Modeling the Spread of Malware on Complex Networks

Chapter · January 2020

DOI: 10.1007/978-3-030-23946-6_12

4 authors, including:

Ángel Martín del Rey
Universidad de Salamanca
151 PUBLICATIONS **855** CITATIONS
SEE PROFILE

Araceli Queiruga-Dios
Universidad de Salamanca
83 PUBLICATIONS **162** CITATIONS
SEE PROFILE

Guillermo Hernández
Universidad de Salamanca
12 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Modeling the Spread of Malware on Complex Networks

A. Martín del Rey[1(✉)], A. Queiruga Dios[1], G. Hernández[2], and A. Bustos Tabernero[3]

[1] Institute of Fundamental Physics and Mathematics, University of Salamanca, Salamanca, Spain
{delrey,queirugadios}@usal.es
[2] BISITE Research Group, University of Salamanca, Salamanca, Spain
guillehg@usal.es
[3] Faculty of Sciences, University of Salamanca, Salamanca, Spain
alvarob97@usal.es

Abstract. Currently, zero-day malware is a major problem as long as these specimens are a serious cyber threat. Most of the efforts are focused on designing efficient algorithms and methodologies to detect this type of malware; unfortunately models to simulate its behavior are not well studied. The main goal of this work is to introduce a new individual-based model to simulate zero-day malware propagation. It is a compartmental model where susceptible, infectious and attacked devices are considered. Its dynamics is governed by means of a cellular automaton whose local functions rule the transitions between the states. The propagation is briefly analyzed considering different initial conditions and network topologies (complete networks, random networks, scale-free networks and small-world networks), and interesting conclusions are derived.

Keywords: Malware · Propagation · Complex networks · Individual-based model · Cellular automata

1 Introduction

Malicious code (also known as malware) can be considered one of the major threats to our digital society since the different types (computer viruses, trojans, computer worms, etc.) are the fundamental tools used in cyberattacks. Special mention must be made to zero-day malware and its role in the Advanced Persistent Threats (APTs for short).

APTs are sophisticated cyber-attacks [19] combining different technologies and methodologies. They are characterized by the following: (1) they have a precise and clear target, (2) they are long-term attacks constituted by several phases of different nature (reconnaissance, incursion, discovery, capture and exfiltration), (3) they implement evasive techniques consisting on stealthy behavior and adaptation to defenders' efforts, and (4) they are complex due to the use

E. Herrera-Viedma et al. (Eds.): DCAI 2019, AISC 1004, pp. 109–116, 2020.
https://doi.org/10.1007/978-3-030-23946-6_12

of different attack methods (zero-day malware, rootkits, etc.) Zero-day malware plays an important role in the implementation of APTs [18]. It is a specimen of malicious code which is characterized by exploiting unknowns (or non-patched) vulnerabilities [15].

The fight against malware is mainly based on the design of Machine Learning techniques and protocols to successfully detect it (see, for example [10,11]). Nevertheless, it is also important to simulate its propagation in an efficient way. In this sense several mathematical models have been proposed in the scientific literature [13]. Most of them are theoretical models of a continuous nature where the spreading environment is defined by a complete network where all devices are in contact with each other [2]. Due to their initial constraints, these models have limited practical application and this drawback has been tried to overcome by considering alternative topologies [1]. In recent years some proposals have appeared dealing with the simulation of malware propagation on complex networks: some of them are deterministic models [4,5,7,9,12], and other are stochastic (see, for example [6,8] and references therein). Although these models give rise to more realistic simulations, they do not take into account the individual characteristics of the devices. This is a very important issue and some attempts have been made using Artificial Intelligence techniques such as agent-based models [3,6] or cellular automata [14,17]. Nevertheless, as far as we know, no individual-based model considering zero-day malware has been proposed.

The main goal of this work is to introduce a novel model to simulate zero-day malware whose dynamics will be governed by means of a cellular automaton. It is an individual-based model where the devices are classified into three compartments: susceptible, infectious and attacked. Its dynamics considers some characteristics of zero-day malware: an infectious device becomes susceptible again if it is not the target of the attack. The model is flexible in the sense that different topological conditions are stated in order to determine the local interactions between the devices.

The rest of the paper is organized as follows: in Sect. 2 the definition of cellular automata is briefly introduced; the model for malware spreading is described in Sect. 3; in Sect. 4 the study of the effects of topology in the propagation is shown, and finally, the conclusions and future work are presented in Sect. 5.

2 Cellular Automata

A cellular automaton (CA for short) is a simple model of computation constituted by n identical memory units called cells [16]: $c_1, c_2, \ldots, c_n$, that are arranged following a certain topology defined by means of a complex network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. As a consequence, each cell of the cellular automaton stands for a vertex of $\mathcal{G}$: $c_i \in \mathcal{V}$ with $1 \leq i \leq n$, and the local interactions between these cells are modeled by means of the edges: there is a connection between c_i and c_j if $(c_i, c_j) \in \mathcal{E}$. Consequently, the neighborhood of the cell c_i is defined as the collection of its adjacent cells:

$$\mathcal{N}_i = \{v \in \mathcal{V} \text{ such that } (v, c_i) \in \mathcal{E}\} = \{c_{\alpha_1}, \ldots, c_{\alpha_i}\} \quad 1 \leq i \leq n. \tag{1}$$

Each cell c_i is endowed with a state at each step of time t: $s_i^t \in \mathcal{S}$, where $\mathcal{S}$ is a finite state set. This state changes accordingly to a local transition function f whose variables are the states of the main cell and the neighbor cells at the previous step of time:

$$s_i^{t+1} = f\left(s_i^t, s_{\alpha_1}^t, \ldots, s_{\alpha_i}^t\right) \in \mathcal{S}. \tag{2}$$

3 Description of the Model

The epidemiological model proposed in this work is a compartmental model where the population of devices is divided into three classes or compartments: susceptible, infectious and attacked. Susceptible devices are those that have not been infected by the malware (the device is free of the malicious code); the infectious devices are characterized because they have been reached by the malware but have not been attacked, and finally, attacked devices are those devices where the malware is carrying out its stealthy and malicious activity.

The dynamics of the propagation process is illustrated in Fig. 1. The coefficients involved are the following: the infection rate $0 \leq h_i \leq 1$ that rules the infection of a device free of malware, the targeted coefficient $0 \leq a_i \leq 1$ which defines the probability that the infectious device c_i will be effectively attacked, the recovery (from infectious) probability $0 \leq b_i << a_i \leq 1$, and the recovery (from attacked) coefficient $r_i (= 0 \text{ or } 1)$ that determines the auto-removing malware process.

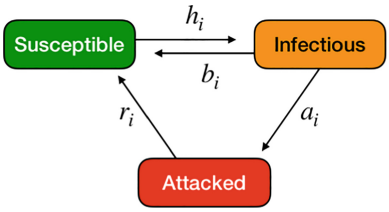

Fig. 1. Flow diagram representing the dynamics of the model.

Note that a susceptible device becomes infectious when the malware reaches it (and, in our model, this depends on both the infection rate and the number of infectious neighbor devices); the infectious device c_i becomes attacked with probability a_i or susceptible with probability $b_i << a_i$ (in this work it is supposed that the malware can remove itself if it does not find any neighbor host or the current host must not be attacked); finally, the attacked devices recover once the attacked period is finished. As a consequence, it is a SIAS model (Susceptible-Infectious-Attacked-Susceptible).

The propagation model is based on a cellular automaton whose main characteristics are the following:

- Each device represents a cell of the CA.
- The state set is formed by three states: susceptible (S), infectious (I), and attacked (A): $\mathcal{S} = \{S, I, A\}$.
- The local transition functions between the compartments are the following:
  - Transition from susceptible to infectious: if $s_i^t = S$ then $s_i^{t+1} = I$ with probability $h_i \cdot N_i^t$, where N_i^t is the number of infectious neighbor devices of c_i at t. Obviously, the device c_i remains susceptible ($s_i^{t+1} = S$) with probability $1 - h_i \cdot N_i^t$.
  - Transition from infectious to attacked: if $s_i^t = I$ then $s_i^{t+1} = A$ with probability a_i.
  - Transition from infectious to susceptible: if $s_i^t = I$ then $s_i^{t+1} = S$ with probability b_i. Note that an infectious device c_i remains infectious at the next step of time with probability $1 - a_i - b_i$.
  - Transition from attacked to susceptible: finally, if $s_i^t = A$ then $s_i^{t+1} = S$ with probability r_i where:

$$r_i = \begin{cases} 1, & \text{if } t = t_i + \tau_i + 1 \\ 0, & \text{if } t_i + 1 \leq t \leq t_i + \tau_i \end{cases} \tag{3}$$

    where t_i is the step of time at which the device becomes attacked, and τ_i is the length of the attack period over the device c_i.

4 Topology Effects on the Propagation of Malware

In what follows we will perform several simulations of the model considering different initial conditions (numerical values of the parameters and characteristics of the topologies that determines the networks). For the sake of simplicity only illustrative examples of the simulations are shown (for each type of topology, the behaviors exhibited by the malware spreading are similar).

Suppose that the four epidemiological coefficients (infection rate, targeted rate, recovery —from infectious device— probability, and duration of the attack period) remain constant for every device: $h_i = \tilde{h}, a_i = \tilde{a}, b_i = \tilde{b}$ and $\tau_i = \tilde{\tau}$ with $1 \leq i \leq n$. Let's examine what happens if the topology is changed.

Assume that there are $n = 100$ devices in the network with only one infectious device at time step $t = 0$: the node with greater degree centrality (that is, with the largest number of neighbors). Set $\tilde{h} = 0.025, \tilde{a} = 0.25, \tilde{b} = 0.025$, and $\tilde{\tau} = 5$. Let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$, and $\mathcal{G}_4$ be a complete network, a random network –with edge probability $p = 0.5$–, a scale-free network –with 2 edges added at each step of the Barabasi-Albert algorithm–, and a small-world complex network –with rewiring probability $p = 0.1$ associated to the Watts-Strogatz algorithm–, respectively (see Fig. 2). Furthermore, some of the most important structural characteristics of these networks (density, diameter, radius and average path length) are shown in Table 1.

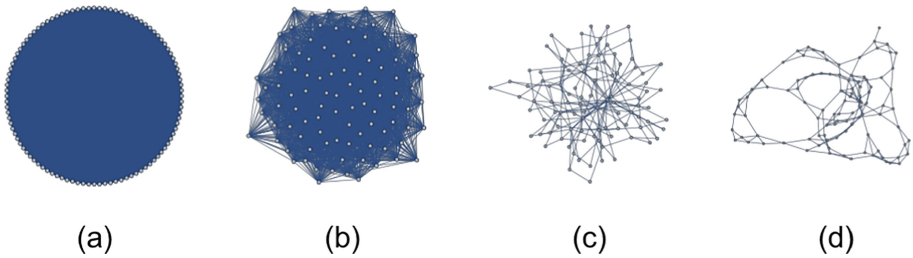

Fig. 2. (a) Complete network $\mathcal{G}_1$. (b) Random network $\mathcal{G}_2$. (c) Scale-free network $\mathcal{G}_3$. (d) Small-world network $\mathcal{G}_4$.

Table 1. Structural characteristics of the complex networks

Metric	$\mathcal{G}_1$	$\mathcal{G}_2$	$\mathcal{G}_3$	$\mathcal{G}_4$
Density	1	0.5103	0.03980	0.04040
Diameter	1	2	5	11
Radius	1	2	3	7
Average path length	1	1.490	3.044	5.042

Illustrative examples of the evolution of the different compartments (susceptible, infectious and attacked devices) are shown in Fig. 3 when the simulation of malware propagation during 90 units of time ($1 \leq t \leq 90$) is computed. Note that in the case of considering the complete topology the evolution of the different compartments exhibits an oscillatory behavior where the great majority of devices are infectious or attacked (see Fig. 3-(a)). The number of targeted devices (infectious and attacked) increases rapidly from the beginning and, in fact, at $t = 5$ the 100% of nodes are affected in some way by the malware. Furthermore, all nodes are infected or attacked during the simulation period.

A (qualitative) similar behavior is obtained when a random network $\mathcal{G}_2$ is considered (see Fig. 3-(b)). The propagation speed is lower but, as in the previous case, all nodes are infected and attacked by the malware. In this case the maximum is reached at $t = 10$ when the 98% of nodes are affected (46% of infectious devices, and 52% of attacked devices). The differences between the different device compartments are not as pronounced as for the complete case and the number of susceptible devices at every step of time is greater than in the precious case. Note that the graph density of $\mathcal{G}_2$ is approximately half of $\mathcal{G}_1$ and the diameter and radius of $\mathcal{G}_2$ is twice that of the complete network.

Very different from the previous dynamics are the evolutions of malware spreading obtained when scale-free or small-world networks are assumed (see Fig. 3-(c) and (d)). In these two cases the malware specimen barely manages to reach a neighbor node of the "patient zero". In fact, when $\mathcal{G}_3$ is considered only one effective contagion occurs; both devices were attacked and after that, the epidemic outbreak dies out. On the other hand, when the small-world network

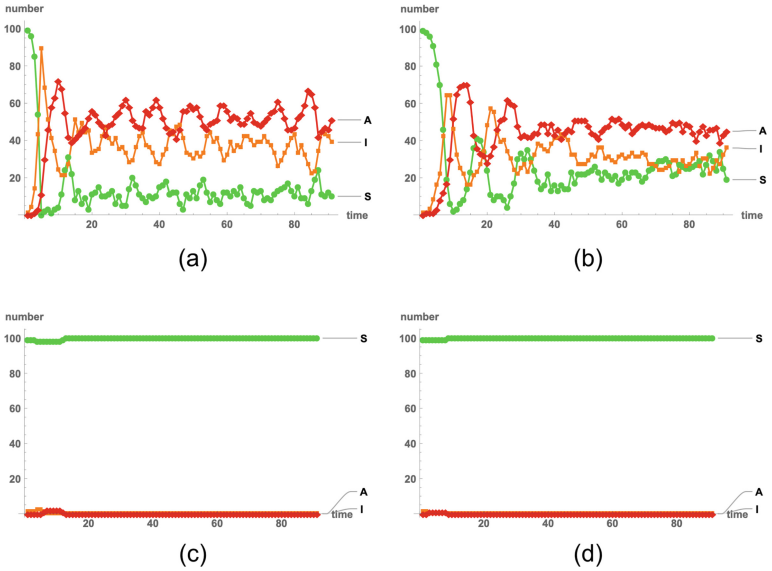

Fig. 3. Global evolution of malware on: (a) a complete network $\mathcal{G}_1$, (b) a random network $\mathcal{G}_2$, (c) a scale-free network $\mathcal{G}_3$, and (d) a small-world network $\mathcal{G}_4$.

$\mathcal{G}_4$ is considered, the malware fails to spread and only one targeted device exists (initially it was infectious and at time step $t = 3$ becomes attacked). This fact should not surprise us if we take into account that the densities of $\mathcal{G}_3$ and $\mathcal{G}_4$ are 0.04 and the average path lengths are much greater than those of $\mathcal{G}_1$ and $\mathcal{G}_2$.

5 Conclusions

In this work a novel computational model to simulate zero-day malware spreading is introduced and analyzed, where the environment is defined by means of a complete network, a typical random network, a scale-free network, and a small-world network. The model is compartmental considering susceptible, infectious and attacked devices. Its dynamics is characterized by the following:

(1) Two classes of targeted devices are considered: infectious (they serve as transmission vectors but they are not the target of the attack) and attacked (they are the main objectives of the attack).
(2) Infectious devices can immediately recover the susceptible status if they are not attacked.
(3) The duration of the malware payload is constant and, consequently, the attacked devices becomes susceptible again when the attacking period finishes.
(4) Only the stealthy behavior of the malware is taken into account. This is reflected when considering the local transition from infectious to susceptible and the short-term attack period.

Several simulations have been performed and only illustrative examples are shown in this paper. From a simple analysis of the same, the following conclusions follow:

- Some structural characteristics of the networks (density, diameter, radius, and average path length) have a direct influence on the propagation process. The malware spreading can be accelerated or decelerated depending on the value of these parameters.
- The type of topology also influences the propagation: complete or random complex networks (with probability greater that 0.5) exhibit a similar evolution: the initial outbreak becomes an epidemic process converging to an oscillatory behavior. On the other hand, scale-free and small-world complex networks slow down the propagation; in fact, the evolution tends to disease-free steady states.

Future work aimed at considering additional zero-day malware features in the design of the cellular automata (states and local transition functions). Furthermore, alternative scenarios must be analyzed in detail considering different choices of "patient zero" based on alternative centrality coefficients (clustering coefficient, betweenness coefficient, etc.) and different measures of the propagation speed.

Acknowledgements. This research has been partially supported by Ministerio de Ciencia, Innovación y Universidades (MCIU, Spain), Agencia Estatal de Investigación (AEI, Spain), and Fondo Europeo de Desarrollo Regional (FEDER, UE) under project with reference TIN2017-84844-C2-2-R (MAGERAN) and the project with reference SA054G18 supported by Consejería de Educación (Junta de Castilla y León, Spain).

A. Bustos Tabernero thanks Ministerio de Educación y Formación Profesional (Spain) for his departmental collaboration grant in the Department of Applied Mathematics (University of Salamanca, Spain).

References

1. Fu, X., Small, M., Chen, G.: Propagation Dynamics on Complex Networks. Wiley, Hoboken (2014)
2. Hernández Guillén, J.D., Martín del Rey, A.: Modeling malware propagation using a carrier compartment. Commun. Nonlinear Sci. Numer. Simul. **56**, 217–226 (2018)
3. Hosseini, S., Azgomi, M.A., Torkaman, A.R.: Agent-based simulation of the dynamics of malware propagation in scale-free networks. Simulation **92**, 709–722 (2016)
4. Hosseini, S., Azgomi, M.A.: A model for malware propagation in scale-free networks based on rumor spreading process. Comput. Netw. **108**, 97–107 (2017)
5. Hosseini, S., Azgomi, M.A.: The dynamics of a SEIRS-QV malware propagation model in heterogeneous networks. Phys. A **512**, 803–817 (2018)
6. Hu, P., Ding, L., Hadzibeganovic, T.: Individual-based optimal weight adaptation for heterogeneous epidemic spreading networks. Commun. Nonlinear Sci. Numer. Simul. **63**, 339–355 (2018)
7. Jackson, J.T., Creese, S.: Virus propagation in heterogeneous bluetooth networks with human behaviors. IEEE Trans. Dependable Secur. Comput. **9**, 930–943 (2012)

8. Karyotis, V., Khouzani, M.H.R.: Malware Diffusion Models for Modern Complex Networks. Morgan Kaufmann-Elsevier, Cambridge (2016)
9. Karyotis, V., Papavassiliou, S.: Macroscopic malware propagation dynamics for complex networks with churm. IEEE Commun. Lett. **19**, 577–580 (2015)
10. Kim, J.-Y., Bu, S.-J., Cho, S.-B.: Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders. Inform. Sci. **460**, 83–102 (2018)
11. Kim, T., Kang, B., Rho, M., Sezer, S., Im, E.G.: A multimodal deep learning method for Android malware detection using various features. IEEE Trans. Inf. Forensic Secur. **14**, 773–788 (2019)
12. Liu, W., Zhong, S.: A novel dynamic model for web malware spreading over scale-free networks. Phys. A **505**, 848–863 (2018)
13. Martín del Rey, A.: Mathematical modeling of the propagation of malware: a review. Secur. Commun. Netw. **8**, 2561–2579 (2015)
14. Martín del Rey, A., Rodríguez Sánchez, G.: A discrete mathematical model to simulate malware spreading. Int. J. Mod. Phys. C **23**, 1–16 (2012). Article number 1250064
15. Rudd, E.M., Rozsa, A., Günter, M., Boult, T.E.: A survey of stealth malware attacks, mitigation measures, and steps toward autonomous open world solutions. IEEE Commun. Surv. Tutor. **19**, 1145–1172 (2017)
16. Sarkar, P.: A brief history of cellular automata. ACM Comput. Surv. **32**, 80–107 (2000)
17. Thomson, B., Morris-King, J.: An agent-based modeling framework for cybersecurity in mobile tactical networks. J. Def. Model. Simulat. **15**, 204–218 (2018)
18. Tounsi, W., Rais, H.: A survey on technical threat intelligence in the age of sophisticated cyber attacks. Comput. Secur. **72**, 212–233 (2018)
19. Winkler, I., Treu Gomes, A.: Advanced Persistent Security. A Cyberwarfare Approach to Implementing Adaptive Enterprise Protection, Detection, and Reaction Strategies. Syngress-Elsevier, Cambridge (2017)