so in the future; moreover, the advanced persistent threat pursues its objectives repeatedly over an extended period of time, adapting to a defender's efforts to resist it, and with determination to maintain the level of interaction needed to execute its objectives". As a consequence, an APT has an specific target as private organizations or governments and/or public agencies. Advanced and sophisticated techniques and highly organized methods are employed to achieve its goals. Moreover, an APT forms a long-term

attack campaign for months or years, and consequently, it has a high ability to remain undetected.

One of the most important techniques used in an APT is constituted by advanced malware that exploits zero-day vulnerabilities. These specimens of malware are highly sophisticated and exhibit the main characteristics of a proper APT. In Table 1, it is summarized the differences between a usual type of malware and zero-day malware used in an APT attack (see [4–7]).

Table 1. Differences between usual malware and malware used in an APT attack.

	Typical Attacks	**Attacks Used in APT**
Cybercriminals	Mainly one person	A group of qualified people
Victim	Any device	Specific institutions and governmental organizations
Aim	Obtaining money or being known	Gain an advantage over their competitors
Characteristics	Fast propagation, and one attempt.	Slow propagation, several attempts and adaptation against countermeasures

The number of APT attacks has increased in recent years [4]. Although the great majority of studies in the scientific literature are devoted to the design of an implementation of efficient algorithms to detect this type of malware [8–10], it is also very important to design and analyze computational models that simulate the propagation of this type of malware, and this is precisely the main goal of this work.

As is well known, there are two types of models that study malware propagation: deterministic and stochastic models. Deterministic models are usually global (that is, they suppose that all devices have the same characteristics and the contact topology is homogeneous) and, consequently, they are based on—deterministic—ordinary differential equations [11]). On the other hand, stochastic models can be also global (and based on stochastic differential equations), although the great majority follows the individual paradigm [12,13] and, consequently, takes into account particular characteristics of devices. All of them are compartmental models where the total population of devices is classified into different classes or compartments (depending on the epidemiological state). In this sense, several different compartments can be considered in a specific model: susceptible devices S, weak susceptible devices W, infectious devices I, carrier devices C, recovered devices R, vaccinated devices V, attacked devices A, damaged devices D, etc. In this way, considering the involved compartments and the dynamics between them, the epidemiological models are classified according to their dynamic: $SCIRAS$ model [14], $SCIRS$ model [15], $SIRA$ model [16], $SEIRS-V$ model [17], $WSIS$ model [18], etc. The model introduced in this work is a $SIDR$ model (susceptible–infectious–damaged–recovered). This considers that susceptible devices can be infected when the advanced malware reaches them, infectious devices can be damaged if they are considered targets by malware, and finally, both infectious devices and damaged devices can be recovered.

Very few models have been proposed in the scientific literature to simulate the propagation of the advanced malware used in APTs. In [14], the authors propose a SCIRAS global and deterministic model based on ordinary differential equations. This is a theoretical proposal where the proposed model can simulate the general evolution of its five compartments (susceptible devices, carrier devices, infectious devices, attacked devices and recovered devices). Moreover, the article analyzed the basic reproductive number considering several parameters, and a qualitative study of the system (computing the equilibrium points and analyzing the stability) is also introduced. In [19], a stochastic model is introduced that simulates advanced malware as well. This model considers different Erdös–Rényi networks as contact topologies in order to study the evolution of infectious

and attacked devices. The centrality measures of the first infected node are also considered to show its impact in the propagation. Other different models to simulate the behavior of advanced malware also appear in some works that study the detection of this type of malware [20].

Our work is focused on article [21]. According to this, the malware used in an APT attack has a set of target devices and its propagation is stealthy and slow. The following three characteristics are considered:

1. The malware has a set of target devices. Then, the main objective of the malware is to infect (and attack) these devices.
2. The propagation of this malware has to be stealthy. Then, the number of infectious devices is smaller. Moreover, this type of malware can obtain information of the security of the system and knows whether a device has efficient countermeasures. This way, the malware tries not to be detected by this type of software. As a consequence, the probability to infect devices with efficient countermeasures is smaller.
3. The propagation of this type of malware is slow. This means that the increase in infectious devices is smaller during the infection period.

In our work, we considered that the dynamics of advanced malware propagation is governed by a dynamic Bayesian network. Consequently, this is a stochastic model that considers both individual characteristics (having efficient countermeasures) and topology features (the particular contact structure of each node/device of the network). The epidemiological coefficients involved can be calculated through different methods, such as parameter learning or structural learning. Then, if we know the propagation of the malware and the individual characteristics of the devices, we can obtain the characteristics of the model (the parameters). This permits to compare the properties of different types of advanced malware. The great majority of proposed models to simulate malware propagation are based on (deterministic) ordinary differential equations and, consequently, they cannot consider in an efficient way neither the individual characteristics of the nodes/devices nor the contact topology. In addition, they cannot differentiate between a target device and a non-target device. Usually, individual-based models are based on (probabilistic) cellular automata or Markov processes. Nevertheless, a small number of models that consider Bayesian networks to simulate malware propagation have appeared in the scientific literature. The aim of this work is to analyze the use of this latter mentioned technique (Bayesian networks) since this notion is clear, well known, and its parameters can be organized easily if we want to consider several characteristics.

A Bayesian network is a probabilistic graphical model so that the nodes depict the random variables and the links represent their conditional dependencies. Moreover, the associated graph is directed and acyclic [22]. Other authors used dynamic Bayesian networks in different applications: to identify faults [23], to predict information diffusion probabilities in social networks [24], etc. Additionally, in epidemiology, Bayesian networks have been used in several applications: to represent a virus infection model [25], to show the structure of cloud components [26], to study possible disease progression mechanisms [27], to predict an epidemic curve [28], etc.

The rest of the paper has the following structure: In Section 2, an introduction of Bayesian networks is presented. In Section 3, the structure of the model is shown. In Section 4, an illustrative example of the model is shown. Finally, in Section 5, the conclusions and future work are exposed.

2. Mathematical Preliminaries of the Model

This section presents a short summary of the mathematical concepts to understand Bayesian networks. A Bayesian network is a directed acyclic graph $G = (V, E)$. $V = \{X_1, X_2, \ldots, X_n\}$ is the set of nodes that represents n random variables, and $E \subset V \times V$ is the set of links that represents the conditional dependencies among the nodes ($|E| = m$). If $(U, W) \in E$, the node U is called the parent of W, $U \in \text{pa}(W)$. Each node follows a conditional probability distribution (CPD) according to the Bayesian network. For

example, if $W \in V$ has several parents $U_1, \ldots, U_n \in V$, the CPD associated to W is $P(W = w_0 | U_1 = u_1, \ldots, U_n = u_n)$.

The probabilistic graphical model used in this work evolves with time, and consequently, the model depends on t. In this way, the vector $X(t) = (X_1(t), \ldots, X_n(t))$ is the set of random variables at time t. Furthermore, the following two conditions are considered:

- Markov's assumption: the variables at the next step of time $t + \Delta t$ only depend on the variables at time t.

$$X(t + \Delta t) \perp \{X(t - k\Delta t), \ldots, X(t - \Delta t)\} | X(t), \quad k \in \mathbb{Z}^+. \tag{1}$$

- Time-invariant: the CPDs of the random variables do not change through time.

$$P(X(t + \Delta t) = A | X(t) = B) = P(X(t + (k + 1)\Delta t) = A | X(t + k\Delta t) = B), \tag{2}$$

for all t and $k \in \mathbb{Z}^+$, where $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ are the vectors of values that the random variables $X = (X_1, \ldots, X_n)$ can take.

Then, the conditional probability distribution can be calculated as follows:

$$P(X(t + \Delta t) = A | X(t) = B) = \prod_{i=1}^{n} P\big(X_i(t + \Delta t) = a_i | \text{pa}(X_i(t + \Delta t)) = \bar{b}_i\big) \tag{3}$$

where $\bar{b}_i$ is the values of the parents of $X_i(t + \Delta t)$ taken from the values B.

3. Structure of the Model

The model proposed in this work involves two different networks:

1. The device network is formed by $r \in \mathbb{N}$ devices. This network takes into account the interactions among the devices—the main feature that influences the propagation process. Then, this type of malware propagates through this network.
2. The Bayesian network. The Bayesian network explains how the characteristics (related to the propagation process) of each device change over time.

It was considered that a device i is endowed with four characteristics at time t: epidemiological state $X_1(i, t)$, target consideration $X_2(i, t)$, efficient security countermeasures $X_3(i, t)$, and the number of infectious nodes in contact with the node i, $X_4(i, t)$. The temporal model based on Bayesian networks is illustrated in Figure 1.

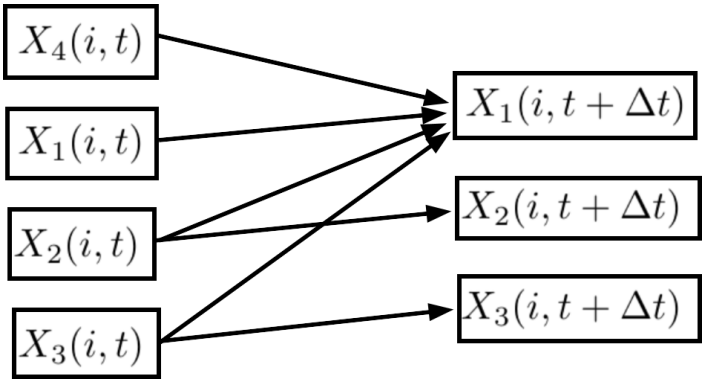


Figure 1. Bayesian network associated to the proposed model.

Then, the evolution of the epidemiological state of a node depends on the four characteristics:

- As with most of the models, the state of the device in the previous step influences the new state.
- The target consideration decides if a node can be damaged since an APT only attacks the target nodes.

- Having efficient security countermeasures leads to more difficult infection and faster recovery.
- The infection process depends on the number of infectious nodes in contact with the node i [19]. If the number is very big, the infection is more likely.

Moreover, nodes can stop having efficient security countermeasures or they can start having them. This happens because one person can install or uninstall efficient anti-malware software in their device. Then, $X_3(i,t)$ can change over time. Therefore, it depends on this variable in the previous epoch.

3.1. Characteristics of Each Node/Device

The node i is endowed with the following characteristics at each step of time t:

- Epidemiological state $X_1(i,t)$. A device can have one of the following four states in each epoch: susceptible, infectious, damaged or recovered. Susceptible devices (denoted by x_{11} = "susceptible") are those devices that are free of malware and can be infected. Infectious devices (denoted by x_{12} = "infectious") are devices that are reached by the malware but they do not suffer its malicious activity. Moreover, these devices can infect other susceptible devices. Damaged devices (denoted by x_{13} = "damaged") are devices that are infected and can suffer malicious activity. These devices can infect other susceptible devices too. According to ATPs, only the targets can be damaged. Finally, recovered devices (denoted by x_{14} = "recovered") are devices that no longer have the malware. In this way, this constitutes a compartmental model whose dynamics consist of the following: if the ATP malware reaches a susceptible device, it becomes infected. An infected device can turn into a recovered one if the malware is removed, or it can become damaged if this is a target device and the malware manages to activate. Finally, a damaged device turns into a recovered device when the malware is removed. Then, the dynamics of this model is $SIDR$ (susceptible–infectious–damaged–recovered). The relations of the different states are represented in the flow diagram shown in Figure 2.

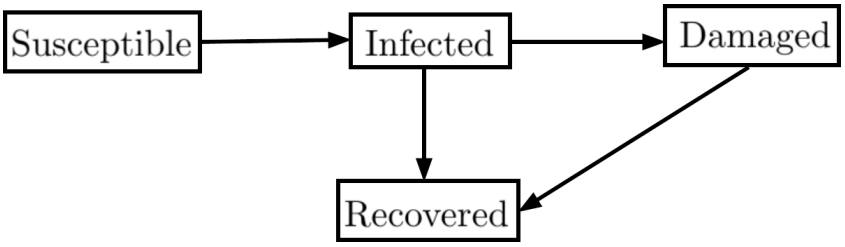


Figure 2. Flow diagram representing the dynamics of the $SIDR$ model.

- Target consideration $X_2(i,t)$. We have considered two types of devices: devices that the ATP wants to damage (the targets), and devices that are not of interest for the malware. Thus, the random variable $X_2(i,t)$ can adopt two values: the device is a target (denoted by x_{21} = "yes") and the device is not a target (denoted by x_{22} = "no").
- Efficient security countermeasures $X_3(i,t)$. We have taken into account two kinds of devices: devices that are endowed with efficient security countermeasures and devices that do not have security efficient countermeasures. On the one hand, the devices with efficient countermeasures (denoted by x_{31} = "yes") can recover easily and can be infected with more difficulty. On the other hand, the devices without efficient countermeasures (denoted by x_{32} = "no") can be easily infected and can recover with more difficulty.
- Number of infectious nodes in contact with a node/device $X_4(i,t)$. If we consider the degree of the node i, k_i, the number of infectious neighbors of device i at t satisfies $0 \leq X_4(i,t) \leq k_i$. For all the nodes, we consider a partition of the interval $[0,K]$, $0 = \gamma_0 < \gamma_1 < \ldots < \gamma_l = K = \max\{k_1, \ldots, k_r\}$, to define the discrete possible values of $X_4(i,t)$. In our model, we considered three possible values:

(1) $x_{41} = 0$, there are no infectious and damaged neighbor devices. Then, the probability of a susceptible device being infectious is 0.
(2) $x_{42} \in \{1, 2, 3, 4\}$, there are some infectious and damaged neighbors. Thus, the probability of a susceptible device being infectious is $p_1 > 0$.
(3) $x_{43} > 4$, there are many infectious and damaged neighbors. This means that the probability of a susceptible device being infectious is p_2 with $p_2 > p_1 > 0$.

3.2. Propagation of the Malware

In order to study the propagation of the malware, we use Bayesian networks. Then, if we take into account Equation (3) and the Bayesian network represented in Figure 1, we obtain the following:

$$\begin{aligned} & P(X(i,t+\Delta t) = A | X(i,t) = B) \\ & = \prod_{j=1}^{n} P\big(X_j(i,t+\Delta t) = a_j | \mathrm{pa}(X_j(i,t+\Delta t)) = \bar{b}_j\big) \\ & = P(X_1(i,t+\Delta t) = a_1 | (X_1(i,t), X_2(i,t), X_3(i,t), X_4(i,t)) = \bar{b}_1) \\ & \cdot P(X_2(i,t+\Delta t) = a_2 | X_2(i,t) = \bar{b}_2) \\ & \cdot P(X_3(i,t+\Delta t) = a_3 | X_3(i,t) = \bar{b}_3) \end{aligned} \tag{4}$$

Then, we only need to determine three CPDs: the CPD of $X_1(i,t+\Delta t)$ (epidemiological state), the CPD of $X_2(i,t+\Delta t)$ (target consideration), and the CPD of $X_3(i,t+\Delta t)$ (efficient countermeasures). The CPD of $X_2(i,t+\Delta t)$ is shown in Table 2.

Table 2. CPD of $X_2(i,t+\Delta t)$.

$X_2(i,t)$	x_{21}	x_{22}
x_{21}	$c_{1,1}$	$c_{1,2}$
x_{22}	$c_{2,1}$	$c_{2,2}$

Note that

$$c_{1,1} = P(X_2(i,t+\Delta t) = \text{'yes'} | X_2(i,t) = \text{'yes'}), \tag{5}$$
$$c_{1,2} = P(X_2(i,t+\Delta t) = \text{'yes'} | X_2(i,t) = \text{'no'}), \tag{6}$$
$$c_{2,1} = P(X_2(i,t+\Delta t) = \text{'no'} | X_2(i,t) = \text{'yes'}), \tag{7}$$
$$c_{2,2} = P(X_2(i,t+\Delta t) = \text{'no'} | X_2(i,t) = \text{'no'}). \tag{8}$$

Table 3 presents the CPD of $X_3(i,t+\Delta t)$:

$$e_{1,1} = P(X_3(i,t+\Delta t) = \text{'yes'} | X_3(i,t) = \text{'yes'}), \tag{9}$$
$$e_{1,2} = P(X_3(i,t+\Delta t) = \text{'yes'} | X_3(i,t) = \text{'no'}), \tag{10}$$
$$e_{2,1} = P(X_3(i,t+\Delta t) = \text{'no'} | X_3(i,t) = \text{'yes'}), \tag{11}$$
$$e_{2,2} = P(X_3(i,t+\Delta t) = \text{'no'} | X_3(i,t) = \text{'no'}). \tag{12}$$

Table 3. CPD of $X_3(i,t+\Delta t)$.

$X_3(i,t)$	x_{31}	x_{32}
x_{31}	$e_{1,1}$	$e_{1,2}$
x_{32}	$e_{2,1}$	$e_{2,2}$

Finally, there is a CPD for $X_1(i,t+\Delta t)$. Inasmuch as the variable $X_1(i,t)$ can have four values, the variable $X_2(i,t)$ can have two values, the variable $X_3(i,t)$ can have two values, the variable $X_4(i,t)$ can have three values, and the CPD for the $X_1(i,t+\Delta t)$ can have

$4 \times 2 \times 2 \times 3 \times 4 = 192$ parameters. Due to there being too many parameters to include in one table, we can consider several tables with fewer values. For example, if we regard each table with the same values of "$X_4(i,t)$" and "$X_1(i,t)$", then there are 12 tables (there are 12 possible combinations of the variables "$X_4(i,t)$" and "$X_1(i,t)$") with 16 parameters in each table as is shown in Table 4.

Table 4. CPD of $X_1(i, t + \Delta t)$.

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	$s_{1,4}$
x_{12}	$s_{2,1}$	$s_{2,2}$	$s_{2,3}$	$s_{2,4}$
x_{13}	$s_{3,1}$	$s_{3,2}$	$s_{3,3}$	$s_{3,4}$
x_{14}	$s_{4,1}$	$s_{4,2}$	$s_{4,3}$	$s_{4,4}$

For example, we can regard Table 4 associated to $X_4(i,t) =$ "more than four devices" and $X_1(i,t) =$ "Infectious". Then, $s_{1,1}$ is the probability of $X_1(i, t + \Delta t) =$ "Susceptible" supposing that $X_2(i,t) =$ "yes" , $X_3(i,t) =$ "yes", $X_4(i,t) =$ "more than four devices", and $X_1(i,t) =$ "Infectious"; $s_{2,2}$ is the probability of $X_1(i, t + \Delta t) =$ "Infectious" supposing that $X_2(i,t) =$ "yes", $X_3(i,t) =$ "no", $X_4(i,t) =$ "more than four devices" and $X_1(i,t) =$ "Infectious"; $s_{2,3}$ is the probability of $X_1(i, t + \Delta t) =$ "Infectious" supposing that $X_2(i,t) =$ "no", $X_3(i,t) =$ "yes", $X_4(i,t) =$ "more than four devices", and $X_1(i,t) =$ "Infectious", and so on.

Once the probabilities are defined, random values are used to simulate malware propagation. First, we need to know which are the probabilities associated to our situation. For example, if we consider a node with the characteristics $X_4(i,t) =$ "more than four devices", $X_1(i,t) =$ "Infectious", $X_2(i,t) =$ "no" and $X_3(i,t) =$ "yes", then we have to take into account the colored columns of Tables 5 and 6.

Table 5. CPD of the epidemiological state.

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	$s_{1,4}$
x_{12}	$s_{2,1}$	$s_{2,2}$	$s_{2,3}$	$s_{2,4}$
x_{13}	$s_{3,1}$	$s_{3,2}$	$s_{3,3}$	$s_{3,4}$
x_{14}	$s_{4,1}$	$s_{4,2}$	$s_{4,3}$	$s_{4,4}$

Table 6. CPDs of the target (on the left) and efficient countermeasures (on the right).

$X_2(i,t)$	x_{21}	x_{22}
x_{21}	$c_{1,1}$	$c_{1,2}$
x_{22}	$c_{2,1}$	$c_{2,2}$
$X_3(i,t)$	x_{31}	x_{32}
x_{31}	$e_{1,1}$	$e_{1,2}$
x_{32}	$e_{2,1}$	$e_{2,2}$

Next, if we apply Equation (3), we obtain that there are 16 possible values for $P(X(i, t + \Delta t) = A | X(i,t) = B)$. These values are: $x_1 = s_{1,3} \times c_{1,2} \times e_{1,1}$, $x_2 = s_{1,3} \times c_{1,2} \times e_{2,1}$, $x_3 = s_{1,3} \times c_{2,2} \times e_{1,1}$, $x_4 = s_{1,3} \times c_{2,2} \times e_{2,1}$, etc. Then, we can form the following intervals:

$$[0, x_1), [x_1, x_1 + x_2), [x_1 + x_2, x_1 + x_2 + x_3), \ldots, [\sum_{l=1}^{15} x_l, \sum_{l=1}^{16} x_l = 1), \tag{13}$$

that are illustrated in Figure 3.

Then, if we choose at random a number within $[0, 1)$, this can be situated in one of the 16 intervals. In this case, if the number is in the first interval, $X_1(i, t + \Delta t) =$ "Susceptible", $X_2(i, t + \Delta t) =$ "yes", and $X_3(i, t + \Delta t) =$ "yes". If the number is in the second interval, $X_1(i, t + \Delta t) =$ "Susceptible", $X_2(i, t + \Delta t) =$ "yes", and $X_3(i, t + \Delta t) =$ "no". The same technique can be used for the rest of the intervals.

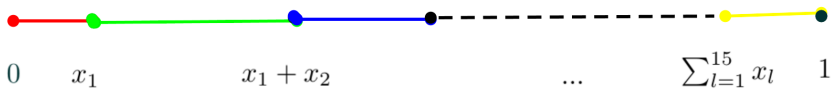


Figure 3. Intervals of the probabilities $P(X(i, t + \Delta t) = A | X(i, t) = B)$.

This way, we can obtain the future characteristics of a node i. Therefore, this method is applied to each node of our network to simulate one step of the model. Finally, we repeat the same process over a certain number of steps.

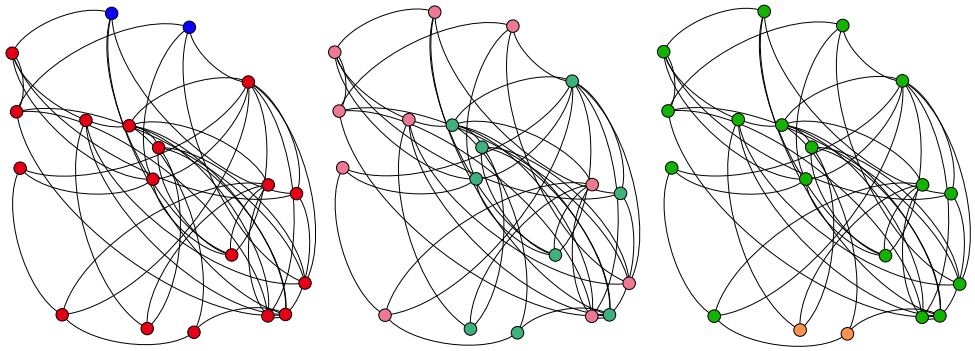
4. Illustrative Example of Malware Propagation

This section shows a temporal model based on the previous Bayesian network with defined parameters.

4.1. Initial Conditions

The following initial conditions were taken into account to obtain the simulations associated to the proposed model:

1. There are 19 devices in the network. Moreover, the network satisfies the following characteristics: the average grade is 2.421, the network diameter is 3 and the network density is 0.269.
2. There are two targets, which are represented with blue in Figure 4a. The rest of the nodes are not considered targets.
3. There are devices with efficient countermeasures (in green) and without efficient countermeasures (in pink), shown in Figure 4b.
4. All of the devices are susceptible, except two, which are infectious. The susceptible devices are shown in green and the infectious devices are illustrated in orange in Figure 4c.



(**a**) (**b**) (**c**)

Figure 4. Initial conditions of the simulations; (**a**) Targets; (**b**) Efficient countermeasures; (**c**) Initial states.

4.2. Determination of the CPDs

The probabilities of the conditional probability distributions (CPDs) in the Bayesian network are defined as follows.

First, it is considered that targets does not change through time. Then, the CPD associated with the future target is shown in Table 7.

According to the countermeasures, the probability of a device having efficient countermeasures is high if this has efficient countermeasures. This happens because efficient countermeasures are usually maintained due to the security awareness of a user, which

usually remain unchanged through time. Similarly, if a device does not have efficient countermeasures, the probability of having efficient countermeasures is low. Therefore, the CPD associated with the future target is shown in Table 7.

Table 7. CPDs of the target (on the left) and efficient countermeasures (on the right).

$X_2(i,t)$	x_{21}	x_{22}
x_{21}	1	0
x_{22}	0	1
$X_3(i,t)$	x_{31}	x_{32}
x_{31}	0.95	0.05
x_{32}	0.05	0.95

The CPD of the future state can be divided into 12 tables:

1. The first table considers that there are no infectious devices around and the node is susceptible ($X_4(i,t) = 0$ and $X_1(i,t) =$ 'susceptible'). If there are no infectious devices, a node cannot be infected. Then, the node must stay in the same state. An example of CPD is shown in Table 8.
2. The second table considers that there are no infectious devices in contact with the node and the node is infectious ($X_4(i,t) = 0$ and $X_1(i,t) =$ 'infectious'). In this table, it is considered that there is a probability to stay in the same state, a probability to be damaged (if the node is a target), and a probability to be recovered. Moreover, it is considered that there is a higher probability to recover if the node has efficient countermeasures. An example of CPD is the Table 8.

Table 8. CPDs of the future state without infectious devices and with the initial states: susceptible (on the left) and infectious (on the right).

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	1	1	1	1
x_{12}	0	0	0	0
x_{13}	0	0	0	0
x_{14}	0	0	0	0
$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	0	0
x_{12}	0.2	0.3	0.5	0.7
x_{13}	0.6	0.6	0	0
x_{14}	0.2	0.1	0.5	0.3

3. The third table takes into account that there are no infectious devices around the node and the node has a damaged state ($X_4(i,t) = 0$ and $X_1(i,t) =$ 'damaged'). Therefore, there is a probability to stay in the damaged state and a probability to turn into a recovered device. Furthermore, because it is considered that only targets can be damaged, the probabilities of the devices that are no targets are erased. An example of CPD is shown in Table 9.
4. The fourth table keeps in mind that there are no infectious devices around, and the node is a recovered device ($X_4(i,t) = 0$ and $X_1(i,t) =$ 'recovered'). Then, the node must stay in the same state due to the flow diagram shown in Figure 1. An example of this table is shown in Table 9.

Table 9. CPDs of the future state without infectious devices and with the initial states: damaged (on the left) and recovered (on the right).

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	-	-
x_{12}	0	0	-	-
x_{13}	0.4	0.5	-	-
x_{14}	0.6	0.5	-	-

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	0	0
x_{12}	0	0	0	0
x_{13}	0	0	0	0
x_{14}	1	1	1	1

5. The fifth table takes into account that there are between one and four infectious devices around and the node is susceptible ($X_4(i,t) \in \{1,2,3,4\}$ and $X_1(i,t) =$ 'susceptible'). Then, there is a probability to be infected and a probability to remain in the same state. If a node has efficient countermeasures, it is more difficult to turn into an infectious device. An instance of CPD is shown in Table 10.
6. The sixth table considers that there are between one and four infectious devices and the node is infectious ($X_4(i,t) \in \{1,2,3,4\}$ and $X_1(i,t) =$ 'infectious'). Therefore, probabilities to turn into damaged and recovered devices exist. There is a probability to remain in the same state too. In this table, the efficient countermeasures are also kept in mind. An instance of CPD is presented in Table 10.

Table 10. CPDs of the future state with one to four infectious devices and with the initial states: susceptible (on the left) and infectious (on the right).

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0.5	0.3	0.5	0.3
x_{12}	0.5	0.7	0.5	0.7
x_{13}	0	0	0	0
x_{14}	0	0	0	0

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	0	0
x_{12}	0.2	0.3	0.5	0.7
x_{13}	0.6	0.6	0	0
x_{14}	0.2	0.1	0.5	0.3

7. The seventh table considers that there are between one and four infectious devices in contact and the node is damaged ($X_4(i,t) \in \{1,2,3,4\}$ and $X_1(i,t) =$ 'damaged'). Therefore, a probability to turn into a recovered state exists. There is also a probability to stay in the same state. The probabilities of the devices that are not targets are erased because only targets can be damaged. An example of CPD is found in Table 11.
8. The eighth table takes into account that there are between one and four devices in contact and the node is recovered ($X_4(i,t) \in \{1,2,3,4\}$ and $X_1(i,t) =$ 'recovered'). There is only a probability of 1 to remain in the same state. An example of CPD is shown in Table 11.

Table 11. CPDs of the future state with one four infectious devices and with the initial states: damaged (on the left) and recovered (on the right).

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	-	-
x_{12}	0	0	-	-
x_{13}	0.4	0.5	-	-
x_{14}	0.6	0.5	-	-
$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	0	0
x_{12}	0	0	0	0
x_{13}	0	0	0	0
x_{14}	1	1	1	1

9. The ninth table regards that there are more than four infectious devices and the node is susceptible ($X_4(i,t) > 4$ and $X_1(i,t) =$ 'susceptible'). Then there is a higher probability to turn into an infectious device due to there being a lot of infectious devices around the node. There are probabilities to remain susceptible too. An instance of CPD is shown in Table 12.
10. The tenth table considers that there are more than four infectious devices and the node is infectious ($X_4(i,t) > 4$ and $X_1(i,t) =$ 'infectious'). This table takes into account that the node can turn into a damaged node (if the node is a target) or a recovered node. An example of CPD is shown in Table 12.

Table 12. CPDs of the future state with more than four infectious devices and with initial states: susceptible (on the left) and infectious (on the right).

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0.3	0.1	0.3	0.1
x_{12}	0.7	0.9	0.7	0.9
x_{13}	0	0	0	0
x_{14}	0	0	0	0
$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	0	0
x_{12}	0.2	0.3	0.5	0.7
x_{13}	0.6	0.6	0	0
x_{14}	0.2	0.1	0.5	0.3

11. The eleventh table considers that more than four infectious devices in contact exists and the node is damaged ($X_4(i,t) > 4$ and $X_1(i,t) =$ 'damaged'). Then, the node can turn into a recovered one or remain in the same state. An instance of CPD is shown in Table 13.
12. Finally, the twelfth table takes into account that there are more than four devices in contact and the node is recovered ($X_4(i,t) > 4$ and $X_1(i,t) =$ 'recovered'). Therefore, there is only the probability to stay in the same state. An example of CPD is shown in Table 13.

Table 13. CPDs of the future state with more than four infectious devices and with initial states: damaged (on the left) and recovered (on the right).

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	-	-
x_{12}	0	0	-	-
x_{13}	0.4	0.5	-	-
x_{14}	0.6	0.5	-	-

$X_2(i,t)$	x_{21}		x_{22}	
$X_3(i,t)$	x_{31}	x_{32}	x_{31}	x_{32}
x_{11}	0	0	0	0
x_{12}	0	0	0	0
x_{13}	0	0	0	0
x_{14}	1	1	1	1

4.3. Simulation of the Model

An example of malware propagation taking into account the previous CPDs is shown in Figures 5 and 6. The program GNU Octave was used to perform this simulation.

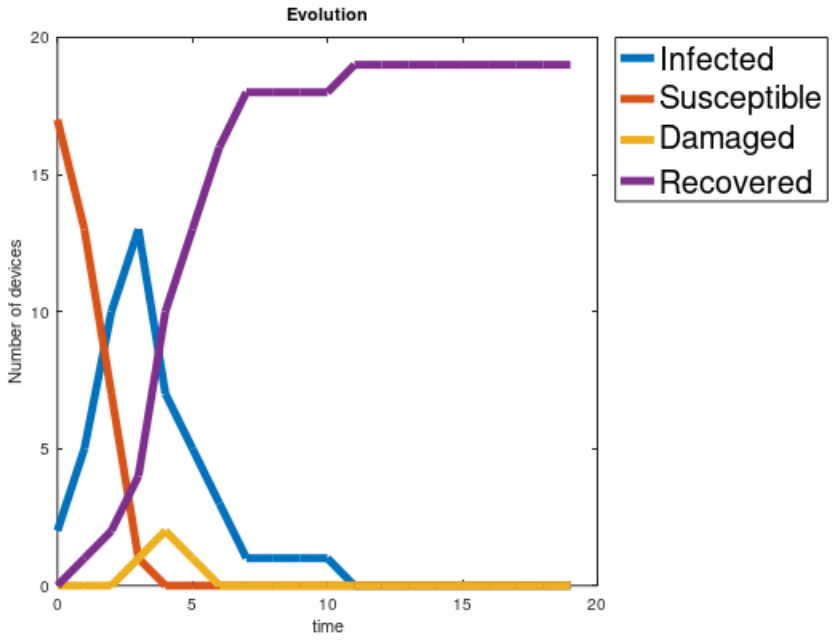


Figure 5. Malware propagation with damaged devices.

In the simulation of Figure 5, one can see how susceptible devices disappear through time. Moreover, recovered devices increase through time. According to the infectious devices, first, these increase and later decrease. Finally, the two targets are damaged.

On the other hand, using the same parameters, we can obtain Figure 6. In this case, the simulation is similar to the previous simulation. However, in this case, any target is reached.

In reference to the evolution, this can end up in two ways:

1. All devices are recovered. This happens when all the susceptible devices are infected. We can observe this in Figure 5.
2. Some devices are recovered and some devices are susceptible. This happens when all the damaged and infected devices are recovered and some susceptible devices remain. We can observe this in Figure 6.

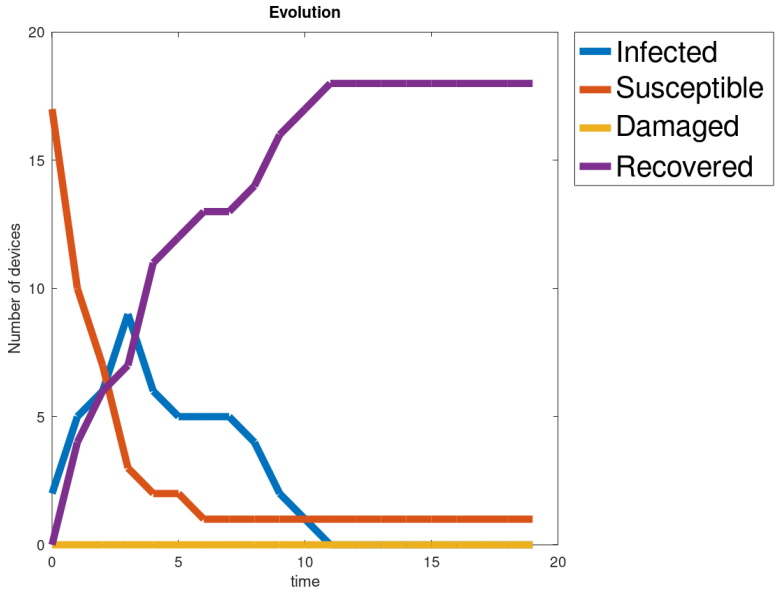


Figure 6. Malware propagation without damaged devices.

After several simulations, we obtained that there is an approximate percentage of 72% to infect a target. This is a high probability, so it would be necessary to change the characteristics of the nodes (increase security countermeasures) or the characteristics of the network (networks with different links) to improve security. For example, if considering more devices with highly efficient security countermeasures ($X_3(i,t) =$ "yes"), then it is harder for the malware to propagate in the network. Another option is to consider more nodes with fewer neighboring nodes. As a consequence, it is harder to infect these nodes and the epidemic propagation is lower.

4.4. Effect of the Efficient Security Countermeasures and Number of Neighboring Infectious Devices

The number of devices with efficient security countermeasures affect malware propagation. In order to show this, we considered three initial conditions:

- All nodes have efficient security countermeasures: all of the devices have some type of efficient software anti-malware.
- Fifty percent of the nodes have efficient security countermeasures: 9 out of 19 devices have some type of efficient anti-malware software.
- All nodes do not have efficient security countermeasures: none of the devices have some type of efficient anti-malware software.

We also considered the average of three characteristics: the sum of the number of infectious devices during the epidemic (total number of infectious), the number of epochs until reaching the peak (epoch of the peak), the duration of the infection and the peak of the infection. After several simulations, we obtained the results shown in Table 14.

Table 14. Effect of the efficient security countermeasures.

	All with Software	50% Software	Without Software
Total number of infectious	32	38	47.5
Epoch of the peak	3.4	3.5	3.8
Peak of the infection	6.8	8.7	9.8
Duration of the infection	8.4	9.5	11.2

Therefore, we obtain that the total number of infectious increases when we remove efficient software of the devices. This can happen because the APT is stealthy and tries to hide from efficient software. Then, if there is a device with efficient software, it is more difficult to infect that device. Moreover, the peak is reached between epochs 3 and 4 in all of the simulations. This implies that the speed of the propagation decreases when all of the devices have efficient software due to the peak being slower. However, if we consider the duration of the infection, we can deduce that the recuperation is faster when all of the devices have efficient software.

In previous simulations, it is defined that the boundary γ_1 for the number of infectious nodes in contact with a device is 4. If we consider different values of γ_1, we can observe that this parameter also affects the model propagation. Considering 50% of devices having efficient software anti-malware, we obtain Figure 7.

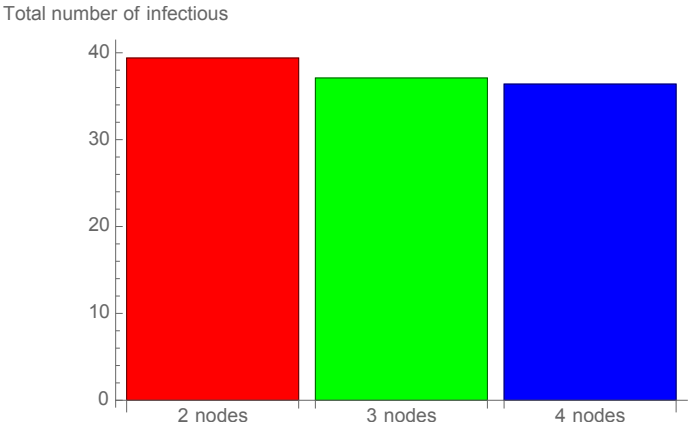


Figure 7. Total number of infectious devices with different values of γ_1.

Then, if the boundary is smaller, the total number of infectious devices in the epidemic is bigger.

5. Conclusions

In this work, a new type of model that simulates malware propagation was created. This model is based on dynamic Bayesian networks and simulates the malware used in APTs. Moreover, considered individual characteristics were considered to define the model, such as efficient countermeasures, epidemiological states, the number of infectious nodes in contact with a node, and being a target.

Under certain characteristics, this type of malware can damage the target devices of our network. With this model, we can calculate the probability of damaging target devices in a network. If there is high probability, the network of devices is not safe. Then, we can improve the efficient countermeasures or change the links of the network to avoid malware damaging the targets. For example, instead of considering that 50% of devices have high efficient countermeasures, we can consider higher percentages, or we can consider fewer links between the devices.

Moreover, this model can be applied to a concrete network. Keeping this in mind this, it would be interesting to study different networks in this model and observe how the networks affect malware propagation. We can take into account other measures instead of the number of infectious that are in contact with a node, such as centrality measures. Moreover, Bayesian networks are a type of machine learning technique. Due to the development of deep learning techniques, it would be intriguing to use these techniques to calculate the parameters of the model and simulate malware propagation. Some of these ideas will be studied in the future.

Author Contributions: Conceptualization, J.D.H.G.; methodology, J.D.H.G., A.M.d.R. and R.C.-V.; software, J.D.H.G.; writing—original draft preparation, J.D.H.G., and A.M.d.R.; writing—review and editing, J.D.H.G., A.M.d.R. and R.C.-V. All authors have read and agreed to the published version of the manuscript.

Funding: J.D. Hernández Guillén is supported by Banco Santander and Universidad de Salamanca (Spain) under a postdoctoral grant.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Li, J.H. Overview of Cyber Security Threats and Defense Technologies for Energy Critical Infrastructure. *J. Electron. Inf. Technol.* **2020**, *42*, 2065–2081.
2. Bhamare, D.; Zolanvari, M.; Erbad, A.; Jain, R.; Khan, K.; Meskin, N. Cybersecurity for industrial control systems: A survey. *Comput. Secur.* **2020**, *89*, 101677. [CrossRef]
3. NIST. Information Security. Special Publication 800–39. 2011. Available online: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-39.pdf (accessed on 22 October 2021).
4. Alshamrani, A.; Myneni, S.; Chowdhary, A.; Huang, D. A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 1851–1877. [CrossRef]
5. Lemay, A.; Calvet, J.; Menet, F.; Fernandez, J.M. Survey of publicly available reports on advanced persistent threat actors. *Comput. Secur.* **2018**, *72*, 26–59. [CrossRef]
6. Chen, P.; Desmet, L.; Huygens, C. A study on advanced persistent threats. In *IFIP International Conference on Communications and Multimedia Security*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 63–72.
7. Chakkaravarthy, S.S.; Sangeetha, D.; Vaidehi, V. A survey on malware analysis and mitigation techniques. *Comput. Sci. Rev.* **2019**, *32*, 1–23. [CrossRef]
8. Fu, Y.; Li, H.; Wu, X.; Wang, J. Detecting APT attacks: A survey from the perspective of big data analysis. *J. Commun.* **2015**, *36*, 1–14.
9. Moon, D.; Im, H.; Kim, I.; Park, J.H. DTB-IDS: An intrusion detection system based on decision tree using behavior analysis for preventing APT attacks. *J. Supercomput.* **2017**, *73*, 2881–2895. [CrossRef]
10. Lu, J.; Chen, K.; Zhuo, Z.; Zhang, X. A temporal correlation and traffic analysis approach for APT attacks detection. *Clust. Comput.* **2019**, *22*, 7347–7358. [CrossRef]
11. Hosseini, S.; Azgomi, M.A. The dynamics of an SEIRS-QV malware propagation model in heterogeneous networks. *Physica A* **2018**, *512*, 803–817. [CrossRef]
12. Kudo, T.; Kimura, T.; Inoue, Y.; Aman, H.; Hirata, K. Stochastic modeling of self-evolving botnets with vulnerability discovery. *Comput. Commun.* **2018**, *124*, 101–110. [CrossRef]
13. Xiao, X.; Fu, P.; Li, Q.; Hu, G.; Jiang, Y. Modeling and validation of SMS worm propagation over social networks. *J. Comput. Sci.* **2017**, *21*, 132–139. [CrossRef]
14. Hernández, J.; del Rey, A.; Casado-Vara, R. Security Countermeasures of a SCIRAS Model for Advanced Malware Propagation. *IEEE Access* **2019**, *7*, 135472–135478. [CrossRef]
15. Hernández, J.; del Rey, A. Modeling malware propagation using a carrier compartment. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *56*, 217–226. [CrossRef]
16. Piqueira, J.R.C.; Batistela, C.M. Considering quarantine in the SIRA malware propagation model. *Math. Probl. Eng.* **2019**, *2019*, 6467104. [CrossRef]
17. Hosseini, S.; Azgomi, M. A model for malware propagation in scale-free networks based on rumor spreading process. *Comput. Networks* **2016**, *108*, 97–107. [CrossRef]
18. Huang, S. Global dynamics of a network-based WSIS model for mobile malware propagation over complex networks. *Physica A* **2018**, *503*, 293–303. [CrossRef]
19. Martín, A.; Hernández, G.; Tabernero, A.B.; Queiruga, A. Advanced malware propagation on random complex networks. *Neurocomputing* **2021**, *423*, 689–696. [CrossRef]
20. Zimba, A.; Chen, H.; Wang, Z.; Chishimba, M. Modeling and detection of the multi-stages of advanced persistent threats attacks based on semi-supervised learning and complex networks characteristics. *Future Gener. Comput. Syst.* **2020**, *106*, 501–517. [CrossRef]
21. Zhou, P.; Gu, X.; Nepal, S.; Zhou, J. Modeling social worm propagation for advanced persistent threats. *Comput. Secur.* **2021**, *108*, 102321. [CrossRef]
22. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; MIT Press: Cambridge, MA, USA, 2009.
23. Cai, B.; Huang, L.; Xie, M. Bayesian networks in fault diagnosis. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2227–2240. [CrossRef]
24. Varshney, D.; Kumar, S.; Gupta, V. Predicting information diffusion probabilities in social networks: A Bayesian networks based approach. *Knowl.-Based Syst.* **2017**, *133*, 66–76. [CrossRef]

25. Kondakci, S. Epidemic state analysis of computers under malware attacks. *Simul. Model. Pract. Theory* **2008**, *16*, 571–584. [CrossRef]
26. Zimba, A.; Chen, H.; Wang, Z. Bayesian network based weighted APT attack paths modeling in cloud computing. *Future Gener. Comput. Syst.* **2019**, *96*, 525–537. [CrossRef]
27. Koch, D.; Eisinger, R.S.; Gebharter, A. A causal Bayesian network model of disease progression mechanisms in chronic myeloid leukemia. *J. Theor. Biol.* **2017**, *433*, 94–105. [CrossRef]
28. Jiang, X.; Wallstrom, G.; Cooper, G.F.; Wagner, M.M. Bayesian prediction of an epidemic curve. *J. Biomed. Inform.* **2009**, *42*, 90–99. [CrossRef] [PubMed]