P2P 网络中激发型蠕虫传播动态建模

冯朝胜[1,2,3],袁 丁[1],卿 昱[2],秦志光[3]
(1.四川师范大学计算科学学院,可视化计算与虚拟现实四川省重点实验室,四川成都 610101;
2.中国电子科技集团公司第 30 研究所,四川成都 610041;
3.电子科技大学计算机科学与工程学院,四川成都 610054)

摘 要: 鉴于激发型蠕虫的巨大危害性,本文在考虑网络动态变化的情况下对激发型蠕虫的传播进行了深入地研究,提出了激发型蠕虫动态传播数学模型和免疫模型,并基于动态传播数学模型推导出了激发型蠕虫不会泛滥的充分条件.大规模仿真实验验证了传播模型的有效性和蠕虫不会泛滥充分条件的正确性.基于传播模型的分析表明,下载率是影响蠕虫传播的关键因素,蠕虫基本繁殖率是衡量蠕虫传播能力的关键指标.基于实测 P2P 网络数据和传播模型,预测和估计了激发型蠕虫的传播能力、传播速度和危害性,指出尽早重视 P2P 激发型蠕虫特别是尽早找到检测和控制方法的重要性和迫切性.

关键词: P2P 网络; 激发型蠕虫; 动态性; 传播建模; 基本繁殖率; 仿真
中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2012) 02-0300-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2012.02.015

Dynamic Modeling of Reactive Worm Propagation in P2P Networks

FENG Chao-sheng[1,2,3], YUAN Ding[1], QING Yu[2], QIN Zhi-guang[3]
(1. *Visual Computing & Virtual Reality Key Laboratory of Sichuan Province, School of Computer Science, Sichuan Normal University, Chengdu Sichuan* 610101, *China*; 2. *The No.30 institute of China Electronic Technology Corporation, Chengdu Sichuan* 610041, *China*; 3. *School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan* 610054, *China*)

Abstract: In this paper, propagation pattern of reactive worms is studied. The propagation model and the immunization model of reactive worms are proposed, in which dynamic factors of P2P networks are taken into account. Further, the sufficient condition of worms not attaining an endemic state is deduced from the model of propagation of reactive worms in applying Epidemiology. Large scale simulation experiments validate the models and the sufficient condition. All the simulations also show that among all P2P-related parameters, the downloading rate is the crucial factor to the propagation of reactive worms, the basic reproduction number of worms is the index of propagation capability of worms, and the sufficient condition is helpful to early warn the presence of an epidemic. In addition, by using data from the Gnutella network, the propagation capability, propagation speed and the risk of reactive worms are predicted and evaluated based on the propagation model. Prediction and evaluation show that it is time that reactive worms should be taken into account.

Key words: P2P networks; reactive worms; dynamics; propagation modeling; basic reproduction number; simulation

1 引言

P2P 网络是同质网络(网络中运行的 P2P 用户软件基本相同),如果 P2P 用户软件有漏洞,那整个 P2P 网络就存在漏洞,这意味着一旦恶意用户发现了某个漏洞并基于该漏洞编制出 P2P 蠕虫,那整个 P2P 网络将很快被攻破而瘫痪[1].P2P 激发型蠕虫就是这样的 P2P 蠕虫.P2P 激发型蠕虫还有有一个重要的特点,那就是攻击的被动性.虽然激发型蠕虫像 P2P 主动型蠕虫一样都是利用漏洞进行传播,但在攻击时却不会像 P2P 主动型蠕虫那样主动发起攻击,而是在正常连接(上传或下载)的激发下才会攻击.正常连接掩盖下进行攻击这一特点使得激发型蠕虫像 P2P 被动型蠕虫一样很难被检测和发现,这使得它具有更大的威胁性和危害性.

收稿日期:2011-03-18;修回日期:2011-09-14
基金项目:国家自然科学基金项目(No.60873075);可视化计算与虚拟现实四川省重点实验室课题(No.J2010N05);四川省科技厅应用基础项目(No.2010JY0125);四川省教育厅重点课题(No.10ZA007);西南交通大学信息编码与传输四川省重点实验室开放研究基金课题(No.2010-05)

尽管 P2P 激发型蠕虫有较强的感染能力和较快的传播速度,然而它并没有受到足够重视.鉴于此,在对 P2P 网络和激发型蠕虫进行深入分析基础上,基于数学流行病学提出了激发型蠕虫传播数学模型和免疫模型.

2 相关研究工作

2.1 相关研究工作及进展

Contagion 蠕虫,这一蠕虫概念是 Staniford 等人[2]在深入分析网络和蠕虫的特点基础上于 2002 年提出的,他们指出 P2P 网络的同质特点决定了它更适合激发型蠕虫的传播.2006 年,Chen 等人[3]在对 P2P 网络分析的基础上,指出 P2P 网络上的蠕虫应分为三类:被动型蠕虫、激发型蠕虫和主动型蠕虫.激发性蠕虫实际上属于 Contagion 蠕虫.

1988 年,Murray[4]率先利用数学流行病学对计算机病毒展开研究.1991 年,Kephart 和 White 将经典流行病模型引入计算机蠕虫传播建模中[5].Zou 等人和 Liljenstam 等人利用数学流行病学于 2002 年和 2004 年分别为互联网上的"红色代码"蠕虫[6]和电子邮件病毒[7]建立了传播模型,较准确地预测了这些病毒的传播趋势和行为.2004 年,Wei[8]提出了利用传统计算机病毒传播模型来研究利用 P2P 网络进行传播的蠕虫的传播方法,之后又对各种扫描策略下蠕虫的传播性能进行了建模和仿真分析[9].2005 年,Dumitriu 等人[10]对污染文件在 P2P 网络上的传播进行了建模.Thommes 等人[11]对 P2P 文件共享网上的病毒传播和感染文件传播分别建立了传播模型.2006 年,Chen 等人[3]指出 P2P 蠕虫为非扫描型蠕虫,并对三类 P2P 蠕虫分别进行了仿真分析;然而,他们并没有给出 P2P 蠕虫传播的数学模型.夏春和等人[12]基于结构化对等网路由表构造方法,建立了 P2P 蠕虫在 Chord、CAN、Pastry 三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影响.同年,Ma 等[13]利用数字模拟的方法分析了 P2P 系统参数对 P2P 被动型蠕虫传播的影响,并构建了蠕虫传播模型.2008 年,Wang 等提出了被动型蠕虫在 Gnutella 网络上的传播模型[14],王跃武等人[15]对 Contagion 蠕虫的传播情况进行了仿真分析.2009 年,应凌云、冯登国等对 P2P 僵尸网络及其防御方法进行了研究,提出了种基于层次化 P2P 网络技术的新型僵尸网络结构[16].

2.2 P2P 文件共享网

在像 Gnutella[17]和 eDonkey2000[18]这样的 P2P 网络中,每个用户都有一个共享文件夹,用户将所有可共享的文件都放到共享文件夹以便其他用户共享,网络中的任何用户都可以从其他任意一个用户的共享文件夹中下载文件.当用户想要下载某个文件时,他会发出搜索文件请求.在 eDonkey2000 中,通过查询服务器来处理这个请求;而在 Gnutella 中,通过邻居不断转发的形式来搜索文件.无论哪种 P2P 文件共享网络,请求文件用户最终都会收到与请求相匹配的文件列表.尽管不同的网络生成文件列表的方式有所不同,但生成的文件列表都是满足用户文件请求的所有 P2P 主机的一部分.获取了文件列表后,用户可以从中选择一个或多个主机来下载该文件,从多个主机上下载文件,被称作多点下载,意味着每个主机都提供文件的一部分.文件下载后被放在共享文件夹,可被网络中其它主机下载.

2.3 P2P 激发型蠕虫

P2P 激发型蠕虫是一种需要"刺激"才会"发起攻击"的蠕虫,这种刺激就是感染主机和未感染主机之间为传送文件而建立起的正常连接,这种连接可以是感染主机为下载文件而向未感染主机发起建立的,也可以是未感染主机为下载文件而向感染主机发起建立的.激发型蠕虫的攻击过程如下.感染主机上运行的蠕虫会监听该主机上的连接.一旦监听到连接被建立,蠕虫利用正常的流量和漏洞将蠕虫副本发送到另一端,并启动副本,于是另一端的主机成为新的感染源.由于该蠕虫利用的是正常连接进行传播,所以蠕虫检测软件难以发现并阻止其传播.激发型蠕虫感染根据感染方向的不同可以分成三类:源感染、目标感染和混合感染.源感染是指仅对发起连接进行下载的客户端进行感染(目标主机为感染主机).目标感染是指仅感染连接的目标主机(源主机为感染主机).混合感染是指发起连接的主机(下载方)和连接目标主机(上传方)都可能被感染.这里,源感染激发型蠕虫和被动型蠕虫很相像,但实际上是不同的,一台易感主机从感染了被动型蠕虫的主机上下载干净文件并不会被感染,而从感染了源感染(或混合)激发型蠕虫的主机上下载任何文件都很可能被感染[3].

3 P2P 激发型蠕虫传播和免疫建模

3.1 建模参数和假设

蠕虫传播建模基于数学流行病学[19].在模型中,主机的状态根据感染情况分成三种:易感的、感染的和免疫的.

(1)易感的:该类主机既没有被蠕虫感染也没有被免疫,所以在下载和上传时都可能被感染.

(2)感染的:该类主机已经感染了蠕虫,当其上传文件或下载文件时都可能致使另一方(下载主机或上传主机)被感染.

(3)免疫的:指安装了专门针对某类蠕虫的补丁的主机.易感主机因为安装了补丁而不会再被该类蠕虫感染,感染主机因为安装了补丁,主机上的蠕虫被清除

并且再也不会被该类蠕虫感染.

为了在建模时考虑网络的动态特征,将网络的主机分为两种:在线主机和离线主机.在线主机是指正运行着 P2P 用户软件的主机;离线主机是指安装了 P2P 用户软件,但没有启动该软件的主机.

为了简化建模,作如下假设:

(1)P2P 网络的规模不变即在线主机和离线主机的数量之和没有发生变化.

(2)主机状态转移都在一个单位时间内完成.这意味着一台主机因为下载而被感染(包括搜寻文件、连接、下载和感染)所用的时间也是一个单位时间.

(3)蠕虫的感染模式为混合式.

为了便于下面的蠕虫建模分析,将建模时要用到的参数和变量列举在表 1 中.

表 1 模型中用到的变量和参数

符号	说明	初值
N	网络中主机总数	10000
$S_{on}(t)$	t 时刻在线易感主机数	8900
$S_{off}(t)$	t 时刻离线主机数	1000
$I_{on}(t)$	在线感染主机数	100
$I_{off}(t)$	离线感染主机数	0
$R_{on}(t)$	在线免疫主机数	0
$R_{off}(t)$	离线免疫主机数	0
λ_d	下载率	0.02
λ_r	恢复率	0.05
λ_{on}	上线率(离线时间的倒数)	0.09
λ_{off}	下线率(在线时间的倒数)	0.01
λ_{sm}	易感主机免疫率	0
λ_{im}	感染主机免疫率	0
p_d	易感主机从感染主机上下载文件被感染的概率	0.6
p_u	感染主机从易感主机上下载时将其感染的概率	0.5

3.2 传播模型

在不考虑免疫的情况下,考查蠕虫的传播情况.在这种情况下,主机要么处于易感状态(S),要么处于感染状态(I).状态转移过程为 S→I→S.显然,两种情况会导致易感主机转变为感染主机.一种是易感主机从感染主机上下载文件,另一种是感染主机从易感主机上下载文件,或者说易感主机向感染主机上传文件.在用户发现主机中毒以后,通过使用升级反病毒软件或运用反病毒知识或者重装系统将蠕虫清除,感染主机就会恢复到易感状态.

(1)在线易感主机变化率

由于激发型蠕虫会被正常连接(即上传或下载文件)所激发,故易感主机在上传或下载时都可能被感染.显然,只有在线易感主机才可能被感染(因为只有上线的主机才能进行上传或下载).

在时刻 t,当易感染主机请求文件时,选中感染主机作为文件源的概率为$\frac{I_{on}(t)}{S_{on}(t)+I_{on}(t)}$,而从感染主机上下载文件被感染的概率为 p_d,所以一台易感主机因为下载而被感染的概率为$\frac{I_{on}(t)p_d}{S_{on}(t)+I_{on}(t)}$.在一个单位时间,$S_{on}(t)$台主机共执行下载 $\lambda_d S_{on}(t)$次,故在一个单位时间里共有$\frac{\lambda_d S_{on}(t)I_{on}(t)p_d}{S_{on}(t)+I_{on}(t)}$台主机因为下载而被感染.

当感染主机请求文件时,任意一台主机被选中作为上传主机的概率为$\frac{1}{S_{on}(t)+I_{on}(t)}$,相应地,不被选中的概率为$(1-\frac{1}{S_{on}(t)+I_{on}(t)})$.在时刻 t,$I_{on}(t)$台主机共执行下载任务 $\lambda_d I_{on}(t)$次.显然,一台主机一次都不被选中作为上传文件主机的概率为$\left(1-\frac{1}{S_{on}(t)+I_{on}(t)}\right)^{\lambda_d I_{on}(t)}$,那么一台主机被选中的概率就为$[1-(1-\frac{1}{S_{on}(t)+I_{on}(t)})^{\lambda_d I_{on}(t)}]$,所以,一台易感主机因给感染主机上传文件而被感染的概率为$p_u\left(1-(1-\frac{1}{S_{on}(t)+I_{on}(t)})^{\lambda_d I_{on}(t)}\right)$,故在一个单位时间里被感染的在线易感主机数为$S_{on}(t)p_u\left(1-(1-\frac{1}{S_{on}(t)+I_{on}(t)})^{\lambda_d I_{on}(t)}\right)$.

当然,部分在线的易感主机还会因为下线而变成离线易感主机,转化的主机数为$\lambda_{off}S_{on}(t)$;而部分离线易感主机会上线变成在线主机,转化的主机数为$\lambda_{on}S_{off}(t)$;另外还有部分在线感染主机因为蠕虫被清除而恢复成易感状态,恢复主机数为 $\lambda_r I_{on}(t)$.

综合以上分析,在线易感主机的变化率为:

$$\begin{aligned}\frac{\mathrm{d}S_{on}(t)}{\mathrm{d}t} = &-\frac{\lambda_d p_d S_{on}(t)I_{on}(t)}{S_{on}(t)+I_{on}(t)}\\ &-S_{on}(t)p_u\left(1-(1-\frac{1}{S_{on}(t)+I_{on}(t)})^{\lambda_d I_{on}(t)}\right)\\ &+\lambda_r I_{on}(t)+\lambda_{on}S_{off}(t)-\lambda_{off}S_{on}(t)\end{aligned}\tag{1}$$

(2)在线感染主机变化率

基于在线易感主机变化率类似的分析,容易得到在线感染主机的变化率为:

$$\begin{aligned}\frac{\mathrm{d}I_{on}(t)}{\mathrm{d}t} = &\frac{\lambda_d p_d S_{on}(t)I_{on}(t)}{S_{on}(t)+I_{on}(t)}\\ &+S_{on}(t)p_u\left(1-(1-\frac{1}{S_{on}(t)+I_{on}(t)})^{\lambda_d I_{on}(t)}\right)\\ &-\lambda_r I_{on}(t)\quad+\lambda_{on}I_{off}(t)-\lambda_{off}I_{on}(t)\end{aligned}\tag{2}$$

(3)离线易感主机变化率

在时刻 t,有 $\lambda_{off}S_{on}(t)$台在线易感主机因为下线转变成了离线主机,并有 $\lambda_r I_{off}(t)$台离线感染主机因为蠕虫被清除而恢复为易感主机;与此同时,有 $\lambda_{on}S_{off}(t)$离

线主机转变成了在线主机.于是,离线易感主机变化率为:

$$\frac{\mathrm{d}S_{off}(t)}{\mathrm{d}t}=\lambda_r I_{off}(t)-\lambda_{on}S_{off}(t)+\lambda_{off}S_{on}(t) \tag{3}$$

(4)离线感染主机变化率

类似地,离线感染主机的变化率为:

$$\frac{\mathrm{d}I_{off}(t)}{\mathrm{d}t}=-\lambda_r I_{off}(t)-\lambda_{on}I_{off}(t)+\lambda_{off}I_{on}(t) \tag{4}$$

方程(1)~(4)组成的方程组就是激发型蠕虫的传播数学模型.

3.3 免疫模型

假设在每个单位时间,易感主机和感染主机的补丁的安装率分别为 λ_{sm}和λ_{im},那么,在线易感主机、在线感染主机、离线易感主机和离线感染主机的免疫数分别为 $\lambda_{sm}S_{on}(t)$、$\lambda_{im}I_{on}(t)$,$\lambda_{sm}S_{off}(t)$和 $\lambda_{im}I_{off}(t)$.结合传播模型的分析,激发型蠕虫的免疫模型为:

$$\begin{aligned}\frac{\mathrm{d}S_{on}(t)}{\mathrm{d}t}=&-\frac{\lambda_d p_d S_{on}(t)I_{on}(t)}{S_{on}(t)+I_{on}(t)}\\&-S_{on}(t)p_u\left(1-\left(1-\frac{1}{S_{on}(t)+I_{on}(t)}\right)^{\lambda_d I_{on}(t)}\right)\\&+\lambda_r I_{on}(t)+\lambda_{on}S_{off}(t)-\lambda_{sm}S_{on}(t)-\lambda_{off}S_{on}(t)\end{aligned} \tag{5}$$

$$\begin{aligned}\frac{\mathrm{d}I_{on}(t)}{\mathrm{d}t}=&\frac{\lambda_d p_d S_{on}(t)I_{on}(t)}{S_{on}(t)+I_{on}(t)}\\&+S_{on}(t)p_u\left(1-\left(1-\frac{1}{S_{on}(t)+I_{on}(t)}\right)^{\lambda_d I_{on}(t)}\right)\\&-\lambda_r I_{on}(t)+\lambda_{on}I_{off}(t)-\lambda_{im}I_{on}(t)-\lambda_{off}I_{on}(t)\end{aligned} \tag{6}$$

$$\frac{\mathrm{d}R_{on}(t)}{\mathrm{d}t}=\lambda_{sm}S_{on}(t)+\lambda_{im}I_{on}(t)+\lambda_{on}R_{off}(t)-\lambda_{off}R_{on}(t) \tag{7}$$

$$\frac{\mathrm{d}S_{off}(t)}{\mathrm{d}t}=\lambda_r I_{off}(t)-\lambda_{on}S_{off}(t)+\lambda_{off}S_{on}(t)-\lambda_{sm}S_{off}(t) \tag{8}$$

$$\frac{\mathrm{d}I_{off}(t)}{\mathrm{d}t}=-\lambda_r I_{off}(t)-\lambda_{on}I_{off}(t)+\lambda_{off}I_{on}(t)-\lambda_{im}I_{off}(t) \tag{9}$$

$$\frac{\mathrm{d}R_{off}(t)}{\mathrm{d}t}=\lambda_{sm}S_{off}(t)+\lambda_{im}I_{off}(t)-\lambda_{on}R_{off}(t)+\lambda_{off}R_{on}(t) \tag{10}$$

这里,

$$N=S_{on}(t)+S_{off}(t)+I_{on}(t)+I_{off}(t)+R_{on}(t)+R_{off}(t)$$

4 蠕虫不会泛滥的充分条件

建立蠕虫传播的数学模型的主要目的是基于它来预测蠕虫的传播趋势并分析影响蠕虫传播的关键因素.在补丁编制出来之前,哪些才是决定蠕虫传播程度的关键因素呢?蠕虫不会泛滥的条件是什么呢?下面基于重要的流行病学理论给出分析.

4.1 重要的流行病理论

根据文献[20],病毒是否能够在网络中流行是由病毒的基本繁殖率 R_0 来决定的.当 $R_0<1$ 时,病毒很快就会在网络中消失,网络处于无病毒的平衡状态.因此,只要求出网络处于无病毒平衡状态的充分条件并通过在蠕虫出现时保证网络满足该条件就能保证蠕虫不会在网络中泛滥或流行(即使有新的主机被感染).文献[21,22]提出了一种求基本繁殖率的方法.在该方法中,将个体状态转移流分成新感染个体进入流和其他流两种,分别用向量 $\boldsymbol{f}$ 和 $\boldsymbol{v}$ 表示.分别求这两个向量对各个状态变量的微分,微分后的向量组成了如下矩阵:

$$\boldsymbol{F}=\left[\frac{\partial f_i}{\partial x_j}(x_0)\right],\quad \boldsymbol{V}=\left[\frac{\partial v_i}{\partial x_j}(x_0)\right],\ 1\leqslant i,j\leqslant m$$

f_i 和 v_i 是 $\boldsymbol{f}$ 和 $\boldsymbol{v}$ 的第 i 个分量,x_i 是第 i 个状态变量并且 $\dot{x}_i=f_i(x)-v_i(x)$,m 表示感染状态变量个数.基本繁殖率 R_0 的值就是矩阵 $\boldsymbol{FV}^{-1}$的最大绝对特征值.

4.2 蠕虫不会泛滥的条件

定理　根据所提出激发型蠕虫传播模型和上节的流行病学重要理论,激发型蠕虫不会泛滥的充分条件是:

$$\frac{\lambda_d(\lambda_r+\lambda_{on})(p_d+p_u)}{\lambda_r^2+\lambda_r(\lambda_{on}+\lambda_{off})}<1$$

证明　根据蠕虫传播模型,知感染状态变量分别 I_{on}和 I_{off},由方程(2)和(4)得:

$$\boldsymbol{f}=\begin{bmatrix}\dfrac{\lambda_d p_d S_{on}(t)I_{on}(t)}{S_{on}(t)+I_{on}(t)}+S_{on}(t)p_u\left(1-\left(1-\dfrac{1}{S_{on}(t)+I_{on}(t)}\right)^{\lambda_d I_{on}(t)}\right)\\0\end{bmatrix}$$

$$\boldsymbol{v}=\begin{bmatrix}\lambda_r I_{on}(t)-\lambda_{on}I_{off}(t)+\lambda_{off}I_{on}(t)\\\lambda_r I_{off}(t)+\lambda_{on}I_{off}(t)-\lambda_{off}I_{on}(t)\end{bmatrix}$$

在网络处于无蠕虫平衡状态时,有

$$\frac{\mathrm{d}S_{on}(t)}{\mathrm{d}t}=\frac{\mathrm{d}I_{on}(t)}{\mathrm{d}t}=\frac{\mathrm{d}S_{off}(t)}{\mathrm{d}t}=\frac{\mathrm{d}I_{off}(t)}{\mathrm{d}t}=0$$

且　$I_{on}=I_{off}=0$

由方程(3)和(4)易得此时在线易感主机和离线易感主机的数量分别为

$$\tilde{S}_{on}=\frac{\lambda_{on}N}{\lambda_{on}+\lambda_{off}},\quad \tilde{S}_{off}=\frac{\lambda_{off}N}{\lambda_{on}+\lambda_{off}}$$

$$\frac{\mathrm{d}\left(\dfrac{\lambda_d p_d S_{on}(t)I_{on}(t)}{S_{on}(t)+I_{on}(t)}\right)}{\mathrm{d}(I_{on}(t))}=\frac{\lambda_d p_d S_{on}(t)(S_{on}(t)+I_{on}(t)-I_{on}(t))}{(S_{on}(t)+I_{on}(t))^2}$$

$$=\frac{\lambda_d p_d S_{on}^2(t)}{(S_{on}(t)+I_{on}(t))^2}$$

$$\because\left(1-\frac{1}{S_{on}(t)+I_{on}(t)}\right)^{\lambda_d I_{on}(t)}=1-\frac{\lambda_d I_{on}(t)}{S_{on}(t)+I_{on}(t)}+\cdots$$

$$\approx 1-\frac{\lambda_d I_{on}(t)}{S_{on}(t)+I_{on}(t)}$$

$$\therefore 1-(1-\frac{1}{S_{on}(t)+I_{on}(t)})^{\lambda_d I_{on}(t)} \approx \frac{\lambda_d I_{on}(t)}{S_{on}(t)+I_{on}(t)}$$

$$\frac{d\left(1-(1-\frac{1}{S_{on}(t)+I_{on}(t)})^{\lambda_d I_{on}(t)}\right)}{d(I_{on}(t))}$$

$$\approx -\frac{d\left(\frac{\lambda_d I_{on}(t)}{S_{on}(t)+I_{on}(t)}\right)}{d(I_{on}(t))}$$

$$=\frac{\lambda_d(S_{on}(t)+I_{on}(t))-\lambda_d I_{on}(t)}{(S_{on}(t)+I_{on}(t))^2}$$

$$=\frac{\lambda_d S_{on}(t)}{(S_{on}(t)+I_{on}(t))^2}$$

向量 $\boldsymbol{f}$ 和 $\boldsymbol{v}$ 分别对 I_{on} 和 I_{off} 求微分，得

$$\boldsymbol{F}=\begin{bmatrix}\frac{\lambda_d p_d S_{on}^2(t)}{(S_{on}(t)+I_{on}(t))^2}+\frac{\lambda_d p_u S_{on}^2(t)}{(S_{on}(t)+I_{on}(t))^2} & 0\\ 0 & 0\end{bmatrix}$$

$$\boldsymbol{V}=\begin{bmatrix}\lambda_r+\lambda_{off} & -\lambda_{on}\\ -\lambda_{off} & \lambda_r+\lambda_{on}\end{bmatrix}$$

在平衡状态 $\{\tilde{S}_{on},0,\tilde{S}_{off},0\}$ 时有

$$\boldsymbol{F}=\begin{bmatrix}\lambda_d(p_d+p_u) & 0\\ 0 & 0\end{bmatrix}$$

$$\because V^{-1}=\frac{1}{\lambda_r^2+\lambda_r(\lambda_{on}+\lambda_{off})}\begin{bmatrix}\lambda_r+\lambda_{on} & +\lambda_{on}\\ \lambda_{off} & \lambda_r+\lambda_{off}\end{bmatrix}$$

$$\therefore R_0=\rho(FV^{-1})=\frac{\lambda_d(\lambda_r+\lambda_{on})(p_d+p_u)}{\lambda_r^2+\lambda_r(\lambda_{on}+\lambda_{off})} \quad (11)$$

根据文献[21]，蠕虫不会泛滥的充分条件为：

$$\frac{\lambda_d(\lambda_r+\lambda_{on})(p_d+p_u)}{\lambda_r^2+\lambda_r(\lambda_{on}+\lambda_{off})}<1$$

5 仿真验证及分析

5.1 实验说明

由于模型中的方程为无法直接求解的非线性方程，故利用数字分析软件 Matlab 提供的组件 Simulink 来求解理论值.为了仿真 P2P 网络，基于专门的 P2P 网络仿真平台 PeerSim 和相关 P2P 网络协议开发出了仿真软件.在仿真时，对每一组参数值运行仿真软件 20 次，将 20 次结果的平均值作为该组参数对应的仿真值.鉴于论文篇幅有限，只展示了部分实验结果，在不作特别说明的情况下，下面图形中涉及到的模型都为传播模型，使用表 1 的参数值作为默认值.

5.2 仿真值和理论值的比较

图 1～图 3(图中的 T 和 S 分别代表理论值和仿真值)分别考查了了不同下载率和恢复率情况下，仿真值和理论值的匹配情况.图中清楚表明，仿真值曲线和理论值曲线走势相当一致，在不同参数条件下所做的其它实验也都反映了这种情况，这就从实验上充分表明，前面所提出的激发型蠕虫动态传播数学模型是有效的，可以用来预测和分析蠕虫的传播趋势.

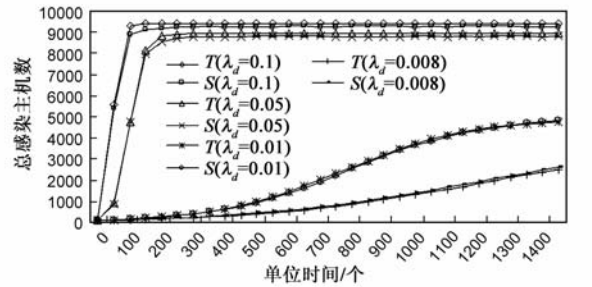

图1 不同下载率下仿真值和理论值的比较(1)

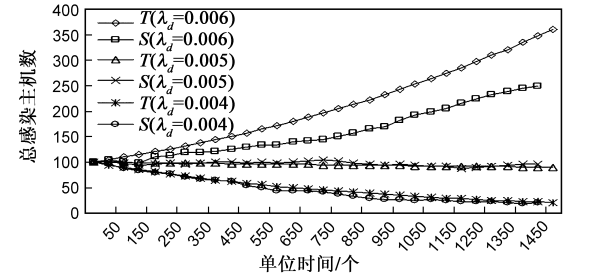

图2 不同下载率下仿真值和理论值的比较(2)

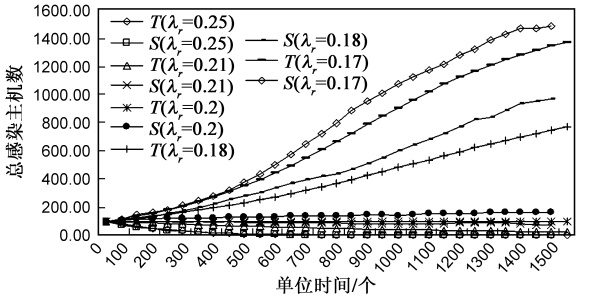

图3 恢复率对蠕虫传播的影响

5.3 蠕虫不会泛滥充分条件的实验证明

图 2～图 5 对应实验用来检验前面推导出来的蠕虫不会泛滥充分条件的正确性.这些实验对应的 R_0 值见表 2.结合表 2 和图 2～图 5 容易看出，在 R_0 小于 1 时，蠕虫逐渐减少，最后消失；在 R_0 很接近于 1 的情况下，蠕虫数量基本保持稳定，而当其大于 1.1 时，蠕虫就呈现出蔓延的态势.这表明推导出的蠕虫不会泛滥的充分条件是正确的.

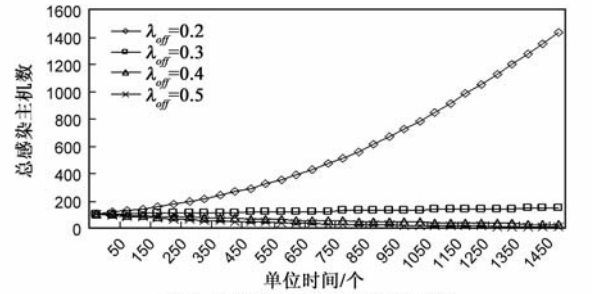

图4 下线率对蠕虫传播的影响

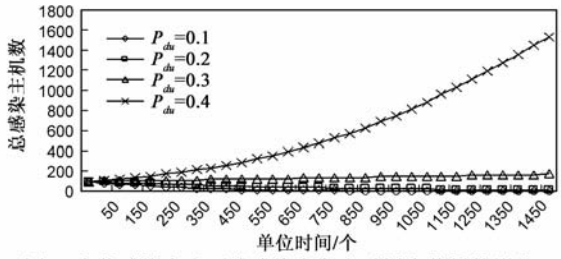

图5　上传感染率和下载感染率之和对蠕虫传播的影响

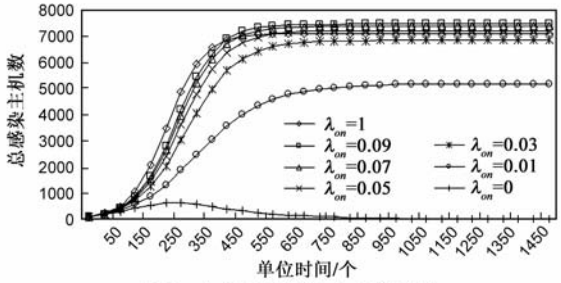

图6　上线率对蠕虫传播的影响

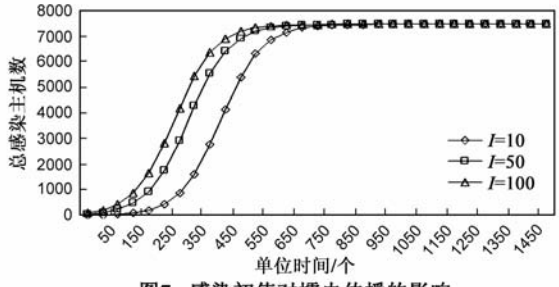

图7　感染初值对蠕虫传播的影响

5.4　P2P 参数对蠕虫传播的影响

从图 1 ~ 图 7 和其它仿真实验结果上看,下载率对蠕虫泛滥程度(或流行程度,其值为平衡状态时总感染主机数与总主机数的比值)的影响最大,其次是恢复率,再次是下载感染率与上传感染率之和,影响最小的是上线率和下线率,初始感染主机数对泛滥程度没有影响.使用默认参数值实验,实验结果表明蠕虫一定会泛滥.在这种情况下,修改上线率再进行实验发现:无论如何改变上线率(大到 1、小到 0.01),蠕虫都会泛滥,所以上线率对蠕虫是否会泛滥影响很小(见图 6),同样结论也适用于下线率.结合表 2 和图 2 ~ 图 5 还能看出,R_0 是一个十分重要的参数,除了可以确定蠕虫是否会泛

表 2　图 2 ~ 图 5 实验的 R_0 值

λ_d	R_0	λ_r	R_0	λ_{off}	R_0	p_{du}	R_0
0.01	1.99	0.025	0.81	0.2	1.42	0.1	0.36
0.008	1.59	0.021	0.96	0.3	1.06	0.2	0.72
0.006	1.19	0.02	1.01	0.4	0.844	0.3	1.08
0.005	0.99	0.018	1.12	0.5	0.70	0.4	1.45
0.004	0.79	0.017	1.18	—	—	—	—

滥外,还表明了蠕虫的传播能力.在蠕虫会泛滥的情况下,R_0 越大,传播越快,到达平衡状态的时间越短;在蠕虫不会泛滥的情况下,R_0 越大,其消失需要的时间越长.

5.5　基于实测数据的分析

为预测和评价激发型蠕虫在现实的 P2P 网络上的传播情况,基于激发型蠕虫传播模型使用 Gnutella 网络实测数据[23~25]对蠕虫传播进行仿真分析.在 Gnutella 网络上,绝大部分用户在线时间不到 20 分钟,而用户下载一个文件的平均时间是 15 分钟,通常 65% 的用户每天仅上网 1 次.根据这些实测数据并结合下载文件经验,估计用户每次在线时间是 20 分钟,每天上线 1 ~ 2 次,每次下载 1 ~ 2 个文件.如果每天用户上线 1 次,相应的离线平均时间是 1420(24 × 60 – 20)分钟;如果每天上线 2 次,每次离线的平均时间是 700(12 × 60 – 20)分钟;.根据这些数据,进行了 4 轮实验.第 1、2 轮实验对应每天上线 1 次的情况下分别下载 1 个、2 个文件的情形;第 3、4 轮实验对应每天上线 2 次的情况下分别下载 1 个、2 个文件的情形.为了保证实验时在线主机数始终为 1000,离线用户主机数在每天上线 1 次和 2 次的情况下分别取为 71000 台和 35000 台(现实网络运行到一定时间后会达到相对平衡状态即在线用户数相对固定).4 轮实验中,变化的参数值和初始变量值见表 3,相同的参数和初始变量值为:$S_{on}=990$, $I_{on}=10$, $I_{off}=0$, $\lambda_r=0.0001$, $\lambda_{off}=1/20=0.05$, $p_d=0.9$, $p_u=0.8$.

表 3　4 轮实验使用的不同参数和变量值

标识	N	λ_d	λ_{on}	S_{off}	R_0
(1,1)	72000	0.05	0.0007	71000	13.46
(1,2)	72000	0.1	0.0007	71000	26.91
(2,1)	36000	0.05	0.0014	35000	25.21
(2,2)	36000	0.1	0.0014	35000	50.43

图 8 和图 9 分别展示了 4 轮实验中总的感染情况和在线主机的感染情况.图 8 表明,在每天上线一次且每次上线下载 1 个文件的情况下,蠕虫感染网络中一半的主机仅需要不到 2 天时间,达到平衡状态要近 6 天时间,此时超过 80% 的主机被感染;在每天上线 2 次每次下载 1 个文件情况下,感染一半的主机仅需 1 天,达到平衡状态需要 3 天,同样有超过 80% 的主机被感染.实验还发现,在 P2P 参数确定的情况下,蠕虫传播趋势和到达平衡状态的时间也是确定的,即蠕虫的传播只与 P2P 参数有关而与网络的规模无关.比如,在每天上线 2 次每次下载一个文件的情况(其它参数不变)下,将网络规模扩大到 1 千万台主机进行实验,实验结果表明,不到 1 天就感染了 5 百万台主机,感染 9 百万台需要 3 天时间,这比 Staniford 估计的一个月要短得多,这说明 P2P 网络中的激发型蠕虫对网络构成的威胁和危害比想象中要大得多.从图 9 容易看出,激发型蠕虫在泛滥过程中有

一个非常明显的短暂的急速增长期,过了该时期后,传播明显减缓.下载率为0.1(每次上线20分钟并下载2个文件)对应的激增期(约65分钟)比下载率为0.05(每次上线20分钟并下载1个文件)短得多(约155分钟),这再次表明下载率是影响蠕虫传播的关键因素.

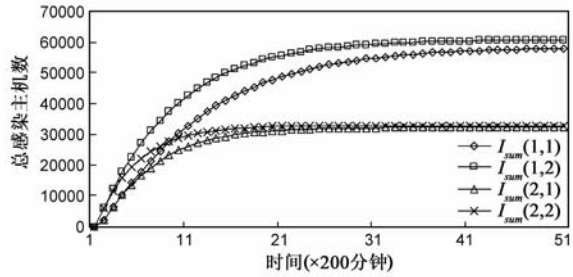

图8 网络中总的感染情况

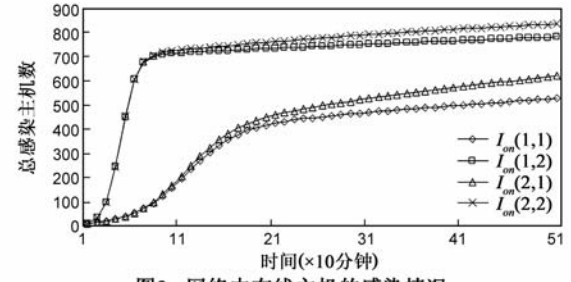

图9 网络中在线主机的感染情况

5.6 蠕虫传播控制

既然影响蠕虫传播的主要因素是下载率、恢复率、下线率、上线率、下载感染率和上传感染率这些参数,而前4个参数是用户可以控制的,那么显然在针对漏洞的补丁未编制出来的情况下似乎通过控制这4个参数就能有效控制蠕虫传播,事实是这样吗?

在保持其它参数不变的情况下,将标识(1,1)(该实验接近现实情况)对应实验的上线率和下线率作各种改变,发现降低上线率、提高下线率能减缓蠕虫的传播,但作用十分有限,无论如何改变它们都无法改变蠕虫会泛滥的状态.

提高恢复率可以明显削弱蠕虫的传播能力.提高恢复率有两种方法:一种是用户利用自己的反病毒知识来清除蠕虫;另一种则是重装系统.即使是对于专门的网络安全人士,要他们在较短的时间内利用自己的网络安全和病毒防范知识来清除计算机上杀毒软件都无法清除的蠕虫(对于新出现的病毒,杀毒软件往往也无能为力)几乎是不可能的,所以提高恢复率的主要方法是重装系统,而这种方法的副作用也是很明显的(数据丢失、用户无法正常工作等).

下载率是影响蠕虫传播最关键的因素,在发现激发型蠕虫的情况下大幅降低下载率就能有效遏制蠕虫传播.在Gnutella这样的给用户带来充分自由的开放网络中,降低下载率的唯一方式是提醒用户网络上存在蠕虫,而这种方式作用十分有限,一方面是时间很紧,仅需3天时间,蠕虫就会蔓延开来,而这段时间用户可能还不知道蠕虫的存在;另一方面,即使用户获知了蠕虫的消息并降低了下载率,但降低后的下载率未必就能保证蠕虫不会蔓延.对于实验(1,1),下载率要降到0.004以下即每13天(上线共250分钟)仅下载1个文件,才能保证蠕虫不会蔓延,而要做到这一点是难度是很大的.在e-Donkey网络中,控制下载率相对较容易,通过查询服务器控制用户对文件的查询就能有效控制下载率.

通过上节分析知道,蠕虫传播有一个较短的急增过程,因此,控制蠕虫的最佳时期就是这段时期.显然,尽早控制蠕虫的前提是尽早发现蠕虫,然而,激发型蠕虫利用正常连接和流量传播的特点决定了要尽早检测到它是很难的.可能的情况是,检测到它的时候也就是它泛滥和爆发的时候,采取控制措施(降低下载率、提高恢复率等)为时已晚.

6 总结与展望

P2P激发型蠕虫,是一种对P2P网络构成极大威胁却未受到重视的蠕虫.针对这种威胁性和危害性极大的蠕虫,该文的主要工作和创新点在于:(1)提出了P2P激发型蠕虫传播的数学模型和免疫模型,模型还考虑了网络的动态特征;(2)基于传播模型和流行病学理论推导出蠕虫不会泛滥的充分条件;(3)通过大规模仿真实验验证了P2P激发型蠕虫的传播模型和蠕虫不会泛滥充分条件的有效性和正确性;(4)基于传播模型分析了各P2P参数对蠕虫传播的影响,指出了下载率是影响蠕虫传播的最关键因素,基本繁殖率是衡量蠕虫传播能力的关键指标;(5)基于实测数据,预测和估计了P2P激发型蠕虫的传播能力、速度和危害性,指出目前尚无检测和控制它的有效方法.

参考文献

[1] Zhou L, Zhang L, McSherry F, et al. A first look at peer-to-peer worms: Threats and defenses[A]. Proc of the 4th Int Workshop on Peer-to-Peer Systems[C]. New York: Springer, 2005. 24 – 35.

[2] Staniford S, Paxson V, Weaver N. How to Own the Internet in Your Spare Time[A]. Proc of the 11th USENIX Security Symposium[C]. San Francisco: ACM, 2002. 149 – 167.

[3] Chen G, Gray R S. Simulating non-scanning worms on peer-to-peer networks[A]. Proc of the 1st Int Conf on Scalable Information Systems[C]. Hong Kong: ACM, 2006.

[4] Murray W H. The application of epidemiology to computer viruses[J]. Computers and Security, 1988, 7: 130 – 150.

[5] Kephart J O, White S R. Directed-graph epidemiological models of computer viruses[A]. Proc of IEEE Symp. Security and Privacy[C]. Oakland: IEEE, 1991. 343 – 359.

[6] Zou C C, Gong W, Towsley D. Code red worm propagation modeling and analysis[A]. Proc of ACM Conf on Computer and Communication Security (CCS'02)[C]. Washington: ACM, 2002.

[7] Liljenstam M, Yuan Y, Premore B, et al. Nicol. Email worm modeling and defense[A]. Proc of IEEE Int. Symp[C]. MASCOTS, Fort Worth, TX, Oct. 2002.

[8] Wei Y. Analyze the Worm-Based Attack. in Large Scale P2P Networks[A]. Proc of the 8th IEEE International Symposium on High Assurance Systems Engineering (HASE'04)[C]. Tampa: IEEE, 2004. 308 – 309.

[9] Wei Y. Analyzing the performance of internet worm attack approaches[A]. Proc of the 13th International Conference on Computer Communications and Networks[C]. Chicaco: IEEE, 2004. 1095 – 2055.

[10] Dumitriu D, Knightly E, Kuzmanovic A, et al. Denial-of-service resilience in peer-to-peer file-sharing systems[A]. Proc of ACM Sigmetrics[C]. Banff, Canada: ACM, 2005.

[11] Thommes R W, Coates M J. Modeling Virus Propagation in Peer-to-Peer Networks[R]. Montreal, Canada: Department of Electrical and Computer Engineering, McGill University, 2005.

[12] 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[J].计算机学报,2006,29(7):952 – 959.
Xia Chun-he, Shi Yun-ping, Li Xiao-jian. Research on Epidemic Modesl of P2P Worms in Structured Peer-to-Peer Networks[J]. Chinese Journal of Computers, 2006, 29(7): 952 – 959. (in Chinese)

[13] Ma J, Chen X M, Xiang G L. Modeling passive worm propagation in peer-to-peer system[A]. Proc of the IEEE 2006 International Conference on Computational Intelligence and Security[C]. Hong Kong: IEEE, 2006. 1129 – 1132.

[14] 王方伟,张运凯,马剑峰.无结构 P2P 网络中被动型蠕虫传播建模和防治[J].天津大学学报,2008,14(1):66 – 72.
Wang Fang-wei, Zhang Yun-kai, Ma Jian-feng. Modeling and defending passive worms over unstructured peer-to-peer networks[J]. Transaction of Tianjin University, 2008, 14(1): 66 – 72. (in Chinese)

[15] 王跃武,荆继武,向继,等. Contagion 蠕虫传播仿真分析[J].计算机研究与发展,2008,45(2):207 – 216.
Wang Yue-wu, Jing Ji-wu, Xiang Ji, et al. Contagion worm propagation simulation and analysis[J]. Journal of Computer Research and Development, 2008, 45(2): 207 – 216. (in Chinese)

[16] 应凌云,冯登国,苏璞睿.基于 P2P 的僵尸网络及其防御[J].电子学报,2009,37(1):31 – 37.
Ying Ling-yun, Feng Deng-guo, Su Pu-rui. P2P-based super botnet: Threat s and defenses[J]. Acta Electronica Sinica, 2009, 37(1): 31 – 37. (in Chinese)

[17] Jovanovic, M. A. Modeling Large-Scale Peer-to-Peer Networks and a Case Study of Gnutella[D]. Cincinnati, Ohio: University of Cincinnati, 2001.

[18] edonkey website[OL]. http://www.donkey.org/, 2009-5-8.

[19] 1 Frauenthal J C. Mathematical Modeling in Epidemiology[M]. New York: Springer, 1980.

[20] Diekmann O, Heesterbeek J A P. Mathematical Epidemiology of Infectious Diseases: Model Builind, Analysis and Interpretation[M]. Wiley, 1999.

[21] Driessche P, Watmough J. Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission[J]. Mathematical Biosciences, 2002, 180: 29 – 48.

[22] Arnio J, Davis J, Hartley D, et al. A multi-species epidemic model with spatial dynamics[J]. Mathematical Medicine and Biology, 2005.

[23] Nicolas C, Andreas S, Chuang J. Content availability, pollution and poisoning in file sharing peer-to-peer networks[A]. Proc of ACM EC'05[C]. Vancouver, British, 2005.

[24] Ilie D. On Unicast QoS Routing in Overlay Networks[D]. Sweden, Karlskrona: Blekinge Institute of Technology, 2008.

[25] Sen S, Wang J. Analyzing peer-to-peer traffic across large networks[J]. IEEE/ACM Transactions on Networking, 2004, 12(2): 219 – 232.

作者简介

冯朝胜　男,1971 年出生四川广元,博士后,副教授,硕士生导师,中国计算机协会高级会员.2010 年获得电子科技大学信息与通信工程博士学位.研究方向为分布式计算、网络与信息安全、恶意代码分析.
E-mail: jamesjiangfeng@163.com

袁　丁　男,1967 年生于四川宜宾,博士,四川师范大学计算机科学学院教授,硕士生导师.2003 年获得西南交通大学工学博士学位.研究方向为网络与信息安全.

卿　昱　女,1970 出生于四川,中国电子科技集团公司第三十研究所研究员,硕士生导师.研究方向为网络安全、软件设计、系统集成.

秦志光　男,1956 年生于四川荣昌,博士,电子科技大学计算机科学与工程学院教授、博士生导师,IEEE 高级会员.研究方向为密码学、网络与信息安全.