基于元胞自动机的复杂信息系统安全风险传播研究*

李钊[1)2)†] 徐国爱[1)2)] 班晓芳[3)] 张毅[3)] 胡正名[1)2)]

1) (北京邮电大学信息安全中心, 北京 100876)

2) (北京邮电大学, 灾备技术国家工程实验室, 北京 100876)

3) (中国信息安全测评中心, 北京 100085)

(2013年5月2日收到; 2013年6月17日收到修改稿)

基于元胞自动机建立复杂信息系统安全风险传播模型, 研究复杂信息系统安全风险在最近邻耦合网络、随机网络, Watts-Strogatz 小世界网络和 Barabasi-Albert 无标度网络四种网络拓扑下的传播问题. 通过研究安全风险传播模型在四种网络拓扑下安全风险的传播阈值, 与现有的传播阈值研究成果进行比较, 验证模型的正确性, 并分析验证网络拓扑结构中度分布的异质化程度越高传播阈值越小的结论. 通过对安全风险的传播演化趋势进行研究, 分析验证网络度分布的异质化程度越高、安全风险影响范围越小、传播速度越快的结论, 并指出度分布的异质化程度越高、模型后期的免疫机制对控制安全风险传播的效果越缓慢. 通过对安全风险在传播最早期就趋于消亡的情况进行研究, 分析得出安全风险在传播之初就趋于消亡的消亡率与传播率之间呈现近似负指数的关系, 并且初期的感染源越多安全风险的消亡率越低. 分析了影响复杂信息系统安全风险传播的关键要素, 对复杂信息系统中安全风险传播的控制具有指导作用.

关键词: 复杂信息系统, 复杂网络, 安全风险传播, 元胞自动机

PACS: 02.50.−r, 05.65.+b, 05.70.Jk **DOI:** 10.7498/aps.62.200203

1 引 言

随着计算机技术与网络技术的快速发展, 传统信息系统不断向大的、复杂的系统演化, 这种演化过程导致了大型复杂信息系统的出现. 复杂信息系统是呈现高度复杂性的信息系统, 其系统规模大、复杂性高并且结构和功能都具备复杂的网络特性, 是一种具有复杂网络结构的信息系统. 社会信息化的不断发展使得复杂信息系统已经逐渐融入社会生活的各个方面, 复杂信息系统的安全性也显得尤为重要. 同时, 针对复杂信息系统的恶意攻击也逐渐复杂化和多样化, 其中网络计算机病毒借助系统软件的安全弱点, 以自动化的方式来蔓延扩散并攻击复杂信息系统 [1]; 而一些组织也在不断地发布各种信息系统的安全弱点数据, 为针对复杂信息系统的攻击提供了便利 [2]. 此类不安全因素正在严重地危及着复杂信息系统的安全. 为了维护复杂信息系统的安全和稳定, 需要适用于复杂信息系统的风险分析及评估技术来对系统进行安全防范, 识别复杂信息系统中潜在的安全威胁, 根据其安全态势及需求来指定安全策略, 避免安全事件的发生.

复杂信息系统是一种具有复杂网络结构的信息系统, 迅速发展的复杂网络理论正快速地增进人们对生物病毒和计算机病毒大规模爆发流行的传染机理的认识. 复杂信息系统中安全风险的传播与计算机病毒和生物病毒在传播的行为上是类似的, 复杂信息系统中受害节点的安全风险可能会 “传染” 原本不直接具有安全风险的节点, 使后者受到安全威胁, 即 “感染” 安全风险, 例如攻击者攻破系统中的某个节点后, 就有可能接着攻击与其相邻的节点, 因此相邻的节点就受到安全威胁, “感染” 安全风险; 同时具有安全风险的节点也在通过更新安全策略或增加安全防护措施等方式以一定概率消

* 国家自然科学基金 (批准号: 60970135, 61170282)、高等学校博士学科点专项科研基金 (批准号: 20120005110017) 和国家科技支撑计划 (批准号: 2012BAH06B02) 资助的课题.

† 通讯作者. E-mail: lizhaoabc@gmail.com

除安全风险, 即 “恢复健康”. 研究生物病毒传播的数学方法可用于研究计算机病毒的传播, 同样可用于研究复杂信息系统中安全风险的传播.

Kephart 等 [3] 基于生物传染病学模型对计算机病毒和蠕虫的传播进行了一系列研究. Okamura 等 [4] 基于病毒传播的节点行为建立随机模型, 研究蠕虫传播的概率行为. Zou 等 [5] 分析 E-mail 网络拓扑对 E-mail 病毒传播行为的影响, 提出了 E-mail 病毒的传播模型, 并且基于人为对抗和网络拥塞等因素提出了蠕虫模型 [6]. 宋玉蓉和蒋国平 [7] 基于元胞自动机对复杂网络恶意软件传播动力学行为进行了一系列研究, 但是其模型忽视了现实中节点在健康状态也可通过人工途径获得免疫的这一事实, 并且该文章中用来验证和研究所建立模型的模拟仿真网络规模较小, 无法充分反映出真实复杂网络中的传播动力学行为. 王亚奇和蒋国平 [8,9] 研究了传播延迟对复杂网络病毒传播过程的影响. Jin 等 [10] 基于元胞自动机提出一种新的传播模型, 研究了病毒的流行程度与元胞邻居大小之间的关系. 此外, 元胞自动机作为研究复杂系统的有效工具, 不仅能够预测病毒的传播趋势, 还能够表现传播中的概率事件, 因此越来越受到人们的重视 [11−17].

本文参照文献 [7] 的方法, 基于元胞自动机建立复杂信息系统中的安全风险传播模型, 分析安全风险在多种网络拓扑下的概率传播行为. 模型按照免疫机理类型分为无免疫机理模型、随机免疫机理模型和熟人免疫机理模型, 针对复杂信息系统节点之间全局交互的特点, 建立各自元胞自动机的元胞空间、有限状态集、元胞邻域以及状态转换规则. 将提出的三种模型用于分析研究最近邻耦合网络 (nearest-neighbor coupled network, NC), Erdos-Renyi (ER) 随机网络 [18], Watts-Strogatz (WS) 小世界网络 [19] 和 Barabasi-Albert (BA) 无标度网络 [20] 等多种网络拓扑下的安全风险传播问题. 本文提出的模型不仅能反映复杂信息系统中安全风险传播的平均趋势, 而且可以描述安全风险的消亡等稀有概率事件, 有效地克服了基于马尔可夫链建立的随机模型不适合描述安全风险传播动态演化特征的缺陷 [7], 也克服了基于平均场方法的确定性模型只能够反映安全风险传播的平均趋势, 只适合对安全风险传播做整体预测的局限性 [7]. 本文还在多种网络拓扑中对安全风险的传播阈值、传播演化以及传播消亡等指标和过程进行了比较研究, 分析影响安全风险传播的关键要素, 对复杂信息系统中安全风险传播的控制和防范具有参考价值和指导意义.

2 安全风险随机传播模型

本文使用元胞自动机建立复杂信息系统安全风险传播的随机模型. 一个元胞自动机可以通过一个四元组来表示:

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

式中 C 表示元胞空间, Q 表示有限状态集, V 表示元胞邻域, f 表示状态转换规则. 考虑免疫机理的类型, 本文建立了三种模型, 分别是: 无免疫机理模型、随机免疫机理模型和熟人免疫机理模型.

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安全状态即健康状态 (susceptible) 和风险状态即感染状态 (infected) 之一, 节点状态变换关系如图 1 所示.

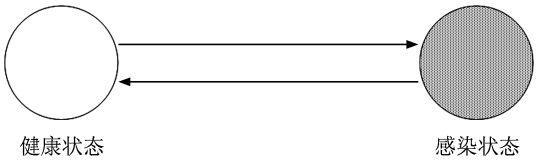

图 1 无免疫机理模型状态转换图

其中健康状态表示节点在复杂信息系统网络中不具有安全风险; 感染状态表示节点在复杂信息系统网络中具有安全风险. 本文将这种模型命名为 Risk-SIS 模型, 简称 R-SIS 模型.

考虑复杂信息系统网络 $G = (N, E)$, 其中 N 表示网络中所有节点的个数, E 表示网络中所有边的集合. 令 $\boldsymbol{A}$ 表示网络 G 的邻接矩阵, 来反映网络的拓扑结构. 根据元胞自动机的四个模型要素, 建立 R-SIS 模型如下.

元胞空间 C: 建立包含 N 个元胞的一维元胞空间, 空间中的一个元胞表示复杂信息系统网络中的一个节点.

有限状态集 Q: R-SIS 模型考虑节点的健康状态和感染状态. 令 $Q = \{0, 1\}$, 0 表示健康状态, 1 表示感染状态. 节点 i 在时刻 t 的状态变量用 $s_i(t)$ ($s_i(t) \in Q$) 表示, 则有:

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

元胞邻域 V: R-SIS 模型中以网络的邻接矩阵 $\boldsymbol{A}$ 来定义元胞之间的邻居关系. 邻接矩阵 $\boldsymbol{A}$ 中的第 i 行的向量表示节点 i 的邻居 V_i, 即 $V_i=\{a_{ij}|a_{ij}\in\boldsymbol{A},\ j=1,2,\cdots,N\}$. 若 $a_{ij}=1$, 表示节点 i 和 j 之间存在通信连接.

状态转换规则 f: R-SIS 模型中任何节点仅能被其邻居感染, 节点 i 在离散时间 t 的状态 $s_i(t)$ 取决于节点 i 在上一时刻 (时刻 $t-1$) 的自身状态 $s_i(t-1)$ 和其邻居的状态 $sv_i(t-1)$. 每个时间间隔内感染节点以概率 α 感染其邻居, 同时每个单位时间内感染节点也以概率 β 恢复健康. 状态转换规则如下所示:

$$s_i(t+1)=\begin{cases}\overline{s_i(t)} & g>0\\ s_i(t) & g\leqslant 0\end{cases},\tag{3}$$

其中上横线表示取反操作; g 为状态转换判断函数, 具体定义如下:

$$g=\overline{s_i(t)}\left(1-(1-\alpha)^{m_i(t)}-r\right)+s_i(t)\,(\beta-r),\tag{4}$$

(4) 式中考虑到节点 i 在时刻 t 分别为健康状态和感染状态的两种情况. 当节点 i 在时刻 t 为健康状态时, $s_i(t)=0$, 则 (4) 式中的第一项起作用, 即

$$g=1-(1-\alpha)^{m_i(t)}-r,\tag{5}$$

(5) 式用来判断在时刻 t 处于健康状态的节点 i, 经过一个离散时间后状态是否改变. 模型中一个健康节点受到感染的概率随着与其相邻的感染节点数量的增加而增加. 在时刻 t 处于健康状态的节点 i 在下一时刻受到感染的概率为 $1-(1-\alpha)^{m_i(t)}$, 其中 $m_i(t)$ 表示在时刻 t 与节点 i 相邻的感染节点的数量:

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

(5) 式中的 r 为 (0, 1) 之间的随机数, 用来与时刻 t 处于健康状态的节点 i 在下一时刻受到感染的概率 $1-(1-\alpha)^{m_i(t)}$ 进行比较, 判断节点 i 是否会受到感染: 若 $1-(1-\alpha)^{m_i(t)}$ 大于 r, 即 (5) 式中 $g>0$, 则节点 i 受到感染, 状态改变, $s_i(t+1)=\overline{s_i(t)}$; 反之若 $1-(1-\alpha)^{m_i(t)}$ 小于等于 r, 即 (5) 式中 $g\leqslant 0$, 则节点 i 没有受到感染, 保持健康状态, $s_i(t+1)=s_i(t)$.

因此, 在时刻 t 为健康状态的节点 i ($s_i(t)=0$) 的状态转换规则符合 (3) 式中定义的规则 (结论 1).

当节点 i 在时刻 t 为感染状态时, $s_i(t)=1$, 则 (4) 式中的第二项起作用, 即

$$g=\beta-r,\tag{7}$$

(7) 式用来判断在时刻 t 处于感染状态的节点 i, 经过一个离散时间后状态是否改变. 模型中一个感染节点以概率 β 恢复健康, (7) 式中的 r 为 (0, 1) 之间的随机数, 用来与时刻 t 处于感染状态的节点 i 在下一时刻恢复健康的概率 β 进行比较, 判断节点 i 是否会恢复健康: 若 β 大于 r, 即 (7) 式中 $g>0$, 则节点 i 恢复健康, 状态改变, $s_i(t+1)=\overline{s_i(t)}$; 反之若 β 小于等于 r, 即 (7) 式中 $g\leqslant 0$, 则节点 i 没有恢复健康, 保持感染状态, $s_i(t+1)=s_i(t)$.

因此, 在时刻 t 为感染状态的节点 i ($s_i(t)=1$) 的状态转换规则符合 (3) 式中定义的规则 (结论 2).

由结论 1 和结论 2 可知, R-SIS 模型中节点 i 的状态转换规则符合 (3) 式中定义的状态转换规则.

用 $S(t)$ 表示在时刻 t 网络中健康节点在所有节点中所占的比率, 用 $I(t)$ 表示在时刻 t 网络中感染节点在所有节点中所占的比率, 那么模型中有以下结果 [7]:

$$I(t)=\frac{1}{N}\sum_{i=1}^{N}s_i(t),\tag{8}$$

$$S(t)+I(t)=1.\tag{9}$$

2.2 随机免疫机理模型

随机免疫是指在网络中随机地选取一部分节点进行免疫. 引入随机免疫机理后的复杂信息系统安全风险传播模型中, 节点的状态分为健康状态、感染状态和免疫状态, 节点状态变换关系如图 2 所示.

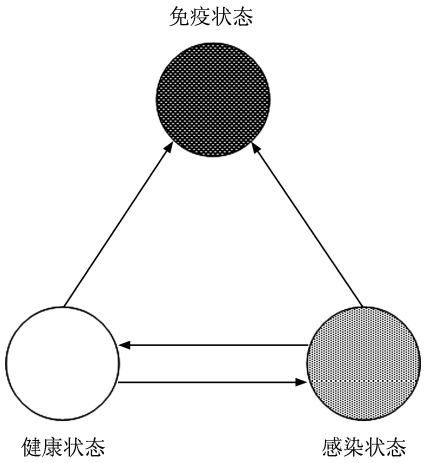

图 2 随机免疫机理模型状态转换图

其中健康状态表示节点在复杂信息系统网络中没有安全风险; 感染状态表示节点在复杂信息系统网络中具有安全风险; 免疫状态表示节点在复杂

信息系统网络中因为进行了安全隔离或者采取了持续的安全防护措施, 从而免于感染安全风险. 本文将这种模型命名为 Risk-SIR-Random 模型, 简称 R-SIR-R 模型. R-SIR-R 模型基于随机免疫机理和生物传染病学中的 SIR 模型建立, 在传播的初始状态 (时刻 0) 对网络中的各节点以一定概率进行随机免疫, 之后依照 SIR 模型, 节点在治愈的同时以一定概率获得免疫.

需要注意的是, 免疫状态的节点不会再被感染节点传染安全风险, 也不会向健康节点传染安全风险. 因此一旦节点被免疫, 免疫节点所连接的边就应该从网络中去除 [7], 同时需要修改网络 G 的邻接矩阵 $\boldsymbol{A}$. 根据元胞自动机的四个模型要素, 建立 R-SIR-R 模型如下.

元胞空间 C: 建立包含 N 个元胞的一维元胞空间, 空间中的一个元胞表示复杂信息系统网络中的一个节点.

有限状态集 Q: R-SIR-R 模型考虑节点的健康状态、感染状态和免疫状态. 令 $Q=\{(0,0),(1,0),(0,1),(1,1)\}$, (0, 0) 表示健康状态; (1, 0) 表示感染状态; (0, 1) 表示免疫状态 [7]. 节点 i 在时刻 t 的状态变量用向量 $\boldsymbol{s}_i(t)$ ($\boldsymbol{s}_i(t)\in Q$) 表示, 向量 $\boldsymbol{s}_i(t)$ 包含两个分量: $\boldsymbol{s}_i(t)=(s_{ix}(t),s_{iy}(t))$, 其中分量 $s_{ix}(t)$ 用来表示节点是否感染, 分量 $s_{iy}(t)$ 用来表示节点是否免疫, 则有 [7]:

$$\boldsymbol{s}_i(t)=\begin{cases}(0,0) & \text{节点 } i \text{ 在时刻 } t \text{ 状态为 susceptible}\\ (1,0) & \text{节点 } i \text{ 在时刻 } t \text{ 状态为 infected}\\ (0,1) & \text{节点 } i \text{ 在时刻 } t \text{ 状态为 removed}\\ (1,1) & \text{状态不存在}\end{cases}. \tag{10}$$

元胞邻域 V: R-SIR-R 模型中以网络在时刻 t 的邻接矩阵 $\boldsymbol{A}(t)$ 来定义在时刻 t 元胞之间的邻居关系. 邻接矩阵 $\boldsymbol{A}(t)$ 中的第 i 行的向量表示节点 i 在时刻 t 的邻居 $\boldsymbol{V}_i(t)$, 即 $\boldsymbol{V}_i(t)=\left\{a_{ij}(t)\middle|a_{ij}(t)\in \boldsymbol{A}(t),\ j=1,2,\cdots,N\right\}$. 若 $a_{ij}(t)=1$, 表示节点 i 和 j 之间在时刻 t 存在可传播安全风险的连接.

状态转换规则 f: 在 R-SIR-R 模型中, 状态转换规则分为两个部分. 第一部分是在初始阶段 (即时刻 $t=0$), 对网络中的各节点以概率 γ 进行随机免疫. 第一部分的状态转换规则如下所示:

$$\begin{aligned}&s_{ix}(0)=0,\\ &s_{iy}(0)=\begin{cases}1, & g>0\\ 0, & g\leqslant 0\end{cases},\end{aligned} \tag{11}$$

其中 g 为状态转换判断函数, 定义如下:

$$g=\gamma-r. \tag{12}$$

(11) 和 (12) 式用来判断节点 i 在初始阶段是否被选中接受免疫. (12) 式中的 r 为 (0, 1) 之间的随机数, 用来与免疫概率 γ 进行比较, 判断节点 i 是否转换为免疫状态: 若 γ 大于 r, 即 (12) 式中 $g>0$, 则节点 i 转换为免疫状态, $s_{ix}(0)=0$, $s_{iy}(0)=1$, $s_i(0)=(0,1)$; 反之若 γ 小于等于 r, 即 (12) 式中 $g\leqslant 0$, 则节点 i 没有免疫, 保持健康状态, $s_{ix}(0)=0$, $s_{iy}(0)=0$, $s_i(0)=(0,0)$. 同时需要修改网络 G 的初始邻接矩阵 $\boldsymbol{A}(0)$, 若 $s_i(0)=(0,1)$, 则 $a_{ij}(0)=a_{ji}(0)=0$, $j=1,2,\cdots,N$.

状态转换规则的第二部分用于传播的演化阶段 (时刻 $t>0$), 在演化阶段中, 每个时间间隔内感染节点以概率 α 感染它的邻居, 同时每个单位时间内感染节点也以概率 β 恢复健康; 感染节点若在下一时刻恢复健康, 则该节点在恢复健康的同时以概率 δ 获得免疫. 第二部分的状态装换规则如下所示:

$$\begin{aligned}s_{ix}(t+1)&=\begin{cases}\overline{s_{ix}(t)} & g_x>0\wedge s_{iy}(t)=0\\ s_{ix}(t) & g_x\leqslant 0\wedge s_{iy}(t)=0\\ 0 & s_{iy}(t)=1\end{cases},\\ s_{iy}(t+1)&=\begin{cases}1 & g_y>0\vee s_{iy}(t)=1\\ 0 & g_y\leqslant 0\wedge s_{iy}(t)=0\end{cases},\end{aligned} \tag{13}$$

当节点 i 时刻 t 的状态为免疫状态, 即 $s_i(t)=(0,1)$ 时, $s_i(t+1)=(0,1)$, 表示节点一旦免疫, 就保持免疫状态不变. (13) 式中上横线表示取反操作; g_x, g_y 为状态转换判断函数, 具体定义如下:

$$\begin{aligned}g_x=&\overline{s_{ix}(t)}\left(1-(1-\alpha)^{m_i(t)}-r\right)\\ &+s_{ix}(t)(\beta-r),\end{aligned} \tag{14}$$

$$g_y=s_{ix}(t)\overline{s_{ix}(t+1)}(\delta-r). \tag{15}$$

(14) 式的解释同 (4) 式类同, 需要做细微调整 [7] 的地方是 $m_i(t)$. (14) 式中用来表示在时刻 t 与节点 i 相邻的感染节点的数量 $m_i(t)$ 调整为

$$m_i(t)=\sum_{j=1}^{N}a_{ij}(t)s_{jx}(t), \tag{16}$$

并且若 $\boldsymbol{s}_i(t)=(0,1)$, 则有 $a_{ij}(t)=a_{ji}(t)=0, j=1,2,\cdots,N$.

(15) 式用来判断在时刻 t 为感染状态并在时刻 t+1 恢复健康的节点 i 是否转换为免疫状态, 其中 r 为 (0, 1) 之间的随机数, 用来与免疫概率 δ 进

行比较. (15) 式中当 $s_{ix}(t)$ 为 1, $s_{ix}(t+1)$ 为 0 并且 $(\delta - r) > 0$ 时, $g_y > 0$, 表示节点 i 在时刻 t 为感染状态, 在时刻 $t+1$ 恢复健康并获得免疫; 其他情况 g_y 均小于等于 0, 表示节点 i 没有获得免疫.

用 $S(t)$ 表示在时刻 t 网络中健康节点在所有节点中所占的比率, $I(t)$ 表示在时刻 t 网络中感染节点在所有节点中所占的比率, $R(t)$ 表示在时刻 t 网络中免疫节点在所有节点中所占的比率, 那么模型中有以下结果 [7]:

$$I(t) = \frac{1}{N}\sum_{i=1}^{N} s_{ix}(t), \tag{17}$$

$$R(t) = \frac{1}{N}\sum_{i=1}^{N} s_{iy}(t), \tag{18}$$

$$S(t) + I(t) + R(t) = 1. \tag{19}$$

2.3 熟人免疫机理模型

熟人免疫是指在网络中随机地选出一定比例的节点, 再从每一个被选出的节点中随机选择一个邻居节点进行免疫. 引入熟人免疫机理后的复杂信息系统安全风险传播模型中, 节点的状态分为健康状态、感染状态和免疫状态, 节点状态变换关系如图 3 所示.

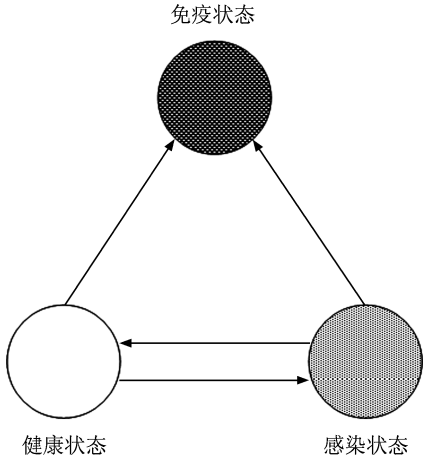

图 3 熟人免疫机理模型状态转换图

熟人免疫机理模型中各节点状态表示的意义与 R-SIR-R 模型中的说明相同. 本文将这种模型命名为 Risk-SIR-Acquaintance 模型, 简称 R-SIR-A 模型. R-SIR-A 模型基于熟人免疫机理和生物传染病学中的 SIR 模型建立, 在传播的初始状态 (时刻 0) 对网络进行熟人免疫, 之后依照 SIR 模型, 节点在治愈的同时以一定概率获得免疫.

同样, 免疫状态的节点不会再被感染节点传染安全风险, 也不会向健康节点传染安全风险. 因此 R-SIR-A 模型中一旦节点被免疫, 免疫节点所连接的边就应该从网络中去除, 同时需要修改网络 G 的邻接矩阵 $\boldsymbol{A}$. 根据元胞自动机的四个模型要素, 建立 R-SIR-A 模型. 其中元胞空间 C、有限状态集 Q 以及元胞邻域 V 的定义与 R-SIR-R 模型相同. 下面定义 R-SIR-A 模型中的状态转换规则 f.

状态转换规则 f: 在 R-SIR-A 模型中, 状态转换规则同样分为两个部分. 第一部分是在初始阶段 (即时刻 $t=0$), 在网络中以概率 p 随机的选择节点, 然后从每一个被选出的节点中随机选择一个邻居节点进行免疫. 第一部分的状态转换规则如下.

首先对于每一个节点 i 判断是否以概率 p 被选中, 用 h_i 来表示: $h_i = 1$ 表示节点 i 被选中, $h_i = 0$ 表示节点 i 没有被选中. 定义 h_i 的规则如下:

$$h_i = \begin{cases} 1 & p - r > 0 \\ 0 & p - r \leqslant 0 \end{cases}, \tag{20}$$

其中 r 为 (0, 1) 之间的随机数, 用来与概率 p 进行比较, 判断节点 i 是否被选中.

若 $h_i = 1$, 表示节点 i 被选中, 接下来需要在节点 i 的邻居中随机选择一个进行免疫. 由模型中元胞邻域 V 的定义可知, 向量 $\boldsymbol{V}_i(0) = \{a_{ij}(0) \,|\, a_{ij}(0) \in \boldsymbol{A}(0), j = 1, 2, \cdots, N\}$ 表示节点 i 在初始时刻和其他节点的邻接状态. 集合 $v_i(0)$ 表示初始时刻节点 i 的所有邻居的集合, $v_i(0) = \{j \,|\, a_{ij}(0) = 1,\ a_{ij}(0) \in \boldsymbol{A}(0)\}$. 若 $v_i(0) = \emptyset$, 则跳过节点 i, 继续对其他节点进行是否选中的判断. 若 $v_i(0) \neq \emptyset$, 则从 $v_i(0)$ 中随机选取一个元素 j_r 进行免疫, 使 $s_{j_r}(0) = (0, 1)$.

若 $h_i = 0$, 表示节点 i 没有被选中, 则跳过节点 i, 继续对其他节点进行是否选中的判断.

同时需要修改网络 G 的初始邻接矩阵 $\boldsymbol{A}(0)$, 若 $s_i(0) = (0, 1)$, 则 $a_{ij}(0) = a_{ji}(0) = 0$, $(j = 1, 2, \cdots, N)$.

状态转换规则的第二部分用于传播的演化阶段 (时刻 $t > 0$). 在演化阶段中, 每个时间间隔内感染节点以概率 α 感染它的邻居, 同时每个单位时间内感染节点也以概率 β 恢复健康; 感染节点若在下一时刻恢复健康, 则该节点在恢复健康的同时以概率 δ 获得免疫. R-SIR-A 模型中状态转换规则的第二部分与 R-SIR-R 模型中演化阶段的状态转换规则相同, 这里就不再赘述.

用 $S(t)$ 表示在时刻 t 网络中健康节点在所有节点中所占的比率, $I(t)$ 表示在时刻 t 网络中感染节点在所有节点中所占的比率, $R(t)$ 表示在时刻 t 网络中免疫节点在所有节点中所占的比率, R-SIR-A 模型也具有 R-SIR-R 模型中 (17)—(19) 式的结果.

3 传播阈值研究

本文参照文献 [7] 的方法对 R-SIS 模型进行验证, 在其基础上改进了网络生成参数, 使得仿真验证环境更加符合复杂信息系统中复杂网络拓扑的实际情况. 通过仿真的方式研究 R-SIS 模型在 NC, ER, WS 和 BA 四种网络拓扑下感染率与传播率 $\lambda = \alpha/\beta$ 之间的对应关系, 与传统 SIS 模型进行比较分析; 研究 R-SIS 模型在四种网络拓扑下的风险传播阈值, 与三种常见的方法 [3,21,22] 得到的传播阈值进行比较分析.

用来比较的三种常见的传播阈值计算方法分别为

$$\lambda_1 = 1/\langle k\rangle, \tag{21}$$

$$\lambda_2 = \langle k\rangle/\langle k^2\rangle, \tag{22}$$

$$\lambda_3 = 1/\tau_{1,\boldsymbol{A}}, \tag{23}$$

(21) 式中 $\langle k\rangle$ 表示网络中各节点的平均度 [3]; (22) 式中 $\langle k\rangle$ 表示网络中各节点的平均度; $\langle k^2\rangle$ 表示网络的平均均方度, 反映网络的异质化程度, 对于均匀网络 $\langle k^2\rangle = \langle k\rangle^2$, 节点度分布越不均匀 $\langle k^2\rangle$ 值越高 [21]; (23) 式中 $\tau_{1,\boldsymbol{A}}$ 表示网络邻接矩阵 $\boldsymbol{A}$ 的最大特征根 [22].

根据 NC 网络定义、ER 随机网络模拟算法 [23], WS 小世界网络模拟算法 [19] 和 BA 无标度网络模拟算法 [20], 生成这四种典型的网络, 各网络参数如表 1 所示.

表 1 仿真网络参数

参数类型	NC 网络	ER 随机网络	WS 小世界网络	BA 无标度网络
节点数量 N	1000	1000	1000	1000
生成参数	—	$p_{\rm er} = 6/999$	$p_{\rm ws} = 0.2$	$m_0 = 8$, $m = 3$, $p_0 = 6/7$
平均度 $\langle k\rangle$	6	6.0140	6	6.0040
平均均方度 $\langle k^2\rangle$	36	42.4320	37.0700	86.7560
聚类系数 C	0.6000	0.0060	0.3072	0.0326
邻接矩阵最大特征根 $\tau_{1,\boldsymbol{A}}$	6	7.2274	6.2373	15.1722

表 1 中 $p_{\rm er}$ 表示 ER 随机图中两个顶点之间有边的概率; $p_{\rm ws}$ 表示 WS 小世界网络构造中的随机重连概率.

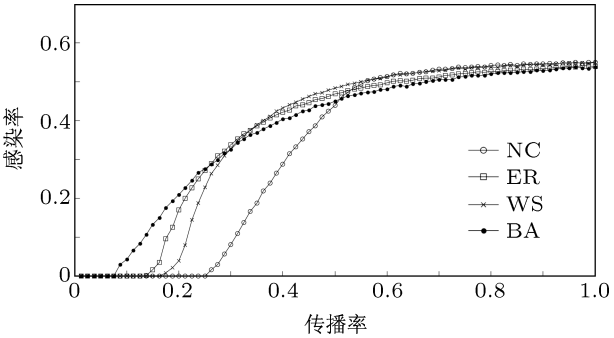

图 4 R-SIS 模型感染率与传播率的对应关系

在这四种仿真网络中研究 R-SIS 模型的传播阈值, 传播参数的取值为: α 初始值为 0, 以步长 0.01 线性增长到 0.8; $\beta = 0.8$; $I(0) = 1/N$. 已知传播率 $\lambda = \alpha/\beta$, 在以上传播参数的设置下运行 R-SIS 模型, 统计运行 100 次中时刻 $t = 250$ 时的感染率的平均值, 建立感染率与传播率 λ 的对应关系, 如图 4 所示.

从图 4 中可以看出, R-SIS 模型在四种网络拓扑中的传播阈值大小情况为: BA < ER < WS < NC. 与表 1 中的参数进行比较, 可以看出 $(\langle k^2\rangle_{\rm BA}) > (\langle k^2\rangle_{\rm ER}) > (\langle k^2\rangle_{\rm WS}) > (\langle k^2\rangle_{\rm NC})$, 说明网络拓扑中度分布的异质化程度越高, 传播阈值越小, 符合文献 [7]. 并且图 4 与传统 SIS 模型感染率与传播率的对应关系曲线 [24] 有着相似的符合, 这点也与文献 [7] 一致. 图 4 中 NC 网络有着最大的传播阈值, 但随着传播率 λ 的增大, NC 网络逐渐有了较大的感染率. 说明安全风险在 NC 网络中最不容易爆发. 此现象是由 NC 网络的拓扑结构所决定的: NC 网络中节点之间的交互具有很强的局域性, 导致风险在扩散过程中能有效感染的邻居数目受到局域交互特征的限制, 因此传播阈值较大. 而同样是因为这种稳定而规则的局域特性, 使得传播率较大时 NC 网络中的风险影响规模依然能够随着传播率的逐渐增大而不断扩大, 最终感染率较

高. 图 4 中 BA 网络有着最小的传播阈值, 但随着传播率 λ 的增大, BA 网络的感染率变得最小. 说明安全风险在 BA 网络拓扑中最容易爆发, 但传播率较大时, 影响规模较小. 此现象是由 BA 网络拓扑的无标度特性所决定的: 在传播率较低时, 度数很大的部分节点依然会比较容易受到感染, 进而感染其邻居节点, 导致安全风险的爆发. 因此 BA 网络中的传播阈值较低. 但是当传播率较高时, BA 网络中度数小的部分节点受到感染的概率依然比较低, 因此 BA 网络中安全风险最终的影响规模相对较小.

表 2 统计了四种网络拓扑下三种常见的传播阈值计算方法的结果以及 R-SIS 模型中的传播阈值.

表 2 传播阈值比较

传播阈值	NC 网络	ER 随机网络	WS 小世界网络	BA 无标度网络
$\lambda_1 = 1/\langle k\rangle$	0.1667	0.1663	0.1667	—
$\lambda_2 = \langle k\rangle/\langle k^2\rangle$	0.1667	0.1417	0.1619	0.0692
$\lambda_3 = 1/\tau_{1,A}$	0.1667	0.1384	0.1603	0.0659
$\lambda_{\text{R-SIS}}$	0.2625	0.1375	0.1750	0.0875
$\Delta\lambda = \lvert\lambda_{\text{R-SIS}} - \lambda_2\rvert$	0.0958	0.0042	0.0131	0.0183
$\Delta\lambda/\lambda_2$	57.47%	2.96%	8.09%	26.45%

表 2 将已有的三种方法得到的传播阈值与 R-SIS 模型得到的仿真阈值在数值上进行比较, 可以看出 ER 网络、WS 网络与 BA 网络中的阈值与第二种方法得到的阈值差异较小, 其中 BA 网络中的阈值与第二种方法在数量上相差并不大, 但是因为 BA 网络中的阈值较小, 因此差异的百分比显得较大. 而 NC 网络中的阈值与第二种方法得到的阈值差异相对较大, 这是因为虽然 NC 网络中的度分布十分均匀, 但其节点的局域交互特性使得 NC 网络中传播阈值显著增加 [7].

4 传播演化研究

安全风险在复杂信息系统中一旦流行, 其传播速度及影响规模会受到多种因素的影响. 在这一部分本文侧重研究四种不同的网络拓扑对安全风险传播的影响. 图 5, 图 6 和图 7 分别描述了 R-SIS 模型、R-SIR-R 模型以及 R-SIR-A 模型下安全风险在 NC 网络、ER 网络、WS 网络和 BA 网络中的传播趋势及影响规模, 网络参数仍如表 1 所示.

在图 5 中, R-SIS 模型安全风险传播稳定后的感染率 BA < ER < WS < NC, 说明网络拓扑中度分布的异质化程度越高, 安全风险传播稳定后的影响范围越小, 符合文献 [7]. 这是因为网络拓扑中度分布的异质化程度越高, 网络中就存在着更多度数非常小的节点. 这些度数小的节点被感染的概率很小, 不容易受到安全风险的影响. 因此度分布的异质化程度越高, 这些不易被影响到的节点越多, 安全风险最终影响到的范围就越小. 在图 5, 图 6 和图 7 中, 安全风险传播速度 BA > ER > WS > NC, 说明网络拓扑中度分布的异质化程度越高, 安全风险传播的速度越快, 符合文献 [7] 中的结论. 这种现象是因为网络拓扑中度分布的异质化程度越高, 网络中就存在着越多度数非常大的节点. 这些节点比较容易受到感染, 进而使得其相邻的节点也容易受到感染, 导致安全风险的传播速度较快.

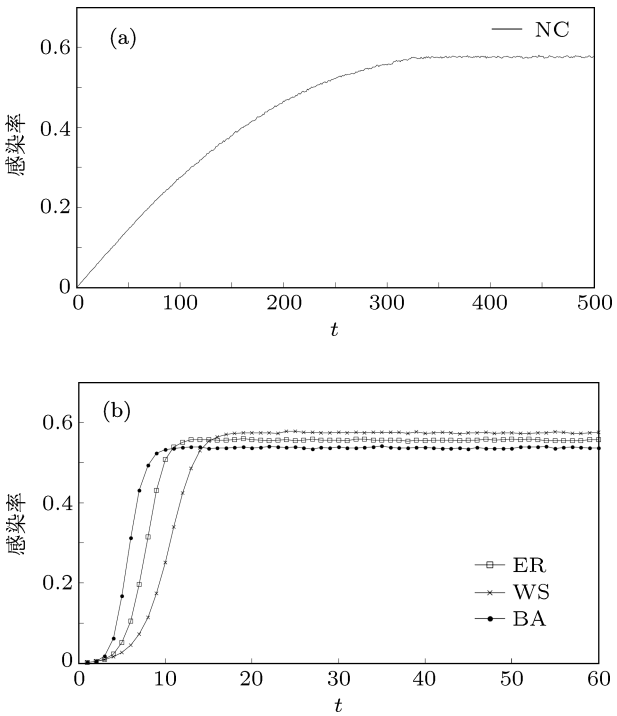

图 5 R-SIS 模型安全风险传播趋势 ($\alpha = 0.3$; $\beta = 0.5$; $I(0) = 2/N$) (a) R-SIS 模型中安全风险在 NC 网络中的传播趋势; (b) R-SIS 模型中安全风险在 ER 网络、WS 网络和 BA 网络中的传播趋势

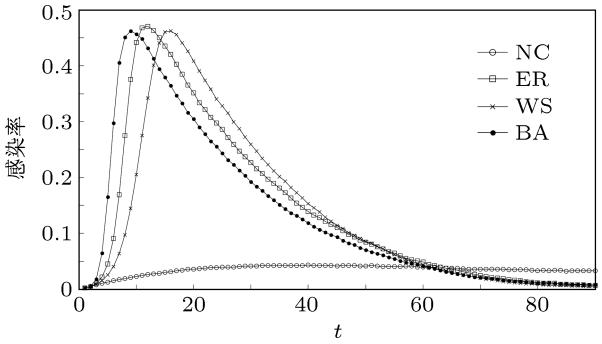

图 6 R-SIR-R 模型安全风险传播趋势 ($\alpha = 0.3$; $\beta = 0.5$; $\gamma = 0.01$; $\delta = 0.1$; $I(0) = 2/N$)

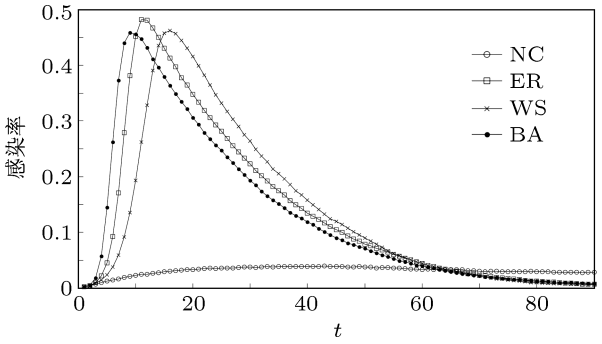

图 7 R-SIR-A 模型安全风险传播趋势 ($\alpha = 0.3$; $\beta = 0.5$; $p = 0.01$; $\delta = 0.1$; $I(0) = 2/N$)

从图 6 和图 7 中 R-SIR-R 模型与 R-SIR-A 模型安全风险传播趋势可以看出, 在考虑免疫机制的条件下, NC 网络风险消亡的速度非常缓慢, 但是免疫机制能够将 NC 网络中安全风险的影响控制在一个很小的范围内. 这是因为模型中的免疫机制能够将 NC 网络演化成一种类似于群落网络的结构, 这种拓扑结构中群落内的连接较多而群落之间的连接很少. 安全风险不容易从一个群落传播到其他群落, 因此能够将安全风险的影响控制在一个很小的范围内. 但是安全风险在各个群落内依然存在并传播, 因此风险消亡的速度非常缓慢. 在另外三种网络拓扑中, 安全风险消亡的速度 BA < ER < WS, 产生这种现象的原因是网络度分布的异质化程度越高, 模型中的免疫机制将度数大的节点免疫的概率越低, 而度数很大的节点容易受到其相邻节点的感染并且容易将其相邻节点感染, 对安全风险的传播非常有利. 因此网络拓扑中度分布的异质化程度越高, 模型后期的随机免疫机制对控制安全风险传播的效果越缓慢. 将图 5 与图 6 和图 7 对比可以看出, 图 6 和图 7 中各图像感染率的最大值小于图 5, 这是因为模型中的免疫机制通过免疫部分节点, 降低了安全风险的影响范围, 也说明在相同的传播参数下, R-SIR-R 模型与 R-SIR-A 模型中的免疫机制可以有效地控制安全风险影响的范围.

5 传播最早期的消亡情况研究

文献 [7] 在四种网络拓扑中对病毒在传播最早期就趋于消亡的情况进行了研究, 指出了恶意软件在传播之初就趋于消亡情形出现的比率与 $I(0)$ 之间呈现负指数关系. 本文接下来对三种模型在四种网络拓扑中安全风险在传播最早期就趋于消亡的情况进行研究, 分析安全风险在传播之初就趋于消亡情形出现的比率与传播率 λ 之间的关系. 图 8, 图 9 和图 10 分别反映了在 R-SIS, R-SIR-R 和 R-SIR-A 模型下安全风险在传播之初就趋于消亡情形出现的消亡率与传播率 λ 之间的关系曲线.

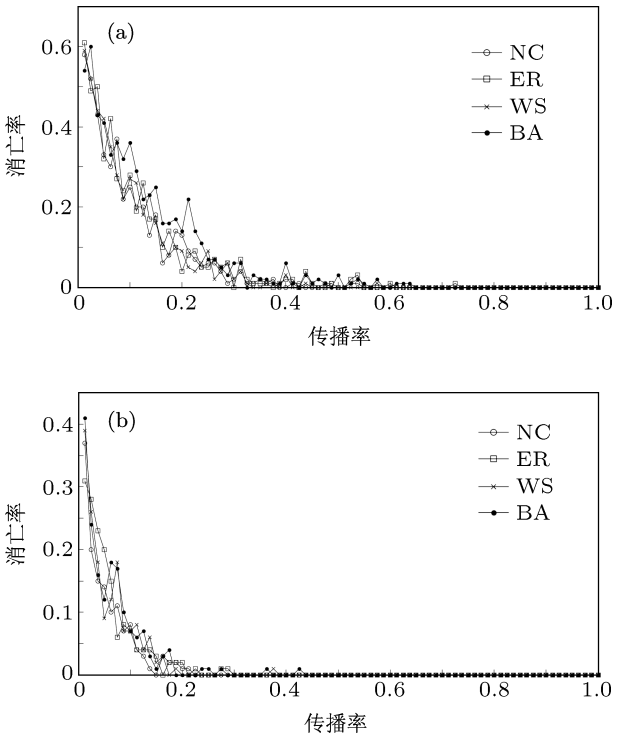

图 8 R-SIS 模型安全风险在传播之初就趋于消亡的消亡率与传播率之间的关系 (a) $I(0) = 2/N$; (b) $I(0) = 4/N$

由图 8, 图 9 和图 10 中的仿真结果可以分析得出, 安全风险在传播之初就趋于消亡的消亡率与传播率之间呈现一种近似于负指数的关系. 安全风险在传播之初就趋于消亡的现象产生的原因是传播之初的感染源没有通过概率 α 感染到其相邻节点, 并且感染源自身以概率 β 恢复了健康, 造成安全风险在传播之初的消亡. 所以 α 越小, β 越大, 消亡率越大, 而传播率 $\lambda = \alpha/\beta$. 因此可得出结论: 随着传播率的增大, 安全风险在传播之初就趋于消亡的消亡率越低. 此外，在图 8, 图 9 和图 10 中通过改变 $I(0)$, 可以得出随着 $I(0)$ 的增大, 消亡率降低的结论, 此结论与文献 [7] 中的研究结果相符.

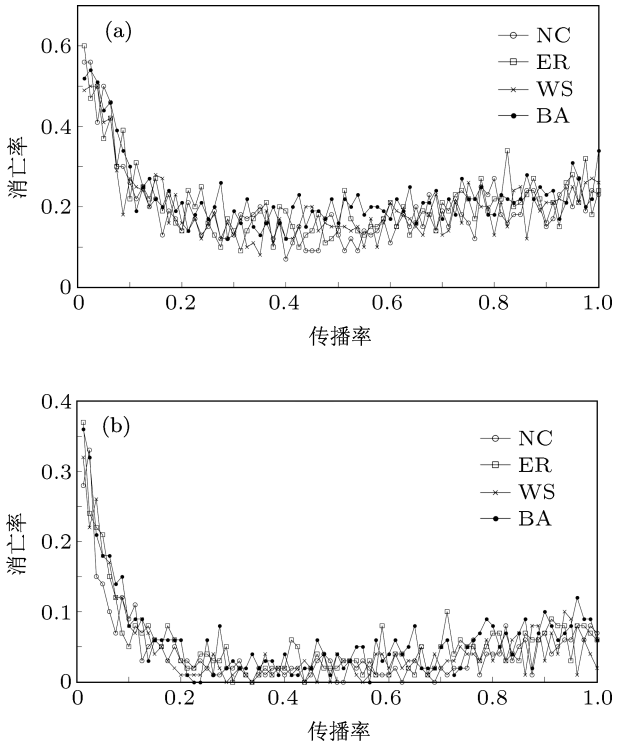

图 9 R-SIR-R 模型安全风险在传播之初就趋于消亡的消亡率与传播率之间的关系 (a) $\gamma = 0.01$; $\delta = 0.1$; $I(0) = 2/N$; (b) $\gamma = 0.01$; $\delta = 0.1$; $I(0) = 4/N$

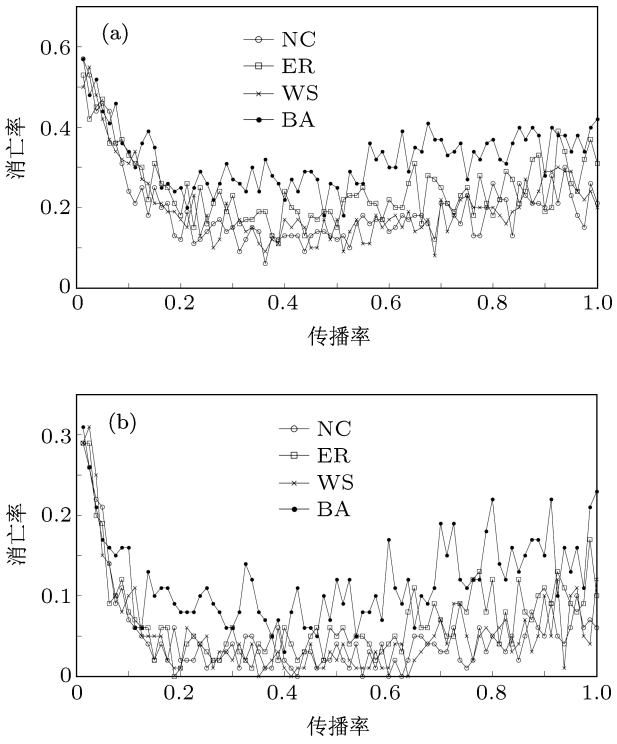

图 10 R-SIR-A 模型安全风险在传播之初就趋于消亡的消亡率与传播率之间的关系 (a) 中 $p = 0.01$; $\delta = 0.1$; $I(0) = 2/N$; (b) $p = 0.01$; $\delta = 0.1$; $I(0) = 4/N$

6 结 论

本文结合复杂信息系统安全风险传播的特点, 使用元胞自动机建立复杂信息系统安全风险传播的三种模型: 无免疫机理 R-SIS 模型, 随机免疫机理 R-SIR-R 模型以及熟人免疫机理 R-SIR-A 模型. 基于复杂信息系统网状结构中节点交互的全局特性, 扩充传统元胞自动机的四元组, 提出复杂信息系统安全风险传播中节点的状态转换规则, 研究在 NC 网络、ER 网络、WS 网络和 BA 网络四种网络拓扑下安全风险传播的概率行为. 通过研究 R-SIS 模型在四种网络拓扑下安全风险的传播阈值, 与现有的传播阈值研究成果进行比较, 所得的传播阈值符合现有研究成果, 验证了 R-SIS 模型的正确性; 并分析验证网络拓扑结构中度分布的异质化程度越高, 传播阈值越小的结论. 通过对 R-SIS 模型在四种网络拓扑下安全风险的传播演化趋势进行研究, 分析验证了网络度分布的异质化程度越高, 安全风险影响范围越小, 传播速度越快的结论; 并从 R-SIR-R 模型与 R-SIR-A 模型在四种网络拓扑下安全风险传播趋势得出网络拓扑中度分布的异质化程度越高, 模型后期的随机免疫机制对控制安全风险传播的效果越缓慢. 通过对安全风险在传播最早期就趋于消亡的情况进行研究, 分析得出安全风险在传播之初就趋于消亡的消亡率与传播率之间呈现一种近似于负指数的关系; 并且感染源 $I(0)$ 越多, 安全风险的消亡率越低. 本文通过研究多种网络拓扑下安全风险的传播阈值、传播演化以及传播消亡等过程, 分析了影响复杂信息系统安全风险传播的关键要素, 对复杂信息系统中安全风险传播的控制具有指导作用.

[1] Feng D, Zhang Y, Zhang Y Q 2004 *J. Commun.* **25** 10 (in Chinese) [冯登国, 张阳, 张玉清 2004 通信学报 **25** 10]

[2] Zhang Y Z, Fang B X, Chi Y, Yun X C 2007 *J. Software* **18** 137 (in Chinese) [张永铮, 方滨兴, 迟悦, 云晓春 2007 软件学报 **18** 137]

[3] Kephart J O, White S R, Chess D M 1993 *IEEE Spectrum* **30** 20

[4] Okamura H, Kobayashi H, Dohi T 2005 *Proceedings of the 16th IEEE International Symposium on Sof tware Reliability Engineering* Chicago, IL, USA, November 8–11, 2005 p149

[5] Zou C C, Towsley D, Gong W B 2007 *IEEE Trans. Depend. Secure Comput.* **4** 105

[6] Zou C C, Gong W, Towsley D 2002 *Proceedings of the 9th ACM Conference on Computer and Communications Security* Washington, DC,

USA, November 18–22, 2002 p10
[7] Song Y R, Jiang G P 2009 *Acta Phys. Sin.* **58** 5911 (in Chinese) [宋玉荣, 蒋国平 2009 物理学报 **58** 5911]
[8] Wang Y Q, Jiang G P 2010 *Acta Phys. Sin.* **59** 6724 (in Chinese) [王亚奇, 蒋国平 2010 物理学报 **59** 6724]
[9] Wang Y Q, Jiang G P 2011 *Acta Phys. Sin.* **60** 080510 (in Chinese) [王亚奇, 蒋国平 2011 物理学报 **60** 080510]
[10] Jin Z, Liu Q X, Mainul H 2007 *Chin. Phys.* **16** 1267
[11] Fuentes M A, Kuperman M N 1999 *Physica* A **267** 471
[12] Sirakoulis G C, Karafyllidis I, Thanailakis A 2000 *Ecol. Model.* **133** 209
[13] White S H, del Rey A M, Sanchez G R 2009 *Appl. Math. Sci.* **3** 959
[14] Gagliardi H, Alves D 2010 *Math. Popul. Stud.* **17** 79
[15] Shan X M, Liu F, Ren Y 2002 *Acta Phys. Sin.* **51** 1175 (in Chinese) [山秀明, 刘锋, 任勇 2002 物理学报 **51** 1175]
[16] Kong L J, Liu M R, Lü X Y 2001 *Acta Phys. Sin.* **50** 1255 (in Chinese) [孔令江, 刘慕仁, 吕晓阳 2001 物理学报 **50** 1255]
[17] Dai S Q, Dong L Y, Xue Y 2001 *Acta Phys. Sin.* **50** 445 (in Chinese) [戴世强, 董力耘, 薛郁 2001 物理学报 **50** 445]
[18] Erdos P, Rnyi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
[19] Watts D J, Strogatz S H 1998 *Nature* **393** 409
[20] Barabsi A L, Albert R 1999 *Science* **286** 509
[21] Pastor-Satorras R, Vespignani A 2002 *Phys. Rev.* E **65** 035108
[22] Chakrabarti D, Wang Y, Wang C 2007 *ACM Trans. Inform. Syst. Secur.* **10** 1
[23] Erdos P, Rnyi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
[24] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev.* E **63** 066117

Complex information system security risk propagation research based on cellular automata*

Li Zhao[1)2)†] Xu Guo-Ai[1)2)] Ban Xiao-Fang[3)] Zhang Yi[3)] Hu Zheng-Ming[1)2)]

1) (*Information Security Center, Beijing University of Posts and Telecommunications, Beijing* 100876*, China*)

2) (*National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing* 100876*, China*)

3) (*China Information Technology Security Evaluation Center, Beijing* 100085*, China*)

(Received 2 May 2013; revised manuscript received 17 June 2013)

Abstract

There models of complex information system security risk propagation are proposed in this paper based on cellular automata, and the probabilistic behaviors of security risk propagation in complex information systems are investigated by running the proposed models on nearest-neighbor coupled network, Erdos-Renyi random graph network, Watts-Strogatz small world network and Barabasi-Albert power law network respectively. Analysis and simulations show that the proposed models describe the behaviors of security risk propagation in the above four kinds of networks perfectly. By researching on the propagation threshold of security risks in four kinds of network topology and comparing with the existing research result, the correctness of the models is verified. The relationship between the heterogeneity of degree distribution and the value of the propagation threshold is analyzed and verified in this paper. Through the research on the evolutionary trends of security risk propagation, the relationship between the heterogeneity of degree distribution and the influence sphere and speed of security risk propagation is analyzed and verified as well. Meanwhile, the relationship between the heterogeneity of degree distribution and the effect of the immune mechanism on controlling security risk propagation is pointed out. Furthermore, the result of simulations describes the negative exponent relationship between security risk extinction rate and the propagation rate. The key factors affecting the security risk propagation are analyzed in this paper, providing the guidance for the control of security risk propagation in complex information systems.

Keywords: complex information system, complex network, security risk propagation, cellular automata

PACS: 02.50.−r, 05.65.+b, 05.70.Jk **DOI:** 10.7498/aps.62.200203

* Project supported by the National Natural Science Foundation of China (Grant Nos. 60970135, 61170282), the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20120005110017), and the National Key Technology Research and Development Program of the Ministry of Science and Technology of China (Grant No. 2012BAH06B02).

† Corresponding author. E-mail: lizhaoabc@gmail.com