社交网络蠕虫仿真建模研究

孙 鑫 刘衍珩 朱建启 李飞鹏

[1)](吉林大学计算机科学与技术学院 长春 130012)

[2)](吉林大学符号计算与知识工程教育部重点实验室 长春 130012)

摘 要 随着互联网虚拟社交网络的发展,借助网络用户进行传播的社交网络蠕虫已经成为威胁网络安全的重大隐患之一.文中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;通过分析网络用户活动的习惯特性,构建了宏观网络上离散的基于用户习惯的社交网络访问模型;从而实现了一个适用于社交网络蠕虫传播研究的仿真系统.实验分析了模型中用户相关因素对蠕虫传播的影响,验证了仿真系统的有效性.最后,利用该系统仿真了社交网络蠕虫在异质社交网络中的传播,揭示了网络拓扑结构和社交网络混杂程度对蠕虫传播的影响,为社交网络蠕虫相关研究提供了重要支持.

关键词 社交网络蠕虫;传播模型;人类行为;博弈理论;仿真

中图法分类号 TP309 **DOI 号**: 10.3724/SP.J.1016.2011.01252

Research on Simulation and Modeling of Social Network Worm Propagation

SUN Xin LIU Yan-Heng ZHU Jian-Qi LI Fei-Peng

[1)](*College of Computer Science and Technology, Jilin University, Changchun* 130012)

[2)](*Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun* 130012)

Abstract Along with the rapid development of social networks, social network worms have constituted one of the major internet security problems. In the paper, the social network worm propagation is discussed from the viewpoint of social engineering, and a human behavior model based on game theory is presented for predicting the expected actions of network user encountered with worm messages. By analyzing network users' diurnal activity behaviors, a discrete social network accessing model is proposed to characterize the general human habit of accessing certain social network. Then a simulation system is constructed with the two models and validated by simulations on the effect of human factors on propagations. Finally, the results of simulation experiments on artificial social networks show the impacts of network topology and number of hybrid nodes on social network worm propagation and indicate that this simulation system can provide great support for social network worm related researches.

Keywords social network worm; propagation model; human behavior; game theory; simulation

1 引 言

社交型网络蠕虫(以下简称"社交网络蠕虫")是利用社会工程学以各种方式诱惑用户点击进行传播的一类蠕虫,具有隐蔽性高、生存周期长和难以根除等特点,其传播难以通过发布漏洞补丁等技术的手段进行有效的控制,因此潜在的危害更为严重.同

收稿日期:2011-01-05;最终修改稿收到日期:2011-06-07.本课题得到国家自然科学基金(60973136、61073164)、国家发改委下一代互联网业务试商用及设备产业化专项(CNGI-09-01-11)、国家科技部国际合作与交流专项(2008DFA12140)、吉林省科技发展计划:青年科研基金(201101033)和吉林大学创新项目(450060445169)资助.**孙 鑫**,男,1984 年生,博士研究生,主要研究方向为网络安全、可信计算.E-mail: sunxin1984@yahoo.com.cn. **刘衍珩**(通信作者),男,1958 年生,博士,教授,博士生导师,主要研究兴趣为网络安全、可信计算、移动 IP 和 QoS、传感器网络等.E-mail: lyh_lb_lk@yahoo.com.cn. **朱建启**,男,1976 年生,博士,讲师,主要研究方向为网络安全、信息隐藏. **李飞鹏**,男,1986 年生,硕士研究生,主要研究方向为网络安全.

时，实际环境中网络管理员以及网络用户安全防范意识的匮乏，为社交网络蠕虫的传播和生存提供了滋生的温床.随着网络用户数量的日益增长和各类形式虚拟社交网络的快速发展，借助网络用户进行传播的社交网络蠕虫已经成为威胁网络安全的重大隐患之一.

目前学术界针对蠕虫传播的研究主要是通过建立仿真模型进行的[1-10]，建立有效的蠕虫传播仿真模型，有利于系统的分析和理解蠕虫的发作特性、传播规律和发展趋势以及提出有效的防御策略.

蠕虫病毒自诞生以来一直是网络安全的重要威胁之一[1]，对蠕虫传播行为建立仿真模型也一直是学术界研究的重点和难点.早期的研究仅仅是考虑计算机蠕虫和生物病毒之间在繁殖传播等方面的相似性，简单地假设为接触即感染的方式进行数学建模.通过将生物病毒传播建模中广为应用的 SIS、SIR 等传播模型引入到计算机蠕虫建模[2-3]，分析和预测蠕虫传播特性和趋势.

然而依赖互联网传播的计算机蠕虫毕竟不同于生物病毒，学术界开始逐渐认识到网络拓扑、线路带宽以及用户对抗措施等外界因素对蠕虫传播的影响.如：Zou 等人[4]考虑到网络用户对蠕虫的对抗措施以及蠕虫迅速传播导致路由器阻塞而引起的时延等因素对蠕虫传播速度的影响，提出了 Two-Factor 模型，将蠕虫感染率、主机免疫率等参数表示为时间 T 的函数，根据网络中被感染主机总数的增加或减少动态调整其取值；Yang 等人[5]以 Rose 邮件蠕虫作为案例，通过建立数学解析模型研究了在不同社交场景（Print Service Office，Internet Café，Friendship Network）下加入免疫因素后蠕虫的传播；夏春和等人[6]抽象出描述 P2P 节点空间结构特征的命题，并将其引入蠕虫传播规律的推导过程，建立了 P2P 蠕虫在 3 种典型结构化对等网中的传播模型，给出刻画 P2P 蠕虫传播能力的函数，并揭示了覆盖网拓扑对蠕虫传播的负面影响.

随着对计算机蠕虫传播机制研究的深入，蠕虫传播方式的差异性逐渐受到重视.卿斯汉等人[7]针对即时通信蠕虫（IM 蠕虫）进行了全面剖析，对 IM 蠕虫的相关研究进行了综述和展望.荆继武等人[8]针对拓扑相关网络蠕虫，提出了一个完整的数据包级仿真模型，并通过仿真拓扑相关蠕虫的传播，分析了逻辑拓扑结构和蠕虫代码启动方式对蠕虫传播的影响.

虽然上述研究已经考虑到了网络拓扑和用户免疫等因素，并对特殊环境下蠕虫行为的仿真建模.但是对于利用社交网络进行传播的蠕虫而言，其特殊性在于感染率更大程度上受到用户自身行为因素的影响，因此以上研究不适合于利用社会工程学进行传播的蠕虫仿真建模.Gray 等人[9]首次研究了利用社会工程学进行传播的蠕虫，并将需要网络用户下载并点击进行传播的一类蠕虫统称为“被动蠕虫”，研究了共享恶意文件数量和文件名受欢迎度的变化对被动蠕虫传播的影响.Zou 等人[10]考虑用户行为具有随机性和不确定性等特点，通过建立实验环境上的仿真模型代替数学理论分析，对用户查收邮件和打开附件的行为进行随机建模，较好地仿真了基于电子邮件的社交网络蠕虫的传播.但在该模型中所有用户查收邮件和打开附件的概率被假设为定值，而事实上，用户作为有主动行为能力的个体，其查收邮件的时间间隔根据用户的习惯有所不同，且随着对蠕虫危害性认识的提高，用户打开蠕虫邮件的概率也随时间逐渐降低，因此该模型在社交网络蠕虫感染行为的模拟上还存在不足.Sun 等人[11]提出了采用博弈理论对蠕虫感染行为建模的思想，研究了用户习惯对蠕虫传播的影响，然而该研究尚未给出合理的计算模型.

近年来，对人类行为的研究受到了广泛的重视，并取得了一定的进展[12-13].本研究认为，社交网络蠕虫的传播与网络用户的行为密切相关，这些行为包括用户的习惯、用户的安全防范意识以及网络中用户的拓扑分布和关联等.结合社会工程学和博弈理论的相关知识，对网络用户与社交网络蠕虫的交互行为以及用户访问社交网络的习惯规律进行建模，并精确地描述网络用户之间的社交关系，是实现仿真社交网络蠕虫传播的重要途径.

本文首先从社会工程学的角度探讨社交网络蠕虫的传播行为，分析并量化了影响社交网络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，得到了离散时间内蠕虫在每个节点的随机感染概率；然后，通过分析网络用户活动的习惯特性，构建了宏观网络上离散的基于用户习惯的社交网络访问模型，用于选择单位时间内社交网络的活动用户，进而实现了一个适用于社交网络蠕虫传播研究的仿真系统.应用 BBV 拓扑生成算法构造了符合真实社交网络环境的加权网络拓扑仿真环境，并根据国际 Internet 数据分析合作组织（CAIDA）的实证研究结果，实验分析了若干用户相关因素（如：用户访问频率和安全意识以及专杀工具

发布及时性)对蠕虫传播的影响,验证了该系统的有效性.最后,利用该系统仿真分析了社交网络蠕虫在异质的社交网络和混杂的社交网络环境中的传播,揭示了网络拓扑和社交网络混杂程度对其传播的影响.

2 社交网络蠕虫传播建模

2.1 社交网络和社交网络蠕虫定义

社交网络.社交网络源自网络社交,其起点是电子邮件[14].BBS、即时通信和 P2P 内容共享推动了网络社交的发展,随着 WEB2.0 技术的推广应用,在线社交网站的出现进一步丰富了社交网络的形式.目前互联网存在的主流社交网络包括:电子邮件网络、P2P 内容共享网络、即时聊天网络和在线社交网站等.

为了明确界定在本文中社交网络蠕虫的研究范围并与其它各类网络蠕虫相区别,将社交网络蠕虫定义为:社交网络蠕虫是一段不依赖于特定系统漏洞的恶意程序,利用社会工程学的手段欺骗网络用户点击才能执行并感染电脑系统.社交网络蠕虫的传播难以通过发布系统漏洞补丁等技术手段进行免疫,只能通过强化用户安全意识等方式进行防范.

2.2 社交网络蠕虫传播机制

社交网络蠕虫的传播包括 3 个阶段:首先利用自身的伪装欺骗用户点击并执行蠕虫文件感染主机;然后对感染用户的好友信息进行采集;最后利用采集到的好友信息通过虚拟社交网络进行蠕虫文件的扩散.

本文将用户状态分为易感染状态和已感染状态.在此不考虑免疫状态是由于社交网络蠕虫难以通过安装系统补丁的方式进行免疫.已感染状态可以被修复并回到易感染状态,但感染的概率会随着用户安全意识的提高而逐渐降低.

2.3 用户行为博弈模型

社会工程学的攻击手段是利用人性弱点通过欺骗等方法骗取对方信任,从而突破安全防御.社交网络蠕虫利用社会工程学手段将自己进行伪装,诱惑用户对含有蠕虫代码的文件进行点击,从而达到感染用户主机的目的.网络用户与蠕虫之间的诱惑与反诱惑的行为,可以视为网络用户与蠕虫代码编写者之间智慧的博弈[15].一方面蠕虫代码编写者会尽可能地伪装蠕虫,加入丰富的诱惑信息,骗取网络用户点击蠕虫文件;另一方面网络用户会根据自己先前的经验判断该文件的安全性,决定是否点击.如 2.2 节所述,对社交网络蠕虫不存在绝对免疫的情况,因此用户打开蠕虫文件的概率,可视为蠕虫在该用户节点的感染率.

本研究通过建立网络用户与蠕虫文件之间的行为博弈模型,从微观上描述社交网络蠕虫感染行为的特点、量化代价和收益,从而得到蠕虫在不同网络节点的感染概率.博弈模型参与对象包括:(1)网络用户:利用互联网进行社交活动的人,包括电子邮件用户、P2P 用户和在线社交网站注册用户等,用户作为具有主观能动性的个体,会权衡打开一个陌生文件的可能收益及损失;(2)蠕虫文件:含有恶意蠕虫代码的计算机文件,利用电子邮件、P2P 系统以及在线社交网站等虚拟社交网络进行传播,通过伪装诱惑用户点击实现对计算机系统的感染.

在网络用户与蠕虫文件的行为博弈模型中,单个节点的实际对抗中只有网络用户是真正具有思考能力的智能体,根据博弈理论在博弈双方只有一方能够做出智慧决策的情况下,对非智能体一方收益值的定义是与博弈无关的[15-16],因此,即便网络用户与网络蠕虫的得失不存在必然的对立平衡状态,我们也可以做如下假设:网络蠕虫的所失即是网络用户的所得,即零和博弈.所以,下面重点对具有理性思考能力的网络用户的代价/收益进行量化.

博弈包括参与对象、参与对象的偏好信息、可获得的策略行动和支付函数.通过分析网络用户行为的代价/收益,给出如下简要的量化表示:

(1)打开带有蠕虫代码的文件,损失量化表示为 $D(Damage)$.

用户 i 打开蠕虫文件的损失也就是蠕虫代码编写者在该节点所期望的回报.对损失 D 进行量化,取决于用户在社交网络中的地位、节点存储信息的重要性以及蠕虫程序的危害程度.用户节点的度 $Degree(i)$ 体现了用户在社交网络中的地位和间接反映该用户存储信息的重要性,同时,由于网络节点存储信息的重要程度带有较大的随机性和不确定性,因此引入高斯函数 $G\infty N(\mu_g,\sigma_g^2)$ 刻画信息存储在网络中的随机分布,得到 $-Degree(i)^G$(负值表示损失);随着安全机构不断发布病毒公告,用户对蠕虫危害程度的认识逐渐提高,对危害程度的量化需要综合考虑蠕虫自身危害性和传播深度(蠕虫首次出现至今所经过的时间)的影响,因此定义用户对蠕虫的危害程度的认识为时间 t 的增函数

$$Harm(t)=HARM\times t^{\varepsilon} \tag{1}$$

$HARM$ 是专家对蠕虫程序危害性的打分;ε 表示用

户对网络安全的关注力度，ε 越大表示用户对网络安全的重视程度越高.

用户 i 的损失定义为

$$D(i) = -Degree(i)^G - Harm(t) \tag{2}$$

(2) 丢弃带有恶意蠕虫代码的文件，奖励量化表示为 $R(Reward)$.

丢弃蠕虫文件的奖励是采取丢弃可疑的蠕虫文件后，计算机系统免于遭受的损失，一般用打开蠕虫文件时计算机系统损失的负值来表示

$$R(i) = -D(i) \tag{3}$$

(3) 打开普通文件的收益量化表示为 $I(Income)$.

用户通过浏览文件可以获取有价值的信息. 假设文件的价值与文件提供者的信誉和双方交往频繁程度相关，那么文件的价值可以由文件提供者的度和用户之间的边权值共同体现

$$I(i) = I + \frac{w(i,j) \times Degree(j)}{w(i,j) + Degree(j)} \tag{4}$$

其中：I 表示打开普通文件的基本收益值；j 为发送节点.

(4) 放弃普通文件的代价为 $C(Cost)$.

打开普通文件收益的负值即是放弃该文件的代价

$$C(i) = -I(i) \tag{5}$$

定义 1.　行为博弈模型表示为三元组 $G=(\Gamma,S,U)$，其中的 3 个要素：

(a) $\Gamma=\{1,2,\cdots,n\}$，参与对象集合；

(b) $S=(S_1,S_2,\cdots,S_n)$，参与对象的策略空间，其中 S_i 表示参与对象 i 的非空策略集合，即 $\forall i\in\Gamma$，$\exists S_i\neq\varnothing$；

(c) $U=(U_1,U_2,\cdots,U_n)$，每位参与对象定义在策略空间 $S=(S_1,S_2,\cdots,S_n)$ 上的收益函数.

定义 2.　如果对任何博弈结果，所有参与人的收益和是零，那么，这个博弈称为零和博弈，即任何策略组合 $s^{(1)}\in S_1,\cdots,s^{(n)}\in S_n$，都有 $\sum_{i=1}^{n}U_i(s^{(1)},s^{(2)},\cdots,s^{(n)})=0$.

图 1 给出了用户行为博弈模型的图表示，为了简便分析，在该模型中网络用户面对陌生文件的选择空间仅包含打开(Open)、丢弃(Discard)两个动作；网络用户需要通过先验知识的积累对陌生文件的可能情况进行判断{蠕虫文件：Worm，普通文件：Normal}.

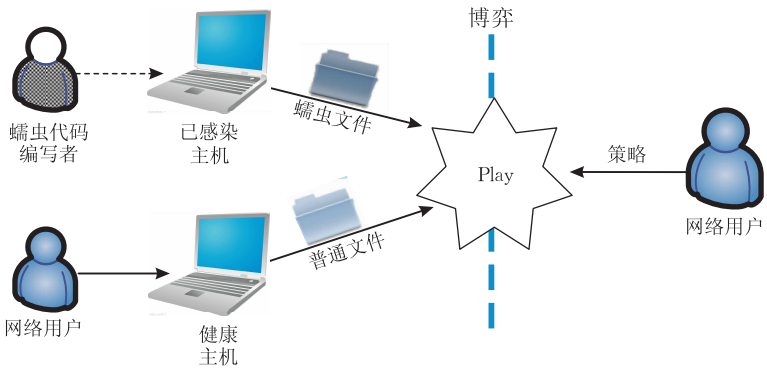


图 1　用户行为博弈模型

在社交网络中节点 i 的行为博弈模型定义如下(收益函数的取值如表 1 所示)：

$\Gamma=\{1,2\}=\{User(i),Worm\}$，

$S=\{Open,Discard,Worm,Normal\}$，

$U=\{U(Open,Worm),U(Discard,Worm),$

$U(Open,Normal),U(Discard,Noraml)\}$.

表 1　博弈收益矩阵

	收益函数	
	Worm	Normal
Open	$U(Open,Worm)=D$	$U(Open,Normal)=I$
Discard	$U(Discard,Worm)=R$	$U(Discard,Noraml)=C$

定义 3.　在一个给定的博弈 $G=(\Gamma,S,U)$ 中，策略组合 s^* 是一个纳什均衡，当且仅当 $\forall i\in\Gamma$，$\forall s_i\in S_i$ 时，有 $u_i(s_i^*,s_{-i}^*)\geq u_i(s_i,s_{-i}^*)$ 或者 $\forall i\in\Gamma$，$s_i^*\in\arg\max_{s_i\in S_i}u_i(s_i,s_{-i}^*)$，其中 $s_{-i}^*=(s_1,\cdots,s_{i-1},s_{i+1},\cdots,s_n)$. 即 s_i^* 是给定其它参与对象选择的策略组合为 s_{-i}^* 的情况下的最优策略，s^* 又称为纯策略纳什均衡.

定义 4.　在一个给定的博弈 $G=(\Gamma,S,U)$ 中，假定参与对象 i 有 m 个纯策略：$S_i=\{s_{i1},s_{i2},\cdots,s_{im}\}$，那么，概率分布 $X_i=\{x_{i1},x_{i2},\cdots,x_{im}\}$ 称为参与对象 i 的一个混合策略，这里，对于所有的 $k=1,2,\cdots,m$，x_{ik} 表示参与对象 i 选择策略 s_{ik} 的概率，$0\leq x_{ik}\leq 1$，且 $\sum_{k=1}^{m}x_{ik}=1$.

混合策略解释了一个参与对象对其它参与对象所采取行动的不确定性,它描述了参与对象在给定信息下以某种概率分布随机选择不同的行动或策略.在复杂多变的网络环境中,经过伪装的蠕虫文件具有较高的隐蔽性和诱惑性,使得网络用户的行为具有随机性和模糊性,因此网络用户是以一定的概率打开陌生文件,当陌生文件寄生有蠕虫代码时,此概率即是蠕虫在该用户节点的感染率.

当参与对象 I 选择混合策略 $\boldsymbol{x}$,参与对象 II 选择混合策略 $\boldsymbol{y}$ 时,参与对象 I 和 II 的预期收益分别为

$$E_{\mathrm{I}}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}\boldsymbol{A}\boldsymbol{y}^{\mathrm{T}} \tag{6}$$

$$E_{\mathrm{II}}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{y}(-\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}^{\mathrm{T}} \tag{7}$$

其中 $\boldsymbol{A}$ 和 $-\boldsymbol{A}^{\mathrm{T}}$ 分别为参与对象 I 和 II 的收益矩阵.

假设在社交网络中用户 i 打开陌生文件的概率为 p_i,那么丢弃该文件的概率为 $1-p_i$;同样当陌生文件是蠕虫文件的概率为 q_i,是普通文件的概率即为 $1-q_i$.因此,网络用户与蠕虫文件的行为博弈模型中策略的定义如下:$\boldsymbol{P}_i=(p_i,1-p_i)$ 和 $\boldsymbol{Q}_i=(q_i,1-q_i)$,收益矩阵如表 1 所示.因此在社交网络节点 i,用户与蠕虫文件的预期收益分别为

$$E_{\mathrm{user}}(\boldsymbol{P}_i,\boldsymbol{Q}_i) = \boldsymbol{P}_i\boldsymbol{A}\boldsymbol{Q}_i^{\mathrm{T}} \tag{8}$$

$$E_{\mathrm{worm}}(\boldsymbol{P}_i,\boldsymbol{Q}_i) = \boldsymbol{Q}_i(-\boldsymbol{A}^{\mathrm{T}})\boldsymbol{P}_i^{\mathrm{T}} \tag{9}$$

在实际环境中,网络用户并不能确定陌生文件是否含有恶意代码,但是依据博弈理论的基本思想[17]:网络用户应假设对手(蠕虫代码编写者)会最大程度地降低网络用户对蠕虫文件的防范,破坏其系统,使得网络用户的收益 $E_{\mathrm{user}}(\boldsymbol{P}_i,\boldsymbol{Q}_i)$ 最小化.相反,网络用户会尽可能地保护自身的系统免受侵犯,使得蠕虫文件的收益 $E_{\mathrm{worm}}(\boldsymbol{P}_i,\boldsymbol{Q}_i)$ 最小化.

因此,通过对式(10)对 $E_{\mathrm{worm}}(\boldsymbol{P}_i,\boldsymbol{Q}_i)$ 关于 q_i 求偏导,可以得到蠕虫在节点 i 的感染率.

$$\frac{\partial E_{\mathrm{worm}}(\boldsymbol{P}_i,\boldsymbol{Q}_i)}{\partial q_i} \tag{10}$$

2.4 基于用户习惯的社交网络访问模型

互联网社交网络中在线用户的规模决定了社交网络蠕虫潜在的感染范围,是影响社交网络蠕虫传播的重要因素.iResearch 调查显示 2009 年中国个人电子邮箱活跃用户规模达 2.18 亿①,各类 P2P 应用也具有规模巨大的用户群体,同时,随着以 SNS 为核心理念的 FACEBOOK 和开心网等在线社交网站的成功运作,截至 2010 年 6 月我国使用社交网站的网民规模达到 2.1 亿②.网络用户在社交网络中的在线规律和行为习惯是影响社交网络蠕虫传播的重要因素,由于社交网络中的节点都是参与互联网的"网民"的抽象,因此,社交网络中在线节点的数量随着"网民"的行为而呈现高度的动态性.

用户访问网络的时间间隔具有随机性和习惯性等特点,统计学中指数分布常被用来表示独立随机事件发生的时间间隔(这类独立随机事件在统计学中通常用指数分布来表示).在文献[10]中 Zou 等人假设用户检查电子邮件的时间间隔服从指数分布,并通过仿真验证了该假设的合理性.用户访问网络的时间间隔除了具有此类独立随机事件特性外,还受社交网络访问习惯的影响,具体表现在不同社交网络中用户平均访问频率不同、同一社交网络中不同用户的访问频率不同,为了能够准确地描述网络用户宏观上在线规律和访问模式,本文提出基于网络用户活动习惯的离散社交网络访问模型.首先假设网络用户 i 访问社交网络的时间间隔 x_i 的概率分布符合指数分布 $f(x_i)$

$$f(x_i) = \begin{cases} \dfrac{1}{\lambda(i)}^{-\frac{1}{\lambda(i)}x_i}, & x_i \geq 0 \\ 0, & x_i < 0 \end{cases} \tag{11}$$

中国互联网络信息中心的调查显示用户使用社交网络的频率具有较大的差异性,而在使用频率较高的用户群中,以维系众多朋友关系为目的的用户居多,具有朋友关系越多访问频率越高的趋势.因此不妨做如下假设:网络用户 i 访问社交网络频率的随机性与网络用户节点的度具有相关性,即 $E(f(x_i))\sim Degree(i)$.

由于指数分布 $f(x_i)$ 的期望 $E(f(x_i))=\lambda(i)$,结合式(11),给出 $\lambda(i)$ 的如下定义

$$\lambda(i) = avgT \times \ln\left(\frac{Degree(i)}{avgDegree}\right) \tag{12}$$

$avgT$ 表示用户访问社交网络时间间隔的期望,不同社交网络的 $avgT$ 值不同;实证表明社交网络的度分布具有幂率特性(如表 2 所示),在双对数坐标系下能够线性拟合,因此对 $\dfrac{Degree(i)}{avgDegree}$ 取对数能够较好的表现某节点用户访问社交网络时间间隔与整体期望的偏差.

定义用户 i 在活动周期 T 内访问社交网络的概率分布向量为

$$\boldsymbol{OnLine}(i) = [p(i)_1,\cdots,p(i)_k,\cdots,p(i)_n]$$

① http://www.iresearch.com.cn
② http://www.cnnic.net.cn

$$n=\frac{T}{\Delta t},\ \Delta t \text{ 为一个时间单位} \tag{13}$$

其中 $p(i)_k$ 表示网络用户 i 在时间段 Δt_k 访问社交网络的概率.

网络用户 i 在时间段 Δt_k 访问社交网络的概率为

$$p(i)_k=\alpha\times\begin{cases}\Delta T-Random(x_i), & \Delta T>Random(x_i)\\ 0, & \Delta T\le Random(x_i)\end{cases} \tag{14}$$

其中 $Random(x_i)$ 为网络用户 i 访问社交网络的随机时间间隔,取值符合式(11)所示的指数分布;ΔT 为距离用户 i 上次访问网络的时间,α 用于宏观上调节社交网络中用户的在线比率.

3　社交网络仿真拓扑构造

社交网络蠕虫所赖以生存和传播的互联网社交网络已经成为一种新质的社会形态,不同介质构成的社交网络在组织方式、拓扑结构和地理分布等方面各不相同.目前,学术界已经对来自不同领域的大量实际网络的拓扑特征进行了广泛的实证性研究[18-22],结果表明在电子邮件网络[18-19]和 P2P 对等网络[20-21]以及社交网站[22]等虚拟社交网络中均表现出幂率特性,如表 2 所示.

真实的社交网络是验证蠕虫传播模型的最佳环境,然而考虑到法律和经济等因素,这样的实验是无法完成的,学术界在蠕虫传播建模的研究中,大多采用无权图的网络拓扑进行仿真实验[10,23].然而这种无权图并不适合于社交网络蠕虫仿真的研究,因为它忽略了用户间信任关系对社交网络蠕虫传播的影响.社交网络中用户间的信任是通过长时间的频繁交往建立起来的,用户间数据的互访越频繁,之间的信任度也就越大.在实际环境中网络用户会对信任用户的数据信息疏于防范,因此社交网络中节点间的权重也是影响社交网络蠕虫传播的重要因素.

表 2　实证社交网络拓扑参数

网络	类型	幂指数
Email[19]	电子邮件	1.5/2.0
Gnutella[20]	P2P	2.30
Flickr[22]	社交网站	2.74
MySpace[22]	社交网站	3.10

Barrat 等人在文献[24]中成功地建立了一个演化加权模型(BBV),该模型通过调节演化参数能够仿真出很多真实网络拓扑,同时节点度、点权(与节点相连边的权值总和)和边权的分布具有幂率特性.在该模型中当在节点新加入一条边后,与该节点相连的所有边的权值根据式(15)进行调整.

$$\Delta w(i,j)=\delta\frac{w(i,j)}{S_i} \tag{15}$$

其中:S_i 是与节点 i 相连的边的权值总和;δ 为演化参数;j 为节点 i 的邻居节点.通过改变式(16)、(17)中演化参数 δ 能够对网络拓扑中节点的度和边的权值进行调节,控制网络拓扑的结构和特性,因此利用参数 δ 值的变化可以构造出不同类型的社交网络拓扑结构.

$$P(k)\sim k^{-\gamma},\ \gamma=\frac{4\delta+3}{2\delta+1} \tag{16}$$

$$P(w)\sim w^{-\alpha},\ \alpha=2+\frac{1}{\delta} \tag{17}$$

本研究考虑到节点间边的权重对社交网络蠕虫传播的影响,应用 BBV 演化加权模型通过调节参数 δ 构造多种符合幂率分布拓扑特性的仿真环境.仿真环境构建步骤为:① 利用公开数据集获取真实社交网络拓扑参数(表 2);② 利用式(16)计算合适的 δ 参数取值;③ 生成加权网络拓扑.

4　社交网络蠕虫传播仿真

为了便于仿真实验的分析和说明,对将要用到的术语及特征量定义如下:

(1)异质社交网络.由不同的应用场景和网络用户群组成的网络,网络用户群和社交方式的不同决定了网络结构的异质性,实证分析表明异质性主要体现为度分布和边权分布的差异;

(2)感染用户总数.经过时间 t,网络中被感染用户的总数;

(3)t 时刻感染用户数.在时刻 t,不存在重复感染情况下,相对于 $t-1$ 时刻网络中新增被感染用户数量;

(4)蠕虫查杀率.随着蠕虫病毒的曝光和专杀工具的推出,蠕虫的查杀率随时间逐渐增大,为时间 t 的增函数:$\gamma(t)=\gamma_0\cdot t^{\eta}$,$\eta$ 体现安全厂商对蠕虫病毒的应急速度.

仿真实验中模型参数分别设为:节点规模 $N=30000$;$G\propto N(\mu_g,\sigma_g^2)$:$\mu_g=0.5$,$\sigma_g^2=0.1$;$HARM=20$;$I=5$;$\gamma_0=0.002$;$\alpha=0.2$;其它参数根据实验目的指定.

4.1　系统有效性验证

由于社交网络蠕虫的传播具有高隐蔽性等特

点，获取蠕虫传播的实证数据是非常困难的. CAIDA 利用 Nyxem 蠕虫成功感染后访问 DNS 服务器的特点，首次对社交网络蠕虫传播进行了实证研究，但是统计结果由于受到正常网络流量和其它恶意流量的影响并非绝对准确，只能从趋势上反映出社交网络蠕虫的传播特点；同时，CAIDA 根据该蠕虫的传播特点总结出网络用户的安全意识和社会的持续关注程度等用户相关因素是影响此类蠕虫传播的关键①，因此是否体现用户行为对社交网络蠕虫传播的影响可以作为评价仿真模型优劣的重要指标. 为了验证本仿真系统的有效性，本节将实验分析社交网络访问频率、用户安全意识和社会应急反应速度等若干用户相关因素对社交网络蠕虫传播的影响.

4.1.1 社交网络访问频率

在图 1 所示的传播状态转换过程中，用户收到蠕虫文件并不必然会使系统感染，系统感染的关键条件在于用户登录社交网络，接触并点击蠕虫文件. 因此用户访问社交网络的频率是影响蠕虫传播速度的关键因素之一. 为了考察仿真系统是否能够体现用户访问频率对蠕虫传播的影响，下面将分别比较 $avgT$(访问社交网络时间间隔的平均期望)的不同取值($avgT=36$，$avgT=24$，$avgT=12$，$avgT=2$)对蠕虫传播的影响，其中：$\delta=1.0$，$\varepsilon=0.05$，$\eta=0$，结如图 2 所示.

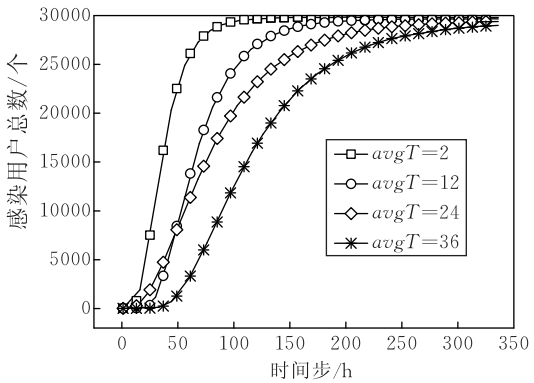


图 2 社交网络访问频率对蠕虫传播的影响

可以看出，较高的访问频率加速了社交网络蠕虫的扩散，当 $avgT=2$ 时，在仿真环境下不到 3 天的时间内能够蔓延整个网络，而较低的访问频率不利于蠕虫的扩散，这能够较为真实地体现社交网络蠕虫的传播特性，验证了模型的有效性.

随着 FACEBOOK 等社交网站各类应用业务的推出，用户访问社交网络的频率也逐渐提高，这类社交网络已经成为互联网用户的一种重要交流手段和联系方式，这为社交网络蠕虫提供了滋生的温床. 因此，在不断吸纳用户和扩展业务的同时，也对社交网站的安全性和健壮性提出了更高的要求.

4.1.2 用户安全意识

强化用户的网络安全意识是应对蠕虫爆发的有效手段. 安全意识的高低直接影响社交网络蠕虫成功欺骗用户并感染系统的概率. 因此，能否反映用户安全意识高低对蠕虫感染的影响是评价仿真模型有效性的重要指标之一. 仿真系统在用户行为博弈模型中，通过式(1)中的参数 ε 控制公众安全意识对蠕虫危害程度的负面影响. 仿真实验在固定参数 $avgT=12$，$\delta=1.0$，$\eta=0$ 的情况下，观察 ε 对仿真结果的影响，如图 3 所示.

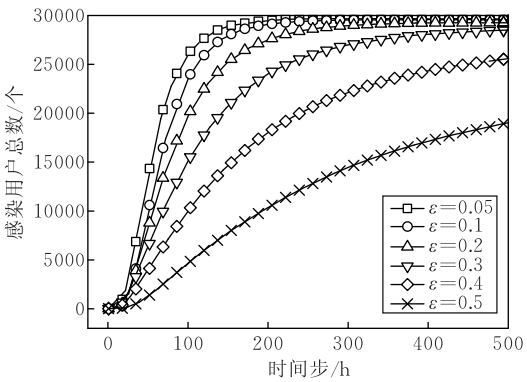


图 3 公众安全意识对蠕虫扩散速度的影响

从图 3 中可以看出 ε 值越高(即用户对蠕虫的警惕性越强)蠕虫扩散越缓慢，因此 ε 值的变化可以正确地体现该用户因素对蠕虫传播的影响.

4.1.3 应急反应速度

专杀工具可以抑制蠕虫的持续扩散，反病毒厂商针对蠕虫发布专杀工具的速度越快，蠕虫造成的影响越小. 蠕虫的查杀率是时间 t 的增函数，仿真系统通过调节参数 η 反映反病毒厂商的应急速度，η 的值越高表示应急能力越强. 图 4 显示了在固定参数 $avgT=12$，$\delta=1.0$，$\varepsilon=0.05$ 的情况下，模型参数

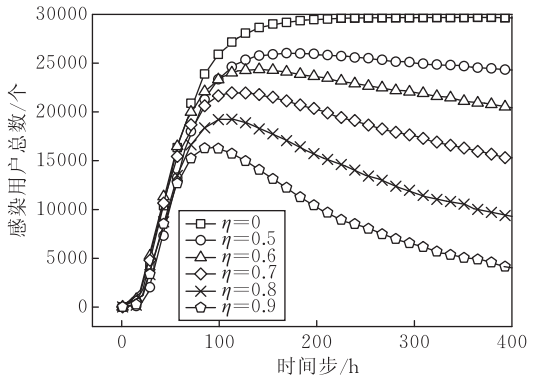


图 4 社会应急反应速度对蠕虫传播的抑制作用

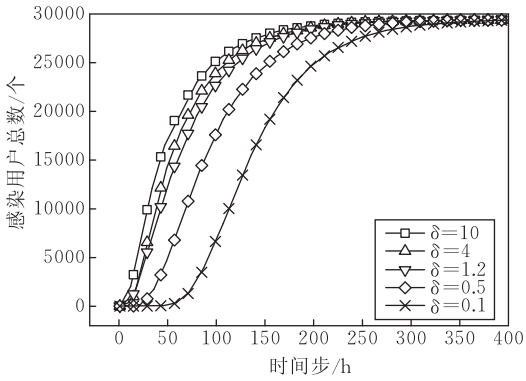
① http://www.caida.org/research/security/blackworm/

η的变化对仿真结果的影响.可以看出,η越高(应急速度越快)对蠕虫扩散的抑制效果越强,仿真模型能够反映出该因素对蠕虫的抑制作用.

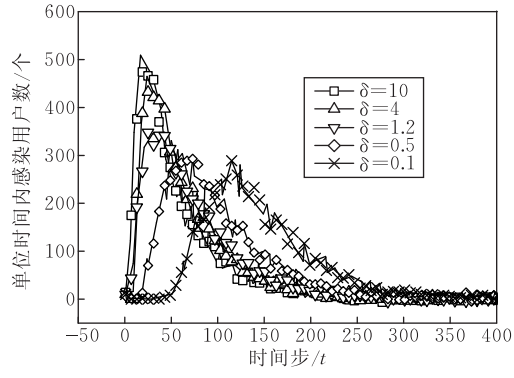
4.2　社交网络蠕虫传播分析

4.2.1　网络拓扑对传播的影响

网络拓扑对蠕虫传播的影响一直是学术界关心的重点问题之一,研究表明幂率特性的度分布对蠕虫传播有着极其深刻的影响.网络用户群和社交方式的差异形成了社交网络拓扑结构的异质性,本文通过调节 BBV 模型中参数 δ 值构造了多个异质社交网络仿真环境,研究网络拓扑对社交网络传播的影响.其中,相关参数的设定如下:$avgT=12$,$\varepsilon=0.05$,$\eta=0$.仿真实验分别在 5 个异质社交网络仿真环境下进行了模拟,结果如图 5 所示.



(a) 蠕虫扩散规模统计



(b)单位时间内感染主机数

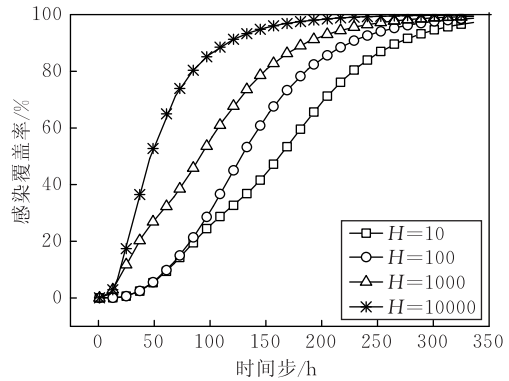
图 5　网络拓扑对蠕虫传播的影响

仿真拓扑环境可分为 2 类场景.场景 1:δ 的取值分别为 1.0、4 和 10 时,度分布和边权值幂率系数取值均在[2,3]之间;场景 2:δ 的取值为 0.1 和 0.5 时,节点度幂率系数落在[2,3]之间,但边权值幂率系数较大,使权值的分布区间较集中,权值分布较为"均匀"[27].通过对比分析两种场景中蠕虫的传播,更有利于揭示边权分布对蠕虫传播的影响.由图 5 可见,网络节点规模相同的情况下,在场景 1 中 δ 值低(幂率系数大),蠕虫的扩散速度相对较缓;在场景 2 中,权值分布趋于均匀,即便存在着大量的 Hub 节点的情况下,由于社交网络蠕虫收到用户信任值较低的影响,使得传播相对缓和.对比分析两种场景下社交网络蠕虫的传播,揭示了边的权值分布对社交网络蠕虫传播的影响,其无标度分布"均衡程度[25]"越低更有利于社交网络蠕虫的传播.

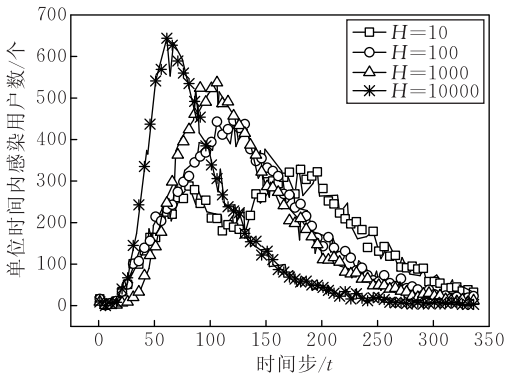
4.2.2　混杂的网络环境对传播的影响

互联网中存在着各种形态的虚拟社交网络,这使用户的身份也变得更为复杂[26],例如一个人在使用电子邮件的同时进行着 P2P 文件共享操作,或者同时拥有多个社交网站的账号进行交友活动等.虚拟社交网络的快速发展丰富人们的生活的同时,也为社交网络蠕虫的生存和传播提供了有利的环境.

为了揭示用户身份多重性对社交网络蠕虫传播的影响,首先构建了由两个相同特性($\delta=1.0$)的社交网络组成的总节点规模为 $N\times 2=60000$ 的仿真环境,其中"社交网络混杂程度(两个网络中具有'多重身份'的用户节点的个数,用 H 表示)"取值分别为 $H=10/100/1000/10000$,其中,相交节点的选取是完全随机的.利用仿真系统在其中一个网络中随机选择初始感染节点,并观察 H 的不同取值对蠕虫在整个网络中传播的影响,实验结果如图 6 所示(相



(a) 蠕虫扩散规模统计



(b) 单位时间内感染主机数

图 6　混杂的网络环境对蠕虫传播的影响

关参数设定为 $\varepsilon=0.05$，$\eta=0$，$avgT=24$).

仿真结果表明，由于“多重身份”用户的存在，社交网络蠕虫在某一个社交网络中爆发后，能够迅速地扩散到相似的社交网络中去，从而实现在整个互联网的蔓延.“多重身份”用户的数量能够明显地影响蠕虫扩散速度，特别是在高度混杂的网络环境下($H=10000$)蠕虫的快速扩散更为显著.

5 结 论

互联网中基于社会工程学方式进行传播的社交网络蠕虫已经成为威胁网络安全的重大隐患. 本文在总结了目前网络蠕虫仿真模型研究进展的基础上，从社会工程学的角度考虑网络用户行为对社交网络蠕虫传播的影响. 首先引入博弈理论对单个节点的网络用户行为进行预测，得到社交网络蠕虫在该节点的可能感染率；其次通过对网络用户访问社交网络的行为习惯进行建模，用于确定单位时间内社交网络中蠕虫潜在的感染范围，进而实现了一个适用于社交网络蠕虫传播研究的仿真系统. 本文主要贡献如下：(1)提出了网络用户与蠕虫文件之间的行为博弈模型，从社会工程学的角度刻画了社交网络蠕虫的感染过程，便于分析网络用户行为因素对蠕虫传播的影响；(2)通过对网络用户访问社交网络的习惯进行建模，提出了基于网络用户活动习惯的离散访问模型，反映了网络用户活动的动态性；(3)利用 BBV 拓扑生成算法构建了符合真实社交网络拓扑特性仿真环境，避免了在经济和法律的制约下大规模传播实验瓶颈，为模型实现提供了可用的仿真平台；(4)使用该系统，实验分析了社交网络蠕虫在社交网络中的传播，揭示了网络拓扑结构和社交网络用户混杂程度对蠕虫传播的影响. 本研究为社交网络蠕虫传播行为的研究提供了一个有效的仿真框架，为相关的检测防御机制研究提供了重要的理论依据.

参 考 文 献

[1] Fosnock C. Computer Worms: Past, Present, and Future. East Carolina University, NC, USA, 2005

[2] Wang Y, Wang C. Modeling the effects of timing parameters on virus propagation//Proceedings of the ACM Workshop on Rapid Malcode. New York, NY, USA, 2003: 61-66

[3] Pastor-Satorras R, Vespignani A. Epidemic dynamics in finite size scale-free networks. Physical Review E, 2002, 65(3): 035108

[4] Zou C C, Gong W, Towsley D. Code red worm propagation modeling and analysis//Proceedings of the ACM Conference on Computer and Communications Security. Washington, DC, USA, 2002: 138-147

[5] Yang S, Jin H, Liao X, et al. Modeling modern social-network-based epidemics: A case study of rose . Autonomic and Trusted Computing, 2008, 5060: 302-315

[6] Xia Chun-He, Shi Yun-Ping, Li Xiao-Jian. Research on epidemic models of P2P worm in structured peer-to-peer networks. Chinese Journal of Computers, 2006, 29(6): 952-959(in Chinese)
(夏春和，石昀平，李肖坚. 结构化对等网中的 P2P 蠕虫传播模型研究. 计算机学报，2006，29(6)：952-959)

[7] Qing Si-Han, Wang Chao, He Jian-Bo, Li Da-Zhi. Research and development of instant messaging worms. Journal of Software, 2006, 17(10): 2118-2130(in Chinese)
(卿斯汉，王超，何建波，李大治. 即时通信蠕虫研究与发展. 软件学报，2006，17(10)：2118-2130)

[8] Wang Yue-Wu, Jing Ji-Wu, Xiang Ji, Liu Qi. Topology aware worm simulation and analysis. Journal of Software, 2008, 19(6): 1508-1518(in Chinese)
(王跃武，荆继武，向继，刘琦. 拓扑相关蠕虫仿真分析. 软件学报，2008，19(6)：1508-1518)

[9] Chen G, Gray R S. Simulating non-scanning worms on peer-to-peer networks//Proceedings of the International Conference on Scalable Information Systems. Hong Kong, China, 2006: 1-13

[10] Zou C C, Towsley D, Gong W. Modeling and simulation study of the propagation and defense of Internet E-mail worms. IEEE Transactions on Dependable and Secure Computing, 2007, 4(2): 105-118

[11] Sun X, Liu Y, Wang J. Modeling email worm propagation using game theory//Proceedings of the International Conference on Multimedia Information Networking and Security. Wuhan, China, 2009: 388-392

[12] Barabasi A L. The origin of bursts and heavy tails in human dynamics. Nature, 2005, 435(7039): 207-211

[13] Dezso Z, Almaas E, Lukacs A. Dynamics of information access on the web. Physical Review E, 2006, 73(6): 066132

[14] Kumar R, Novak J, Tomkins A. Structure and evolution of online social networks//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, NY, USA, 2006: 611-617

[15] Stahl S. A Gentle Introduction to Game Theory. Washington, DC, USA: American Mathematical Society, 1999

[16] Sallhammar K, Knapskog S J, Helvik B E. Using stochastic game theory to compute the expected behavior of attackers//Proceedings of the Symposium on Applications and the Internet Workshops. Trento, Italy, 2005: 102-105

[17] Owen G. Game Theory. 3rd Edition. New York, USA: Academic Press, 1995

[18] Newman M. The structure and function of complex networks. SIAM Review, 2003, 45(2): 167-256

[19] Ebel H, Mielsch L I, Bornholdt S. Scale-free topology of E-mail networks. Physical Review E, 2002, 66(3): 035103

[20] Saroiu S, Gummadi P K, Gribble SD. A measurement study of peer-to-peer file sharing systems//Martin G K, Prashant J S. Proceedings of the Multimedia Computing and Networking. SPIE, New York, USA, 2001: 156-170

[21] Ripeanu M, Foster I, Iamnitchi A. Mapping the Gnutella network. IEEE Internet Computing, 2002, 6(1): 50-57

[22] Ahn Y Y, Han S, Kwak H, Moon S, Jeong H. Analysis of topological characteristics of huge online social networking services//Proceedings of the 16th International Conference on World Wide Web. New York, NY, USA, 2007: 835-844

[23] Wang F W, Zhang Y K, Ma J F. Defending passive worms in unstructured P2P networks based on healthy file dissemination. Computers and Security, 2009, 28(7): 628-636

[24] Barrat A, Barthelemy M, Vespignani A. Weighted evolving networks: Coupling topology and weight dynamics. Physical Review Letters, 2004, 92(22): 28701

[25] Xiao B, Liu L D, Guo X C, Xu K. Modeling the IPv6 Internet AS-level topology. Physica A: Statistical Mechanics and its Applications, 2009, 388(4): 529-540

[26] Xu W, Zhang F, Zhu S. Toward worm detection in online social networks//Proceedings of the 26th Annual Computer Security Applications Conference. Austin, Texas, 2010: 11-20

SUN Xin, born in 1984, Ph. D. candidate. His major research interests include network security, complex network.

LIU Yan-Heng, born in 1958, professor, Ph. D. supervisor. His major research interests include network security, mobile IP, network management and wireless sensor networks.

ZHU Jian-Qi, born in 1976, Ph. D., lecturer. His major research interests is computer security.

LI Fei-Peng, born in 1986, M. S. candidate. His major research interests is computer security.

Background

Social network services provide means for users to interact over the internet, such as E-mail, instant messaging and social networking sites. Along with the rapid development of social networks, social network worms have constituted one of the major internet security problems. Unlike random scanning active worms, social network worms spread by sending messages that contained the worm via the social networks and entice users to click on the malicious messages to infect the computer system. Accordingly, the worm spreading greatly depends on human behaviors. There is few researches considering the impact of human behaviours on worm spreads and no model till date that claims to construct models based on human behaviours. In this work, the social network worm propagation is discussed from the viewpoint of social engineering, and a human behavior model based on game theory is presented for predicting the expected actions of network user encountered with worm messages. By analyzing network users' diurnal activity behaviors, a discrete social network accessing model is proposed to characterize the general human habit of accessing certain social network. Finally, a simulation system is constructed with the aforementioned two models. This work is supported by the National Natural Science Foundation of China (No. 60973136, 61073164), National Development and Reform Commission(CNGI-09-01-11), China-BC ICSD (No. 2008DFA12140), Science Foundation for Youth of Jilin Province (201101033) and Science Foundation for Young Teachers of Jilin University (450060445169).