Scientific Research Publishing

Communications and Network, 2023, 15, 120-143
https://www.scirp.org/journal/cn
ISSN Online: 1947-3826
ISSN Print: 1949-2421

A New Model for Spreading Malware over SMS Using Network Automata

Erick Iván Medina-Salas[1], Ana Lilia Laureano-Cruces[1,2,3], Ma. Elena Lárraga-Ramírez[1,4]

[1]Posgrado en Ciencia e Ingeniería de la Computación, UNAM, México City, México
[2]Departamento de Sistemas, UAM-A, México City, México
[3]Laboratoire Informatique d'Avignon, Avignon, France
[4]Instituto de Ingeniería, UNAM, México City, México
Email: e.ivan.ms@hotmail.com, clc@azc.uam.mx, mlarragar@iingen.unam.mx

How to cite this paper: Medina-Salas, E.I., Laureano-Cruces, A.L. and Lárraga-Ramírez, Ma.E. (2023) A New Model for Spreading Malware over SMS Using Network Automata. *Communications and Network*, **15**, 120-143.
https://doi.org/10.4236/cn.2023.154008

Received: October 5, 2023
Accepted: November 21, 2023
Published: November 24, 2023

Abstract

By the year 2026, it is estimated that the number of smartphone users in Mexico will be approximately 118.1 million. Each smartphone has the functionality of sending and receiving SMS (Short Message Service) messages, which pose a significant threat to all users, as it makes any device vulnerable to a malware attack. In particular, worm-type malware takes advantage of this means of communication in order to spread. Studying the dynamics of malware propagation can help understand and prevent massive contagion between mobile devices. In this work, a model based on Network Automata and compartmental epidemiological models is presented, aiming to simulate, analyze and study the spread of worm-like malware through sending SMS on smartphones.

Keywords

Cellular Automaton, Malware Spread, Complex Systems, Network Automaton

1. Introduction

The use of mobile devices, such as laptops, electronic tablets or smartphones, has surprisingly increased. Particularly, in Mexico, according to Statista [1], it is estimated that the total number of users who will have a smartphone in 2026 will be 118.1 million people. Each smartphone user has the short message service or simple message service, better known as SMS (for the acronym of the English Short Message Service), which allows the sending of short messages with a limited number of characters to other mobile phones. Many advantages SMS offers us, such as reading or sending messages at any time, are used by malware or malicious software writers.

There are different types of malware such as: viruses, worms, and Trojans, among others. Regarding mobile devices, worm-type malware is potentially dangerous, since it exploits the topology of the networks to which they are connected to spread, making SMS an important attack vector. Malware infection on a device can be as serious as the attacker intends, from information loss to information theft, which is detrimental to the victim (**Figure 1**).

The objective of this work is to present a new mathematical and computational model, based on individuals using Cellular Automata theory, compartmental epidemiology, and scale-free networks to analyze, understand, evaluate, and prevent attacks through worm-like malware between smartphones, with SMS-based [2]. Unlike other models developed in recent years, such as models based on differential equations and models based on stochastic processes, important features that determine the dynamics of worm spread are taken into account, such as the local interactions that arise between mobile devices.

2. State of the Art

Studying the behavior of any infectious disease has been of vital importance for health. Epidemiology is a discipline in charge of studying the distribution, frequency, as well as severity of health problems and what causes them [3]. Being able to predict behavior, as well as taking measures to minimize the damage caused by these diseases, is carried out by researchers in the area of medicine. Epidemiology in conjunction with mathematics has given better results to study any case of disease.

In 1926, Anderson Gray McKendrick published an article called: "Applications of mathematics to medical problems" [4], where he introduced a new continuous mathematical model to model the behavior of epidemics taking into account stochastic processes of infection and recovery. The model considered three states for the population, that is, in three compartments:

Figure 1. Spread of worm malware.

- **Susceptible:** denoted by S, they are individuals considered healthy that can be infected.
- **Infected:** denoted by I, they are individuals who have been infected and can spread the disease.
- **Recovered:** denoted by R, are individuals who have recovered from the disease, taking some measure against it.

The previous model was denoted SIR by the acronym of its states.

Later, in 1927 [5], 1932 [6], and 1933 [7], William Ogilvy Kermack in collaboration with McKendrick, wrote several documents entitled: "Contributions to the mathematical theory of epidemics", where they studied deterministic epidemiological models, based on differential equations.

Epidemiological models can be classified according to the type of tool or mathematical method used by each one [8] [9], as shown below:

- **Models based on differential equations (or deterministic models):** such as the Kermack and McKendrick model, were among the first models to be developed. They are represented by differential equations. The size of the susceptible and infected population is assumed to be a defined function of time.
- **Models based on stochastic processes (or stochastic models):** the population in this type of model is represented by stochastic processes, these models can be described by interrelationships of their probability distribution.
- **Models based on individuals (or spatio-temporal models):** these models strongly use the Cellular Automata theory, which is described in the next section. Basically, a large number of simple components with local interactions are assumed and they are capable of simulating complete systems in their space-time evolution process. Important features of the epidemiological models are summarized in Table 1.

Table 1. Comparative table of the different types of epidemiological models.

Type	Based on Differential Equations	Based on Stochastic Processes	Based on Individuals
Theory	Differential Equations	Markovian Processes	Cellular Automata
Time	Continuous	Continuous or discrete	Discrete
Status Individual	Continuous	Discrete	Discrete
Individual Interaction	No	No	Yes
Adaptive range	Individuals with random movement	Small number of individuals	Large number of individuals
Modeldescription	Differential equations	Markov chains in continuous or discrete time	Stochastic evolution rules

Models based on differential equations can describe the dynamic interrelationships between rates of change and population size. In addition to the fact that the mathematical theory of this type of model has been very well developed and they are suitable for making predictions. However, the local interactions that arise between individuals are not considered, being a macroscopic analysis, therefore, they cannot simulate the individual dynamics of each individual; In addition, they consider that all individuals should be infected homogeneously, an event that is not very close to reality.

Models based on stochastic processes are suitable for studying small populations. However, like the models based on differential equations, they do not consider local interactions.

Individual-based models become an important tool to study the space-time evolution of self-organizing systems and to characterize complex systems based on local evolution rules. In this way, it can cover the deficiencies that other types of models have, giving each individual their own characteristics.

Currently, a wide variety of mathematical models can be found to describe the spread of worm-like malware through SMS messages or some type of network (social or complex).

Table 2, shows a comparison of the most important works in this regard, most of the models available in the literature do not consider factors such as:

Table 2. Comparative table of the different types of epidemiological models available in the literature. Where DE means differential equations, SP means stochastic processes and BI means based on individuals.

Model	Theory	Individual differences	Network topology	Message review	Relationship between nodes	Device security
Signes [10]	BI	Yes	No	No	No	No
Xiao [11]	BI	Yes	No	Yes	Yes	Yes
Xiao [12]	DE	No	No	No	No	No
Xiao [13] [14]	DE	No	No	No	No	No
Jia [15]	BI	Yes	No	No	Yes	Yes
Zhang [16]	DE	No	Yes	No	No	No
Liu [16]	DE	No	Yes	No	No	No
Hosseini [17] [18] [19]	DE	No	Yes	No	No	Yes
Gan [20]	DE	No	No	No	No	No
Huang [21]	DE	No	Yes	No	No	No
Liu [22] [23]	DE	No	Yes	No	No	No
Selvam [24]	DE	No	Yes	No	No	No
Yun [25]	SP	No	Yes	Yes	Yes	No
Peng [26] [27]	SP	No	No	Yes	Yes	Yes

message-checking dynamics, security awareness, relationship between nodes or a topology to represent best form the network formed by the list of users or contacts. In addition, there is a message delivery time limit in SMS, if the mobile device has no signal for a long time, the message will never reach its destination. In this way, there are other factors that can influence the spread of malware and that it is important to take into account. Finally, a better representation of the social network (network formed between the contacts of each mobile device) would lead to a better representation of the propagation dynamics, this problem could be solved using a scale-free network or SFN (for its acronym in English Scale-Free Network).

3. Theoretical Framework

In this section, the theory used for the developed model is explained.

3.1. Cellular Automata

A Cellular Automaton (CA) is a discrete dynamic system in time and space, it is composed of cells and the evolution of the system depends on the previous state, that is, each cell will influence the following evolution. It is basically a type of finite-state machine that is capable of simulating complex systems efficiently and effectively. The state of each cell will define the general state of the system [28]. Therefore, local interactions are considered and, therefore, individual aspects, making it possible to simulate each individual in a very simple way.

A CA can be considered as a system composed of a cell array *A*. Each cell c_i represents a finite automaton with a set of states *Q*, in an alphabet Σ and a transition function $\delta : Q \times \Sigma \to Q$. The input of the alphabet Σ is given by all the possible combinations of the state of the adjacent cells of each cell. Each cell c_i and its adjacent cells are considered all together and represented as the unique set $N = \{c_i\} \cup N^{-c_i}$ where N^{-c_i} is the set of cells considered as neighbors of an arbitrary cell c_i. Then, the Cellular Automaton can be defined as a 4-tuple $M = \{A, Q, \delta, N\}$ [29].

The components of a CA are:

- **Cellular space:** it is the physical space where the automaton evolves, being an arrangement, it has dimension *D* or infinite.
- **Transition function:** this function can be a rule or a set of rules that will define an evolution in time, causing a change of state in the system, so the set of rules δ define the dynamics of the CA. Given a cell *i* in a state k_i and the set of states of the neighbors *N* of *i*, at an instant of time *t*, δ is deterministic and calculates the next state of time $t + 1$ for cell *i*.
- **Time variable:** the dynamics of the cellular system develops over a discrete time.
- **Neighborhood:** are the cells that will influence the evolution of a cell. So, to each cell *i* it is necessary to assign a set of cells called neighbors, including itself.

- **Border:** in the case that the cellular space is finite, the border is the condition in which the cells that are in the perimeter of the cellular space are found.
- **States:** is the set of possible states Q in which the system can be, in particular the cells, at time *t*.

From the definition of CA, modifications have been made to them such as modifying the cellular space, thus emerging a new theory of CA. In particular, if we change the cell space from an array to a graph, we get a new automaton called the Network Automaton.

3.2. Graphs and Network Automata

Being able to model systems is a really complicated task, but when trying to model complex systems it is even more complicated. A complex system is one that is composed of a variety of entities, including another system or another individual. Many systems have been represented through graphs, remembering that a graph G can be represented through a tuple (V, E) where V is the set of nodes and E is the set of edges. Due to the need to represent complex systems, modifications have been made to the so-called simple graphs, obtaining graphs such as: directed graphs, labeled graphs, multigraphs, hypergraphs, among others. Many combinations of graphs have resulted in what are called complex networks.

A complex network is a network, represented by a graph, which has statistics and specific characteristics in its structure or topology. Some examples of these types of networks are:

- **Scale Free Networks (SFN):** they are graphs where the degree of their nodes follows a Power Law distribution [30], that is, most nodes are connected to few nodes while the minority is highly connected to many nodes. The distribution can be described as follows: $P(k)=k^{-\gamma}$, where $\gamma > 0$. According to Barabási Albert [31], we can build a scale-free network by executing his algorithm step by step. It has been found suitable to use an SFN to represent a social network [30] [32].
- **Small world networks:** they are graphs where the distance between any pair of nodes is relatively small while at the same time the level of transitivity or clustering is relatively high [33].

Then, a Network Automaton N can be defined as a 3-tuple (G, Q, $\left\{ f_i \middle| i \in N \right\}$) [34], where G is a graph in N, Q is a set of states and $f_i : Q^{|U_i|} \to Q$ a mapping, called the transition rule associated with vertex *i*. $U_i = \left\{ j \in N \middle| \left\{ j,i \right\} \in E \right\}$ is the neighborhood of *i*, *i.e.* the set of vertices connecting *i* and $|U_i|$ denotes the number of vertices joining U_i. The graph G is assumed locally infinite.

A Network Automaton follows the definition of a CA, as it differs only in cell space. Local interactions are still considered, that is when one node interacts with another, through the transition function. Once the transition function has been executed, it will indicate to which state it will evolve. Therefore, it continues to preserve the important characteristics of a CA, as well as simplicity.

4. Model

Once the necessary concepts regarding epidemiology have been reviewed, as well as works related to the spread of worm-type malware, a new model for the spread of malware through SMS messages is proposed, adding features that were not considered non-existent in previous works. they were raised as a whole as: individual differences (Cellular Automata), network topology (Scale-Free Network), message review activity, the relationship between nodes, and device security.

Thus, in this section, a new probabilistic model is described, based on the Network Automaton theory, with the aforementioned characteristics. The goal is to simulate the spread of worm-type malware by sending SMS messages to cell phones.

4.1. Model Formulation

The model consists of a Network Automaton, denoted as $N=(G,Q,F)$, where each node represents an agent that emulates a user and his device (smartphone) connected in the cellular network, denoted as *G*, each smartphone attribute varies over time. Each node will belong to a single state of the set of states, denoted as *Q*, at an instant of time *t*. The next state to which the device will belong will be defined by the transition function, denoted as *F*, which evaluates the parameters and current state of each node to obtain the next state. For practical purposes, it will be referred to in the following sections as a smartphone or device, both to the user and to their device.

4.1.1. Characteristics of the Cellular Network

The cellular network, formed by each node, will be formed according to the telephone directory of each device, so the cellular space can be defined as a directed graph. The outgoing degree of each node in *G*, that is, the number of edges that come out of each of them, will be defined by the Power Law distribution, converting the cellular space into an SFN. The neighborhood of each node will be defined by the edges connected to it, as follows: any node v_j that has v_i added to its contact list will be defined as an incoming neighbor of v_i; for node vi any node v_j that it has added to its contact list is defined as node v_i, outgoing neighbor.

4.1.2. Epidemiological States

It is necessary to define the set of epidemiological states *Q*. Each state will divide the population at an instant of time *t*, according to its condition against the disease, in this case, against the worm. Each node will belong to a single state at time *t*, the next state will be determined by the transition function.

Q is defined as the following set of states:

- **Susceptible (S):** nodes that have not been infected and are susceptible to receiving a message with a malicious link.
- **Waiting (W):** nodes to which an SMS message has been sent and are waiting

to receive it, this is due to the latency time of delivery of a message.

- **Exposed (E):** susceptible nodes that received an SMS message with a malicious link, but said message has not been opened or read. The exposed devices are not capable of starting a contagion process. If more than two messages are received at the same time, only the message from the source user with the highest degree of trust will be considered.
- **Vulnerable (V):** exposed nodes that opened a malicious message, but the malicious link has not yet been opened. Vulnerable devices are not capable of initiating infections.
- **Latent (L):** vulnerable nodes that clicked on the malicious link and downloaded the file, *i.e.* a copy of the worm, but for reasons of compatibility with the Operating System (OS), it cannot be executed and therefore cannot initiate an infection process. This is a terminal state.
- **Infected (I):** Vulnerable nodes that clicked on the malicious link, downloaded and executed the file, thus the worm is in operation and sends SMS text messages with malicious links to all your contacts inyour phone book. An infected node can only infect nodes in a susceptible state.
- **Recovered (R):** nodes in which the worm was detected and permanently removed from the device, through a vaccine thanks to the antivirus it carries. This is a terminal state.
- **Immune (Im):** Vulnerable nodes that become immune due to the user's security awareness. This is a terminal state.
- **Inactive (In):** infected nodes which have stopped sending malicious messages to their contacts due to an event that does not allow the device to continue sending SMS messages, however, it can resume sending messages if the event ends.

4.1.3. Transition Rules

The state transition diagram of N is shown in **Figure 2**, which only shows the state changes between each of them.

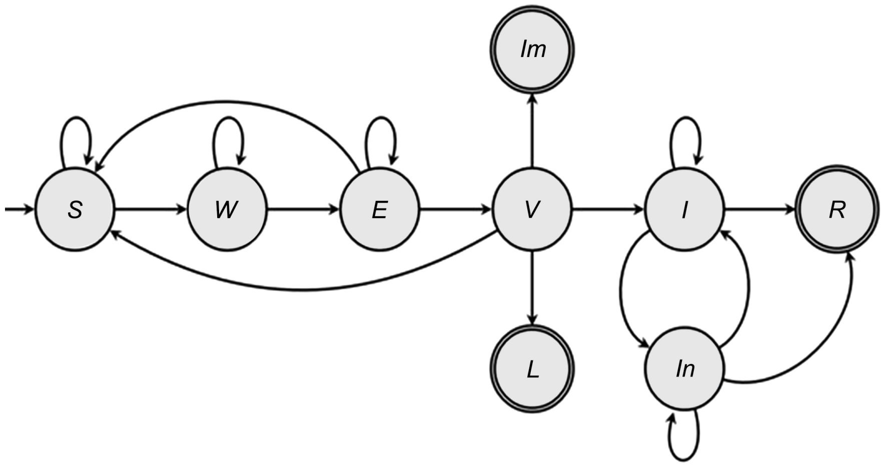

Figure 2. Transition diagram of the States of the Network Automaton.

The rules that define each change in the diagram are described below:

- A device *i* in state **S** will pass to state **W** at the next instant of time, if the probability P_{cont}, is fulfilled, which is denoted as follows:

$$P_{cont} = \beta \frac{I_i(t)}{N_i}(1 - P_F)$$

where:

- ○ β is the degree of infection;
- ○ $I_i(t)$ is the number of incoming neighbors of *i* infected at time *t*;
- ○ N_i is the total number of incoming neighbors of *i*, and;
- ○ P_F is the probability that a message will not reach its destination, due to an error in the cellular network.

- A device *i* in state **W** will pass to state **E** if it meets a time *TL*, which is the time it takes for an SMS message to reach its destination.
- A device in state **E** will go to state **V** if the user of the device decides to open and read the message, because he trusts the sending device that sent it, that is, a probability $DC(i, j)$ is met, such that $DC(i, j)$ is the degree of trust that devices *i* and *j* have in each other(as obtained in [26]). In case the message is never opened, that is, after a *TR* period the user who received the message never opened it, the device will return to state **S**.
- A device in state **V** has the possibility of clicking on the malicious link in the SMS message. There are four possibilities for this event: the device trusts the link, clicks and a copy of the worm is downloaded, affecting the OS, becoming infected (**I**), that is, a probability of $1 - CR_i$ is fulfilled and the OS of the worm is compatible with the OS of the device, such that CR_i is the degree of risk awareness of user *i*; that the device trusts the link and downloads the copy of the worm, but for reasons of compatibility with the OS it is not affected (**L**); that the user does not trust the link and does not click on the link, that is, it returns to a state **S**, and; that the device does not trust any suspicious links due to its high degree of risk awareness, leaving it in an **Im** state.
- A device in state **I** may move to a state **R** if the antivirus acts on the device thanks to a vaccine, that is, a P_R probability is met, or; the device suffered some event so it does not allow you to continue sending malicious messages to your contacts, leaving it in an **In** state, that is, a P_{In} probability is met.
- If a device is in state **In**, it will be a case similar to state **I**, since the device can send malicious messages again, if a P_I probability is met, that is, it recovers from the event that caused the inactivity, or; recover from the infection using the antivirus.

4.2. General Considerations

Listed below are some important considerations, which directly affect the propagation dynamics in the model:

- Only the sending of SMS messages is considered an infection vector.
- It is considered that the worm that infects the device will always have access

to the contact list.

- A population with different security characteristics is considered, that is, a heterogeneous population. Depending on the case study, the population can carry the same or different OS (Android, iOS).
- The size of the population remains constant over time.
- The size of each compartment or type of population, according to its state before the disease at each instant of time, is considered a whole quantity, not a fraction.
- Every second the transition function will be applied.
- It is considered that there is only one worm mounted on the server, so it affects devices with the same OS as the worm.
- Once a phone is recovered, that is, it was in an infected or inactive state and a vaccine was applied due to antivirus, or the user of the device took some action that reset it, the device will be protected against the worm so cannot be re-infected.
- **Devices in states:** Recovered, Immune and Latent, will remain in the same state respectively at the next time state, since they are terminal states.
- The initial number of infected devices is *only one*.

4.3. General Dynamics of the Model

The dynamics of the worm spread is executed as follows:

1) Devices are assigned to the cellular space G and their attributes are initialized, as well as individual and global parameters of each device in the cellular space, such as: the telephone directory, the registry of messages sent between devices, security awareness of the devices, among others.

2) The transition function of the automaton is executed, each device must collect the information of its neighbors in order to apply the transition function and obtain its next state after evaluating the function.

3) Variables for each device are updated for use in the next time state.

Steps 2 and 3 are executed in each evolution of the system until the criterion for stopping or ending the simulation is met. For our model, the stopping criterion is defined when the system is in an equilibrium state, that is, the number of devices in each compartment does not vary.

5. Analysis of Results

In this section, the simulation results obtained by varying the parameters with different values are presented to analyze the behavior of the system in response to them. What is sought is to emulate the performance of different types of malware and the effects they generate in the spread of the infection through SMS.

5.1. Information from the Simulations

For the different case studies, a population $|V|$ of 10,000 nodes, the population remains constant for all cases and with an initial number of infected nodes

$I(0)=1$. To establish the telephone directory of each node, that is, the way in which they will be connected in the network, a Power Law is followed, in such a way that the majority of the nodes will be highly connected, while the minority of them will be connected. with very few, thus building an SFN. To build the SFN, the Barabási Albert Graph algorithm was used, which is implemented in the Networks module [35] in Python.

The values of degrees of trust between nodes are obtained by filling out a message log, which indicates how many messages have been sent between two nodes in a period of one week, considering that the maximum number of messages that a node can send in one week is 30 and that this does not vary in subsequent weeks, in accordance with [36] [37]. Regarding cybersecurity awareness, to find out how aware the user is of the existence of malware, a normal distribution was used to establish its value. The risk threshold is set to a fixed value of 0.70.

The probability that a device goes to an inactive state is set to 0.01, this is considering that most devices remain mostly active over time. In the event that a device goes into the inactive state, it will return to the infected state again if a probability of 0.95 is met, since the user will do everything possible to keep the cell phone working in its entirety, as it happens in reality.

For the rest of the model parameters, the values used are specified in each case study. The dynamics of the system evolves in time steps of one second. For all the case studies presented, the results are the average of 10 simulations, which correspond to an average behavior of the system under study. In addition, the accumulated number of infected is considered as the sum of the devices in an Infected state and those in an Inactive state; since both are devices infected by the worm.

5.2. Infection Rate Variation

Table 3, describes the values of the parameters used for the simulations carried out by varying the infection rate. In particular, four values for the infection rate are considered: 0.25, 0.50, 0.75 and 1.0. The results obtained from the different simulations, taking into account the mentioned values, are shown in **Figure 3**. As can be seen in the image, as the value of the infection rate is higher, the malware infection spreads faster, as is to be expected since the probability of contagion is also.

5.3. Variation of Delivery Latency Time

According to [38], when SMS messages are transmitted there are different types of delay, which affect their delivery time. **Table 4**, describes the values of the parameters used for the simulations carried out, varying the TL message delivery latency time in three values: 1, 2, and 3. The results obtained from the different simulations taking into account the mentioned values are shown in **Figure 4**. As can be seen in the image, the message latency time, as its name indicates, generates a delay in the time required to infect the population; in such a way that the

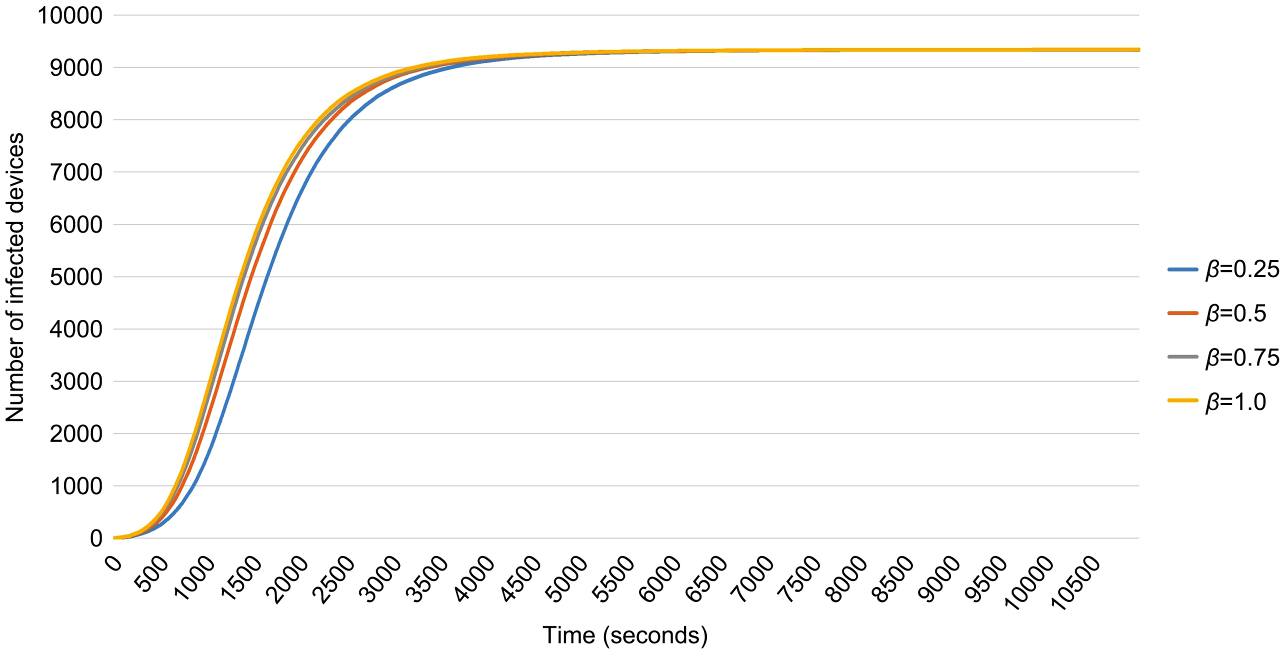

Figure 3. Comparison of number of infected devices over time due to vary β.

Table 3. Parameters used in the variation of the infection rate.

Parameter	Value
Infection rate (β)	0.25, 0.50, 0.75 and 1.0
Failure probability	0.05
Delivery latency time	3 s
Reading time	180 s
Risk awareness	average = 0.56
Probability of recovering	0.0
OS	Homogeneous population

Table 4. Parameters used in the variation of latency time.

Parameter	Value
Infection rate (β)	0.25
Failure probability	0.05
Delivery latency time	1, 2 y 3 s
Reading time	180 s
Risk awareness	average = 0.56
Probability of recovering	0.0
OS	Homogeneous population

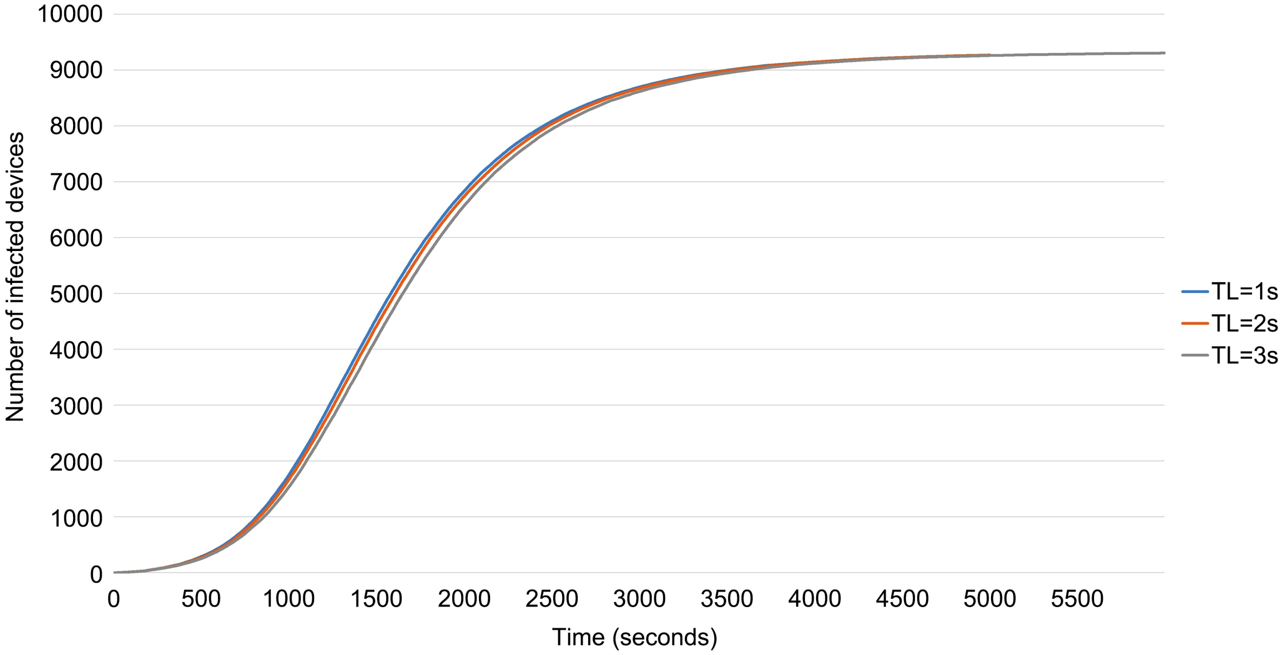

Figure 4. Comparison of the number of infected devices for different latency times *TL*.

greater the latency time, the time to infect the population of devices will also be greater. This is because a greater delay in the delivery of SMS messages will cause the propagation to be slower.

5.4. Variation of Message Delivery Failure

According to [38], there is a minimum loss of messages between a range of 5% to 20%. **Table 5**, describes the values of the parameters used for the simulations carried out, varying the message delivery failure P_F in four values: 0.05, 0.10, 0.15, and 0.20. The results obtained from the different simulations, taking into account the mentioned values, are shown in **Figure 5**. As can be seen in the image, as the probability of message delivery failure increases, the speed of propagation is lower, and therefore, the time required to infect the population is increased.

5.5. Variation of Risk Awareness

According to Statista [39], on average 56% of people are aware that malware exists and, therefore, are aware of the risk of clicking on a link. **Table 6**, describes the values of the parameters used for the simulations carried out, varying the risk awareness of the users, the following values are considered: 0.25, 0.50, 0.56 and 0.75. The results obtained from the different simulations, taking into account the values mentioned, are shown in **Figure 6**. As a higher percentage of the population is aware of the risk involved in clicking on a link, the population that can be infected by it is smaller. malware, even when there is no antivirus or other measure to prevent contagion. This implies that the spread of malware is slower. Particularly, when considering the case where 25% of the population is aware of the risk, the speed of spread is higher. On the contrary, when it is

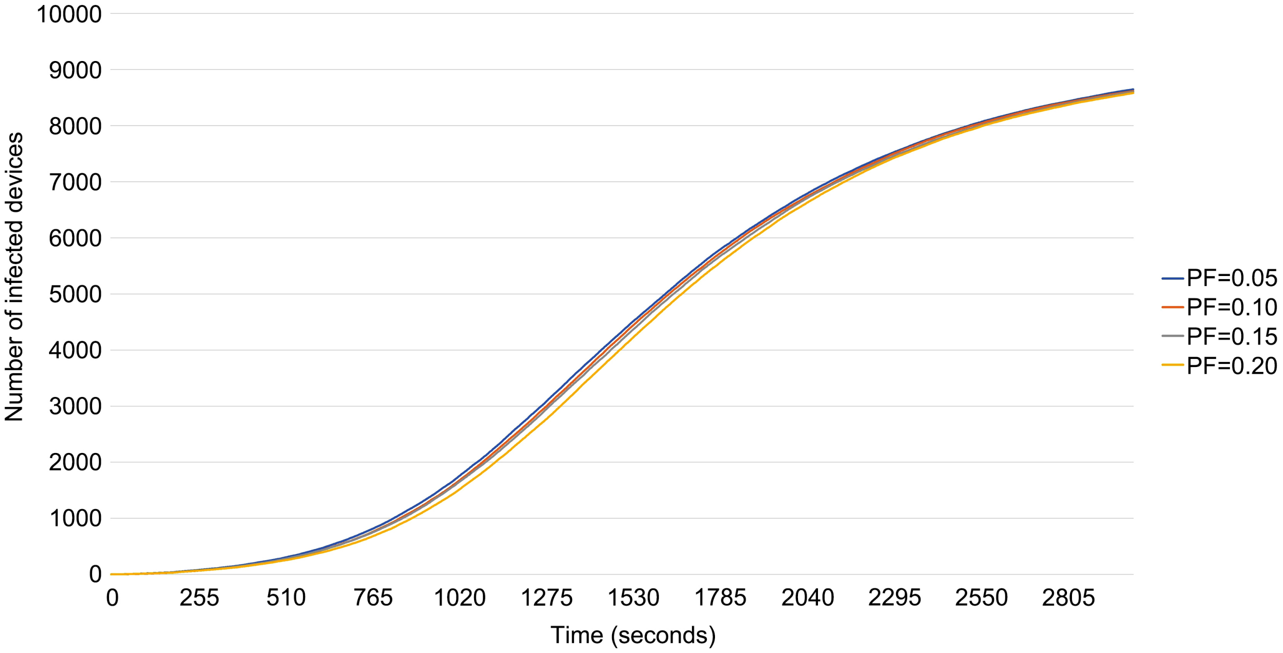

Figure 5. Comparison of the number of infected devices for different values of the probability of message delivery failure P_F.

Table 5. Parameters used in the message delivery failure variation.

Parameter	Value
Infection rate (β)	0.25
Failure probability	0.05, 0.10, 0.15 y 0.20
Delivery latency time	3 s
Reading time	180 s
Risk awareness	average = 0.56
Probability of recovering	0.0
OS	Homogeneous population

Table 6. Parameters used in the variation of security awareness.

Parameter	Value
Infection rate (β)	0.25
Failure probability	0.05
Delivery latency time	3 s
Reading time	180 s
Risk awareness	average = 0.25, 0.50, 0.56, 0.75
Probability of recovering	0.0
OS	Homogeneous population

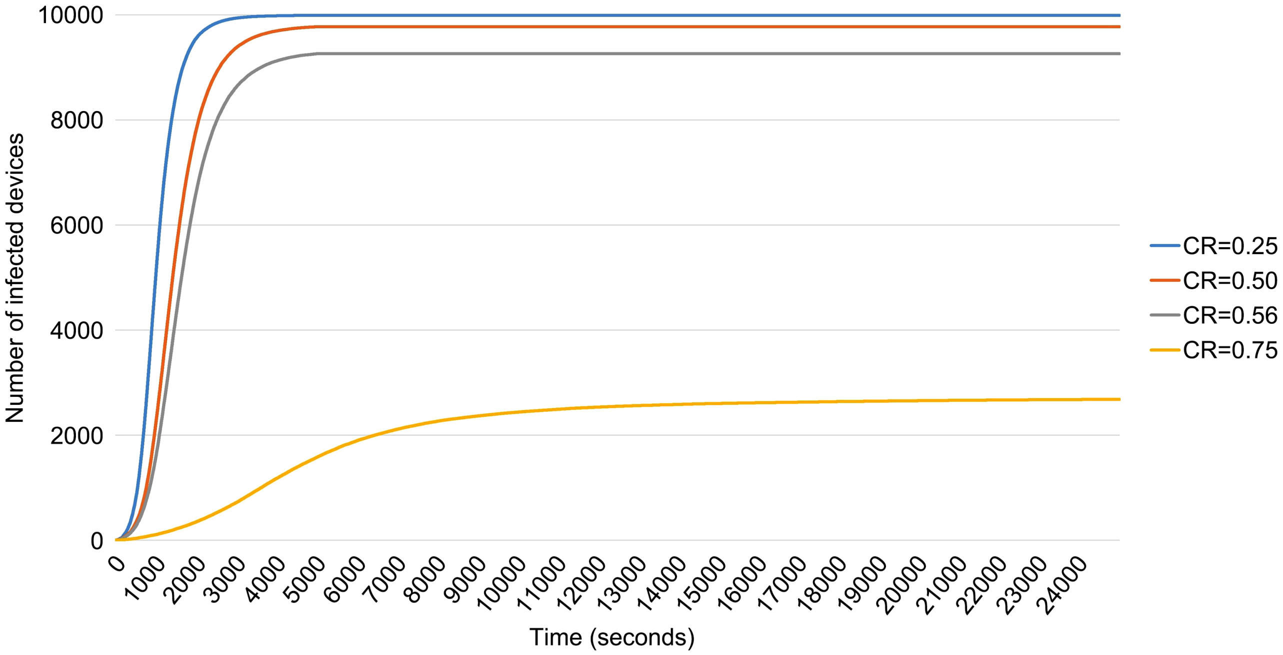

Figure 6. Comparison of the number of infected devices considering different values of risk awareness *CR*.

considered that 75% of the population is aware, the speed of spread is reduced, as well as the number of devices to be infected. On the other hand, a risk threshold *UR* is considered, such that if a risk-aware user *i* has a CR_i is greater than the threshold, it means that user *i* is highly aware, so he will not open links to malicious messages he receives. Therefore, if the user reaches or exceeds the *UR*, the person will not perform any action that puts their device atrisk. With greater risk awareness, the number of immune devices will grow, as seen in **Figure 7**.

5.6. Reading Time Variation

According to Tatango, 90% of users read a received text message within the first 3 minutes (180 seconds). **Table 7**, describes the values of the parameters used for the simulations carried out, varying the average reading time TR in four values: 90, 180, 270 and 360 seconds. The results obtained from the different simulations, taking into account the mentioned values, are shown in **Figure 8**. As can be seen in the **Figure 8**, a lower reading frequency implies a higher propagation speed, since users check their device more constantly. checking if there are new messages so the risk of infection increases.

5.7. Variation of the Probability of Recovery

Table 8, describes the values of the parameters used for the simulations carried out, varying the probability of recovery P_R in four values: 0.00025, 0.00050, 0.00075 and 0.0010. The results obtained from the different simulations, taking into account the mentioned values, are shown in **Figure 9** and **Figure 10** (infection and recovery curves). As can be seen in the figures, the lower the probability of recovery, the maximum number of devices to infect is greater and the speed of

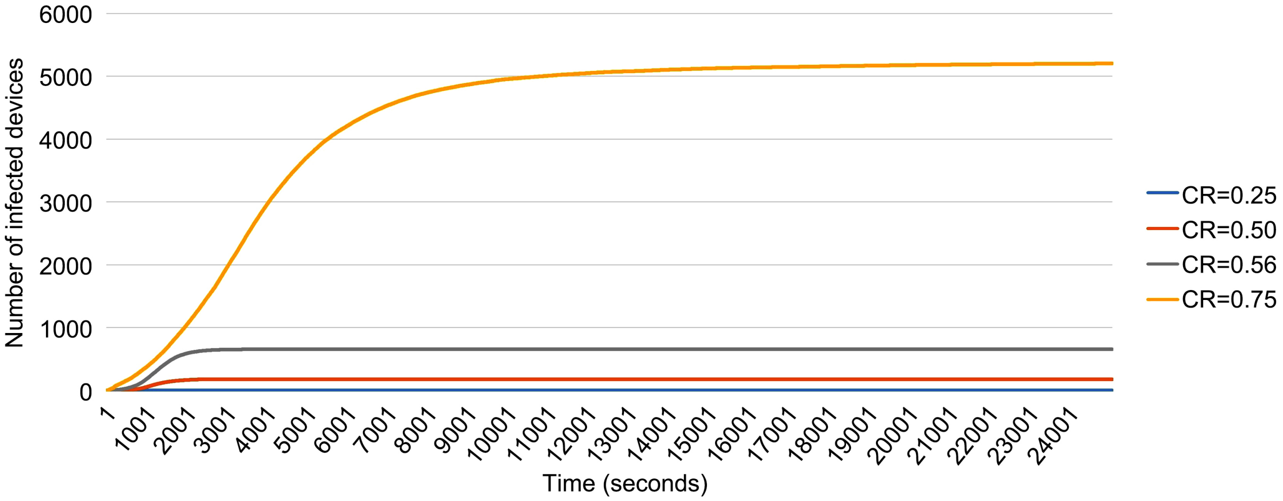

Figure 7. Comparison of the number of immune devices considering different values of risk awareness *CR.*

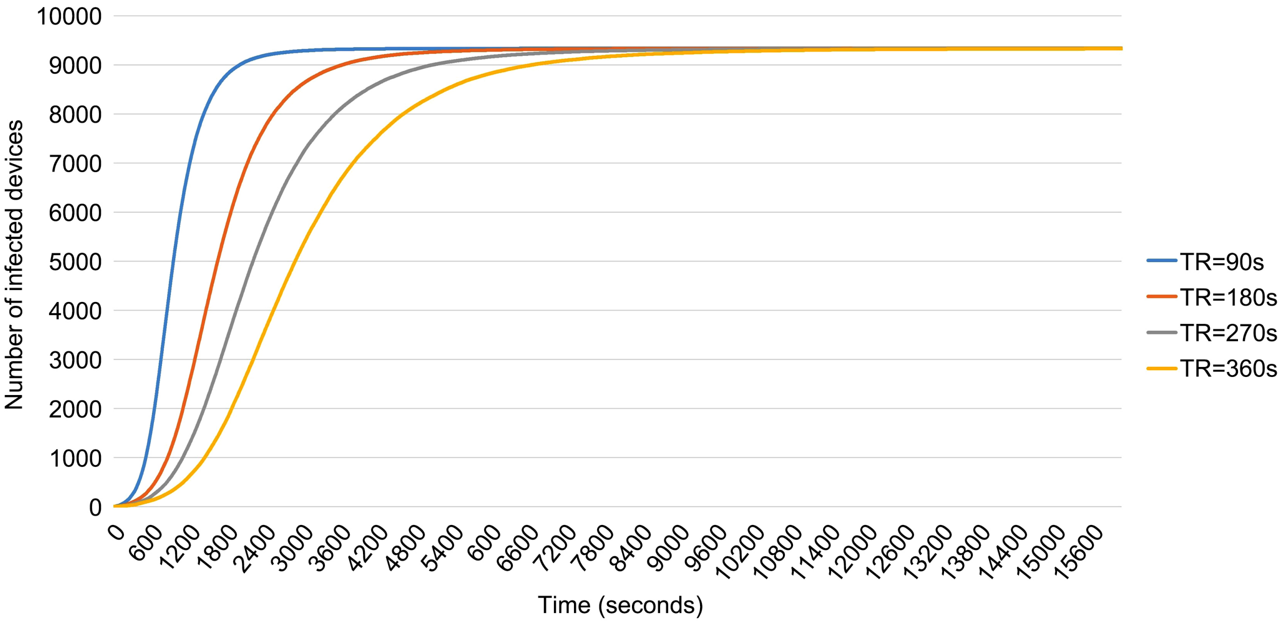

Figure 8. Comparison of the number of numbers of infected devices for different values of read time *TR.*

Table 7. Parameters used in the reading time variation.

Parameter	Value
Infection rate (β)	0.25
Failure probability	0.05
Delivery latency time	3 s
Reading time	90, 180, 270 y 360 s
Risk awareness	average = 0.56
Probability of recovering	0.0
OS	Homogeneous population

Table 8. Parameters used in the reading time variation.

Parameter	Value
Infection rate (β)	0.25
Failure probability	0.05
Delivery latency time	3 s
Reading time	180 s
Risk awareness	average = 0.56
Probability of recovering	0.00025, 0.00050, 0.00075 and 0.0010
OS	Homogeneous population

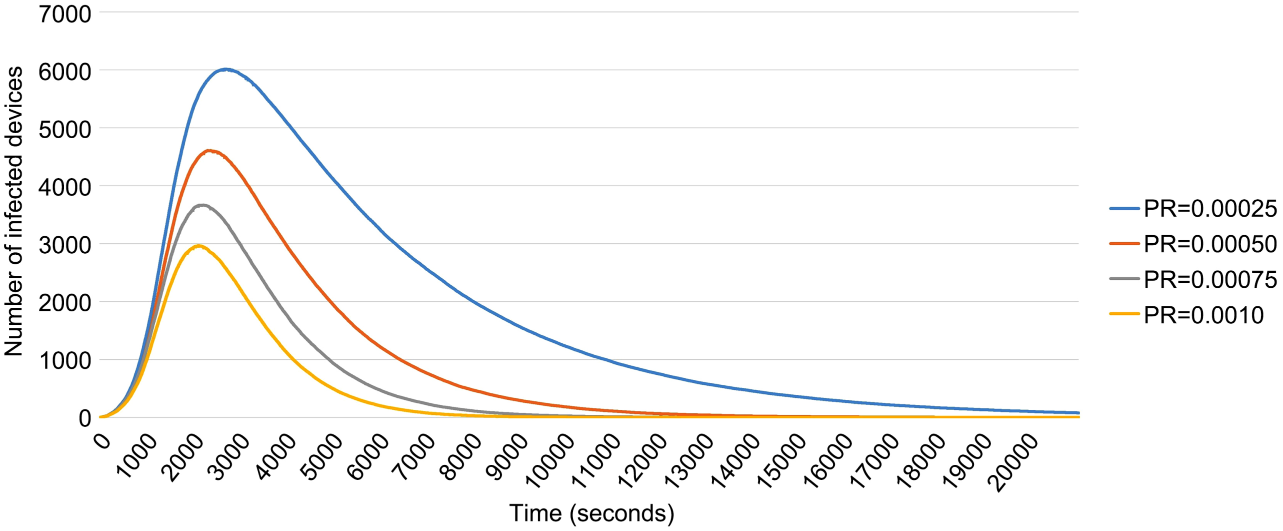

Figure 9. Comparison of the number of numbers of infected devices for different values of the probability of recovery P_R.

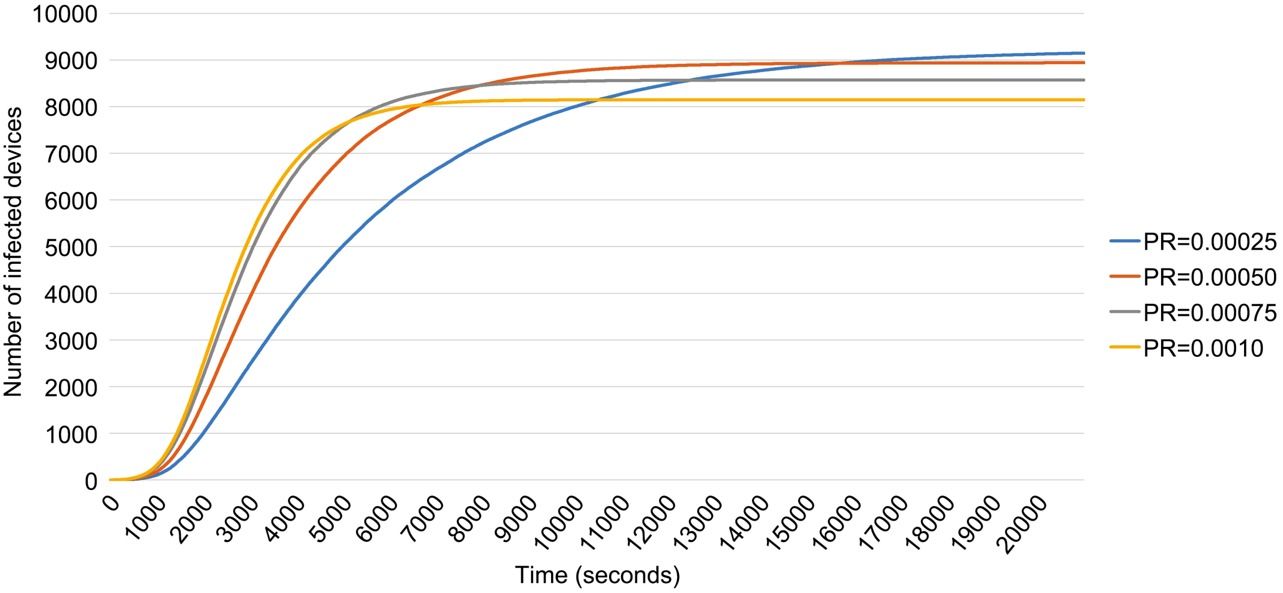

Figure 10. Comparison of the number of numbers of recovered devices for different values P_R.

spread is higher, since the time required to infect the entire maximum possible population increases depending on the number of devices to be infected. of this probability of recovery. As a consequence, also the recovery process of a device will be slower.

5.8. Variation of Degree of Confidence

The Degree of confidence (DC) or relationship indicates how much one user trusts another. Table 9, describes the values of the parameters used for the simulations carried out, considering four values of the probability of the degree of confidence: 0.20, 0.40, 0.60 and 0.80. The results obtained from the different simulations, taking into account the mentioned values, are shown in Figure 11. As can be seen, if users have a higher degree of trust among themselves, they will click on the malicious link sent more quickly, for therefore, the speed of propagation

Table 9. Parameters used in the variation of DC.

Parameter	Value
Infection rate (β)	0.25
Failure probability	0.05
Delivery latency time	3 s
Reading time	180 s
Risk awareness	average = 0.56
Probability of recovering	0
OS	Homogeneous population
Degree of confidence	P = 0.20, 0.40, 0.60 and 0.80

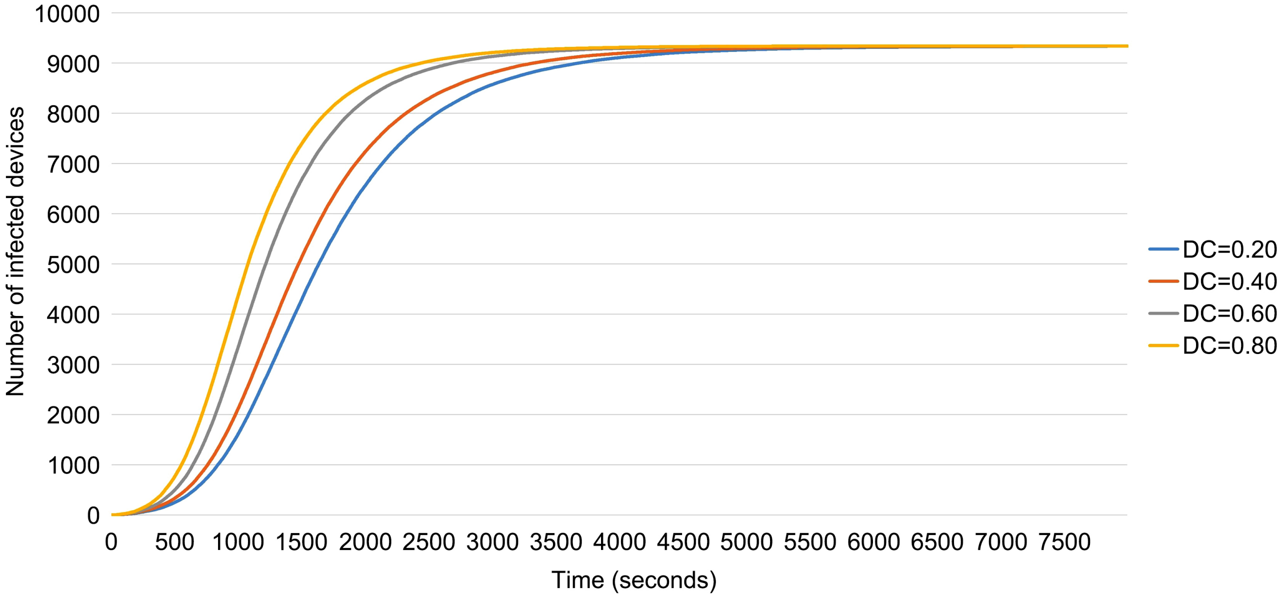

Figure 11. Comparison of the number of number of infected devices for different values of *DC*.

will be higher. This behavior is justified because users depend on the degree of trust to decide whether or not to read any message received from a contact. It is important to point out that although the network topology is formed by an SFN, the DC implies how many SMS messages were sent between two devices in a given period.

5.9. Variation of the Operating System

According to [40] [41], on average 84.4% of devices purchased between 2019 and 2020 have an Android-type OS, while on average 15.5 have an iOS-type OS, and less than 0.1% have a different OS than the previous ones. Table 10, describes the values of the parameters used for the simulations carried out, considering two types of device populations (Android and iOS). The results obtained from the different simulations, taking into account the mentioned values, are shown in Figure 12 and Figure 13. As can be seen in Figure 11, the number of

Table 10. Parameters used in the OS variation.

Parameter	Value
Infection rate (β)	0.25
Failure probability	0.05
Delivery latency time	3 s
Reading time	180 s
Risk awareness	average = 0.56
Probability of recovering	0
OS	84.4% Android, 15.5% iOS

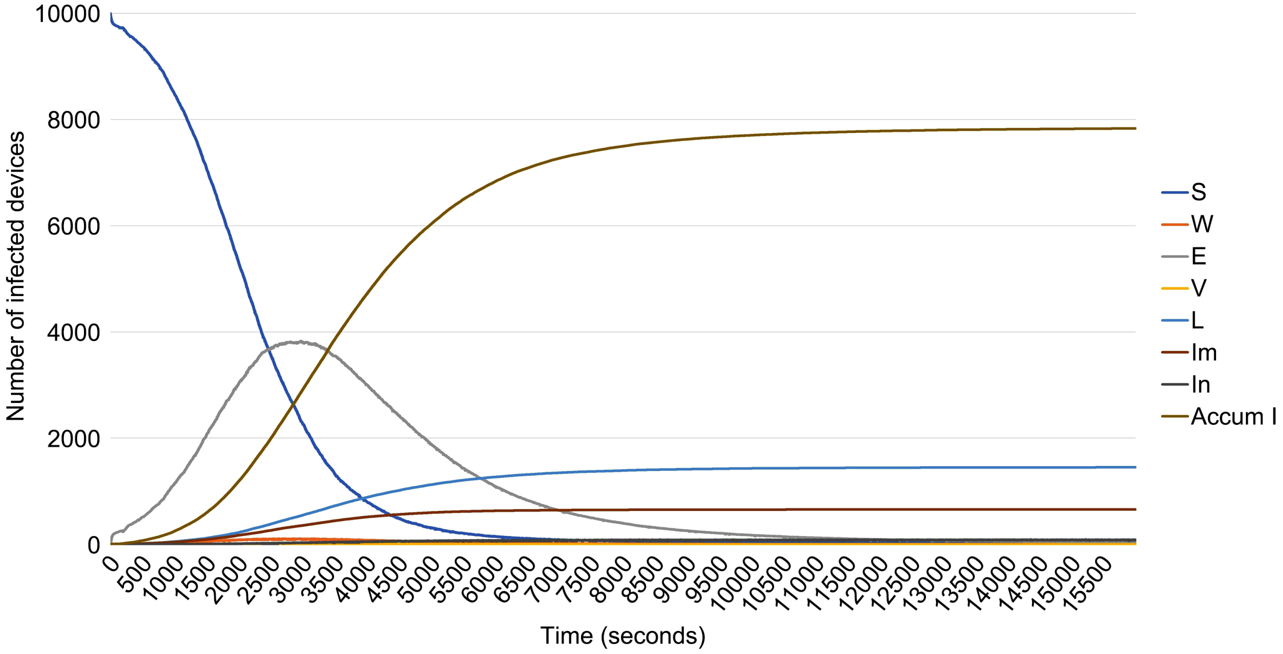

Figure 12. Curves of each compartment with Android OS.

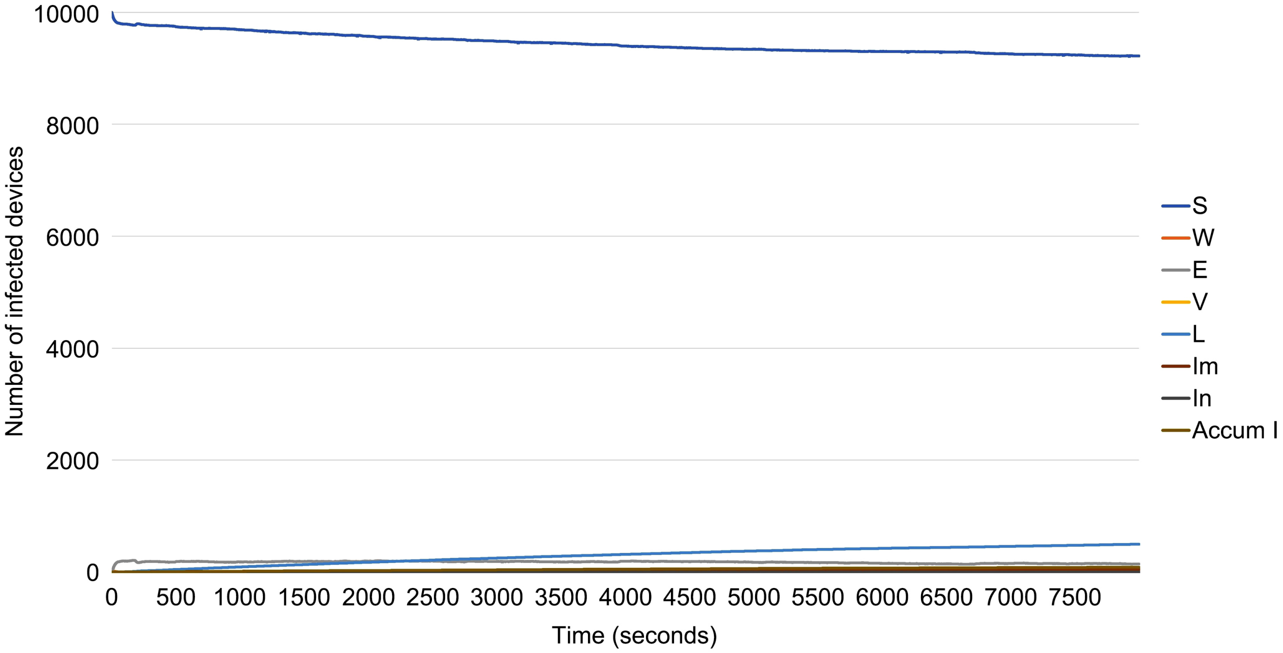

Figure 13. Curves of each compartment with iOS OS.

infected Android devices is close to 8000 devices, while the number of latent devices remains below 2000. In **Figure 12**, it can be seen that the number of infected iOS devices is minimal, since most of the infected devices are of another type of OS, in addition to the fact that the number of latent devices is higher than the number of infected ones, so it is significantly reduced that users with this OS can infect others of the same type.

6. Conclusions and Future Work

In this work, a new discrete and probabilistic model, based on Network Automata, is presented. The proposed model is based on the concepts of compartmental epidemiological models and scale-free networks to define the propagation dynamics between smartphones. The model takes into account characteristics related to user interaction and behavior, as well as the behavior of the worm, which are crucial to adequately reproduce the propagation dynamics.

An analysis of the behavior of the proposed model was carried out by varying the values of the different parameters and under specific scenarios through computational simulation. The results indicate that the proposed model is able to adequately replicate the spreading behavior of SMS-based worm-like malware.

Although there are physical or external factors, such as failure in the telephone network (delay or failure in message delivery) or incompatibility with the OS, it is important to note that the determining factors in the worm's spread dynamics were human factors, particularly: the degree of trust, the reading time of a message and risk awareness.

The results obtained and the analysis carried out confirm that the best way to protect against malware attacks, leaving aside the vulnerabilities of the devices, is to make users aware of the risks that exist on the Internet, the existence of malware and the damage they cause, as well as the proper use of the SMS service.

In future work, we will work with a computer with greater characteristics, which allows integrating a network with a greater number of mobile devices, as it really is, parallelizing the proposed model. It is intended to transpose this model to one that involves fuzzy logic such as fuzzy cognitive maps [42]; for which the different, behaviors will be simulated independently.

On the other hand, we will try to obtain real data or data obtained from at least one simulator of specific scenarios in order to better parameterize the model and better evaluate its performance. Finally, we will work on a model that considers SMS in conjunction with Bluetooth and/or MMS, as well as defining users with different behaviors.

Acknowledgements

This work is the product of the research carried out by Erick Iván Medina-Salas, (CVU: 926556) to obtain the degree of Master in Computer Science and Engineering at Posgrado en Ciencia e Ingeniería de la Computación-Universidad Nacional Autónoma de México and supported by the Conahcyt-Mexico. It is also part of the divisional project *Design of intelligent interfaces for simulating the behavior of living or animate organisms*, from the Universidad Autónoma Metropolitana-Azcapotzalco.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Statista (2022) Número de usuarios de smartphones en México 2021. https://es.statista.com/estadisticas/1077622/usuarios-de-smartphone-en-mexico/

[2] Medina-Salas, E.I. (2021) Modelación de la dinámica de propagación de malware en smartphones por SMS basado en autómatas celulares en redes. Master's Thesis, National Autonomous University of Mexico, Mexico City. http://132.248.9.195/ptd2021/enero/0806114/Index.html

[3] Santillán, M.L. (2022) Epidemiologí a, útil para describir e investigar la saluddelapoblació n. http://ciencia.unam.mx/leer/887/epidemiologia-util-para-describir-la-salud-de-la-poblacion

[4] McKendrick, A.G. (1925) Applications of Mathematics to Medical Problems. *Proceedings of the Edinburgh Mathematical Society*, **44**, 98-130. https://doi.org/10.1017/S0013091500034428

[5] Kermack, W.O. and McKendrick, A.G. (1927) A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **115**, 700-721. https://doi.org/10.1098/rspa.1927.0118

[6] Kermack, W.O. and McKendrick, A.G. (1932) Contributions to the Mathematical

Theory of Epidemics. II.—The Problem of Endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **138**, 55-83. https://doi.org/10.1098/rspa.1932.0171

[7] Kermack, W.O. and McKendrick, A.G. (1933) Contributions to the Mathematical Theory of Epidemics. III.—Further Studies of the Problem of Endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **141**, 94-122. https://doi.org/10.1098/rspa.1933.0106

[8] del Rey, A.M. (2015) Mathematical Modeling of the Propagation of Malware: A Review. *Security and Communication Networks*, **8**, 2561-2579. https://doi.org/10.1002/sec.1186

[9] Peng, S., Yu, S. and Yang, A. (2013) Smartphone Malware and Its Propagation Modeling: A Survey. *IEEE Communications Surveys & Tutorials*, **16**, 925-941. https://doi.org/10.1109/SURV.2013.070813.00214

[10] Signes-Pont, M.T., Cortés-Castillo, A., Mora-Mora, H. and Szymanski, J. (2018) Modelling the Malware Propagation in Mobile Computer Devices. *Computers & Security*, **79**, 80-93. https://doi.org/10.1016/j.cose.2018.08.004

[11] Xiao, X., Fu, P., Li, Q., Hu, G. and Jiang, Y. (2017) Modeling and Validation of SMS Worm Propagation over Social Networks. *Journal of Computational Science*, **21**, 132-139. https://doi.org/10.1016/j.jocs.2017.05.011

[12] Xiao, X., Fu, P., Hu, G., Sangaiah, A.K., Zheng, H. and Jiang, Y. (2017) SAIDR: A New Dynamic Model for SMS-Based Worm Propagation in Mobile Networks. *IEEE Access*, **5**, 9935-9943. https://doi.org/10.1109/ACCESS.2017.2700011

[13] Xiao, X., Fu, P., Dou, C., Li, Q., Hu, G. and Xia, S. (2017) Design and Analysis of SEIQR Worm Propagation Model in Mobile Internet. *Communications in Nonlinear Science and Numerical Simulation*, **43**, 341-350. https://doi.org/10.1016/j.cnsns.2016.07.012

[14] Jia, P., Liu, J., Fang, Y., Liu, L. and Liu, L. (2018) Modeling and Analyzing Malware Propagation in Social Networks with Heterogeneous Infection Rates. *Physica A: Statistical Mechanics and Its Applications*, **507**, 240-254. https://doi.org/10.1016/j.physa.2018.05.047

[15] Zhang, C. and Huang, H. (2016) Optimal Control Strategy for a Novel Computer Virus Propagation Model on Scale-Free Networks. *Physica A: Statistical Mechanics and Its Applications*, **451**, 251-265. https://doi.org/10.1016/j.physa.2016.01.028

[16] Liu, W., Liu, C., Yang, Z., Liu, X., Zhang, Y. and Wei, Z. (2016) Modeling the Propagation of Mobile Malware on Complex Networks. *Communications in Nonlinear Science and Numerical Simulation*, **37**, 249-264. https://doi.org/10.1016/j.cnsns.2016.01.019

[17] Hosseini, S. and Azgomi, M.A. (2016) A Model for Malware Propagation in Scale-Free Networks Based on Rumor Spreading Process. *Computer Networks*, **108**, 97-107. https://doi.org/10.1016/j.comnet.2016.08.010

[18] Hosseini, S. and Azgomi, M.A. (2018) The Dynamics of a SEIRS-QV Malware Propagation Model in Heterogeneous Networks. *Physica A: Statistical Mechanics and Its Applications*, **512**, 803-817. https://doi.org/10.1016/j.physa.2018.08.081

[19] Hosseini, S., Abdollahi Azgomi, M. and Rahmani Torkaman, A. (2016) Agent-Based Simulation of the Dynamics of Malware Propagation in Scale-Free Networks. *Simulation*, **92**, 709-722. https://doi.org/10.1177/0037549716656060

[20] Gan, C., Yang, M., Zhang, Z. and Liu, W. (2017) Global Dynamics and Optimal Control of a Viral Infection Model with Generic Nonlinear Infection Rate. *Discrete Dynamics in Nature and Society*, **2017**, Article ID: 7571017.

https://doi.org/10.1155/2017/7571017

[21] Huang, S. (2018) Global Dynamics of a Network-Based WSIS Model for Mobile Malware Propagation over Complex Networks. *Physica A*: *Statistical Mechanics and Its Applications*, **503**, 293-303. https://doi.org/10.1016/j.physa.2018.02.117

[22] Liu, W. and Zhong, S. (2018) A Novel Dynamic Model for Web Malware Spreading over Scale-Free Networks. *Physica A*: *Statistical Mechanics and Its Applications*, **505**, 848-863. https://doi.org/10.1016/j.physa.2018.04.015

[23] Liu, X., Li, T., Xu, H. and Liu, W. (2019) Spreading Dynamics of an Online Social Information Model on Scale-Free Networks. *Physica A*: *Statistical Mechanics and Its Applications*, **514**, 497-510. https://doi.org/10.1016/j.physa.2018.09.085

[24] Selvam, A.G.M., Janagaraj, R., Jones, G.M., Thandalam, C. and India, S. (2018) Dynamics in a Fractional Order SIQR Model of Worm Propagation. *International Journal of Pure and Applied Mathematics*, **119**, 549-558.

[25] Yun, X., Li, S. and Zhang, Y. (2015) SMS Worm Propagation over Contact Social Networks: Modeling and Validation. *IEEE Transactions on Information Forensics and Security*, **10**, 2365-2380. https://doi.org/10.1109/TIFS.2015.2455413

[26] Peng, S., Wang, G. and Yu, S. (2013) Modeling Malware Propagation in Smartphone Social Networks. 2013 12*th IEEE International Conference on Trust*, *Security and Privacy in Computing and Communications*, Melbourne, 16-18 July 2013, 196-201. https://doi.org/10.1109/TrustCom.2013.28

[27] Peng, S., Wu, M., Wang, G. and Yu, S. (2014) Propagation Model of Smartphone Worms Based on Semi-Markov Process and Social Relationship Graph. *Computers & Security*, **44**, 92-103. https://doi.org/10.1016/j.cose.2014.04.006

[28] Kroc, J., Sloot, P. and Hoekstra, A. (2010) Simulating Complex Systems by Cellular Automata. Springer, Berlin. https://doi.org/10.1007/978-3-642-12203-3

[29] Tissera, P.C., Printista, A.M. and Errecalde, M.L. (2007) Evacuation Simulations Using Cellular Automata. *Journal of Computer Science and Technology*, **7**, 14-20.

[30] Barabási, A.L. and Bonabeau, E. (2003) Scale-Free Networks. *Scientific American*, **288**, 60-69. https://doi.org/10.1038/scientificamerican0503-60

[31] Held, P., Dockhorn, A. and Kruse, R. (2014) On Merging and Dividing of Barabási-Albert Graphs. 2014 *IEEE Symposium on Evolving and Autonomous Learning Systems* (*EALS*), Orlando, 9-12 December 2014, 17-24. https://doi.org/10.1109/EALS.2014.7009499

[32] Barabási, A.L., Albert, R. and Jeong, H. (2000) Scale-Free Characteristics of Random Networks: The Topology of the World-Wide Web. *Physica A*: *Statistical Mechanics and Its Applications*, **281**, 69-77. https://doi.org/10.1016/S0378-4371(00)00018-2

[33] Watts, D. and Strogatz, S. (1998) Collective Dynamics of 'Small-World' Networks. *Nature*, **393**, 440-442. https://doi.org/10.1038/30918

[34] Boccara, N. and Cheong, K. (1992) Automata Network SIR Models for the Spread of Infectious Diseases in Populations of Moving Individuals. *Journal of Physics A*: *Mathematical and General*, **25**, 2447-2461. https://doi.org/10.1088/0305-4470/25/9/018

[35] NetworkX (2020) NetworkX Documentation. https://networkx.github.io

[36] Miklas, A.G., Gollu, K.K., Chan, K.K., Saroiu, S., Gummadi, K.P. and De Lara, E. (2007) Exploiting Social Interactions in Mobile Systems. In: Krumm, J., Abowd, G.D., Seneviratne, A. and Strang, T., Eds., *UbiComp* 2007: *Ubiquitous Computing*, Springer, Berlin, 409-428. https://doi.org/10.1007/978-3-540-74853-3_24

[37] Shahyad, S., Pakdaman, S., Hiedary, M., Miri, M., Asadi, M., Nasri, A. and Alipour,

A.S. (2011) A Comparison of Motivation, Frequency and Content of S.M.S. Messages Sent in Boys and Girls High School Student. *Procedia—Social and Behavioral Sciences*, **15**, 895-898.
http://www.sciencedirect.com/science/article/pii/S1877042811003867
https://doi.org/10.1016/j.sbspro.2011.03.207

[38] Meng, X., Zerfos, P., Samanta, V., Wong, S. and Lu, S. (2007) Analysis of the Reliability of a Nationwide Short Message Service. *IEEE INFOCOM* 2007—26*th IEEE International Conference on Computer Communications*, Anchorage, 6-12 May 2007, 1811-1819. https://doi.org/10.1109/INFCOM.2007.211

[39] Statista Research Department (2016) Awareness of Internet Security Risks according to Internet Users Worldwide as of August 2016.
https://www.statista.com/statistics/463767/awareness-of-online-security-risks/

[40] Chau, M. and Reith, R. (2023) Smartphone Market Share.
https://www.idc.com/promo/smartphone-market-share/os

[41] StatCounter (2022) Mobile Operating System Market Share Worldwide.
https://gs.statcounter.com/os-market-share/mobile

[42] Mora-Torres, M., Laureano-Cruces, A.L., Ramírez-Rodríguez, J. and Espinosa-Paredes, G. (2009) Analysis and Design of the Knowledge Representation for the Implementation of a Distributed Reasoning. *Revista de Matemática*: *Teoría y Aplicaciones*, **16**, 267-281. https://www.redalyc.org/articulo.oa?id=45326951007
https://doi.org/10.15517/rmta.v16i2.306