第 7 期 电 子 学 报 Vol. 45 No. 7
2017 年 7 月 ACTA ELECTRONICA SINICA Jul. 2017

面向分层网络的社交蠕虫仿真建模研究

王天博[1,2],夏春和[1,2],贾 琼[1,2]
(1. 北京市网络技术重点实验室,北京 100191; 2. 北京航空航天大学计算机学院,北京 100191)

摘 要: 随着社交网络的普及,社交蠕虫已经成为了威胁社会的主要隐患之一. 这类蠕虫基于拓扑信息和社会工程学在因特网中快速传播. 先前的学者们对社交蠕虫的传播建模与分析主要存在两个问题:网络拓扑的不完整性和传播建模的片面性;因而导致对社交蠕虫感染规模的低估和人类行为的单一化建模. 为了解决上述问题,本文提出了社交蠕虫传播仿真模型,该模型使用分层网络能更准确地抽象社交逻辑层与实际物理层之间的关系,以及利用人类移动的时间特性能更全面地刻画社交蠕虫的传播行为. 实验结果表明,该仿真模型揭示了用户行为、网络拓扑参数以及不同的修复过程对社交蠕虫传播造成的影响. 同时,文中对社交蠕虫的传播能力做出了定性分析,为网络防御提供了重要的理论支持.

关键词: 社交蠕虫; 传播行为; 人类移动; 仿真建模
中图分类号: TP309. 05 **文献标识码:** A **文章编号:** 0372-2112 (2017)07-1722-09
电子学报 URL: http://www. ejournal. org. cn **DOI:** 10. 3969/j. issn. 0372-2112. 2017. 07. 024

Modeling and Simulation Study of Social Worms Oriented Hierarchical Networks

WANG Tian-bo[1,2], XIA Chun-he[1,2], JIA Qiong[1,2]
(1. *Beijing Key Laboratory of Network Technology*, *Beijing* 100191, *China*;
2. *School of Computer Science and Engineering*, *Beihang University*, *Beijing* 100191, *China*)

Abstract: With the popularity of social networks, social worms have constituted one of the major security threats. This type of worm spreads fast in Internet networks based on topology information and social engineering. There are two critical problems of previous works, which result in underestimating the scale of infected network and lacking a comprehensive and accurate abstraction for human behaviors. These are structural imperfection of network topology and one-sidedness of the propagation model. To address the above problems, this paper proposed a novel simulation model of social worms. We more accurately and comprehensively abstracted the relationship between social logical layer and actual physical layer in a hierarchical network, and characterized the social worm propagation through the temporal characteristic of human mobility. Experimental results show the impacts of user behaviors, topology parameters and different recovery processes on the propagation. Meanwhile, we give the qualitative analysis of the spread ability.

Key words: social worm; propagation behavior; human mobility; simulation modeling

1 引言

近年来,社交蠕虫已经成为网络安全主要威胁之一. 根据网络安全威胁的官方报告[1],社交蠕虫传播的频率和危害都有显著地提升. 一方面,社交网络一直是一个很好的信息传播平台;另一方面,它们借助社会工程学,提高网络攻击的成功率. 因此,社交蠕虫已经成为网络安全领域的一个研究热点.

先前的学者在研究社交蠕虫传播机制和防御方案方面做了大量的工作. 一方面,主要将传染病学的思想引入到蠕虫传播的领域,提出了一些基于状态转变的传播模型[2~3],分析蠕虫传播的传染阈值和平衡解. 另一方面,分析拓扑型和扫描型两类蠕虫传播行为特征[4~6]. 但是他们研究成果存在网络拓扑不完整和传播

收稿日期:2015-09-15;修回日期:2016-07-26;责任编辑:孙瑶
基金项目:北京航空航天大学博士研究生卓越学术基金;国防基础科研计划(No. A2120110006);北京市教育委员会共建项目建设计划(No. JD100060630);中航工业产学研(No. CXY2011BH07)

建模片面的问题. 前者导致感染主机数量被低估,后者导致了传播建模的单一化,即只考虑消息检查行为. 国内学者冯朝胜等人用平均场理论分别研究了 P2P 网络主动与被动蠕虫的传播过程,但是此方法无法刻画蠕虫传播过程的离散事件[7,8]. 文献[9]基于进程代数研究良性蠕虫与恶意蠕虫对抗过程,进而分析蠕虫传播特性. 文献[10]主要分析社交网络之间好友关系对传播行为的影响,并建立了基于信任的蠕虫防御机制. 但是上述文献没有考虑人类移动的时间特性对于社交蠕虫传播的影响.

本文对上述的两个问题进行了深入地分析,构建了分层网络的社交蠕虫传播场景,提出了基于人类移动时间特性的社交蠕虫仿真模型. 同时,理论分析得出了影响社交蠕虫传播能力的定性指标,为今后的网络防御奠定了坚实的基础.

2　相关工作

近十年来,蠕虫传播领域的研究成果和发现主要集中在两种类型的蠕虫:扫描型蠕虫和拓扑型蠕虫. 前者是凭借不同的扫描策略去感染脆弱性主机;后者是依靠拓扑结构关系去感染他们的邻居节点主机. 最近研究人员更多地关注通过建立传播模型来检测和评估扫描型蠕虫的传播行为方式. Yu 等人对一种能够通过自我训练实现自动调节扫描率以躲避检测的蠕虫进行了研究. 引入博弈论思想形式化蠕虫传播者和防御者之间的对抗,给出了一种有效的蠕虫检测和取证分析的集成方案[11]. 次年,他们调查研究了伪装蠕虫的传播特征,并设计了一个基于频谱的检测方案[12]. Manna 等人提出了一种基于相互依赖的微分方程数学模型,能够准确地分析置换扫描型蠕虫的传播模式[13].

随着信息技术的不断发展,社交软件已经成为一种主要的网络交流工具. 本文主要以这类拓扑型蠕虫作为研究对象. 社交蠕虫的研究主要分为两类:解析建模和传播仿真. 解析建模主要是基于微分和差分方程建立数学解析模型,研究社交蠕虫的传播特性. 例如,Chen 等人提出了一种基于概率图的马尔科夫模型,使用随机过程描述传播的统计依赖性(联合概率的计算)[4]. Wen 等人针对时间动态性(网络用户消息检测触发时刻是不同的)和空间依赖性(传播环导致感染概率的重复计算)的问题,提出了一个"健康-感染-免疫"SII 模型[6]. Faghani 等人对跨脚本攻击的 XSS 蠕虫传播行为进行了数学建模,提出了一种选择性监控检测方案[14]. 但是他们是面向社会逻辑层进行蠕虫传播相关研究,不符合实际的分层网络拓扑结构. 传播仿真主要是通过仿真模型对蠕虫传播行为进行分析. 例如,Zou 等人通过实验环境中的仿真模型对用户检测邮件的行为进行随机化建模,反映了用户行为的随机性和不确定性的特点,较好地仿真了电子邮件蠕虫的传播[5]. Yan 等人基于真实的在线社交网络位置信息数据集,建立社交蠕虫传播仿真环境,分析拓扑结构、用户点击行为和活动模式对社交蠕虫传播动态的影响[15]. Fan 等人分析了社交网络中群组交流模式,能够缩短用户之间的网络距离,促进社交蠕虫的传播. 并通过仿真系统验证了相应的结论[16].

3　问题描述

社交蠕虫传播所在的分层网络拓扑是由社会逻辑层和实际物理层构成,如图 1(a)所示.

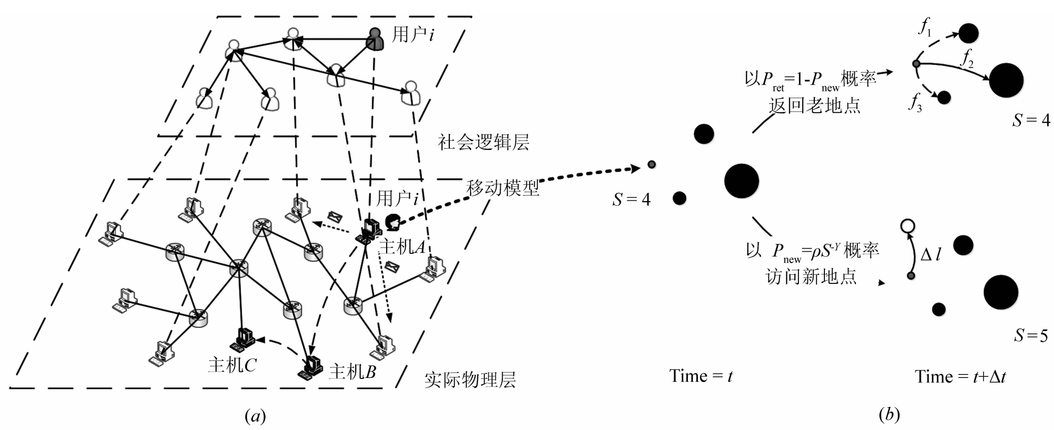


图1　分层网络中社交蠕虫的传播场景

3.1　网络拓扑结构的不完整性问题

人类在社会中扮演着不同的社会角色,使得人们可以在不同的地点进行工作娱乐活动. 这一现象决定了用户可以在不同的地点使用相应的主机处理相关的事务,从而导致传统的社交蠕虫传播拓扑结构与现实的社交蠕虫传播拓扑结构不一致. 前人的模型基于社会逻辑层研究社交蠕虫的感染趋势[5,6,17],拓扑结构默认每个用户只有一台主机. 但是现实中不同地点的主机可以属于同一个用户. 因此,先前的模型存在感染规模低估的问题.

3.2 传播建模的片面性问题

先前的学者针对用户行为主要集中在用户的消息检测行为对社交蠕虫传播的影响. 但是他们忽略了人类移动行为的时间特性对社交蠕虫传播也构成了直接的影响. 同时,移动行为也是分层网络中不同层的元素之间对应关系形成的基础. 可见,前人的研究对社交蠕虫传播行为的认识是不准确和不全面的. 为了解决此问题,本文引入人类移动模型,如图 1(b)所示[18].

4 社交蠕虫传播仿真模型

4.1 社交蠕虫的传播机制

社交蠕虫的传播过程主要包括 3 个阶段:首先恶意代码伪装成正常的邮件附件或者网络链接,欺骗用户点击查看并执行蠕虫文件感染主机;其次对感染用户的好友关系信息进行采集;最后利用好友关系信息通过社交网络进行蠕虫传播.

本文将主机的状态分为:易感状态(Susceptible State)、感染状态(Infected State)和免疫状态(Immune State). 通过用户对杀毒软件的升级,或者对恶意代码和软件的警觉性增强,网络中的易感或感染主机可以被免疫修复,变成免疫状态. 在上述防御修复过程中,仿真模型引入修复函数 $r(t)$ 表示网络主机被免疫修复的过程:一个易感的主机可以对社交蠕虫产生免疫,同时一个感染的主机可以通过用户清除或者修复手段,使得不会被同一种蠕虫所感染.

4.2 网络节点、拓扑和相关事件描述

本节对基于人类移动时间特性的社交蠕虫的传播行为建模. 首先,引入 $M\times M$ 的传播矩阵表示社会逻辑层的传播拓扑.

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ \vdots & \vdots & p_{ij} & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{pmatrix} \tag{1}$$

其中,p_{ij}表示从用户 i 到用户 j 的传播概率,M 表示社会逻辑层的用户规模.

其次,采用 $M\times L_i^t$ 的移动矩阵 $\boldsymbol{N}_{M\times L_i^t}$表示在时刻 t 用户 i 访问过的所有地点次数.

$$\begin{pmatrix} N_{11}(t) & \cdots & N_{1L_i^t}(t) \\ \vdots & N_{ij}(t) & \vdots \\ N_{M1}(t) & \cdots & N_{ML_i^t}(t) \end{pmatrix} \tag{2}$$

其中,$N_{ij}(t)$表示截止到时刻 t 用户 i 访问过地点 j 的次数,L_i^t 表示在时刻 t 用户 i 访问过不同地点的个数. 随机变量 T_{rest}^i表示用户 i 停留在某一地点的逗留时间,则用户 i 在地点 j 的逗留时间由 $P(T_{\text{rest}}^i)$ 分布决定. $p_i^j(t)$ 表示用户 i 返回地点 j 的概率,则有

$$p_i^j(t) = \frac{N_{ij}(t)}{\sum_{j=1}^{L_i^t} N_{ij}(t)} \tag{3}$$

一方面,本文引入随机变量 T_{check}^i 表示用户 i 的消息检测周期. 它主要由用户行为习惯决定,每个特定的时间间隔用户会查看社交账户中的新消息. 另一方面,引入变量 ε 记录从上一次检查新消息到当前时刻 t 之间(不包括 t)的新消息到达的任意时间.

4.3 社交蠕虫传播模型

本文引入动态向量(Dynamic Vector,**DV**)数据结构表示物理主机的使用情况,如图 2 所示. 本文把用户使用的主机分成私有主机和公共主机两种类型. 规定用户 i 在地点 1 使用的主机 host 1 设为私有主机只能为自己使用,其他地点使用的主机设为公共主机可被其他人使用.

社交蠕虫的传播过程:

①初始化:移动矩阵(2)和动态向量 **DV** 中除了初始时刻所在地点 1 和 host 1 对应的元素设为 1,其他的元素都设为 0,传播矩阵(1)中 p_{ij}服从指数分布,邻接矩阵(1)中 A_{ij}按照社会逻辑层节点 i 和节点 j 是否存在边的情况赋值 1 或 0.

②根据图 1(b)的人类移动模型,用户移动到不同的地点使用相应的主机,并更新移动矩阵的对应元素值.

③更新动态向量 **DV** 的元素值. 根据“共享优先”原则,决定用户对公共主机的使用情况. 具体描述如下:如图 2 所示,假设用户 1 以概率 P_{new}到达新地点 4,依照“共享优先”原则,首先判断是否可能使用其他用户的公共主机. 设 $P_{i,j}^n(t)$表示用户 i 和用户 j 在同一地点 n 的概率.

$$P_{i,j}^n(t) = \zeta\cdot\frac{L_j^t}{\sum_{n=1}^{L_j^t} T_{\text{rest}}^j(n)}\cdot\frac{1}{k} \tag{4}$$

$P_{i,j}^n(t)$与用户访问过的地点总数 L_j^t 成正比,与不同地

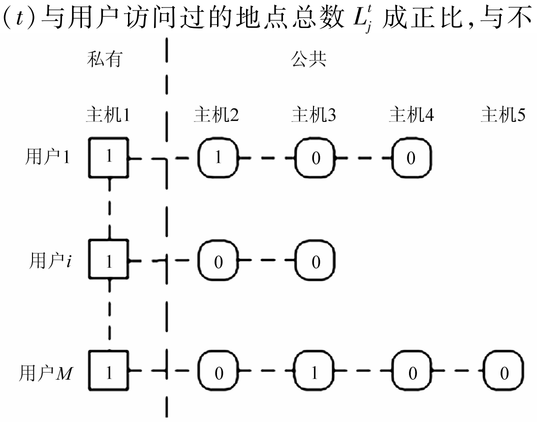


图2 动态向量数据结构

点的逗留时间 $T^{j}_{\mathrm{rest}}(n)$ 总和以及最短路径距离 k 成反比. 比例系数 ζ 为:

$$\zeta = \frac{1}{\sum_{k=1}^{7}\sum_{\substack{j:\mathrm{neighbor\ of\ }i,\\ d^{ij}_{\min}=k}} \frac{L^t_j}{\sum_{n=1}^{L^t_j} T^i_{\mathrm{rest}}(n)} \cdot \frac{1}{k}} \tag{5}$$

其中,$d^{ij}_{\min}=k$ 表示用户 i 是用户 j 的最短路径距离为 k 的邻居.

$P_{j|i}(k)$ 表示最短路径距离为 k 的用户 j 使用邻居 i 的公共主机的概率.

$$P_{j|i}(k) = \frac{1}{\deg_j} \cdot \frac{1}{\deg^2_{k-1}} \cdots \frac{1}{\deg^2_1} \cdot \frac{1}{\deg_i} \tag{6}$$

$\deg_i$ 表示节点 i 的度. 式(6)表示用户 i 和用户 j 路径上所有边的权重乘积,边的权重等于边上两个节点度数的倒数乘积.

计算邻接矩阵(1)搜索用户 1 所有最短距离为 k($1\leqslant k\leqslant 7$,k 的上限为 7——依据六分度理论得出)的邻居 i,依次计算邻居 i 对应的每个公共主机 host n 的概率 $P^n_{i,1}(t)\cdot P_{1|i}(k)$是否发生. 如果发生,则在相应 host n 的元素置为 1,并操作主机 host n. 如果所有邻居对应的所有地点的公共主机的概率 $P^n_{i,1}(t)\cdot P_{1|i}(k)$没有发生,则在自己的地点 4 的加入公共主机 host 4 节点,并将对应元素置为 1,操作 host 4 主机. 如果用户 1 以概率 P_{ret}返回老地点,同样按照“共享优先”原则执行. 注意:用户到达地点 1 直接使用私有主机将对应的元素置为 1,操作自己的私有主机,不执行步骤③.

④ 重复执行步骤②和③,实现社交蠕虫的传播过程. 此外,通过遍历动态向量 **DV** 中数值为 1 的元素,计算当前时刻 t 网络感染主机的数量.

5　仿真模型验证

5.1　社交网络拓扑生成

本文研究人类移动行为时间特性对蠕虫传播的影响,不涉及空间特性. 因此,社交蠕虫传播与实际物理层的拓扑结构和空间位置无关,由于它是通过社会逻辑层的拓扑关系与用户行为去感染实际物理层中的主机,则社会逻辑层的拓扑生成是本文的一项重要工作. 具体拓扑生成算法描述如算法 1 所示.

算法 1　The algorithm for social network topology generation

```
/* step 1:拓扑初始化:通过十字链表构建拓扑 */
1. initialize the amount of nodes N;
2. create node_head[N] array;
/* step 2:初始化节点的度和边的权重 */
1. InitDegrees By PowerLaw(alpha,max,min);
2. initialize the amount of edges E = 0
  for i = 1 to N do
      E + = node_head[i].degree;
  end
3. indegree[N] = outdegree[N] = node_head[N].degree;
/* step 3:插入双向边 */
RE = reciprocity rate × E
  while (RE > 0)
    do
      n1 = rand() % E, s = find(n1);
      n2 = rand() % E, d = find(n2);
    while (s = d)
      addEdge(s,d,p);
    addEdge(d,s,p);
    RE - = 2, indegree[s] - = 1, outdegree[d] - = 1;
  end
/* step 4:插入单向边 */
  DE = E = (1 - reciprocity rate) × E
  for i = 1 to DE do
      for j = 1 to indegree[i] do
        n = rand() % E, d = find(n);
        addEdge(i,d,p);
        E - = 1, outdegree[d] - = 1;
      end
  end
```

5.2　模型的有效性验证

前人的研究是通过仿真和解析模型分析验证他们的结果[4~6]. 本节将与前人的模型作对比,同时分析用户行为、网络拓扑以及不同修复过程对社交蠕虫传播的影响. 仿真环境是通过 C++ 和 Matlab R2014a 实现的,试验中通过 C++TR1 扩展库生成随机数. 分层网络中社会逻辑层的规模为 100000 个节点,实际物理层的规模为 300000+ 个节点. 实验以 100 次重复运行的平均值作为最终结果. 每次运行初始时刻在网络中随机选取两个感染节点,节点间的距离为 6,以此反映聚集系数的影响[6].

5.2.1　本文仿真模型与真实蠕虫传播数据对比

官方 CADIA 对数据集描述[19]:Nyxem 蠕虫尝试禁用不同种类的反病毒软件,使用不同的主题字段和附件,通过扫描邮件联系人列表自动地在网络中传播恶意代码. 为了能够估算该蠕虫的感染规模,我们对网络中的每个 IP 检测两个关键值:易受攻击的浏览器类型的数量和来自 IP 地址发回探针的总数. 前者代表感染节点数量的下界,后者代表感染节点数量的上界. 感染节点规模在区间[469,507,946,835]之间,数据获取时间从 2006 年 2 月 1 日星期三 05:00:12 到 2006 年 1 月 15 日星期日 23:40:54,持续 365 个小时分布在 200 多个城市. 但是,研究人员没有公布上述数据集详细准确的实际社交网络的结构和规模. 因此,我们通过分析实际感染节点的数量,设置社交拓扑节点数的均值大概

为 500,000.

图 3 显示了本文模型的感染规模曲线与 Nyxem 社交蠕虫的感染规模的均值曲线非常的接近. 这说明了本文模型能够准确地刻画社交蠕虫的传播行为. 两条曲线存在偏差的原因主要在于官方没有提供准确地社交拓扑数据. 因此,实际应用中如果能够获得观察地区的相关拓扑数据,本文模型能够准确预测社交蠕虫的传播趋势,具有实际应用价值. 注:图中的两个子图分别代表了实际蠕虫感染规模的上界和下界.

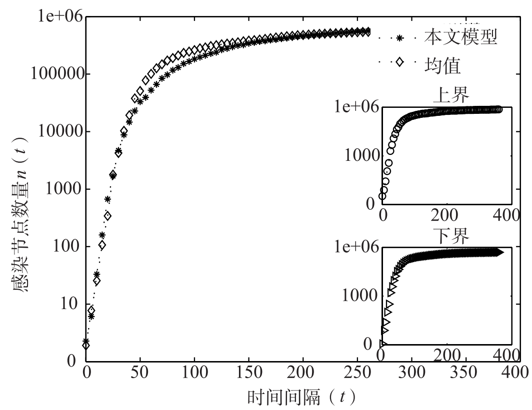


图3 本文模型与实际蠕虫感染数据对比

5.2.2 本文仿真模型与经典模型对比

为了评估仿真模型的精确性,本节对不同的参数进行了设置[5,6,18]. 一方面,为了排除修复过程对传播模型的影响,本节设置修复函数($r(t)=0$). 另一方面,网络拓扑节点度服从幂指数 $\alpha=2.5$ 的幂律分布,节点平均度 $<k>=6.8$,回旋律 $\lambda=0.23$,传播概率 p_{ij}服从高斯分布 $N(0.5,0.2^2)$,检测时间 T^i_{check}服从指数分布 exp(1/40),逗留时间 T^i_{rest}服从幂指数 $\beta=1.8$ 的幂律分布.

图 4 显示了先前的模型都存在网络感染节点规模低估的问题[4-6],节点数差值最大可以达到 28000 左右. 主要原因在于它们是基于社会逻辑层分析社交蠕虫传播行为,没有考虑分层网络拓扑的传播场景. 可见

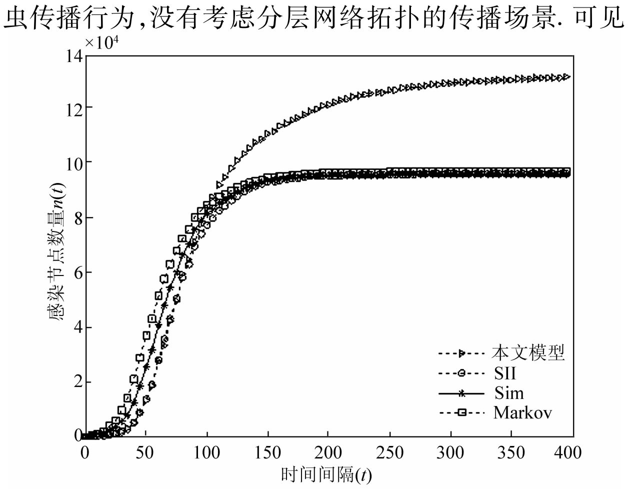


图4 不同模型的之间对比

先前的模型在刻画现实社交蠕虫传播感染趋势存在严重的偏差.

5.2.3 本文仿真模型的性能评估

(1) T^i_{rest}服从幂律分布时参数 $x_{\min}$和 β 的变化在大规模的网络拓扑下(例如国家或者城市之间),T^i_{rest}服从幂律分布($P(T^i_{\text{rest}})\propto(T^i_{\text{rest}})^{-\beta}$)[18],采用逆变换的方法生成随机样本. 假设随机变量 U 服从(0,1]之间的均匀分布,则随机变量 $T^i_{\text{rest}}=\frac{x_{\min}}{U^{1/(\beta-1)}}$[20]. 为了评估仿真模型的精确性,实验分别设置 $x_{\min}=5,15,45$ 和 $\beta=1.9,1.75,1.6$,结果如图 5 和图 6 所示.

实验结果可以看出:随着参数 $x_{\min}$减少和 β 增大,网络的感染节点规模会不断变大. 这是因为两个参数的变化情况说明逗留时间会集中分布在相对较小的数值区间,暗示了用户会更频繁地在不同地点之间活动. 因此,主机被使用的概率会增大,被感染的概率也会相应的增加.

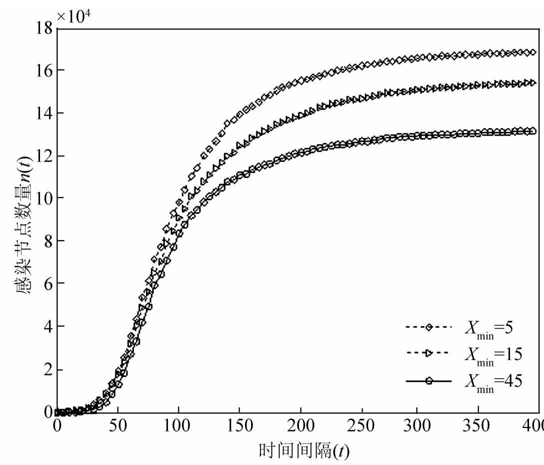


图5 参数$x_{\min}$对蠕虫传播的影响

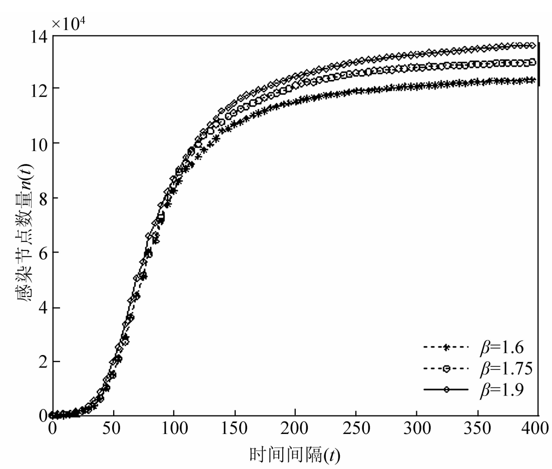


图6 参数β对蠕虫传播的影响

(2) T^i_{rest}服从指数分布时参数 θ 的变化

在小规模的网络拓扑下(例如城市内部),T^i_{rest}服

从指数分布[21]. 分别参数 $\theta = 1/20, 1/40, 1/80$,研究社交蠕虫的传播情况. 如图 7 所示. 从图 7 中可以看出:随着参数 θ 的增大,网络感染节点数目成上升的趋势. 主要原因在于参数 θ 增大表示用户将会花费更少的时间停留在某一个地点,这意味着他会在不同的地点频繁使用相应的主机,提高了主机被感染的概率.

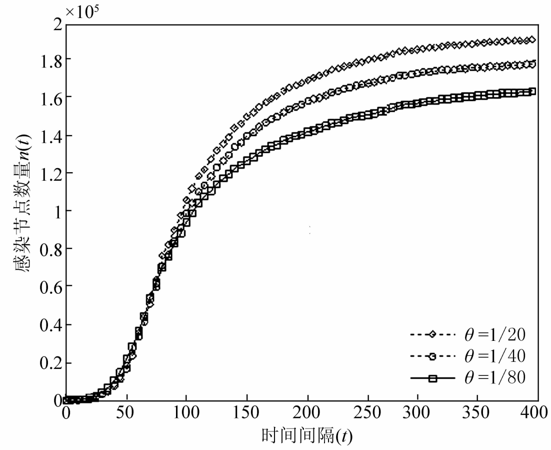


图7 参数θ对蠕虫传播的影响

(3) T^{i}_{check}服从指数分布时参数 σ 的变化

以上表述,主要分析了人类移动时间特性 T^{i}_{rest} 相关参数如何影响社交蠕虫在网络中的传播行为. 本节分析消息检测行为 T^{i}_{check} 对社交蠕虫传播的影响. T^{i}_{check} 服从指数分布,其中 $\sigma = 1/20, 1/40, 1/80$.

从图 8 所示的实验结果可以看出:用户检测消息的频率越大,这种行为会导致社交蠕虫的传播速度越快,但是感染规模会减少. 传播速度快是因为恶意的附件或者链接会更频繁地被点击和浏览. 同时,感染规模减少是因为当用户停留在某一地点未离开前,如果邮箱中有多封恶意邮件,则用户频繁的消息检测行为会导致这些邮件只会感染这个地点正在使用的主机,抑制了社交蠕虫对不同地点主机的感染效果.

(4) 不同网络拓扑对比

本节实验主要考察仿真模型被应用到不同的网络拓扑时,对社交蠕虫传播产生的影响. 如图 9 所示,以($\alpha = 2.5, \langle k \rangle = 4.5, \lambda = 0.23, p_{ij} \sim N(0.5, 0.2^2)$)为参考标准. 情况 1,如果用户更愿意把彼此加入各自的联系人列表,即 λ 变大. 情况 2,如果用户在社交网络中有更多的朋友,即 α 变小且 $\langle k \rangle$ 变大. 情况 3,如果网络防御的脆弱性提高,p_{ij} 即越大. 可以看出:以上的三种情况都能够促进社交蠕虫的传播,本文提出的仿真模型能够更加准确地反映现实世界的社交蠕虫传播情况.

5.3 加入防御手段后模型的有效性验证

本节对防御手段实现了数学形式化描述,通过仿真实验定量的分析了它们对社交蠕虫传播的影响.

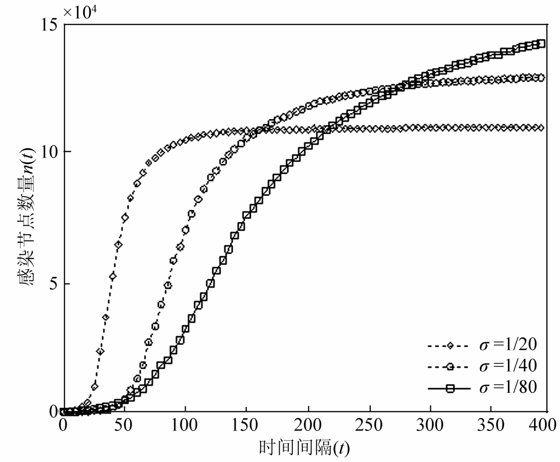


图8 参数σ对蠕虫传播的影响

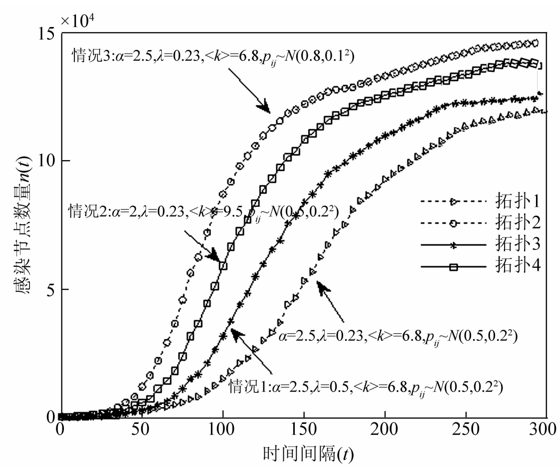


图9 不同网络拓扑对蠕虫传播的影响

5.3.1 客观防御手段

客观防御手段是一种网络相关的管理人员针对可控的网络实施的网络安全防御方案的部署,或者官方机构定期地发布补丁和升级程序来加强计算机的防御能力的防御行为. 它主要被描述为以下 4 种修复函数:① ‘Constant recovery rate’[4]:$r(t) = \rho$. ② ‘Ratio’[22]:$r(t) = \dfrac{n(t)}{\Omega}$,其中 $n(t)$ 表示时刻 t 网络的感染节点规模,Ω 表示网络节点总规模. ③ ‘Composite’[23]:$r(t) = r_0\left[1 + \dfrac{i(t)}{\Omega}\right]^{\varphi}$,其中 $i(t)$ 表示时刻 t 网络的免疫节点规模,r_0 表示初始修复率,φ 表示修复调节因子. ④ ‘Qualys’[24]:

$$r(t) = \begin{cases} 0, & t < d \\ 1 - 0.5^{\frac{t-d}{D}}, & t \geqslant d \end{cases} \tag{7}$$

其中,d 表示蠕虫首次发现时间,D 表示半衰期,即感染节点数下降 50% 的时间间隔[24].

从图 10 可以看出:即使修复函数的差别很大,但是仿真模型都能够很好地反映网络防御过程. 图 11 说明

了如果社交蠕虫能够很容易地被清除,即半衰期 D 较小时,网络节点感染数量会下降的很快. 图12 说明了如果网络相关人员能够很快地发现社交蠕虫的存在,并采取相应防御手段,即首次发现时间 d 较小时,蠕虫的传播会较早地被抑制住.

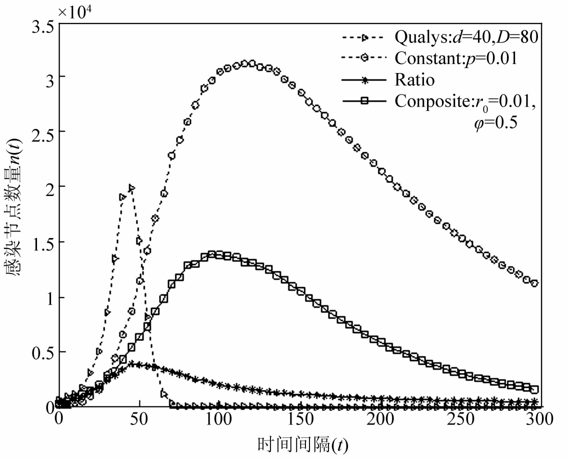


图10　不同修复函数对蠕虫传播的影响

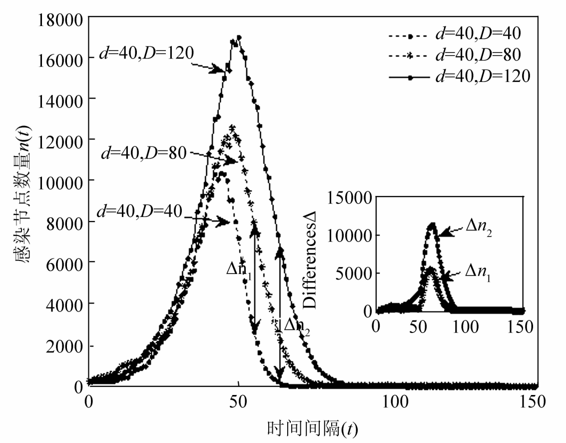


图11　参数D对蠕虫传播的影响

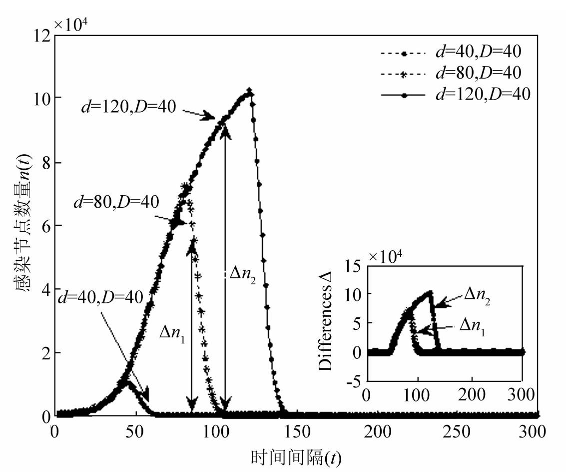


图12　参数d对蠕虫传播的影响

5.3.2　主观防御手段

主观防御手段是一种用户登录社交账户查看网页,或者登录邮箱点击附件或链接等的社交用户行为. 这些行为反映了用户的安全意识和警觉性对社交蠕虫传播的影响,如图 13 所示. 本节引入随机变量 $C_i(t)$ 表示用户点击附件或者链接的行为. 如果 $C_i(t)=1$ 表示用户点击附件或者查看了相关网页. 否则,$C_i(t)=0$. 根据文献[5],

$$P(C_i(t)=1)\sim N(\mu,\sigma^2) \tag{8}$$

一般 $\mu=0.5$,$\sigma=0.3$,$P(C_i(t)=1)$ 的概率值反映用户的安全意识.

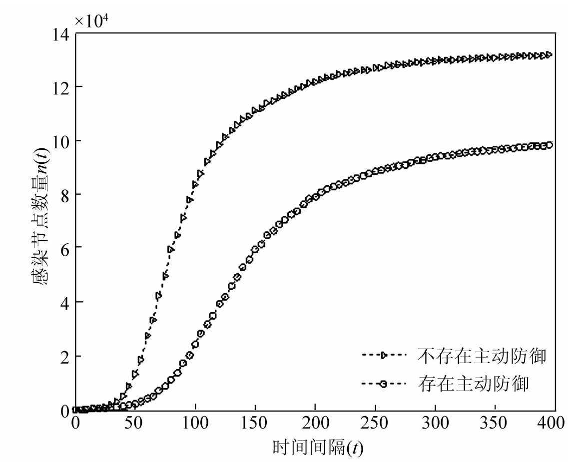


图13　用户的主观防御行为对蠕虫传播的影响

从图 13 可以看出:用户主观的防御行为可以有效地抑制社交蠕虫的传播,降低蠕虫的传播速度和感染规模. 因此,提高人们使用社交网络的安全意识,对于净化网络环境有着积极的作用.

6　理论分析

本节从理论角度分析了模型有效性验证部分的相关参数能够促进社交蠕虫传播的原因,并给出了影响社交蠕虫传播的定性化结论.

假设 $\rho_k(t)$ 表示度为 k 的节点被感染的概率,根据平均场理论:

$$\frac{\mathrm{d}\rho_k(t)}{\mathrm{d}t}=-\rho_k(t)+\eta k[1-\rho_k(t)]\Theta(\rho(t)) \tag{9}$$

式(10)考虑了单位恢复速率并且忽略高阶项($\rho_k(t)\ll 1$). η 表示传播率,$\Theta(\rho(t))$ 表示度为 k 的节点的任意邻居被感染的概率. 因此,

$$\Theta(\rho(t))=\sum_{k'}P(k'\mid k)\rho_{k'}(t) \tag{10}$$

在稳态下,$\rho_k(t)$ 关于 η 的函数,这样 $\Theta(\rho(t))$ 就是关于 η 的隐含数. 当稳态条件 $\frac{\mathrm{d}\rho_k(t)}{\mathrm{d}t}=0$,则

$$\rho_k=\frac{\beta k\Theta(\eta)}{1+\beta k\Theta(\eta)} \tag{11}$$

可知 $\rho_k \propto \Theta(\eta)$. 对于无关联的非均匀网络中,有 $P(k'|k) = kP(k)/<k>$. 因此式(10)被修改为:

$$\Theta(\eta) = \sum_{k'} \frac{k'P(k')\rho_{k'}(t)}{\sum_k kP(k)} = \sum_{k'} \frac{k'P(k')\rho_{k'}(t)}{<k>} = \frac{1}{<k>} \sum_{k'} \frac{\beta k'^2 P(k')\Theta(\eta)}{1+\beta k'\Theta(\eta)} \tag{12}$$

式(13)可以用连续化的方法近似求得 $\Theta(\eta)$,可知 $\Theta(\eta) \propto \eta$. 最后,网络中任意一个节点被感染的概率 ρ 为

$$\rho = \sum_k P(k)\rho_k \tag{13}$$

综上所述,可以得出结论 $\rho \propto \eta$. 根据 BA 模型中获得结果 $\rho \sim e^{-1/m\eta}$ 能够验证以上结论[21]. 因为上节实验部分中用户频繁的移动和检查消息,以及人类两种行为的时间同步性和网络拓扑参数的改变都能够增强蠕虫的传播率,因此会增大主机被感染的概率,进而增强了社交蠕虫在网络中的传播率,即扩大网络主机的感染规模.

定义 1 社交蠕虫的传播能力 A_{spread} 定义为一个二元组 $<V_{\text{spread}}, M_{\text{infect}}>$. 其中,$V_{\text{spread}}$ 表示社交蠕虫的传播速度,M_{infect} 表示社交蠕虫的最终感染规模.

根据实验仿真分析,可以得出

$$T^i_{\text{rest}} \propto f_r(\hat{\beta}, \hat{\theta}) \tag{14}$$

$$M_{\text{infect}} \propto f_M(\hat{T}^i_{\text{rest}}, S(D_{T^i_{\text{check}}}, D_{T^i_{\text{rest}}}), T^i_{\text{check}}, \text{topo}(\hat{\alpha}, \lambda, <k>, p_{ij}), \hat{A}_{\text{defense}}) \tag{15}$$

其中,$\text{topo}(\hat{\alpha}, \lambda, <k>, p_{ij})$ 是关于网络拓扑参数的隐含数. A_{defense} 表示网络防御能力的量化值. 假设 $D_{T^i_{\text{check}}}$ 和 $D_{T^i_{\text{rest}}}$ 分别表示消息检测时间分布和逗留时间分布,则 $S(D_{T^i_{\text{check}}}, D_{T^i_{\text{rest}}})$ 表示消息检测时间分布与逗留时间分布之间样本的相似度,$S(D_{T^i_{\text{check}}}, D_{T^i_{\text{rest}}}) \in (0,1]$.

$$V_{\text{spread}} \propto f_V(\hat{T}^i_{\text{check}}, S(D_{T^i_{\text{check}}}, D_{T^i_{\text{rest}}}), \text{topo}(\hat{\alpha}, \lambda, <k>, p_{ij}), \hat{A}_{\text{defense}}) \tag{16}$$

注意:假设变量 x 和 $\hat{y}$,则上述分析的所有隐函数 $f(x, \hat{y})$ 都是关于 x 增函数,关于 y 减函数. 基于式(15)和(16),本文能够获得社交蠕虫传播能力 $A_{\text{spread}} = \langle V_{\text{spread}}, M_{\text{infect}} \rangle$ 的定性结论.

7 结论

在互联网技术高速发展的今天,社交蠕虫的传播已经具有跨学科性和多维性的特点. 本文深入地研究了分层网络拓扑结构和人类移动时间特性对社交蠕虫的影响,提出了一种新的社交蠕虫传播行为的仿真模型. 该模型解决了先前研究存在的两个重要问题:网络拓扑的不完整性和传播建模的片面性,更全面真实地反映社交蠕虫的传播行为. 同时,本文做了大量的仿真实验分析了人类消息检测行为、移动行为、网络拓扑结构以及网络修复过程对社交蠕虫传播造成的影响. 此外,本文从理论角度给出了社交蠕虫传播能力的定性分析,为网络防御提供了有力的支持.

参考文献

[1] FOSSI M B J. Symantec Internet Security Threat Report 2013 [R]. San Francisco:Symantec Corporation,2014. 63 –75.

[2] ZOU C C, GONG W T D. Code red worm propagation modeling and analysis [A]. Proceedings of the 9th ACM Conference on Computer and Communications Security [C]. New York:ACM,2002. 138 – 147.

[3] KIM J, RADHAKRISHNAN S D S K. Measurement and analysis of worm propagation on Internet network topology [A]. Proceedings of the 13th International Conference on Computer Communications and Networks [C]. New Jersey:IEEE,2004. 495 – 500.

[4] CHEN Z J C. Spatial-temporal modeling of malware propagation in networks [J]. IEEE Transactions on Neural Networks,2005,16(5):1291 – 1303.

[5] ZOU C C, TOWSLEY D G W. Modeling and simulation study of the propagation and defense of internet e-mail worms [J]. IEEE Transactions on Dependable and Secure Computing,2007,4(2):105 – 118.

[6] WEN S, ZHOU W, ZHANG J, et al. Modeling propagation dynamics of social network worms [J]. IEEE Transactions on Parallel and Distributed Systems, 2013, 24(8):1633 – 1643.

[7] 冯朝胜,袁丁,卿昱,等. P2P 网络中激发型蠕虫传播动态建模[J]. 电子学报,2012,40(2):300 – 307.
FENG Chao-sheng, YUAN Ding, QING Yu, et al. Dynamic modeling of reactive worm propagation in P2P network [J]. Acta Electronica Sinica,2012,40(2):300 – 307. (in Chinese)

[8] 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 Chao-sheng, QIN Zhi-guang, YUAN Ding, et al. Modeling propagation and immunization of passive worms in peer-to-peer networks [J]. Acta Electronica Sinica, 2013,41(5):884 – 889. (in Chinese)

[9] 严博,吴晓平,廖巍,等. 基于随机进程代数的 P2P 网络蠕虫对抗传播特性分析[J]. 电子学报,2012,40(2):293 – 299.
YAN Bo, WU Xiao-ping, LIAO Wei, et al. Propagation Characteristics analysis of worm-anti-worm in P2P network based on stochastic process algebra [J]. Acta Electronica Sinica,2012,40(2):293 – 299. (in Chinese)

[10] 刘昕,贾春福,石乐义,等.基于社会计算的 IM 恶意代码防御机制[J].电子学报,2012,41(6):1130－1139.
LIU Xin,JIA Chun-fu,SHI Le-yi,et al. A collaborative defending scheme based on social computing against IM malicious codes [J]. Acta Electronica Sinica,2012,41(6):1130－1139.(in Chinese)

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

[12] YU W,WANG X,CALYAM P,et al. Modeling and detection of camouflaging worm [J]. IEEE Transactions on Dependable and Secure Computing,2011,8(3):377－390.

[13] MANNA P K,CHEN S R S. Inside the permutation-scanning worms:Propagation modeling and analysis [J]. IEEE/ACM Transactions on Networking,2010,18(3):858－870.

[14] FAGHANI M R N U. A study of XSS worm propagation and detection mechanisms in online social networks [J]. IEEE Transactions on Information Foresics and Security,2013,8(11):1815－1826.

[15] YAN G,CHEN G,EIDENBENZ S L N. Malware propagation in online social networks:nature,dynamics,and defense implications [A]. Proceedings of the 22nd Annual Conference on Computer Security Applications [C]. New Jersey:IEEE,2011.245－256.

[16] FAN W,YEUNG K W K. Assembly effect of groups in online social networks [J]. Physica A:Statistical Mechanics and its Applications,2013,392(5):1090－1099.

[17] PASTOR-SATOrras R V A. Epidemic dynamics in finite size scale-free networks [J]. Physical Review E,2002,65(3):35108.

[18] SONG C,KOREN T,WANG P B A S. Modelling the scaling properties of human mobility [J]. Nature Physics,2010,6(10):818－823.

[19] MOORE D,SHANNON C. The Nyxem Email Virus:Analysis and Inferences [R]. San Diego:CAIDA,2006.1－10.

[20] TANIZAKI H. Computational Methods in Statistics and Econometrics [M]. Boca Raton:CRC Press,2004.122－149.

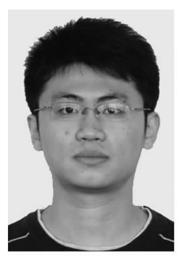
[21] LIANG X,ZHAO J,DONG L X K. Unraveling the origin of exponential law in intra-urban human mobility [J]. Scientific Reports,2013,3:1－7.

[22] BOGUN A M A N,PASTOR-SATORRAS R V A. Epidemic spreading in complex networks with degree correlations [A]. Proceedings of the XVIII Sitges Conference on Statistical Mechanics [C]. Berlin:Springer,2003.1－23.

[23] ZHANG D W Y. SIRS:Internet worm propagation model and application [A. Proceedings of 2010 International Conference on Electrical and Control Engineering [C]. Wuhan:IEEE,2010.3029－3032.

[24] KANDEK W. The Laws of Vulnerabilities [R]. Paris:Qualys Inc,2009.1－18.

作者简介



王天博 男,1986 年出生,黑龙江齐齐哈尔人,北京航空航天大学在读博士,计算机学会学生会员.主要研究方向为网络与信息安全,入侵检测技术和信息对抗.
E-mail:tianbo_king@163.com

夏春和 男,1965 年出生,江苏海安人,北京航空航天大学教授.主要研究方向为网络与信息安全,信息对抗,云安全和网络策略.
E-mail:xch@buaa.edu.cn