UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

Modelación de la dinámica de propagación de malware en smartphones por SMS basado en autómatas celulares en redes

T E S I S

QUE PARA OPTAR POR EL GRADO DE
MAESTRO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

PRESENTA:
ERICK IVÁN MEDINA SALAS

Asesor:
Dra. María Elena Lárraga Ramírez
Instituto de Ingeniería

Ciudad Universitaria, Cd. Mx. Enero 2021

Agradecimientos

A mis padres, Ana y Javier, que con mucho amor y apoyo me han acompañado durante todos mis estudios y sobre todo me han enseñado que con mucho esfuerzo todo se puede lograr. Gracias por confiar en mí.

A mi asesora, la Dra. María Elena Lárraga Ramírez por su dedicación, tiempo y esfuerzo puestos durante la realización de este trabajo.

A la Dra. Ana Lilia C. Laureano-Cruces por su interés, apoyo y sabios consejos que me brindó desde que inicie hasta que terminé mis estudios de maestría.

A mis sinodales, los doctores Héctor Benítez Pérez, Javier Gómez Castellanos, Luis Francisco García Jiménez y Luis Agustín Álvarez-Icaza Longoria por tomarse el tiempo de leer este trabajo y a sus conocimientos, los cuales enriquecieron el mismo.

Al Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, de la Universidad Nacional Autónoma de México, por la oportunidad que me brindó de estudiar un posgrado de alta excelencia académica.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca que se me otorgó durante la realización de mis estudios de maestría.

Y finalmente, al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IN112619. Agradezco a la DGAPA-UNAM la beca recibida para concluir este trabajo.

Resumen

En 2020, el número de usuarios de smartphones en México se estimó en aproximadamente 80.9 millones. Se prevé que el número de usuarios de estos dispositivos móviles crezca de manera continua. En particular, los ataques mediante SMS (Short Message Service) plantean una amenaza significativa para todos los usuarios móviles. El malware de tipo gusano aprovecha este medio de comunicación para poder propagarse. Así, la vulnerabilidad de estos dispositivos a diversas amenazas plantea un grave riesgo para los usuarios. Por lo que estudiar la dinámica de propagación del malware puede ayudar a entender y prevenir un contagio masivo entre dispositivos móviles. En el presente trabajo, se desarrolla un modelo basado en Autómatas Celulares (AC) en red y modelos epidemiológicos compartimentales, con el objetivo de simular, analizar y estudiar la propagación de malware tipo gusano a través del envío de mensajes SMS en smartphones. El modelo contempla factores reales del comportamiento del usuario y características del malware que influyen en la dinámica de propagación del gusano: el grado de relación entre personas, el grado de conciencia de riesgo, y la frecuencia de revisión de mensajes de los usuarios. Para lograr que el modelo reproduzca de forma fiel y real el sistema. El modelo preserva la simplicidad que caracteriza a los modelos basados en AC, mejorando a su vez los modelos existentes. Al ser un modelo basado en AC, cubre las desventajas que modelos basados en ecuaciones diferenciales o en procesos estocásticos tienen para describir comportamientos individuales.

Resultados obtenidos de diferentes simulaciones de computadora muestran que el modelo reproduce de manera fiel la dinámica de propagación del gusano mediante mensajes SMS, por lo que es adecuado para analizar y evaluar estrategias contra ataques mediante este tipo de malware. Al utilizar AC, el trabajo desarrollado puede ser implementado para simulaciones utilizando paralelización, lo que aumenta la eficiencia en cuanto a tiempo de ejecución y al manejo de una gran cantidad de datos (población grande de dispositivos).

Abstract

Nowadays, smartphones play an important role in daily life. In 2020, the number of users has already reached 80.9 million in Mexico. This trend will continue in the upcoming years. However, this significant usage of smartphones has also attracted worm writers to commit their malicious goals by exploiting their vulnerabilities. Simple messaging service (SMS) has become one of the most attractive means to spread mobile malware. Therefore, to characterize the propagation dynamics of SMS/MMS-worms has become an important topic to understand and prevent a massive infection between mobile devices. In this work, a new Cellular Automata (CA) network-based model is proposed, whose dynamics is developed by the compartmental epidemiological models and scale-free networks. The aim is to simulate, analyze, and study the spread of malware of the type worm on smartphones when different user interactions and malware characteristics. In such a way that the proposed model will be able to reproduce the realistic behavior of the worm propagation in smartphone-based on SMS, but it also preserves the simplicity that characterizes the AC-based model. Simulations results from the proposed model under different scenarios indicate that this model reproduces the propagation dynamics of the malware type worm by SMS, making it suitable to analyze, evaluate, and prevent its spreading among smartphones.

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

Desde que los seres humanos comprendieron que necesitaban más conocimiento, más acceso a la información y procesar de forma rápida todo en un equipo que sea pequeño, accesible y fácil de trasladar, la expansión de los dispositivos móviles se ha incrementado de manera continua. Como consecuencia, el desarrollo de dispositivos móviles desde laptops, tablets, smartphones (que actualmente han llegado a ser oficinas portátiles y escuelas en línea) se ha vuelto fundamental, ya que todos estos equipos no sólo facilitan la vida de los usuarios sino que además conectan a cada persona en el mundo.

Particularmente, en lo que a smartphones se refiere, a nivel mundial hay 5190 millones de usuarios de teléfonos celulares [4], y el 73 % de las personas están conectadas y comparten su tiempo desde un teléfono inteligente o smartphone. En México, de acuerdo con el INEGI [5], el número total de usuarios que disponen de un teléfono inteligente creció de 64.7 millones de personas en 2017 a 80.9 millones en 2020. Y además, el 71 % de los smartphones que poseen los mexicanos fueron adquiridos en los últimos 18 meses.

Un servicio disponible en los teléfonos móviles es el servicio de mensajes cortos o servicio de mensajes simples, más conocido como SMS (por las siglas del inglés Short Message Service), que permite el envío de mensajes cortos con un número limitado de caracteres, entre teléfonos móviles. Actualmente, los mensajes SMS son un medio de comunicación importante entre personas en todo el mundo, se ocupan cotidianamente para consultar el saldo telefónico, enviar alertas bancarias o publicidad, hacer compras en línea, recordatorios de pagos, entre otros usos. Por lo que miles de millones de mensajes SMS se envían todos los días, convirtiéndose en un generador de ingresos para operadores inalámbricos [6]. SMS ofrece grandes ventajas, como leer o enviar en cualquier momento un mensaje. Debido a ello, una gran cantidad de aplicaciones se construyen ahora sobre esta tecnología y cada vez se desarrollan más [7].

Por otro lado, conectar cada uno de nuestros dispositivos a Internet es cada vez más fácil, las compañías telefónicas y de TV privada ofrecen la conexión para este servicio. Los smartphones no sólo se conectan a Internet, sino que permanecen conectados a las redes telefónicas móviles mediante un operador móvil; de esta manera los dispositivos permiten la comunicación aunque no se tenga acceso a la gran red, que es Internet.

1.1. Justificación

A pesar de las ventajas que da el servicio SMS, su amplio uso ha generado que sea un vector de ataque importante de los escritores de malware o software malicioso, quienes engañan a las personas aprovechándose de distintas vulnerabilidades que existe en la red y en los dispositivos mismos. Las consecuencias pueden ser tan graves, como se lo proponga el desarrollador del malware; desde una simple pérdida de información hasta el robo de la misma; siendo perjudicial para la víctima. Instalar aplicaciones para proteger los dispositivos, como antivirus, no garantiza que estarán seguros todo el tiempo.

Existen distintos tipos de malware como virus, gusanos, troyanos, entre otros. Particularmente, el malware tipo gusano es potencialmente peligroso para los dispositivos móviles, pues explotan la topología de la red para poder propagarse hacia los demás dispositivos, especialmente utilizando mensajes SMS. Por lo que en los últimos años, los investigadores se han dado a la tarea de estudiar, entender ya analizar el comportamiento del malware en dispositivos móviles para poder predecir a futuro, como se propagaría el nuevo malware.

Así, en los últimos años, se han desarrollado diversos modelos relacionados al estudio de la propagación de gusanos en dispositivos móviles através de SMS, como los modelos basados en ecuaciones diferenciales y modelos basados en procesos estocásticos. La mayoría de estos modelos utilizan la teoría de epidemiología compartimental, donde se agrupa la población en compartimentos dependiendo de las características que se generen; por ejemplo, cuando se hablan de infecciones de enfermedades se tiene un compartimiento de susceptibles, uno de infectados y otro de recuperados (Modelo SIR). Sin embargo, la mayoría de los modelos presentados no toman en cuenta algunas características importantes que determinan la dinámica de propagación del gusano, como las interacciones locales que surgen entre los dispositivos móviles. Ello ocurre pues se trata de un análisis de forma macroscópica, por lo que, no pueden simular la dinámica individual de cada dispositivo; además, consideran que todos los dispositivos pueden infectarse homogéneamente, evento que no es muy apegado a la realidad.

Recientemente los modelos basados en Autómatas Celulares (referido como AC) para la propagación de malware han tomado mucho auge. Los modelos basados en autómatas celulares pueden solucionar las deficiencias que los demás modelos tienen, dando características propias a cada individuo, en el caso de la propagación de malware tipo gusano, a cada dispositivo móvil [8]. Sin embargo, aún los modelos desarrollados por esta nueva teoría, la mayoría no contemplan características reales que influyen directamente en la propagación de gusanos, mediante mensajes SMS, como lo son las siguientes:

- La relación entre dispositivos, es decir, que tanto confía un dispositivo de otro;
- La topología de la red, pues los modelos existentes utilizan topologías aleatorias para representar la red formada por la lista de contactos de cada dispositivo y no representa fielmente la realidad;
- La actividad de revisar los dispositivos de los usuarios, o;

- Fallas en la red telefónica, como lo son la pérdida o retraso en la entrega de mensajes.

1.2. Objetivo

El objetivo de este trabajo de tesis se enfoca en el desarrollo de un modelo matemático y computacional nuevo y basado en individuos para analizar, entender, evaluar y prevenir ataques a través de malware tipo gusano entre smartphones, con base en SMS.

Para este propósito, el nuevo modelo es discreto y probabilista, y su definición se basa en el uso de Autómatas Celulares en Red, redes libres de escala y modelos epidemiológicos compartimentales para definir su dinámica de propagación del malware. De tal manera que el modelo propuesto toma en cuenta tanto características referentes a la interacción del usuario y su comportamiento, y el comportamiento del gusano; que son determinantes para reproducir de mejor manera el comportamiento de la propagación del malware.

1.3. Metas

Las metas que se persiguen con este trabajo son:

- Desarrollar un modelo basado en autómatas celulares en red para simular la propagación de malware tipo gusano en dispositivos móviles a través del envío de mensajes SMS, considerando características de la interacción del usuario, de la red y del comportamiento del malware;
- Utilizar una red libre de escala para representar de mejor forma la relación entre personas;
- Hacer la implementación del modelo propuesto en el lenguaje de programación Python, para realizar una simulación bajo escenarios específicos;
- Realizar un análisis de los resultados obtenidos de diferentes simulaciones, debido a la variación de distintos parámetros de entrada, y;
- Evaluar el desempeño del modelo propuesto.

1.4. Hipótesis

La hipótesis que orienta este trabajo es que, los autómatas celulares en red conjugados con redes libre de escala son adecuados para modelar la propagación de malware a través de SMS. Debido a que permiten tomar en cuenta para la definición de la reglas de evolución del sistema, características del comportamiento; tanto de los usuarios como del gusano, que son determinantes para la representación de la dinámica de la propagación de malware tipo gusano de manera simple.

1.5. Contribución

La contribución de este trabajo es el desarrollo de un modelo nuevo para simular la dinámica de propagación de malware a través de SMS. Que aporte a la literatura de estudio, pero que la vez sirva como herramienta para el análisis, estudio y evaluación de la propagación del malware ante escenarios específicos. Que sirva como precedente para el estudio del comportamiento de otros vectores de ataque.

1.6. Organización del trabajo

El resto del trabajo de tesis se organiza de la siguiente forma. En el capítulo 2, se muestra un breve marco teórico relacionado con el servicio de mensajes cortos (SMS), así como malware, para facilitar la lectura del trabajo presentado. En el capítulo 3, se presenta una breve revisión del estado del arte acerca de modelos de propagación de malware, asi como teoría utilizada para la realización de cada modelo presentado. En el capítulo 4, se presenta un modelo nuevo basado en el paradigma de Autómata Celular en red para la propagación de gusano mediante el envío de mensajes SMS. En el capítulo 5, se presentan la validación y verificación del modelo, a través de los resultados obtenidos de simulaciones obtenidas de diferentes casos de estudio. Por último, en el capítulo 6, se presentan las conclusiones de este trabajo de tesis y algunas propuestas para trabajo futuro.

Capítulo 2

SMS y Malware

En este capítulo se introducen los conceptos necesarios relacionados con el SMS, así como de el malware y los diferentes tipos de malware que existen, en especial los del tipo gusano teniendo como medio de propagación en los mensajes SMS.

Hoy en día, la mayoría de las personas posee al menos un teléfono inteligente, ya que son útiles en distintas tareas, principalmente en la comunicación, así como en el almacenamiento de información. Además, los teléfonos inteligentes evolucionan rápidamente integrando mayores tecnologías que son atractivas para los usuarios por lo que el uso de estos dispositivos cada vez se vuelve mayor. En México, el número total de usuarios que disponen de un celular inteligente creció de 64.7 millones de personas en 2017 a 80.9 millones en 2020.

Los smartphones, están mayormente conectados a las redes móviles que son ofrecidas por alguna compañía telefónica. De esta forma, es posible acceder a Internet, gracias a los servicios que ofrecen las mismas compañías o a través de una conexión entre el dispositivo y algún punto de acceso a esta. Sin embargo, actualmente se vuelve un riesgo potencial debido a las diferentes amenazas, así como vulnerabilidades que hay en la red; como lo es el software malicioso o malware.

Aunque no siempre se está conectado a Internet, los smartphones estarán enlazados a las redes móviles mientras alguna antena del operador telefónico le brinde servicio, es decir, el teléfono recibe señal de la antena. Realizar una llamada o enviar un mensaje de texto SMS (Short Message Service), son unas de las principales funciones que pueden hacer los teléfonos inteligentes sin necesidad de utilizar Internet, teniendo solamente como referencia un número celular.

Debido a lo anterior, los estafadores aprovechan la vulnerabilidad de los dispositivos al permanecer conectados a una red móvil, pues con un simple mensaje de texto es posible enviar un enlace de un sitio web donde se almacene algún malware, que al ser accedido por el usuario se abre un mensaje malicioso que infecta a su dispositivo. Las consecuencias pueden ser muy molestas y terribles para los usuarios afectados, desde drenar la batería del dispositivo hasta el robo de información. Una vez que un dispositivo es infectado por el malware, este se puede propagar entre dispositivos utilizando la red móvil y gracias a que el uso de dispositivos móviles se ha incrementado considerablemente, es aquí donde se vuelve importante poder predecir así como prevenir un ataque, analizando su propagación en estos dispositivos.

2.1. Servicio de Mensajes Cortos (SMS)

Actualmente, los mensajes SMS siguen siendo un medio de comunicación importante entre personas en todo el mundo, se ocupan cotidianamente para consultar el saldo telefónico, enviar alertas bancarias o publicidad, hacer compras en línea, recordatorios de pagos, entre otros usos. Miles de millones de mensajes SMS se envían todos los días, convirtiéndose en un generador de ingresos para operadores de red inalámbricos. El servicio SMS, es una tecnología inalámbrica que permite el envío y recibo de mensajes entre dispositivos móviles; pertenece al estándar de telefonía celular GSM (Global System for Mobile communication). En diciembre de 1992, el primer mensaje fue enviado de una Computadora Portátil (PC) a una Estación Móvil, en la red GSM de Vodafone, en Reino Unido [9].

2.1.1. Sistema Global para comunicación Móvil (GSM)

GSM pertenece a la Segunda Generación de telefonía digital (2G). En un principio operó en la banda de los 900 MHz [10], divida en dos de 25 MHz: 890-915 MHz para el enlace de bajada y 935-960 MHz para el enlace de subida; cada banda es dividida en canales de 200 KHz [11], como se muestra en la Figura 2.1, teniendo un total de 125 canales para usar, donde uno pertenece al de control. Posteriormente GSM se adaptó, en Estados Unidos, en la banda de los 1800 MHz [12], divida en dos bandas de 75 MHz: 1710-1785 MHz para recibir y 1805-1880 MHz para enviar; cada banda es dividida igualmente en canales de 200 MHz teniendo un total de 375 canales.

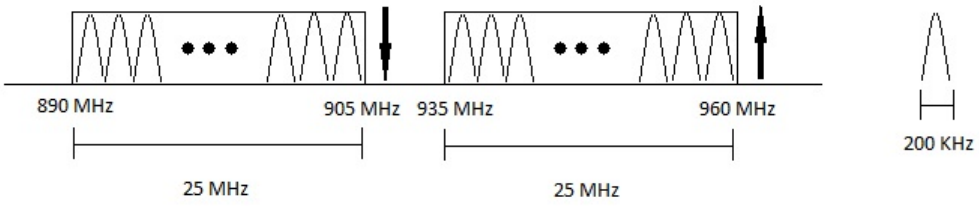


Figura 2.1: Banda de los 900 MHz en GSM.

A diferencia de la Primera Generación de telefonía celular (1G) que utilizó un Acceso Múltiple por División de Frecuencia (FDMA), GSM utiliza un Acceso Múltiple por División de Tiempo (TDMA) teniendo mayores ventajas que la 1G, entre estas: incrementa la capacidad de atención a usuarios, incrementa la calidad del audio en una llamada e incremento en la velocidad de transmisión. En la Figura 2.2 se muestra la principal diferencia entre estas técnicas de modulación, donde se representa a varios usuarios atendidos en un llamada, siendo TDMA la técnica que atiende al mayor número de usuarios. En GSM, cualquier dispositivo celular, como un teléfono, es llamado Estación Móvil (EM). Para que una EM opere es necesario que tenga instalado un Módulo Identificador de Subscriptor (SIM), más conocido como tarjeta SIM. La tarjeta

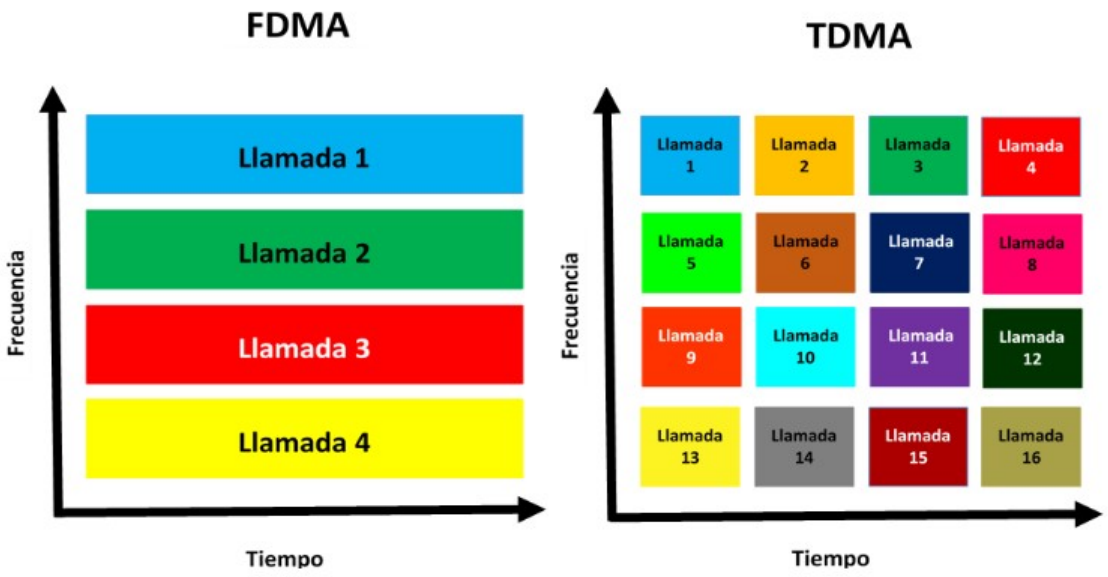


Figura 2.2: Diferencia entre TDMA y FDMA.

SIM contiene información importante que autentica y valida el dispositivo móvil que se intenta conectar a la red de la Estación Base (BS). Ademas, la tarjeta SIM permite almacenar información del usuario como de sus aplicaciones o de sus contactos telefónicos [13]. La BS se encarga de proporcionar los servicios de telefonía, gestionando de la mejor manera posible los recursos de la red a través del canal de control. El canal de control a la vez se divide en diferentes subcanales, cada uno encargado de alguna tarea en específica.

GSM y SMS fueron estándares desarrollados originalmente por European Telecommunications Standars Institute (ETSI) [14, 15].

2.1.2. Funcionamiento del envío de un mensaje

Como se muestra en la Figura 2.3, en GSM existen diferentes canales, de los cuales dos son importantes para la transmisión de información de control [16]:

- El Canal de Control Dedicado Independiente (SDCCH).
- El Canal de Control Asociado Lento (SACCH).

El SDCCH transporta información de señalización después del establecimiento de la conexión móvil a la red. El SACCH siempre está asociado con un canal de tráfico o un SDCCH y se asigna al mismo canal físico. El SACCH transporta información general del móvil a la red, como detalles de la intensidad de la señal [17]. Los mensajes SMS, en GSM, se pueden transmitir de dos formas, a través del SDCCH o del SACCH, dependiendo del tráfico:

- Cuando no hay tráfico en el canal, el mensaje se enviará usando el SDCCH.

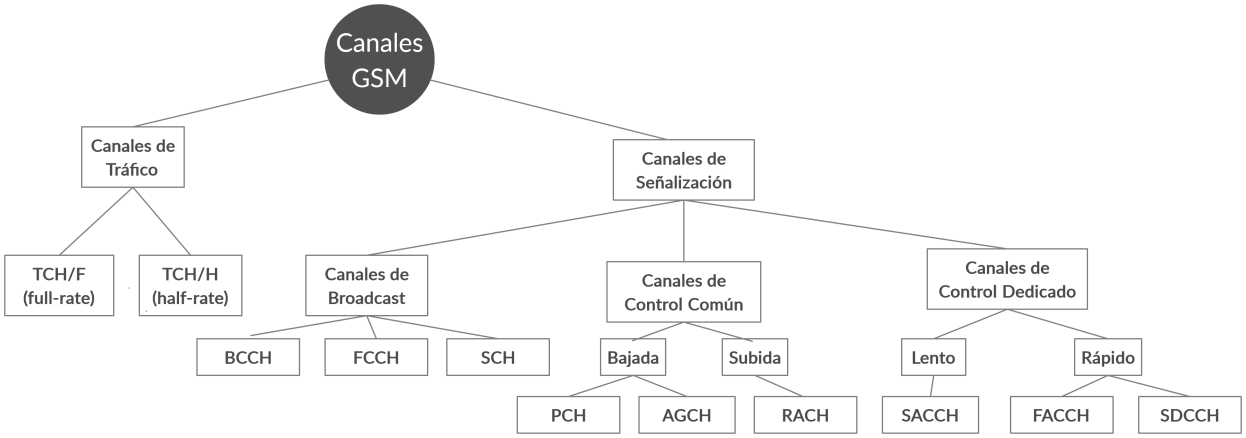


Figura 2.3: Canales de GSM.

- Si hay tráfico en el canal durante el envío del mensaje, se detiene el envío y se continuará en el SACCH.
- Si desde el inicio hay tráfico se envía en el SACCCH.
- Cuando el tráfico termina, se puede decidir enviar en el SACCH o en el SDCCH.

Una vez que es enviado un mensaje SMS desde una terminal móvil, este es recibido por el Centro SMS (SM-SC) [18]. El SM-SC es responsable del manejo de operación SMS en una red. Una vez recibido el mensaje SMS, este lo reenviará al destino. El principal objetivo de un SM-SC es rutear los mensajes SMS y regular su proceso, por ejemplo, si el dispositivo móvil destino no está disponible, el SM-SC lo guardará y enviará una vez que este disponible. El cliente SMS debe conocer el número del SM-SC (operador de red), el cual está guardado en la tarjeta SIM. Todo mensaje SMS, al ser enviado, tiene asociado un periodo de validez. Un mensaje SMS es almacenado en un SM-SC si el dispositivo no está disponible. El periodo en que tarda en ser borrado del SM-SC se llama periodo de validez, puede ser especificado por el usuario aunque está predefinido por el operador telefónico generalmente.

Por lo anterior, el envío de SMS consiste básicamente en dos servicios:

- Transmisión de una Terminal Móvil al SM-SC.
- Transmisión del SM-SC a una Terminal Móvil.

En la Figura 2.4 se muestra el envío de un mensaje entre dos terminales móviles cuando pertenecen al mismo operador, es decir, solo involucra un SM-SC. Durante el envío de un mensaje, se añade además información adicional que indica el estado de cada mensaje [18], esta información es llamada reporte. Los reportes asociados a un mensaje SMS son:

- Reporte de estado del mensaje: informa a la terminal móvil origen si el mensaje SMS ya se entregó a la terminal móvil destino.

- Reporte de envío de mensaje: informa a la terminal móvil origen si el mensaje fue recibido correctamente por el SM-CS.
- Reporte de entrega de mensaje: informa al SM-SC si el mensaje fue recibido correctamente por el móvil destino.

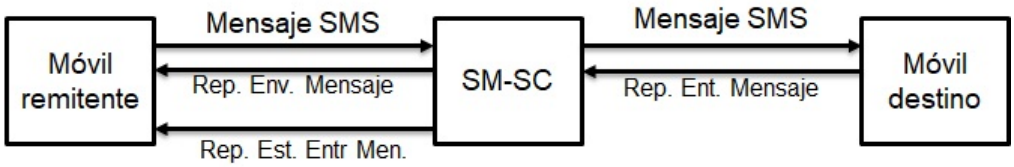


Figura 2.4: Envío de un mensaje entre un mismo operador.

En la Figura 2.5 se muestra el envío de un mensaje cuando involucra más de dos operadores, es decir, más de dos SM-SC. Por lo que, la comunicación entre los dos SM-SC dependerá de la características de las redes de ambos operadores, teniendo los siguientes casos [19]:

- Cuando las redes son similares, sucede el mismo proceso que cuando es un solo operador, solo que se utilizan interconexiones de señalización.
- Cuando las redes son diferentes, los centros SMS son conectados a través de un Gateway SMS o con un protocolo de comunicación que soporten ambos centros.

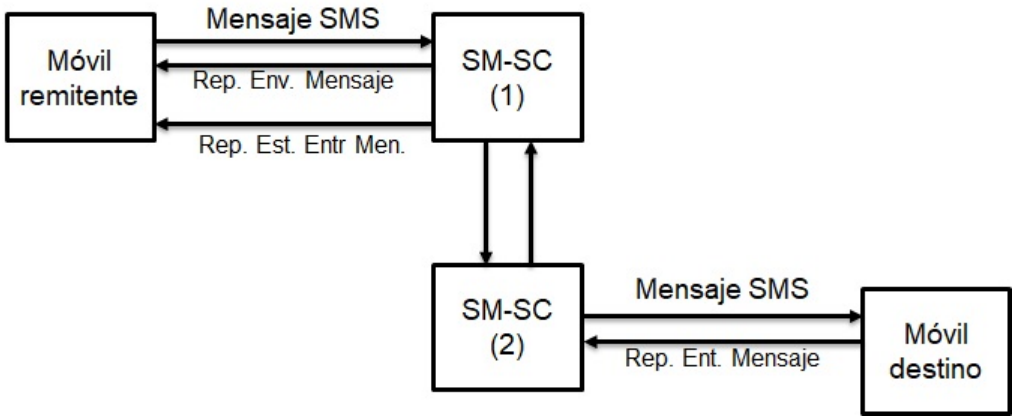


Figura 2.5: Envío de un mensaje entre dos operadores.

El Gateway SMS [19] se encarga de comunicar los dos SM-SC, como se muestra en la Figura 2.6, traduce el protocolo de un SM-SC a otro. De esta manera se permite el intercambio de mensajes SMS entre estos. Existen diferentes protocolos, algunos de ellos son:

- Short Message Peer-to-Peer (SMPP): es un protocolo abierto, creado en la industria especialmente para el intercambio de mensajes SMS.

- HyperText Transfer Protocol (HTTP): es un protocolo de comunicación, utilizado mayormente para intercambio de información en la World Wide Web (WWW).
- Universal Computer Protocol (UCP) / External Machine Interface (EMI): EMI es un protocolo extendido de UCP, soporta otro tipo de telefonía digital como: GPRS, CDMA y UMTS.
- Simple Object Access Protocol (SOAP): Es un protocolo web, utilizado para el intercambio de datos entre aplicaciones.

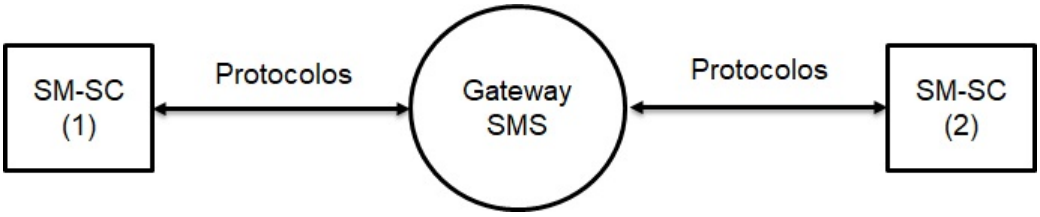
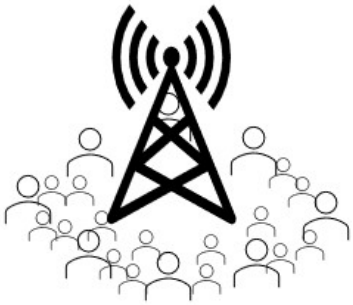


Figura 2.6: Comunicación entre SM-SC a través de un Gateway.

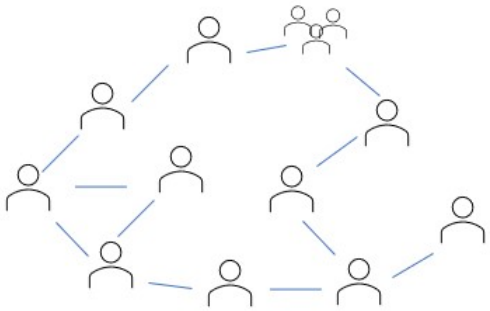
2.1.3. Características de los mensaje SMS

Los mensajes SMS cuentan con las siguientes características [10]:

- Son mensajes de paginación alfanuméricos, es decir, se envían a través de páginas soportando letras y números.
- La longitud máxima de cada mensaje es de 160 caracteres, ocupando 7 bits para la representación de un carácter; o 140 octetos, para una codificación de caracteres latinos como el inglés 70 caracteres, ocupando 16 bits para la representación de un carácter, si se utiliza una codificación Unicode UCS2, como caracteres chinos.
- Soportan cualquier lenguaje, debido a la codificación Unicode.
- Además de texto, también soportan datos binarios. Es posible enviar ringtones, imágenes, animaciones, contactos y hasta configuración de un celular, a través de WBXML (Web Binary Extensible Markup Lenguage).
- Soportan la concatenación de mensajes.
- Existen dos tipos de mensajería, como se muestra en la Figura 2.7 :
  - Servicio de Difusión Celular (Cell Broadcast Service): periódicamente se entregan mensajes a subscritores en el área.
  - Servicio Punto a Punto (PP): se entrega un mensaje a un específico usuario.



(a) Servicio de Difusión Celular



(b) Servicio Punto a punto

Figura 2.7: Tipos de mensajería.

2.1.4. Ventajas de utilizar los mensajes SMS

Los mensajes SMS son usados debido a las grandes ventajas que ofrecen:

- Pueden ser leídos y enviados en cualquier momento.
- Pueden ser enviados a un teléfono móvil no disponible, es decir, que este apagado o no tenga señal
- Permite una comunicación más fácil, por ejemplo, en una biblioteca no se permite hacer ruido o hablar por lo que enviar un mensaje SMS es lo ideal.
- Todos los dispositivos móviles actuales soportan GSM.

Además, muchas empresas de marketing utilizan como herramienta el envío de SMS, ya que otros métodos de comunicación, como el correo electrónico, se necesita ingresar a Internet e iniciar sesión; en cambio el recibir un mensaje SMS no necesita Internet ni registrarse o entrar a alguna cuenta, debido a que se recibe directamente en el teléfono móvil.

Gracias a estas ventajas, una gran cantidad de aplicaciones se construyen ahora sobre esta tecnología y cada vez se están desarrollando más [7]. Sin embargo, algunas de estas ventajas son aprovechados para hacer vulnerables los dispositivos electrónicos, por ejemplo, a ataques de malware.

2.2. Malware

Malware suele referirse a cualquier programa creado para llevar a cabo actividad no autorizada por el usuario, siendo perjudicial para el sistema que lo porta. Al almacenarse en la memoria del sistema y ejecutarse, su actividad consiste desde algo simple como borrar archivos o uso de datos privados, por ejemplo, información de sitios web visitados, lista de contactos, contraseñas, números de cuenta, entre otros; hasta llegar a extorsionar

a los usuarios, pues los escritores de dichos programas pueden llegar a "secuestrar" el dispositivo, es decir, bloquear el equipo y solo lo desbloquearán a cambio de que se ofrezca algo de valor, como dinero.

2.2.1. Breve revisión histórica del Malware

En 1949, John von Neumman escribió un documento llamado: "Theory of self-reproducing automata" [20, 21], siendo la primera persona en referirse a la teoría de virus computacional, aunque este documento fue publicado hasta el año de 1966. Creeper fue considerado como el primer virus computacional, escrito por Bob Thomas [22]. Creeper era capaz de auto replicarse en otras computadoras cada vez que era usado, mostrando el mensaje "Soy Creeper, ¡atrápame si puedes!", por lo que no era dañino para los equipos. Después aparecieron Rabbit y Animal, siendo Animal el primer malware de tipo troyano.

Fue hasta el año de 1983, donde Frederick Cohen quien realizaba una investigación acerca de la protección de computadoras, creo un prototipo de código malicioso demostrando que las computadoras eran vulnerables ante él. Cohen junto con su profesor Leonard Adleman, nombraron al programa: "virus informático" [23].

2.2.2. Vectores de propagación

Un vector de propagación, de ataque o de infección, es la forma en que un escritor de código malicioso puede hacer vulnerable un dispositivo, como una computadora, un sistema o una red. Básicamente, aprovechan las vulnerabilidades de los sistemas [24], es decir, una debilidad o un fallo en un sistema de información que pone en riesgo la seguridad de la información [25]. Debido a lo anterior, la mayoría de los fabricantes de distintos dispositivos trabajan día con día para encontrar dichas vulnerabilidades; una vez encontradas, aplican actualizaciones de seguridad (tanto a software como hardware) para resolver dicha vulnerabilidad, las cuales son importante recibir en nuestros dispositivos [26]. Sin embargo, mientras los fabricantes trabajan para encontrar alguna solución para dichas vulnerabilidades, los atacantes de malware aprovechan ese tiempo para atacar.

La propagación de malware se lleva a cabo por diferentes vectores de ataque, como:

- Dispositivos extraíbles: los dispositivos extraíbles son aquellas unidades de almacenamiento que permiten guardar información de forma externa, de tal manera que se pueda llevar la información a cualquier otro lugar. Esta característica es aprovechada por los escritores de malware infectando archivos o ejecutables que, al ser abiertos o ejecutados en alguna otra computadora o sistema este queda infectado también.

- Envío de archivos o enlaces infectados a través del correo electrónico, Bluetooth o SMS/MMS: los escritores de malware alojan en algún servidor público el archivo malicioso, de tal forma que se envía un enlace donde se puede descargar tal archivo ya sea a través del correo electrónico o SMS/MMS. Las personas que reciban dichos enlaces reciben el enlace mediante un mensaje trampa, es decir, pueden

ser engañados y motivados para que abran dicho enlace. En la subsección 2.3, se detalla de mejor forma la manera en que se propaga un gusano mediante mensajes SMS. Además, es posible enviar algún archivo malicioso a través del correo electrónico o Bluetooth, una vez descargado el archivo puede quedar el sistema involucrado infectado.

- Descarga de archivos infectados por sitios web maliciosos: en este caso, los escritores de malware también alojan un archivo malicioso en un servidor web, cuando alguna persona visite tal sitio, este archivo se descargará infectando al sistema involucrado.

Aunado a lo anterior, distintos vectores de ataque son utilizados por la manera en que se usa o se creo, por ejemplo: los dispositivos extraíbles se crearon para poder llevar información a cualquier lugar sin necesidad de cargar una computadora; en estos casos, las acciones que realiza el usuario son fundamentales (como no abrir archivos que pongan en riesgo la seguridad del dispositivo) para que el escritor de malware no puede realizar un ataque, ya que no existen actualizaciones de seguridad que puedan solucionar estos problemas.

2.2.3. Tipos de malware

Existen distintos tipos de malware, se puede clasificar [8] como:

- Virus: son capaces de replicarse a sí mismo y propagarse a otras computadoras, debido a la acción del usuario. Se propagan infectando programas para que al momento de que el usuario ejecute dicho programa se propague a la computadora o al dispositivo conectado, por ejemplo: el virus infecta un programa ejecutable en una memoria, si el usuario ejecuta dicho programa en cada computadora que utilice, cada computadora quedará infectada.

- Gusanos: se propaga sobre redes de computadoras. A diferencia de un virus, un gusano puede auto replicarse y propagarse por sí solo, sin la necesidad de que el usuario intervenga físicamente en cada dispositivo. Por ejemplo, si un usuario abre un archivo infectado con un gusano, el gusano se transmitirá por si solo a los demás dispositivos conectados en la misma red.

- Troyanos: imitan un programa auténtico. Cuando se ejecute, se pide autorización para descargar e instalar software adicional, en este caso malware, otorgando acceso al sistema. En muchos casos el acceso es remoto, haciendo vulnerable al sistema.

- Rootkits: se utilizan para acceder remotamente a una computadora, derivado de un troyano.

- Spyware: su principal función es espiar la actividad del usuario, sin su consentimiento.

- Adware, bombas lógicas, keyloggers, rasonware, entre otros.

2.3. Gusanos en SMS

La infección utilizando como vector de ataque SMS, se realiza a través de malware tipo gusano, los cuales aprovechan las estructuras de red formadas por los dispositivos móviles. Básicamente, la dinámica de infección (ver Figura 2.8) consiste en alojar el gusano en un servidor web público (1), el atacante envía un mensaje trampa a algunos primeros usuarios mediante SMS (2), el mensaje suele ser atractivo para el usuario y contiene un enlace hacia el servidor que aloja el malware, en caso de que el usuario abra (3) y descargue el contenido quedará infectado (4), el gusano utilizará la lista de contactos para enviar un mensaje SMS de la misma forma que lo hizo el atacante (5), esto se repite hasta que los usuarios queden infectados (6,7) o dejen de abrir los mensajes SMS.

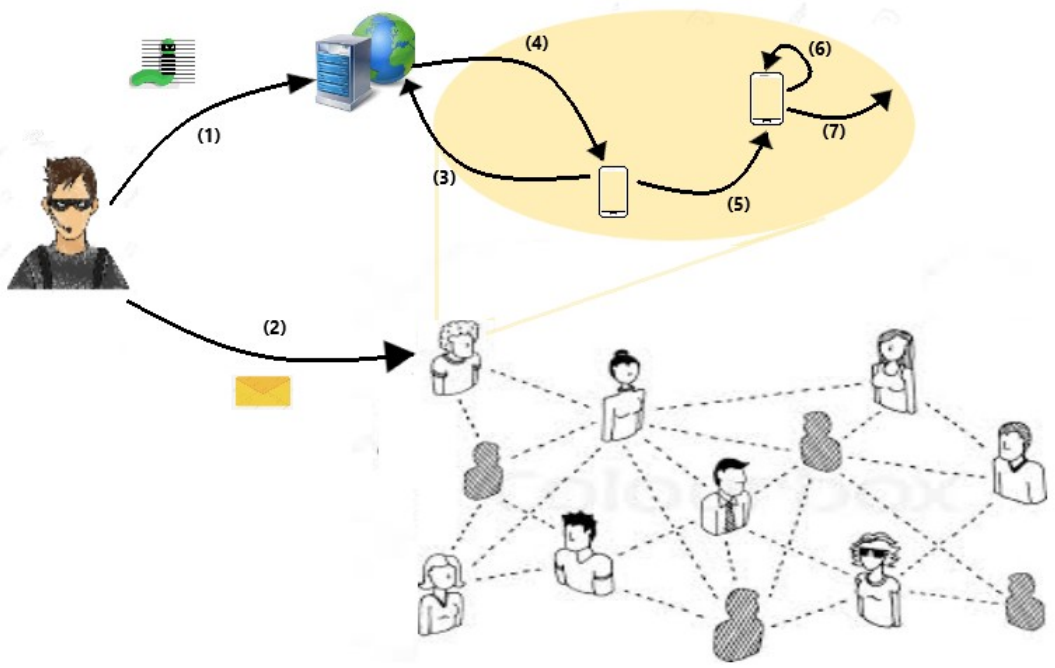


Figura 2.8: Dinámica de propagación de gusano sobre SMS.

Es importante señalar qué para que un dispositivo móvil sea infectado, debe estar conectado tanto a una red telefónica móvil como a Internet (puede ser a través de la misma red telefónica), pues el dispositivo necesita poder recibir y enviar mensajes SMS, así como descargar archivos de un servidor web.

2.3.1. Breve revisión actual del malware a través de SMS

En el año 2016 [27], se detectó un malware tipo gusano llamado Mazar, el cual se propagaba a través de mensajes SMS utilizando la lista de contactos, una vez infectado el atacante podía borrar toda la información del celular o robarla; se estimó que fueron enviados mensajes SMS a más de 100,000 celulares con el enlace de descarga infectado. En el año 2018 [28], un investigador de malware de ESET detectó una aplicación la cual

enviaba mensajes SMS para que descargarán la aplicación, una vez que se infectaba el dispositivo podían robar información bancaria del dispositivo . También en el año 2018 [29], se observaron nuevas técnicas de infección de dispositivos móviles y un aumento en el uso de esquemas de distribución (por ejemplo, spam de SMS), centrándose en ataques a cuentas bancarias a través de dispositivos móviles, aplicaciones de adware, Droppers y Risktool. También, productos y tecnologías de Kaspersky Lab detectaron [30]: 5,321,142 instalaciones de paquetería maliciosa, 151,359 nuevos bancos de troyanos móviles y 60,176 nuevos tipos de troyanos ransomware móviles . Particularmente, México quedo como sexto país con mayores detecciones de malware a nivel internacional [31], siendo el primer lugar en Latinoamérica. Aunado a lo anterior, se encuentra la importancia de estudiar y predecir el comportamiento del malware.

Una vez presentados los conceptos relacionados a SMS y malware, en el siguiente capítulo se presenta el estado del arte del área de estudio de este trabajo, se muestra una breve descripción de los modelos existentes, así como características de cada uno de ellos.

Capítulo 3

Modelos de propagación de malware

En este capítulo se presenta una revisión del estado del arte de modelos epidemiológicos relacionados a la propagación de malware, empleando: teoría de Autómatas Celulares, teoría de grafos y la combinación de anteriores.

Actualmente, estudiar el comportamiento de cualquier enfermedad infecciosa, ha sido de vital importancia para la salud. La epidemiología es una disciplina encargada de estudiar la distribución, la frecuencia, así como la gravedad de los problemas de salud y qué los causan [32]. Poder predecir el comportamiento, así como tomar medidas para minimizar los daños provocados por estas enfermedades es llevado por investigadores en el área de la medicina.

La combinación de distintas áreas, como lo son las matemáticas y la epidemiología, han dado mejores resultados para estudiar cualquier caso de enfermedad. Entre las décadas 1650 y 1670, el inglés Thomas Sydenham utilizó estadísticas para hacer una clasificación de la disentería, la malaria, la viruela, la gota, la sífilis y la tuberculosis, reconociendo a cada enfermedad como distinta de la otra [33]. Más adelante en 1747, analizó las causas que provoca el escorbuto, publicando un trabajo acerca de ello en 1753 [34]. Finalmente en 1760, Daniel Bernoulli publicó un trabajo relacionado con la viruela, encontró que la variolación (método usado antes de la existencia de la vacuna) protegía contra esta y de por vida [35]. Sin embargo, los trabajos publicados describían características propias de las enfermedades, en algunos casos para clasificación o relacionado con el tratamiento o cura. Hasta el momento, no había un estudio formal acerca del comportamiento de propagación de las enfermedades en la población.

Fue hasta 1926, donde Anderson Gray McKendrick publicó un artículo llamado: "Aplicaciones de matemáticas para problemas médicos" [36], donde introdujo un nuevo modelo continuo matemático para modelar el comportamiento de epidemias tomando en cuenta procesos estocásticos de infección y recuperación. El modelo consideró tres estados en los cuales la población puede estar, es decir, en tres compartimentos:

- Susceptible: denotado por S, son individuos considerados sanos que pueden ser infectados.
- Infectado: denotado por I, son individuos que han sido infectados y pueden propagar la enfermedad.

- Recuperado: denotado por R, son individuos que se han recuperado de la enfermedad, tomando alguna medida contra esta.

Además de una población inicial de tamaño N, con una persona infectada.

Todo lo anterior, se pudo reflejar en la propagación de malware, en especial del tipo gusano, debido a que el gusano puede ser visto como una enfermedad de tipo infección y que es propagada por algún vector de ataque. Empezando de esta forma a considerar utilizar teoría epidemiológica compartimental (dividir a la población en compartimentos o grupos) para poder comprender, estudiar y analizar la dinámica de propagación del gusano en dispositivos.

3.1. Modelos epidemiológicos

William Ogilvy Kermack en colaboración con McKendrick, escribieron varios documentos titulados: "Contributions to the mathematical theory of epidemics", publicando la primera versión en 1927 [37], la segunda en 1932 [38] y la tercera en 1933 [39], donde estudiaron modelos epidemiológicos determinísticos. El primer artículo consideraba una población de tamaño N muy grande, con los mismos estados posibles que en el artículo de McKendrick de 1926 (susceptible, infectado y recuperado), dónde $S(t)$, $I(t)$ y $R(t)$ denota al número de personas en cada diferente estado respectivamente. El modelo puede ser simplificado a través de tres ecuaciones diferenciales:

$$\begin{aligned}\frac{dS}{dt} &= -aSI,\\ \frac{dI}{dt} &= aSI - bI,\\ \frac{dR}{dt} &= bI.\end{aligned}$$

con la condición inicial $S(0) > 0$, $I(0) > 0$ y $R(0) = 0$.

Denotando al modelo anterior SIR por la sigla de sus estados, siendo Kermack y McKendrick los que asentaron las bases de la epidemiología compartimental mediante estos artículos. Cualquier modelo puede ser representado mediante un diagrama de transición de estados, como se muestra en la Figura 3.1. La epidemiología compartimen-

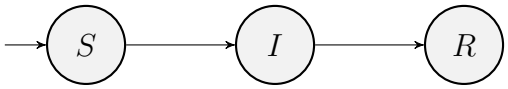


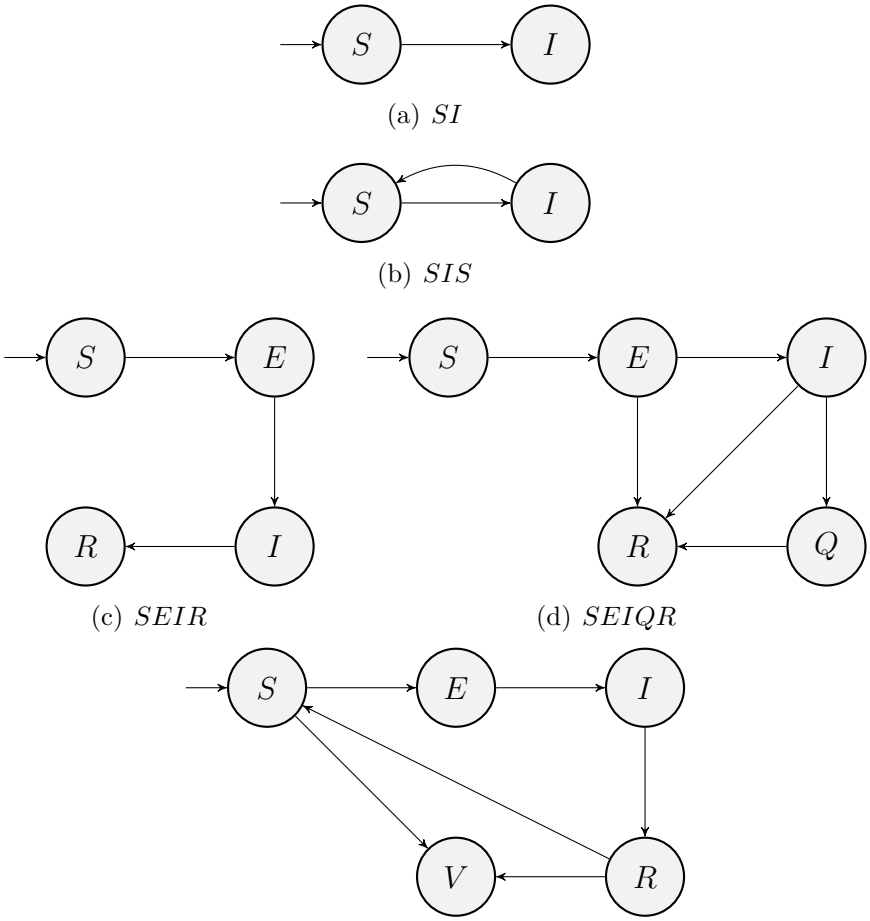
Figura 3.1: Diagrama de transición de estados del modelo SIR.

tal divide a la población estudiada en diferentes grupos o compartimentos, dependiendo de su estado ante la enfermedad. Como se muestra en la Figura 3.2, otros modelos determinísticos existentes son:

- SI: Susceptible-Infectado
- SIS: Susceptible-Infectado-Susceptible

- $SEIR$: Susceptible-Expuesto-Infectado-Recuperado
- $SEIQR$: Susceptible-Expuesto-Infectado-Cuarentena-Recuperado
- $SEIRS-V$: Susceptible-Expuesto-Infectado-Recuperado-Susceptible y Vacunado

A partir de estos modelos, han surgido combinaciones o variantes de los mismos.



(a) SI (b) SIS (c) $SEIR$ (d) $SEIQR$ (e) $SEIRS-V$

Figura 3.2: Diagrama de transición de estados de modelos epidemiológicos existentes.

Cada modelo se caracteriza por hacer suposiciones acerca de las variables (características que cambian en el tiempo), los parámetros (características que no cambian en el tiempo) y relaciones funcionales entre las variables y parámetros que gobiernan la dinámica de las variables. De esta manera los modelos abarcan las hipótesis de los modelos estudiados y se pueden comparar estas hipótesis con los datos empíricos.

3.1.1. Clasificación de modelos

Por lo anterior, cada modelo utiliza aplicaciones matemáticas así como estadísticas para poder ser más fiel y apegarse a lo que sucede en la realidad, en cuestión con

alguna enfermedad. Sin embargo, podemos clasificar a estos modelos de acuerdo al tipo de herramienta o método matemático que utiliza cada uno. La clasificación es la siguiente [8, 40]:

- Modelos basados en ecuaciones diferenciales (o modelos determinísticos): como el modelo de Kermack y McKendrick, fueron de los primeros modelos en ser desarrollados. Son representados mediante ecuaciones diferenciales. Se asume que el tamaño de la población susceptible e infectada es un función definida en el tiempo.

- Modelos basados en procesos estocásticos (o modelos estocásticos): la población en este tipo de modelos es representada mediante procesos estocásticos, estos modelos pueden ser descritos mediante interrelaciones de su distribución de probabilidad.

- Modelos basados en individuos (o modelos espacio-temporal): estos modelos utilizan fuertemente la teoría de Autómatas Celulares (AC), que se describe en la siguiente sección 2.3 de este capítulo. Básicamente, se asume un número grande de componentes simples con interacciones locales y son capaces de simular sistemas completos en su proceso de evolución espacio-temporal.

Se resume en el Cuadro 3.1 las diferentes características de los tipos de modelos epidemiológicos.

Tipo	Basado en Ecs. Dif.	Basado en Procesos Estocásticos	Basado en Individuos
Teoría	Ecuaciones Diferenciales	Procesos Markovianos	Autómatas Celulares
Espacio	Continuo	Continuo	Discreto
Tiempo	Continuo	Continuo o discreto	Discreto
Estado individual	Continuo	Discreto	Discreto
Interacción individual	No	No	Si
Alcance adaptativo	Individuos con movimiento aleatorio	Número pequeño de individuos	Número grande de individuos
Descripción del modelo	Ecuaciones diferenciales	Cadenas de Markov en tiempo continuo o discreto	Reglas de evolución estocásticas

Cuadro 3.1: Cuadro comparativo de los diferentes tipos de modelos epidemiológicos.

3.1.2. Ventajas y desventajas

Los modelos basados en ecuaciones diferenciales pueden describir las interrelaciones dinámicas entre las tasas de cambio y tamaño de la población. Además de que la teoría

matemática de este tipo de modelos ha sido muy bien desarrollada y son adecuados para hacer predicciones. Sin embargo, no consideran las interacciones locales que surgen entre los individuos, siendo un análisis de forma macroscópica, por lo que, no pueden simular la dinámica individual de cada individuo; además, consideran que todos los individuos deben infectarse homogéneamente, evento que no es muy apegado a la realidad.

Los modelos basados en procesos estocásticos son adecuados para estudiar poblaciones pequeñas. Sin embargo, al igual que los modelos basados en ecuaciones diferenciales, no consideran las interacciones.

Los modelos basados en individuos se convierten en una importante herramienta para estudiar la evolución espacio-temporal de sistemas auto-organizables y caracterizar los sistemas complejos basados en reglas de evolución locales. De esta forma, puede cubrir las deficiencias que los demás tipos de modelos tienen, dando características propias a cada individuo.

3.2. Autómatas Celulares

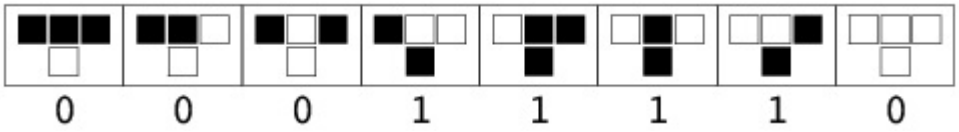
Al principio de la década de los 50's, John von Neumann comenzó a interesarse en los procesos de auto-replicación. Estudió los mecanismos que proporcionan a una máquina los medios de replicarse a sí mismos. Von Neumann definió el constructor universal, una máquina capaz de construir cualquier otra máquina, dando su descripción. La idea de von Neumann fue lograr la auto-replicación mediante el constructor universal y su propia descripción [41]. Para su universo, von Neumann escogió una matriz infinita, donde cada célula es una máquina de estados finitos. Después de varios intentos, definió 29 estados y su regla de transición. Demostró que dando una configuración su autómata lograba auto-replicarse.

Un Autómata Celular (AC) es un sistema dinámico discreto, en tiempo y espacio; está compuesto por células y la evolución del sistema depende del estado anterior, es decir, cada célula influirá en la siguiente evolución. Es básicamente un tipo de máquina de estados finito que es capaz de simular sistemas complejos de manera eficiente y eficaz. El estado de cada célula definirá el estado general del sistema [42]. Por lo anterior, es que se consideran interacciones locales y, por lo tanto, aspectos individuales haciendo posible una simulación de cada individuo de una manera muy simple. Existen muchas aplicaciones de los autómatas celulares en áreas como: Ciencias de la Computación, Biología, Bioinformática, Criptografía o Ingeniería.

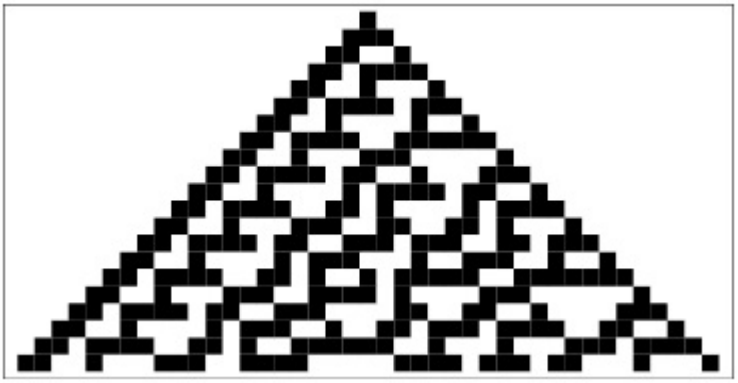
Un AC puede ser considerado como un sistema compuesto por un arreglo de células A. Cada célula c_i representa un autómata finito con un conjunto de estados Q, en un alfabeto Σ y una función de transición $\delta : Q \times \Sigma \to Q$. La entrada del alfabeto Σ está dada por todas las posibles combinaciones del estado de las células adyacentes de cada célula. Cada célula c_i y sus célula adyacentes son consideradas todas juntas y representadas cómo el único conjunto $N = \{c_i\} \cup N^{-c_i}$, donde N^{-c_i} es el conjunto de células consideras como vecinas de una arbitraria célula c_i. Entonces, se puede definir al Autómata Celular como una 4-tupla $M = \{A, Q, \delta, N\}$ [43]. Los componentes de un Autómata Celular son:

- Espacio celular: es el espacio físico donde evoluciona el autómata, al ser un arreglo

este tiene dimensión D o infinita. Los autómatas con $D = 1$ son llamados autómatas lineales, un ejemplo de ellos son los AC elementales [44], como se muestra en la Figura 3.3. Los autómatas con $D = 2$ representan una superficie plana, generalmente representada por una cuadricula, se muestra un ejemplo en la Figura 3.4.



(a) Regla de transición del AC elemental



(b) Evolución del AC elemental en el tiempo

Figura 3.3: Ejemplo de un AC elemental. Imágenes tomadas de [1].

- Función de transición: está función puede ser una regla o un conjunto de reglas que definirán una evolución en el tiempo, provocando un cambio de estado en el sistema, por lo que el conjunto de reglas δ definen la dinámica del AC. Dada una célula i en un estado k_i y el conjunto de estados k_{N_i} de los vecinos N de i, en un instante de tiempo t, δ es determinista y calcula el siguiente estado de tiempo $t + 1$ para la célula i.

- Variable de tiempo: La dinámica del sistema celular se desarrolla a lo largo de un tiempo discreto.

- Vecindad: son las células que influirán en la evolución de una célula. Por lo que a cada célula i es necesario asignar un conjunto de células llamadas vecinas, incluyéndose a sí misma. En caso de espacios bidimensionales, las vecindades más comunes son: Moore y von Neumann, las cuales se les puede asignar un radio, como se muestra en la Figura 3.5.

- Frontera: en el caso de que el espacio celular sea finito, la frontera es la condición en que se encuentran las células en el están en el perímetro del espacio celular. Se tienen diferentes tipos de frontera:

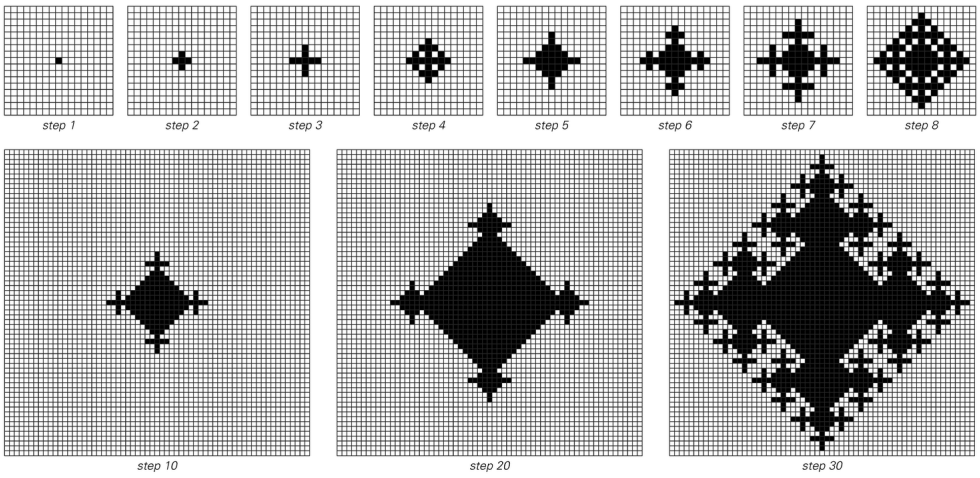
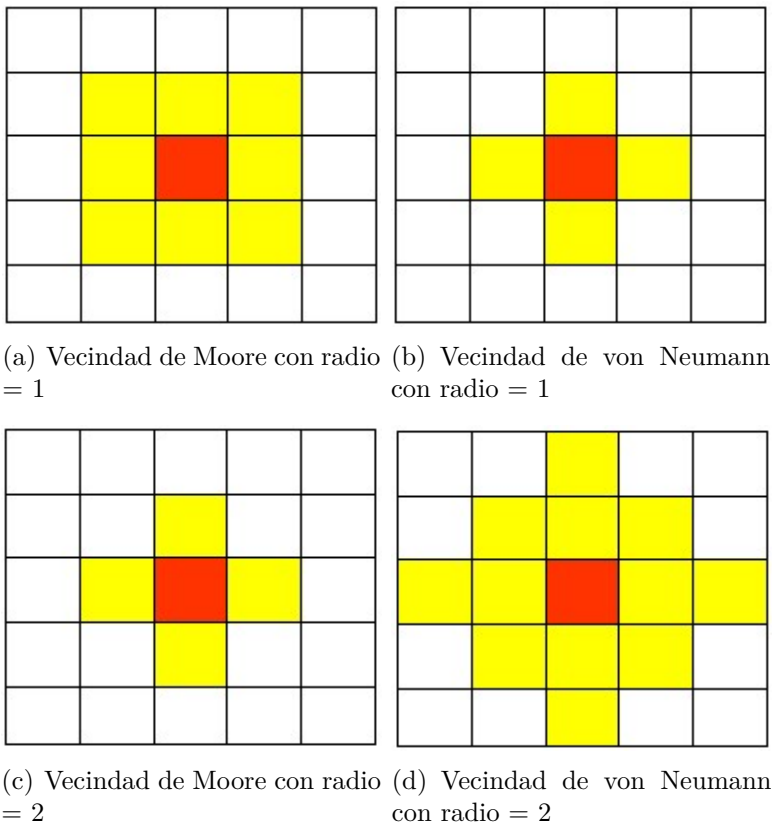


Figura 3.4: Ejemplo de evolución en el tiempo de un AC con $D = 2$. Imagen tomada de [2].



(a) Vecindad de Moore con radio = 1

(b) Vecindad de von Neumann con radio = 1

(c) Vecindad de Moore con radio = 2

(d) Vecindad de von Neumann con radio = 2

Figura 3.5: Ejemplo de vecindades con espacio bidimensional.

- Periódica: el espacio celular es enlazado del principio a fin, por ejemplo, cuando se tiene una vecindad de Moore, las células en la orilla del espacio celular se verán afectadas por las células en el otro extremo del espacio. En la Figura 3.6, se observa un ejemplo de frontera periódica, donde la célula roja es la célula de interés.

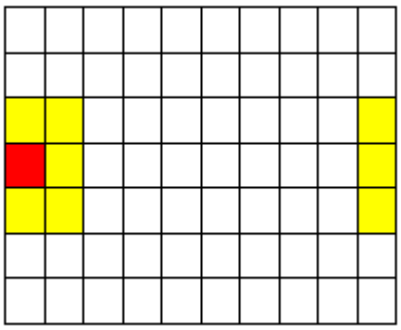


Figura 3.6: Frontera de tipo periódica.

- Asignada: las células en la perímetro del espacio celular tienen un estado fijo durante el paso del tiempo. Se muestra un ejemplo en la Figura 3.7, donde las célula roja es la célula de interés, las células amarillas es la vecindad y las células verdes ya tienen un estado fijo.

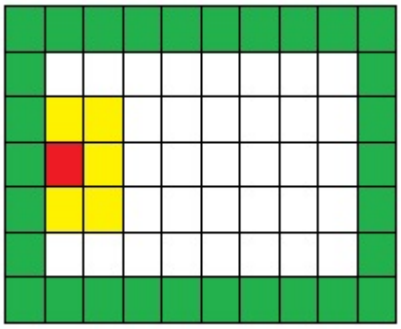


Figura 3.7: Frontera de tipo asignada.

- De reflexión: considerando una vecindad de Moore, una célula en la frontera se verá afectada por las células que estén dentro del espacio, las células faltantes serán aquellas tomadas en espejo o reflexión. En la Figura 3.8 se observa un ejemplo, las células amarillas y naranjas son parte de la vecindad de la célula roja, mientras las células verdes son aquellas que reflejarán su estado de las células naranjas.
- Sin frontera: el espacio celular puede crecer tanto como sea necesario. Como se observa en la Figura 3.9, puede haber casos donde una célula "se mueva" por lo que es necesario aumentar el espacio celular.

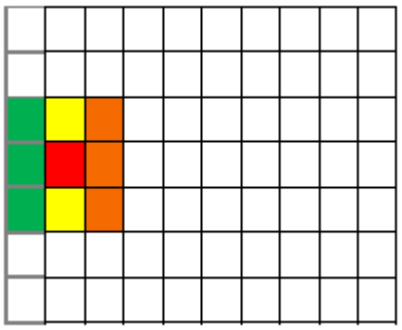
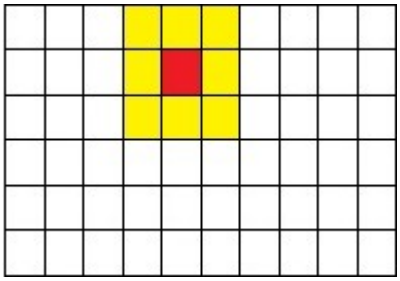
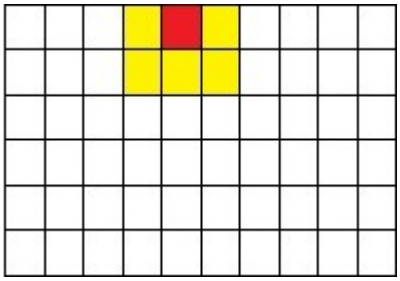


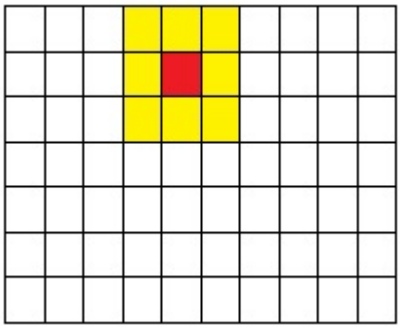
Figura 3.8: Frontera de tipo reflexión.



(a) Ejemplo de una célula que se va a mover por el espacio celular.



(b) Ejemplo de una célula que se movió por el espacio celular.



(c) Ejemplo del crecimiento del espacio celular debido al movimiento de una célula.

Figura 3.9: Ejemplo de un espacio celular sin frontera.

- Adiabática: se asigna a la frontera el estado de otras células arbitrarias, es decir, se copia su valor. En la Figura 3.10 se observa un ejemplo de este tipo de frontera.

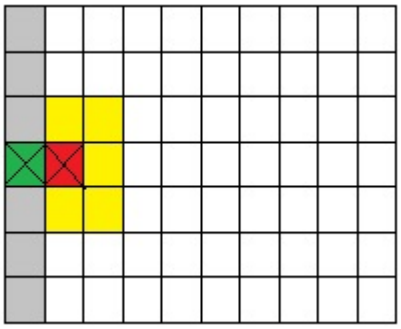


Figura 3.10: Frontera de tipo adiabática.

- Abierta: No existe una condición en la frontera, suele utilizarse para simular espacios celulares infinitos teniendo un espacio finito.

- Estados: es el conjunto de estados posibles en los que el sistema puede estar, en particular las células, en el tiempo t. El conjunto de estados más simple son aquellos llamados biestables, donde solamente pueden tener dos posibilidades: 0 y 1, como los AC elementales.

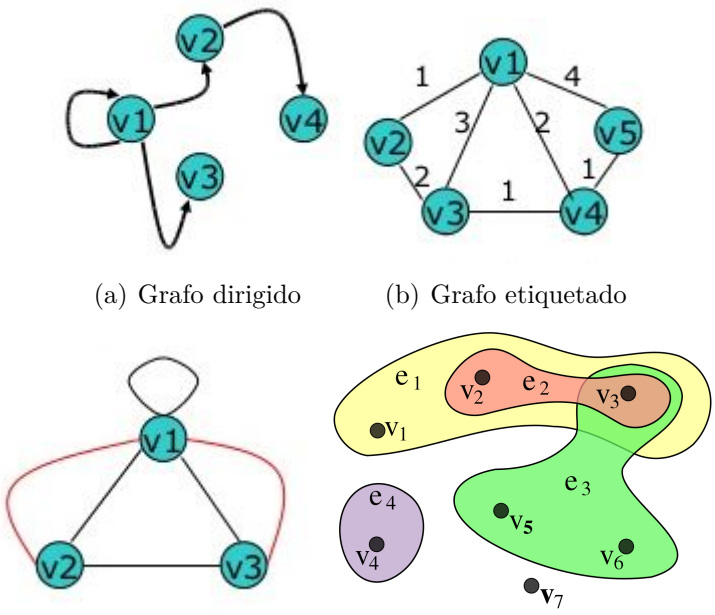
A partir de la definición de AC, se han hecho modificaciones a los mismos como modificar el espacio celular, surgiendo así una nueva teoría de AC. Particularmente, si cambiamos el espacio celular de un arreglo a un grafo, obtenemos un nuevo autómata llamado Autómata Network.

3.3. Grafos y Autómatas Networks

Poder modelar sistemas es una tarea realmente complicada, pero cuando se intenta modelar sistemas complejos es una tarea aún más complicada. Un sistema complejo, es aquel sistema que está compuesto por diversas entidades, pudiendo ser otros sistemas o individuos. A cada entidad se le define un comportamiento, pero al momento en que estas entidades empiezan a interactuar, no se sabe como se comportará todo el sistema, dando como resultado nuevas propiedades llamadas propiedades emergentes del sistema. Por lo anterior, un sistema complejo no es fácil de describir, debido a que no se sabe cual será el estado del sistema en cierto tiempo.

Muchos sistemas han podido ser representados a través de grafos, recordando que un grafo G puede representarse a través de una tupla $\{V, E\}$, donde V es el conjunto de nodos y E es el conjunto de aristas. Debido a la necesidad de representar sistemas complejos, se han realizado modificaciones a los grafos llamados simples, obteniendo grafos (ver Figura 3.11) como:

- Grafos dirigidos: Las aristas tienen una dirección, son representadas por flechas.
- Grafos etiquetado: Las aristas tienen un valor o peso.
- Multigrafos: Se permite más de una arista entre dos nodos.
- Hipergrafo: Las aristas pueden tener más de un extremo.



(a) Grafo dirigido (b) Grafo etiquetado

(c) Multigrafo (d) Hipergrafo. Imagen tomada de [45]

Figura 3.11: Ejemplos de otros tipos de grafos.

Además de los ejemplos anteriores, han surgido combinaciones de los mismos. Como son las llamadas redes complejas. Una red compleja es una red, representada mediante un grafo el cual tiene estadísticas y características específicas en su estructura o topología. Algunos ejemplos de este tipo de redes son:

- Redes libres de escala (SFN): son grafos donde el grado de sus nodos sigue una distribución de Ley de Potencia [46] (ver Figura 3.12), es decir, la mayoría de los nodos se encuentran conectados a pocos nodos mientras que la minoría se encuentra altamente conectado a muchos nodos, en la Figura 3.13 se muestra un ejemplo de una SFN.
  La distribución se puede describir de la siguiente forma:

  $$P(k) = k^{-\gamma}$$

  donde $\gamma > 0$.

  De acuerdo al Algoritmo de Barabási Albert [47], se puede construir una red sin escala, de la siguiente forma:

1. Se comienza la red con m nodos.
2. Conectar todos los nodos en el grafo para crear un grafo completo.
3. Crear un nuevo nodo i.
4. Elegir un nodo j del grafo de forma aleatoria, estableciendo una probabilidad

$$P = \frac{k_i}{\Sigma_j k_j}.$$

   donde: k_i y k_j son los grados de los nodo i y j, respectivamente.
5. Obtener un número real aleatorio uniforme R entre 0 y 1. Si $P > R$ entonces se conecta el nodo i con j.
6. Repetir del paso 3 al 5 hasta completar N nodos.

Se ha encontrado que es adecuado utilizar una SFN para representar una red social [46, 48].

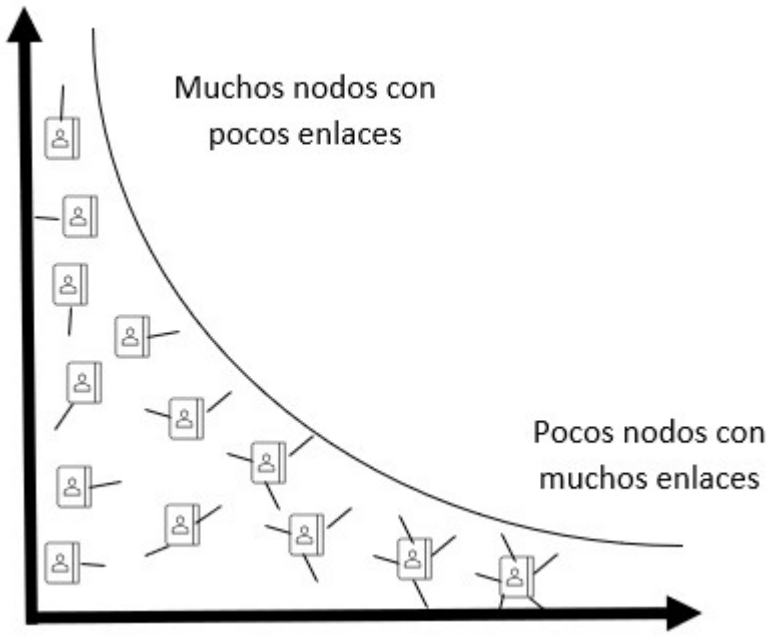


Figura 3.12: Ley de la Potencia.

- Redes de mundo pequeño: son grafos donde la distancia entre cualquier par de nodos es relativamente pequeña mientras que al mismo tiempo el nivel de transitividad o clustering es relativamente alto [49]. En la Figura 3.14 se muestra un ejemplo.

Entonces, se puede definir un Autómata Network N, como una 3-tupla $(G, Q, \{f_i \| i \in N\})$ [50], donde G es un grafo en N, Q es un conjunto de estados y $f_i : Q^{|U_i|} \to Q$ un mapeo, llamado regla de transición asociada al vértice i. $U_i = \{j \in N | \{j, i\} \in E\}$ es

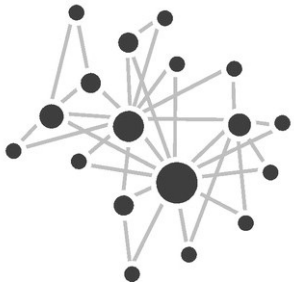


Figura 3.13: Ejemplo de una SFN. Imagen tomada de [3].

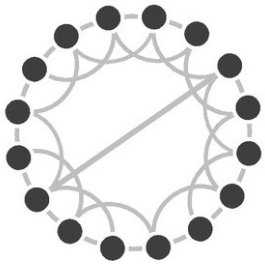


Figura 3.14: Ejemplo de una red de mundo pequeño. Imagen tomada de [3].

la vecindad de i, por ejemplo, el conjunto de vértices que conectan a i y $|U_i|$ denota el número de vértices que unen a U_i. El grafo G se asume es localmente infinito, por ejemplo, para todo $i \in N$, $|U_i| < \infty$.

Un Autómata Network sigue la definición de un AC, pues solamente es cambiado el espacio celular. Se siguen contemplando las interacciones locales, esto es cuando un nodo interactúa con otro, a través de la función de transición. Y una vez ejecutada la función de transición, indicará a que estado se cambiará. Por lo anterior, sigue preservando las características importantes de un AC, así como simplicidad.

Una vez revisados conceptos fundamentales sobre modelos epidemiológicos, así como las herramientas necesarias para modelarlos, en la siguiente sección se presenta una breve revisión del estado del arte enfocado a la propagación de malware en dispositivos móviles mediante SMS. Debido a que se utiliza la lista de contactos de cada dispositivo en la propagación, SMS puede ser representado como una red social, particularmente una SFN, ya que por lo anterior, una SFN es adecuada para representar este tipo de redes. Entonces, se presentan también modelos basados en redes complejas, como es el caso de las SNF.

3.4. Modelos de propagación de gusanos

El modelado de la dinámica de propagación de malware ha sido un tema importante en estos últimos años, cada vez existen y abundan más dispositivos conectados a las redes telefónicas. Hoy en día, existen diversos tipos de modelos relacionados al estudio de la propagación de malware en dispositivos móviles, estos modelos están acompañados

por el enfoque de epidemiología compartimental utilizando algún tipo de herramienta matemática como: ecuaciones diferenciales, procesos estocásticos o autómatas celulares. Estos modelos han sido de gran ayuda para poder predecir, analizar y determinar el comportamiento de algún malware en especial, gracias a las simulaciones computacionales.

Actualmente, se puede encontrar una gran variedad de modelos matemáticos para describir la propagación de malware tipo gusano a través de mensajes SMS o en algún tipo de red (social o compleja). A continuación se resumen algunos de los trabajos más importantes al respecto.

En 2018, Teresa Signes et al. [51] presentaron un modelo SEIRS (Susceptible - Expuesto - Infectado - Recuperado - Susceptible) considerando equipos enlazados a través de Internet. Utilizaron AC para el modelado, los equipos son conectados si están en la misma vecindad, utilizando diferentes vecindades como la de Moore o de von Neumann. Sin embargo, al representar una red como lo es Internet, utilizar un AC es poco fiel a la realidad pues no se establece una conexión por la cercanía entre dispositivos, además de que se limita la conexión entre los mismos por la vecindad definida.

Xi Xiao et al., en 2017 [52, 53], presentaron un modelo SAIDR (Susceptible - Afectado - Infectado - Latente - Recuperado), básicamente es el modelo SIR agregando dos estados. El estado A representa a dispositivos afectados, es decir, dispositivos que recibieron el enlace malicioso mediante un mensaje SMS pero no lo han abierto. El estado D representa a dispositivos en estado latente, es decir, dispositivos que no pueden enviar mensajes maliciosos a otros. Consideraron tres factores para el proceso de infección: revisión de mensajes, la relación de amistad entre dispositivos y la consciencia de seguridad en los usuarios. Sin embargo, el primer trabajo presentado fue basado en individuos, mientras que el segundo trabajo fue basado en ecuaciones diferenciales. En ambos no consideraron utilizar una topología para representar de mejor forma la red formada por SMS. También, en el mismo año [54], presentaron un nuevo modelo SEIQR basado en ecuaciones diferenciales, agregando un estado de cuarentena (Q) al modelo SEIR. Sin embargo, se presentan las mismas observaciones que en los trabajos anteriores.

Peng Jia et al., en 2018 [55], elaboraron un modelo HSID (Heterogéneo - Susceptible - Infectado - Latente). La letra H corresponde a la consideración de tasas de infección homogéneas para cada nodo. También, consideraron la consciencia de seguridad de usuarios así como el tiempo de infección de un nodo. Las tasas de infección siguen una distribución normal o ley de potencia. Sin embargo, al igual que el trabajo anterior, no consideran utilizar una topología de red para representar la red.

En 2016, Chunming Zhang y Haitao Huang [56] propusieron un modelo SLBOS, clasificando a los nodos conectados en Internet en cuatro tipos, divididos a la vez en dos: estando dentro del sistema (libre de virus (S), con virus latente (L) y con virus activo(B)) y fuera del sistema (O). Ocuparon una SFN para representar la topología de la red. Sin embargo, el factor de infección es homogéneo en todos los nodos, dependiendo únicamente solamente de los nodos infectados vecinos o alrededor.

Wanping Liu et al., en 2016 [57], presentaron un modelo WSIS, considerando dos niveles de protección para los nodos: débilmente protegidos (W) y fuertemente protegidos (S), en conjunto con los estados susceptibles (S) e infectado (I). Utilizan una SFN para

representar la topología de la red. Sin embargo, no se considera utilizar otros factores como la relación entre nodos o la consciencia de seguridad de los usuarios.

Soodeh Hosseini y Mohammad Abdollahi Azgomi, en 2016 [58], elaboraron un modelo SEIR-V (Susceptible - Expuesto - Infectado - Recuperado - Vacunado), consideraron que un nodo pueda ser recuperado así como vacunado, debido a la consciencia de seguridad de los usuarios o por un antivirus. Utiliza una topología de SFN, hacen un caso de estudio cuando los nodos se desconectan y conectan. En 2018 [59], presentaron un modelo SEIRS-QV (Susceptible-Expuesto-Infectado-Recuperado-Susceptible-Cuarentena-Vacunado) basado en ecuaciones diferenciales. Utilizaron una SFN para la topología de la red agregando los estados de cuarentena (Q) y vacuna (V). Sin embargo, no contempla la consciencia de seguridad en los usuarios.

Soodeh Hosseini et al., en 2016, [60] presentaron un modelo SEIRS, relacionado con el trabajo anterior. Agregan un nuevo estado R (rígido) al modelo SEIS (Susceptible - Expuesto - Infectado - Susceptible). El nodo es llamado rígido si pierde la tendencia a ser infectado por malware después de ya haber recibido el mismo. Sin embargo, no contemplan utilizar alguna topología para representar su red, así como la relación de usuarios o consciencia de seguridad en estos.

Chenquan Gan et al., en 2017 [61], presentaron un modelo SIES (Susceptible - Infectado - Externo - Susceptible), considerando el modelo SIS agregando un estado nuevo, llamado externo. El estado externo se utiliza para los nodos que están desconectados de la red, en este caso de Internet. Sin embargo, no utilizan alguna topología en la red para representar de mejor forma la realidad. No contemplan la relación entre usuarios.

En 2018, Shouying Huang [62] presentó un trabajo basado en [57], considerando una tasa de inmunidad heterogénea en los nodos y una topología de red (SFN). Sin embargo, presenta las mismas observaciones que el trabajo presentado por Wanping Liu.

Xiongding Liu et al., en 2018 [63], presentan un modelo SDIRS (Susceptible - Latente - Infectado - Recuperado - Susceptible) basado en ecuaciones diferenciales, consideran una distribución de ley de la potencia para los grados de los nodos de la red, es decir, una SFN. También, consideran un estado de latencia para los nodos así como que la red se mantenga estática, es decir, que no se desconecten o conecten nuevos nodos. Sin embargo, presenta las observaciones que otros trabajos al ser basados por ecuaciones diferenciales. En 2019, [64] presentaron un modelo ICST basado en ecuaciones diferenciales, se consideraba transmitir información entre usuarios, los cuales eran marcados por cuatro clases: los que ignoran (I), los que comentan (C), los que comparten información (S) y los que no responden (T). Utilizaron una SFN para la topología de la red. Sin embargo, presente las mismas observaciones que trabajos anteriores.

Maria Selvam et al., en 2018 [65], elaboraron un modelo SIQR basado en ecuaciones diferenciales, consideran agregar un nuevo estado cuarentena (Q) al modelo SIR ya trabajado. Se basan en una red de sensores, por lo que la topología de la red es aleatoria. Sin embargo, la topología de la red ya está definida y no sigue alguna distribución pues esta debería de estar determinada por el espacio físico.

En 2015, Xiaochun Yun et al. [66] elaboraron un modelo analítico RTSS basado en proceso estocásticos y en el modelo SIR, consideraron la reputación de sus nodos (R), el grado de creencia o de verdad (T) así como dos estados susceptibles (SS). Los

estados susceptibles consideran la dinámica de checar mensajes de los usuarios, por lo que el primer mensaje define a usuarios que están en un estado de no revisar mensajes, considerándolos inmunes y el segundo estado define a usuarios que están en un estado de revisar mensajes por lo que están susceptibles a abrir y leer un mensaje. Además, el proceso de infección se hace a través de diferentes dominios. El modelo contempla una topología en la red para representar de mejor forma la red (SFN), y además al ser basado en procesos estocásticos solo considera poblaciones de individuos pequeñas y tampoco interacciones locales.

Sancheng Peng et al., en 2013 [67], presentaron un modelo SIR basado en procesos semi-Markovianos, utilizaron un grafo para representar la red social, siguiendo un patrón de comunicación basado en los mensajes enviados entre nodos. Consideran que los factores de infección, así como de resistencia, son diferentes para cada nodo, esto debido a la relación de los nodos mismos. En 2014 [68], presentaron un modelo SEIR basado en proceso Markovianos, consideraron utilizar un grafo con peso en sus aristas, las cuales estaban relacionadas con la cantidad de mensajes enviados entre nodos. Sin embargo, los dos trabajos presentan las mismas observaciones que el trabajo anterior.

Modelo	Teoría	Diferencias Individuales	Topología de red	Revisión de mensajes	Relación nodos	Seguridad dispositivos
Signes [51]	AC	SÍ	NO	NO	NO	NO
Xiao [52]	BI	SÍ	NO	SÍ	SÍ	SÍ
Xiao [53]	ED	NO	NO	NO	NO	NO
Xiao [54]	ED	NO	NO	NO	NO	NO
Jia [55]	BI	SÍ	NO	NO	SÍ	SÍ
Zhang [56]	ED	NO	SÍ	NO	NO	NO
Liu [57]	ED	NO	SÍ	NO	NO	NO
Hosseini [58–60]	ED	NO	SÍ	NO	NO	SÍ
Gan [61]	ED	NO	NO	NO	NO	NO
Huang [62]	ED	NO	SÍ	NO	NO	NO
Liu [63, 64]	ED	NO	SÍ	NO	NO	NO
Selvam [65]	ED	NO	SÍ	NO	NO	NO
Yun [66]	PE	NO	SÍ	SÍ	SÍ	NO
Peng [67, 68]	PE	NO	NO	SÍ	SÍ	SÍ

Cuadro 3.2: Cuadro comparativo de los diferentes tipos de modelos epidemiológicos disponibles en la literatura. Donde ED significa ecuaciones diferenciales, PE significa procesos estocásticos y BI basado en individuos.

En el Cuadro 3.2 se muestra un cuadro comparativo de los modelos antes mencionados. Aunado a lo anterior, la mayoría de los modelos que se tienen disponibles en la literatura no contemplan factores como: dinámica de chequeo de mensajes, consciencia de seguridad, relación entre nodos o una topología para representar de mejor forma la red formada por la lista de contactos. Además, en SMS, existe un tiempo límite de entrega de un mensaje; si el dispositivo móvil no tiene señal durante un largo tiempo, el mensaje nunca llegará a su destino. De esta manera, existen otros factores que pueden influir en la propagación de malware y que es importante tomar en cuenta. Por último, una mejor representación de la red social conllevaría a una mejor representación de la dinámica de propagación, este problema se podría resolver utilizando una SFN.

En el siguiente capítulo se introduce un modelo nuevo basado en el paradigma de Autómata Network para la propagación de gusano mediante el envío de mensajes SMS. La dinámica del modelo utiliza epidemiología compartimental y se ajusta para representar la interacción entre nodos. El modelo toma en cuenta aspectos importantes para la propagación en teléfonos inteligentes, que a la fecha no se han considerado por otros modelos de existentes.

Capítulo 4

Un nuevo modelo para la propagación de malware a través de SMS

Una vez revisado los conceptos necesarios respecto a la epidemiología, así como trabajos relacionados a la modelación de la propagación de gusanos, es posible plantear un nuevo modelo para la propagación de malware a través de mensajes SMS, agregando nuevas características que en trabajos anteriores no se consideraron o no se plantearon en conjunto, como: diferencias individuales (Autómatas Celulares), topología de red (Red Libre de Escala), chequeo de mensajes, relación de nodos y seguridad en los dispositivos.

En las siguientes secciones se propone un modelo probabilista, basado en la teoría de Autómata Network, con las características antes mencionadas.El objetivo es simular la propagación de malware tipo gusano, a través del envío de mensajes SMS en teléfonos celulares.

4.1. Consideraciones generales

Antes de describir de manera detallada el modelo propuesto en este trabajo, primero se listan algunas consideraciones importantes para su definición que se detallarán más adelante.

- Se considera como vector de infección el envío de mensajes SMS, con un enlace de descarga malicioso, no se contempla el envío de gusanos mediante WBXML.
- Se considera que el gusano que infecta dispositivos siempre tendrá acceso a la lista de contactos.
- Se considera una población con diferentes características de seguridad y de Sistema Operativo (SO), es decir, una población heterogénea.
- El tamaño de la población se mantiene constante en el tiempo.

- El tamaño de cada compartimento o tipo de población, de acuerdo a su estado ante la enfermedad en cada instante de tiempo, se considera una cantidad entera, no fraccionaria.
- Cada segundo se aplicará la función de transición al modelo.
- Se considera un tiempo de latencia máximo de TL segundos para que un dispositivo reciba un mensaje, es decir, que pase de un estado Susceptible a En espera.
- Se considera el grado de infección, la probabilidad de fallo o error en el envío y recibo, así como el tiempo de latencia de entrega, de un mensaje SMS, para que un nodo pueda pasar de un estado susceptible a expuesto, es decir, el nodo haya recibido de manera exitosa un mensaje con una liga maliciosa de un nodo vecino infectado.
- Se considera la actividad de revisar mensajes, así como el grado de confianza entre dispositivos, para que un nodo en estado expuesto pueda pasar a vulnerable, es decir, abra y lea un mensaje malicioso.
- Para un nodo en estado vulnerable, se considera el grado de consciencia de riesgo y el tipo de sistema operativo, para que pueda pasar a un estado infectado, es decir, confíe en el vecino que envió el mensaje, confíe en el enlace malicioso y de clic al mismo.
- Se considera que hay solo un gusano montado en el servidor, por lo que solo afecta a dispositivos con el mismo Sistema Operativo que el gusano.
- Únicamente los teléfonos que estén en un estado vulnerable, podrán ser contagiados por un teléfono inteligente en estado infectado, para ello el dispositivo sano deberá dar clic a la liga con el enlace malicioso.
- Un dispositivo infectado pasará a un estado de inactividad si el dispositivo deja de enviar mensajes SMS a su lista de contactos, por ejemplo: el dispositivo perdió la conexión con la red móvil. Esta transición sucede si se cumple una probabilidad P_{In}.
- Un dispositivo inactivo pasará a un estado de infectado si el dispositivo vuelve a enviar mensajes SMS a su lista de contactos. Esta transición sucede si se cumple una probabilidad P_I.
- Un teléfono inteligente expuesto revisará a lo más TR segundos los nuevos mensajes que reciba, en caso de que sobrepase el tiempo límite el usuario revisará el mensaje.
- Aquellos dispositivos que rebasen el umbral de riesgo UR, es decir, que su consciencia de riesgo CR sea mayor que el umbral, quedarán inmunes ante cualquier mensaje malicioso, debido a que el usuario nunca los abrirá.

- Una vez que un teléfono sea recuperado, es decir, estuvo en un estado infectado o inactivo y se aplicó una vacuna debido a un antivirus o el usuario del dispositivo realizó alguna acción que reestableció el dispositivo, el dispositivo estará protegido contra el gusano por lo que no se podrá volver a infectar.

- Los dispositivos en estados: Recuperado, Inmune y Latente, quedarán en el mismo estado respectivamente al siguiente estado de tiempo, ya que son estados terminales.

- El número inicial de dispositivos infectados es solo uno.

4.2. Formulación del Modelo

El modelo consiste de un Autómata Network, denotado como $N = (G, Q, F)$, donde cada nodo representa a un agente que emula a un usuario y su dispositivo (teléfono inteligente) conectado en la red celular, denotada como G, cada atributo del teléfono inteligente varía en el tiempo. Cada nodo pertenecerá a un único estado del conjunto estados, denotado como Q, en un instante de tiempo, t. El siguiente estado al cual pertenecerá el dispositivo será definido por la función de transición, denotada como F, la cual evalúa los parámetros y estado actual de cada nodo para obtener el siguiente estado. Para cuestiones prácticas se referirá en las siguientes secciones como teléfono inteligente o dispositivo, tanto al usuario como a su dispositivo.

4.2.1. Características de la red celular

Sea el grafo $G = \{V, E\}$ de la red celular de N, donde V representa los nodos conectados en la red y E la relación entre estos. Cada arista en E representa una relación de agregación, de acuerdo al directorio telefónico de cada dispositivo, es decir, existirá una arista $e_{i,j}$ (que va dirigida del nodo v_i al nodo v_j) si el nodo v_i tiene añadido en su lista de contactos al nodo v_j; definiendo el espacio celular como un grafo dirigido, como se muestra en la Figura 4.1. El grado saliente de cada nodo en G, es decir el número de aristas que salen de cada uno, estará definido por la distribución Ley de la Potencia, convirtiendo el espacio celular en una SFN.

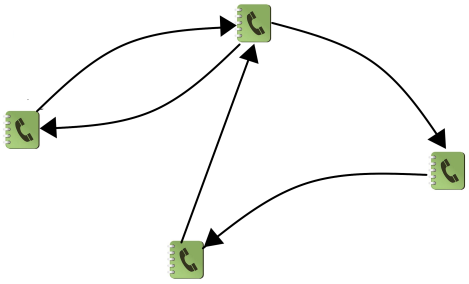


Figura 4.1: Grafo dirigido a través del directorio telefónico.

La vecindad de cada nodo está dada por las aristas conectadas a este, de la siguiente forma: cualquier nodo v_j que tenga agregado en sus lista de contactos a v_i será definido como vecino entrante de v_i; para el nodo v_i, cualquier nodo v_j que tenga agregado en su lista de contactos se define como vecino saliente del nodo v_i.

Sea M una matriz de dimensiones $|V| \times |V|$, se define a M como la matriz de adyacencia entre los dispositivos en el espacio celular G, donde: $|V|$ es el número de nodos en G del Automata Network; por lo que, M representa todas las relaciones del directorio telefónico o lista de contactos de cada dispositivo en G. Cada elemento de M estará definido de la siguiente manera:

$$M[i,j] = \begin{cases} 1, & \text{si } v_i \text{ tiene como vecino saliente a } v_j \\ 0, & \text{en otros casos.} \end{cases}$$

Se muestra un ejemplo en la Figura 4.2.

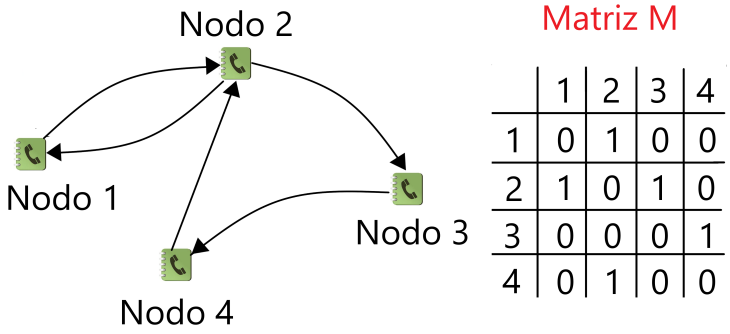


Figura 4.2: Matriz de adyacencia de un grafo dirigido.

4.2.2. Función de transición: Estados epidemiológicos

Definido el grafo G, es necesario definir también el conjunto de estados Q, basados en epidemiología compartimental. Los estados epidemiológicos dividirán a la población (teléfonos inteligentes) en un instante de tiempo t, de acuerdo a su condición frente a la enfermedad, en este caso, frente al gusano. Cada nodo pertenecerá al tiempo t a un único estado, el siguiente estado será determinado por la función de transición.

Entonces, se define Q como el siguiente conjunto de estados:

- **Susceptible(S)**: Nodos que no han sido infectados y están susceptibles a recibir un mensaje con un enlace malicioso.
- **En espera(En)**: Nodos a los cuales se les ha enviado un mensaje SMS y están a la espera de recibirlo, esto es debido al tiempo de latencia de entrega de un mensaje.

- **Expuesto(E)**: Nodos susceptibles que recibieron algún mensaje SMS con un enlace malicioso, pero no ha sido abierto o leído dicho mensaje. Los dispositivos expuestos no son capaces de iniciar un proceso de contagio. En caso de recibir más de dos mensajes al mismo tiempo, será considerado únicamente el mensaje del usuario origen que tenga mayor grado de confianza.
- **Vulnerable(V)**: Nodos expuestos que abrieron algún mensaje malicioso pero aún no ha sido abierta la liga maliciosa. Los dispositivos vulnerables no son capaces de iniciar contagios.
- **Latente(L)**: Nodos vulnerables que dieron clic a la liga maliciosa y descargaron el archivo, es decir, una copia del gusano, pero que por motivos de compatibilidad con el Sistema Operativo (SO), este no puede ser ejecutado y por lo tanto no puede iniciar un proceso de infección. Este es un estado terminal.
- **Infectado(I)**: Nodos vulnerables que dieron clic a la liga maliciosa, descargaron y ejecutaron el archivo, por lo que, el gusano esta en operación y envía mensajes de texto SMS con enlaces maliciosos a todos sus contactos en su directorio telefónico. Un nodo infectado solo puede contagiar a nodos en estado susceptible.
- **Recuperado(R)**: Nodos a los cuales se les detectó y se les removió el gusano del dispositivo de forma permanente, mediante una vacuna gracias al antivirus que porta. Este es un estado terminal.
- **Inmune(Im)**: Nodos vulnerables que se vuelven inmunes debido a la Consciencia de Seguridad de usuario. Este es un estado terminal.
- **Inactivo(In)**: Nodos infectados los cuales dejaron de enviar mensajes maliciosos a sus contactos debido a que ocurrió algún evento el cual no le permite al dispositivo seguir enviando mensajes SMS, como por ejemplo que se le acabó la batería y se apaga; sin embargo, puede volver a seguir enviando mensajes si el evento termina.

4.2.3. Función de transición: Reglas de transición

El diagrama de transición de estados de N es mostrado en la Figura 4.3, el cual muestra únicamente los cambios de estados entre cada uno de ellos. Las reglas que definen cada cambio en el diagrama se describen a continuación.

Transición: Susceptible (S) a En espera (En)

La transición del estado **S** al estado **En** ocurre cuando un dispositivo "sano" es seleccionado por un vecino entrante infectado, el cual le enviará un mensaje SMS, es decir, el dispositivo infectado tiene en su lista de contactos al dispositivo sano y lo selecciona para enviar un mensaje SMS con un enlace malicioso.

Se asume que todo dispositivo infectado siempre está posibilitado a enviar mensajes SMS y que todo dispositivo susceptible siempre está posibilitado a recibir mensajes. Entonces, sea P_{cont} la probabilidad de que un dispositivo i en estado **S** pase al estado **En** en el siguiente instante de tiempo, es denotada de la siguiente manera:

$$P_{cont} = \beta \frac{I_i(t)}{N_i}(1 - P_F) \tag{4.1}$$

donde:

- β es el grado de infección;
- $I_i(t)$ es el número de vecinos entrantes infectados al tiempo t;
- N_i es el número total de vecinos entrantes de i, y;
- P_f es la probabilidad de que un mensaje no llegue a su destino, debido a un error en la red celular.

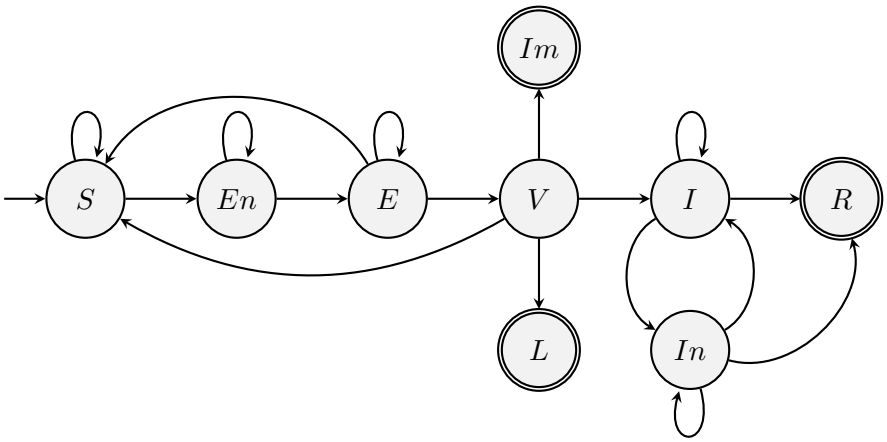


Figura 4.3: Diagrama de transición de los Estados del Autómata Network.

Por lo anterior, la transición del estado susceptible de un teléfono inteligente i al estado expuesto, se define como:

if $rand() <= P_{cont}$ **then**
  $\omega_i(t+1) = En$
else
  $\omega_i(t+1) = S$
end if

donde: $rand()$ es un número aleatorio real, tal que $0 \leq rand() \leq 1$.

Transición: En espera a Expuesto

La transición del estado En espera al estado Expuesto representa cuando el dispositivo "sano" está en espera de recibir el mensaje que el envío algún contacto infectado, esto es debido al tiempo de latencia que presenta la red celular.

Entonces, la transición del estado en espera de un teléfono inteligente i al estado expuesto, se define como:

if $(t - TEM_i) == TL$ **then**
  $\omega_i(t+1) = E$
else
  $\omega_i(t+1) = En$
end if

donde:

- TEM_i es el tiempo en el cual el dispositivo vecino infectado de i le envío un mensaje malicioso y,
- TL es el tiempo de latencia que tarda en llegar un mensaje SMS a su destino.

Transición: Expuesto a Vulnerable o Expuesto a Susceptible

Estas transiciones representan cuando un dispositivo que recibió un mensaje tiene la posibilidad de leerlo o no. El primer evento sucede cuando el usuario del dispositivo i, decide abrir y leer el mensaje, porque confía en el dispositivo vecino j, que se lo envío (Expuesto a Vulnerable); o que el usuario decide no abrir nunca el mensaje (Expuesto a Susceptible). Para ello, sea $F^t_{L_i}$ la función lógica usada para determinar si el mensaje que un smartphone i recibe desde un nodo j, se lee o no, tal que $F^t_{L_i}$ puede tomar los valores de 1 o 0, respectivamente. Sean TL^t_i, TM^t_i, y $GCP^t_{i,j}$ variables lógicas que indican que un mensaje podría leerse antes de transcurrir el periodo de tiempo TR desde que se recibió el mensaje, se ha alcanzado el tiempo máximo para que un mensaje se lea, y se satisface el grado de confianza de lectura de un mensaje, respectivamente. Entonces, se define la función lógica $F^t_{L_i}$ para determinar la transición de un estado Expuesto a Vulnerable como:

$$F^t_L(i) = (TL^t_i \vee TM^t_i) \wedge GCP^t_{i,j} \tag{4.2}$$

donde:

$$TL^t_i = \begin{cases} 1, & \text{con probabilidad } P_1 = \dfrac{1}{TR} \\ 0, & \text{con probabilidad } (1 - P_1) \end{cases} \tag{4.3}$$

$$TM^t_i = \begin{cases} 1, & \text{si } (t - TLM_i == TR) \\ 0, & \text{en otro caso} \end{cases} \tag{4.4}$$

$$GCP^t_{i,j} = \begin{cases} 1, & \text{con probabilidad } GC(i,j) \\ 0, & \text{con probabilidad } (1 - GC(i,j)) \end{cases} \tag{4.5}$$

aquí:

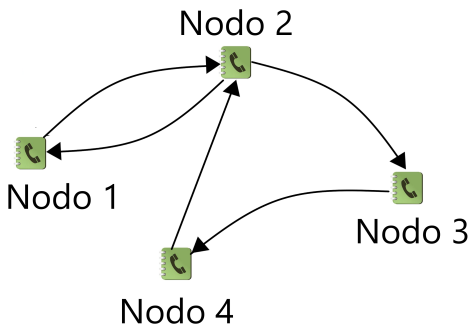
- TLM_i es el tiempo en el cual el dispositivo i recibió un mensaje malicioso.
- TR es el periodo que los usuarios tardan en revisar nuevos mensajes en su dispositivo. Revisar los mensajes que se reciben en un dispositivo móvil es un comportamiento complejo, los usuarios pueden revisar a cualquier hora del día el dispositivo. En [66] han estudiado este comportamiento y observaron que, a diferentes horas del día, la actividad de revisar mensajes puede variar respecto a la frecuencia y constancia con que los usuarios revisan. Por lo que, TR podrá incrementar o decrementar de acuerdo a las horas del día, aunque en este trabajo se toma con valor constante.
- $$GC(i,j) = \frac{NME(i,j)}{MaxNME} \tag{4.6}$$

  donde $NME(i,j)$ denota el número de mensajes intercambiados entre el dispositivo i y j y $MaxNME$ el número máximo de mensajes intercambiados entre dos dispositivos cualesquiera en la red celular; tal que $0 \leq GC \leq 1$, es una función que representa al grado de confianza o nivel de confianza que los usuarios de los dispositivos i y j se tienen (como se obtiene en [67]). Así, el usuario i abrirá un mensaje SMS sólo si confía en el dispositivo j y viceversa; representando un mayor grado de confianza cuando el valor sea cercano a uno y un menor grado de confianza cuando el valor sea cercano a 0.

  El valor de GC se obtiene así, a través de un registro de mensajes RM, el cual contiene el número de mensajes que un dispositivo i recibió de un dispositivo j en toda la red celular, en el periodo de una semana. Se muestra un ejemplo de RM en la Figura 4.4. Entonces, se suma la cantidad de mensajes intercambiados entre un dispositivo i y j, para cada dispositivo en la red celular, como se muestra en la Figura 4.5, y se obtiene el máximo número de mensajes intercambiados entre dos dispositivos. Para ejemplificarlo, véase el ejemplo de la Figura 4.5, donde el máximo número de mensajes es 7. Entonces, la suma de los mensajes enviados entre un dispositivo i y j se divide entre el máximo número de mensajes enviados en toda la red para obtener GC. Así, mediante el uso de GC se genera un grafo con peso en las aristas que indica el grado de confianza entre dos nodos, siempre que exista una arista entre ellos, y el peso es mayor que 0 si en algún momento los nodos han intercambiado al menos con un mensaje SMS. En la Figura 4.6, se muestra el grafo con pesos GC para la red mostrada en la Figura 4.4.

Como resultado, las transiciones del estado Expuesto a Vulnerable o Susceptible, se define de la siguiente forma:

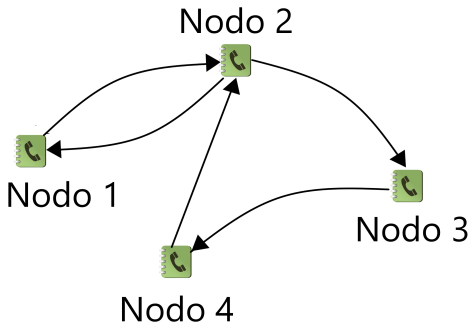
$$\omega_i(t+1) = \begin{cases} V, & \omega_i(t) = E, F_L^t(i) == 1 \\ E, & \omega_i(t) = E, F_L^t(i) == 0 \\ S, & \omega_i(t) = E, F_{NL}^t(i) == 1 \end{cases} \tag{4.7}$$



Matriz RM

	1	2	3	4
1	0	3	0	0
2	4	0	1	0
3	0	0	0	2
4	0	1	0	0

Figura 4.4: Ejemplo de Registro de Mensajes RM.



i - j	# msj
1 - 2	7
2 - 3	1
2 - 4	1
3 - 4	2

Figura 4.5: Cantidad de mensajes intercambiados entre los dispositivos de una red celular.

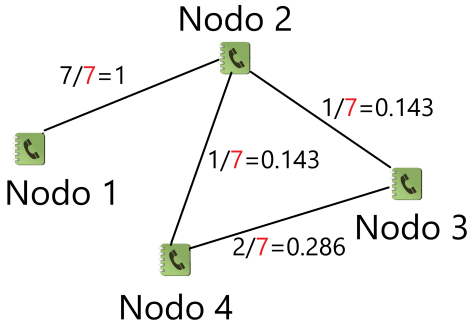


Figura 4.6: Ejemplo de un grafo con GC como peso en las aristas.

donde $F^t_{NL}(i) == 1$ es la función lógica para que un mensaje no se lea en el periodo de tiempo TR al no satisfacerse el grado de confianza, en cuyo paso el dispositivo regresa al estado susceptible y se define como sigue:

$$F^t_{NL}(i) = (TL^t_i \vee TM^t_i) \wedge \neg GCP^t_{i,j} \tag{4.8}$$

Transición: Vulnerable a Infectado o Vulnerable a Susceptible o Vulnerable a Latente o Vulnerable a Inmune

Estas transiciones representan cuando un dispositivo en estado vulnerable tiene la posibilidad de dar clic a la liga maliciosa de un mensaje SMS. Existen cuatro posibles eventos para esta transición: que el dispositivo vulnerable confíe en el enlace, dé clic y se descargue una copia del gusano, afectando al SO, quedando infectado (Vulnerable a Infectado); que el dispositivo confíe en el enlace y se descargue una copia del gusano, pero por motivos de compatibilidad con el SO no sea afectado(Vulnerable a Latente); que el usuario no confíe en el enlace y no dé clic a la liga (Vulnerable a Susceptible); y que el dispositivo no confíe en ningún enlace sospechoso debido a su alto grado de consciencia de riesgo (Vulnerable a Inmune).

La consciencia de riesgo o de seguridad se refiere al grado en que los usuarios de los dispositivos están conscientes de la existencia de malware, si un usuario es altamente consciente no intentará realizar alguna acción que ponga en riesgo el dispositivo, en este caso, abrir un enlace sospechoso. De modo que, esta transición puede ser representada como:

if $UR < CR_i$ **then**
  $\omega_i(t+1) = Im$
else
  if $rand() <= (1 - CR_i)$ **then**
    if $SO_i = SO_w$ **then**
      $\omega_i(t+1) = I$
    else
      $\omega_i(t+1) = L$
    end if
  else
    $\omega_i(t+1) = S$
  end if
end if

donde:

- $CR_i \in [0, 1]$ es el grado de consciencia de riesgo del usuario i:
- $UR \in [0, 1]$ es el umbral de riesgo. Si un usuario i con consciencia de riesgo CR_i este es mayor que el umbral, significa que el usuario i es altamente consciente por lo que no abrirá las ligas de mensajes maliciosos que reciba;
- SO_w es el tipo de SO que afecta el gusano, y;
- SO_i es el tipo de SO del dispositivo vulnerable.

Transición: Infectado a Recuperado o Infectado a Inactivo

Esta transición representa dos posibles eventos para un dispositivo que está infectado. El primer evento sucede cuando al dispositivo se le remueve de forma permanente el gusano, gracias a una vacuna por parte del antivirus (Infectado a Recuperado); el segundo evento sucede cuando el dispositivo es incapaz de seguir enviando mensajes debido a algún evento, como el hecho que el dispositivo se apagó, el dispositivo perdió la conexión a la red móvil, entre otros eventos, por lo que el dispositivo no puede seguir infectando a otros dispositivos.(Infectado a Inactivo). Por lo que, esta transición se representa como se describe a continuación:

if $rand() <= P_R$ **then**
  $\omega_i(t+1) = R$
else
  if $rand() <= P_{In}$ **then**
    $\omega_i(t+1) = In$
  else
    $\omega_i(t+1) = I$
  end if
end if

donde:

- $P_R \in [0, 1]$ es la probabilidad de que el antivirus detecte y remueva al gusano en el dispositivo infectado;
- $P_{In} \in [0, 1]$ es la probabilidad de que el dispositivo pase a un estado de inactividad;

Transición: Inactivo a Infectado o Inactivo a Recuperado

Esta transición representa un caso similar al estado infectado, ya que el dispositivo puede volver a enviar mensajes maliciosos o recuperarse, pues el dispositivo sigue infectado por el malware. El primer evento sucede, al igual que el estado infectado, cuando al dispositivo se le remueve de forma permanente el gusano (Inactivo a Recuperado); el segundo evento sucede cuando el dispositivo se encuentra de nuevo en condiciones que le permiten seguir enviando mensajes, por lo que seguirá infectando más dispositivos (Infectado a Inactivo). Tal que, esta transición se representa como se describe a continuación:

if $rand() <= P_R$ **then**
  $\omega_i(t+1) = R$
else
  if $rand() <= P_I$ **then**
    $\omega_i(t+1) = I$
  else
    $\omega_i(t+1) = In$
  end if
end if

donde: $P_I \in [0, 1]$ es la probabilidad de que el dispositivo regrese al estado Infectado;

Una vez descrito los cambios de transición entre estados, lo anterior define de manera general como es que los nodos evolucionan en el tiempo t al siguiente estado de tiempo $t + 1$.

4.3. Características y parámetros considerados en el Modelo

4.3.1. Características de los nodos

Como se mencionó antes, cada nodo en la red es representado mediante el uso de un agente. Definido el autómata, es posible obtener los atributos que caracterizan a cada nodo, estos son descritos en el Cuadro 4.1.

Atributo	Variable	Descripción
Estado Actual	$\omega_i(t)$	Estado al que pertenece el nodo i en el tiempo t. Es necesario para poder aplicar la función de transición y evolucionar al estado siguiente.
Estado Siguiente	$\omega_i(t+1)$	Estado al que pertenecerá el nodo i de acuerdo con la función de transición definida.
Tipo de Sistema Operativo	SO_i	Sistema Operativo que está instalado en el dispositivo, se considera sean: Android o iOS. El funcionamiento del gusano dependerá completamente del tipo de Sistema Operativo.
Tiempo de Llegada de Mensaje	TLM_i	Tiempo al cual el dispositivo i recibió un mensaje malicioso.

Cuadro 4.1: Atributos de un nodo representado por un teléfono inteligente

4.3.2. Parámetros de entrada

La dinámica del modelo se verá afectada por distintos parámetros, los cuales definen el comportamiento del Autómata. Estos parámetros se dividen en dos tipos: parámetros globales y parámetros individuales.

Parámetros Globales

Los Parámetros Globales definen un comportamiento general que afectan por igual a todos los individuos que forman parte de la red celular del Autómata, estos no cambian

o varían durante el paso de tiempo de ejecución. Estos parámetros se muestran en el Cuadro 4.2.

Parámetro	Variable
Número de teléfonos inteligentes	$\|V\|$
Número inicial de teléfonos inteligentes infectados	$I(0)$
Índice de infección	β
Probabilidad de fallo de entrega de un mensaje SMS	P_F
Tiempo de latencia de entrega de un mensaje SMS	TL
Umbral de Riesgo	UR
Probabilidad de detección y remoción del gusano en un dispositivo infectado	P_R
Periodo de tiempo de lectura de mensajes	TR
Probabilidad de inactividad de un dispositivo infectado	P_{In}
Probabilidad de regresar a infectado de un dispositivo inactivo	P_I

Cuadro 4.2: Parámetros Globales del Modelo

Parámetros Individuales del Modelo

Los Parámetros Individuales definen un comportamiento individual, es decir, un comportamiento a cada individuo i en el espacio celular, pudiendo afectar de diferente manera a cada uno de ellos. Los Parámetros Individuales se muestran en el Cuadro 4.3.

Parámetro	Variable
Grado de confianza	$GC(i,j)$
Consciencia de Riesgo	CR_i

Cuadro 4.3: Parámetros Individuales del Modelo

4.4. Dinámica General del Modelo

La dinámica de propagación del gusano, se ejecuta de la siguiente manera:

1. Se asignan los dispositivos al espacio celular G y se inicializan sus atributos, así como parámetros individuales y globales de cada dispositivo en el espacio celular como: el directorio telefónico, el registro de mensajes enviados entre dispositivos, consciencia de seguridad de los dispositivos, entre otros.

2. Se ejecuta la función de transición del autómata, cada dispositivo debe recolectar la información de sus vecinos para poder aplicar la función de transición y obtener su siguiente estado después de evaluar la función.

3. Se actualizan las variables de cada dispositivo para su uso en el siguiente estado de tiempo.

Se ejecutan los pasos 2 y 3 en cada evolución del sistema hasta que se cumpla el criterio de paro o finalización de la simulación. Para nuestro modelo, el criterio de paro es definido cuando el sistema se encuentre en un estado de equilibrio, es decir, el número de dispositivos en cada compartimento no varíe.

En el capítulo 4 se mostrarán y detallarán los resultados de simulación obtenidos por diferentes casos de estudios debido a la variación de parámetros de entrada que afectan al sistema en general.

Capítulo 5

Análisis de Resultados

En este capítulo, se presentan los resultados de simulación obtenidos del modelo presentado en el capítulo previo. Los resultados analizan el comportamiento del modelo cuando diferentes valores de los parámetros se varían y se analiza el comportamiento de sistema ante los mismos. Lo que se busca es emular el desempeño de diferentes tipos de malware y los afectos que generan en la propagación de la infección a través de SMS.

5.1. Información de las simulaciones

Para los diferentes casos de estudio se considera una población $|V|$ de 10,000 nodos, la población se mantiene constante para todos los casos y con un número de nodos infectados inicial $I(0) = 1$. Para establecer el directorio telefónico de cada nodo, es decir la manera en que estarán conectados en la red, se sigue una Ley de potencia, del tal manera que algunos nodos están altamente conectados, es decir, poseen un gran número de enlaces a otros nodos, aunque el grado de conexión de casi todos los nodos es bastante bajo. De tal manera que se construye una SFN. Para ello, se utilizó el algoritmo de Barabasi Albert Graph descrito en el capítulo 1 y que está implementando en el módulo Networkx [69] en Python.

Los valores de los grados de confianza entre dos nodos, se obtienen de llenar el registro de mensajes RM, que indica cuántos mensajes se han enviado entre dos nodos en un periodo de una semana, considerando que el número de mensajes máximo que un nodo puede enviar en una semana es 30 y que este no variará en las semanas posteriores, de acuerdo con [70, 71]. Para la mayoría de los casos de estudios analizados, el grado de confianza se varía usando una distribución uniforme, excepto un caso de estudio que usa una distribución binomial.

En lo referente a la consciencia de seguridad cibernética para cada nodo, es decir, que tan consciente es el usuario del nodo de la existencia de malware, se utilizó una distribución normal para establecer su valor. En cada caso de estudio que se presenta, se específica que valores se consideraron. El umbral de riesgo se establece en un valor fijo de 0.70.

La probabilidad de que un dispositivo pase a un estado de inactividad se establece en 0.01, esto es considerando que la mayoría de los dispositivos permanecen mayormen-

te activos en el tiempo. En caso de que un dispositivo pase al estado de inactividad, regresará de nuevo al estado infectado si cumple una probabilidad de 0.95, pues el usuario siempre hará lo posible por que el celular se mantenga funcionando en su totalidad, como sucede en la realidad.

En referente al SO, se establecen de dos tipos: población homogénea, cuando la población es del mismo tipo de SO y, población heterogénea, cuando la población esta dividida en dos tipos de SO (Android y iOS). Para el resto de los parámetros del modelo, en cada caso de estudio se especifican los valores utilizados. La dinámica del sistema evoluciona en pasos de tiempo de un segundo.

Para todos lo casos de estudio presentados, los resultados son el promedio de 10 simulaciones, que corresponden a un comportamiento promedio del sistema bajo estudio. Además, se considera al número acumulado de infectados como la suma de los dispositivos en un estado infectado y aquellos en un estado inactivo; ya que ambos son dispositivos infectados por el gusano.

5.1.1. Variación del índice de infección

En la actualidad, la epidemiología ha estudiado distintas enfermedades infecciosas. Cada una puede presentar un índice de infección diferente, siendo algunas enfermedades más infecciosas que otras; es decir, el contagio es mayor en unas enfermedades. En el caso de propagación de malware, se puede representar este comportamiento a través del parámetro β, que denota el índice de infección. Aquí se analiza la variación del mismo.

Parámetro	Valor
Índice de infección (β)	0.25, 0.50, 0.75 y 1.0
Probabilidad de falla	0.05
Tiempo de latencia de entrega	3s
Tiempo de lectura	180s
Consciencia de riesgo	media = 0.56
Probabilidad de recuperación	0
SO	Población homogénea

Cuadro 5.1: Parámetros utilizados en la variación del índice de infección

En el Cuadro 5.1, se describen los valores de los parámetros utilizados para las simulaciones llevadas a cabo. Particularmente, se consideran cuatro valores para el índice de infección: 0.25, 0.50, 0.75 y 1. Los resultados obtenidos de las diferentes simulaciones tomando en cuenta los valores antes mencionados se muestran en la Figura 5.1. Donde se muestra la evolución del número de dispositivos infectados en el tiempo para diferentes valores de β. Como puede observarse de esta imagen, a medida que el valor del índice de infección es mayor la infección del malware se propaga más rápido, como es de esperarse ya que la probabilidad de contagio también lo es. Es importante notar que, aunque el gusano trate de infectar al 100 % de los dispositivos, debido a que se toman en cuenta distintas barreras o protecciones, como lo es el grado de confianza y la cons-

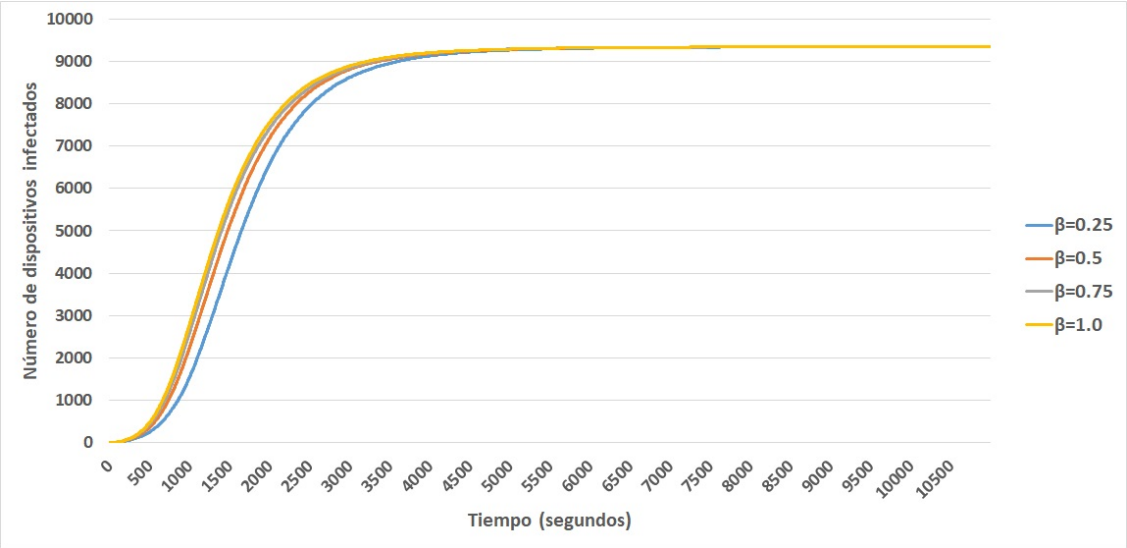


Figura 5.1: Comparación de número de dispositivos infectados en el tiempo debido a variar β.

ciencia de riesgo, esto no es posible. La manera en que los factores anteriores influyen en el comportamiento general del sistema se presentan en las siguientes subsecciones.

5.1.2. Variación del tiempo de latencia de entrega

La comunicación inalámbrica evoluciona año con año, lo que hace posible que cada vez sea más rápida la transmisión de información. No obstante lo anterior, los tiempos de retraso siempre estarán presentes al trabajar con redes inalámbricas debido al medio en el que se transmite. De acuerdo con [72], existen distintos tiempos de retraso, los cuales afectan la entrega de mensajes SMS. Para tomar en cuenta este comportamiento se introduce el parámetro TL que representa el tiempo de latencia de entrega de un mensaje SMS.

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	0.05
Tiempo de latencia de entrega	1, 2 y 3s
Tiempo de lectura	180s
Consciencia de riesgo	media = 0.56
Probabilidad de recuperación	0.0
SO	Población homogénea

Cuadro 5.2: Parámetros utilizados en la variación del tiempo de latencia.

En el Cuadro 5.2, se muestran los parámetros utilizados para las simulaciones llevadas a cabo, variando el tiempo de latencia de entrega en tres valores: 1, 2 y 3 s. Los

resultados obtenidos de las diferentes simulaciones tomando en cuenta los valores antes mencionados, se muestran en la Figura 5.2. Como puede observarse de esta figura, el tiempo de latencia de entrega de mensajes, como su nombre lo indica, genera un retardo en el tiempo requerido para infectar a la población; de tal manera que a mayor valor de TL, el tiempo requerido para infectar a la población de smartphones también se incrementa. Ello debido a que una latencia mayor un retraso en la entrega de los mensajes SMS y por lo tanto una propagación de la infección más lenta.

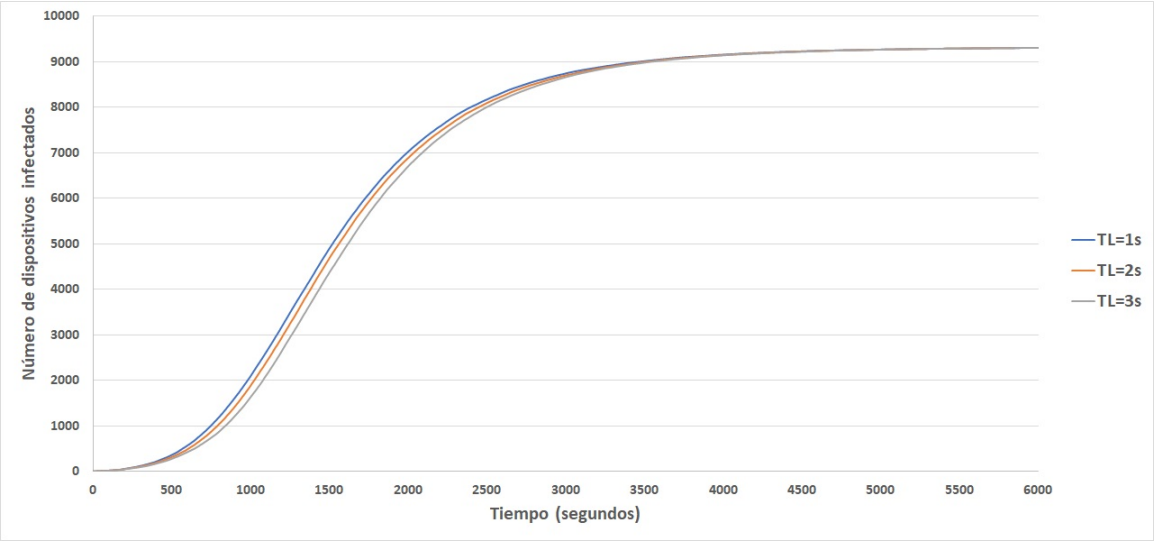


Figura 5.2: Comparación de número de dispositivos infectados en el tiempo para diferentes valores del TL

Así, si un smartphone infectado envía un mensaje SMS al tiempo t con $TL = 1$, el dispositivo destino recibirá el mensaje en el siguiente instante de tiempo, es decir, al tiempo $t + 1$. Sin embargo, para tiempos de latencia mayores, si se envía un mensaje SMS es probable que no se entregue al siguiente instante de tiempo, generando que el propagación de la infección por malware sea más lenta, conforme el valor de TL es mayor; ya que los mensajes tardan mucho más tiempo en llegar a sus destinos.

5.1.3. Variación de la falla de entrega de mensaje

Otra falla importante que sucede en las comunicaciones inalámbricas es la pérdida de información. Aunque las compañías telefónicas trabajan diariamente para evitar este evento, no pueden proporcionar un servicio 100 % confiable debido al medio con el que trabajan, por lo que es posible haya pérdida de mensajes durante la comunicación entre dos dispositivos. De acuerdo con [72], hay una pérdida mínima de mensajes, entre un rango de 5 % a 20 %. En el modelo propuesto, la pérdida de mensajes se representa a través del parámetro P_F que representa la probabilidad de que un mensaje no llegue a su destino.

En el Cuadro 5.3 se describen los parámetros utilizados para las simulaciones llevadas a cabo, variando la falla de entrega de mensaje en cuatro valores: 0.05, 0.10, 0.15

y 0.20. Los resultados obtenidos de las diferentes simulaciones tomando en cuenta los valores antes mencionados, se muestran en la Figura 5.3. Debido a que la variación entre las diferentes curvas de la Figura es mínima, se presenta la comparación de las curvas del tiempo 0 al 3000 en la Figura 5.4.

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	0.05. 0.10. 0.15. 0.20
Tiempo de latencia de entrega	3s
Tiempo de lectura	180s
Consciencia de riesgo	media = 0.56
Probabilidad de recuperación	0
SO	Población homogénea

Cuadro 5.3: Parámetros utilizados en la variación de falla de entrega del mensaje

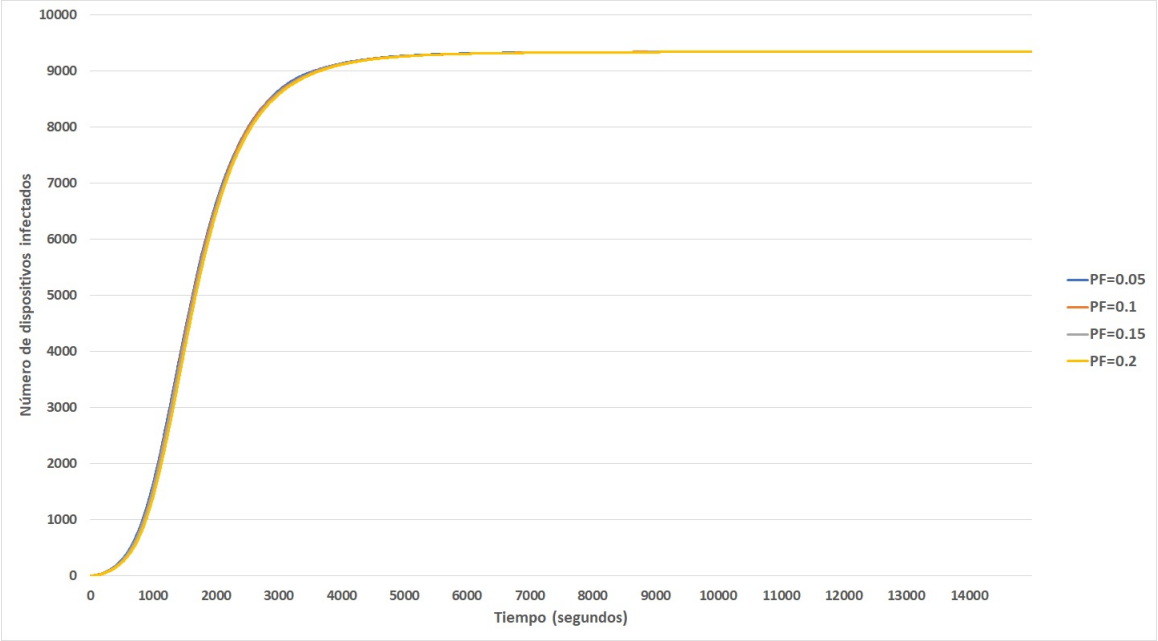


Figura 5.3: Comparación de número de dispositivos infectados para diferentes valores de la probabilidad de fallo de entrega de mensaje P_F.

Como puede notarse en la Figura 5.4, a medida que la probabilidad de falla de en la entrega del mensaje se incrementa, la velocidad de la propagación es menor y por lo tanto, tiempo requerido para infectar a la población se incrementa. Por lo tanto, contar con sistemas de red que funcionen adecuadamente, sin duda permite que las funcionalidades de los smartphones sean más eficientes, pero también compromete más al propagación de malware. Dado que no se considera un medio de recuperación de los dispositivos, todos los casos considerados infectan a la población de manera semejante, aunque con diferente velocidad de propagación del software dañino.

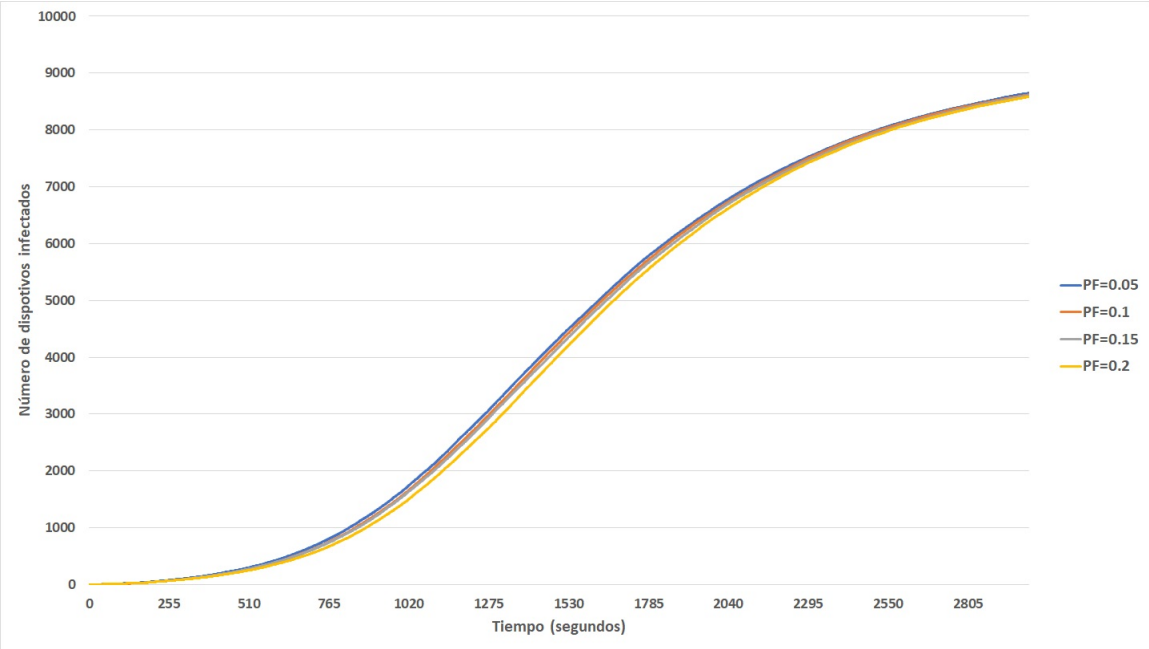


Figura 5.4: Comparación de número de dispositivos infectados debido a variar P_F del tiempo 0 al 3000.

5.1.4. Variación de la consciencia de riesgo

La consciencia de riesgo es un factor importante en la propagación del malware, es un medio importante para evitar que un malware se propague entre los dispositivos. De acuerdo con Statista [73], en promedio el 56 % de las personas están conscientes de que existen malwares y si son conscientes del riesgo que implica dar clic sobre una liga, a los escritores de malware les costará más trabajo que estas personas realicen alguna acción que pongan en riesgo su dispositivo. Con la finalidad de analizar los efectos de la conciencia de seguridad en al propagación del malware mediante SMS, en lo siguiente se analiza la variación de la misma.

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	0.05
Tiempo de latencia de entrega	3s
Tiempo de lectura	180s
Consciencia de riesgo	media =0.25, 0.50, 0.56, 0.75
Probabilidad de recuperación	0
SO	Población homogénea

Cuadro 5.4: Parámetros utilizados en la variación de la consciencia de seguridad

En el Cuadro 5.4 los parámetros utilizados para las simulaciones llevadas a cabo, variando la consciencia de riesgo de los usuarios, se consideran los siguientes valores: 0.25, 0.50, 0.56 y 0.75. Los resultados obtenidos de las diferentes simulaciones tomando en cuenta los valores antes mencionados, se muestran en la Figura 5.5.

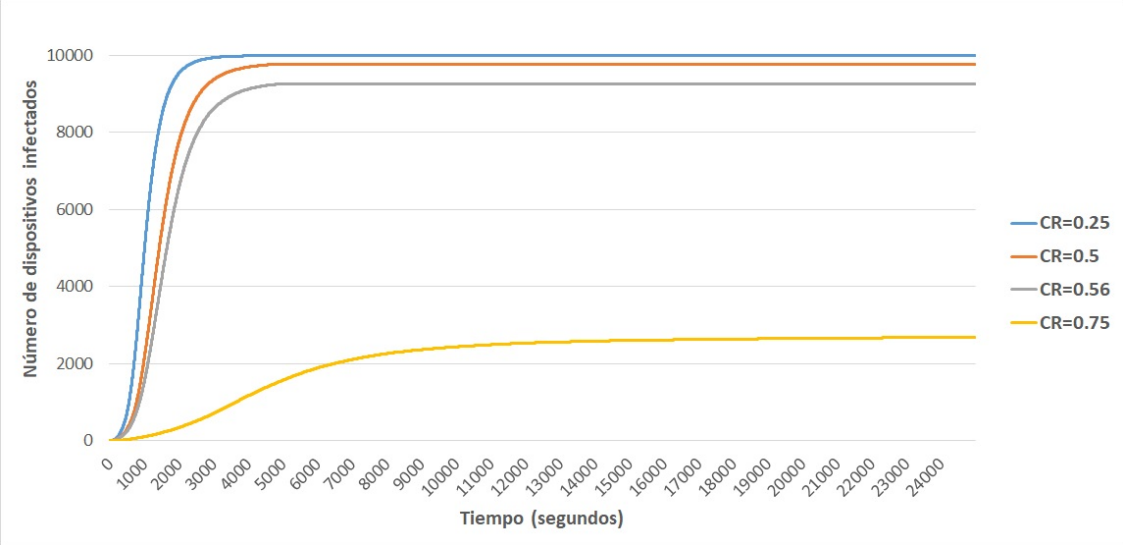


Figura 5.5: Comparación de número de dispositivos infectados en el tiempo considerando diferentes valores de la consciencia de riesgo CR.

Como puede notarse en la Figura 5.5, a medida que un mayor porcentaje de la población es consciente del riesgo que implica dar clic sobre una liga, es menor la población que puede infectarse por el malware, aún cuando no se cuente con un antivirus que evite el contagio. Además, que la propagación del malware es más lenta. Particularmente, cuando se considera el caso donde 25 % de la población es consciente del riesgo de accesar una liga recibida a través de un mensaje, la velocidad de propagación es mayor. Por el contrario, cuando se considera que 75 % de la población está consciente del riesgo que implica recibir un mensaje SMS y acceder sus ligas, no solamente se reduce la velocidad de programación, sino también la cantidad de dispositivos que se pueden infectar. Por lo que la consciencia de riesgo es importante para reducir la propagación de un malware y sus efectos negativos. Por otra parte, en el modelo se considera un umbral de riesgo UR, tal que si un usuario i con consciencia de riesgo CR_i mayor que el umbral, significa que el usuario i es altamente consciente por lo que no abrira las ligas de mensajes maliciosos que reciba (como si tuviera un antivirus). De tal manera que si la conciencia de riesgo de una persona alcanza o rebasa el UR, la persona no realizará ninguna acción que ponga en riesgo su dispositivo por lo que quedará inmune. Así, en la Figura 5.6, se muestra el número de dispositivos que se vuelven inmunes en el tiempo para diferentes valores de CR y $UR = 0.70$. Se puede observar de esta figura que a mayor consciencia de riesgo, el número de dispositivos inmunes crecerá, pues los usuarios de estos dispositivos no realizarán alguna acción que ponga en riesgo el dispositivo. Por lo tanto, el UR es un factor importante al variar la CR, mientras el valor de la media de CR sea menor y alejado del valor de UR el número de dispositivos infectados crecerá

notablemente y el número de dispositivos inmunes será mínimo; como ocurre cuando se tiene una media en CR de 0.25 y 0.50. Cuando el valor de CR sea cercano, igual o mayor que UR, el número de dispositivos inmunes crece y el número de dispositivos infectados será mínimo; como es el caso de una media en CR de 0.556 y 0.75.

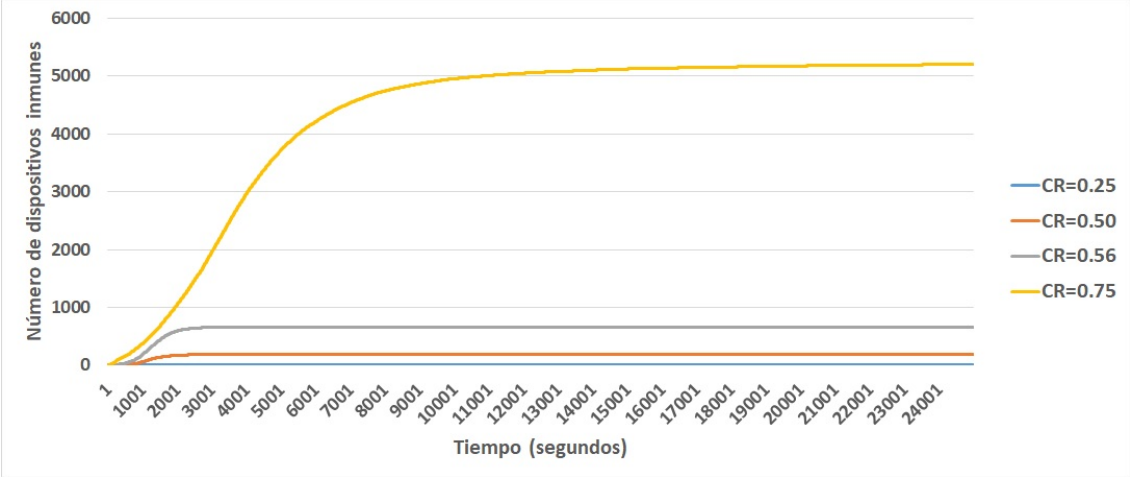


Figura 5.6: Comparación de número de dispositivos inmunes en el tiempo para diferentes valores de consciencia de riesgo CR.

5.1.5. Variación de tiempo de lectura

Las personas a lo largo del día se encuentran haciendo distintas actividades, como trabajando, estudiando, viajando, entre otras, las cuales ocurren de manera simultánea a las mismas horas del día. En algunas ocasiones puede ocurrir que las personas estén sin hacer alguna actividad o desocupados, por lo que el uso del dispositivo móvil varía de acuerdo a las horas de cada día. Por lo que la actividad de revisar el dispositivo es un factor que afecta directamente la dinámica de como se transmite el malware. Aunque en este trabajo no se establecen diferentes probabilidades de lectura de mensaje en función del día, si se realiza un análisis de los efectos que genera en la propagación del malware el intervalo con que regularmente una persona lee los mensajes recibidos. De acuerdo con Tatango [74], el 90 % de los usuarios leen un mensaje de texto recibido dentro de los primeros 3 minutos (180 segundos). En el modelo presentado se representa el tiempo de lectura máximo para que una persona lea un mensaje a través de la variable TR, la cual indica el periodo máximo en el cual un persona que va a leer un mensaje SMS lo hará.

En el Cuadro 5.5 se describen los parámetros utilizados para las simulaciones llevadas a cabo, variando la media del tiempo de lectura en cuatro valores: 90, 180, 270 y 360s. Los resultados obtenidos de las diferentes simulaciones tomando en cuenta los valores antes mencionados, se muestran en la Figura 5.7. Como se puede notar de esta figura, una frecuencia de lectura menor implica una mayor velocidad de propagación; ya que los usuarios revisan más constantemente su dispositivo verificando si hay nuevos mensajes y es riesgo de infección se incrementa.

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	0.05
Tiempo de latencia de entrega	3s
Tiempo de lectura	90, 180, 270 y 360s
Consciencia de riesgo	media =0.56
Probabilidad de recuperación	0
SO	Población homogénea

Cuadro 5.5: Parámetros utilizados en la variación de tiempo de lectura

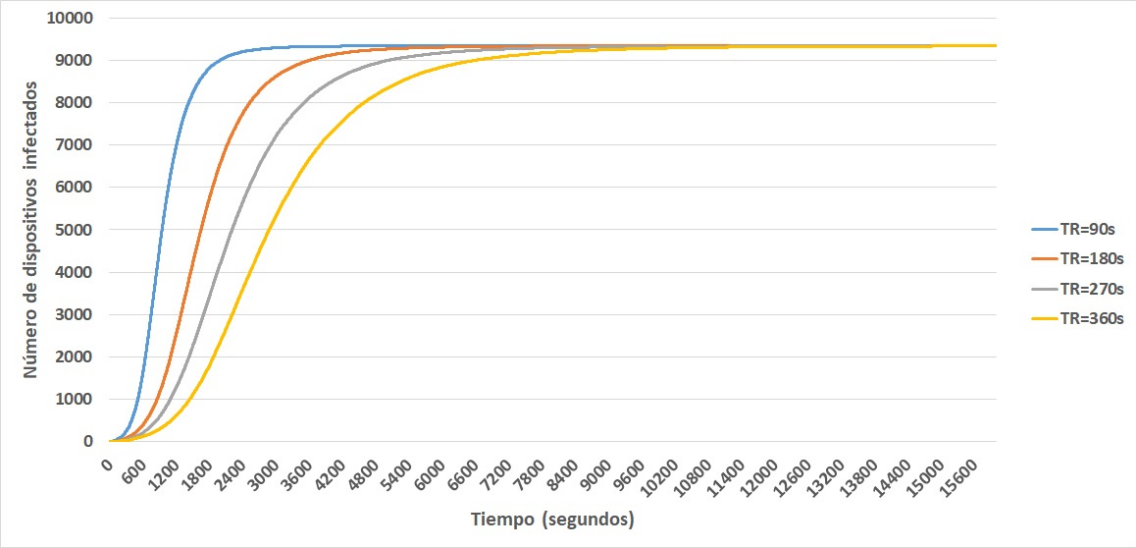


Figura 5.7: Comparación de número de dispositivos infectados en el tiempo para diferentes valores del tiempo de lectura TR.

5.1.6. Variación de probabilidad de recuperación

En la actualidad, existe una gran variedad de antivirus, cada uno de ellos trata de ofrecer la mayor protección al dispositivo en el que se encuentra. Existen distintos antivirus dependiendo de sus características, diferenciados desde el precio hasta el tipo de dispositivo en el que se instala. En el caso de los dispositivos móviles, y debido a que los usuarios no son conscientes sobre los riesgos que existen, la mayor parte de los usuarios no instala un antivirus o una protección ante el malware. A pesar de que los dispositivos tengan instalados algún antivirus, los dispositivos se exponen al nuevo malware que surge día con día, por lo que la protección que ofrecen los mismos puede no ser efectiva. Pero no solamente los antivirus pueden realizan la tarea de recuperar o limpiar un dispositivo infectado, los usuarios también pueden realizar alguna acción que ponga en un estado seguro su dispositivo (como un *Hard Reset*), provocando que quede limpio y sano del malware que lo infectó. Por lo anterior, en este trabajo se presentan resultados cuando se considera un factor de recuperación.

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	5 %
Tiempo de latencia de entrega	3s
Tiempo de lectura	180s
Consciencia de riesgo	media = 0.56
Probabilidad de recuperación	0.00025, 0.00050, 0.00075 y 0.0010
SO	Población homogénea

Cuadro 5.6: Parámetros utilizados en la variación de la probabilidad de recuperación.

En el Cuadro 5.6 se muestran los parámetros utilizados para las simulaciones llevadas a cabo. Cuatro valores de la probabilidad de recuperación de un dispositivo se consideran: 0.00025, 0.00050, 0.00075 y 0.0010. Los resultados obtenidos de las diferentes simulaciones tomando en cuenta los valores antes mencionados, se muestran en las Figuras 5.8 y 5.9 (curvas de infección y recuperación).

En la Figura 5.8 se puede observar que a menor probabilidad de recuperación, el número de dispositivos máximo que se pueden infectar es mayor y la velocidad de propagación es más alta. Debido a ello, el tiempo que se requiere para infectar a la población máxima posible, se incrementa en función de esta probabilidad de recuperación: un valor de P_R menor implica un tiempo mayor, ya que el número de dispositivos que pueden infectarse es más alto. Como consecuencia, el proceso de recuperación es más lento para P_R menor, como se puede observar de la Figura 5.9, que muestra el número de dispositivos recuperados en el tiempo para diferentes valores de P_R. Por ejemplo, si un dispositivo que tiene en su directorio telefónico a 100 contactos se infecta y recupera rápidamente, debido a que el factor P_R es alto, no alcanzará a infectar a todos los dispositivos existentes; por lo que existirán menor cantidad de dispositivos que se deben recuperar.

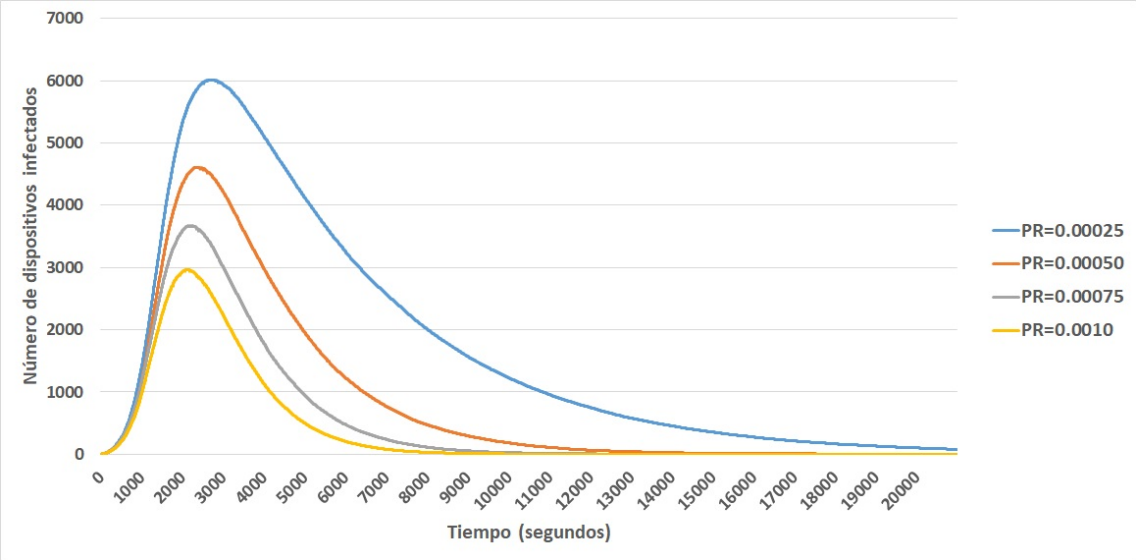


Figura 5.8: Comparación de número de dispositivos infectados en el tiempo para diferentes valores la probabilidad de recuperación P_R.

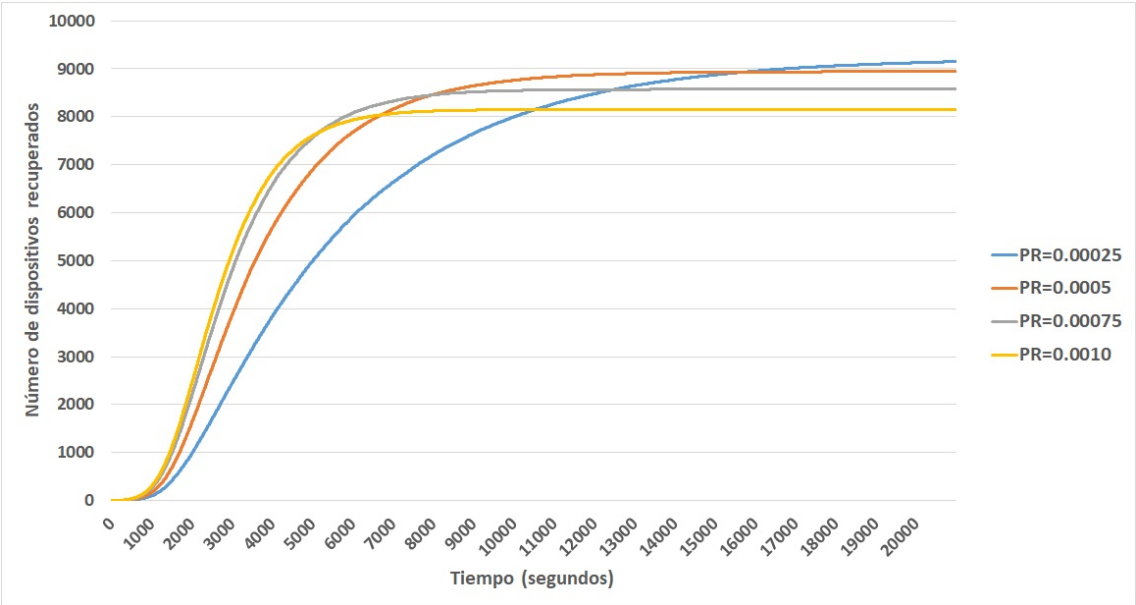


Figura 5.9: Comparación de número de dispositivos recuperados en el tiempo para diferentes valores de P_R.

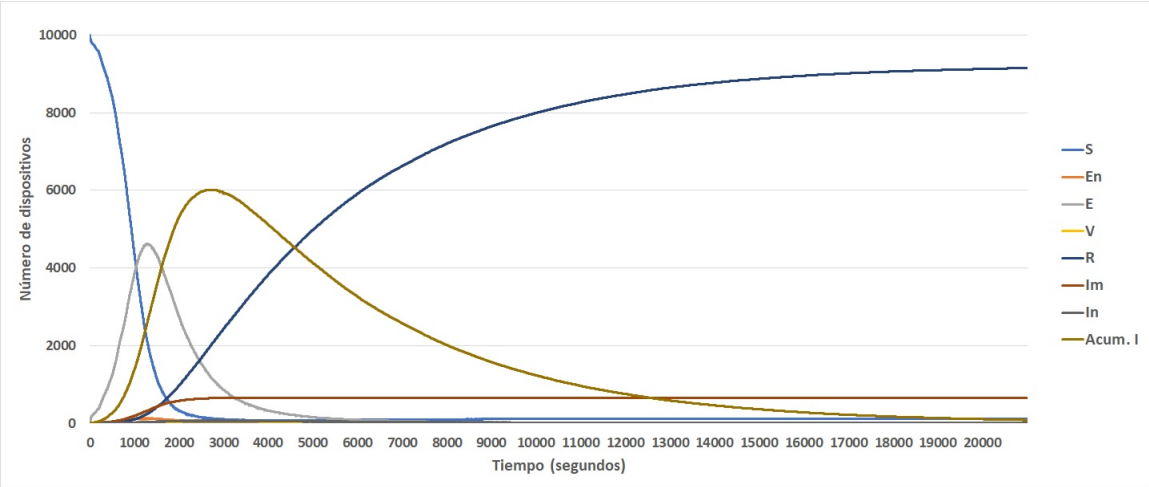


Figura 5.10: Número de dispositivos en cada compartimiento con respecto al tiempo para $P_R = 0.00025$.

Nótese que a diferencia del caso cuando la probabilidad $P_R = 0$, donde la curva de infección crece hasta propagarse a la mayor población posible y se mantiene en el tiempo, cuando algún medio de recuperación del malware se considera, el comportamiento de la curva de infección cambia (ver Figura 5.10 que muestra el número de dispositivos por compartimento en el tiempo para $P_R = 0.025$). Al inicio, la mayoría de los dispositivos son susceptibles, pero con el tiempo esta curva desciende debido a que la población se empieza a infectar hasta que un máximo de dispositivos infectados de alcanza, y entonces la curva desciende y la cantidad de dispositivos recuperados se incrementa (no se considera factor de renovación). Este desempeño observado en la Figura 5.10 corresponde con el observado en los modelos epidemiológicos.

5.1.7. Variación de grado de confianza

El Grado de Confianza (GC) o de relación indica que tanto confía un usuario de otro. De acuerdo a como se explica en el capítulo anterior, se determina mediante la cantidad de mensajes que se envían entre sí dos dispositivos en una semana. Por lo anterior, la cantidad de veces que un dispositivo se comunicó con otro mediante un mensaje SMS, determina el grado de confianza entre los dispositivos. En los casos de estudio de las subsecciones anteriores, el GC se obtuvo a través de inicializar de forma aleatoria el RM, donde el máximo número de mensajes que un dispositivo puede enviar a otro es de 30 en una semana. Sin embargo, para generar un sistema inicial más adecuado, en los resultados que se presentan en esta sección, una distribución binomial se usa para asignar el grado de confianza GC en la población en general.

En el Cuadro 5.7 se muestran los parámetros utilizados para las simulaciones llevadas a cabo. Los resultados se obtienen de considerar cuatro valores de la probabilidad del grado de confianza: 0.20, 0.40, 0.60 y 0.80. Los resultados obtenidos de las diferentes simulaciones tomando en cuenta los valores antes mencionados, se muestran en la Figura 5.11 que corresponde al número de dispositivos infecciosos en el tiempo, para diferentes

valores de GC.

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	5 %
Tiempo de latencia de entrega	3s
Tiempo de lectura	180s
Consciencia de riesgo	media = 0.56
Probabilidad de recuperación	0
SO	Población homogénea
Grado de confianza	P = 0.20, 0.40, 0.60 y 0.80

Cuadro 5.7: Parámetros utilizados en la variación de GC.

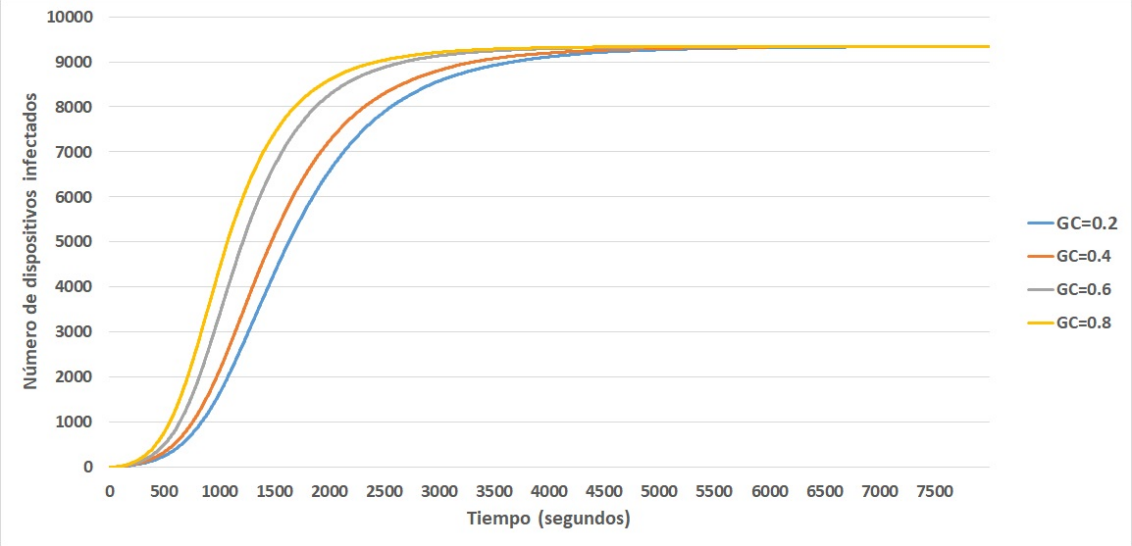


Figura 5.11: Comparación de número de dispositivos infectados en el tiempo para diferentes valores de GC.

Como se puede notar de la Figura 5.11, si los usuarios tienen mayor grado de confianza entre los mismos, darán clic más rápidamente a la liga enviada para infectarse, que cuando se tiene un menor grado de confianza y por lo tanto, la velocidad de propagación es mal alta. Este comportamiento se justifica porque los usuarios dependen del grado de confianza para decidir si leen o no algún mensaje recibido de un contacto: los usuarios leerán más rápido un mensaje si el grado de confianza que tiene con el vecino infectado que le envío el mensaje es mayor y el malware se propaga más rápidamente.

Aunque la red telefónica está formada mediante una SFN, es importante señalar que el GC es independiente a la topología de la red, pues GC implica cuantos mensajes SMS se enviaron de un dispositivo a otro en un periodo determinado.

5.1.8. Variación del Sistema Operativo

Por otra parte, los malware muchas veces sólo son capaces de infectar a aquellos dispositivos con un sistema operativo particular, por lo que un dispositivo con sistema operativo diferente recibe el mensaje SMS no será capaz de infectarse.

Actualmente el Sistema Operativo instalado en la mayoría de los teléfonos inteligentes es Android. De acuerdo con [75, 76] y como se muestra en la Figura 5.12, en promedio el 84.4 % de dispositivos comprados, entre el 2019 y 2020, cuenta con un SO de tipo Android, mientras que en promedio el 15.5 % cuenta con un SO de tipo iOS y menos del 0.1 % cuenta con un SO diferente a los anteriores. Con los datos anteriores, en esta subsección se realiza un análisis de la propagación del malware cuando se considera que se cuenta con dispositivos con diferentes SO. Es importante mencionar que el tipo de SO del gusano, se supone será el mismo tipo que el tipo de SO que tiene instalado el dispositivo inicial infectado.

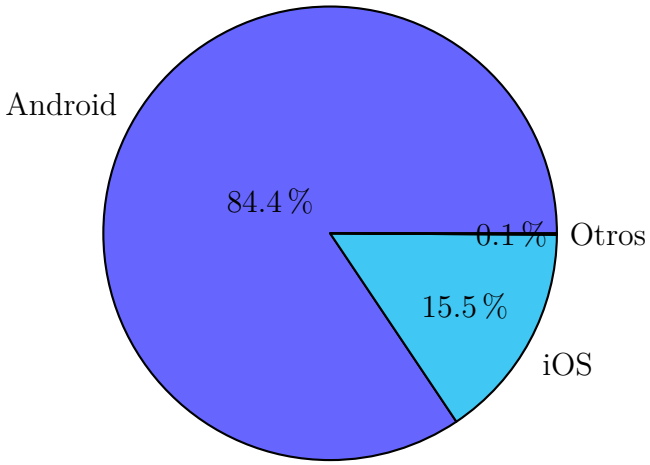


Figura 5.12: Porcentaje de dispositivos comprados con diferente SO vendidos entre 2019 y 2020.

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	0.05
Tiempo de latencia de entrega	3s
Tiempo de lectura	180s
Consciencia de riesgo	media = 0.56
Probabilidad de recuperación	0
SO	84.4 % Android, 15.5 % iOS

Cuadro 5.8: Parámetros utilizados en la variación del SO.

En el Cuadro 5.8, se muestran los valores de los parámetros que se utilizan para las simulaciones llevadas a cabo, variando el SO en Android y iOS. Los resultados obtenidos de las dos simulaciones tomando en cuenta los valores antes mencionados, se muestran en la Figuras 5.13 y 5.14.

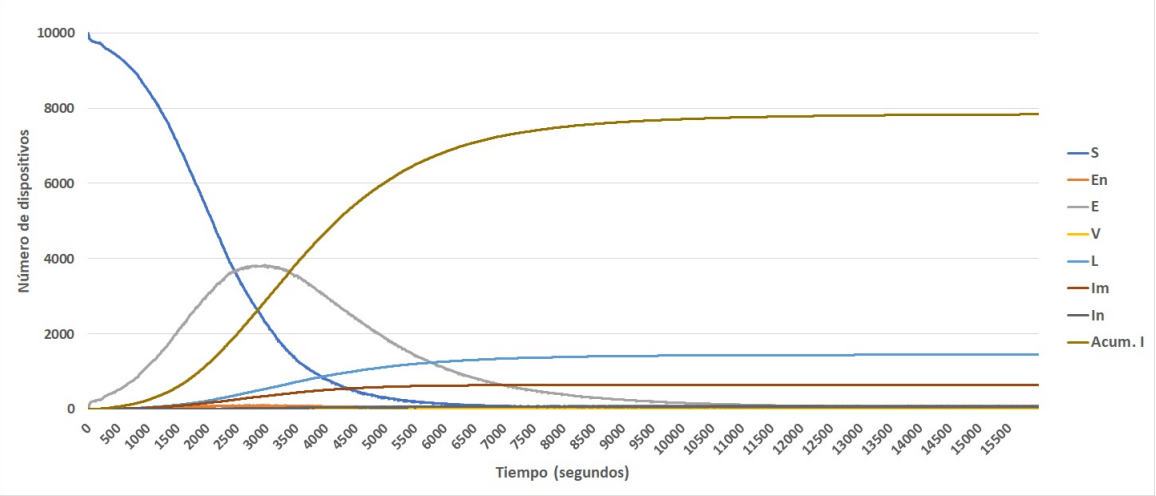


Figura 5.13: Curvas de cada compartimento con SO Android.

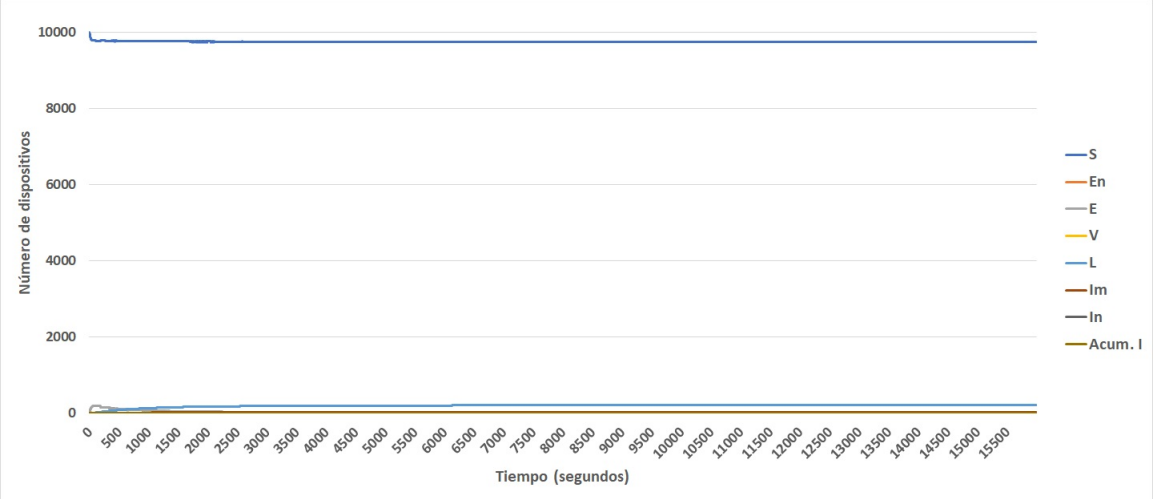


Figura 5.14: Curvas de cada compartimento con SO iOS.

Como se observa en la Figura 5.13, cuando se infectan dispositivos de tipo Android (84.5 % de la población total), el número de dispositivos infectados se aproxima a 8000 dispositivos, mientras el número de dispositivos latentes se mantiene abajo de los 2000 dispositivos. Recordar que existen otros factores que impiden que el 100 % de la población se infecte.

En caso contrario, como se muestra en la Figura 5.14, cuando se infectan dispositivos de tipo iOS (15.5 % de la población total), el número de dispositivos infectados es mínimo; pues la mayoría de los dispositivos son del otro tipo de SO. El número

de dispositivos latentes es superior respecto al número de infectados, lo cual reduce significativamente que los usuarios con este SO puedan contagiar a otros del mismo tipo.

5.2. Simulación gráfica

Para los casos de estudio anteriores, no es posible visualizar la red debido a que se cuenta con un gran número de dispositivos. Por lo anterior, utilizar un número de dispositivos menor es ideal para poder visualizar como se infectan de manera gráfica los dispositivos en la red telefónica. Para la siguiente simulación se utiliza una red telefónica con 100 dispositivos, los cuales son fáciles de mostrar en una red. Se utiliza el módulo de Plotly [77], el cual ayuda a visualizar información de manera gráfica, el módulo se encuentra implementado en Python. Los parámetros utilizados para esta simulación se muestran en el Cuadro 5.9.

Los resultados obtenidos de la simulación tomando en cuenta los valores antes mencionados, se muestran en la Figura 5.15.

Se puede observar el avance del proceso de infección en las Figuras 5.16 - 5.28. Es importante señalar que en la simulación gráfica se puede observar que identificador tiene cada dispositivo, así como a que estado pertenece, como se muestra en la Figura 5.29. El color de cada nodo varía de acuerdo a su estado, siendo la relación de colores y estados como se muestra a continuación:

- Verde - Susceptible
- Amarillo - En espera
- Anaranjado - Expuesto
- Azul - Vulnerable
- Café - Latente
- Rojo - Infectado
- Verde Claro - Recuperado
- Rosa - Inactivo
- Morado - Inmune

Parámetro	Valor
Índice de infección (β)	0.25
Probabilidad de falla	5 %
Tiempo de latencia de entrega	3s
Tiempo de lectura	180s
Consciencia de riesgo	media = 0.50
Probabilidad de recuperación	0
SO	Población homogénea

Cuadro 5.9: Parámetros utilizados para la simulación gráfica.

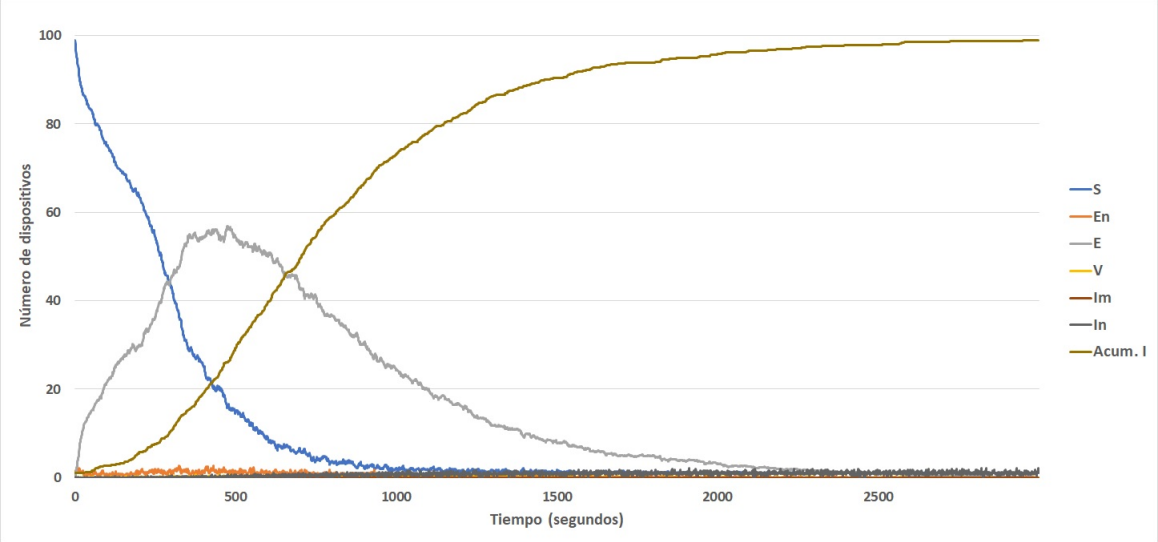


Figura 5.15: Número de dispositivos en cada compartimiento con $N = 100$.

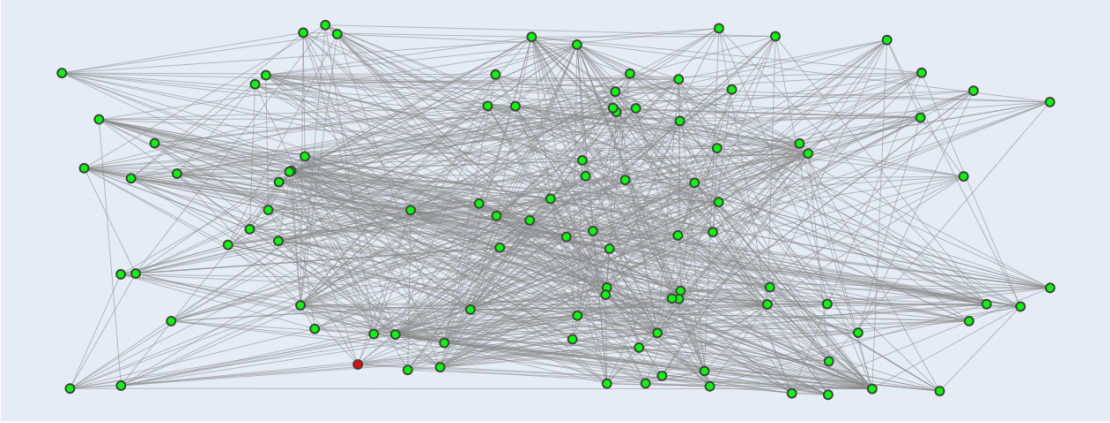


Figura 5.16: Simulación con $N = 100$ en el tiempo $t = 0$.

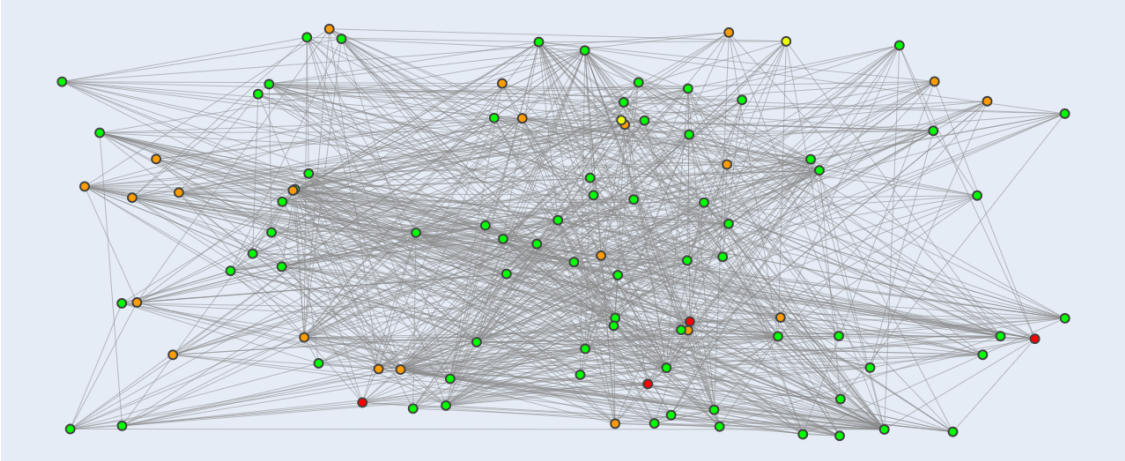


Figura 5.17: Simulación con $N = 100$ en el tiempo $t = 250$.

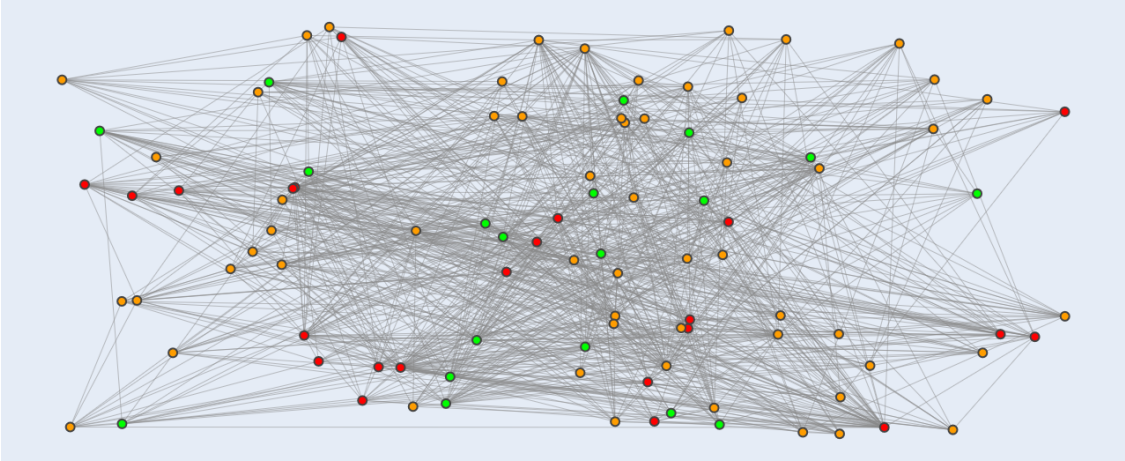


Figura 5.18: Simulación con $N = 100$ en el tiempo $t = 500$.

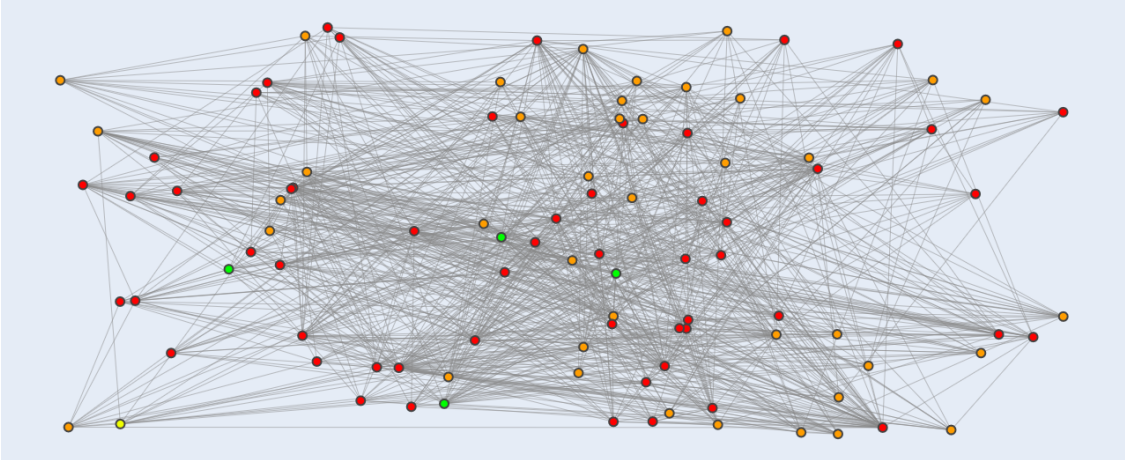


Figura 5.19: Simulación con $N = 100$ en el tiempo $t = 750$.

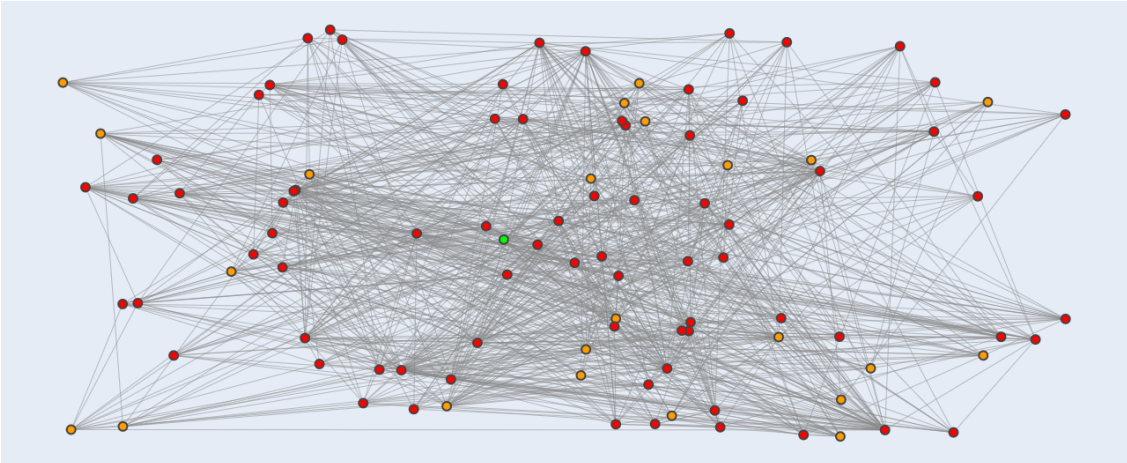


Figura 5.20: Simulación con $N = 100$ en el tiempo $t = 1000$.

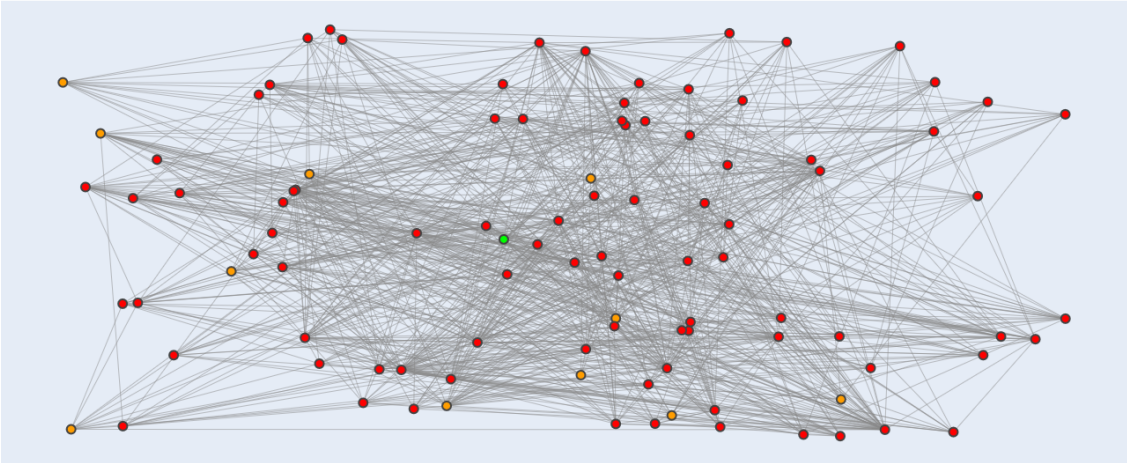


Figura 5.21: Simulación con $N = 100$ en el tiempo $t = 1250$.

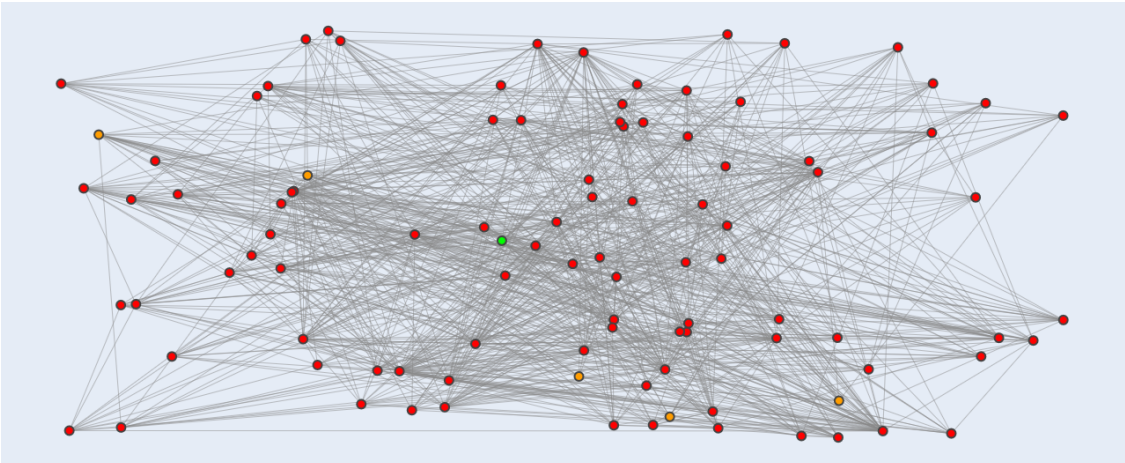


Figura 5.22: Simulación con $N = 100$ en el tiempo $t = 1500$.

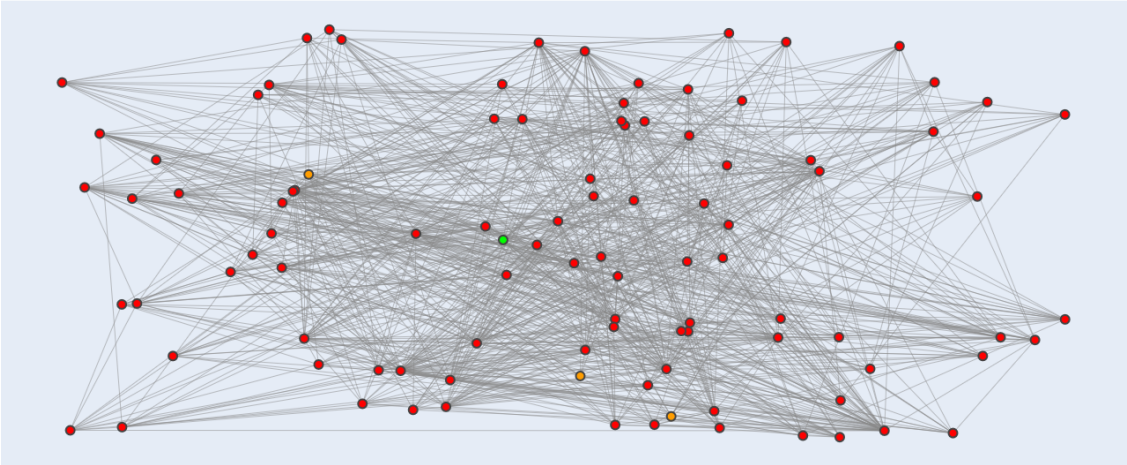


Figura 5.23: Simulación con $N = 100$ en el tiempo $t = 1750$.

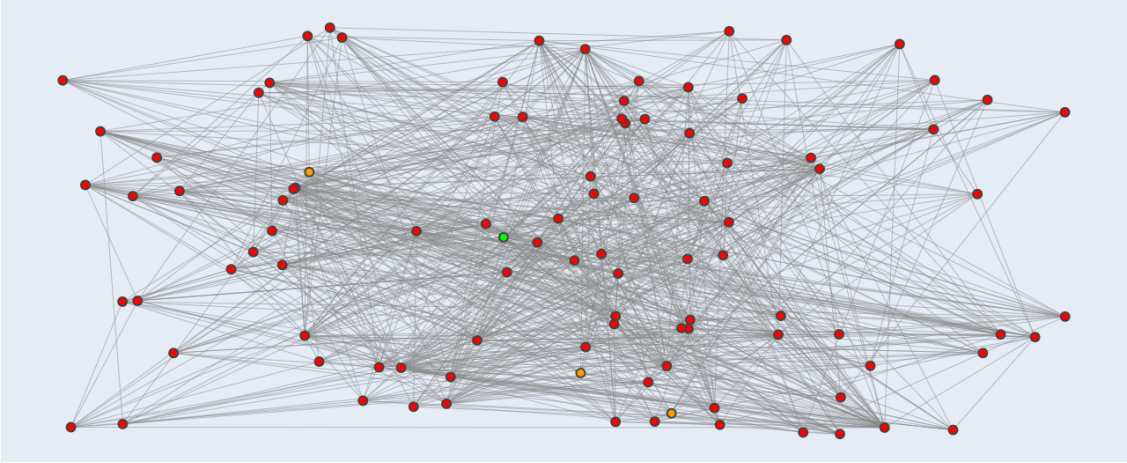


Figura 5.24: Simulación con $N = 100$ en el tiempo $t = 2000$.

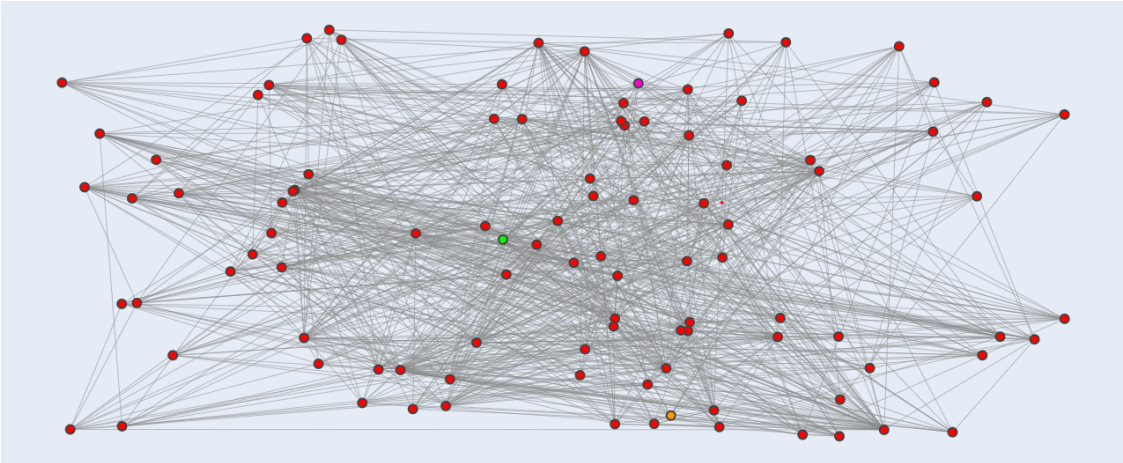


Figura 5.25: Simulación con $N = 100$ en el tiempo $t = 2250$.

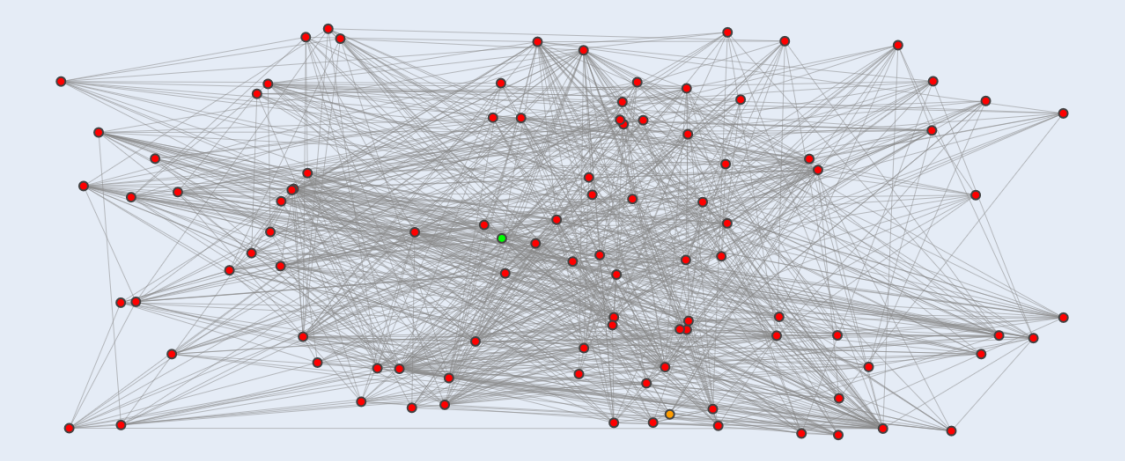


Figura 5.26: Simulación con $N = 100$ en el tiempo $t = 2500$.

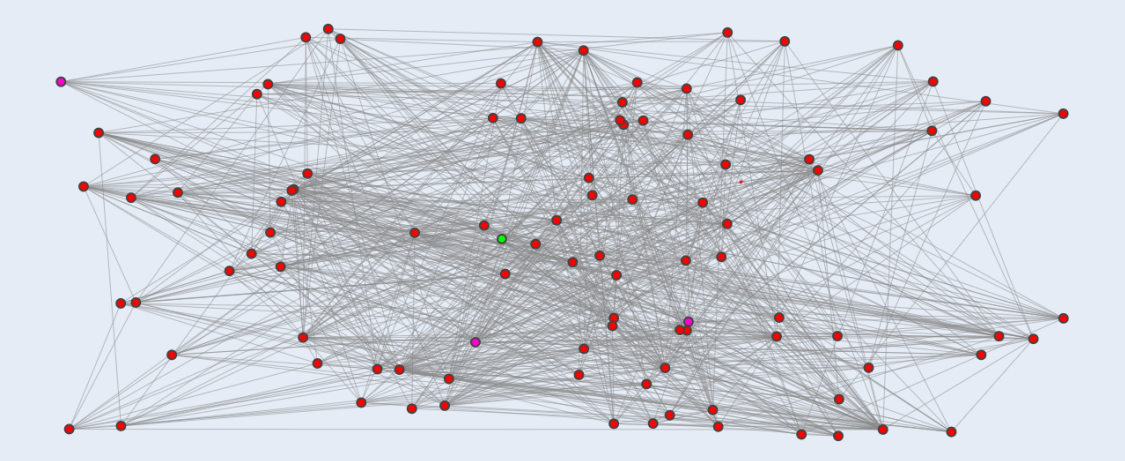


Figura 5.27: Simulación con $N = 100$ en el tiempo $t = 2750$.

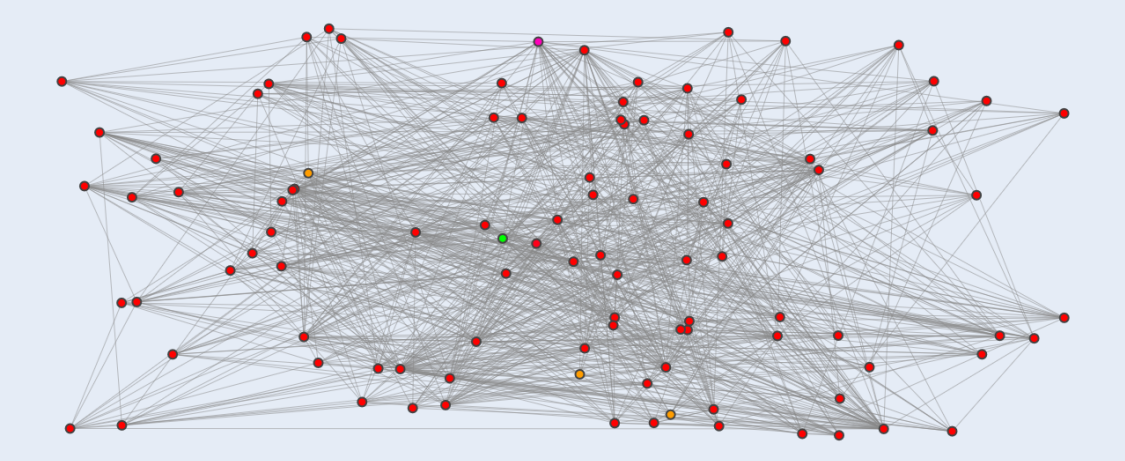


Figura 5.28: Simulación con $N = 100$ en el tiempo $t = 3000$.

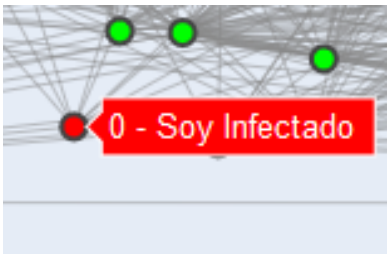


Figura 5.29: Identificador y estado de un nodo en la red celular.

Capítulo 6

Conclusiones y Trabajo Futuro

El uso de dispositivos móviles va en aumento día con día. De acuerdo con el INEGI se prevé que al final del 2020, el número de usuarios de teléfonos inteligentes (smartphones) alcanzará los 80.9 millones de usuarios, ya que se han vuelto un medio indispensable para mantenerse conectado a Internet, incluso en los trabajos y el dispositivo para poder realizar operaciones bancarias y compras. Sin embargo, el aumento en el uso viene acompañado de una explosión de código malicioso, malware, diseñado para atacar smartphones, lo que ha generado que los usuarios estén constantemente en riesgo.

Particularmente, uno de los medios de ataque más usados para la propagación de malware entre smartphone es el SMS, principalmente a través de malware tipo gusano que un tipo de ataque donde los usuarios determinan completamente si el atacante logra su objetivo; es decir, infecta a todos los usuarios conectados en la red telefónica o la mayoría de ellos. Por lo que analizar, predecir y prevenir ataques a través de malware tipo gusano entre smartphones con base en SMS es fundamental en nuestros días.

En este trabajo de tesis, se presentó un nuevo modelo discreto y probabilista, basado en Autómatas Celulares en red. El modelo propuesto se basa en los conceptos de modelos epidemiológicos compartimentales y redes de libre escala para definir la dinámica de propagación entre los smartphones. El modelo toma en cuenta características referentes a la interacción del usuario y su comportamiento y el comportamiento del gusano, que son determinantes para reproducir de manera adecuada la dinámica de propagación.

Particularmente, el modelo toma en cuenta para su definición:

- La topología de la red, para determinar la relación entre los usuarios y su comportamiento, mediante el uso de redes libres de escala;
- Características referentes al usuario: tiempo de lectura de un mensaje, el grado de confianza, la consciencia de riesgo y el umbral de de seguridad;
- Características referentes a los smartphones: tipo de sistema operativo, actividad;
- Fallas en la red telefónica (falla y retraso en la entrega de mensajes), problemas que existen debido al medio en que se transmite; y
- Características propias de gusanos (índice de infección o tipo de SO).

Se realizó un análisis del comportamiento del modelo propuesto variando los valores de los diferentes parámetros y bajo escenarios específicos mediante simulación computacional. Los resultados indican que el modelo propuesto es capaz de replicar de manera adecuada el comportamiento propagación de malware tipo gusano basado en SMS. Particularmente, cuando el grado de confianza, el tiempo de lectura de un mensaje y la conciencia de riesgo se varían, los tiempos de propagación y la cantidad de la población infectada también cambian. Cuando los usuarios son altamente conscientes del riesgo, respecto a la existencia de malware y los riesgos de no usar adecuadamente su smartphone al dar clic sobre las ligas contenidas en sus mensajes SMS, la propagación del gusano se reduce en tiempo y cantidad de la población, de tal manera que es semejante a tener un antivirus para el gusano. Por el contrario, cuando estos factores se consideran en menor medida, la velocidad de propagación de incrementa y se infecta la mayoría de la población, lo que puede comprometer ampliamente la seguridad cibernética de los usuarios y su información. Además, los resultados indican que el reducir la frecuencia con que los usuarios leen sus mensajes retarda la propagación de la infección entre los dispositivos. De tal forma, que el comportamiento del usuario es determinante en la propagación de malware tipo gusano.

En lo referente a las características del gusano, se varió la probabilidad de contagio del malware, encontrándose que a medida que esta se reduce los tiempos de propagación son más lentos, pero si la probabilidad de contagio es alta rápidamente puede infectarse a la población. Una forma que los escritores de software incrementan este riesgo, es cuando escriben un malware cuyo mensaje es sumamente llamativo para el usuario. También, la heterogeneidad de los sistemas operativos de los teléfonos es un medio para reducir la población infectada por el gusano, ya que muchos de los malware existentes comúnmente se desarrollan para afectar a dispositivos con un tipo de sistema operativo; los resultados indican que esta heterogeneidad genera una especia de cerco sanitario para frenar la propagación del gusano. Desafortunadamente, el mercado está dominado por un tipo de sistema operativo, Android, lo que compromete más la seguridad de los usuarios.

Por último, se realizó un análisis de los efectos que generan las fallas ocasionadas en la red telefónica debido al medio en que se transmite sobre la propagación del gusano. Los resultados indican que debido a la manera como opera el envío de mensajes SMS, esto no frena la población infectada por el gusano, sólo genera un retardo en la velocidad de propagación. Por otro lado, cuando la falla refiere a la pérdida de información, que es cuando los mensajes no llegan a su destino, tampoco se evita se infecte toda la población posible, al menos que se cuente con antivirus que evite ello pase, ya que el dispositivo infectado sigue enviando mensajes a su lista de contactos de manera repetitiva.

Los resultados obtenidos y el análisis realizado confirman que la mejor manera de protegerse ante ataques de malware, dejando a lado las vulnerabilidades de los dispositivos, es concientizar a los usuarios acerca de los riesgos que existen en Internet, la existencia de malware, los daños que ocasionan y el uso adecuado del servicio SMS. Actualmente, muchas empresas capacitan a sus empleados sobre este tema, con la finalidad de evitar ataques que comprometan la seguridad de la misma y proteger su información y la los usuarios.

El modelo propuesto es computacionalmente simple, pero también se puede para-

lelizar con la finalidad de trabajar con poblaciones de smartphones es una escala muy grande que permita preservar la dinámica del modelo basada en las interacciones locales de los usuarios.

Se pensaba comparar las simulaciones con resultados reales, pero se solicitaron las bases de datos y éstas eran erróneas, y no se obtuvo una respuesta por parte de los responsables de los mismo, por lo que no se pudo parametrizar el modelo con base en los mismos, ni comparar los resultados obtenidos contra aquellos de un escenario real, como se había deseado al inicio de este trabajo de tesis.

6.1. Trabajo Futuro

Durante el desarrollo de este trabajo, se tuvieron distintas limitaciones debido al equipo con el que se trabajo, respecto al número de dispositivos que formaron la red telefónica. Valdría la pena trabajar más adelante en un equipo con muchas mayores características y recursos que permitan integrar una red con una mayor cantidad de dispositivos móviles, como lo es realmente y paralelizar el modelo propuesto.

Además, de que se obtuvieron datos obtenidos de la simulaciones, valdría la pena intentar conseguir datos reales u obtenidos al menos de un simulador para escenarios específicos, con la finalidad de parametrizar mejor el modelo y evaluar mejor su desempeño.

Aunque el modelo propuesto se enfoca sólo en SMS como vector de ataque, en la realidad existen otros medios de ataque como MMS y Bluetooth y como trabajo futuro seria interesante e importante desarrollar un modelo que considere de manera conjunta dichos vectores de ataque. Así como definir escenarios que consideren a la vez usuarios con comportamientos diferentes y se valúe la propagación del malware.

Bibliografía

[1] S. Wolfram, "Announcing the rule 30 prizes," accessed 2020-11-09. [Online]. Available: https://writings.stephenwolfram.com/2019/10/announcing-the-rule-30-prizes/

[2] ——, "A new kind of science," accessed 2020-11-09. [Online]. Available: https://www.wolframscience.com/nks/p171--cellular-automata/

[3] K. Selvarajoo, "Large scale-free network organization is likely key for biofilm phase transition," *Engineering Biology*, vol. 3, 08 2019.

[4] G. Intelligence, "Extensive datasets with a global reach," accessed 2020-11-09. [Online]. Available: https://www.gsmaintelligence.com/

[5] INEGI, "COMUNICADO DE PRENSA NÚN. 179/19," Apr. 2019, accessed 2019-09-05. [Online]. Available: https://www.inegi.org.mx/datos

[6] J. Ángel Plaza López, "La vida secreta del sms: 4.100 millones de mensajes que nadie envía pero todos recibimos," accessed 2020-12-05. [Online]. Available: https://retina.elpais.com/retina/2019/03/04/tendencias/1551697832_364060.html

[7] "Short Message Service / SMS Tutorial," 2019, accessed 2019-04-10. [Online]. Available: https://www.etsi.org/technologies/smart-cards/sim

[8] A. M. del Rey, "Mathematical modeling of the propagation of malware: a review," *Security and Communication Networks*, vol. 8, no. 15, pp. 2561–2579, 2015.

[9] CBC, "Can you guess the first text message, sent 25 years ago? — CBC News," Dec. 2017, accessed 2019-04-18. [Online]. Available: https://www.cbc.ca/news/technology/text-message-anniversary-1.4430659

[10] A. Vaha-Sipila and E. Wilde, "URI Scheme for Global System for Mobile Communications (GSM) Short Message Service (SMS)," RFC 5724, Jan. 2010. [Online]. Available: https://rfc-editor.org/rfc/rfc5724.txt

[11] S. Wang and X. Duan, "Rtp payload format for gsm-hr speech codecs," accessed 2020-11-10. [Online]. Available: https://tools.ietf.org/html/draft-wang-avt-rtp-gsm-hr-00

[12] M. Sauter, *Global System for Mobile Communications (GSM).* John Wiley & Sons, Ltd, 2017, ch. 1, pp. 1–70. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119346913.ch1

[13] ETSI, "ETSI - Sim Card Technology - Sim Application Toolkit, Smart Card Platform (SCP)," accessed 2019-09-08. [Online]. Available: https://www.etsi.org/technologies/smart-cards/sim

[14] Sabine Dahmen-Lhuissier, "ETSI - 2g - Global System for Mobile Communication (GSM)," accessed 2019-04-18. [Online]. Available: https://www.etsi.org/technologies/mobile/2g

[15] ETSI, "ETSI TS 123 040 V9.3.0," Oct. 2010, accessed 2019-04-19. [Online]. Available: https://www.etsi.org/

[16] ——, "Digital cellular telecommunication system (phase 2+);point-to-point (pp) short message service (sms)support on mobile radio interface(gsm 04.11)," accessed 2020-11-10. [Online]. Available: https://www.etsi.org/deliver/etsi$_$gts/04/0411/05.01.00$_$60/gsmts$_$0411v050100p.pdf

[17] P. J. Marnick and R. G. Russell, "48 - personal communications networks," in *Telecommunications Engineer's Reference Book*, F. Mazda, Ed. Butterworth-Heinemann, 1993, pp. 48–1 – 48–12. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B978075061162650054X

[18] ETSI, "Digital cellular telecommunications system (phase 2+);universal mobile telecommunications system (umts);technical realization of short message service (sms) (3gpp ts 23.040 version 4.10.0 release 4)," accessed 2020-11-10. [Online]. Available: https://www.etsi.org/deliver/etsi$_t$s/123000$_$123099/123040/04.10.00$_$60/$_$123040v041000p.pdf

[19] ETSI, "Short message service (sms) for fixed networks;network based solution (nbs);part 2: Architecture and functional entities," accessed 2020-11-10. [Online]. Available: https://www.etsi.org/deliver/etsi_es/202000_202099/20206002/01.01.01_60/es_20206002v010101p.pdf

[20] J. Neumann *et al.*, *Theory of self-reproducing automata.* University of Illinois Press, 1966, vol. 1102024.

[21] A. Mohanta, M. Hahad, and K. Velmurugan, *Preventing ransomware.* Packt Publishing, 2018.

[22] 2019. [Online]. Available: https://latam.kaspersky.com/resource-center/threats/a-brief-history-of-computer-viruses-and-what-the-future-holds

[23] S. Pagnotta, "Antimalware day: el origen de los virus... y de la protección contra ellos — welivesecurity," 2019. [Online]. Available: https://www.welivesecurity.com/la-es/2017/10/31/antimalware-day-origen-virus/

[24] M. Scheau, L. Arsene, and G. DINCĂ, "Infection vectors – risk factors for financial transactions," *International Journal of Information Security and Cybercrime*, vol. 4, pp. 73–85, 12 2015.

[25] INCIBE, "Amenaza vs vulnerabilidad, ¿sabes en qué se diferencian?" accessed 2020-11-09. [Online]. Available: https://www.incibe.es/protege-tu-empresa/blog/amenaza-vs-vulnerabilidad-sabes-se-diferencian

[26] Norton, "La importancia de las actualizaciones generales de software y los parches," accessed 2020-11-09. [Online]. Available: https://mx.norton.com/internetsecurity-how-to-the-importance-of-general-software-updates-and-patches.html

[27] BBC Mundo, "El mensaje de texto que borra el contenido de tu celular y amenaza tus datos bancarios," 2016, accessed 2019-04-15. [Online]. Available: https://www.bbc.c

[28] Stefanko, Luis, "Video analysis of Android SMS worm spying on victims," 2018, accessed 2019-04-15. [Online]. Available: https://lukasstefanko.com/2018/10/video-analysis-of-android-sms-worm-spying-on-victims.html

[29] IT Digital Security, "El volumen de ataques de malware móvil se duplicó en 2018," accessed 2020-11-10. [Online]. Available: https://www.itdigitalsecurity.es/endpoint/2019/03/el-volumen-de-ataques-de-malware-movil-se-duplico-en-2018

[30] Chebyshev, Victor , "Mobile malware evolution 2018," 2019, accessed 2019-04-15. [Online]. Available: https://securelist.com/mobile-malware-evolution-2018/89689/

[31] Denise Giusto Bilic, "Seguridad en dispositivos móviles: resumen de lo que fue el 2018," 2018, accessed 2019-04-15. [Online]. Available: https://www.welivesecurity.com/la-es/2018/12/21/seguridad-dispositivos-moviles-resumen-2018/

[32] M. L. Santillán, "Epidemiología, útil para describir e investigar la salud de la población," accessed 2020-11-11. [Online]. Available: http://ciencia.unam.mx/leer/887/epidemiologia-util-para-describir-e-investigar-la-salud-de-la-poblacion

[33] J. R. Guzman, "Historia de las estadísticas de salud," 2006.

[34] J. Arriagada S., "James lind (1716–1794)," *Revista Médica Clínica Las Condes*, vol. 30, no. 1, p. 99, 2019, tema central: Investigación clínica aplicada. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0716864019300124

[35] S. López-Moreno, F. Garrido-Latorre, and M. Hernández-Avila, "Desarrollo histórico de la epidemiología: su formación como disciplina científica," *Salud Pública de México*, vol. 42, pp. 133–143, 2000.

[36] A. G. M'Kendrick, "Applications of mathematics to medical problems," *Proceedings of the Edinburgh Mathematical Society*, vol. 44, p. 98–130, 1925.

[37] W. O. Kermack and A. G. McKendrick, "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*, vol. 115, no. 772, pp. 700–721, 1927.

[38] ——, "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*, vol. 138, no. 834, pp. 55–83, 1932.

[39] ——, "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*, vol. 141, no. 843, pp. 94–122, 1933.

[40] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 925–941, 2013.

[41] T. de Camino, "Un lenguaje para la especificación de autómatas celulares con aplicaciones en biología," Ph.D. dissertation, Instituto Tecnológico de Costa Rica, 05 2000.

[42] J. Kroc, P. Sloot, and A. Hoekstra, *Simulating Complex Systems by Cellular Automata*. Springer, 06 2010.

[43] P. C. Tissera, A. M. Printista, and M. L. Errecalde, "Evacuation simulations using cellular automata," *Journal of Computer Science and Technology*, vol. 7, no. 01, pp. 14–20, 2007.

[44] S. Wolfram, "Statistical mechanics of cellular automata," *Reviews of modern physics*, vol. 55, no. 3, p. 601, 1983.

[45] I. Levin, "Exact graph-based analysis of scientific articles on clinical trials," accessed 2020-11-09. [Online]. Available: https://www.biorxiv.org/content/10.1101/164475v1.full

[46] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific american*, vol. 288, no. 5, pp. 60–69, 2003.

[47] P. Held, A. Dockhorn, and R. Kruse, "On merging and dividing of barabási-albert-graphs," in *2014 IEEE Symposium on Evolving and Autonomous Learning Systems (EALS)*, 2014, pp. 17–24.

[48] A.-L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web," *Physica A: statistical mechanics and its applications*, vol. 281, no. 1-4, pp. 69–77, 2000.

[49] D. Watts and S. Strogatz, "Collective dynamics of "small-world" networks. nature, 393: 440–442," *View Article*, 1998.

[50] N. Boccara and K. Cheong, "Automata network SIR models for the spread of infectious diseases in populations of moving individuals," *Journal of Physics A: Mathematical and General*, vol. 25, no. 9, pp. 2447–2461, may 1992.

[51] M. T. Signes-Pont, A. Cortés-Castillo, H. Mora-Mora, and J. Szymanski, "Modelling the malware propagation in mobile computer devices," *Computers & Security*, vol. 79, pp. 80–93, 2018.

[52] X. Xiao, P. Fu, Q. Li, G. Hu, and Y. Jiang, "Modeling and validation of sms worm propagation over social networks," *Journal of computational science*, vol. 21, pp. 132–139, 2017.

[53] X. Xiao, P. Fu, G. Hu, A. K. Sangaiah, H. Zheng, and Y. Jiang, "Saidr: A new dynamic model for sms-based worm propagation in mobile networks," *IEEE Access*, vol. 5, pp. 9935–9943, 2017.

[54] X. Xiao, P. Fu, C. Dou, Q. Li, G. Hu, and S. Xia, "Design and analysis of seiqr worm propagation model in mobile internet," *Communications in nonlinear science and numerical simulation*, vol. 43, pp. 341–350, 2017.

[55] P. Jia, J. Liu, Y. Fang, L. Liu, and L. Liu, "Modeling and analyzing malware propagation in social networks with heterogeneous infection rates," *Physica A: Statistical Mechanics and its Applications*, vol. 507, pp. 240–254, 2018.

[56] C. Zhang and H. Huang, "Optimal control strategy for a novel computer virus propagation model on scale-free networks," *Physica A: Statistical Mechanics and its Applications*, vol. 451, pp. 251–265, 2016.

[57] W. Liu, C. Liu, Z. Yang, X. Liu, Y. Zhang, and Z. Wei, "Modeling the propagation of mobile malware on complex networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 37, pp. 249–264, 2016.

[58] S. Hosseini and M. A. Azgomi, "A model for malware propagation in scale-free networks based on rumor spreading process," *Computer Networks*, vol. 108, pp. 97–107, 2016.

[59] ——, "The dynamics of an seirs-qv malware propagation model in heterogeneous networks," *Physica A: Statistical Mechanics and its Applications*, vol. 512, pp. 803–817, 2018.

[60] S. Hosseini, M. Abdollahi Azgomi, and A. Rahmani Torkaman, "Agent-based simulation of the dynamics of malware propagation in scale-free networks," *Simulation*, vol. 92, no. 7, pp. 709–722, 2016.

[61] C. Gan, M. Yang, Z. Zhang, and W. Liu, "Global dynamics and optimal control of a viral infection model with generic nonlinear infection rate," *Discrete Dynamics in Nature and Society*, vol. 2017, 2017.

[62] S. Huang, "Global dynamics of a network-based wsis model for mobile malware propagation over complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 503, pp. 293–303, 2018.

[63] W. Liu and S. Zhong, "A novel dynamic model for web malware spreading over scale-free networks," *Physica A: Statistical Mechanics and its Applications*, vol. 505, pp. 848–863, 2018.

[64] X. Liu, T. Li, H. Xu, and W. Liu, "Spreading dynamics of an online social information model on scale-free networks," *Physica A: Statistical Mechanics and its Applications*, vol. 514, pp. 497–510, 2019.

[65] A. G. M. Selvam, R. Janagaraj, G. M. Jones, C. Thandalam, and S. INDIA, "Dynamics in a fractional order siqr model of worm propagation," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 3, pp. 549–558, 2018.

[66] X. Yun, S. Li, and Y. Zhang, "Sms worm propagation over contact social networks: Modeling and validation," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2365–2380, 2015.

[67] S. Peng, G. Wang, and S. Yu, "Modeling malware propagation in smartphone social networks," in *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, July 2013, pp. 196–201.

[68] S. Peng, M. Wu, G. Wang, and S. Yu, "Propagation model of smartphone worms based on semi-markov process and social relationship graph," *Computers & security*, vol. 44, pp. 92–103, 2014.

[69] Networkx.github.io, "NetworkX — NetworkX documentation," 2020. [Online]. Available: https://networkx.github.io/

[70] A. G. Miklas, K. K. Gollu, K. K. Chan, S. Saroiu, K. P. Gummadi, and E. De Lara, "Exploiting social interactions in mobile systems," in *International Conference on Ubiquitous Computing*. Springer, 2007, pp. 409–428.

[71] S. Shahyad, S. Pakdaman, M. Hiedary, M. Miri, M. Asadi, A. Nasri, and A. S. Alipour, "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*, vol. 15, pp. 895 – 898, 2011, 3rd World Conference on Educational Sciences - 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042811003867

[72] X. Meng, P. Zerfos, V. Samanta, S. Wong, and S. Lu, "Analysis of the Reliability of a Nationwide Short Message Service," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, 06 2007, pp. 1811 – 1819.

[73] Statista, "Awareness of internet security risks according to internet users worldwide as of August 2016 — Statista," accessed 2020-04-02. [Online]. Available: https://www.statista.com/statistics/463767/awareness-of-online-security-risks/

[74] Tatango, "90 % of SMS Marketing Messages Read in 3 Minutes — Tatango," accessed 2020-04-15. [Online]. Available: https://www.tatango.com/blog/90-of-text-messages-are-read-within-3-minutes/

[75] M. Chau and R. Reith, "Smartphone Market Share," accessed 2020-08-26. [Online]. Available: https://www.idc.com/promo/smartphone-market-share/os

[76] StatCounter, "Mobile Operating System Market Share Mexico," accessed 2020-08-26. [Online]. Available: https://gs.statcounter.com/os-market-share/mobile

[77] Plotly, "Plotly — Graphing Libraries," 2020. [Online]. Available: https://plotly.com/graphing-libraries/