Modelos matemáticos para la propagación de malware: estado del arte y algunas reflexiones

A. Martín del Rey

Resumen Este capítulo está dedicado al análisis de modelos matemáticos para la propagación de malware en diferentes tipos de redes de comunicación (desde grandes como Internet hasta más específicas como las redes IoT). Específicamente, se presenta una revisión detallada del estado del arte estudiando tanto modelos deterministas como estocásticos, continuos y discretos, y modelos globales y basados en individuos. Finalmente, se describirán en detalle diferentes líneas de investigación futura.

1. Introducción

Gracias al desarrollo de las ciencias de la computación y a la aparición y uso generalizado de Internet, desde la década de los 90 del siglo XX se ha producido un impresionante desarrollo tecnológico: la primigenia *Internet de los Contenidos* (E-mail, HTML) dió paso a la *Internet de los Servicios* (comercio electrónico, servicios electrónicos, etc.), que a su vez propició la *Internet de la Gente* (redes sociales), dando lugar a la *Internet de las Cosas*: Industria 4.0, ciudades inteligentes, e-Health, agricultura inteligente, transporte inteligente, edificios inteligentes, etc. Vivimos pues en una sociedad altamente dependiente de los sistemas y servicios informáticos tanto a nivel individual (los ciudadanos particulares pueden realizar trámites on-line de todo tipo: compras, pago de impuestos, tramitación de matrículas escolares, solicitud de permisos, presentación de solicitudes, etc.), como a nivel colectivo (multitud de servicios tanto públicos como privados -entre los que se incluyen los servicios esenciales y las infraestructuras críticas- se encuentran gestionados por sistemas informáticos recayendo en ellos su buen funcionamiento y disponibilidad).

A. Martin del Rey
Universidad de Salamanca, Departamento de Matemática Aplicada, IUFFyM, 37008-Salamanca (España). e-mail: delrey@usal.es

Como no podía ser de otra manera, este escenario se encuentra permanentemente sometido a ciberamenazas de todo tipo y complejidad [1, 36, 38]: desde los clásicos ataques de physing hasta las complejas *Amenazas Persistentes Avanzadas* (APT). Este hecho ha propiciado el desarrollo de múltiples algoritmos y metodologías para garantizar, en la medida de las posibilidades, la ciberseguridad de estos entornos [8, 46]. Una de las principales herramientas utilizadas en la inmensa mayoría de estos ciberataques es el código malicioso (comúnmente conocido como *malware*) [15]. Según la *Kaspersky IT Encyclopedia* "malware refers to any program that is deliberately created to perform an unauthorized, often harmful, action" [29]. Así, los diferentes especímenes de malware aprovechan las potenciales vulnerabilidades de las redes y sus dispositivos para infiltrarse en ellas y desarrollar su labor maliciosa a través de los procesos de propagación por la red y de infección en los dispositivos.

La actividad de la comunidad científica en relación al malware se ha centrado principalmente en el desarrollo de modelos basados en Aprendizaje Automático para detectar de forma eficiente su presencia en una cierta red informática. De esta manera en la literatura científica se pueden encontrar multitud de algoritmos basados en diferentes técnicas y procedimientos (véase, por ejemplo, [5, 21, 39, 41]). Otra aproximación al estudio del malware (mucho más minoritaria) se basa en estudiar y tratar de predecir la propagación del malware en una red. De esta forma se han ido proponiendo modelos matemáticos y computacionales para simular la difusión de diferentes especímenes de malware en distintos entornos [43, 44].

El desarrollo (y estudio) de estos modelos se fundamenta en las técnicas de la Epidemiología Matemática, que es la disciplina científica cuyo objetivo original era el diseño y análisis de modelos matemáticos para simular la propagación de agentes biológicos (virus, bacterias, hongos, etc.) De forma más concreta, las tres cuestiones básicas que se trata de responder la Epidemiología Matemática son las siguientes [10]: (1) ¿un brote infeccioso derivará en un proceso epidémico?, (2) en caso de que así ocurra... ¿a qué velocidad crecerá el número de individuos infectados?, y (3) ¿cuál será el número total de personas infectadas a lo largo de toda la epidemia? Ahora bien, desde finales del siglo pasado y debido al gran desarrollo de las tecnologías y redes de comunicaciones, la noción de "agente propagable" ha cambiado dejando de ser algo de naturaleza exclusivamente biológica para pasar a representar también otro tipo de entes: información, rumores, ideologías, marketing viral, malware, etc. Así, hasta donde llega mi conocimiento, el primer trabajo que sugería describir la transmisión de ideas como un proceso epidémico apareció en 1964 y fue debido a W. Goffman y V.A. Newill [19]; al año siguiente se propone un modelo para simular la propagación de rumores [9]; el primer modelo matemático para simular la propagación de virus computacionales es publicado en 1989 [17]; la propagación de comportamientos fanáticos utilizando modelos epidemiológicos es propuesta por C. Castillo-Chaves y B. Song en 2003 [7]; en 2007 se empieza a trabajar en el estudio de la diseminación de marketing viral utilizando modelos matemáticos [32]; el primer trabajo sobre la predicción de resultados electorales basada en modelos epidemiológicos aparace en 2020 [50].

En este trabajo nos centraremos en revisar y analizar cómo se ha utilizado la metodología para el diseño de modelos de propagación de agentes biológicos (Epi-

demiología Matemática clásica) en el desarrollo de modelos de difusión de código malicioso en redes informáticas. Se realizará una exposición clara de los diferentes tipos de modelos que han aparecido hasta el momento y, a partir de un análisis crítico de los mismos, se propondrán nuevas vías de investigación para el desarrollo de modelos mejorados y más realistas.

El resto del capítulo se organiza como sigue: en la Sect. 2 se presentan las bases de la Epidemiología Matemática clásica; las consideraciones generales sobre la modelización matemática de la propagación del malware se muestran en la Sect. 3; en la Sect. 4 se detalla el proceso clásico del diseño de este tipo de modelos (modelos globales) mostrando los ejemplos más importantes; la descripción del paradigma de la modelización basada en el individuo, como alternativa a la modelización clásica, es mostrada en la Sect. 5.2; finalmente, las conclusiones son presentadas en la Sect. **??**

2. Fundamentos básicos de la Epidemiología Matemática

La Epidemiología Matemática tiene una larga tradición pues el primer estudio matemático fue realizado por Daniel Bernoulli en 1760 y tenía como objetivo analizar el proceso de propagación del virus de la viruela para cuantificar las ventajas de la variolización [3]. En 1889 P.D. En'ko propone el que se considera como primer modelo moderno cuyo objetivo se centró en estudiar la propagación del sarampión en San Petersburgo [12]. No obstante, los fundamentos de la moderna Epidemiología Matemática se establecieron en el primer tercio del siglo XX. Así, en 1906 W.H. Hamer reformula en el ámbito de la epidemiología el concepto de *ley de acción de masas* (proveniente de la Química), proponiendo de esa manera que el aumento del número de contagios es proporcional al número de individuos tanto infecciosos como susceptibles [22]. Posteriormente, en 1910, R. Ross utiliza la idea planteada por Hamer para diseñar un modelo de propagación de la malaria considerando a los mosquitos como vector de transmisión y en el que demuestra que la reducción de la población de tales mosquitos por debajo de un cierto valor umbral sería suficiente para detener el brote epidémico (*mosquito theorem*) [47]. La importancia de este trabajo radica en que es la primera vez en la que se menciona la existencia de un coeficiente umbral para que se produzca una epidemia (crecimiento de infectados), lo cual daría lugar con posterioridad al concepto de *número reproductivo básico*. En 1917 el propio Ross y H. Hudson caracterizan matemáticamente por primera vez el concepto de *incidencia* (número de nuevos infectados por unidad de tiempo) de la siguiente manera [48]:

$$\iota(t) = \frac{S(t)}{N} \int_0^\infty c\, F(t,\tau)\, d\tau, \tag{1}$$

donde N es el número total de individuos de la población, $S(t)$ es el número de individuos susceptibles en el instante de tiempo t, $c > 0$ es el coeficiente de

infectividad, τ es el tiempo transcurrido desde que un individuo se infectó ("edad de infección"), y $F(t,\tau)$ es el número de individuos contagiados en el instante $t-\tau$ que siguen estando infectados en el instante t.

Ahora bien, es en 1927 cuando los escoceses W.O. Kermack y A.G. McKendrick diseñan el primer modelo epidemiológico compartimental de naturaleza matemática [31]. Este ha sido la fuente de inspiración de gran parte de los modelos que han aparecido posteriormente en la literatura científica [23]. Se trata del famoso modelo compartimental SIR en el que la población se divide en tres clases (o compartimentos): susceptibles S, infecciosos I, y recuperados R (véase la Fig. 1). Es un modelo determinista de naturaleza global puesto que las variables a estudiar representan la densidad (o número) de individuos que se encuentran en alguno de los tres estados mencionados anteriormente en un cierto instante de tiempo: $0 \leq S(t), I(t), R(t) \leq 1$. Su dinámica viene descrita por un sistema de ecuaciones diferenciales ordinarias cuya versión original es la siguiente:

$$\begin{cases} S'(t) = -S(t)\left(\int_0^t A(\tau)\, v(t-\tau)\, d\tau + A(t)\, i_0\right) \\ I(t) = \int_0^t B(\tau)\, v(t-\tau)\, d\tau + B(t)\, i_0 \\ R'(t) = \int_0^t C(\tau)\, v(t-\tau)\, d\tau + C(t)\, i_0 \\ v(t) = -S'(t) \end{cases} \tag{2}$$

$$S(0) = s_0, I(0) = i_0, R(0) = r_0 \tag{3}$$

donde la población se mantiene constante a lo largo del tiempo, $S(t) + I(t) + R(t) = 1$, $\forall t$, y

$$A(\tau) = \phi(\tau)\, B(\tau), \quad B(\tau) = \exp\left[-\int_0^\tau \psi(z)\, dz\right], \quad C(\tau) = \psi(\tau)\, B(\tau), \tag{4}$$

de manera que $\phi(\tau)$ es la tasa de infección en la "edad de infección" τ, y $\psi(\tau)$ es la tasa de recuperación en τ. Obsérvese que $v(t)$ representa la incidencia en el instante t.

Si suponemos que $\phi(\tau) = \kappa \in \mathbb{R}^+$ y $\psi(\tau) = \delta \in \mathbb{R}^+$ son funciones constantes, entonces un simple cálculo muestra que el sistema (2) se reduce a la siguiente descripción matemática simplificada del fenómeno:

$$\begin{cases} S'(t) = -\alpha I(t) S(t) \\ I'(t) = \alpha I(t) S(t) - \beta I(t) \\ R'(t) = \beta I(t) \end{cases} \tag{5}$$

$$S(0) = s_0, I(0) = i_0, R(0) = 1 - s_0 - i_0 \tag{6}$$

donde $\alpha \geq 0$ es la tasa de contagio, $\beta \geq 0$ es la tasa de recuperación.

Figura 1 Diagrama de flujo representativo del modelo epidemiológico SIR.

La aportación más importante de Kermack y McKendrick fue el llamado *Teorema Umbral* en el que el valor de un parámetro, $\mathcal{R}_0 = \frac{\alpha}{\beta}$, determina la evolución del brote infeccioso:

Teorema 1 (Teorema Umbral) *Existe un valor umbral para el número de individuos susceptibles en el instante inicial, s_0, que determina la evolución del número de infecciosos. Así:*

- *Si $s_0 \leq \frac{\beta}{\alpha}$ entonces $I(t)$ es monótona decreciente tal que $\lim_{t\to\infty} I(t) = 0$.*
- *Si $s_0 > \frac{\beta}{\alpha}$ entonces $I(t)$ inicialmente crecerá hasta alcanzar el valor máximo*

$$I_{max} = 1 - \frac{\beta}{\alpha}\left(1 + \log\left(\frac{\alpha s_0}{\beta}\right)\right) \tag{7}$$

y posteriormente decrecerá de manera que $\lim_{t\to\infty} I(t) = 0$.

En consecuencia si la población inicial es enteramente susceptible y $\mathcal{R}_0 \leq 1$, entonces $I(t)$ decrecerá hasta extinguirse, mientras que si $\mathcal{R}_0 > 1$ $I(t)$ crecerá hasta un cierto máximo y posteriormente decrecerá hasta extinguirse.

Adicionalmente un sencillo estudio analítico nos muestra que el sistema anterior posee un único punto de equilibrio, denominado punto de equilibrio *libre de infección* $P_0^* = (s_\infty, 0, 1 - s_\infty)$, de manera que s_∞ es la única solución en el intervalo $[0, 1]$ de la siguiente ecuación:

$$s_\infty = s_0 e^{-\frac{\alpha}{\beta}(1-s_\infty)}. \tag{8}$$

En la Fig. 2-(a) se puede observar la gráfica con la evolución temporal de las densidades de los diferentes comparimentos en el caso $\alpha = 0{,}68, \beta = 0{,}33$, y $s_0 = 0{,}99$. En esta situación $\mathcal{R}_0 \approx 2{,}06 > 1$, $I_{max} \approx 0{,}17$, y $P_{0*} \approx (0{,}18, 0, 0{,}82)$. En la Fig.2-(b) se representa el diagrama de fases correspondiente al modelo: se observa como todas las órbitas tienen hacia algún punto en el eje horizontal caracterizado por $I = 0$.

La determinación de la *incidencia* (número/densidad de nuevos infectados por unidad de tiempo) es, sin duda alguna, el elemento clave en el diseño de cualquier modelo epidemiológico de naturaleza matemática. En el caso más sencillo esta vendrá dada por el siguiente término:

$$\text{incidencia} = \alpha I(t) S(t) = \frac{qc}{N} I(t) S(t), \tag{9}$$

donde q es la probabilidad de que un contacto adecuado acabe en contagio, N es el número total de individuos de la población y c es la tasa de contacto que "mide" el

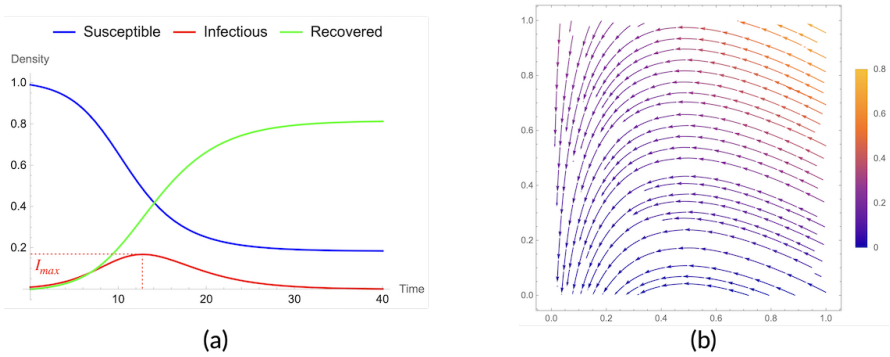

Figura 2 (a) Evolución temporal de las densidades de los compartimentos del modelo SIR. (b) Diagrama de fases S-I del modelo SIR.

número de contactos por unidad de tiempo de un individuo con el resto de individuos de la población. Su determinación es de capital importancia y normalmente la suponemos dependiente del número de individuos: $c = c\,(N)$ de manera que puede adoptar diferentes formas, entre las que podemos destacar las siguientes:

- Tasa de contacto bilineal: $c = c\,(N) = c_0 N$ con $c_0 > 0$.
- Tasa de contacto estándar: $c = c_0$ con $c_0 > 0$.
- Tasa de contacto con saturación respecto a N: $c = c\,(N)$ es una función tal que verifica las siguientes condiciones:

$$c\,(0) = 0, \quad \text{(sin población no hay contactos)} \tag{10}$$

$$c'\,(N) \geq 0, \quad \text{(es una función creciente)} \tag{11}$$

$$\lim_{N\to\infty} c\,(N) = c_0, \quad \text{(aumenta hacia un valor límite)} \tag{12}$$

$$\frac{d}{dN}\left(\frac{c\,(N)}{N}\right) \leq 0 \quad \text{(la media de contactos no debe crecer).} \tag{13}$$

Adicionalmente, y dependiendo de las características del agente biológico a estudiar y del medio por el que se propaga, se podrían considerar tasas de contacto saturadas respecto de la densidad de infecciosos $I(t)$, por ejemplo.

Con el ánimo de mejorar y hacer más realistas los modelos, la anterior familia se puede modificar teniendo en cuenta que la topología de contactos viene definida por una distribución de grado, $P(k)$, correspondiente a una red compleja no completa (red aleatoria, red de mundo pequeño, red libre de escala, etc.). De esta forma se modifica la incidencia incluyendo de manera adecuada dicha distribución de grado y dividiendo los distintos compartimentos en los que se clasifica la población en diferentes subcompartimentos en función del número de contactos de los individuos que los forman; es decir, se tiene en cuenta el grado de los nodos (individuos) que forman la red compleja [28]. Así, el modelo básico SIR sobre una red compleja (derivado del modelo de Kermack y McKendrick) se define por el siguiente sistema

de ecuaciones diferenciales:

$$\begin{cases} S_k'(t) = -qk\Theta(t) S_k(t) \\ I_k'(t) = qk\Theta(t) S_k(t) - \beta I_k(t) \\ R_k'(t) = \beta I_k(t) \end{cases} \tag{14}$$

$$S_k(0) = 1 - i_{k,0}, I_k(0) = i_{k,0}, R_k(0) = 0. \tag{15}$$

$$S_k(t) + I_k(t) + R_k(t) = 1. \tag{16}$$

donde $S_k(t)$, $I_k(t)$ y $R_k(t)$ representan, respectivamente, las densidades de k-nodos susceptibles, infecciosos y recuperados que existen en el instante de tiempo t, con $1 \leq k \leq N-1$. Así

$$S(t) = \sum_{k=1}^{N-1} S_k(t), I(t) = \sum_{k=1}^{N-1} I_k(t), R(t) = \sum_{k=1}^{N-1} R_k(t). \tag{17}$$

Por otra parte la función $\Theta(t)$ representa la densidad media de nodos infecciosos que se encuentran conectados en el instante de tiempo t con un k-nodo susceptible:

$$\Theta(t) = \frac{1}{\langle k \rangle} \sum_{k=1}^{N-1} (k-1) P(k) I_k(t), \tag{18}$$

donde $\langle k \rangle$ es el grado medio de la red compleja.

Estos siguen siendo modelos deterministas, globales y cuya dinámica viene descrita por sistemas de ecuaciones diferenciales ordinarias. Al igual que en el caso anterior, el análisis cualitativo de los mismos permite obtener condiciones de estabilidad para los puntos de equilibrio en función de un parámetro umbral [16]:

Teorema 2 *Sea* $\mathcal{R}_0 = \frac{q\langle k \rangle^2}{\beta \langle k \rangle}$, *donde* $\langle k \rangle^2 = \sum_{k=1}^{N-1} k^2 P(k)$. *Se verifica que:*

(1) Si $\mathcal{R}_0 \leq 1$ *el sistema evoluciona hacia el estado de equilibrio libre de infección* $P_0^* = (1, 0, 0)$.

(2) Si $\mathcal{R}_0 > 1$ *el sistema evoluciona hacia el estado de equilibrio libre de infección*

$$Q_0^* = \left(\sum_{k=1}^{N-1} P(k) e^{-qk\phi_\infty}, 0, \sum_{k=1}^{N-1} P(k) \left(1 - e^{-qk\phi_\infty}\right) \right), \tag{19}$$

donde ϕ_∞ *es la única solución no trivial de la ecuación:*

$$\phi_\infty = \frac{1}{\beta} - \frac{1}{\beta \langle k \rangle} \sum_{k=1}^{N-1} kP(k) e^{-qk\phi_\infty}. \tag{20}$$

El paradigma de las redes complejas no sólo puede ser empleado en Epidemiología Matemática del modo anterior sino que, y teniendo en mente conseguir modelos lo mas realistas posibles, la red compleja puede representar un conjunto de comunidades (nodos) que se encuentran interconectadas entre sí (enlaces) mediante

diferentes tipos de medios de transporte posibilitando de esta manera el flujo poblacional entre ellas. Así pues, la dinámica viene descrita por un sistema de ecuaciones diferenciales ordinarias cuyas variables son las fracciones de individuos de cada tipo en cada comunidad y en cada instante de tiempo. Por ejemplo C.G. Antonopoulos *et al.* propusieron un modelo compartimental SVIR (Susceptible-Vaccinated-Infected-Recovered) en donde se simulaba la propagación de un agente biológico entre un grupo de N comunidades - metapopulation model - y se evaluaban diferentes estrategias de vacunación [2]:

$$\begin{cases} S_i'(t) = -\alpha_i S_i(t) I_i(t) - \phi_i S_i(t) + \delta_i R_i(t), \\ V_i'(t) = \phi_i S_i(t) - \rho_i \alpha_i I_i(t) V_i(t), \\ I_i'(t) = \alpha_i S_i(t) I_i(t) + \rho_i \alpha_i I_i(t) V_i(t) - \lambda_i I_i + \beta \sum_{j=1}^{N} L_{ij} I_j, \\ R_i'(t) = \lambda_i I_i - \delta_i R_i(t), \end{cases} \tag{21}$$

$$1 \leq i \leq N,$$

$$N = \sum_{i=1}^{N} \left(S_i(t) + V_i(t) + I_i(t) + R_i(t) \right). \tag{22}$$

donde los coeficientes asociados a la comunidad i-ésima son los siguientes: α_i tasa de contagio, ϕ_i es la fracción de población susceptible que es vacunada por unidad de tiempo, δ_i representa la tasa de pérdida de inmunidad, ρ_i mide la eficacia de la vacunación, λ_i es la tasa de recuperación, β es la "fuerza" de conexión entre las diferentes comunidades (flujo de individuos) y $L = \left(L_{ij}\right)_{1 \leq i,j \leq N}$ es la matriz Laplaciana asociada a la red compleja. Obviamente se puede realizar un estudio cualitativo de estos modelos y además es posible obtener simulaciones no sólo a nivel global sino a nivel "local" de la evolución en cada una de las comunidades.

3. Consideraciones sobre la modelización matemática de la propagación de malware

Desde un punto de vista estrictamente matemático, originariamente la modelización de la propagación de malware se ha fundamentado en la Epidemiología Matemática ya que, aunque los agentes "propagables" (código malicioso y agentes biológicos) son de diferente naturaleza (los procesos de progación e infección pueden llegar a ser muy distintos entre ellos), la descripción matemática de los distintos términos que intervienen en la dinámica (principalmente la incidencia) son similares.

El fenómeno de la propagación de un especimen de malware en una red informática involucra a una serie de agentes/actores cuyas características y comportamientos determinan la evolución del brote epidémico. Así, entre dichos actores podemos destacar los siguientes

(1) El propio especimen de código malicioso: sus características definirán tanto el proceso de propagación por la red (determinando los coeficientes epidemiológicos

asociados a dicho proceso y, en consecuencia, el término de la incidencia) como el proceso de infección en el dispositivo huésped (determinando los posibles estados epidemiológicos de los dispositivos y caracterizando los coeficientes epidemiológicos asociados -recuperación, transición entre estados-).

(2) Los dispositivos que conforman la red: sus especificaciones en cuanto a movilidad, transmisión y seguridad determinarán, entre otras cosas, la topología de contactos (red compleja estática o dinámica) y ciertos coeficientes epidemiológicos como el asociado a la resistencia frente al contagio, a la recuperación, etc.

(3) Los usuarios de dichos dispositivos: sus características relacionadas con la movilidad y la concienciación (preocupación) con la seguridad influirán en la determinación de la topología de contactos (al igual que en el caso anterior, si viene descrita por una red compleja estática o dinámica)

(4) La red de dispositivos: su estructura y especificaciones definirán el tipo de red compleja que describe la topología de contactos, los protocolos de comunicación establecidos, y los protocolos de seguridad implementados, etc.

La descripción matemática de este fenómeno se puede realizar a nivel local, considerando algunos de dichos agentes (usualmente los dispositivos) como protagonistas y estudiando la evolución de los estados epidemiológicos particulares de los mismos. Este tipo de modelos son los llamados modelos de naturaleza individual y suelen seguir en su diseño un enfoque "Bottom-Up". Por otro lado, se puede describir la propagación a nivel global de manera que no se estudien los agentes (dispositivos) propiamende dichos sino estructuras (colectividades de agentes) más grandes formadas por dispositivos que comparten un mismo estado epidemiológico. Así el objetivo de estos modelos es simular la evolución de una (o varias) características asociadas a dichas estructuras: el tamaño o densidad de las mismas habitualmente. Estos son los denominados modelos globales y, obviamente, fundamentan su diseño en un enfoque "Top-down".

Al igual que ocurre en la Epidemiología Matemática clásica, todos ellos (o la inmensa mayoría) son modelos compartimentales, esto es, la población de agentes estudiados se divide en distintas clases o grupos -compartimentos- atendiendo a sus características epidemiológicas. Como hemos dicho, lo habitual es que los modelos se centren en los dispositivos, con lo que estos se pueden clasificar en susceptibles S (libres de la infección pero no inmunes a ella), expuestos o latentes L (infectados por un especimen de malware que no está activo -ni realiza su actividad maliciosa en el dispositivo huésped, ni trata de propagarse-), infecciosos I (infectados en los que el malware tiene capacidad e intención de propagarse), atacados A (infectados en los que el malware está realizando su actividad maliciosa), recuperados R (libres de malware e inmunizados contra dicho especimen), dañados D (infectados que han sido incapacitados por acción del malware), etc. En el caso de los modelos globales, estos compartimentos se corresponden con las colectividades anteriormente mencionadas de manera que se caracterizan por simular la evolución del tamaño/densidad de dichos compartimentos. En los modelos individuales estos compartimentos determinan los posibles estados en los que se pueden encontrar los dispositivos en cada instante de tiempo.

Dependiendo de los compartimentos existentes y de la dinámica entre ellos, los modelos compartimentales pueden clasificarse en muy diversos tipos: modelos SI (los dispositivos susceptibles pasan a infecciosos de manera que el malware queda "acantonado" en ellos de manera indefinida), modelos SIS (los dispositivos susceptibles se infectan y permanecen temporalmente en dicho estado hasta que el malware desaparece y puden volver a ser infectados), modelos SIR como el de Kermack y McKendrick introducido en la Sect. 2 (los dispositivos susceptibles se infectan de manera que el malware puede ser eliminado, adquiriendo los mismos inmunidad permanente frente a nuevas infecciones por el mismo espécimen de código malicioso), etc.

Como ya se ha comentado el objetivo de los modelos matemáticos de propagación de malware es simular la diseminación del código malicioso en una red calculando o bien la cantidad de dispositivos que se encuentran en cada estado epidemiológico, o bien el propio estado epidemiológico de cada uno de los dispositivos que forman la red. En el primer caso las variables dependientes del modelo serán los tamaños (o densidades) de los compartimentos, mientras que en el segundo representarán el estado de cada individuo (el compartimento al que pertenece).

Lo habitual es diseñar modelos (ya sean globales o individuales) que estudien la diseminación temporal del malware [24, 27] de manera que las anteriores variables dependientes dependan del tiempo t. También han aparecido modelos (fundamentalmente globales) cuyo objetivo es simular la propagación espacio-temporal de manera que dichas variables dependerán de una variable independiente temporal t y una o dos variables independientes espaciales: x e y [18, 51].

La descripción de la evolución de los estados (ya sea temporal o espacio-temporal) se puede realizar de manera continua o discreta en las variables independientes: evolución temporal a pasos discretos de tiempo o de forma continua, evolución espacial en un número finito de porciones del espacio (teselación del espacio de interés) o de manera continua. De esta manera podemos clasificar los modelos en contínuos y discretos [25] y ello determinará el tipo de técnicas matemáticas a emplear: ecuaciones diferenciales, ecuaciones en recurrencias [33], etc.

Finalmente, y dependiendo de las herramientas matemáticas empleadas para describir la dinámica del proceso epidemiológico, podemos tener modelos deterministas (ante las mismas condiciones iniciales y valores de los coeficientes epidemiológicos la simulación siempre es la misma) o modelos estocásticos. Los primeros suelen hacer uso de ecuaciones diferenciales [49], mientras que los segundos se basan habitualmente en ecuaciones diferenciales estocásticas [34] o cadenas de Markov [6].

4. Modelización clásica: modelos de naturaleza global

Como se ha comentado y siguiendo la tradición de la Epidemiología Matemática clásica, el diseño de modelos matemáticos (de naturaleza global) para simular la propagación de malware se ha centrado fundamentalmente en el estudio de la evolución

temporal del tamaño de los compartimentos en que se dividen los individuos, con lo que en los modelos se considera una única variable independiente: el tiempo t, y tantas variables dependientes como compartimentos: $x_1(t), x_2(t), \ldots$. Esto no quiere decir que no se pueda tener en cuenta en los mismos la influencia de la disposición espacial de los individuos como en el caso de los metapopulation models [30]. En mucha menor medida se ha estudiado la propagación espacio-temporal de un determinado espécimen de malware, consistiendo dicho estudio en la determinación del tamaño de los compartimentos -variables dependientes- en función de dos o tres variables independientes: el tiempo t, y una o dos variables espaciales: x and y. En este caso las técnicas matemáticas utilizadas van desde cellular automata using a lattice-based arrangement of sites [4] to partial differential equations [35].

En lo que sigue describiremos detalladamente el proceso de diseño de modelos globales tanto para el caso de la simulación temporal como el caso de la simulación espacio-temporal. Se ilustrará este desarrollo mediante sendos ejemplos de modelos que han aparecido en la literatura científica.

4.1. Estudio de la evolución temporal del malware usando modelos globales

4.1.1. Fundamentos matemáticos

Como hemos dicho, la Epidemiología Matemática moderna (y, por tanto, el diseño teórico de modelos de propagación de malware) se cimentó sobre el modelo de Kermack y McKendrick. De esta manera la expresión general para los modelos de naturaleza global es como sigue: supongamos que la población de dispositivos se puede clasificar en $n + m$ compartimentos, $X_1, \ldots, X_{n+m}$, de manera que los n primeros compartimentos, $X_1, \ldots, X_n$, caracterizan diferentes tipos de dispositivos infectados, mientras que los m restantes compartimentos clasifican a los dispositivos que no se encuentran infectados: $X_{n+1}, \ldots, X_{n+m}$. Supongamos que $x_i(t) \in [0, 1]$ representa la densidad de individuos del compartimento i-ésimo, $1 \leq i \leq n + m$, en el instante de tiempo t, de manera que $x(t) = (x_1(t), \ldots, x_{n+m}(t))$. Entonces la dinámica del modelo se describe matemáticamente de la siguiente manera:

$$\begin{cases} x_1'(t) = \mathcal{F}_1\left(x(t)\right) + \mathcal{V}_1\left(x(t)\right), \\ \vdots \\ x_{n+m}'(t) = \mathcal{F}_{n+m}\left(x(t)\right) + \mathcal{V}_{n+m}\left(x(t)\right), \end{cases} \tag{23}$$

donde $\mathcal{F}_i\left(x(t)\right)$ representa la tasa a la que la incidencia modifica el tamaño del compartimento i-ésimo, y $\mathcal{V}_i\left(x(t)\right)$ se corresponde con la tasa a la que o bien la progresión de la infección en el huesped, o bien la propia dinámica del proceso de propagación va haciendo cambiar de estado epidemiológico de dicho dispositivo huésped. Así, podemos escribir

$$\mathcal{V}_i\left(x(t)\right) = \mathcal{V}_i^+\left(x(t)\right) - \mathcal{V}_i^-\left(x(t)\right) \tag{24}$$

donde $\mathcal{V}_i^+\left(x(t)\right)$ y $\mathcal{V}_i^-\left(x(t)\right)$ representan respectivamente el flujo de entrada y el flujo de salida en el compartimento i-ésimo. Estas funciones deben poseen las siguientes características:

(1) $\mathcal{F}_i\left(x(t)\right) > 0$ if $1 \leq i \leq n$, and $\mathcal{F}_i\left(x(t)\right) = 0$ if $n+1 \leq i \leq n+m$.
(2) $\mathcal{V}_i^+\left(x(t)\right) \geq 0, \mathcal{V}_i^-\left(x(t)\right) \geq 0$ for $1 \leq i \leq n+m$. In fact, if $x_i(t) = 0$ then $\mathcal{V}_i^-\left(x(t)\right) = 0$.
(3) If $x_j(t) = 0$ with $1 \leq j \leq n$, then $\mathcal{F}_i\left(x(t)\right) = 0$, and $\mathcal{V}_i^+\left(x(t)\right) = 0$ for $1 \leq i \leq n$.

Tanto en el análisis cualitativo de estos modelos como en el estudio de las estrategias más eficientes de control de los procesos epidémicos que simulan, juega un papel extremadamente importante el denominado número reproductivo básico $\mathcal{R}_0$. Grosso modo este parámetro umbral representa el número de contagios directos producidos por un único individuo infeccioso a lo largo de todo su periodo infeccioso en una población enteramente susceptible. Consecuentemente si $\mathcal{R}_0 > 1$ entonces el número de dispositivos infectados tenderá a crecer, mientras que si $\mathcal{R}_0 < 1$ el número de dispositivos infectados tenderá a decrecer. Este concepto fue formalizado por Van den Driessche y Watmough en [13], y O. Dieckmann, J. Heesterbeek y J. Metz propusieron el llamado *Next Generation Method* [11] para calcularlo explícitamente en el caso de modelos deterministas globales y compartimentales como el dado en (23). Concretamente el $\mathcal{R}_0$ es el radio espectral de la matriz de siguiente generación $\left(\mathcal{F}\cdot\mathcal{V}^{-1}\right)_{P_0^*}$, donde $P_0^* = \left(x_{1,0}^*, \ldots, x_{n+m,0}^*\right)$ es el punto de equilibrio del sistema libre de infección (es decir, no existen dispositivos infecciosos -que puedan transmitir el código malicioso-) y

$$\mathcal{F} = \begin{pmatrix} \frac{\partial \mathcal{F}_1}{\partial x_1} & \cdots & \frac{\partial \mathcal{F}_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{F}_n}{\partial x_1} & \cdots & \frac{\partial \mathcal{F}_n}{\partial x_n} \end{pmatrix}, \quad \mathcal{V} = \begin{pmatrix} \frac{\partial \mathcal{V}_1}{\partial x_1} & \cdots & \frac{\partial \mathcal{V}_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{V}_n}{\partial x_1} & \cdots & \frac{\partial \mathcal{V}_n}{\partial x_n} \end{pmatrix}. \tag{25}$$

Como se ha comentado, este coeficiente juega un papel transcendental en el análisis de la dinámica de los modelos globales y en el estudio del comportamiento de sus puntos de equilibrio [10]. Recordemos que los puntos de equilibrio son las soluciones del sistema:

$$0 = \mathcal{F}_i\left(x(t)\right) + \mathcal{V}_i\left(x(t)\right), \quad 1 \leq i \leq n+m, \tag{26}$$

y habitualmente obtenemos el ya mencionado punto de equilibrio libre de infección P_0^*, y el punto de equilibrio endémico $P_e^* = \left(x_{1,e}^*, \ldots, x_{n+m,e}^*\right)$ caracterizado por representar un estado en el que siempre existen dispositivos infecciosos. El estudio cualitativo del sistema (23) nos indica que P_0^* es local y asintóticamente estable cuando $\mathcal{R}_0 < 1$, mientras que P_e^* es estable cuando $\mathcal{R}_0 > 1$.

En la siguiente subsección ilustraremos este desarrollo teórico mediante un modelo global que ha aparecido muy recientemente en la literatura científica.

4.1.2. Un ejemplo ilustrativo: el modelo de V. Madhusudanan *et al.*

Muy recientemente V. Madhusudanan *et al.* han propuesto un novedoso modelo para estudiar la difusión de un gusano computacional en una red inalámbrica de naturaleza IoT [37]. En concreto se trata de un modelo compartimental SEIRV donde los dispositivos que forman la red se clasifican en cuatro compartimentos: Susceptibles ($X_1 = S$), Expuestos ($X_2 = E$), Infecciosos ($X_3 = I$), Recuperados ($X_4 = R$) y Vacunados ($X_5 = V$). Específicamente se trata de un modelo con retardo con incidencia no lineal, y dotado de dinámica poblacional: en cada instante de tiempo aparecen dispositivos con densidad A y desaparecen dispositivos de la red siguiendo una tasa uniforme δ_0 con independencia del estado epidemiológico del dispositivo. Los dispositivos susceptibles se infectan y pasan a ser dispositivos expuestos según una tasa α y una incidencia no lineal . Los dispositivos expuestos pasan a ser infecciosos con tasa δ_1 sobre los dispositivos expuestos τ unidades de tiempo antes. Los dispositivos infecciosos se recuperan o bien cuando el malware los abandona con tasa δ_2 (sin intervención de las medidas de seguridad), o bien cuando son intervenidos y el software de seguridad es capaz de eliminar el malware con tasa β. Finalmente, los dispositivos susceptibles pueden ser "vacunados" (se instala en ellos los parches de seguridad necesarios para que no se puedan infectar) segun una tasa μ de manera que esto les confiere una inmunidad temporal que pierden según una tasa η. En la Fig. 3 se ilustra el diagrama de flujo de esta dinámica que viene descrita por el siguiente sistema de ecuaciones ordinarias con retardo:

$$\begin{cases} S'(t) = A - \delta_0 S(t) - \frac{\alpha S(t)^2 I(t)}{S(t)^2 + cI(t)^2} + \eta V(t) - \mu S(t) \\ E'(t) = \frac{\alpha S(t)^2 I(t)}{S(t)^2 + cI(t)^2} - \delta_0 E(t) - \delta_1 E\left(t-\tau\right) \\ I'(t) = \delta_1 E\left(t-\tau\right) - \left(\delta_0 + \delta_2 + \delta_3\right) I(t) - \frac{\beta I(t)}{I(t)+a} \\ R'(t) = \delta_2 I(t) - \delta_0 R(t) + \frac{\beta I(t)}{I(t)+a} \\ V'(t) = \mu S(t) - \left(\delta_0 + \eta\right) V(t) \end{cases} \tag{27}$$

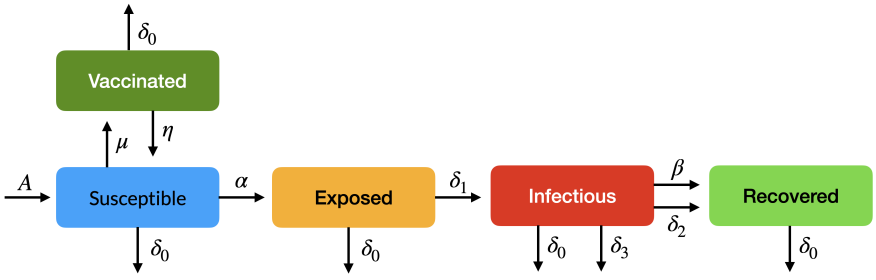

Figura 3 Diagrama de flujo representativo del modelo epidemiológico propuesto por Madhusudanan *et al.*.

Dado que el compartimento de los recuperados no interviene en la variación del resto de compartimentos, podemos simplifica el sistema anterior quedándonos con el siguiente:

$$\begin{cases} S'(t) = A - \delta_0 S(t) - \frac{\alpha S(t)^2 I(t)}{S(t)^2 + cI(t)^2} + \eta V(t) - \mu S(t) \\ E'(t) = \frac{\alpha S(t)^2 I(t)}{S(t)^2 + cI(t)^2} - \delta_0 E(t) - \delta_1 E\left(t-\tau\right) \\ I'(t) = \delta_1 E\left(t-\tau\right) - \left(\delta_0 + \delta_2 + \delta_3\right) I(t) - \frac{\beta I(t)}{I(t)+a} \\ V'(t) = \mu S(t) - \left(\delta_0 + \eta\right) V(t) \end{cases} \tag{28}$$

donde se supone que $S(t)+E(t)+I(t)+V(t) = N$, de manera que un sencillo cálculo nos muestra la siguiente acotación para el número total de dispositivos:

$$N(t) \leq N_0 e^{-\delta_0 t} + \frac{A}{\delta_0}\left(1 - e^{-\delta_0 t}\right). \tag{29}$$

Consecuentemente se toma como condición inicial $N(0) = \frac{A}{\delta_0}$ tal que la región factible $\Omega = \{(S,E,I,V) \in \mathbb{R}_+^4 \colon 0 \leq S+E+I+V \leq N(0)\}$ es una región positivamente invariante. Un sencillo cálculo nos muestra que este modelo tiene dos puntos de equilibrio: el punto de equilibrio libre de infección:

$$P_0^* = \left(S_0^*, E_0^*, I_0^*, V_0^*\right) = \left(\frac{A\left(\delta_0+\eta\right)}{\delta_0\left(\delta_0+\eta+\mu\right)}, 0, 0, \frac{A\mu}{\delta_0\left(\delta_0+\eta+\mu\right)}\right), \tag{30}$$

y en el punto de equilibrio endémico:

$$P_e^* = \left(S_e^*, E_e^*, I_e^*, V_e^*\right), \tag{31}$$

donde:

$$S_e^* = I_e^* \sqrt{\frac{c\left(\delta_0+\delta_1\right)\left[\left(\delta_0+\delta_2+\delta_3\right)\left(I_e^*+a\right)+\beta\right]}{\left[\alpha\delta_1 - \left(\delta_0+\delta_1\right)\left(\delta_0+\delta_2+\delta_3\right)\right]\left(I_e^*+a\right) - \beta\left(\delta_0+\delta_1\right)}}, \tag{32}$$

$$E_e^* = \frac{\left(\delta_0+\delta_1+\delta_3\right) I_e^*}{\delta_1} + \frac{\beta I_e^*}{\delta_1\left(I_e^*+a\right)}, \tag{33}$$

$$V_e^* = \frac{\mu S_e^*}{\delta_0+\eta}, \tag{34}$$

siendo I_e^* la raíz real positiva de la ecuación $0 = \sum_{i=0}^{6} \Gamma_i x^{6-i}$, donde las expresiones explícitas de los coeficientes son las siguientes:

$$\Gamma_0 = c\alpha^2\delta_1^2k^2l^2 - c\alpha\delta_1k^3l^3 \tag{35}$$

$$\Gamma_1 = ac\alpha^2\delta_1^2k^2l^2 - ac\alpha\delta_1k^3l^3 - c\alpha\beta\delta_1k^3l^2 + 2ac\alpha^2\delta_1^2k^2 - 2ac\alpha\delta_1k^3, \tag{36}$$

$$\begin{aligned}\Gamma_2 = {} & A^2c\alpha^2\delta_1^2 - A^2c\alpha\delta_1 + c\alpha^2\delta_1^2k^2l^2a^2 - c\alpha\delta_1k^3l^3a^2 + 2c\alpha^2\delta_1^2a^2k^2 \\ & -2c\alpha\delta_1a^2k^3l - 2ac\alpha\delta_1k^3\beta - m^2c^2\alpha\delta_1kl,\end{aligned} \tag{37}$$

$$\begin{aligned}\Gamma_3 = {} & A^2ac\alpha^2\delta_1^2 - A^2ackl\alpha\delta_1 + 2aA^2c\alpha^2\delta_1^2 - 2aA^2c\alpha\delta_1kl \\ & +ca^3\alpha^2\delta_1^2k^2l^2 - ca^3\alpha\delta_1k^3l^3 - c\alpha\beta\delta_1k^3l^2a^2 - 2Akl\alpha^2{\delta_1}^2 \\ & -2Ak^3l^3\alpha\delta_1 + 4A\alpha\delta_1k^2l^2 - m^2c^2\alpha\delta_1kal - m^2c^2\alpha\beta\delta_1k,\end{aligned} \tag{38}$$

$$\begin{aligned}\Gamma_4 = {} & A^2a^2c\alpha^2\delta_1^2 - A^2a^2c\alpha\delta_1kl - 2A^2a^2c\alpha\delta_1kl - 2A^2ack\alpha\beta\delta_1 \\ & -4aAkl\alpha^2\delta_1^2 + 4aAk^3l^3 + 8Aa\alpha\delta_1k^2l^2 + 4A\alpha\beta k^2l\delta_1 \\ & -4A\beta k^3l^2 + 2aAk^3l^3 - 4A\alpha\delta_1k^2l^2 + 2Ak\beta\alpha^2\delta_1^2 + 2A\beta k^3l^2 \\ & -4A\beta\alpha\delta_1k^2l,\end{aligned} \tag{39}$$

$$\begin{aligned}\Gamma_5 = {} & A^2a^3c\alpha^2\delta_1^2 - A^2a^3klc\alpha\delta_1 - A^2c\alpha\delta_1a^2\beta k - 2Akla^2\alpha^2\delta_1^2 \\ & -2Ak^3l^3a^2 - 2\beta^2Ak^3l + 4Ak^2l^2a^2\alpha\delta_1 - 4Ak^3l^2a + 4aA\beta k^2l\alpha\delta_1 \\ & +4Akl\alpha^2\delta_1^2a^2 + 4Aa^2k^3l^3 + 4Aa\beta k^3l^2 + 4Aak\beta\alpha^2\delta_1^2 - 4Aak^2l\alpha\beta\delta_1 \\ & +4A\beta ak^3l^2 - 8Aa\alpha\delta_1k^2\beta l + 4A\beta^2k^3l - 4A\beta^2k^2\alpha\delta_1,\end{aligned} \tag{40}$$

$$\begin{aligned}\Gamma_6 = {} & 2a^2A^2c\alpha^2\delta_1^2 + 2Akl\alpha^2\delta_1^2a^3 + 2Ak^3l^3a^3 + 2Ak^3l\beta^2a \\ & -4\alpha\delta_1Ak^2l^2a^3 - 4\alpha\delta_1k^2l^2a^2 + 4aA\beta k^3l^2 - 4A\beta k^2la^2\alpha\delta_1 \\ & +2Ak\beta\alpha^2\delta_1^2a^2 + 2Ak^3l^2\beta a^2 + 2Ak^3\beta^3 - 4A\alpha\delta_1k^2l\beta a + 4Ak^3\beta^2la \\ & -4A\beta k^2\alpha\delta_1a,\end{aligned} \tag{41}$$

$$k = \delta_0 + \delta_1, \tag{42}$$

$$l = \delta_0 + \delta_2 + \delta_3, \tag{43}$$

$$m = \frac{\delta_0^2\left(\delta_0 + \eta + \mu\right)^2}{\left(\delta_0 + \eta\right)^2}. \tag{44}$$

Los autores demuestran que el punto de equilibrio endémico es local y asintóticamente estable si $0 \leq \tau < \tau_0$ para cierto τ_0, y el sistema experimenta una bifurcación de Hopf alrededor de P_e^* en $\tau = \tau_0$ [37].

En la Figura 4 se muestran dos simulaciones numéricas para ilustrar el modelo. Específicamente, en la Figura 4-(a) se muestra la evolución temporal de los diferentes compartimentos cuando los coeficientes considerados son los siguientes: $A = 2, \alpha = 0{,}27, c = 0{,}01, \eta = 0{,}2, \mu = 0{,}003, \delta_0 = 0{,}02, \delta_1 = 0{,}2, \delta_2 = 0{,}045, \delta_3 = 0{,}03, a = 1, \beta = 0{,}05$. Además, $s(0) = 99, e(0) = 0, i(0) = 1, r(0) = 0$, y $v(0) = 0$. Obsérvese que en este caso el sistema evoluciona hacia un punto de equilibrio endémico: $P_e^* \approx (1{,}49, 8{,}95, 18{,}36, 43{,}60, 0{,}02)$. Por otro lado, en la Figura 4-(b) se muestra la evolución temporal de los compartimentos de dispositivos expuestos e infecciosos cuando $\alpha = 0{,}1$ (en lugar de $\alpha = 0{,}27$). En este caso, se alcanza el punto de equilibrio libre de infección $P_0^* \approx (98{,}34, 0, 0, 0{,}23, 1{,}34)$.

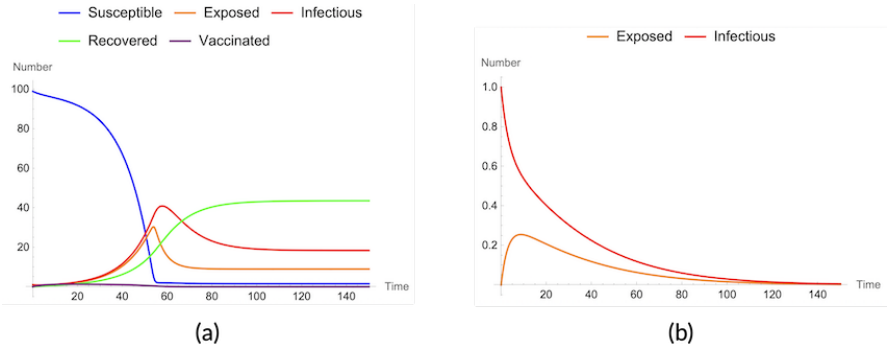

Figura 4 (a) Evolución temporal de los compartimentos alcanzándose el punto de equilibrio endémico. (b) Evolución temporal de los compartimentos de expuestos e infecciosos cuando se tiende hacia el punto de equilibrio libre de infección.

4.2. Propagación espacio-temporal del malware

4.2.1. Fundamentos matemáticos

Como ya se ha comentado previamente, la inmensa mayoría de los modelos simulan la evolución temporal de la propagación de malware en una determinada red. Ahora bien, en determinadas ocasiones puede resultar interesante estudiar desde un punto de vista global la dinámica espacio-temporal del fenómeno. En esta situación hemos de distinguir dos casos: (1) los dispositivos o los usuarios que portan los dispositivos (que hospedan el espécimen de código malicioso) se encuentran en movimiento; y (2) los dispositivos permanence inmóviles.

En el primer caso la diseminación del malware entre la población de dispositivos viene determinada tanto por el propio movimiento de los dispositivos (que hace que la topología de contactos vaya cambiando con el tiempo) como por el propio proceso de propagación del malware (que hace que éste se intente propagar desde el nodo huesped hacia aquellos otros nodos que se encuentran a su alcance -y que están definidos por la topología de contactos anteriormente mecionada-). Por otra parte, cuando los dispositivos permanen inmóviles la evolución espacial viene sólo condicionada por el propio proceso de propagación del código malicioso. Consecuentemente, sin pérdida de generalidad, podemos suponer que el "medioambiente" en el que existen tanto los dispositivos como el malware es una región del plano, $\Omega \subseteq \mathbb{R}^2$. Entonces, si nos interesa conocer la cantidad (o densidad) de dispositivos que se encuentran en un determinado estado epidemiológico (es decir, el tamaño de los compartimentos) en un instante de tiempo y en una determinada posición, las variables dependientes del modelo (global) deben ser de la forma $X_i\,(t, x, y)$, con $1 \leq i \leq n + m$.

Las variables independientes (tanto la temporal como las espaciales) pueden evolucionar de manera discreta o continua. Si la evolución temporal es discreta la dinámica del modelo podría venir descrita por ecuaciones en diferencias, mientras

que si la evolución temporal es continua se podrían utilizar ecuaciones en derivadas parciales -si la propagación espacial se comporta de manera continua también- o técnicas híbridas (metapopulation models, for example) si la propagación espacial es discreta.

La estructura matemática de los modelos de propagación espacio-temporal basados en ecuaciones en derivadas parciales (tiempo y espacios continuos) es la siguiente:

$$\begin{aligned} &\frac{\partial x_i}{\partial t} = d\nabla^2 x_i + \mathcal{F}_i\left(x(t)\right) + \mathcal{V}_i\left(x(t)\right), \quad 1 \leq i \leq n+m, \\ &(x,y) \in \Omega, t \geq 0. \end{aligned} \tag{45}$$

donde seguimos la nomenclatura y estructura introducidas en la subsección 4.1.1, ∇^2 denota el Laplaciano, $d\nabla^2 x_i$ es el término de la difusión asociado al compartimento de dispositivos i-ésimo, y habría que establecer las oportunas condiciones iniciales y de contorno.

Por otra parte, en los modelos híbridos (tiempo continuo y espacio discreto), la región de interés está, por ejemplo, teselada de manera uniforme según un conjunto finito de células: $\Omega = \bigsqcup_{1 \leq a \leq L_0, 1 \leq b \leq L_1} C_{ab}$. Entonces el modelo se plantea según las siguientes ecuaciones:

$$\begin{aligned} &x_i'\left(t,a,b\right) = \mathcal{F}_i\left(x(t,a,b), x(t,\alpha,\beta)\right) + \mathcal{V}_i\left(x(t,a,b)\right), \\ &1 \leq a, \alpha \leq L_0, 1 \leq b, \beta \leq L_1, \\ &1 \leq i \leq n+m, \end{aligned} \tag{46}$$

donde $C_{\alpha,\beta}$ es una célula adyacente a la célula $C_{a,b}$, de manera que los nodos infecciosos de dicha célula adyacente pudieran infectar a nodos susceptibles emplazados en la célula considerada $C_{a,b}$.

En lo que sigue veremos un par de ejemplos ilustrativos de estas dos aproximaciones

4.2.2. El primer ejemplo ilustrativo: el modelo propuesto por Zhu *et al.*

L. Zhu, H. Zhao y X. Wang propusieron en 2015 un modelo de reacción-difusión para simular la propagación espacio-temporal de malware en redes de sensores inalámbricos móviles (MWSNs) [53]. Específicamente, se trata de un modelo compartimental SIR con retardo cuya dinámica se describe mediante el siguiente sistema de ecuaciones en derivadas parciales:

$$\begin{cases} \frac{\partial S}{\partial t} = d\nabla^2 S + rS\left(1 - \frac{1}{k}S\right) - \beta SI(t-\tau) - \epsilon_1 S - \eta S \\ \frac{\partial I}{\partial t} = d\nabla^2 I + \beta SI(t-\tau) - \epsilon_2 I - \eta I \\ \frac{\partial R}{\partial t} = d\nabla^2 R + \epsilon_1 S - \eta R \end{cases} \tag{47}$$

$$(x, y) \in \Omega = (0, L) \times (0, L), t > 0, \tag{48}$$

$$\frac{\partial S}{\partial \phi} = \frac{\partial I}{\partial \phi} = \frac{\partial R}{\partial \phi} = 0, \quad x, y = 0, L, \quad t \geq 0. \tag{49}$$

dotado de las siguientes condiciones iniciales:

$$S(t, x, y) = \psi_1(t, x, y) \geq 0, \tag{50}$$

$$I(t, x, y) = \psi_2(t, x, y) \geq 0, \tag{51}$$

$$R(t, x, y) = \psi_3(t, x, y) \geq 0, \tag{52}$$

$$(t, x, y) \in [-\tau, 0] \times [0, L] \times [0, L]. \tag{53}$$

En (47) el coeficiente $d > 0$ representa la capacidad de difusión de los nodos móviles, $k > 0$ es la capacidad de carga más grande de los nodos, $r > 0$ denota la tasa de crecimiento intrínseco de los nodos susceptibles, ϵ_1 es la probabilidad con la que un nodo susceptible se recupera, η representa la tasa de desaparición de la red, y ϵ_2 representa la probabilidad con la que un nodo infeccioso se recupera. Además, las condiciones de frontera de Neumann, $\frac{\partial S}{\partial \phi} = \frac{\partial I}{\partial \phi} = \frac{\partial R}{\partial \phi} = 0$, establecen que no hay comunicación entre nodos a través de la frontera de la región de interés (ϕ es el vector normal exterior a $\partial\Omega$). Finalmente, las densidades iniciales de nodos susceptibles, infecciosos y recuperados vienen dadas, respectivamente, por ψ_1, ψ_2, y ψ_3.

Obsérvese que como las dos primeras ecuaciones del sistema (47) no dependen de la variable $R(t, x, y)$, se puede considerar solo dichas dos primeras ecuaciones del sistema para describir la dinámica del mismo:

$$\begin{cases} \frac{\partial S}{\partial t} = d\nabla^2 S + rS\left(1 - \frac{1}{k}S\right) - \beta SI(t-\tau) - \epsilon_1 S - \eta S \\ \frac{\partial I}{\partial t} = d\nabla^2 I + \beta SI(t-\tau) - \epsilon_2 I - \eta I \end{cases} \tag{54}$$

$$(x, y) \in (0, L) \times (0, L), t > 0, \tag{55}$$

$$\frac{\partial S}{\partial \phi} = \frac{\partial I}{\partial \phi} = 0, \quad x, y = 0, L, \quad t \geq 0. \tag{56}$$

Se demuestra que este sistema presenta tres puntos de equilibrio:

- El trivial $P^* = (S^*, I^*) = (0, 0)$ que existe siempre.
- El punto de equilibrio libre de infección:

$$P_0^* = \left(S_0^*, I_0^*\right) = \left(\frac{k(r - \eta - \epsilon_1)}{r}, 0\right), \tag{57}$$

que existe si $r > \eta + \epsilon_1$.

- El punto de equilibrio endémico:

$$P_e^* = \left(S_e^*, I_e^*\right) = \left(\frac{\eta + \epsilon_2}{\beta}, \frac{kr\beta - k\beta\eta - k\beta\epsilon_1 - r\eta - r\epsilon_2}{k\beta^2}\right), \tag{58}$$

que existe si $k\beta\left(r - \eta - \epsilon_1\right) - 3r\left(\eta + \epsilon_2\right) > 0$.

Además, se obtienen los siguientes resultados:

Teorema 3 *El punto de equilibrio libre de infección P_0^* es local y asintóticamente estable para $\tau \geq 0$ si se satisfacen las siguiente condiciones:*

$$0 < r\left(r + \epsilon_2 - \epsilon_1\right) + k\beta\left(\eta + \epsilon_1 - r\right), \tag{59}$$

$$0 < k\beta\left(\eta + \epsilon_1 - r\right) + r\left(\eta + \epsilon_2\right). \tag{60}$$

Teorema 4 *El punto de equilibrio endémico P_e^* es local y asintóticamente estable para $\tau \in [0, \tau_0^0)$ si se satisfacen las siguientes condiciones:*

$$0 < r\left(\eta + \epsilon_2\right) + k\beta\left(\epsilon_1 - \epsilon_2\right), \tag{61}$$

$$0 < r^2\left(\eta + \epsilon_2\right) + k\beta\left(\epsilon_1 - \epsilon_2\right)\left(2r + k\beta\right), \tag{62}$$

$$0 < d^2 + k^2\beta^2\left(\eta + \epsilon_1 - r\right)\left(\eta + \epsilon_2\right). \tag{63}$$

4.2.3. El segundo ejemplo ilustrativo: el modelo propuesto por Zhang *et al.*

Zhang *et al.* propusieron el llamado modelo MDBCA para estudiar la propagación de malware en redes de sensores inalámbricos [52]. Con el fin de simular la evolución espacio-temporal, este modelo combina autómatas celulares bidimensionales con sistemas de ecuaciones diferenciales. Se trata de un modelo compartimental en el que la población de nodos sensores se clasifica en cinco compartimentos diferentes: susceptibles S, expuestos E, infecciosos I, recuperados R, y nodos dañados D. Por lo tanto, puede considerarse como un modelo compartimental SEIRD cuya dinámica se describe de la siguiente manera: un nodo sensor susceptible se vuelve expuesto o infeccioso según tasas $0 < \lambda \leq 1$, y $0 < \tau \leq 1$ respectivamente, de modo que $0 < \lambda + \tau \leq 1$. Los sensores expuestos se vuelven infecciosos según una tasa $0 < \sigma \leq 1$ y los nodos infecciosos se recuperan siguiendo una tasa $0 < \epsilon \leq 1$. Además, el espécimen de malware es capaz de dañar los nodos sensores; en este sentido, se asume que un sensor expuesto o infeccioso se daña según una tasa $0 \leq \mu \leq 1$. De manera similar, el consumo de energía puede dejar inutilizados a los dispositivos susceptibles y recuperados con tasa $0 < \mu \leq 1$. Finalmente, todos los nodos sensores dañados son reemplazados en la WSN, es decir, la población permanece constante a lo largo del tiempo.

Se supone que los nodos sensores están desplegados en una retícula bidimensional constituida por $L \times L$ celdas cuadradas idénticas, lo que determina el espacio celular $C = \{C_{ab}, 0 \leq a, b \leq L\}$ de un autómata celular. El despliegue se realiza

aleatoriamente, de modo que en cada celda C_{ab} habrá N_{ab} nodos sensores. Sea $0 \leq S^t_{ab} \leq 1$, $0 \leq E^t_{ab} \leq 1$, $0 \leq I^t_{ab} \leq 1$, $0 \leq R^t_{ab} \leq 1$, y $0 \leq D^t_{ab} \leq 1$ las fracciones de nodos susceptibles, expuestos, infecciosos, recuperados y dañados ubicados en la celda C_{ab} en el paso de tiempo t, entonces $S^t_{ab} + L^t_{ab} + I^t_{ab} + R^t_{ab} + D^t_{ab} = 1$.

El conjunto de estados del autómata celular es $\mathcal{S} = [0,1] \times \overset{(5.}{\ldots} \times [0,1]$ y, en consecuencia, el estado de la celda $C_{ab} \in C$ está definido por $Q^t_{ab} = \left(S^t_{ab}, L^t_{ab}, I^t_{ab}, R^t_{ab}, D^t_{ab}\right) \in \mathcal{S}$. Además, la vecindad de $C_{ab} \in C$ es

$$\mathcal{V}ab = \{Cij \in C : |a-i| \leq r, |b-j| \leq r\} = \{C_{i_1 j_1}, \ldots, C_{i_\alpha j_\alpha}\} \subseteq C, \tag{64}$$

donde r representa el radio de comunicación (constante) de las celdas.

La función de transición local que rige la dinámica del autómata celular se define como $Q^{t+1}_{ab} = \mathcal{F}\left(Q^t_{i_1 j_1}, \ldots, Q^t_{i\alpha j_\alpha}\right)$ tal que:

$$\begin{aligned} S^{t+1}_{ab} &= S^t_{ab} - (\lambda+\tau)\, S^t_{ab} I^t_{ab} - (\lambda+\tau) \sum_{C_{ij} \in \mathcal{V}_{ab}} m \frac{N_{ij}}{N_{ab}} S^t_{ab} I^t_{ij} \\ &\quad +\mu\left(1 - S^t_{ab}\right), \end{aligned} \tag{65}$$

$$L^{t+1}_{ab} = L^t_{ab} + \left(\lambda S^t_{ab} I^t_{ab} + \lambda \sum_{C_{ij} \in \mathcal{V}_{ab}} m \frac{N_{ij}}{N_{ab}} S^t_{ab} I^t_{ij}\right) - (\sigma+\mu)\, L^t_{ab}, \tag{66}$$

$$\begin{aligned} I^{t+1}_{ab} &= I^t_{ab} + \left(\tau S^t_{ab} I^t_{ab} + \tau \sum_{C_{ij} \in \mathcal{V}_{ab}} m \frac{N_{ij}}{N_{ab}} S^t_{ab} I^t_{ij}\right) \\ &\quad +\sigma L^t_{ab} - \epsilon I^t_{ab} - \mu I^t_{ab}, \end{aligned} \tag{67}$$

$$R^{t+1}_{ab} = R^t_{ab} + \epsilon I^t_{ab} - \mu R^t_{ab}, \tag{68}$$

$$D^{t+1}_{ab} = D^t_{ab} + \mu\left(S^t_{ab} + L^t_{ab} + I^t_{ab} + R^t_{ab}\right), \tag{69}$$

donde $0 \leq m \leq 1$ representa el flujo de comunicación entre celdas adyacentes.

Suponiendo un despliegue constante y uniforme de nodos sensores sobre la región, es decir, $N_{ab} = N$ para todas las $C_{ab} \in C$, y considerando los dos primeros términos de la expansión de Taylor del sistema (65)-(69), la dinámica de propagación del malware en cada celda puede ser descrita por el siguiente sistema de ecuaciones diferenciales ordinarias:

$$S'_{ab}(t) = -(\lambda+\tau)\, S_{ab}(t) I_{ab}(t) - (\lambda+\tau) \sum_{C_{ij}\in\mathcal{V}_{ab}} m S_{ab}(t) I_{ij}(t) \tag{70}$$
$$+\mu\,(1 - S_{ab}(t))\,,$$
$$L'_{ab}(t) = \lambda S_{ab}(t) I_{ab}(t) + \lambda \sum_{C_{ij}\in\mathcal{V}_{ab}} m S_{ab}(t) I_{ij}(t) - (\sigma+\mu)\, L_{ab}(t), \tag{71}$$
$$I'_{ab}(t) = \tau S_{ab}(t) I_{ab}(t) + \tau \sum_{C_{ij}\in\mathcal{V}_{ab}} m S_{ab}(t) I_{ij}(t) \tag{72}$$
$$+\sigma L_{ab}(t) - \epsilon I_{ab}(t) - \mu I_{ab}(t),$$
$$R'_{ab}(t) = \epsilon I_{ab}(t) - \mu R_{ab}(t), \tag{73}$$
$$D'_{ab}(t) = \mu, \tag{74}$$

Obsérvese que este sistema se puede reducir ya que la ecuación (74) puede resolverse fácilmente y $D_{ab}(t)$ no afecta a las otras variables dependientes.

Se verifican los siguientes resultados:

Proposición 1 *Existen dos puntos de equilibrio asociados al sistema (70)-(73): el punto de equilibrio libre de infección* $P_0^* = \left(S_0^*, L_0^*, I_0^*, R_0^*\right) = (1,0,0,0)$*, y el punto de equilibrio endémico* $P_e^* = \left(S_e^*, L_e^*, I_e^*, R_e^*\right)$*, donde*

$$S_e^* = \frac{(\mu+\sigma)(\mu+\epsilon)}{\lambda\sigma + \tau(\mu+\sigma)}, \tag{75}$$
$$L_e^* = \frac{\lambda\,(\mu+\epsilon)}{\sigma\,(\lambda+\tau) + \mu\tau} I_e^*, \tag{76}$$
$$I_e^* = \mu \frac{\sigma(\lambda+\tau) - \mu^2 + \mu(\tau-\sigma) - \epsilon(\mu+\sigma)}{(\lambda+\tau)(\mu+\sigma)(\mu+\epsilon)}, \tag{77}$$
$$R_e^* = \frac{\epsilon}{\mu} I_e^*. \tag{78}$$

Proposición 2 *El número reproductivo básico asociado al sistema (70)-(73) es*

$$\mathcal{R}_0 = \frac{(1+m)\lambda\sigma + \tau\,(\mu+\sigma)}{(\epsilon+\mu)\,(\mu+\sigma)}. \tag{79}$$

Teorema 5 *El punto de equilibrio libre de infección* P_0^* *es local y asintóticamente estable si* $\mathcal{R}_0 < 1$*. El punto de equilibrio endémico* P_e^* *es local y asintóticamente estable si* $\mathcal{R}_0 > 1$*.*

5. Discusión

5.1. Análisis crítico de los modelos globales

Los dos modelos ilustrativos descritos en las anterior sección son ejemplos claros en donde se ponen de manifesto las posibles deficiencias que exhiben los modelos de naturaleza global. Sin lugar a ninguna duda se trata de modelos con una sólida fundamentación matemática que permite un completo análisis teórico de los mismos (fundamentalmente un estudio cualitativo, aunque se han realizado también análisis numéricos [26] y estadísticos [14]). Por regla general los autores, tras un proceso de modelización del fenómeno no justificado plenamente, se enfrentan al desafío del estudio de la estabilidad, tanto local como global, de los puntos de equilibrio asociados al sistema. Esta es una tarea muchas veces difícil y compleja y que, sin duda, repercute beneficiosamente en el entendimiento de la dinámica de la propagación en tanto en cuanto es posible extraer conclusiones relevantes sobre el comportamiento de la evolución del número de dispositivos en cada estado epidemiológico. No obstante, la sensación que subyace en el ambiente es que el desafío original consistente en desarrollar un modelo que simule de la manera más realista posible la diseminación de un determinado espécimen de código malicioso en una determinada red, ha sido sustituido por el desafío del estudio teórico/matemático de la estabilidad de un sistema de ecuaciones diferenciales, olvidándonos de la "aplicabilidad" del mismo.

Así, por ejemplo, el estudio matemático realizado en [37] es de gran complejidad y calidad pero se plantean preguntas relevantes en cuanto a la descripción matemática del fenómeno:

- ¿Cómo se tiene en cuenta que es un gusano computacional el agente que se está propagando?
- ¿Dónde se tiene en cuenta que el "medioambiente" por el que se propaga dicho gustado es una red IoT?
- ¿Cómo se definen/determinan los coeficientes epidemiológicos empleados?
- La heterogeneidad de las redes IoT hace que los dispositivos desplegados sean muy diferentes entre sí (actividad, capacidad de cómputo y/o transmisión, etc.) ¿Cómo se tiene en cuenta esta particularidad?

En los modelos de naturaleza global las peculiaridades propias y particulares de los dispositivos que forman la red no son tenidas en cuenta (prácticamente todas las interacciones entre un dispositivo infectado y uno susceptible son indistinguibles entre sí). Por una parte la topología de contactos se especializa, como mucho, a nivel de comunidad o de subcompartimentos formados por dispositivos con el mismo número de contactos (k-nodos) y nunca a nivel del agente particular) y la variación de dicha topología se expresa a través de las tasas de contacto definidas en incidencias o dependientes de la población total, o saturadas con respecto a dicha población total o al número total de infectados o susceptibles en cada instante de tiempo. Por otra parte, los coeficientes epidemiológicos empleados son generales y, al igual que en el caso anterior, no atienden a las particularidades de cada dispositivo

que pudieran influir de un modo u otro en el proceso de propagación del malware: sistema operativo, recursos computacionales, utilidad, capacidades del usuario, etc.

5.2. Una alternativa: la modelización basada en el individuo

Consecuentemente, para diseñar modelos más realistas, en el sentido de intentar capturar cuantas más propiedades particulares mejor, se puede recurrir a la denominada modelización basada en el individuo [45]. En este paradigma el análisis del sistema no se centra en el estudio de la evolución del número de dispositivos de un determinado compartimento o clase durante el periodo de tiempo $T \subset \mathbb{R}$ considerado, sino en la evolución del estado de cada individuo a lo largo de dicho periodo de tiempo. Consecuentemente las variables estudiadas en los modelos globales: $S(t), I(t), \ldots$, son sustituidas por variables que hacen referencia a la situación epidemiológica de cada uno de los dispositivos que intervienen en el sistema: así, por ejemplo, $s_i(t) \in \mathcal{S}$ representará el estado del nodo/dispositivo i-ésimo en el instante de tiempo t, con $1 \leq i \leq N$, y siendo $\mathcal{S}$ el conjunto de todos los posibles estados en los que se puede encontrar cada individuo -que habitualmente viene definido por los compartimentos en que se ha clasificado a la población-. Obviamente con esta nueva concepción es posible también calcular la evolución de cada uno de los compartimentos, por ejemplo: $I(t) = \sharp\{i \in \{1, 2, \ldots, N\} \text{ such that } s_i^t = \text{“infectious”}\}$.

Se trata de modelos muy flexibles en donde la dinámica temporal puede ser continua (existe $s_i(t) \in \mathcal{S}$ para todo $t \in T$) o discreta (existe $s_i(t)$ si $t = t_0, t_0 + \Delta t, t_0 + 2\Delta t, \ldots$) y el conjunto de estados puede ser un conjunto finito, $\mathcal{S} \sim \{k : k \leq n\}$, o infinito, usualmente $\mathcal{S} = [0, 1]^n$. Por otra parte, el cambio del estado de cada individuo vendrá determinado por una función de transición local que dependerá de los estados del propio individuo considerado y de sus “contactos”:

$$s_i(t + \Delta t) = \mathcal{F}_i\left(s_{j_1}(t), \ldots, s_{j_{k_i(t)}}(t)\right) \in \mathcal{S}, \quad 1 \leq i \leq N, \tag{80}$$

donde $\mathcal{N}_i(t) = \{j_1, j_2, \ldots, j_{k_i(t)}\} \subseteq \{1, 2, \ldots, N\}$ es el conjunto de individuos/dispositivos adyacentes al nodo i-ésimo en el instante de tiempo t (es decir, aquellos individuos con los el nodo i-ésimo mantiene un contacto adecuado en el instante de tiempo t). Obsérvese que estamos suponiendo que la topología de contactos viene definida por una red compleja temporal $\mathcal{G} = \{\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t)), t \in T\}$ donde el número de nodos se mantiene constante: $\sharp V = N$. Estas funciones de transición pueden ser de naturaleza determinista o estocástica y venir definidas por herramientas continuas como ecuaciones diferenciales [40] o discretas como autómatas celulares [20] o modelos basados en agentes [42].

Referencias

1. Al-Hawawreh, M., Alazab, M., Ferrag, M.A., Hossain, M.S.: Securing the industrial internet of things against ransomware attacks: A comprehensive analysis of the emerging threat landscape and detection mechanisms. J. Netw. Comput. Appl. **223**, 103809 (2024). DOI 10.1016/j.jnca.2023.103809
2. Antonopoulos, C.G., Akrami, M.H., Basios, V., Latifi, A.: A generic model for pandemics in networks of communities and the role of vaccination. Chaos **32**(6), 063127 (2022). DOI 10.1063/5.0082002
3. Bernoulli, D.: Essai d'une nouvelle abalyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. In: Mémoires de Mathématiques et de Physique, pp. 1–45. Académie Royale des Sciences, Paris (1760)
4. Bouaine, A., Rachik, M.: Modeling the impact of immigration and climatic conditions on the epidemic spreading based on cellular automata approach. Ecol. Inform. **46**, 36–44 (2018). DOI https://doi.org/10.1016/j.ecoinf.2018.05.004
5. Brown, A., Gupta, M., Abdelsalam, M.: Automated machine learning for deep learning based malware detection. Comput. Secur. **137**, 103582 (2024). DOI 10.1016/j.cose.2023.103582
6. Carnier, R., Li, Y., Fujimoto, Y., Shikata, J.: Exact markov chain of random propagation of malware with network-level mitigation. IEEE Internet Thing J. **10**(12), 10933–10947 (2023). DOI 10.1109/JIOT.2023.3240421
7. Castillo-Chavez, C., Song, B.: Models for the transmission dynamics of fanatic behaviors. In: Bioterrorism: Mathematical Modeling Applications in Homeland Security, pp. 155–172. Frontiers in Applied Mathematics, SIAM (2003). DOI 10.1137/1.9780898717518.ch7
8. Corallo, A., Lazoi, M., Lezzi, M., Luperto, A.: Cybersecurity awareness in the context of the industrial internet of things: A systematic literature review. Comput. Ind. **137**, 103614 (2022). DOI 10.1016/j.compind.2022.103614
9. Daley, D., Kendall, D.: Stochastic rumours. IMA J. Appl. Math. **1**(1), 42 – 55 (1965). DOI 10.1093/imamat/1.1.42
10. Diekmann, O., Heesterbeek, H., Britton, T.: Mathematical tools for understanding infectious disease dynamics. Princeton University Press, Princeton, NJ (2013)
11. Diekmann, O., Heesterbeek, J., Metz, J.: On the definition and the computation of the basic reproduction rate r_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. **28**, 365–382 (1990)
12. Dietz, K.: The first epidemic model: A historical note on p.d. en'ko. Aust. J. Stat. **30A**(1), 56–65 (1988). DOI https://doi.org/10.1111/j.1467-842X.1988.tb00464.x
13. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**, 29–48 (2002)
14. Fang, Z., Zhao, P., Xu, M., Xu, S., Hu, T., Fang, X.: Statistical modeling of computer malware propagation dynamics in cyberspace. J. Appl. Stat. **49**(4), 858–883 (2022). DOI 10.1080/02664763.2020.1845621
15. Ferdous, J., Islam, R., Mahboubi, A., Islam, M.Z.: A review of state-of-the-art malware attack trends and defense mechanisms. IEEE Access **11**, 121118–121141 (2023). DOI 10.1109/ACCESS.2023.3328351
16. Fu, X., Small, M., Chen, G.: Propagation Dynamics on Complex Networks: Models, Methods and Stability Analysis. John Wiley & Sons, UK (2014)
17. Gleissner, W.: A mathematical theory for the spread of computer viruses. Comput. Secur. **8**, 35–41 (1989). DOI 10.1016/0167-4048(89)90037-0
18. Godoi, A., Piqueira, J.: Spatio-temporal malware containment model with alert. Chaos Solitons Fractals **173**, 113618 (2023). DOI https://doi.org/10.1016/j.chaos.2023.113618
19. Goffman, W., Newill, V.: Generalization of epidemic theory: An application to the transmission of ideas. Nature **204**, 225–228 (1964). DOI 10.1038/204225a0
20. González, G., Lárraga, M., Álvarez Icaza, L., Gómez, J.: Bluetooth worm propagation in smartphones: Modeling and analyzing spatio-temporal dynamics. IEEE Access **9**, 75265–75282 (2021). DOI 10.1109/ACCESS.2021.3081482

21. Gorment, N.Z., Selamat, A., Cheng, L.K., Krejcar, O.: Machine learning algorithm for malware detection: Taxonomy, current challenges, and future directions. IEEE Access **11**, 141045–141089 (2023). DOI 10.1109/ACCESS.2023.3256979
22. Hamer, W.: Epidemic disease in england. Lancet **1**, 733–739 (1906)
23. Hethcote, H.W.: Mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
24. Hoang, M.: Global asymptotic stability of some epidemiological models for computer viruses and malware using nonlinear cascade systems. Bol. Soc. Mat. Mex. **28**(2), 39 (2022). DOI 10.1007/s40590-022-00432-9
25. Hoang, M.: Dynamical analysis of two fractional-order siqra malware propagation models and their discretizations. Rend. Circ. Mat. Palermo **72**(1), 751–771 (2023). DOI 10.1007/s12215-021-00707-6
26. Hoang, M.T., Ngo, T.K.Q., Hurg Tran, D.: Dynamically consistent nonstandard numerical schemes for solving some computer virus and malware propagation models. Math. Found. Comput. **6**(4), 704–727 (2023). DOI 10.3934/mfc.2022042
27. Hosseini, S., Azgomi, M., Torkaman, A.: Agent-based simulation of the dynamics of malware propagation in scale-free networks. Simul.-Trans. Soc. Model. Simul. Int. **92**(7), 709–722 (2016). DOI 10.1177/0037549716656060
28. Karyotis, V., Khouzani, M.H.R.: Malware Diffusion Models for Modern Complex Networks: Theory and Applications. Morgan Kaufmann, Cambridge, MA (2016)
29. Kaspersky: Kaspersky IT Encyclopedia. Website. URL: https://encyclopedia.kaspersky.com/. Accessed on 2024-01-24
30. Keeling M.J., R.P.: Modeling Infectious Diseases in Humans and Animals. Princeton University Press, Princeton, NJ (2008)
31. Kermack, W., McKendrick, A.: Contributions to the mathematical theory of epidemics, part i. Proc. Roy. Soc. London A **115**, 700–721 (1927)
32. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. ACM Trans. Web **1**(1), Article number 5 (2007). DOI 10.1145/1232722.1232727
33. Liu, W., Liu, C., Liu, X.: A discrete dynamic model for computer worm propagation. In: Springer Proceedings in Mathematics and Statistics, vol. 150, p. 119 – 131. Springer (2015). DOI 10.1007/978-3-319-24747-2_9
34. Llamazares-Elías, S., Tocino, A.: Stability analysis of a stochastic malware diffusion seir model. Lect Notes Netw. Syst. **748**, 197 – 204 (2023). DOI 10.1007/978-3-031-42519-$6_1$9
35. Lohner, R., Antil, H., Srinivasan, A., Idelsohn, S., Onate, E.: High-fidelity simulation of pathogen propagation, transmission and mitigation in the built environment. Arch. Comput. Method Eng. **28**(6), 4237–4262 (2021). DOI 10.1007/s11831-021-09606-6
36. Lu, Y., Xu, L.D.: Internet of things (iot) cybersecurity research: A review of current research topics. IEEE Internet Things J. **6**(2), 2103–2115 (2019). DOI 10.1109/JIOT.2018.2869847
37. Madhusudanan, V., Geetha, R., Murthy, B.S.N., Dao, N.N., Cho, S.: Analysis of delay-aware worm propagation model in wireless iot systems with ratio-dependent functional response. IEEE Access **11**, 34968–34976 (2023). DOI 10.1109/ACCESS.2023.3264978
38. Makhdoom, I., Abolhasan, M., Franklin, D., Lipman, J., Zimmermann, C., Piccardi, M., Moghadam, N.S.: Detecting compromised iot devices: Existing techniques, challenges, and a way forward. Computers and Security **132**, 103384 (2023). DOI 10.1016/j.cose.2023.103384
39. Manzil, H.H.R., Naik, S.M.: Detection approaches for android malware: Taxonomy and review analysis. Expert Syst. Appl. **238**(F), 122255 (2024). DOI 10.1016/j.eswa.2023.122255
40. Mei, W., Mohagheghi, S., Zampieri, S., Bullo, F.: On the dynamics of deterministic epidemic propagation over networks. Annu. Rev. Control **44**, 116–128 (2017). DOI 10.1016/j.arcontrol.2017.09.002
41. Miranda-Garcia, A., Rego, A.Z., Pastor-Lopez, I., Sanz, B., Tellaeche, A., Gaviria, J., Bringas, P.G.: Deep learning applications on cybersecurity: A practical approach. Neurocomputing **563**, 126904 (2024). DOI 10.1016/j.neucom.2023.126904
42. Mwangi, K., Masupe, S., Jeffrey, M.: Modelling malware propagation on the internet of things using an agent-based approach on complex networks. Jordanian J. Comput. Info. Technol. **6**(1), 26–40 (2020). DOI 10.5455/jjcit.71-1568145650

43. Nwokoye, C.H., Madhusudanan, V.: Epidemic models of malicious-code propagation and control in wireless sensor networks: An indepth review. Wirel. Pers. Commun. **125**(2), 1827–1856 (2022). DOI 10.1007/s11277-022-09636-8
44. Peng, S., Yu, S., Yang, A.: Smartphone malware and its propagation modeling: A survey. IEEE Commun. Surv. Tutor. **16**(2), 925 – 941 (2014). DOI 10.1109/SURV.2013.070813.00214
45. Railsback, S., Grimm, V.: Agent-Based and Individual-Based Modeling: A Practical Introduction, Second Edition. Princeton University Press (2019)
46. Raval, K.J., Jadav, N.K., Rathod, T., Tanwar, S., Vimal, V., Yamsani, N.: A survey on safeguarding critical infrastructures: Attacks, ai security, and future directions. Int. J. Crit. Infrastruct. Prot. **44**, 100647 (2024). DOI 10.1016/j.ijcip.2023.100647
47. Ross, R.: The prevention of malaria, 2nd ed. Murray, London, UK (1911)
48. Ross, R., Hudson, H.: An application of the theory of probabilities to the study of a priori pathometry - part iii. Proc. R. Soc. London Ser. A-Math. Phys. Eng. Sci. **43**, 225–240 (1917). DOI 10.1098/rspa.1917.0015
49. Verma, C., Gupta, C.: Effect of vaccination on stability of wireless sensor network against malware attack: An epidemiological model. SN Comput. Sci. **5**(2) (2024). DOI 10.1007/s42979-023-02532-3
50. Volkening, A., Linder, D.F., Porter, M.A., Rempala, G.A.: Forecasting elections using compartmental models of infection. SIAM Rev. **62**(4), 837 – 865 (2020). DOI 10.1137/19M1306658
51. Xiao, M., Chen, S., Zheng, W., Wang, Z., Lu, Y.: Tipping point prediction and mechanism analysis of malware spreading in cyber–physical systems. Commun. Nonlinear Sci. Numer. Simul. **122**, 107247 (2023). DOI https://doi.org/10.1016/j.cnsns.2023.107247
52. Zhang, H., Shen, S., Cao, Q., Wu, X., Liu, S.: Modeling and analyzing malware diffusion in wireless sensor networks based on cellular automaton. Int. J. Distrib. Sens. Netw. **16**(11), 1550147720972944 (2020). DOI 10.1177/1550147720972944
53. Zhu, L., Zhao, H., Wang, X.: Stability and bifurcation analysis in a delayed reaction-diffusion malware propagation model. Comput. Math. Appl. **69**(8), 852–875 (2015). DOI 10.1016/j.camwa.2015.02.004