Accepted Manuscript

Modelling the malware propagation in mobile computer devices

María Teresa Signes Pont , Antonio Cortés Castillo ,
Higinio Mora Mora , Julian Szymanski

Modelling the malware propagation in mobile computer devices

María Teresa Signes Pont[1*], Antonio Cortés Castillo[2], Higinio Mora Mora[3], Julian Szymanski[4]

[1, 3] Departamento de Tecnología Informática y Computación, Universidad de Alicante, Spain
[2] Departamento de Informática, Universidad de Panamá, Panamá
[4] Department of Computer Systems Architecture, Gdansk University of Technology, Poland

teresa@dtic.us.es; antonio.cortes@up.ac.pa; hmora@dtic.ua.es; julian.szymanski@eti.pg.gda.pl

Abstract- Nowadays malware is a major threat to the security of cyber activities. The rapid development of the Internet and the progressive implementation of the Internet of Things (IoT) increase the security needs of networks. This research presents a theoretical model of malware propagation for mobile computer devices. It is based on the susceptible-exposed-infected-recovered-susceptible (SEIRS) epidemic model. The scheme is based on a concrete connection pattern between nodes defined by both a particular neighbourhood which fixes the connection between devices, and a local rule which sets whether the link is infective or not. The results corroborate the ability of our model to perform the behaviour patterns provided by the ordinary differential equation (ODE) traditional method.

Key-words- Epidemic model, Malware propagation, Mobile devices, Local rules, Von Neumann neighbourhood, Moore neighbourhood.

1. Introduction

Nowadays malware is a major threat to the security of cyber activities. The rapid development of the Internet and the progressive implementation of the Internet of Things (IoT) increase the security needs of networks. Several malicious objects (virus, worm, Trojan horse…) attack users and causes a great variety of damages: not only information resources are the targets of the cyber-attacks, humans can also be vulnerable even without being aware [1, 2]. Messages with malware can arrive from many sources and in a variety of forms. They can be sent as an attachment to spam email, embedded in a file or hidden in a link within the body of a message. The extent of the damage depends on the targets of the virus and sometimes the results of its activity are imperceptible for the users of a machine. The scenarios created by the apparition of many new instances can be classified among existing threat intelligence types [3, 4]. Malware epidemiology can be considered as a particular case of the propagation process of a disease expansion, since biological viruses, bacteria, fungi or prions, behave like computer viruses and worms [5]. The Kermack and McKendrick susceptible-infected-recovered (SIR) model and its variants susceptible-infected-susceptible (SIS), susceptible-infected-recovered-susceptible (SIRS) and susceptible-exposed--infected-recovered-susceptible (SEIRS) are dynamical approaches that assume the population is confined in different compartments; these are, S for susceptible people, I for infected people and R for recovered people when acquired immunity is permanent [6-11]. When the infection is not instantaneous, the E compartment stands for exposed people [12]. This approach can be improved by considering births and deaths in population instead of constant population. These models have traditionally been solved by means of ODE which provide accurate results for large populations. Stochastic models generally use Markov chains and are more suitable for small populations [13]. Both, deterministic and stochastic models consider the population as a whole and cannot assess local interactions between individuals because they do not take into account the individual behaviour of the nodes [14]. These drawbacks can be overcome by means of discrete models, which depict the local characteristics of the spreading processes [15-17]. This paper presents a time-space framework to model the malware propagation, based on local Boolean rules that fix the

evolution of the nodes placed in a square grid. The interactions between nodes are fixed through different connection types, such as Von Neumann, Moore or chess horse jumping neighbourhoods. This model succeeds in obtaining the same behavioural patterns than the simulation of ODE system. Following the Introduction, Section 2 presents an analysis of related work, section 3 is entirely devoted to the exposition of our method; the connectivity and local rule application are explained and illustrated. Section 4 compares our method with the proposal from W. Liu and S. Zhong [1]. Finally, Section 5 summarizes and presents some concluding remarks and future work.

2. Related work

In this section we specially pay attention to the recent works that model the worm propagation, specifically if they are inspired on the Kermack and McKendrick disease expansion model and operate in finite and discrete space and time. A theoretical model that characterizes the propagation dynamics of worms in smartphones is presented by S. Peng et al. in [18]. N smartphones (nodes) are randomly deployed on a square 2-D grid. The smartphones are in one out of five states, susceptible-exposed-infected-diagnosed-recovered (SEIDR). The model incorporates the spatial distribution of the population by use of the Von Neumann or the Moore neighbourhood. The transition rule is structured in five steps and each node collects the information of its neighbours in order to calculate the next state of the node. The model provides two important factors, the infection factor, in order to evaluate the degree of the spread of infected nodes, and the resistance factor, which measures the resistance degree of a node on infection from other nodes. In [19], Martin et al. apply the SIS model to study the impact of mobile malware on cell phones. They aim to help users prepare in the event of a future mobile malware epidemic in Washington DC metropolitan area by proposing some preventative measures. Feng et al. [11] proposed a time-delayed SIRS model which introduces two parameters: temporal immunity, variable infection rate and explore the impact of the variable infection rate on the scale of malware outbreak. In [20], Pradip De et al. model the process of compromise spreading from a single node to the whole network which has been modelled as a random graph. The approach is analysed in two scenarios: with or without node recovery, depending on whether infected nodes will be recovered by external measures such as key revocation, immunization, or not. In [21] W. Xia builds a theoretical model based on five compartments susceptible- exposed-infected-recovered-dormancy (SEIRD) for the Bluetooth and MMS hybrid spread mode. The simulation keeps good correlation with the theoretical analysis. This approach analyses the influence of the propagation parameters such as user gather density in groups, moving velocity of smart phone, the time for the worm to replicate itself and provides some feasible control strategies. In [22] B. K. Mishra and N. Jha propose a susceptible-exposed-infectious-quarantined-recovered (SEIQR) model for the transmission of malicious objects in computer network. The model has a constant recruitment of nodes and an exponential natural and infection-related death (crashing) of the nodes. Thresholds, equilibria, and their stability are also found with cyber mass action incidence and the effect of quarantine on recovered nodes is analysed. The simulation of the system is numerically solved. In [23], F. Wang et al. propose a novel epidemic model which combines both vaccinations and dynamic quarantine methods (SEIQV). Using this model, the authors obtain the basic reproduction number that governs whether or not a worm threat is extinct. The impact of different parameters on this model is studied. Simulation results show the model has the capability to decrease the number of infected hosts and reduce the worm propagation speed. In [24], A. Martin del Rey and G. Rodríguez consider the case of the malware propagated using Bluetooth connections, which infects devices in its proximity similar to a biological virus. They use a compartmental model where the mobile devices are classified into four types: susceptible-carriers-exposed-infectious (SCEI) and its dynamic is governed by means of a couple of two-dimensional cellular automata (CA). The simulation of the model allows to determine how a mobile malware might spread under different conditions. In [25] G. González García et al present a compartmental

epidemiological model to study the space-time propagation dynamics of Bluetooth worms. The model is based on CA. The model considers the local interactions between the smartphones and is able to simulate the individual dynamic of each device and the effect of mobility of their users on the infection propagation. In [26], Mickler et al. present a global stochastic CA paradigm. The interaction measure between the cells is a function of population density and Euclidean distance, and extends to include geographic, demographic and migratory constraints. This model improves both the traditional CA paradigm and the classic SIR model and facilitates optimal deployment of public health resources for prevention, control and surveillance of infectious diseases.

3. The method

As mentioned in Section 2, generic epidemic models are very useful to deal with the dynamics of malware propagation [27-29]. In this section we present the SEIRS epidemic model to analyse the propagation behaviour of the malware in mobile computing devices. Our space-time framework is based on a concrete connection pattern between nodes defined by both a particular neighbourhood which fixes the connection between devices, and a local rule which sets whether the link is infective or not.

3.1 Connectivity and local rule

In this work, we have considered three types of connectivity defined by a neighbourhood relationship on a square NxN-sized grid. Each cell in the grid represents a node. The Von Neumann neighbourhood is composed of a central cell and its four adjacent cells (4-neighbours, horizontal and vertical connection). The Moore neighbourhood is composed of a central cell and the eight cells surrounding it (8-neighbours, horizontal, vertical and diagonal connection). Finally, we have the "L" neighbourhood, moving two squares horizontally then one square vertically, or moving one square horizontally then two squares vertically.

Let R be a local Boolean rule implemented by a binary operation shown by Equation 1.

$$R : \{0,1\} \times \{0,1\} \to \{0,1\}$$
$$(x, y) \to R(x, y) = a_i \tag{1}$$

Without loss of generality we set: $R(0, 0) = a_3$, $R(1, 0) = a_2$, $R(0, 1) = a_1$ and $R(1, 1) = a_0$

We can define $2^4 = 16$ different local rules depending on the values of the sequence $a_3\, a_2\, a_1\, a_0$. Let R_m denote a particular rule, $m \in [0, 2^4-1]$. The binary representation of m is: $m = a_3\, a_2\, a_1\, a_0$.

In order to model the dynamics of SEIRS, we set:
$R(1, 0) = a_2 = 1$, an infected node (with value = 1) propagates infection to an exposed neighbour node (with value = 0), so this value turns to 1.
$R(1, 1) = a_0 = 1$, an infected node has no effect on another infected node.

$R(0, 1) = a_1 = 0$, a recovered node promotes the healing of an infected neighbour node, before it becomes susceptible again

$R(0, 0) = a_3 = 0$, exposed, or recovered or susceptible nodes (with value = 0) have no effect on exposed, or recovered or susceptible nodes.

So, the suitable local rule to depict this case is R_5. We also use colours to discern between susceptible and exposed:

Cells (nodes): yellow colour for susceptible, grey colour for exposed
Number (generation): red colour for infected, green colour for recovered.
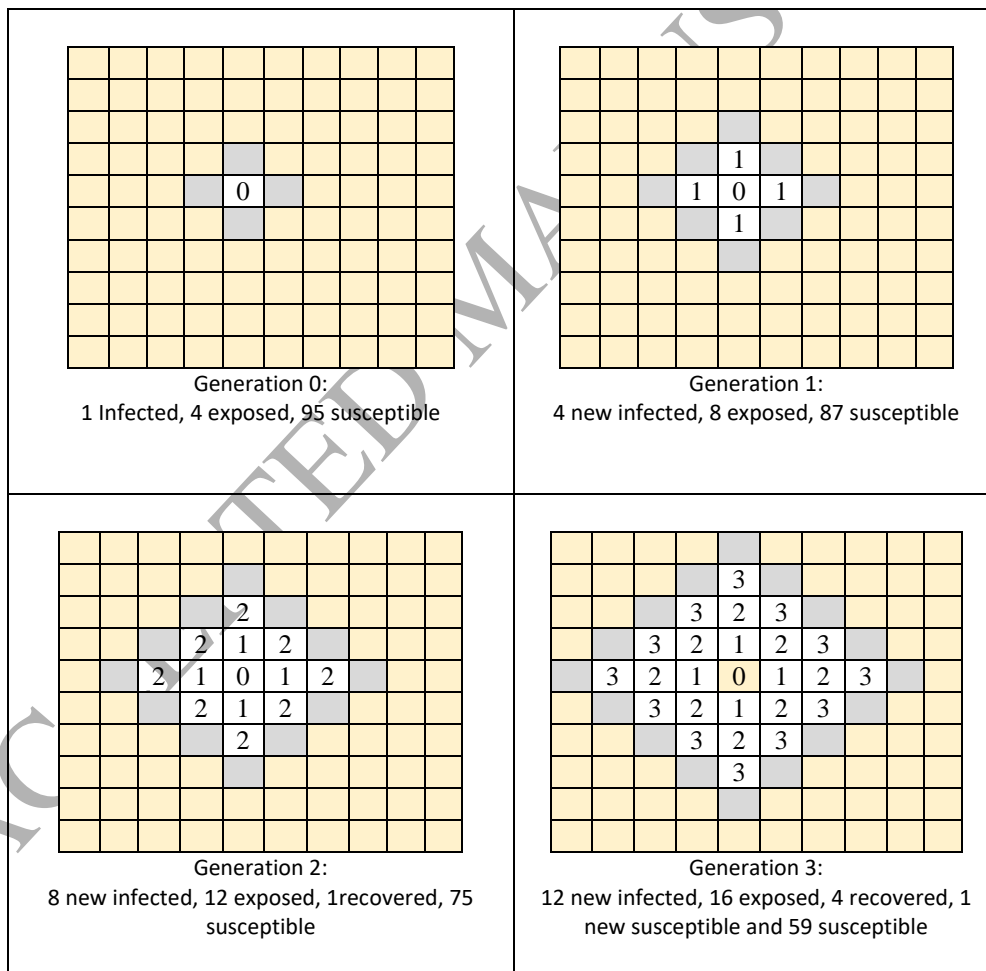We set:

- A node recovers two generations after it is infected
- A node is exposed when it is in the neighbourhood of an infected node.
- A node is able to infect again three generations after it is recovered, after being susceptible and exposed.

As follows, we apply the rule R_5 to the different connectivity patterns.

a) The Von Neumann neighbourhood

Figure 1 shows the evolution of the infection process of 100 nodes placed on a 10x10 grid, where 99 nodes are susceptible and the one at the centre is infected. The Von Neumann neighbours of the infected nodes are exposed and the rest are susceptible nodes. The number stands for the generation.

Generation 0:
1 Infected, 4 exposed, 95 susceptible

Generation 1:
4 new infected, 8 exposed, 87 susceptible

Generation 2:
8 new infected, 12 exposed, 1 recovered, 75 susceptible

Generation 3:
12 new infected, 16 exposed, 4 recovered, 1 new susceptible and 59 susceptible

8	7	6	5	4	5	6	7	8	9
7	6	5	4	3	4	5	6	7	8
6	5	4	3	2	3	4	5	6	7
5	4	3	2	1	2	3	4	5	6
4	3	2	1	0	1	2	3	4	5
5	4	3	2	1	2	3	4	5	6
6	5	4	3	2	3	4	5	6	7
7	6	5	4	3	4	5	6	7	8
8	7	6	5	4	5	6	7	8	9
9	8	7	6	5	6	7	8	9	10

Generation 12
8 new infected, 8 new exposed, 1 recovered, 12 exposed, 12 new second infected, 18 new recovered, 16 susceptible, 4 new susceptible

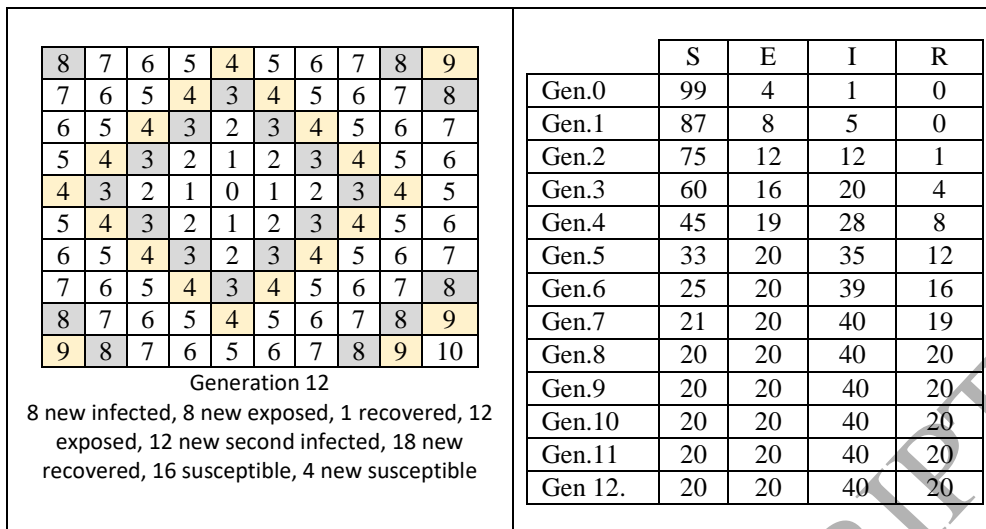|	S	E	I	R
Gen.0	99	4	1	0
Gen.1	87	8	5	0
Gen.2	75	12	12	1
Gen.3	60	16	20	4
Gen.4	45	19	28	8
Gen.5	33	20	35	12
Gen.6	25	20	39	16
Gen.7	21	20	40	19
Gen.8	20	20	40	20
Gen.9	20	20	40	20
Gen.10	20	20	40	20
Gen.11	20	20	40	20
Gen 12.	20	20	40	20

Figure 1. Schedule of the propagation of malware for a SEIRS model, implemented by R_5 and the Von Neumann neighbourhood (grid 10x10) (generations 4-11 are not shown)
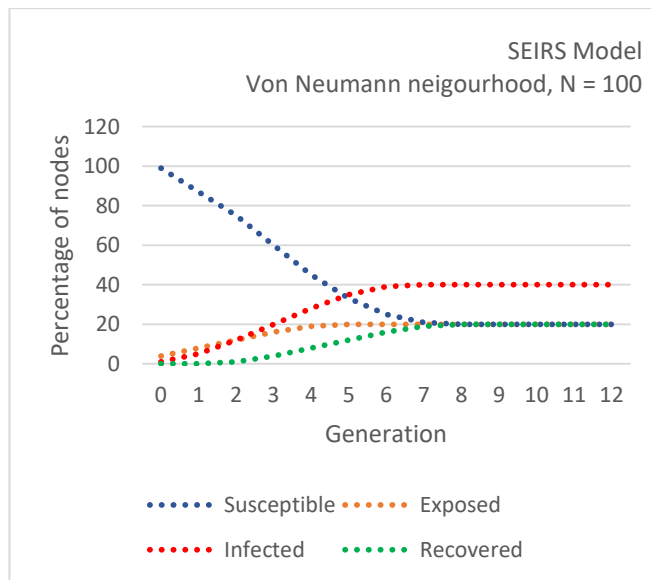

Figure 2. Evolution of the propagation of malware. SEIRS model for the Von Neumann neighbourhood and the local rule R_5 (N = 100)

For the Von Neumann neighbourhood, we observe an asymptotic behaviour for S, E, I and R plots. The number of Infected, Recovered and Exposed increases smoothly and reaches the equilibrium value from generation 7; the equilibrium value is equal to 40 nodes (40%) for I, and 20 nodes (20%) for the others. The number of Susceptible decreases smoothly and achieve the equilibrium value which is 40 nodes (40%).

b) The Moore neighbourhood

Figure 3 shows the evolution of the infection process of 100 nodes placed on a 10x10 grid, where 99 nodes are susceptible and the one at the centre is infected. The Moore neighbours of the infected nodes are exposed and the rest are susceptible nodes. The number stands for the generation.
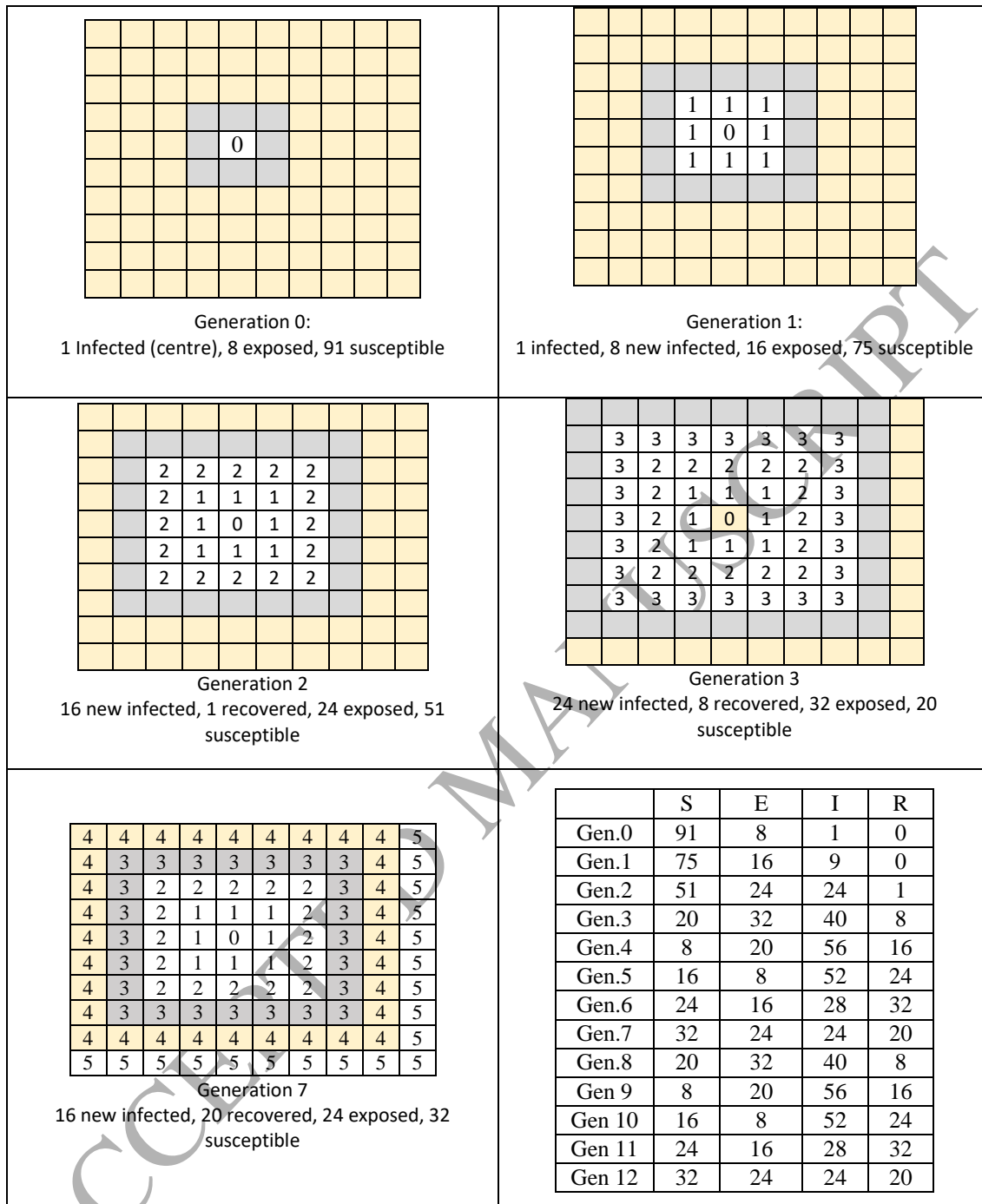
Generation 0:
1 Infected (centre), 8 exposed, 91 susceptible

Generation 1:
1 infected, 8 new infected, 16 exposed, 75 susceptible

Generation 2
16 new infected, 1 recovered, 24 exposed, 51 susceptible

Generation 3
24 new infected, 8 recovered, 32 exposed, 20 susceptible

Generation 7
16 new infected, 20 recovered, 24 exposed, 32 susceptible

	S	E	I	R
Gen.0	91	8	1	0
Gen.1	75	16	9	0
Gen.2	51	24	24	1
Gen.3	20	32	40	8
Gen.4	8	20	56	16
Gen.5	16	8	52	24
Gen.6	24	16	28	32
Gen.7	32	24	24	20
Gen.8	20	32	40	8
Gen 9	8	20	56	16
Gen 10	16	8	52	24
Gen 11	24	16	28	32
Gen 12	32	24	24	20

Figure 3. Schedule of the propagation of malware, related to the nodes state, for a SEIRS model implemented by R_5 and the Moore neighbourhood (grid = 10x10) (Generations 4-6 are not shown)
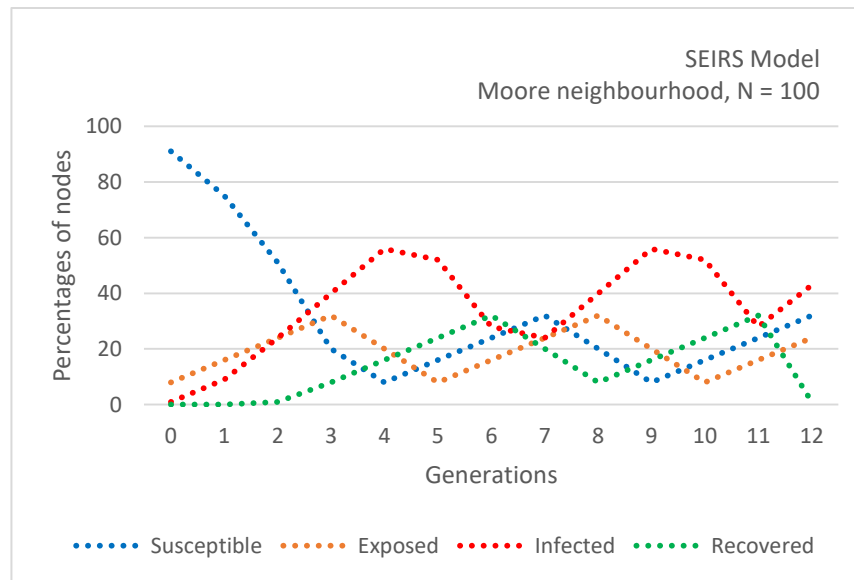
Figure 4. Evolution of the propagation of malware. SEIRS model for the Moore neighbourhood and the local rule R_5 (N = 100)

For the Moore neighbourhood, we observe a cyclic behaviour for S, E, I and R. The evolution of Susceptible, Recovered and Exposed is very similar from generation 5: they exhibit oscillations almost in phase (the phase difference between them is only one generation), with an average value of 20%, an amplitude of 10% and a frequency of 5 generations. The Infected behaviour has the same dynamics but the oscillation is out of phase related to Susceptible; the average value is almost 40%, the amplitude is 18% approximately and the frequency is 5 generations.

c) The L neighbourhood

Figure 5 shows the evolution of the infection process of 100 nodes placed on a 10x10 grid, where 99 nodes are susceptible and the one at the centre is infected. The L-neighbours of the infected nodes are exposed and the rest are susceptible nodes. The number stands for the generation.
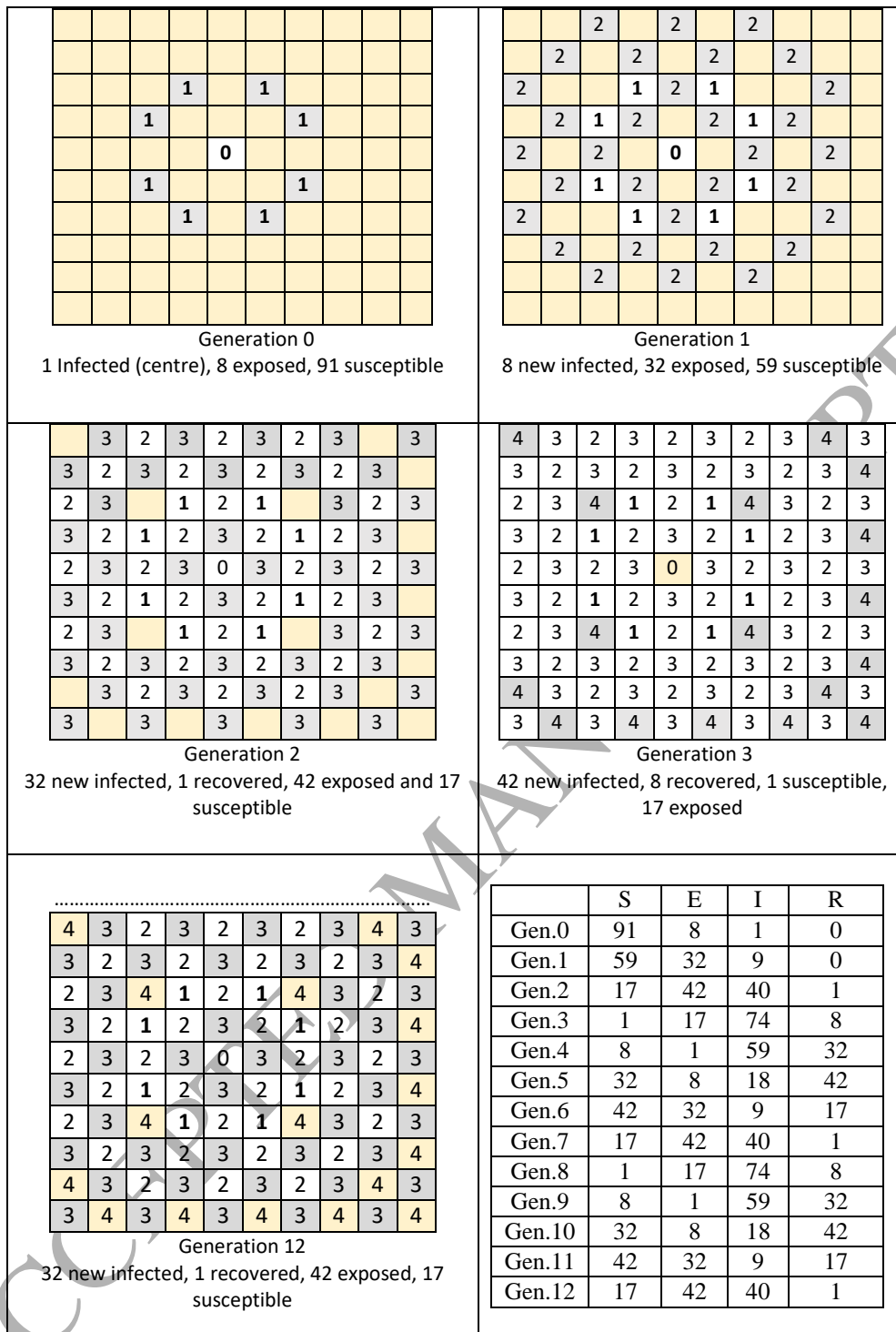
Figure 5. Schedule of the propagation of malware, related to the nodes state, for a SEIRS model implemented by R_5 and the L neighbourhood (grid 10x10) (generations 4-11 are not shown).
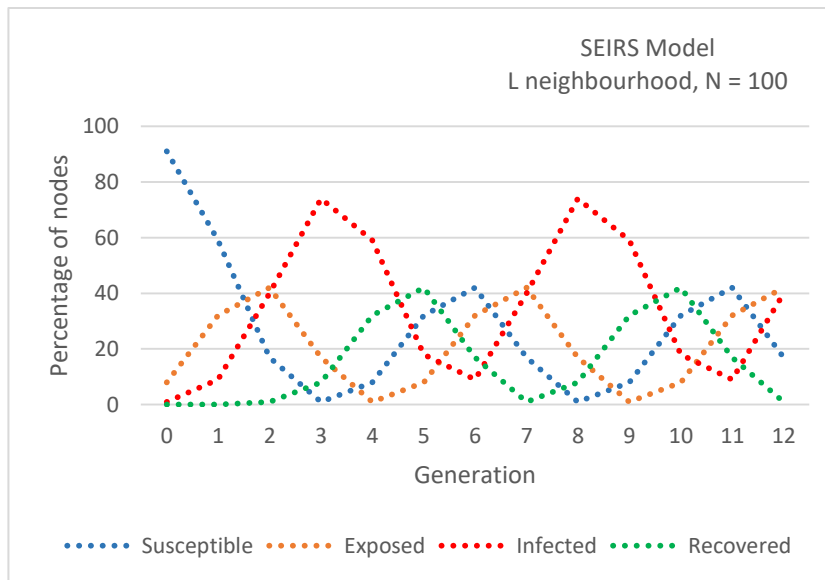
Figure 6. Evolution of the propagation of malware. SEIRS model for the L neighbourhood and the local rule R_5, (N = 100).

For the L neighbourhood, we also observe a cyclic behaviour for S, E, I and R. The evolution of Susceptible, Recovered and Exposed is very similar from generation 5: they exhibit oscillations almost in phase (the phase difference between them is only one generation), with an average value of 20%, an amplitude of 20% and a frequency equal to 5 generations. The Infected behaviour has the same dynamics but the oscillation is out of phase related to Susceptible; the average value is almost 40%, the amplitude is 35% approximately and the frequency is 5 generations.

As follows we present the time evolution of S, E, I and R when N=100, 2500 and 10000, for Von Neumann, Moore and L neighbourhoods. See Figure 7.

Evolution of the SUSCEPTIBLE nodes
N=10000
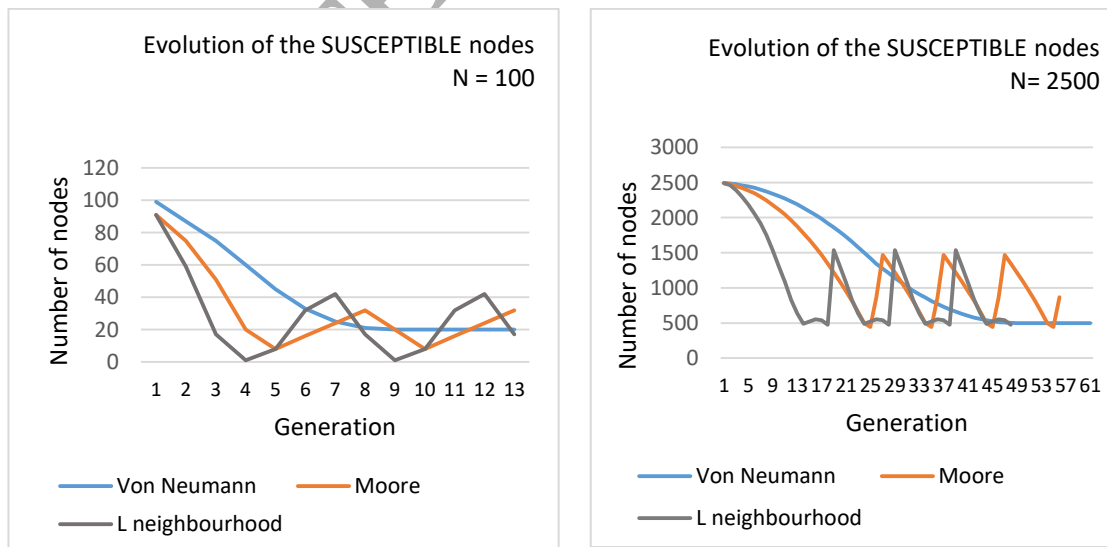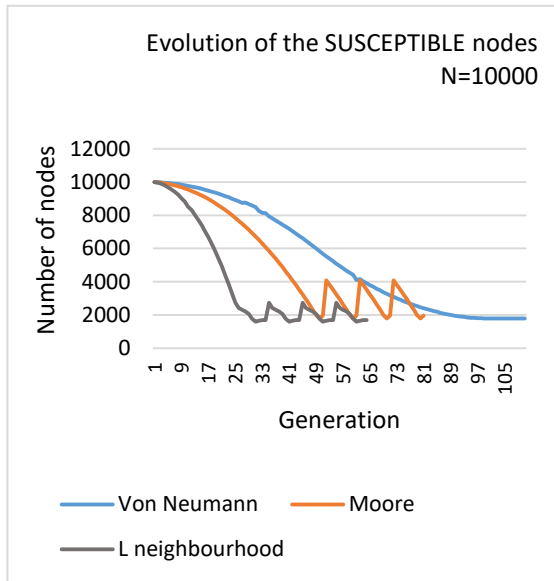
The number of susceptible nodes decreases initially when generations pass. The dynamics are quite different depending on the connectivity pattern.

For the Von Neumann neighbourhood, the decrease is continuous and very smooth, and the equilibrium value always reaches 20%, for any N.
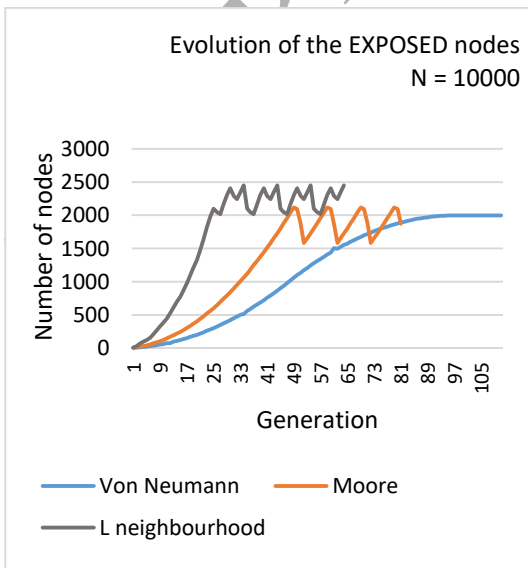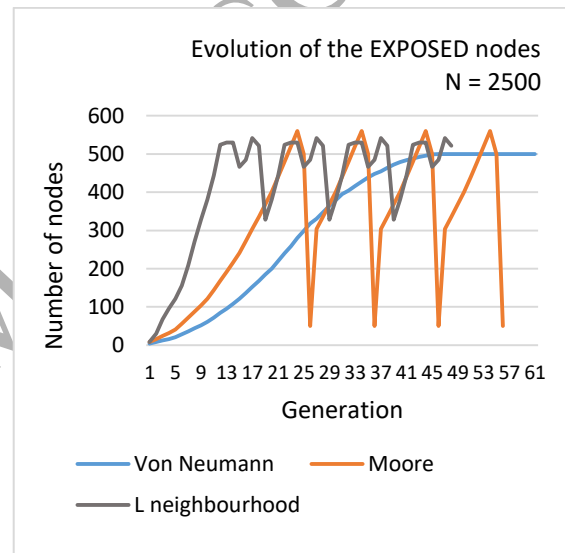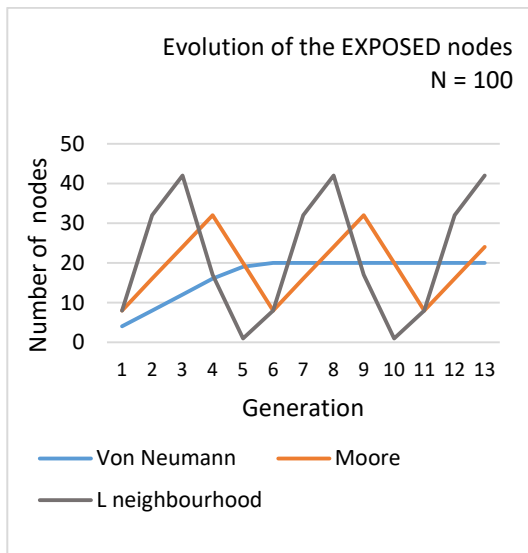
The number of susceptible nodes decreases more dramatically in the case of the Moore and L neighbourhoods and when the equilibrium value is reached an oscillatory regular cycle appears, with different average values:

Moore:
20% (N = 100), 40% (N = 2500), 30% (N = 10000)
L:
20% (N = 100), 40% (N = 2500), 20% (N = 10000)

Evolution of the EXPOSED nodes
N = 100

Evolution of the EXPOSED nodes
N = 2500

Evolution of the EXPOSED nodes
N = 10000

The number of exposed nodes increases initially when generations pass. The dynamic is quite different depending on the connectivity pattern.

For the Von Neumann neighbourhood, the increase is continuous and very smooth, and the equilibrium value always reaches 20%, for any N.
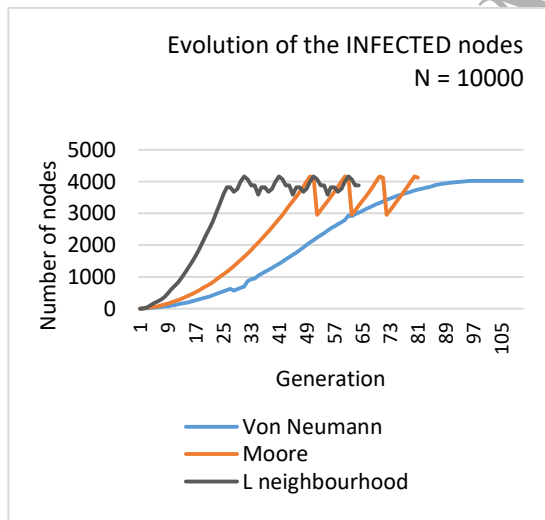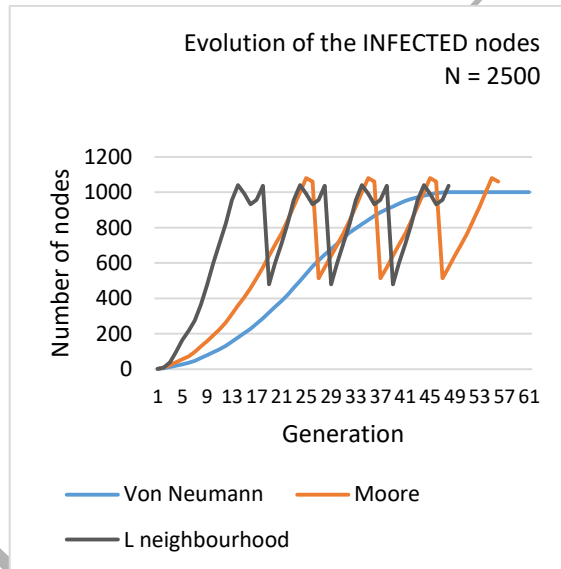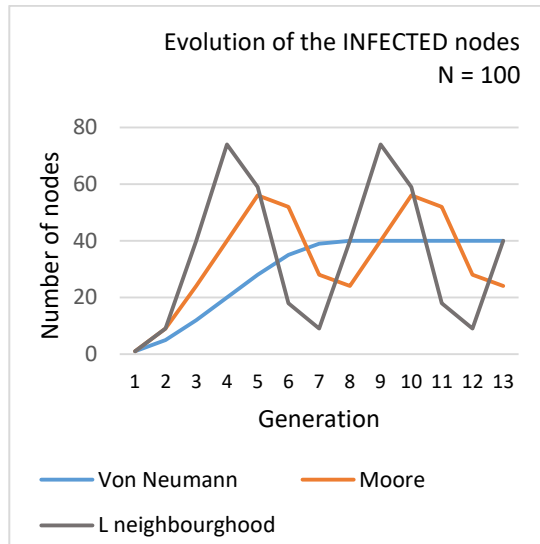
The percentage of exposed nodes increases more dramatically in the case of the Moore and L neighbourhoods and then, when the equilibrium value is reached, an oscillatory regular cycle appears with different average values:

Moore:
20% (N = 100), 12% (N = 2500), 17,5% (N = 10000)
L:
 20% (N = 100), 17,5% (N = 2500), 22% (N = 10000)

Evolution of the INFECTED nodes
N = 100

Evolution of the INFECTED nodes
N = 2500

Evolution of the INFECTED nodes
N = 10000

The number of infected nodes increases initially when generations pass. The dynamic is quite different depending on the connectivity pattern.

For the Von Neumann neighbourhood, the increase is continuous and very smooth, and the equilibrium value always reaches 40% for any N.

The percentage of infested nodes increases more dramatically in the case of the Moore and L neighbourhoods and when the equilibrium value is reached, an oscillatory regular cycle appears with different average values:

Moore:
40% (N = 100), 28% (N = 2500), 35% (N = 10000)
L:
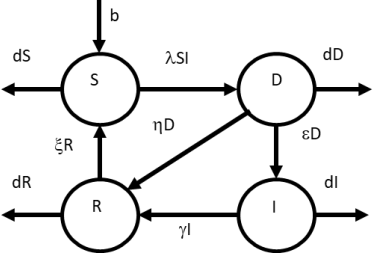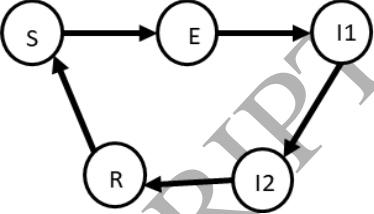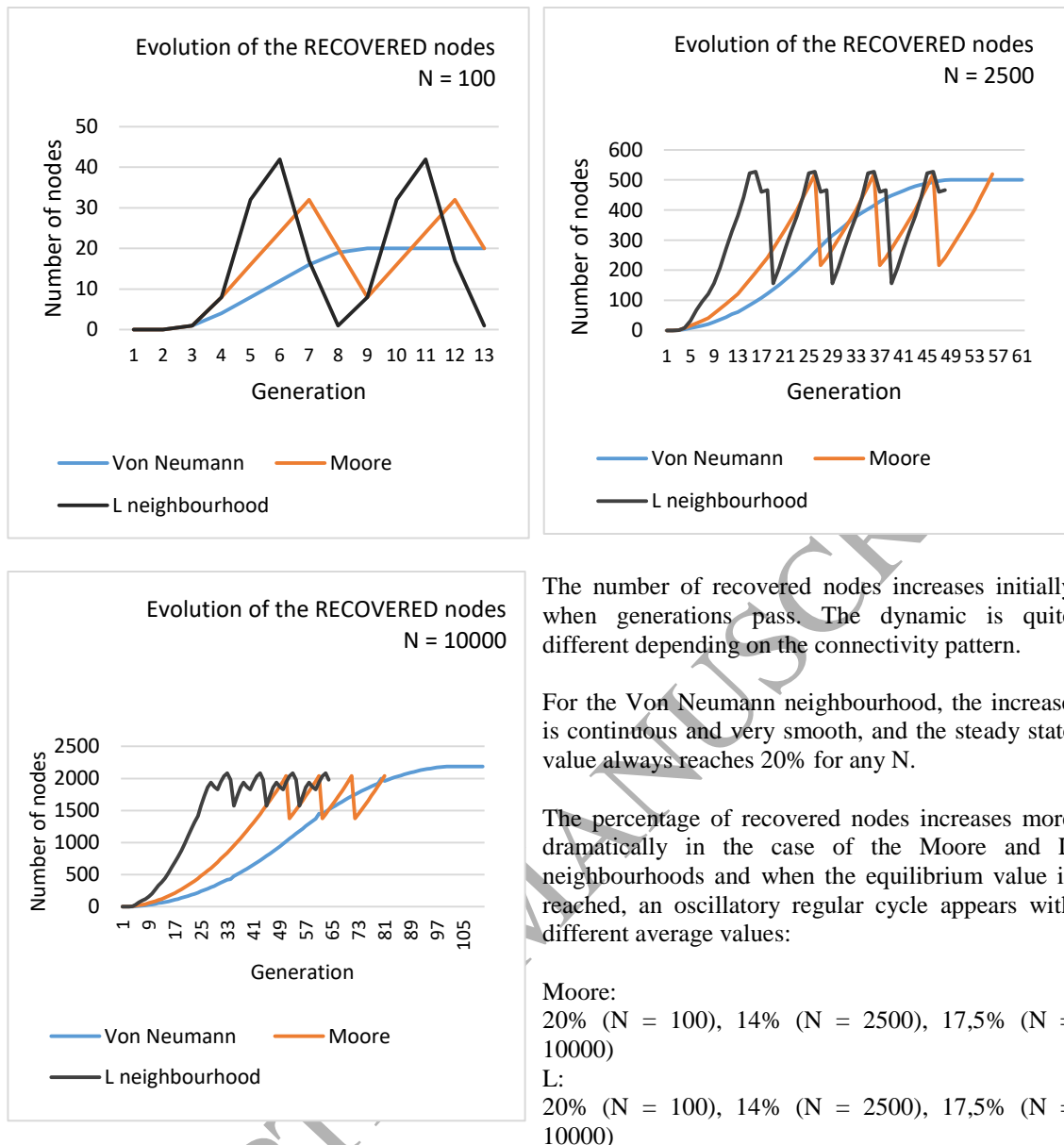40% (N = 100), 28% (N = 2500), 40% (N = 10000)

The number of recovered nodes increases initially when generations pass. The dynamic is quite different depending on the connectivity pattern.

For the Von Neumann neighbourhood, the increase is continuous and very smooth, and the steady state value always reaches 20% for any N.

The percentage of recovered nodes increases more dramatically in the case of the Moore and L neighbourhoods and when the equilibrium value is reached, an oscillatory regular cycle appears with different average values:

Moore:
20% (N = 100), 14% (N = 2500), 17,5% (N = 10000)
L:
20% (N = 100), 14% (N = 2500), 17,5% (N = 10000)
.

Figure 7. SEIRS Model. Evolution of S, E, I and R nodes for the three neighbourhoods, for N=100, 2500 and 10000

4. A comparative study

In this section, we compare the work by W. Liu and S. Zhong presented in [1] with ours. Although this work concretely aims to implement malware immunization strategies, we are now interested in the previous numerical simulations of the ODE system in order to provide an initial discussion about two different theoretical approaches. Table 1 presents a comparison between the statements of the two models.

Table 1. Comparison between the statements of the two models

	W. Liu and S. Zhong model	Our model

Computational tool Mathematical/	$\dfrac{dS(t)}{dt} = b - \lambda S(t)I(t) + \zeta R(t) - dS(t)$ $\dfrac{dD(t)}{dt} = \lambda S(t)I(t) - \eta D(t) - \varepsilon D(t) - dD(t)$ $\dfrac{dI(t)}{dt} = \varepsilon D(t) - \gamma I(t) - dI(t)$ $\dfrac{dR(t)}{dt} = \eta D(t) + \gamma I(t) - \zeta R(t) - dR(t)$	Grid NxN Neighbourhood + local rule Von Neumann (VN), Moore (M), L (L) These neighbourhoods define the interactions between nodes R_5 is the local Boolean rule, which fixes the result of the interaction between two neighbour nodes.
Dynamics	 S = susceptible D = delitescent I = infected R = recovered λ: how fast nodes move from being susceptible to delitescent ε: how fast nodes move from being delitescent to infected γ: how fast nodes move from being infected to resistant ζ: how fast nodes move from being resistant to susceptible	 S = susceptible (0) yellow cell E = exposed (0) grey cell I = infected (1) red number R = recovered (0) green number We assume The E compartment is equivalent to the D (delitescent) compartment of the ODE system.
Conditions	In order to make easier the comparison with our research, we set: b=d=0 (static network) η=0 (we do not consider the impact of the human decision).	* we assimilate the groups D and E, since they are almost equivalent. * the central cell in the grid is set to I * the neighbourhood cells of I cells are set to E * the other cells are set to S * the infection lasts 2 generations, states I_1 and I_2, (it could have been set otherwise) * the generation is a generic measure of time
Analysis	fourth-order Runge-Kutta algorithm	Eclipse IDE 2, Release (4.5.2)

As follows, Tables 2a-c present the results obtained by the two models (ODE and ours) for different population size (nodes) and different connectivity patterns.

Table 2-a. Comparison of the results obtained by the two models for N = 100 nodes

N = 100			
Neighbourhood	Von Neumann	Moore	L
Initial conditions	S = 95; E = 4; I = 1; R = 0	S = 91; E = 8; I = 1; R = 0	
	S and I curves intersect at generation 5 (value = 35 nodes). For generation > 7 the equilibrium is achieved	S and I curves intersect at generation 3 (value = 40 nodes). For generation > 5	S and I curves intersect at generation 2 (value = 40 nodes). For generation > 5 S, E, I and R exhibit oscillations

	The equilibrium values are: I ≈ 40 nodes (40%) S, E and R ≈ 20 nodes (20%)	S, E, I and R exhibit oscillations (cycle = 5 generations) S, E and R are almost in phase. average value = 20 nodes (20%), amplitude =12 nodes (12%) S and I out of phase average value = 40 nodes (40%) amplitude value = 16 nodes (16%)	(cycle = 5 generations). S, E and R are almost in phase. average value = 20 nodes (20%) amplitude = 20 nodes (20%) S and I out of phase average value = 41 nodes (41%) amplitude value = 33 nodes (33%)
ODE system (λ = 0,9; ε = 0,5; γ = 0,2; ζ = 0,6)			
Initial conditions	S = 95; D = 4; I = 1; R = 0	S = 91; D = 8; I = 1; R = 0	
	S and I curves intersect at t = 12 (value = 35 nodes). For t >16 the equilibrium is achieved (constant values for S, D, I and R). The equilibrium values are: I ≈ 43 nodes (43%) S, D and R ≈ 20 nodes (20%)	S and I curves intersect at t = 9 (value = 35 nodes). For t >14 the equilibrium is achieved (constant values for S, D, I and R). The equilibrium values are: I ≈ 43 nodes (43%) S, D and R ≈ 20 nodes (20%)	
Observations	For N = 100 *The comparison between the Von Neumann approximation and the ODE solution for the same initial conditions show very similar results, although generation and time are different units. *The Moore and L approximations have almost the same results, the only difference is the amplitude of the oscillation of S and I that is greater in L than in the Moore neighbourhood. The main difference of these approaches with ODE system is the existence or not of oscillations, but the equilibrium value (average value) in the systems is the same, at least for I.		

Table 2-b. Comparison of the results obtained by the two models for N = 2500 nodes

N = 2500			
Neighbourhood	Von Neumann	Moore	L
Initial conditions	S = 2495; E = 4; I = 1; R = 0	S = 2491; E = 8; I = 1; R = 0	

	S and I curves intersect at generation 34 (value = 844 nodes). For generation > 43 the equilibrium is achieved The equilibrium values are: I ≈ 1000 nodes (40%) S, E and R ≈ 500 nodes (20%)	S and I curves intersect at generation 22 (value = 841 nodes). For generation > 25 S, E, I and R exhibit oscillations (cycle = 10 generations). E, I and R are almost in phase. average and amplitude: E ≈ 305 and 205 nodes (8,2%) I ≈ 796 and 283 nodes (11,3%) R ≈ 368 and 152 nodes (6%) S and I are out of phase average value of S ≈ 954 nodes (38,16%) amplitude of S ≈ 512 nodes (20,48%)	S and I curves intersect at generation 12 (value = 824 nodes). For generation >14 S, E, I and R exhibit oscillations (cycle = 10 generations). E, I and R are almost in phase. average and amplitude: E ≈ 429 and 101 nodes (4%) I ≈ 700 and 280 nodes (11,2%) R ≈ 339 and 156 nodes (6,24%) S and I are out of phase average value of S ≈ 1012 nodes (40,48%) amplitude of S ≈ 525 nodes (21%)
ODE system (λ = 0,9; ε = 0,5; γ = 0,2; ζ = 0,6)			
Initial conditions	S = 2495; D = 4; I = 1; R = 0	S = 2491; D = 8; I = 1; R = 0	
	S and I curves intersect at t = 22 (value = 875 nodes). For t >28 the equilibrium is achieved (constant values for S, D, I and R). The equilibrium values are: I ≈ 1100 nodes (40%) S, D and R ≈ 500 (20%)	S and I curves intersect at t = 19 (value = 875 nodes). For t >25 the equilibrium is achieved (constant values for S, D, I and R). The equilibrium values are: I ≈ 1100 nodes (40%) S, D and R ≈ 500 (20%)	
Observations	For N = 2500 *The comparison between the Von Neumann approximation and the ODE solution for the same initial conditions show very similar results, although generation and time are different units. *The Moore and L approximations have similar results but also have some differences: the Moore is slower than the L approximation and the amplitude of the oscillation of the E nodes is more important in Moore than in L neighbourhood. The main difference of these approaches with ODE system is the existence of oscillations, but the equilibrium value (average value) in the systems is almost the same, at least for S and I.		

Table 2-c. Comparison of the results obtained by the two models for N = 10000 nodes

N = 10000			
Nieghbourhood	Von Neumann	Moore	L
Initial conditions	S = 9995; E = 4; I = 1; R = 0	S = 9991; E = 8; I = 1; R = 0	

	S and I curves intersect at generation 70 (value = 3328 nodes). For generation > 94 the equilibrium is achieved The equilibrium values are: I ≈ 4000 nodes (40%) S, E and R ≈ 2000 nodes (20%)	S and I curves intersect at generation 45 (value = 3382 nodes). For generation > 51 S, E, I and R exhibit oscillations (cycle = 10 generations) E, I and R are almost in phase. average and amplitude: E ≈ 1781 and 340 (3,4 %) I ≈ 3535 and 584 (5,84 %) R ≈ 1708 and 332 (3,32%) S and I are out of phase average value of S ≈ 2984 (29,40%) amplitude of S ≈ 1880 (18,80%)	S and I curves intersect at generation 24 (value = 3292 nodes). For generation >29 S, E, I and R exhibit oscillations (cycle = 10 generations). E, I and R are almost in phase. average and amplitude: E ≈ 2235 and 219 (2,19 %) I ≈ 3873 and 286 (2,86 %) R ≈ 1830 and 255 (2,55%) S and I are out of phase average value of S ≈ 2174 (21,74 %) amplitude of S ≈ 566 (5,66%)
ODE system (λ = 0,9; ε = 0,5; γ = 0,2; ζ = 0,6)			
Initial conditions	S = 9995; E = 4; I = 1; R = 0	S = 9991; E = 8; I = 1; R = 0	
	S and I curves intersect at t = 25 (value = 4500 nodes). For t>28 the equilibrium is achieved (constant values for S, D, I and R). The equilibrium values are: I ≈ 4500 nodes (45%) S, D and R have almost the same equilibrium value, which is about 2000 (20%)	S and I curves intersect at t = 22 (value = 3500 nodes). For t>28 the equilibrium is achieved (constant values for S, D, I and R). The equilibrium value are I ≈ 4500 nodes (45%) S, D and R have almost the same equilibrium value, which is about 2000 (20%)	
Observations	For N = 10000 *The comparison between the Von Neumann approximation and the ODE solution for the same initial conditions shows very similar results, although generation and time are different units. * When N increases although the Moore and L approximations follow the same trend, they are less and less similar. The Moore is slower than the L approximation and the amplitude of the oscillation of the S nodes is more important in Moore than in L neighbourhood. The main difference of these approaches with ODE system is the existence of oscillations, and more, the equilibrium value (average value) are quite different, at least for S and I.		

From Table 2a-c, some general trends can be highlighted. In our model, the connectivity and the number of nodes are crucial parameters to the evolution of S, E, I and R. In the ODE system, the parameters λ, ε, γ and ζ are crucial to the evolution of S, D, I and R. A qualitative analysis leads to the following equivalences as shown in Table 3.

Table 3. Equivalences between the parameters of the two approaches

Behavioural pattern	ODE	Our model
S and I intersect	λ↑ and/or γ↓	always
S and I reach the equilibrium value without intersection	λ↓ and/or γ↑	never
the equilibrium value of I ↑	When γ↓	When N ↑
the equilibrium value of I ↓	When λ↓	When N ↓
the equilibrium value of S ↓	When γ↓	When N ↓
the equilibrium value of S ↑	When λ↓	When N ↑
S, D, I and R present oscillations	When ζ, ↓	Von Neumann: never oscillates

with increasing amplitude		Moore: always oscillates and L: always oscillate the amplitude of the oscillations ↑ when N ↑ (although it represents a lower percentage of the value) the amplitude of the oscillations is larger in Moore than in L neighbourhood.
Speeding up the intersection between S and I	When S ↓ and/or D ↑	When N ↓ Von Neumann is always the slowest and L is always the fastest in relation to the intersection of S and I

The comparison between the two different theoretical approaches shows the capability of our proposal to meet the behavioural patterns provided by the ODE in the case of intersection between S and I, that is to say for high values for λ and low values for γ. The best match is with the Von Neumann neighbourhood for any N. For N = 100, the matching could also be acceptable for Moore and L neighbourhood, at least for compartment I. For N = 10000, the matching could also be acceptable for L neighbourhood, for compartments I and S.

5. Conclusion

In this paper, we present a discrete time-space version of the SEIRS model of disease spread to address the expansion of malware in mobile computing devices. We have studied the tuning parameters and have compared the scope of our approach with the traditional ODE model. The results of this preliminary work corroborate the ability of our model to perform some behavioural patterns provided by the ODE system. As a future work we plan to improve our model. We may build more progressive local rules, by combining different neighbourhoods. We should consider the impact of the human awareness and decision and therefore some exposed nodes could recover before they become infected. This could be useful to envisage social tools to provide control capability to our model. Finally, it may be interesting to complete our study by a comparison between our model and the discrete-time SEIRS epidemic model in scale-free networks.

References

[1] W. Liu & S. Zhong. Web malware spread modelling and optimal control strategies. *Sci. Rep.* **7**, (2017).
[2] D. Ki-Aries & S. Faily. Persona-centred information security awareness. *Computers & security* 70 (2017) 663–674.
[3] W. Tounsi & H. Rais A survey on technical threat intelligence in the age of sophisticated cyber attacks *Computers & security* 72 (2018) 212–233.
[4] P. Burnapa et al. Malware classification using self organising feature maps and machine activity data. *Computers & security* 73 (2018) 399–410.
[5] M. Meisel et al. A taxonomy of biologically inspired research in computer networking. *Computer Networks,* vol.54, pp. 901-016, 2010.
[6] W.O. Kermack & A.G. McKendrick. A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. 115, 1927, pp. 700-721.
[7] S. Hosseini et al. On the Global Dynamics of an SEIRS Epidemic Model of Malware Propagation. Proceedings of the 7th International Symposium on Telecommunications (IST' 2014), pp.646-651.

[8] R. Isea & K.E. Lonngren. On the Mathematical Interpretation of Epidemics by Kermack and McKendrick. *Gen. Math. Notes*, Vol. 19, No. 2, December 2013, pp. 83-87.

[9] P. Van den Driessche. Deterministic Compartmental Models: Extension of Basic Models. Lecture Notes in Mathematics, Mathematical Epidemiology, Springer, 2008.

[10] S. Tang and B.I. Mark. Analysis of virus spread in wireless sensor networks: an epidemic model. Proceedings of the 7th International Workshop on Design of Reliable Communication Networks (DRCN, 2009) pp. 86-91.

[11] L. Feng et al. Dynamical analysis and control strategies on malware propagation models. *Applied Mathematical Modelling*, vol. 37 (16), pp. 8225-8236, 2013.

[12] M. V. José et al. A Discrete SEIRS Model for Pandemic Periodic Infectious Diseases. Advanced Studies in Biology, Vol. 4, nº 4, pp. 153 – 174, 2012.

[13] V. Karyotis, A. Kakalis and S. Papavassiliou. Malware-propagative mobile ad hoc networks: asymptotic behaviour analysis. *J. Comput. Sci. Tech*. 23(3) 2008, pp. 389-399.

[14] A. Fuster Sabater et al. Simulación de la propagación del malware: Modelos continuos vs. modelos discretos. Actas de la XIII Reunión Española sobre Criptología y Seguridad de la Información RECSI 2014.

[15] S. H. White et al. Modeling epidemics using cellular automata. *Appl. Math. Comput*. 186 (1) 2007, pp. 193-202.

[16] A. Martín del Rey et al. Mathematical modeling of the propagation of malware: a review. Security and Communications Networks, 2015, pp. 2561–2579.

[17] Y. Song and G. Jiang. Research of malware propagation in complex networks based on 1-D cellular automata. *Acta Phys*. Sinica 58 (9) 2009 pp. 5901-5908.

[18] S. Peng et al. Modeling the dynamics of worm propagation using two-dimensional cellular automata in smartphones. *Journal of Computer and Systems Sciences* 79 (2013), pp. 586-595.

[19] J. C. Martin et al. Modelling the spread of mobile malware. International. J. Comput. Aided Eng. Technol. 2(2), 2010, pp. 3-14.

[20] P. De, Y. Liu, and S.K. Das. Modeling node compromise spread in wireless sensor networks using epidemic theory. Proceedings of the 2006 International Symposium on World of Wireless, Mobile and Multimedia Networks. 2006. IEEE Computer Society.

[21] W. Xia et al. Commwarrior worm propagation model for smart phone networks. J. China Univ. Posts Telecomm. 15(2), 2008, pp.60-66.

[22] B. K. Mishra and N. Jha SEIQRS model for the transmission of malicious objects in computer network. Applied Mathematical Modelling. Vol. 34(3), pp. 710-715, 2010.

[23] F. Wang et al. Stability analysis of a SEIQV epidemic model for rapid spreading worms. Computers & Security, Vol. 29 ($) pp. 410-418, June 2010.

[24] A. Martin del Rey and G. Rodríguez Sánchez. A CA Model for Mobile Malware Spreading Based on Bluetooth Connections. Herrero Á. et al. (eds) International Joint Conference SOCO'13-CISIS'13-ICEUTE'13. Advances in Intelligent Systems and Computing, vol 239. Springer, 2014.

[25] G. González García et al. Worm Propagation Modeling Considering Smartphones Heterogeneity and People Mobility. Proceedings of the 2017 International Conference on Applied Mathematics, Modeling and Simulation (AMMS 2017), Advances in Intelligent Systems Research

[26] A. R. Mickler et al. Modeling infectious diseases using global stochastic cellular automata. *J. Biol. Systems* 13 (4) 2005 pp. 421-439.

[27] S. Peng, S. Yu and A. Yang. Smartphone Malware and Its Propagation Modeling: A Survey. IEEE COMMUNICATIONS SURVEYS & TUTORIALS, Vol 16, nº2 Second Quarter 2014

[28] M.T. Signes Pont et al. The Susceptible-Infectious Model of disease expansion analysed under the scope of connectivity and neighbour rules. Proceedings of the CSITA conference, Computer Science & Information Technology (CS & IT) 7 (1): 1-10 (January 2017).

[29] M.T. Signes Pont et al. A discrete approach of the Susceptible-Infectious-Susceptible (SIS) Model of Disease Expansion. International Journal of Computers, Volume 2, 2017.

Maria Teresa Signes Pont received the BS degree in Computer Science at the Institut National des Sciences Appliquées, Toulouse (France) and the BS degree in Physics at the Universidad Nacional de Educación a Distancia (Spain). She received the PhD in Computer

Science at the University of Alicante in 2005. Since 1996, she is a member of the Computer Science and Technology department of the same university where she is currently an associate professor and researcher. Her areas of research at the Specialized Processors Architecture Laboratory include computer arithmetic and the design of floating point units, approximation algorithms related to VLSI design and natural systems modelling.

Antonio Cortes Castillo is a computer engineer trained in the Latin American University of Science and Technology (ULACIT), Costa Rica, 1995. He obtained his bachelor's degree in computer engineering with an emphasis in Management Information Systems at the National University of Heredia, Costa Rica, 2002. Adquiere his Master's degree in Computer Science (Telematics) at the Technological Institute of Costa Rica. He is now teaching at the University of Panama and is with his PhD. at the University of Alicante.

Higinio Mora received the BS degree in computer science engineering and the BS degree in business studies in University of Alicante, Spain, in 1996 and 1997, respectively. He received the PhD degree in computer science from the University of Alicante in 2003. Since 2002, he is a member of the faculty of the Computer Technology and Computation Department at the same university where he is currently an associate professor and researcher of Specialized Processors Architecture Laboratory. His areas of research interest include computer modelling, computer architectures, high performance computing, embedded systems, internet of things and cloud computing paradigm. He has participated in many conferences and most of his work has been published in international journals and conferences, with more than 60 published papers.

Julian Szymański research interests falls into domain of cognitive science and information retrieval. As part of the research I develop algorithms to more accurately extract information from text. In my approach to this I use natural language processing tools and computational intelligence methods. Cognitive theories of human knowledge serve me as an inspiration for the implementation of the algorithms able to process the data in the form of text, in a way similar to how it would be done by humans. At the moment I'm working on the methods of semantic documents categorization. The results of the research are used to organize knowledge in Wikipedia.