doi:10.11959/j.issn.1000-436x.2015111

基于马尔可夫链的网络蠕虫传播模型

周翰逊[1]，郭薇[2]，刘建[1]，贾大宇[3]
(1. 辽宁大学 信息学院，辽宁 沈阳 100036；2. 沈阳航空航天大学 计算机学院，辽宁 沈阳 110136；
3. 东北大学 信息科学与工程学院，辽宁 沈阳 110819)

摘 要:提出了网络蠕虫的随机传播模型。首先，基于马尔可夫链对于网络蠕虫进行了建模，并且讨论了模型的极限分布以及平稳分布的存在性。然后，讨论了网络蠕虫在传播初期灭绝的充要条件以及在传播后期灭绝的必要条件。最后，讨论了网络蠕虫的传播规模。仿真实验对于模型进行了验证,讨论了模型中传播参数，时间参数以及漏洞主机数等相关参数对于网络蠕虫传播的影响，并且与 G-W 模型进行了数据对比，说明了本模型的优势。
关键词：网络安全；蠕虫；马尔可夫链；蠕虫的随机模型
中图分类号：TP393.08 **文献标识码**：A

Internet worm propagation model based on Markov chain

ZHOU Han-xun[1], GUO Wei[2], LIU Jian[1], JIA Da-yu[3]
(1. School of Information Science, Liaoning University, Shenyang 100036, China;
2. School of Computer, Shenyang Aerospace University, Shenyang 110136, China;
3. School of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract: A stochastic model of Internet worms is presented. Firstly, the propagation of worms is modeled based on Markov chain. The limit distribution and invariant distribution of the model is discussed. Then, necessary and sufficient conditions of the worm propagation in the initial stage and sufficient conditions of the worm propagation in the late stage are discussed. Finally, the scale of the worm propagation is discussed. The simulation validates the model. And the effect of propagation, time and vulnerable host parameter on the spread of worms is discussed. Furthermore, it is compared to the G-W model, and the advantage of it is illustrated.
Key words: network security; worm; Markov chain; stochastic model of Internet worms

1 引言

自从 1988 年 Morris 蠕虫[1]爆发以来，网络蠕虫就在不断地威胁着网络的安全。然而，直到 2001 年 code red 蠕虫事件爆发后[2]，人们才开始关注蠕虫这个领域。这是由于直到 21 世纪初，网络才与人们的经济和生活紧密的联系起来，因此蠕虫对于网络造成的危害就是对于人们的经济生活造成的危害。为了能够提供好的蠕虫抑制方法，人们利用蠕虫的传播模型来揭示蠕虫的传播规律，并且指导人们抑制蠕虫。

理想的蠕虫传播模型能够充分反映蠕虫的传播过程,预测蠕虫可能带来的威胁，指导人们设计蠕虫的防御检测方法[3~6]。文献[7]利用传染病学的经典 SEM 模型对网络蠕虫进行了建模，然而该模型不能够反映蠕虫后期的传播规律。邹长春等[8]通过考虑蠕虫在传播的后期人们对其防治的 2 个因素，在 KM 模型的基础上得到了两因素模型，该

收稿日期：2014-03-15；**修回日期**：2014-07-23
基金项目：国家自然科学基金资助项目(61300233, 61402298)；辽宁省博士启动基金资助项目(20131086)；沈阳市科技计划基金资助项目(F13-316-1-35)；辽宁省交通科技基金资助项目(201502)
Foundation Items: The National Natural Science Foundation of China (61300233, 61402298); Liaoning Province Research Fund for the Doctoral Program of Higher Education of China (20131086); Shenyang Science and Technology Development (F13-316-1-35); The Transportation Technology Project of Liaoning Province (201502)

模型可以反映蠕虫传播后期的规律。文献[9]提出了刻画采用随机扫描策略网络蠕虫的传播模型 AAWP(analytical active worm propagation)。Yu 等[10]对于可以改变扫描率的网络蠕虫进行了建模。在拓扑蠕虫的建模方面，冯朝胜等[11]提出了 P2P 网络中被动型蠕虫的传播模型。孙鑫等[12]从社会工程学的角度研究社交网络蠕虫的传播机制，通过量化影响用户行为的若干因素，提出了微观节点上的基于用户安全意识的行为博弈模型。文献[13]通过博弈模型表明多种蠕虫检测方法的整合才能有效地检测故意降低传播速度来降低被检测的概率的网络蠕虫。张伟等[14]针对云安全体系环境，基于经典 SIR 模型提出了一种新的病毒传播模型，该模型重点分析了网络中云安全的部署程度和信息收集能力对蠕虫传播模型的影响。Jennifer 等[15]对于在蓝牙网络环境下网络蠕虫的传播过程进行了建模。虽然文献[16]利用马尔可夫模型对于网络蠕虫进行了建模，然而并没有考虑到网络蠕虫主机的移去状态，也没有对于模型的稳定性等性质进行数学证明。文献[17]利用 G-W 分支过程对于网络蠕虫传播模型进行了建模，然而在数学模型中也没有考虑到网络蠕虫主机移去的可能性，只是在仿真实验中加入了该因素。

但是，目前人们建立的网络蠕虫传播模型大多是对于某一特殊蠕虫的建模，使用确定性模型的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，此类模型无法表述传播过程中的概率事件[14]，此外，确定性模型忽视了个体之间的交互行为。本文研究网络蠕虫的随机模型，分析蠕虫病毒在大量主机上传播时表现出来的特征，由于基于马尔可夫链对于网络蠕虫的传播过程进行建模，可以考虑网络蠕虫传播过程中的概率事件，因此对于网络蠕虫的传播刻画的更贴近其真实传播情况。

2 模型结构

模型假设除感病特征外，主机间没有差异，蠕虫传播时采用具有代表性的随机扫描策略。为了对网络蠕虫进行建模，将涉及的主机划分成 3 类，分别为易感状态、感染状态和移去状态。处于易感状态的主机没有被蠕虫感染，但是具有蠕虫可以感染主机的漏洞；处于感染状态的主机是由于网络蠕虫通过感染易感状态的主机将其转化为感染状态的主机。移去状态：处于易感状态的主机是处于感染状态的主机经过杀毒软件或者人工的操作将蠕虫进行删除，从而转化为移去状态的主机。表 1 列出了下文模型中出现的主要参数符号及说明。

表 1 模型参数

符号	说明
N	蠕虫可以感染的漏洞主机
β	蠕虫的感染率
γ	蠕虫的移去率
t	时间 t
$p_{ij}(t)$	t 时刻网络蠕虫传播从 i 状态到 j 状态的概率
$I(i)$	处于感染状态的主机数目

2.1 基于马尔可夫链的网络蠕虫传播模型

在蠕虫的传播过程中，由于网络蠕虫的扩散速度以及具有可以感染的漏洞主机数目的不同等因素，每个时间点网络蠕虫的传播是随机的，因此在时间 t 上的每个时间点，网络蠕虫感染主机的数目 $I(t)$是一个随机变量，其具有的状态为：0,1,2,⋯,N，其中，N 为网络蠕虫的漏洞主机数目。离散的随机变量 $I(t), t\in[0,\infty)$，取值于状态空间 S={0,1,2,⋯,N}，相关的概率函数 $P(t)=(P_0(t),\cdots,P_n(t))^{\mathrm{T}}$。

$$P_i(t)=P\{I(t)=i\} \tag{1}$$

其中，i=0,1,2,⋯,N 并且 t=0, Δt ,2 Δt ,⋯，马尔可夫过程需要满足如下条件：$t_1,t_2,\cdots,t_n,t_{n+1}$，及任意的状态 $i_1,i_2,i_3,\cdots,i_n,i_{n+1}\in S$，均有：$P\{I(t_{n+1})=i_{n+1}|I(t_1)=i_1, I(t_2)=i_2,\cdots,I(t_n)=i_n\}=P\{I(t_{n+1})=i_{n+1}|I(t_n)=i_n\}$，则称此随机过程为马尔可夫链。也就是说，该随机过程 t_{n+1} 时刻的状态仅取决于 t_n 时刻的状态，由于网络蠕虫 t_{n+1} 时刻的传播数目仅与 t_n 时刻的传播数目有关，因此网络蠕虫的传播过程符合马尔可夫链的传播条件，因此即为马尔可夫链。

下面讨论其一步转移概率矩阵。假设在足够小的时间 Δt 里，网络蠕虫只能增加一个感染主机，减少一个感染主机或者保持不变，如下

$$i\to i+1, i\to i-1, i\to i$$

由于在足够小的时间 Δt 里，转移概率被认为是无穷小的转移概率 $o(\triangle t)$，因此 $o(\triangle t)$ 在定义式子 $\lim\limits_{x\to\infty} o(\triangle t)/\triangle t=0$，那么根据传染病模型的原理，可以得到

$$p_{ij}(\Delta t)=\begin{cases}\dfrac{\beta i(N-i)}{N}\Delta t, i=j+1\\ \gamma i\Delta t, i=j-1\\ 1-(\dfrac{\beta i(N-i)}{N}+\gamma i)\Delta t, i=j\\ o(\Delta t), i\neq j, j+1, j-1\end{cases} \tag{2}$$

也就是说，当 $i\to i$+1 时，其转移率为 $\dfrac{\beta i(N-i)}{N}\Delta t$；当 $i\to i$−1 时，其转移率为 $\gamma i\Delta t$；根据马尔可夫链的性质，那么当 $i\to i$ 时，其转移率为 $1-(\dfrac{\beta i(N-i)}{N}+\gamma i)\Delta t$。假设 $P\{I(0)=k\}=1$, 那么 $p_{i,i0}(\Delta t)=p_i(\Delta t)$，令 $b(i)=\dfrac{\beta i(N-i)}{N}$, $d(i)=\gamma i$，因此可以得到如下等式。

$$p_i(t+\Delta t)=p_{i-1}(t)b(i-1)\Delta t+p_{i+1}(t)d(i+1)\Delta t+ p_i(t)(1-(b(i)+d(i))\Delta t) \tag{3}$$

式(3)以矩阵的形式来进行表示，其中，$\boldsymbol{p}(t)=\{p_0(t),\cdots,p_N(t)\}^{\mathrm{T}}$，矩阵定义如下

$$\boldsymbol{Q}=\begin{pmatrix}1 & 0 & 0 & 0\cdots 0\\ d(1)\Delta t & 1-(b(1)+d(1))\Delta t & b(1)\Delta t & 0\cdots 0\\ 0 & d(2)\Delta t & 1-(b(2)+d(2))\Delta t & b(2)\Delta t\cdots 0\\ \vdots & \vdots & \vdots & \\ 0 & 0 & \cdots & 1-d(N)\Delta t\end{pmatrix} \tag{4}$$

定理 1　网络蠕虫的马尔可夫链模型存在极限分布以及平稳分布，并且该马尔可夫链是正常返的。

证明　根据式(4),可以得知 $P_{j,j-1}+P_{j,j}+P_{j,j-1}=1$，可知各状态都相通，根据其状态转移图，利用平衡原理，可以得到式(5)。

$$\begin{cases}\pi_0 p_{01}=\pi_1 p_{10}\\ \pi_1(p_{10}+p_{12})=\pi_0 p_{01}+\pi_2 p_{21}\\ \cdots\\ \pi_j(p_{jj-1}+p_{jj+1})=\pi_{j-1}p_{j-1j}+\pi_{j+1}p_{j+1j}\\ \cdots\end{cases} \tag{5}$$

由此可得

$$\pi_1=\frac{P_{01}}{P_{10}}\pi_0,\pi_2=\frac{P_{01}P_{12}}{P_{10}P_{21}}\pi_0,\cdots,\pi_N=\frac{P_{01}P_{12}\cdots P_{j-1j}}{P_{10}P_{21}\cdots P_{jj-1}}\pi_0 \tag{6}$$

利用 $\sum_{i=0}^{N}\pi_k=1$，可以得到式(7)。

$$\pi_0=(1+\sum_{j=0}^{N}\frac{P_{01}P_{12}\cdots\in P_{j-1j}}{P_{10}P_{21}\cdots P_{jj-1}})^{-1},\pi_N=\frac{P_{01}P_{12}\cdots P_{j-1j}}{P_{10}P_{21}\cdots P_{jj-1}}\pi_0 \tag{7}$$

由式(7)可知，$\pi_0,\cdots,\pi_N$ 存在且均为一个常量，因此，该马尔可夫链的极限分布以及平稳分布均存在，又由于 $(1+\sum_{j=0}^{N}\frac{P_{01}P_{12}\ldots P_{j-1j}}{P_{10}P_{21}\ldots P_{jj-1}})^{-1}>0$，因此，网络蠕虫的马尔可夫链模型是正常返的。

2.2　灭绝概率

随着网络蠕虫的不断传播，此种群究竟是不断壮大，还是规模保持不变或者最终走向灭亡是一个值得研究的问题。由于蠕虫的灭绝可以净化网络环境，减小网络蠕虫对于网络的破坏，因此研究网络蠕虫的灭绝概率就成了一个重要问题。下面给出网络蠕虫传播初期灭绝的充要条件。

性质 1　在蠕虫传播的初期，网络蠕虫的传播不会灭绝的充要条件是 $\beta>\gamma$。

证明　由于网络蠕虫在传播过程中是否会灭绝，当且仅当它在 t 时刻的数学期望不大于 1[18]。因此，下面来推导网络蠕虫在 t 时刻的传播数目的数学期望。

令 $I_F(t)=\mathrm{E}\{I(t)\}=\sum_{n=0}^{N}np_n(t)=\sum_{n=1}^{N}np_n(t)$

$$\mathrm{E}\{I(t+\Delta t)\}=\sum_{n=1}^{N}np_n(t+\Delta t)$$

$$=\sum_{n=1}^{N}np_{n-1}(t)b(n-1)\Delta t+\sum_{n=0}^{N-1}np_{n+1}(t)d(n+1)\Delta t-$$

$$\sum_{n=0}^{N}np_n(t)b(n)\Delta t-\sum_{n=0}^{N}np_n(t)d(n)\Delta t+\sum_{n=0}^{N}np_n(t)$$

$$=\sum_{n=1}^{N}p_{n-1}(t)\frac{\beta(n-1)(N-n+1)}{N}\Delta t-\sum_{n=0}^{N-1}p_{n+1}(t)\gamma(n+1)\Delta t+$$

$\mathrm{E}\{I(t)\}$

因此，$\dfrac{E\{I(t+\Delta t)\}-E\{I(t)\}}{\Delta t}=\sum_{n=1}^{N}p_{n-1}(t)\dfrac{\beta(n-1)(N-n+1)}{N}-\sum_{n=1}^{N-1}p_{n+1}(t)\gamma(n+1)$

当 $\Delta t\to 0$，得到式(8)。

$$\frac{\mathrm{d}E(I(t))}{\mathrm{d}t}=\sum_{n=1}^{N}p_{n-1}(t)\frac{\beta(n-1)(N-n+1)}{N}-\sum_{n=1}^{N-1}p_{n+1}(t)\gamma(n+1) \tag{8}$$

在蠕虫的传染初期由于网络中的漏洞主机远大于网络中的感染主机，即 $N>>n$，因此可以得到 $\dfrac{(N-n+1)}{N}=1-\dfrac{n-1}{N}\approx 1$。因此可以得到如下结论。

原式 $\approx \sum_{n=1}^{N} p_{n-1}(t)\beta(n-1) - \sum_{n=1}^{N-1} p_{n+1}(t)\gamma(n+1)$，假设扫描蠕虫在传播的过程中的扫描率为 s，当 $N>>n$ 时，并且 $ns<N$（由于蠕虫扫描初期，网络中的扫描蠕虫较少而且扫描率一般为 100 左右，例如红色代码蠕虫的扫描率为 100），也就是说目前的网络蠕虫在时刻 t 发出的扫描，即使全部攻击成功，也不可能全部感染网络中的 N 个主机，因此，$P_M(t)=0$，所以有

$$\sum_{n=1}^{N} p_{n-1}(t)\beta(n-1) = \sum_{n=1}^{N} p_n(t)\beta n$$

$$\sum_{n=1}^{N-1} p_{n+1}(t)\gamma(n+1) = \sum_{n=1}^{N} p_{n+1}(t)\gamma n$$

因此，

$$\frac{\mathrm{d}I_F(t)}{\mathrm{d}t} = \sum_{n=1}^{N} p_n(t)\beta n - \sum_{n=1}^{N} p_{n+1}(t)\gamma n$$

$$= (\beta-\gamma)\sum_{n=1}^{N} n p_n(t)$$

$$= (\beta-\gamma)\ I_F(t)$$

若初始条件 $I_F(t)=i$，即可求得结果

$$I_F(t) = \mathrm{E}\{I(t)\} = ie^{(\beta-\gamma)t} \tag{9}$$

蠕虫传播过程中的灭绝条件为 $I_F(t) \leqslant 1$，即 $ie^{(\beta-\gamma)t} \leqslant 1$，因此 $\ln i(\beta-\gamma)t \leqslant 0$，又由于 $t \geqslant 0$，$\ln i>0$,因此只有 $\beta \leqslant \gamma$ 的时候，$I_F(t) \leqslant 1$。也就是说，网络蠕虫的传播才会终止。因此，在蠕虫传播的初期，网络蠕虫的传播不会灭绝的充要条件是 $\beta > \gamma$，因此性质得到证明。

虽然该性质只是证明了在蠕虫传播初期网络蠕虫不会灭绝的条件，但是在网络蠕虫的中后期，由于网络蠕虫的传播受到漏洞主机的数目越来越少，2 个蠕虫感染一台主机的概率越来越大，传播速度一定会低于传播的初期，因此该性质给出了在网络蠕虫传播的中后期不会灭绝的上限条件。具体证明请参阅 2.3 节。

2.3 蠕虫传播的规模

由 2.2 节可知，在网络蠕虫传播的初期，网络蠕虫的传播规模为式(9)，但是在网络蠕虫传播的中后期，就不能用式(9)来进行推算了，下面来推导 $\mathrm{E}\{I(t)\}$的解析解。

根据 2.2 节可知

$$\frac{\mathrm{dE}(I(t))}{\mathrm{d}t} = \sum_{n=1}^{N} p_{n-1}(t)\frac{\beta(n-1)(N-n+1)}{N} - \sum_{n=1}^{N-1} p_{n+1}(t)\gamma(n+1) \tag{10}$$

因此

$$\frac{\mathrm{d\,E}(I(t))}{\mathrm{dt}} = (\beta-\gamma)\mathrm{E}(I(t)) - \frac{\beta}{N}\mathrm{E}(I^2(t)) \tag{11}$$

由于微分方程中有 $\mathrm{E}(I^2(t))$，因此无法对式(13)得到解析解，只能得到式(13)的数值解。由于 $\mathrm{E}(I^2(t)) \geqslant \mathrm{E}^2(I(t))$，因此可以得到如下不等式

$$\frac{\mathrm{dE}(I(t))}{\mathrm{d}t} \leqslant (\beta-\gamma)\mathrm{E}(I(t)) - \frac{\beta}{N}\mathrm{E}^2(I(t))$$

$$= \frac{\beta}{N}(N-\mathrm{E}(I(t)))\mathrm{E}(I(t)) - \gamma\mathrm{E}(I(t)) \tag{12}$$

性质 2 网络蠕虫的传播不会灭绝的必要条件是 $\beta > \gamma$。

证明 在式(11)中,由于网络蠕虫的感染数目是非负的，如果 $\beta \leqslant \gamma$，那么 $(\beta-\gamma)\mathrm{E}(I(t)) \leqslant 0$；由于 $\mathrm{E}(I^2(t))$ 是非负的，若 $\mathrm{E}(I^2(t))=0$，由于 $I(t)$ 非负，那么 $\mathrm{E}(I(t))=0$，又由于 $\frac{\mathrm{dE}(I(t))}{\mathrm{d}t}=0$，即 $\mathrm{E}(I(t))$ 的增量为 0，那么 $\mathrm{E}(I(t))=0$，也就是说网络蠕虫灭绝了；若 $\mathrm{E}(I^2(t))>0$,则 $-\frac{\beta}{N}\mathrm{E}(I^2(t)) \leqslant 0$，因此 $\frac{\mathrm{dE}(I(t))}{\mathrm{d}t}<0$，网络蠕虫的数目将不断降低，最终导致灭绝。也就是说,当 $\beta \leqslant \gamma$ 时，网络蠕虫将会灭绝。因此，性质 2 得到证明。

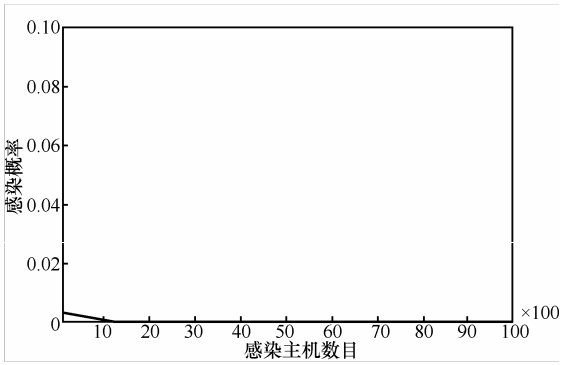
3 仿真验证及分析

为了进一步揭示网络蠕虫的传播规律，模拟随机扫描蠕虫的传播策略编写了一个模拟器仿真验证了传播模型, 对于每项试验进行了 1 000 次的仿真实验，通过发生的频率与实验次数的比值计算其概率，并且得到了该模型的传播图像，如图 1 所示。图像的横轴为感染主机的数目，纵轴代表相应蠕虫感染主机数目的感染概率。下面对于模型的参数进行详细讨论。

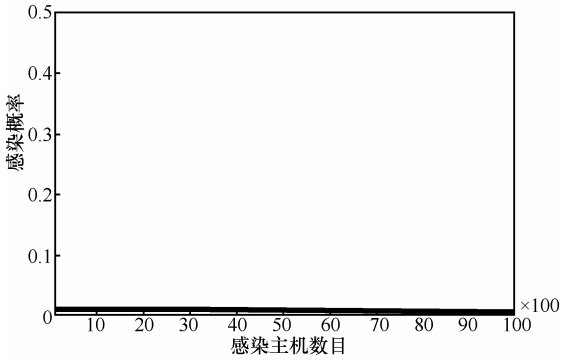
3.1 传播参数

图 1 是关于传播参数 β 和 γ 的讨论。其中图 1(a)和图 1(b)分别为 $\beta=\gamma$ 以及 $\gamma=2\beta$。在这 2 组数据下，网络蠕虫的传播主机数目的相应概率均为 0。

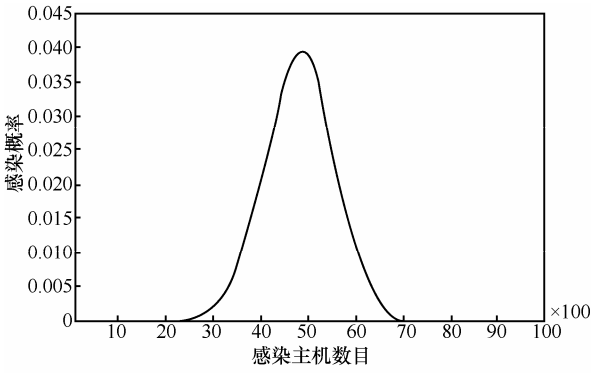
同时测试了其他参数组合，得到了类似的结论，由于篇幅所限，这里就不加叙述。也就是说，当$\beta \leqslant \gamma$的时候，网络蠕虫就会灭绝。这也证明了在 2.2 节对于灭绝概率的讨论。图 1(c)和图 1(e)分别为β=1 和γ_c=10γ_e，图 1(d)和图 1(f)分别是图 1(c)和图 1(e)的累积概率图，即概率分布函数。可以看到在图 1(c)中，网络蠕虫数目为 5 000 的时候的概率为峰值 0.04，在图 1(d)中，可以看到网络蠕虫传播小于 7 000 的概率为 1；在图 1(d)中，网络蠕虫在 8 000 的概率为峰值 0.085，在图 1(d)中，可以看到网络蠕虫传播小于 9 200 的概率为 1。也就是说，当网络蠕虫的传播速度比移除网络蠕虫的速度越快，那么网络蠕虫的传播数目较大的概率就越大。但是，即使是概率的峰值也不过只有 0.04 或者 0.085，因此，网络蠕虫传播的随机性很大。但是，从图像中又可以看到在图 1(c)中网络蠕虫的传播值在[2 000, 7 300]之间概率大于 0，其他值等于 0；在图 1(d)中网络蠕虫的传播值在[6 200,9 200]之间概率大于 0，其他值等于 0。因此，网络蠕虫在传播过程中，只能知道在某个区间的传播可能性较大，而无法确切的指出其传播的值。因此，定义该区间为传播区间，传播区间指出了网络蠕虫在传播过



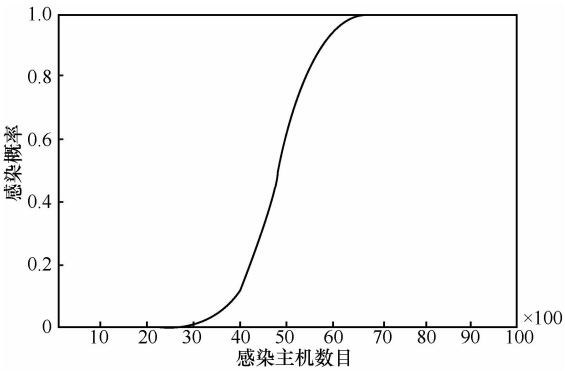
(a) $\beta = \gamma$



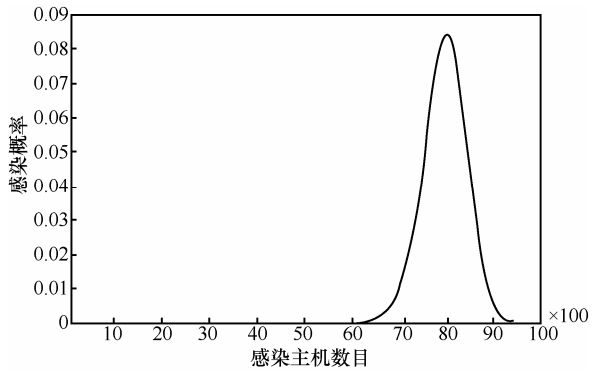
(b) γ=2 β



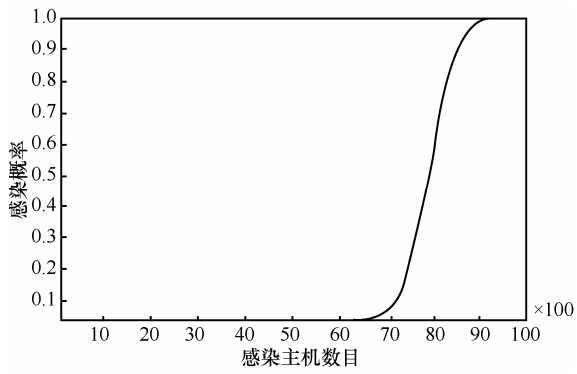
(c) β =1



(d) 图 1(c)累积概率



(e) γ_c =10 γ_e



(f) 图 1(e)累积概率

图 1　传播参数

程中可能的感染状态。

3.2 时间参数

图 2 是在传播时间分别为 750、1 000 和 2 000 个单位时间时，网络蠕虫的传播图像。当时间点不同时，网络蠕虫的传播区间也会发生变化，因此将网络蠕虫的传播区间定义为三元组 $WP(t,\ \min,\ \max)$。

$$WP(t,\min,\max)=\begin{cases}P(t,0,\min)=0\\P(t,\max,N)=0\\P(t,\min,\max)>0\end{cases} \tag{13}$$

其中，t 为时间，min 为传播区间的最小值，max 为传播区间的最大值。

因此，若 $WP(2\,000, 2\,000, 7\,000)$表明网络蠕虫在时间为 2 000 时，网络蠕虫的传播值在[2 000, 7 000]之间概率大于 0，其他值等于 0。从图中我们不难得出另外 2 个传播区间 $P(750,0,6\,000)$，$P(1\,000, 0, 7\,500)$。从实验结果不难发现，随着网络蠕虫传播时间的增加，网络蠕虫感染较大主机的数目的概率也会不断增大。

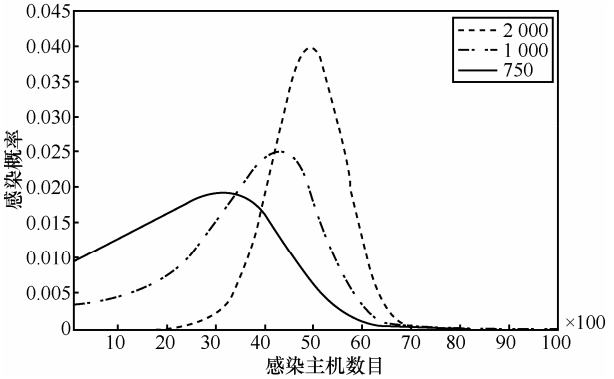
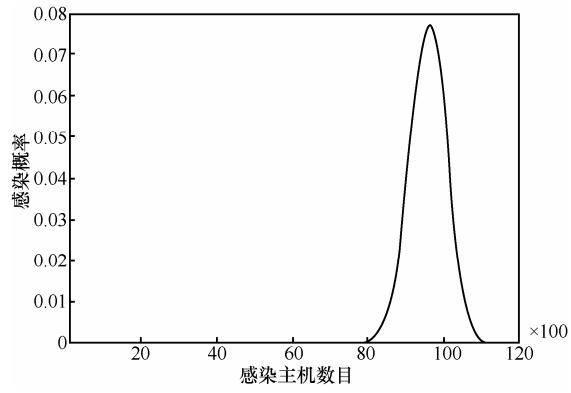


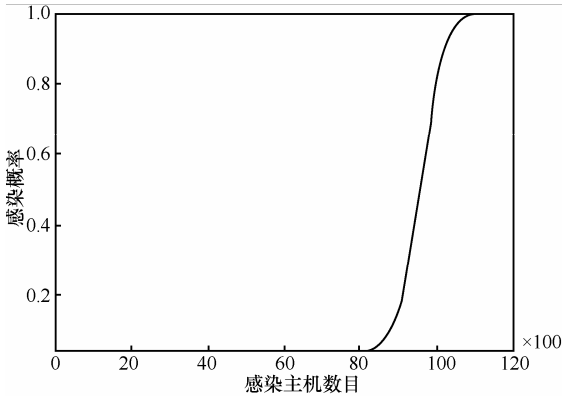
图 2 时间参数

3.3 漏洞主机

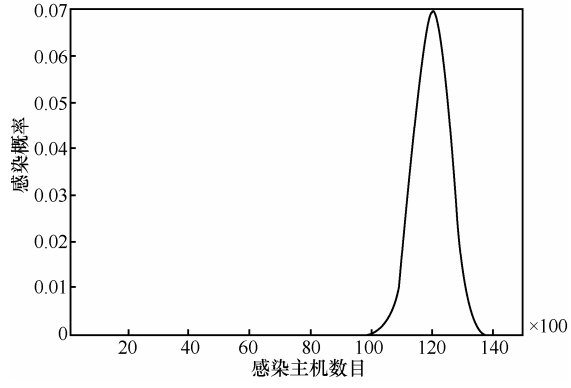
图 3(a)、图 3(b)是漏洞主机为 12 000 时，网络蠕虫的传播概率图以及相应的概率分布函数。图 3(c)、图 3(d)是漏洞主机为 15 000 时，网络蠕虫的传播概率图以及相应的概率分布函数。从图中不难发现，当漏洞主机为 12 000 时，其传播区间为 $WP(2\,000,8\,000,11\,000)$；当漏洞主机为 15 000 时，其传播区间为 $WP(2\,000,10\,000,14\,000)$。在图 3(a)中，网络蠕虫数目为 95 000 的时候的概率为峰值 0.078，在图 3(b)中，网络蠕虫传播小于 11 000 的概率为 1；



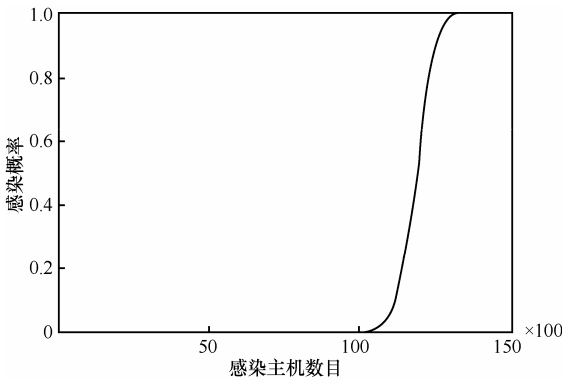
(a) 漏洞主机为 12 000 时网络蠕虫的传播概率



(b) 漏洞主机为 12 000 时网络蠕虫的概率分布函数



(c) 漏洞主机为 15 000 时网络蠕虫的传播概率



(d) 漏洞主机为 15 000 时网络蠕虫的概率分布函数

图 3 漏洞主机的影响

在图 3(c)中，网络蠕虫在 12 000 的概率为峰值 0.07，在图 3(d)中，可以看到网络蠕虫传播小于 14 000 的概率为 1。即使是概率的峰值也不过只有 0.07 或者 0.078，因此，网络蠕虫传播的随机性还是很大。因此，随着网络蠕虫漏洞主机数目的增加，网络蠕虫感染较大主机的数目的概率也会不断增大。

3.4　模型比较

图 4(a)为本文提出的基于马尔可夫链的网络蠕虫传播模型与文献[17]提出的基于G-W分支过程的网络蠕虫传播模型(N=10 000)的对比，图 4(b)为本文模型与文献[16]提出的未考虑主机移除率的马尔可夫链的网络蠕虫模型(N=10 000)的对比。本文提出的传播模型其传播区间为 WP(2 000, 2 000, 7 000)，基于 G-W 分支过程的网络蠕虫传播模型的传播区间为 WP(2 000, 80 000, 10 000)，未考虑主机移除率的马尔可夫链蠕虫模型的传播区间为 WP(2 000, 70 000, 95 000)。这主要是由于基于 G-W 分支过程的网络蠕虫传播模型以及文献[16]的传播模型并没有考虑到网络蠕虫主机的移除的可能性，因此存在几乎感染全部漏洞主机的可能性。此外，本文的网络蠕虫模型传播概率的峰值为 0.04；G-W 分支过程的网络蠕虫传播模型传播概率的峰值为 0.125；未考虑主机移除率的马尔可夫链模型的传播概率的峰值为 0.093。3 个模型都表明即使是概率峰值也不大，因此网络蠕虫的传播具有较强的随机性。

4　结束语

本文提出了网络蠕虫的随机传播模型。首先，基于马尔可夫链对于网络蠕虫进行了建模，并且讨论了模型的极限分布以及平稳分布的存在性。然后，讨论了网络蠕虫在传播初期灭绝的充要条件以及在传播后期灭绝的必要条件。最后，讨论了网络蠕虫的传播规模。仿真实验对于模型进行了验证。由于本文研究网络蠕虫传播的随机特性，分析蠕虫病毒在大量主机上传播时表现出来的特征，可以考虑网络蠕虫传播过程中的概率事件，因此对于网络蠕虫的传播刻画得更贴近其真实传播情况。

参考文献：

[1] SPAFFORD E H. The Internet Worm Program: an Analysis[R]. Technical Report, CSD-TR-823, West Lafayette: Department of Computer Science, Purdue University, 1988. 1-29.

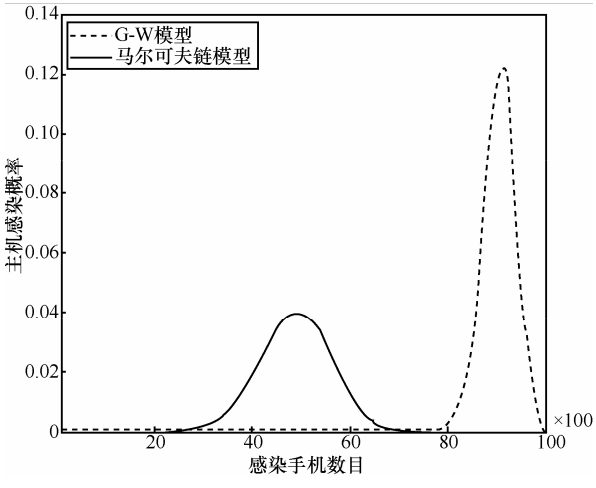
[2] MOORE D, SHANNON C, BROWN J. Code-Red: a case study on the spread and victims of an Internet worm[A]. Proceedings of the Second ACM SIGCOMM Workshop on Internet Measurement[C]. 2002. 273-284.

[3] 刘玉岭, 冯登国, 吴丽辉等. 基于静态贝叶斯博弈的蠕虫攻防策略绩效评估[J]. 软件学报, 2012, 23(3): 712-723.
LIU Y L, FENG D G, WU L H, *et al*. Performance evaluation of worm attack and defense strategies based on static Bayesian game[J]. Journal of Software, 2012,23(3):712-723.

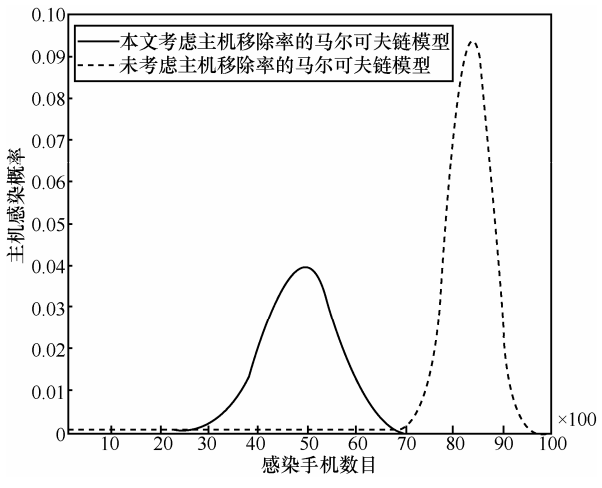
[4] 汪洁, 王建新, 刘绪崇. 基于近邻关系特征的多态蠕虫防御方法[J]. 通信学报, 2011,32(8): 150-158.
WANG J, WANG J X, LIU X C. Novel approach based on neighborhood relation signature against polymorphic internet worms[J]. Journal on Communications,2011,32(8):150-158.

[5] 洪征, 吴礼发. 基于阳性选择的蠕虫检测系统[J]. 软件学报, 2010,21(4): 816-826.
HONG Z,WU L F. Worm detection system based on positive selection[J]. Journal of Software, 2010, 21(4):816-826.

[6] 唐勇, 诸葛建伟, 陈曙晖等. 蠕虫正则表达式特征自动提取技术研究[J]. 通信学报, 2013, 34(3):141-147.



(a) 比较 1



(b) 比较 2

图 4　模型对比

TANG Y, ZHUGE J W, CHEN S H, *et al*. Automatic generating regular expression signatures for real network worms[J]. Journal on Communications, 2013, 34(3): 141-147.

[7] STANIFORD S, PAXSON V, WEAVER N. How to own the Internet in your spare time[A]. Proc of the 11th Usenix Security Symp[C]. San Francisco, 2002.

[8] ZOU C C, GONG W, TOWSLEY D. Code Red worm propagation modeling and analysis[A]. Proc of the 9th ACM Symp. on Computer and Communication Security[C]. Washington, 2002. 138-147.

[9] CHEN Z, GAO L, KWIAT K. Modeling the spread of active worms[A]. IEEE INFOCOM 2003[C]. 2003.1890-1900.

[10] YU W, WANG X, PRASAD C, XUAN D, ZHAO W. Modeling and detection of Camouflaging worm[J]. IEEE Transaction on Dependable and Secure Computing, 2011, 8(4): 377-390.

[11] 冯朝胜, 秦志光, 袁丁等. P2P 网络中被动型蠕虫传播与免疫建模[J]. 电子学报, 2013, 41(5):884-889.
FENG C S, QIN Z G, YUAN D, *et al*. Modeling propagation and immunization of passive worms in peer-to-peer networks[J]. Acta Electronica Sinica,2013,41(5):884-889.

[12] 孙鑫, 刘衍珩, 朱建启等. 社交网络蠕虫仿真建模研究[J]. 计算机学报, 2011, 34(7): 1252-1261.
SUN X, LIU Y H, ZHU J Q, *et al*. Research on simulation and modeling of social network worm propagation[J].Chinese Journal of Computers,2011,34(7):1252-1261.

[13] YU W, ZHANG N, FU X W, *et al*. Self-disciplinary worms and countermeasures:modeling and analysis[J]. IEEE Transactions on Parallel and Distributed Systems, 2010, 21(10): 1501-1514.

[14] 张伟, 王汝传, 李鹏. 基于云安全环境的蠕虫传播模型[J]. 通信学报, 2012, 33(4):17-24.
ZHANG W, WANG R C, LI P. Worm propagation modeling in cloud security[J].Journal on Communications, 2012, 33(4):17-24.

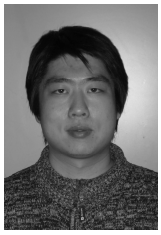
[15] JENNIFER T. Jackson and sadie creese, virus propagation in heterogeneous bluetooth networks with human behaviors[J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(6):930-943.

[16] 刘烃, 郑庆华, 管晓宏等. 基于随机实验的蠕虫传播预测研究[J]. 通信学报, 2007, 28(12):72-77.
LIU J, ZHENG Q H, GUAN X H, *et al*. Research of worm-propagation prediction based on stochastic experiment[J]. Journal on Communications, 2007, 28(12):72-77.

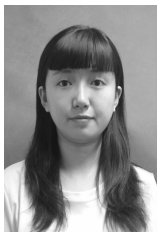
[17] SARAH S, NESS B S, SAURABH B. Modeling and automated containment of worms[J]. IEEE Transactions on Dependable and Secure Computing, 2008, 5(2): 528-537.

[18] ROSS S. Stochastic Processes[M]. John Wiley & Sons, 1996.

作者简介：

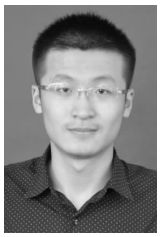


周翰逊（1981-），男，辽宁沈阳人，博士，辽宁大学讲师，主要研究方向为网络安全和图像处理。



郭薇（1983-），女，辽宁沈阳人，博士，沈阳航空航天大学讲师，主要研究方向为网络安全和图像处理。

刘建（1979-），男，辽宁铁岭人，辽宁大学讲师，主要研究方向为网络安全、无线传感器网络和数据挖掘。



贾大宇（1990-），男，辽宁沈阳人，东北大学硕士生，主要研究方向为网络安全和大数据。