Simulación de la propagación del malware: Modelos continuos vs. modelos discretos

Amparo Fúster Sabater
Instituto de Tecnologías Físicas y de la Información, C.S.I.C.
Email: amparo@iec.csic.es

Ángel Martín del Rey
Departamento de Matemática Aplicada
IUFFyM, Universidad de Salamanca
Email: delrey@usal.es

Gerardo Rodríguez Sánchez
Departamento de Matemática Aplicada
IUFFyM, Universidad de Salamanca
Email: gerardo@usal.es

***Resumen*—La gran mayoría de modelos matemáticos propuestos hasta la fecha para simular la propagación del malware están basados en el uso de ecuaciones diferenciales. Dichos modelos son analizados de manera crítica en este trabajo, determinando las principales deficiencias que presentan y planteando distintas alternativas para su subsanación. En este sentido, se estudia el uso de los autómatas celulares como nuevo paradigma en el que basar los modelos epidemiológicos, proponiendo una alternativa explícita basada en ellos a un reciente modelo continuo.**

***Palabras clave*—Autómatas Celulares (*cellular automata*). Código malicioso (*malware*). Ecuaciones diferenciales (*differential equations*). Epidemiología matemática (*mathematical epidemiology*). Modelización matemática (*mathematical modelling*).**

I. Introducción

El malware es una de las principales amenazas a la Seguridad de la Información con la que nos enfrentamos en la actualidad. Esta amenaza (y los efectos causados), lejos de disminuir, se acrecentará en los próximos años debido fundamentalmente al perfeccionamiento de sus técnicas y fines (APT, Crimeware, etc.) y a la progresiva implantación de la Internet de las Cosas. La lucha contra el malware se lleva a cabo en diferentes frentes: desde la concienciación del usuario para que adopte medidas de seguridad, hasta el desarrollo de software antimalware por parte de las empresas especializadas, pasando por el establecimiento de políticas de seguridad adecuadas en los distintos organismos y compañías, etc. El gran olvidado en este escenario es el desarrollo de software simulador de la propagación de malware. Este tipo de aplicaciones, tan usadas en otros campos como en la propagación de enfermedades infecciosas o de incendios forestales, sería de gran utilidad para el gestor ya que le permitiría simular el comportamiento de la propagación del código malicioso en una red, probar la efectividad de contramedidas y, en definitiva, tomar decisiones adecuadas para la contención de la propagación o, al menos, la minimización de sus efectos nocivos. El software de simulación se deriva de la implementación computacional de un determinado modelo matemático. Así pues el desarrollo de este tipo de modelos que traten de explicar el comportamiento de la propagación del código malicioso es básico.

Existen muy pocos modelos publicados en la literatura científica cuyo propósito sea el mencionado anteriormente; la gran mayoría de ellos se basan en el paradigma heredado de la Epidemiología Matemática y más concretamente en el modelo de Kermack y McKendrick ([5]) que hace uso de un sistema de ecuaciones diferenciales. Aunque estos modelos continuos poseen una sólida base matemática que posibilita un estudio cualitativo muy detallado, presentan serios problemas a la hora de aplicarlos en determinadas situaciones reales. Ello hace conveniente explorar otro tipo de herramientas matemáticas de naturaleza discreta (autómatas celulares, modelos basados en agentes, etc.) que posibiliten el diseño de modelos más eficaces.

El objetivo fundamental de este trabajo es realizar un análisis crítico de los modelos matemáticos propuestos, determinando los puntos fuertes y débiles y, a partir de ello, proponer paradigmas alternativos que permitan soslayar los problemas planteados por los existentes. En este sentido, y con la finalidad de ilustrar las conclusiones obtenidas, se propone un modelo basado en autómatas celulares para el estudio de la propagación de malware, alternativo al desarrollado por Feng *et al.* ([3]) que se basa en ecuaciones diferenciales ordinarias.

El resto del trabajo está organizado como sigue: en la sección II se analizan los modelos matemáticos que se han desarrollado para simular la propagación del malware, determinando ventajas y desventajas; en la sección III se detalla el modelo continuo debido a Feng *et al.*. La propuesta y análisis de la alternativa discreta es presentada en la sección IV; finalmente, en la sección V se presentan las conclusiones.

II. Modelos matemáticos basados en ecuaciones diferenciales

Los modelos matemáticos desarrollados para estudiar la propagación de malware se basan en los modelos diseñados para estudiar la diseminación de las enfermedades infecciosas; ello es debido a las similitudes entre el comportamiento de los virus biológicos, bacterias, hongos o priones y el del malware (virus computacionales, gusanos, etc.) Así pues muchas de las propiedades y características de los primeros se traducen y tienen su reflejo en los segundos (véase [10], [14]), a saber: clases en que se divide la población, la naturaleza del modelo y mecanismos que rigen la dinámica de la infección. Los modelos epidemiológicos de carácter matemático son modelos compartimentales, esto es, la población se divide en diferentes tipos (o compartimentos) teniendo en cuenta las características de la enfermedad: susceptibles, expuestos (con

ISBN: 978-84-9717-323-0

o sin síntomas), infectados, infecciosos, recuperados, en cuarentena, vacunados, aislados, etc. Así pues nos podemos encontrar con modelos SIS (Susceptible-Infectado-Susceptible), modelos SIR (Susceptible-Infectado-Recuperado), modelos SEIR (Susceptible-Expuesto-Infectado-Recuperado), modelos SEIQR (Susceptible-Expuesto-Infectado-Cuarentena-Recuperado), etc.

Consecuentemente, en los modelos cuyo objeto de estudio es el malware podemos encontrar estos mismos compartimentos; así se han propuesto modelos SIS (véase, por ejemplo, [1]), SIR (véase, por ejemplo, [13]), SEIR (véase, por ejemplo, [12]), SEIRS (véase [15]), etc. Se puede ver cómo no existe un tipo de modelo compartimental que centre el mayor número de trabajos sino que se observa una cierta homogeneidad en cuanto a los modelos compartimentales propuestos. Estos modelos se pueden clasificar también atendiendo a la naturaleza y a las herramientas matemáticas en las que se basan. En este sentido nos podemos encontrar con modelos deterministas (véase, por ejemplo [11], [17]) o con modelos estocásticos ([2], [6]). Los modelos deterministas están basados en ecuaciones diferenciales, mientras que los modelos estocásticos hacen uso fundamentalmente de las cadenas de Markov (sobre tiempo y estados continuos o discretos). Los modelos deterministas proporcionan buenos resultados cuando la población es muy grande, mientras que los modelos estocásticos se muestran más eficaces cuando se intenta simular la propagación de malware en redes pequeñas de ordenadores. La gran mayoría de los modelos propuestos (ya sean deterministas o estocásticos) se pueden calificar como modelos globales ya que estudian la dinámica de la población en su conjunto sin tener en cuenta las interacciones locales entre los individuos más allá de lo reflejado en los parámetros. Por el contrario existen muy pocos modelos de carácter individual; todos ellos basados en autómatas celulares (véase [4], [7], [8], [9]).

El objetivo de la inmensa mayoría de los modelos propuestos es el estudio de la dinámica de los diferentes compartimentos en que se divide la población, es decir, el conocimiento del número de ordenadores susceptibles, expuestos, infectados, etc. que hay en cada instante de tiempo y cuál es su tendencia.

Todo modelo matemático viene caracterizado por tres elementos: las variables que se estudian, los parámetros que se utilizan y las relaciones funcionales que rigen la dinámica considerando las variables y parámetros. En el caso de la simulación de la propagación del malware, las variables utilizadas son el número de ordenadores que se encuentran en alguno de los tipos considerados. Los parámetros que se utilizan en la modelización suelen ser los siguientes (el uso de unos u otros depende del modelo implementado y del tipo de malware considerado): tasa de infección, tasa de recuperación (debida al efecto de los antivirus), índice de eliminación de un ordenador de la red, índice de aparición de nuevos ordenadores en la red, probabilidades de paso de un compartimento a otro, probabilidad de adquisición de inmunidad (temporal o indefinida), periodo de latencia, periodo de inmunidad, etc. La evolución de los diferentes compartimentos viene regida por las relaciones funcionales que tienen en cuenta los parámetros introducidos en el modelo. Estas relaciones se pueden articular en torno a diferentes herramientas matemáticas, siendo la más utilizada las ecuaciones diferenciales.

El pilar sobre el que se fundamentan los modelos basados en ecuaciones diferenciales es el modelo de Kermack y McKendrick ([5]). Se trata de un modelo SIR en el que el tamaño de la población se mantiene constante e igual a N y se consideran dos parámetros: el índice de transmisión a, y el índice de recuperación b. La dinámica del modelo se rige según el siguiente sistema de ecuaciones diferenciales ordinarias:

$$\begin{cases} S'(t) = -\frac{a}{N} S(t) I(t) \\ I'(t) = \frac{a}{N} S(t) I(t) \\ R'(t) = bI(t) \end{cases} \tag{1}$$

donde $S(t), I(t)$ y $R(t)$ representan el número de ordenadores susceptibles, infectados y recuperados en el instante t, respectivamente.

El uso de ecuaciones diferenciales permite realizar un detallado análisis matemático del modelo en cuestión. El comportamiento de estos modelos depende fundamentalmente de un parámetro umbral llamado número reproductivo básico, R_0, el cual determinará la estabilidad del equilibrio sin infección (*disease-free equilibrium*) y del equilibrio endémico (*endemic equilibrium*). El número reproductivo básico se define como el número de infecciones secundarias causadas por un único ordenador infectado en una población enteramente susceptible. De esta manera, se demuestra que si $R_0 < 1$ la infección se irá reduciendo (el número de ordenadores infectados decrecerá hasta erradicarse) alcanzándose un estado de equilibrio estable sin infección; si, por el contrario, se verifica que $R_0 > 1$, entonces la infección se propagará (el número de individuos infectados crecerá) llegándose a un estado de equilibrio endémico estable.

Se trata pues de modelos bien fundamentados y coherentes desde el punto de vista matemático y con un detallado estudio de las principales características de su dinámica: estabilidad, equilibrio, etc. No obstante presentan algunos inconvenientes que pasaremos a detallar a continuación y que son debidos a su propia naturaleza:

(1) No tienen en cuenta las interacciones locales entre los ordenadores que forman la red. Se utilizan parámetros como la tasa de infección, la tasa de recuperación, etc. que son de carácter general: el valor del parámetro es constante para todos los elementos de la red o, en algunos casos, sigue una determinada distribución de probabilidad. Consecuentemente no se contempla el uso de parámetros individualizados para cada uno de los ordenadores.

(2) Suponen que los ordenadores que forman la red (a través de la que se propaga el código malicioso) están homogéneamente distribuidos y conectados todos entre sí. Cuando se analiza la propagación del código malicioso de manera macroscópica (en toda Internet, por ejemplo) los resultados que se obtienen dan una aproximación bastante buena de lo que ocurre en la realidad; ahora bien, si analizamos dicha propagación en redes locales,

intranets, etc. los resultados obtenidos son manifiestamente mejorables ya que a escala microscópica la dinámica es muy sensible a las interconexiones locales.

(3) No es posible simular la dinámica individual de cada uno de los elementos de la red. Bien es cierto que, cuando el tamaño de la red es muy grande, el comportamiento general obtenido puede ser muy similar (en cuanto a tendencias) a lo que se produce en la realidad pero se omite el uso de información fundamental: por ejemplo aquellas computadoras cuyo sistema operativo sea Mac OS no se deberían ver afectadas (en el sentido de ser infectadas) por el código malicioso diseñado para sistemas que utilicen Windows (aunque podrían considerarse expuestas), etc.

Consecuentemente, en los modelos basados en ecuaciones diferenciales podemos obtener buenos resultados acerca del comportamiento global aunque no tendremos información sobre el comportamiento individual de cada una de los ordenadores de la red. Estas tres deficiencias fundamentales que presentan estos modelos podrían ser subsanadas si utilizáramos otro tipo de modelos como los basados en autómatas celulares. En éstos es posible tener en cuenta las características individuales de cada una de las computadoras o dispositivos que se encontraran conectados a la red; además podríamos considerar diferentes topologías de red e incluso variarlas con el tiempo. De esta manera tendríamos definido un modelo en el que la dinámica variara en función de los distintos parámetros individuales.

III. El modelo de Feng *et al.*

En [3] Feng *et al.* propusieron un modelo SIRS basado en un sistema de ecuaciones diferenciales ordinarias con retardo para simular la propagación de un determinado código malicioso. Este modelo se caracteriza porque en él se considera una tasa de infección $\beta(t)$ variable, un cierto periodo de inmunidad temporal τ tras la eliminación satisfactoria del malware, y se supone que el número total de ordenadores puede variar con el tiempo: $S(t)+I(t)+R(t)=N(t)$.

La dinámica del mismo viene definida por las siguientes consideraciones:

(1) Un ordenador susceptible pasa a estar infectado con tasa de infección $\beta(t)$. Este índice depende de múltiples factores como: número de ordenadores susceptibles, daños causados por el malware, etc.

(2) Un ordenador susceptible (*resp.* infectado) pasa a estar recuperado con tasa de inmunidad ϕ (*resp.* γ) si sobre él se tienen implementadas diferentes medidas de seguridad: software antivirus, firewall, sistema de detección de intrusos, etc.

(3) Un ordenador recuperado pasa a ser susceptible según la tasa δ después de un cierto periodo de tiempo τ.

Concretamente, las ecuaciones que rigen el modelo son las siguientes:

$$\begin{aligned} S'(t) &= p\Lambda-\beta(t)S(t)I(t)-(\mu+\phi)S(t) \\ &\quad +\delta R(t-\tau) \\ I'(t) &= \beta(t)S(t)I(t)-(\mu+\gamma)I(t) \\ R'(t) &= (1-p)\Lambda+\phi S(t)+\gamma I(t)-\delta R(t-\tau) \\ &\quad -\mu R(t). \end{aligned} \tag{2}$$

Tabla I: Parámetros del modelo debido a Feng *et al.*

Parámetro	Descripción
p	Porcentaje de ordenadores susceptibles
Λ	Número de nuevos nodos
δ	Tasa de pérdida de inmunidad
$\beta(t)$	Tasa de infección en el instante t
μ	Tasa de reposición de ordenadores
ϕ	Tasa de inmunidad proporcionada por el software antivirus
γ	Tasa de recuperación de la infección

de manera que los parámetros utilizados se muestran en la tabla I.

Obsérvese que éstos son parámetros globales, es decir, el valor de cada uno de ellos es constante sobre todos los ordenadores de la red.

Un laborioso cálculo matemático (véase [3]) demuestra que el número reproductivo básico asociado es:

$$R_0=\frac{\beta_0(p\mu+\delta)\Lambda f'(0)}{\mu(\mu+\gamma)(\mu+\delta+\phi)}, \tag{3}$$

donde $f(t)=\frac{\beta(t)I(t)}{\beta_0}$, siendo β_0 la tasa inicial de infección. Además, si $R_0\leq 1$ se obtiene el estado de equilibrio sin infección dado por $E_0^*=(S_0^*,I_0^*,R_0^*)$, donde:

$$S_0^*=\frac{(p\mu+\delta)\Lambda}{\mu(\mu+\delta+\phi)}, I_0^*=0, R_0^*=\frac{(1-p)\Lambda+\phi S_0^*}{\delta+\mu}. \tag{4}$$

Se verifica que E_0^* es globalmente asintóticamente estable para cualquier τ si $R_0<1$. Por otro lado, si $R_0>1$ entonces se alcanza un estado de equilibrio endémico. Se demuestra que dicho estado es localmente asintóticamente estable si $\tau<\tau_0$ e inestable cuando $\tau>\tau_0$, donde τ_0 es un cierto parámetro umbral.

IV. El modelo basado en autómatas celulares

IV-A. *Descripción del modelo*

En el modelo de Feng *et al.* descrito en la Sección III se emplean parámetros generales sin atender a las características específicas de cada uno de los ordenadores que se encuentra en la red ni a las posibles conexiones entre ellos. A fín de tener en cuenta estos condicionantes, proponemos un modelo alternativo basado en autómatas celulares cuyos resultados globales son los similares a los obtenidos por el modelo original pero que, al mismo tiempo, permite incorporar las características particulares de cada uno de los ordenadores y obtener, adicionalmente, la evolución temporal de los mismos.

Los autómatas celulares son modelos simples de computación (un tipo particular de modelos basados en agentes) que

son capaces de simular de manera eficaz y eficiente sistemas complejos (véase [16]). Están formados por un número finito de unidades de memoria denominadas células que se encuentran conectadas entre sí según una cierta topología definida por un grafo, de tal manera que en cada instante de tiempo cada célula está en un estado de entre un número finito de ellos. Este estado va cambiando con el paso discreto del tiempo de acuerdo una regla de transición local cuyas variables son los estados en el instante anterior de la propia célula y sus vecinas (aquellas células adyacentes a la dada).

En el caso que nos ocupa supondremos que cada célula representará un ordenador de la red considerada y que la vecindad de la misma vendrá definida por el conjunto de ordenadores que se encuentran conectados de manera que sea posible la transmisión del malware entre ellos (vía correo electrónico, bluetooth, etc.) A este respecto denotaremos por $[i]$ al i-ésimo ordenador de la red y por $V_i = \{[j_{i,1}], [j_{i,2}], \ldots, [j_{i,v_i}]\}$ a su vecindad. El estado de la célula/ordenador i-ésimo en el instante de tiempo t se denotará por $E_i(t)$ y tomará alguno de los siguientes tres valores: S (susceptible), I (infectado), o R (recuperado). La transición entre dichos estados vendrá regida por las siguientes suposiciones:

(1) *Transición de susceptible a infectado*: El ordenador susceptible $[i]$ pasará estar infectado cuando exista un ordenador vecino infectado, en cuyo caso dicha infección se producirá con probabilidad $\beta_i(t)$. La existencia de un vecino infectado vendrá dada por la siguiente variable booleana:

$$r_{i,j}(t) = \begin{cases} 1, & \text{si } [j] \text{ está infectado en } t \\ 0, & \text{si } [j] \text{ no está infectado en } t \end{cases} \tag{5}$$

donde $[j] \in V_i$.

(2) *Transición de susceptible a recuperado*: El ordenador susceptible $[i]$ pasará a estar recuperado cuando se tomen las medidas necesarias para que el malware no le afecte. Ello se producirá con probabilidad $\phi_i(t)$.

(3) *Transición de infectado a recuperado*: El ordenador infectado $[i]$ pasará a estar recuperado cuando tenga software antivirus instalado, en cuyo caso la recuperación se producirá con probabilidad $\gamma_i(t)$. La existencia de software de protección vendrá dada por la siguiente variable booleana:

$$s_i(t) = \begin{cases} 1, & \text{si } [i] \text{ tiene antivirus instalado en } t \\ 0, & \text{si } [i] \text{ no tiene antivirus instalado en } t \end{cases} \tag{6}$$

(4) *Transición de recuperado a susceptible*: Un ordenador recuperado se mantendrá en este estado durando un cierto periodo de tiempo: τ_i unidades temporales discretas. Posteriormente pasará a encontrarse en estado susceptible con probabilidad $\delta_i(t)$.

Consecuentemente, las respectivas funciones de transición local serán las siguientes:

$$E_i(t+1) = \begin{cases} S & \text{si } E_i(t) = S \text{ y } f_{S\mapsto I}(t) = 0 \\ S & \text{si } E_i(t) = S \text{ y } f_{S\mapsto R}(t) = 0 \\ S & \text{si } E_i(t) = R \text{ y } f_{R\mapsto S}(t) = 1 \\ I & \text{si } E_i(t) = S \text{ y } f_{S\mapsto I}(t) = 1 \\ I & \text{si } E_i(t) = I \text{ y } f_{I\mapsto R}(t) = 0 \\ R & \text{si } E_i(t) = S \text{ y } f_{S\mapsto R}(t) = 1 \\ R & \text{si } E_i(t) = I \text{ y } f_{I\mapsto R}(t) = 1 \\ R & \text{si } E_i(t) = R \text{ y } f_{R\mapsto S}(t) = 0 \end{cases} \tag{7}$$

donde:

$$f_{S\mapsto I}(t) = \bigwedge_{[j]\in V_i} r_{i,j}(t) \vee \Omega_i(t), \tag{8}$$

$$f_{S\mapsto R}(t) = \begin{cases} 1, & \text{con probabilidad } \phi_i(t) \\ 0, & \text{con probabilidad } 1-\phi_i(t) \end{cases} \tag{9}$$

$$f_{R\mapsto S}(t) = \begin{cases} 1, & \text{con probabilidad } \delta_i(t) \\ 0, & \text{con probabilidad } 1-\delta_i(t) \end{cases} \tag{10}$$

$$f_{I\mapsto R}(t) = s_i(t) \vee \Gamma_i(t), \tag{11}$$

siendo:

$$\Omega_i(t) = \begin{cases} 1, & \text{con probabilidad } \beta_i(t) \\ 0, & \text{con probabilidad } 1-\beta_i(t) \end{cases} \tag{12}$$

$$\Gamma_i(t) = \begin{cases} 1, & \text{con probabilidad } \gamma_i(t) \\ 0, & \text{con probabilidad } 1-\gamma_i(t) \end{cases} \tag{13}$$

IV-B. Simulaciones

A continuación se realizarán una serie de simulaciones para poder comparar los dos modelos. En estos casos y para simplificar, no tendremos en cuenta la dinámica poblacional (aparición y desaparición de ordenadores). En ellas se considerarán $n = 500$ ordenadores en la red y se supondrán que inicialmente hay 5 ordenadores infectados ($I(0) = 5$).

En primer lugar se tiene en cuenta un escenario homogéneo (condiciones de los modelos continuos), esto es, se considerará que todos los ordenadores se encuentran conectados entre sí en todo momento (es decir, la topología asociada al autómata celular viene definida por un grafo completo), y se supondrá que todos los ordenadores poseen los mismos valores de los parámetros:

$$\beta_i(t) = \beta(t), \phi_i(t) = \phi, \delta_i(t) = \delta, \gamma_i(t) = \gamma, \quad \forall i. \tag{14}$$

Concretamente usaremos los mismos valores de los parámetros que los empleados por Feng *et al.* en [3], esto es:

$$\gamma = 0{,}2, \beta = 0{,}8, \phi = 0{,}46, \delta = 0{,}7, \tau = 10. \tag{15}$$

La simulación obtenida con el modelo de Feng *et al* se muestra en la figura 1-(a), mientras en que la figura 1-(b) se presenta la simulación obtenida con el modelo discreto. Como se puede apreciar en estas simulaciones las tendencias globales obtenidas en ambos casos son similares aunque en la conseguida a partir del modelo basado en autómatas celulares (escenario individual) se puede observar cómo es más sensible a las interconexiones entre los diferentes elementos de la red. Como se ha comentado anteriormente, el modelo discreto permite obtener también la evolución individual de

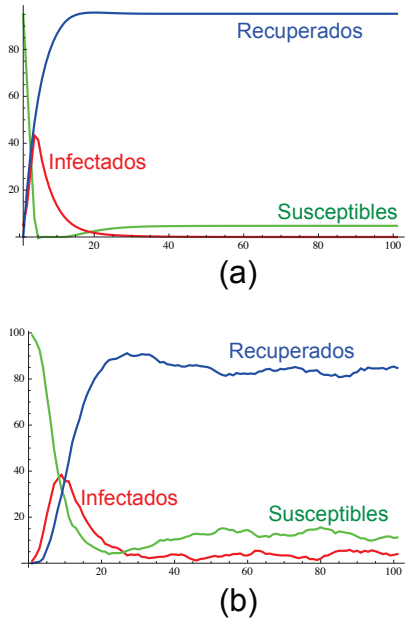

Figura 1: Evolución global de las diferentes clases en un escenario homogéneo. (a) Modelo continuo de Feng *et al.* (b) Modelo discreto propuesto en este trabajo.

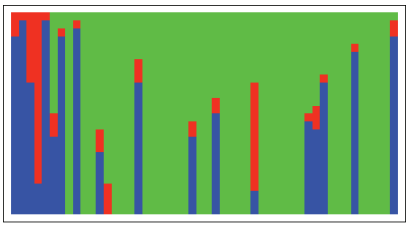

Figura 2: Evolución individual de una colección de ordenadores: cada columna representa un ordenador de manera que el estado susceptible se representa en color verde, el estado infectado en color rojo, y el estado recuperado en color azul.

cada uno de los ordenadores de la red. En la figura 2 se puede observar el diagrama de evolución de estados de una serie de ordenadores: cada columna representa la evolución de un ordenador diferente.

Por otro lado, y dentro del escenario individualizado, supondremos en primer lugar que se mantienen las conexiones según un grafo completo (todo ordenador está conectado con el resto) y que la población se divide en dos grupos atendiendo a las características y prácticas de seguridad que presentan y tienen tanto los ordenadores como sus usuarios. El tipo A estará definido por aquellos ordenadores y usuarios asociados que se preocupen por la seguridad (tengan sistemas operativos y software antivirus instalado y actualizado, tengan prácticas seguras en el uso de Internet, etc.), mientras que el tipo B lo constituirán aquellos dispositivos y usuarios con prácticas más relajadas en temas de seguridad. En la tabla II se muestra el rango de valores numéricos asignados a cada uno de los parámetros para cada uno de los tipos (estos

Tabla II: Valores de los parámetros en el escenario individualizado

Parámetro	Valores (usuarios tipo A)	Valores (usuarios tipo B)
δ_i	$0{,}25 \leq \delta_i \leq 0{,}5$	$0{,}5 \leq \delta_i \leq 0{,}75$
β_i	$0{,}25 \leq \beta_i \leq 0{,}5$	$0{,}5 \leq \beta_i \leq 0{,}75$
ϕ_i	$0{,}5 \leq \phi_i \leq 1$	$0 \leq \phi_i \leq 0{,}5$
γ_i	$0{,}5 \leq \gamma_i \leq 0{,}75$	$0{,}1 \leq \gamma_i \leq 0{,}4$
τ_i	$1 \leq \tau_i \leq 10$	$1 \leq \tau_i \leq 10$

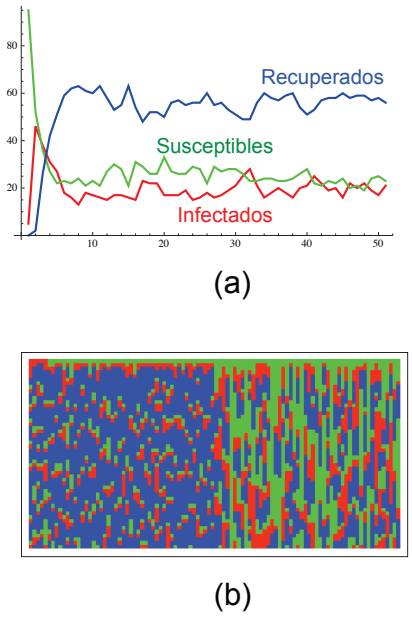

Figura 3: Evolución de las diferentes clases en un escenario individual con topología definida por un grafo completo. (a) Dinámica global (b) Dinámica individual.

valores son meramente ilustrativos). Se supondrá además que los ordenadores se reparten por igual entre los dos tipos.

En la figura 3 se puede observar la evolución tanto global (figura 3-(a)) como individual (figura 3-(b)) de los ordenadores de la red. Obsérvese que los ordenadores correspondientes al tipo A (cuya evolución viene representada por la primera mitad de columnas de la figura 3-(b)) se infectan prácticamente en la misma proporción que el resto pero se recuperan antes y permanecen en dicho estado mucho más tiempo que el resto.

Por otra parte, y dentro del escenario individualizado, supondremos a continuación que la topología de la red de ordenadores no viene definida por un grafo completo sino por el grafo que se muestra en la figura 4 (en gris se encuentran representados los ordenadores del tipo A, mientras que en negro se colorean los ordenadores correspondientes al tipo B). En este caso supondremos que los parámetros siguen lo establecido en la tabla II. En la figura 5 se ilustra la situación que presenta la red en tres instantes de tiempo: $t = 0,3$ y $t = 6$.

Obsérvese que los ordenadores del cúmulo de la izquierda (todos ellos pertenecientes al tipo A) tardan más tiempo en

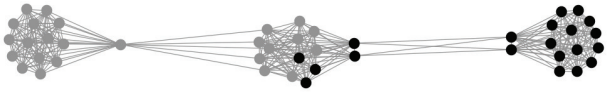

Figura 4: Grafo que define la topología de la red.

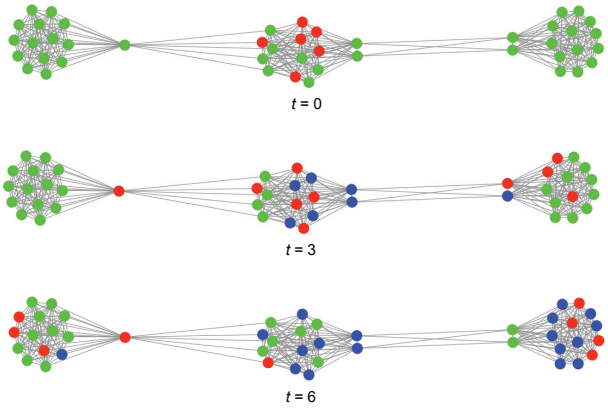

Figura 5: Evolución individual de las diferentes clases.

infectarse que el resto de los ordenadores. Asimismo, se comprueba cómo más del $50\,\%$ de los ordenadores del cúmulo de la derecha (todos ellos son del tipo B) se infectan en los 5 primeros pasos de tiempo.

V. Conclusiones

Los modelos matemáticos diseñados para simular la propagación de malware en redes de ordenadores son eminentemente de naturaleza determinista y continua, y su dinámica se basa en sistemas de ecuaciones diferenciales ordinarias.

Estos modelos, debido al paradigma en el que se fundamentan, presentan las siguientes deficiencias:

(a) Consideran que todos los ordenadores se encuentran conectados entre sí. Consecuentemente no se tienen en cuenta la conexiones locales entre los elementos de la red.

(b) No tienen en cuenta las características individuales de los ordenadores que forma la red, esto es, los parámetros de los que depende la dinámica del malware, son globales: se utilizan los mismos para todos los ordenadores.

Estos problemas se pueden solventar si basamos los modelos en otro tipo de herramientas matemáticas que permitan incorporar las características propias de cada uno de los ordenadores, a saber: tipo de sistema operativo instalado y frecuencia con la que se actualiza, tipo de software de seguridad instalado (firewall, software antivirus, etc.), concienciación del usuario en temas de seguridad, prácticas del usuario, etc.

Así se considera el uso de los autómatas celulares como posible herramienta para el diseño de dichos modelos. En este sentido se estudia el modelo basado en ecuaciones diferenciales propuesto por Feng *et al.* y se propone una alternativa al mismo basada en un autómata celular booleano. Se comprueba que las simulaciones obtenidas en el caso homogéneo (en el que se suponen las condiciones de los modelos continuos) son similares en ambos modelos cuando el número de ordenadores es elevado. Asimismo, se han realizado también simulaciones en el caso individual (cuando los valores de los parámetros varían con el ordenador) y se han mostrado tanto la evolución global como la individual (cosa que no es posible con el modelo basado en ecuaciones diferenciales). Se comprueba como el modelo basado en autómatas celulares es más sensible a las conexiones locales entre los diferentes elementos de la red; asimismo, produce resultados más ajustados a la realidad que el modelo basado en ecuaciones diferenciales cuando el número de ordenadores es pequeño.

Agradecimientos

Este trabajo ha sido subvencionado por el Ministerio de Economía y Competitividad bajo el proyecto TUERI (TIN2011-25452) y por la Consejería de Educación de la Junta de Castilla y León.

Referencias

[1] J. Amador, J.R. Artalejo, "Modeling computer virus with the BSDE approach,"*Computer Networks*, vol. 57, pp. 302-316, 2013.

[2] J. Amador, J.R. Artalejo, "Stochastic modeling of computer virus spreading with warming signals,"*Journal of the Franklin Institute*, vol. 350, pp. 1112-1138, 2013.

[3] L. Feng, X. Liao, Q. Han, H. Li, "Dynamical analysis and control strategies on malware propagation model," *Applied Mathematical Modelling*, vol. 37, pp. 8225–8236, 2013.

[4] J. Hao, J. YIN, B. Zhang, "Modeling viral agents and their dynamics with persistent turing machines and cellular automata,"*Lecture Notes in Computer Science*, vol. 4088, pp. 690-695, 2006.

[5] W.O. Kermack, A.G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics,"*Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 115, pp. 700-721, 1927.

[6] S. Kondakci, "Epidemic state analysis of computers under malware attacks,"*Simulation Modelling Practice and Theory*, Vol. 16, pp. 571-584, 2008.

[7] A. Martín del Rey, "A Computer Virus Spread Model Based On Cellular Automata of Graphs,"*Lecture Notes in Computer Science*, vol. 5518, pp. 503-506, 2009.

[8] A. Martín del Rey, "A SIR e-Epidemic model for computer worms based on cellular automata,"*Lecture Notes in Artificial Intelligence*, vol. 8109, pp. 228-238, 2013.

[9] A. Martín del Rey, G. Rodríguez Sánchez, "A discrete mathematical model to simulate malware spreading,"*International Journal of Modern Physics C*, vol. 23, paper id. 1250064, 2012.

[10] M. Meisel, V. Pappas, L. Zhang, "A taxonomy of biologically inspired research in computer networking,"*Computers Networks*, vol. 54, pp. 901-916, 2010.

[11] B.K. Mishra, N. Keshri, "Mathematical model on the transmission of worms in wireless sensor network,"*Applied Mathematical Modelling*, vol. 37, pp. 4103-4111,2012.

[12] B.K. Mishra, D.K. Saini, "SEIRS epidemic model with delay for transmission of malicious objects in computer network,"*Applied Mathematics and Computation*, vol. 188, pp. 1476-1482, 2007.

[13] J. Ren, X. Yang, L.X. Yang, Y. Xu, F. Yang, "A delayed computer virus propagation model and its dynamics,"*Chaos, Solitons & Fractals*, vol. 45, pp. 74-79, 2012.

[14] M. Rice, J. Butts, R. Miller, S. Shenoi, "Applying public health strategies to the protection of cyberspace,"*International Journal of Critical Infrastructure Protection*, vol. 3, pp.118-127, 2010.

[15] O.A. Toutonji, S.M. Yoo, M. Park, "Stability analysis of VEISV propagation modeling for network worm attack,"*Applied Mathematical Modelling*, vol. 36, pp. 2751-2761, 2012.

[16] S. Wolfram, "A New Kind of Science," Champaign, IL: Wolfram Media Inc., 2002.

[17] Y. Yao, L. Guo, H. Guo, G. Yu, F.X. Gao, X.J. Tong, "Pulse quarantine strategy of internet worm propagation: Modeling and analysis,"*Computers and Electrical Engineering*,vol. 38, pp. 1047-1061, 2012.