第 58 卷 第 9 期 2009 年 9 月
1000-3290/2009/58(09)/5911-08

物 理 学 报
ACTA PHYSICA SINICA

Vol.58,No.9,September.,2009
©2009 Chin. Phys. Soc.

基于一维元胞自动机的复杂网络恶意软件传播研究*

宋玉蓉[1)†] 蒋国平[1)2)]

1)(南京邮电大学控制与智能技术研究中心，南京 210003)

2)(南京邮电大学自动化学院，南京 210003)

(2008 年 11 月 9 日收到；2008 年 11 月 26 日收到修改稿)

基于一维元胞自动机，研究复杂网络恶意软件传播行为.利用信息网络节点全局交互的特点，建立元胞自动机邻域和状态转换函数，提出恶意软件传播模型，研究在多种网络拓扑下恶意软件传播的概率行为.研究表明，该模型能够准确描述在最近邻耦合网络(nearest-neighbor coupled network，NC)，Erdos-Renyi(ER)随机网络，Watts-Strogatz(WS)小世界网络和 Barabasi-Albert(BA)幂率网络等拓扑下的传播动力学行为，不仅能反映恶意软件传播的平均趋势，而且可以描述病毒消亡和渗透等稀有概率事件，有效克服基于平均场方法建立的微分方程模型只能反映传播的平均趋势，只适合对传播作整体预测的局限性.同时，研究指出网络中度分布的异质化程度和网络的局域空间交互特征是影响传播及免疫行为的关键要素.

关键词：复杂网络，恶意软件传播，元胞自动机，状态转换函数

PACC：0250，0565，0570J

1. 引 言

迅速发展的复杂网络理论正有效地增进人们对爆发大规模生物和计算机病毒流行的传染机理的认识[1—9].尤其，基于通信网络诸如 Internet，WWW，P2P 和 E-mail 等信息技术网络的恶意软件传播模型的建立及动力学分析问题被一些研究者所关注[4,10—12].恶意软件传播模型大多基于流行病学的研究，计算机病毒和蠕虫与生物病毒在自复制和传播行为上是类似的，因此由研究生物传染病学发展起来的数学方法都可用于计算机病毒及蠕虫的传播.

在信息技术网络的病毒和蠕虫建模领域，Kephart 等人[13]基于流行病学模型已经进行了一系列研究.Staniford 等[14]提出随机常数传播模型研究，该模型与所观察到的 Code Red 的传播增长部分的数据匹配较好.Zou 等在考虑人为对抗措施以及由蠕虫引起的网络拥塞等因素影响基础上提出“双因素”蠕虫模型[15].他们还构建电子邮件病毒传播模型[4]，认为电子邮件网络拓扑对电子邮件病毒的传播行为起到重要影响.Hiroyuki 等[16]把病毒传播的节点行为当作生灭过程，由此建立随机模型研究网络蠕虫传播的概率行为.Chen 等人[17]使用空-时随机过程描述了恶意软件在任意网络拓扑下的统计依赖.

在流行病学领域，生物病毒传播模型分为两类：确定性模型和随机模型.确定性模型使用平均场方法简化问题并用微分方程描述病毒传播的平均趋势，不考虑概率事件.这导致此类模型无法表述传播过程中的概率事件，例如病毒消亡或突发事件，此外确定性模型忽视了个体之间的交互行为.文献[4，13—15]中建立的就是基于平均场方法的确定性模型.随机模型多基于马尔可夫链，此类模型缺乏空间概念，且通常事先固定的状态转移矩阵不适合描述病毒传播的动态演化.文献[16,17]中建立的是随机模型.此外，网络拓扑对恶意软件的传播产生重要影响，然而现有的传播模型大多基于网络同质性的假设忽视了拓扑对传播行为的影响或只针对一种网络拓扑进行研究[4].

元胞自动机(cellular automata，CA)是一个具有简单构造但产生复杂自组织行为的离散动力学系统，CA 能有效克服基于平均场方法建立的微分方程

* 国家教育部新世纪优秀人才支持计划(批准号：NCET-06-0510)，国家自然科学基金(批准号：60874091)，江苏省普通高校研究生科研创新计划(批准号：CX08B_081Z).

† 通讯联系人.E-mail：songyr@njupt.edu.cn

模型和基于马尔可夫链建立的随机模型所表现出来的缺陷,是一种研究传染病传播动力学的有效替代方法.基于元胞自动机对生物传染病学的研究已经受到广泛关注[18—22],将 CA 应用到 Internet,WWW,P2P 等信息技术网络中研究恶意软件传播的动力学行为将是一种极为有效的研究方法.生物传染病学强调接触传染的局域交互特性,而上述 Internet,WWW,P2P 等信息网络的虚拟空间特性使得节点与节点交互趋于全局特性,即节点交互往往和空间距离无关,和节点之间是否有通信连接有关.直接将生物传染病学的有关研究结果用于信息技术网络并不适合 Internet,WWW,Email 和 P2P 等技术网络中的恶意软件传播问题.

本文基于一维元胞自动机建立信息网络中的恶意软件传播模型,分析恶意软件在多种网络拓扑下的概率传播行为.模型考虑个体间简单交互行为而产生复杂系统行为这一复杂系统的本质特征,针对信息网络全局交互的特点,建立元胞自动机邻域和转换规则.将提出的模型用于分析研究最近邻耦合(NC)网络,Erdos-Renyi(ER)随机网络[23],Watts-Strogatz(WS)小世界网络[24]和 Barabasi-Albert(BA)无尺度幂率网络[25]等多种网络拓扑下的恶意软件传播问题.研究表明,提出的模型不仅能反映恶意软件传播的平均趋势,而且可以描述病毒消亡和渗透等稀有概率事件,有效地克服了平均场方法建立的微分方程模型只能反映传播的平均趋势,只适合对传播作整体预测的局限性,也克服了基于马氏链建立的随机模型缺乏空间概念、模型复杂、不适合描述恶意软件动态演化的特征.此外,本文还对恶意软件的传播阈值、传播速度、传播范围等指标在多种网络拓扑中进行了比较研究,指出网络的度分布异质化程度及网络的局域交互特征是影响传播特征的关键要素.

2. 随机恶意软件传播模型

我们使用 CA 建立恶意软件传播的随机模型,该模型的建立由网络固有的性质及病毒传播机理中固有的随机特性所决定.

一个 CA 可以通过一个四元组定义:

$$\mathrm{CA} = (C, Q, V, f), \tag{1}$$

C 表示元胞空间,Q 表示有限状态集,V 表示节点的邻域,f 代表状态转换规则函数.使用 CA 建立传播模型,事实上就是要定义不同特征条件下(包括恶意软件攻击特性、传播环境等)元胞空间、元胞状态、元胞邻居和状态转换规则集等模型要素,建立恶意软件传播的动态演化模型.

考虑是否具有免疫机理,我们建立了两种模型.

2.1. 无免疫机理模型

该模型考虑系统中节点状态只能处于健康状态(susceptible)和感染状态(infected)之一,节点状态变换关系:susceptible→infected→susceptible.我们命名这种模型为 SIS-CA 模型.

考虑网络 $G=(N;E)$,N 表示网络中节点的个数,E 表示网络中节点与节点的边,令 A 表示网络的邻接矩阵,它反映网络的拓扑信息.根据元胞自动机四要素,建立模型:

元胞空间 C:建立包含 N 个元胞的一维元胞空间,一维元胞空间中的一个元胞即代表网络中的一个节点;

邻域 V:传统 CA 定义中,元胞邻域只由邻接元胞或以空间距离为半径确定邻居,如 Von Neumann 型,Moore 型邻居等[26].这种邻域定义强调生物网络中的节点交互局域特征,这显然不能满足诸如 Internet,WWW,P2P 和 E-mail 等信息技术网络中节点交互的全局特性,即节点交互往往和空间距离无关,和节点之间的通信连接有关.为表达这种空间全局特性,在我们的模型中,元胞邻域的定义突破标准元胞自动机中邻域的定义,以网络的邻接矩阵 A 直接定义各元胞邻居关系,所以节点 i 的邻域 V_i 就被定义为 A 中的第 i 行的向量,即 $V_i = \{a_{ij} \mid a_{ij} \in A, j = 1,2,\cdots,N\}$ 若 $a_{ij}=1$,表示节点 i 和 j 之间存在连接.

状态集 Q:本系统基于 SIS 模型,仅考虑节点的健康、感染两种状态,健康用状态 0 表示,感染以状态 1 表示,令 $Q=\{0,1\}$,节点 i 在 t 时刻的状态变量用 $s_i(t)$($s_i(t) \in Q$)表示,则有

$$s_i(t) = \begin{cases} 1, & \text{节点 } i \text{ 在时刻 } t \text{ 为 infected}, \\ 0, & \text{节点 } i \text{ 在时刻 } t \text{ 为 susceptible}. \end{cases} \tag{2}$$

本地转换函数 f:也称转换规则,它是 CA 的核心,CA 的转换规则有多种形式,根据不同的应用目的需要定义不同的转换规则,最为著名的转换规则是‘生命游戏’转换规则[26],该规则建立在节点局部交互的特点之上.本模型基于信息网络节点交互的全局特性及病毒传播机理中固有的随机特性,建立了数学表达完整的本地转换函数.

任何节点仅能被其邻居感染，节点 i 在离散时刻 t 的状态 $s_i(t)$ 依赖于节点在上一时刻的自身状态 $s_i(t-1)$ 和其邻居的状态 $s_{V_i}(t-1)$，每个时间间隔内感染节点试图以概率 β 感染它的邻居，同时感染节点也以概率 δ 恢复健康，由此建立如下转换函数：

$$\begin{aligned} s_i(t+1) &= \max(f_\delta(s_i(t))(1-\delta_i)), \\ & f_\beta(\overline{s_i(t)})(1-(1-\beta_i)^{m_i(t)}))), \end{aligned} \tag{3}$$

这里，$\overline{\cdot}$ 表示取反操作，(3)式中第一项 $f_\delta(s_i(t))(1-\delta_i))$ 表示先前处于感染状态节点，经过一个离散时间 t 后，其状态改变结果：感染节点以 δ 概率治愈，以 $1-\delta$ 概率维持原有的感染状态，即 $f_\delta(x)$ 为治愈过程的状态转换子函数，具体定义如下：

$$f_\delta(x) = \begin{cases} 1, x \geqslant \delta, \\ 0, x < \delta. \end{cases} \tag{4}$$

(3)式中第二项 $f_\beta(\overline{s_i(t)})(1-(1-\beta_i)^{m_i(t)}))$ 表示处于健康状态节点，经过一个离散时间 t 后的状态改变结果．即 $f_\beta(x)$ 为感染过程的状态转换子函数，

$$f_\beta(x) = \begin{cases} 1, x \geqslant (1-\beta), \\ 0, x < (1-\beta). \end{cases} \tag{5}$$

一个感染节点以概率 β 感染处于健康状态的邻居节点，即一个健康节点受到处于感染状态的邻居的感染．这个健康节点获得感染的概率随着处于感染状态的邻居数目的增加而增加，这个概率表示为 $1-(1-\beta_i)^{m_i(t)}$，其中，$m_i(t)$ 表示在 t 时刻，节点 i 的邻居中处于感染状态的邻居数目，

$$m_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t). \tag{6}$$

(3)式中的第一项和第二项结果中的最大值即为节点经过一个时间 t 后的结果状态．

设 $I(t)$ 表示 t 时刻网络中受感染节点的比率，$S(t)$ 表示处于健康状态节点比率；$I(0)$ 表示初始时刻的主机感染数目．显然有下列统计结果：

$$\begin{aligned} I(t) &= \frac{1}{N}\sum_{i=1}^{N} s_i(t), \\ I(t) + S(t) &= 1. \end{aligned} \tag{7}$$

2.2. 带免疫机理模型

考虑免疫机理后，节点状态处于 susceptible，infected 和免疫状态(removed)之一，节点状态变换关系：susceptible→infected→removed．我们命名这种模型为 SIR-CA 模型．特别注意的是，免疫节点意味着不会再遭受感染也不会再去感染其他健康节点，所以网络中的免疫节点意味着对恶意软件的传播是没有贡献的，它等同于从传播网络中移除了此节点．因此，模型中，节点 i 处于免疫状态后，需要修改表示节点连通状态的邻接矩阵 A，去除与节点 i 相连的任何边．即邻接矩阵随着时间演化，免疫节点的增加使得传播网络呈现动态演化特征．

在 SIR-CA 模型中，元胞、元胞空间和邻域的定义与 SIS-CA 模型中的定义是类似的，在此不再赘述，重点阐述元胞状态集 Q 和本地转换函数 f 的定义．

状态集 Q：令 $Q = \{(0,0),(0,1),(1,0),(1,1)\}$，定义节点 i 在 t 时刻的状态向量为 $s_i(t) \in Q$，该向量包含两个分量，即 $s_i(t) = (s_{ix}(t), s_{iy}(t))$，第一分量与是否感染有关，第二分量与是否免疫有关，则有

$$s_i(t) = \begin{cases} (0,0), \text{节点 } i \text{ 在时刻 } t \text{ 为 susceptible}, \\ (1,0), \text{节点 } i \text{ 在时刻 } t \text{ 为 infected}, \\ (0,1), \text{节点 } i \text{ 在时刻 } t \text{ 为 removed}, \\ (1,1), \text{不存在}. \end{cases} \tag{8}$$

本地转换函数 f：在 SIR-CA 模型中，$s_i(t)$ 遵循如下转换规则：

$$\begin{aligned} s_{ix}(t+1) &= \max(f_\delta(s_{ix}(t))(1-\delta_i)), \\ & f_\beta(\overline{s_{ix}(t)})(1-(1-\beta_i)^{m_i(t)}))), \\ s_{iy}(t+1) &= f_\gamma(\overline{s_{ix}(t+1)}\, s_{ix}(t)\gamma_i), \end{aligned} \tag{9}$$

(9)式中关于第一分式解释同(3)式类同，在此不再赘述．第二分式中 $f_\gamma(\overline{s_{ix}(t+1)}\, s_{ix}(t)\gamma_i)$ 项表示节点 i 若在 t 时刻为感染状态在下一时刻即 $t+1$ 时刻获得治愈则同时该节点在 $t+1$ 时刻以概率 γ 获得免疫．$f_\gamma(x)$ 就为免疫过程的状态转换子函数，

$$f_\gamma(x) = \begin{cases} 1, x \leqslant \gamma, \\ 0, x > \gamma. \end{cases} \tag{10}$$

为简单起见，本模型中采用了随机免疫策略，实际上，根据恶意软件传播特征及网络拓扑结构，可以设计并应用各种免疫策略．

特别需要说明的是，表示节点 i 在 t 时刻处于感染状态的邻居数目的变量 $m_i(t)$，需要做轻微修正，

$$m_i(t) = \sum_{j=1}^{N} a_{ij}(t)\, s_{jx}(t). \tag{11}$$

这里，如果 $s_i(t) = (0,1)$，则 $a_{ij}(t) = a_{ji}(t) = 0, j = 1,2,\cdots,N$．我们有如下统计结果：

$$I(t) = \frac{1}{N}\sum_{i=1}^{N} s_{ix}(t),$$

$$R(t) = \frac{1}{N}\sum_{i=1}^{N} s_{iy}(t),$$

$$I(t) + R(t) + S(t) = 1, \qquad (12)$$

$R(t)$ 表示处于免疫状态的节点比例.

3. 传播阈值研究

为验证提出的模型,依据最近邻耦合(nearest-neighbor coupled,NC)网络定义,Erdos-Renyi(ER)随机网络模型算法[27],Watts-Strogtz(WS)小世界网络模型算法[24]和 Barabasi-Albert(BA)无标度网络模型算法[25](幂率指数 $\alpha = 3$),我们生成如表 1 所示的这四种典型的网络,应用提出的模型 SIS-CA 于这些网络中,通过仿真分析恶意软件在不同网络拓扑下的传播动力学特性.

Kephart 等[13] 基于同质网络(homogenous network)假设提出基于流行病模型建立网络病毒传播模型,得到病毒传播域值为 $\lambda_{KW} = 1/\langle k\rangle$. 抛开网络同质性假设,Pastor-Satorras 等[3,5,6,28] 研究了复杂异质网络(heterogeneous network)下病毒的爆发,得到传播临界值 $\lambda_{SV} = \langle k\rangle/\langle k^2\rangle$. Chakrabarti 和 Wang 等[29,30] 提出了一个称之为 NLDS 的病毒传播模型,得到无向图的传播临界值为 $\lambda_{CW} = 1/\tau_{1,A}$,其中 $\tau_{1,A}$ 为网络邻接矩阵 A 的最大特征根.

这一部分我们基于提出的 SIS-CA 模型,通过仿真研究了各种网络下的传播临界值问题,并与上述三种方法得到的临界值进行比较分析.

表 1 网络及网络参数

网络类型	NC	ER	WS	BA
网络规模 N	200	200	200	200
网络生成参数	–	–	$p = 0.2$(随机重连概率)	$m_0 = 8, m = 3$
网络平均度$\langle k\rangle$	6	6.01	6	6.025
* 网络平均均方度$\langle k^2\rangle$	36	42.05	37.1	78.725
聚类系数 C	0.6	0.0301	0.3072	0.0976
最大特征根 $\tau_{1,A}$	6	7.17	6.26	12.45

* 此参数反映网络度分布的异质化(heterogeneous)程度,节点度分布越不均匀$\langle k^2\rangle$值越高,反之,该值越低,对于均匀网络$\langle k^2\rangle \to \langle k\rangle^2$.

仿真中的网络及相关网络参数如表 1 所示,传播参数取为:$\delta = 0.8$;$I(0) = 1$;β 从 0 以步长 0.01 线性增长到 1;已知传播值 $\lambda = \beta/\delta$. 在以上参数设置下运行 SIS-CA 模型 100 次后统计平均建立感染密度与传播值 λ 的对应关系如图 1 所示,表 2 统计了应用不同方法四种网络拓扑下传播临界值结果.

表 2 传播域值比较

网络类型	RG	ER	WS	BA
$\lambda_{KW} = 1/\langle k\rangle$	0.1667	0.1664	0.1667	–
$\lambda_{SV} = \langle k\rangle/\langle k^2\rangle$	0.1667	0.1429	0.1617	0.0828
$\lambda_{CW} = 1/\tau_{1,A}$	0.1667	0.1395	0.1597	0.0803
λ_c (SIS-CA 仿真结果)	0.2000	0.1375	0.1625	0.0625
$\Delta\lambda = \lvert\lambda_c - \lambda_{SV}\rvert/\lambda_{SV}$	19.98%	3.78%	0.49%	24.52%

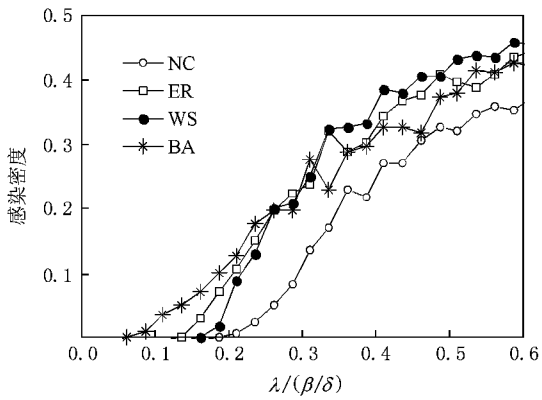

图 1 SIS-CA 模型感染密度与传播值 λ 的对应关系

首先从图 1 中可看出 NC 具有最大的传播临界值,WS 和 ER 网络其次,BA 网络最小,即网络度分布的异质化程度越高,临界值越小,图 1 中曲线与文献[5]中得到的曲线(如图 2)有极为相似地符合. BA 网络有最小的传播阈值和较小的传播规模,最近邻耦合网络有最大的阈值却有最小的传播规模,说明病毒在最近邻耦合网络中最不易爆发. 这是因为最近邻耦合网络中节点与节点之间呈现很强的局域空间交互特征,这导致病毒在扩散过程中受局域交互特征的影响,每个处于感染状态的节点实际能有效

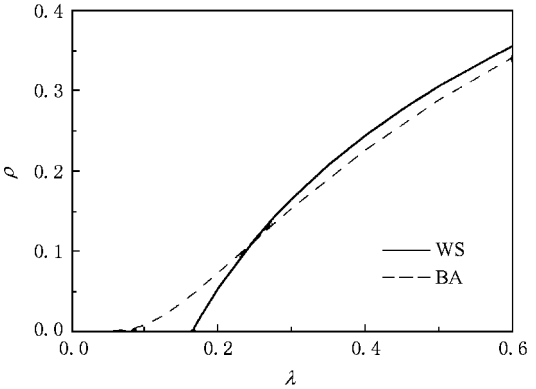

图 2　SIS 模型感染密度与传播值的对应关系曲线(取自文献[5])

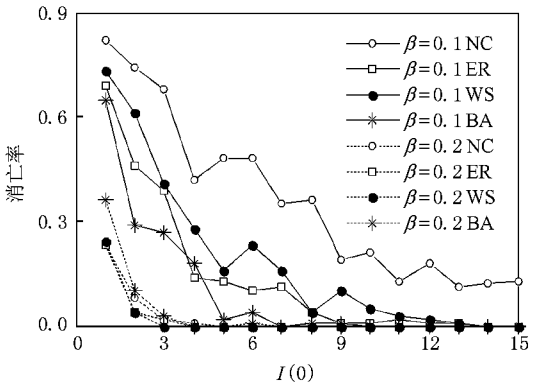

图 3　SIS-CA 模型演示消亡率与 $I(0)$和 β 关系，$\delta = 0.5$

感染的邻居数目要远少于节点的平均度.

表 2 从具体数值上将通过 SIS-CA 模型得到的仿真阈值与已有的三种方法得到阈值进行比较.在 ER，WS 网络拓扑下，SIS-CA 模型得到的仿真阈值结果与 λ_{SV}，λ_{CW}的结果都有较好地符合，在 NC 和 BA 网络中，我们所提出的模型则与已知方法的阈值结果有较大差异，NC 网络中尽管度分布十分均匀，但其局域交互特性使得传播阈值显著增加.BA 网络中的结果说明病毒更加容易在度分布异质化程度更高的网络中传播.

4. 传播演化

基于表 1 中的四种网络，我们对 SIS-CA 模型和 SIR-CA 模型进行仿真试验，通过模型可以看到同样初始参数设置下，恶意软件的传播行为不尽相同，在传播值大于传播阈值时($\lambda > \lambda_c$)时，恶意软件有可能持续感染健康主机最后达到稳定的感染规模，也有可能在病毒传播最早期就趋于消亡，而并没有在网络中流行，这两种情形在以上四种网络中都可观察到.SIS-CA 模型中，取仿真参数 $\beta = 0.2$，$\delta = 0.5$，$I(0) = 1$，此时 $\lambda = 0.4$，参考表 2，该值均大于四种网络传播阈值.这四种网络分别执行 100 次仿真试验，观察到 ER 网络中共有 23 次出现消亡，77 次病毒大规模流行；WS 网络中出现消亡的概率占到了 24%，BA 网络为 36%，NC 网络为 23%；SIR-CA 模型中，取仿真参数 $\gamma = 0.01$，当 $I(0) = 2$ 时，四种网络呈现部分消亡趋势，NC 网络中共出现 25 次消亡，ER 网络中共出现 22 次消亡，WS 出现 27 次，BA 网络中出现了 27 次.

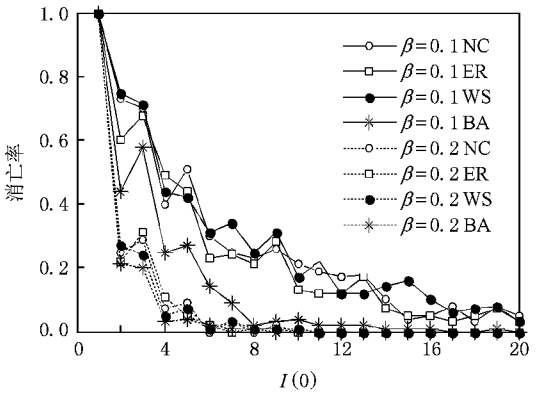

图 4　SIR-CA 模型演示消亡率与 $I(0)$和 β 关系，$\delta = 0.5$，$\gamma = 0.01$

改变 $I(0)$和病毒感染率 β 的值进一步仿真分析，发现恶意软件在传播之初就趋于消亡情形出现的比率与$I(0)$呈负指数关系，且随感染率降低，消亡比率增大，图 3 和图 4 分别反映了在 SIS-CA 和 SIR-CA 模型下传播趋于消亡的比率与 $I(0)$和感染率之间的关系曲线.然而，当 $\lambda > \lambda_c$ 时，病毒传播最早期就趋于消亡的情形在基于平均场方法建立的只能反映病毒传播平均趋势的确定性模型中是无法反映的.

病毒一旦流行，其传播规模及传播速度受多种因素影响，这里我们侧重研究网络拓扑对恶意软件传播的影响.图 5 和图 6 分别描述了两种模型下恶意软件在 NC 网络、ER 网络、WS 网络和 BA 网络病毒传播和流行趋势，网络参数仍如表 1 所示.从传播速度而言，BA 网络 > ER 网络 > WS 网络 > NC 网络.SIS-CA 模型中，传播稳定后的主机感染数目即流行范围来看，BA 网络 < ER 网络 < WS 网络 < NC 网络.在我们的模型中，NC 网络由于其鲜明的空间局域交

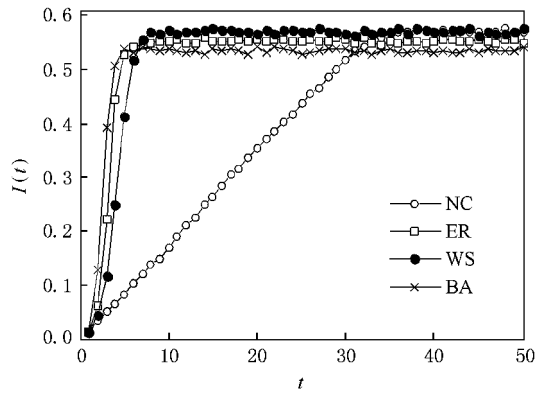

图 5 SIS-CA 模型恶意软件传播趋势($\beta = 0.3, \delta = 0.5, I(0) = 1$)

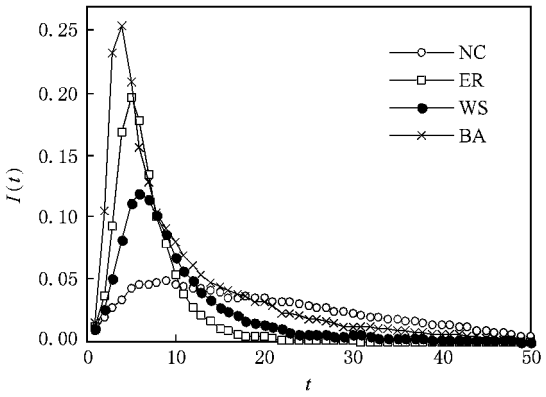

图 6 SIR-CA 模型恶意软件传播趋势($\beta = 0.2, \delta = 0.5, \gamma = 0.01$, $I(0) = 2$)

互特征,它的传播速度远小于其他三种网络环境.在 SIR-CA 模型中,显然随机免疫策略对 ER 拓扑最为有效,可使感染节点数目迅速回落,尽管 NC 网络中受到感染的节点数目最少,但随机免疫策略对 NC 拓扑效果很差,病毒消亡趋势极为缓慢,同样由于 NC 网络的局域化特性,使得该网络拓扑对随机免疫表现出很差的特性.

比较 ER 网络、小世界网络和无标度网络这三种不同网络拓扑下,恶意软件的传播行为,发现网络的异质化程度主导网络的传播特性.例如,不考虑免疫机理小世界网络具有比随机网络更慢的传播速度,但稳定后所达到的感染规模要高于随机图网络;应用随机免疫策略后,同样发现 WS 网络传播速度低于随机网络和 BA 网络.参考表 1 中所建三种网络的相关参数$\langle k^2 \rangle$和网络最大特征根,这两个值越大说明网络的异质化程度越高,同质化程度越低,反之则代表网络度分布的同质化程度越高,异质化程度越低.故我们得出一般性结论,网络异质化程度越高,病毒传播越快,无免疫策略应用时,同质化程度越高的网络最终感染波及的范围越大.而考虑随机免疫机理后,网络度分布的异质化程度和网络的局域交互特征决定网络病毒趋于消亡的时间.

5. 结 论

本文提出使用一维元胞自动机建立恶意软件随机传播模型 SIS-CA 和 SIR-CA,基于诸如 Internet, WWW,P2P 和 E-mail 等信息技术网络中节点交互的全局特性,即节点交互往往和空间距离无关,和节点之间的通信连接有关.扩充传统元胞自动机中定义,提出表达完整的状态转换规则,研究在任意网络拓扑下恶意软件传播的概率行为.通过使用 SIS-CA 模型对 NC 网络、ER 网络、WS 网络和 BA 网络中的传播临界值研究,并与现有对临界值研究的成果进行分析比较,提出的模型所得传播临界值符合了现有结论,网络度分布异质化程度越高,传播临界值越趋于消失.同时发现,网络的局域交互特征会增大传播临界值,使得恶意软件较难在此类网络中传播 提出的模型有效克服基于平均场方法建立的微分方程模型只能反映传播的平均趋势,只适合对传播作整体预测的局限性,提出的模型不仅能反映恶意软件传播的平均趋势,而且可以描述病毒消亡和渗透等稀有概率事件 恶意软件在各种网络下并不总是会爆发,有可能在感染之初就趋于消亡,消亡概率与$I(0)$呈指数下降关系,与感染率也有关系.从传播趋势看网络度分布的异质化程度越高,恶意软件传播速度越快,传播规模越小,网络的局域交互特征极大减缓了传播速度 考虑随机免疫机理后,网络度分布的异质性和网络的局域交互特征主导病毒消亡的速度.

[1] Chen Z Q, Pei W D, Yuan Z Z 2008 *Chin. Phys.* B **17** 373

[2] Moreno Y, Gomez J B, Pacheco A F 2003 *Phys. Rev.* E **68** 035103

[3] Moreno Y, Pastor-Satorras R, Vespignani A 2002 *European Physical Journal* B **26** 521

[4] Zou C C, Towsley D, Gong W B 2007 *IEEE Transactions on*

Dependable and Secure Computing **4** 105
[5] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev.* E **63** 066117
[6] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
[7] Li X, Wang X F, Xu D 2007 *Acta Phys. Sin.* **56** 1313 (in Chinese) [李　翔、汪小帆、许　丹 2007 **物理学报 56** 1313]
[8] Qiu B, Jiao J, Li Y *et al* 2005 *Chin. Phys.* **14** 2153
[9] Pei W D, Liu Z X, Chen Z Q, Yuan Z Z 2008 *Acta Phys. Sin.* **57** 6777 (in Chinese) [**裴伟东、刘忠信、陈增强、袁著祉** 2008 **物理学报 57** 6777]
[10] Song Y, Jiang G P 2008 *Proceedings of IEEE ICNNSP*2008 Zhengjiang, China 623
[11] Newman M E J, Forrest S, Balthrop J 2002 *Phys. Rev.* E **66** (3)
[12] Thommes R W, Coates M J. 2005 *Proceedings of the Fifth International Conference on Information, Communications and Signal Processing* 981
[13] Kephart J O, White S R, Chess D M 1993 *IEEE Spectrum* **30** 20
[14] Staniford S, Paxson V, Weaver N 2002 *Usenix Security*.
[15] Zou C C, Gong W, Towsley D 2002 *Proceedings of the* 9*th ACM Conference on Computer and Communications Security* 10
[16] Okamura H, Kobayashi H, Dohi T 2005 *Proceedings of the* 16*th IEEE International Symposium on Software Reliability Engineering* 149
[17] Chen Z S, Ji C Y 2005 *IEEE Transactions on Neural Networks* **16** 1291
[18] Jin Z, Liu Q X, Mainul H 2007 *Chin. Phys.* **16** 1267
[19] Liu Q X, Jin Z 2005 *Chin. Phys.* **14** 1370
[20] White S H, Rey A M d, Sanchez G R 2007 *Applied Mathematics and Computation* **186** 193
[21] Fuentes M A, Kuperman M N 1999 *Physica* A **267** 471
[22] Jin Z, Liu Q X 2006 *Chin. Phys.* **15** 1248
[23] Erdos P, Rényi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
[24] Watts D J, Strogatz S H 1998 *Nature* **393** 409
[25] Barabási A L, Albert R 1999 *Science* **286** 509
[26] Kari J 2005 *Theoretical Computer Science* **334** 3
[27] Erdos P, Rényi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
[28] Pastor-Satorras R, Vespignani A 2002 *Phys. Rev.* E **65** 035108
[29] Chakrabarti D, Wang Y, Wang C *et al*. 2007 *ACM Transactions on Information and System Security* **10** 1
[30] Wang Y, Chakrabarti D, Wang C *et al* 2003 *Proceedings of the* 22*nd International Symposium on Reliable Distributed Systems* 25

Research of malware propagation in complex networks based on 1-D cellular automata *

Song Yu-Rong[1)†] Jiang Guo-Ping[1)2)]

1) (*Center for Control & Intelligence Technology, Nanjing University of Posts and Telecommunications, Nanjing* 210003, *China*)

2) (*College of Automation, Nanjing University of Posts and Telecommunications, Nanjing* 210003, *China*)

(Received 9 November 2008; revised manuscript received 26 November 2008)

Abstract

In this paper, based on 1-D cellular automata, the probabilistic behaviors of malware propagation in complex networks are investigated. Neighborhood and state transition functions with integrated expression are established and two models of malware propagation are proposed to evaluate the probabilistic behavior of malware propagation in various networks. We run the proposed models on nearest-neighbor coupled network (NC) and Erdos-Renyi (ER) random graph network and Watts-Strogatz (WS) small world network and Barabasi-Albert (BA) power law network respectively. Analysis and simulations show that, the proposed models describe perfectly the dynamic behaviors of propagation in the above networks. Furthermore, the proposed models describe not only the average tendency of malware propagation but also the rare events such as saturation and extinction of malware, and overcome the limitation occurring in a deterministic model based on mean-field method that describes only the average tendency of malware propagation and neglects the probabilistic event. Meanwhile, the result of simulations shows that the heterogeneity of degree distribution and local spatial interaction are key factors affecting the malware propagation and immunization.

Keywords: complex network, malware propagation, cellular automata, transition function

PACC: 0250, 0565, 0570J

* Project supported by the Program for New Century Excellent Talents in University of China (Grant No. NCET-06-0510) and the National Natural Science Foundation of China (Grant No. 60874091) and the Scientific Innovation Program for University Research Students in Jiangsu Province (Grant No. CX08B-081Z).

† Corresponding author. E-mail: songyr@njupt.edu.cn