邮件蠕虫传播与防御的分析研究 *

罗卫敏[1,2]，张凤荔[2]
(1. 重庆三峡学院 数学与计算机科学学院，重庆 万州 404000；2. 电子科技大学 计算机科学与工程学院，成都 610054)

摘 要：邮件蠕虫利用 e-mail 在具有 power-law 结构特点的网络中进行传播，使得传统的蠕虫防御策略失效。结合 power-law 网络拓扑结构的特点，引入节点免疫和邮件服务器参与两种防御策略，分别对重复感染与非重复感染两种类型的邮件蠕虫传播进行了实验仿真。结果表明，节点的优先免疫类型、免疫起始时间、邮件服务器参与防御时间及蠕虫邮件识别正确率都与邮件蠕虫的传播有着紧密联系。

关键词：网络安全；邮件蠕虫；幂律法则；免疫
中图分类号：TP309.5 **文献标志码**：A **文章编号**：1001-3695(2009)04-1532-03

Analysis and research of propagation and defense of e-mail worm

LUO Wei-min[1,2], ZHANG Feng-li[2]
(1. *College of Mathematics & Computer Science, Chongqing Three Gorges University, Wanzhou Chongqing* 404000, *China*; 2. *School of Computer Science & Engineering, University of Electronic Science & Technology of China, Chengdu* 610054, *China*)

Abstract: E-mail worm spreads in the power-law network by sending e-mail. The policies which were used to defend traditional worm were improper to defend e-mail worm. This paper presented the e-mail worm simulation based on power-law topology on two e-mail worm styles including repeated infection and unrepeated infection by considering two defense policies included node immunization and e-mail server's filter function. The results show that many factors included the first immunized node style, the start-time of immunization, the start-time of e-mail server's filter and the filter preciseness have great influence on e-mail worm propagation.

Key words: network security; e-mail worm; power-law; immunization

0 引言

邮件蠕虫属于蠕虫的一个分支，以电子邮件为传播媒介进行传播，感染主机后，会依据邮件联系簿上的地址列表主动发送自身拷贝，引发大量的垃圾邮件流量。著名的邮件蠕虫有 1999 年的“Melissa”、2000 年的“Lover Letter”、2001 年的“W32/Sircam”、2003 年的“SoBig”、2004 年的“Bagle”“Netsky”“Plexus”、2005 年的“MyDoom”、2006 年的风暴蠕虫(Storm worm)等。邮件蠕虫主要运用社会工程技巧诱使邮件查阅者点击蠕虫附件，一旦点击，蠕虫将自动尝试链接远程计算机并转发大量蠕虫邮件。

邮件蠕虫发展到现在，呈现出病毒、蠕虫和木马的融合发展趋势，具有巨大的破坏力。2006 年底和 2007 年初的风暴蠕虫，综合了木马后门、僵尸网络、攻击工具包、加密及 P2P 网络等技术，设计愈加复杂化。风暴蠕虫会在感染主机上安插后门程序，将大量的感染主机组成僵尸网络(botnet)，继而引发海量的垃圾邮件攻击、分布式拒绝服务攻击以及其他以金融为目的的攻击行为。

邮件蠕虫与其他蠕虫相比有着明显的特点：不扫描、被动触发、基于邮件列表传播、传播网络呈现幂律特征等，传统蠕虫防御策略并不适用于邮件蠕虫，对邮件蠕虫的防御更多地依赖于用户的安全意识和机器免疫。

本文运用仿真实验，从节点免疫和邮件服务器参与两个方面对邮件蠕虫的传播进行观察和研究，给出了邮件蠕虫模型，并进行实验仿真和分析。

1 相关工作

2001 年红色代码蠕虫爆发后，Staniford 等人[1]研究了 Internet 蠕虫的传播模型，许多学者相继在此基础上对蠕虫的传播模型进行了深入研究[2~8]。但是，当时研究的传播限于随机扫描型蠕虫，而这些传播模型并未表现出邮件蠕虫的传播特征，如后文所述，邮件蠕虫不采用随机扫描策略，同时感染的拓扑网络图呈现 power-law 特点。

Pastor-Satorras 等人[9]研究了在稳定状态下的感染极限，通过节点的出入度数来体现蠕虫在拓扑逻辑图上的传染概率，修改了 Susceptible-infectious-susceptible(SIS)模型。

Moreno 和 Boguna 等人[10~12]提出了 Susceptible-infectious-recovered(SIR)模型，研究了在不稳定状态下蠕虫的传播行为，但是基于网络中为同类系统的假设，使 SIR 模型高估了蠕虫的传播速度。

Wang 等人[13]研究了基于串、树的网络拓扑结构中蠕虫的传播模型，蠕虫在没有人工干涉的情况下呈匀速传播，显然这不符合邮件蠕虫的传播情况。文献[14]中 Wong 等人在对某校园网监控的基础上分析了 SoBig 蠕虫和 MyDoom 蠕虫的传

收稿日期：2008-06-16；**修回日期**：2008-07-17 **基金项目**：国家“242”信息安全计划资助项目(2006C27)

作者简介：罗卫敏(1977-)，男，重庆万州人，硕士研究生，主要研究方向为网络安全(wmluo123@yahoo.com.cn)；张凤荔(1963-)，女，教授，博导，博士，主要研究方向为移动数据管理及其应用、网络安全.

播。王长广等人[15]用有向图描述了电子邮件网络的结构,并分析了电子邮件网络的无尺度特性,在此基础上,通过用户检查邮件的频率和打开邮件附件的概率建立了一种电子邮件蠕虫的传播模型。Newman 等人[16]基于校园网用户的 e-mail 地址簿中地址的数量得出了在校园网内,e-mail 用户网络呈指数分布这一结论,但是结论的前提是只针对单个用户地址簿,没有考虑到多数邮件蠕虫在感染主机后,会向主机上所有用户的地址簿中所有地址进行传播的情况。

针对病毒和蠕虫的传播,许多学者采用了免疫策略进行防御。所谓免疫,是指在网络中有一部分节点经过预先处理,在蠕虫传播时不会被感染。Wang 等人[13]指出在类似树型结构的网络中,选择性的免疫可以有效减缓蠕虫的传播。Zou 等人[17]给出了在 power-law 网络中实行免疫策略,减缓了蠕虫的传播速度,降低了蠕虫范围,但是未考虑免疫是一个连续的、动态的过程。

2　邮件蠕虫模型

本文以文献[18]所提出的邮件蠕虫模型为基础,采用无向图描述邮件网络拓扑结构:$G=\langle V,E\rangle$。其中:$\forall v\in V$,v 表示网络中某一个邮件用户;$\forall e=(u,v)\in E$,e 表示邮件用户 u 拥有邮件用户 v 的邮件地址;邮件用户拥有 e 的多少称为度 d,d 的大小意味着该邮件用户的好友数量;$|V|$ 表示网络内邮件用户总数。

当用户的邮件地址中包含有电子邮件组时,用户即同时拥有电子邮件组内所有的邮件地址,这就意味着即使用户邮件地址中包含的数量很少,也会因为邮件组的关系,导致用户在邮件网络中的度非常高,使邮件网络呈现出重尾分布的特点,也即幂律拓扑特征。

邮件蠕虫不需要利用系统漏洞,它通过邮件进行传播,使用社会工程学相关知识诱使用户点击蠕虫附件进行感染。这其中有两种行为与用户被感染有关,即检查邮件的频率 F_i 和被诱使点击附件的概率 P_i。

本文使用邮件的检查时间间隔 $T_i(i=1,2,\cdots,|V|)$ 表示邮件的检查频率 $F_i(i=1,2,\cdots,|V|)$,用 $E[T_i]$ 表示用户 i 的检查时间间隔的平均值。T_i 可能遵循不同的分布,这由用户习惯决定:当用户用固定时间或固定时间间隔进行邮件检查,那 T_i 就是一个常量;当用户的邮件检查是随机性的,则 T_i 是一个指数分布,用户的检查行为就是一个泊松过程。对于概率 P_i,则取决于用户的安全意识和蠕虫使用的社会工程学相关知识的熟练程度。

3　仿真实验

仿真实验使用的拓扑网络节点总数为 100 000 个,平均节点度为 8,幂律系数 1.7。仿真实验有如下假设:

a)邮件网络中的节点都是同构的,邮件蠕虫在传播中对网络没有搅动行为。

b)基于邮件列表发送邮件的时间是同时的,且没有考虑邮件的传送时间。

c)用户检查邮件时会检查所有新邮件。

d)用户对相同的邮件蠕虫的打开概率 P_i 不变。

e)不考虑利用邮件程序漏洞进行自动传播的蠕虫。

本文用 $E[N_t]$ 表示任意时刻 t 感染用户的平均数,仿真实验使用不同的种子产生随机数进行 100 次仿真,对结果 N_t 取平均值后得到 $E[N_t]$;用户检查邮件的时间间隔 T_i 由高斯分布的随机变量 T 生成,$T\sim N(40,20^2)$($T<0$,$E[T_i]=0$);P_i 由高斯分布的随机变量 P 产生,$P\sim N(0.5,0.3^2)$($P<0$,$P_i=0$;$P>1$,$P_i=1$);初始感染节点数为 2,每次仿真实验的初始感染节点是随机选择的。

3.1　两种感染情况分析

邮件蠕虫感染时分为两种情况:a)重复感染。用户在已经感染的情况下,再次打开蠕虫附件时依然会向邮件地址中所有用户发送蠕虫邮件,如 W32/Sircam 蠕虫。b)非重复感染。用户只在首次感染时会向邮件地址中所有用户发送蠕虫邮件,如 Melissa 和 Love Letter 蠕虫。图 1 为两种类型的蠕虫传播情况。

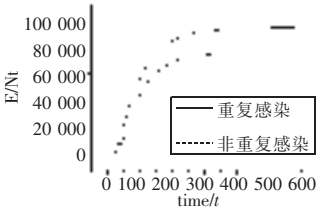

图 1　重复感染与非重复感染两种情况下的蠕虫传播

从图 1 可以看出,重复感染比非重复感染传播速度快、传播范围广。这主要是因为,虽然用户打开概率 P_i 不变,但在收到 n 个蠕虫邮件时的感染概率为 $1-(1-P_i)^n$,用户收到的蠕虫邮件数量越多,就越增加了用户感染的概率。显然,可靠在重复感染情况下用户收到的蠕虫邮件要比在非重复感染情况下多得多。

3.2　邮件服务器参与防御的时间分析

设定在 t 分别为 50、100、150 时开始发挥邮件服务器的邮件过滤功能,设定蠕虫邮件识别率为 100%,邮件蠕虫传播如图 2 和 3 所示。

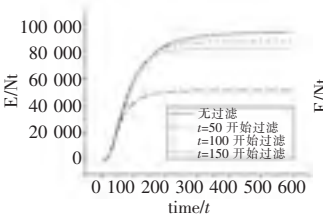

图 2　重复感染情况下不同参与时间的蠕虫传播

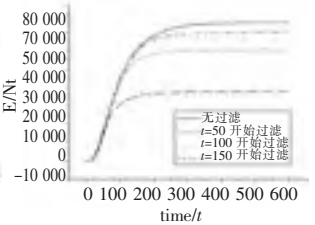

图 3　非重复感染情况下不同参与时间的蠕虫传播

从图 2、3 可以看出,邮件服务器越早参与蠕虫防御,就越能将邮件蠕虫遏制在有限的范围内。在重复感染的情况下,邮件服务器参与时间一旦滞后,对蠕虫的遏制效果就不再那么明显,过滤拦截功能基本失去作用,而对于非重复感染蠕虫,参与时间的前后所造成的遏制效果还是有一定区别。

3.3　邮件服务器过滤有效性分析

图 2、3 都是假设邮件服务器过滤功能的有效性达到 100%,但事实上,与过滤拦截垃圾邮件一样,不可能对所有的蠕虫邮件达到过滤拦截。实验中设定蠕虫邮件识别率分别为 40%、60%、80%、100%,在 t 为 50 时邮件服务器开始进行过滤拦截,邮件蠕虫传播如图 4、5 所示。

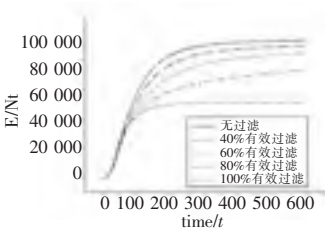

图 4 重复感染情况下不同过滤效果的蠕虫传播

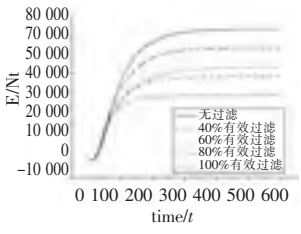

图 5 非重复感染情况下不同过滤效果的蠕虫传播

从图 4、5 可以看出,邮件服务器的蠕虫邮件识别技术越好,遏制邮件蠕虫传播的效果就越好。在非重复感染情况下,感染主机只发送一次蠕虫邮件,所以即使有效性不高,蠕虫传播也可以较平稳地控制在某个范围;而在重复感染情况下,由图 4 可以看出,在海量的蠕虫邮件发送下,邮件服务器的过滤拦截功能的有效性高低对最终蠕虫感染范围无特别明显的影响,只能起到减缓蠕虫传播速度的效果。

3.4 不同时间开始节点的免疫

文献[17]中所描述的免疫场景是个理想状态,Zou 等人假定免疫节点在初期已经散布在网络中,但现实情况是免疫节点的增加是逐步的、连续的过程,节点的免疫是伴随着邮件蠕虫的爆发与传播同时发生的。

考虑到免疫需要占用一定时间,实验中指定每 3 个时间单位进行节点免疫,在 t 分别为 50、100、150 开始进行免疫,每次选取随机 100 个节点进行免疫,邮件蠕虫传播如图 6、7 所示。

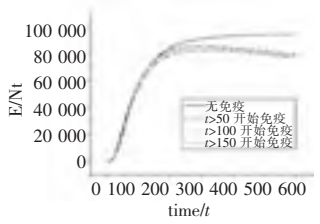

图 6 重复感染情况下不同时间开始免疫

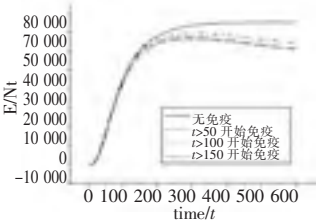

图 7 非重复感染情况下不同时间开始免疫

从图 6、7 可以发现,随机性的节点免疫即使开始时间不同,但对于蠕虫的传播并不能起到良好的遏制效果,不同时间之间的区别并不明显,而且蠕虫依然会迅速地传播到高峰状态,随后才会随着免疫节点的增多,感染节点逐渐减少。

3.5 固定时间不同数量节点的免疫

设定在 t 为 50 时开始随机节点的免疫,依然考虑节点的免疫时间,观察每次对不同数量的节点进行免疫对邮件蠕虫传播的遏制效果,邮件蠕虫传播如图 8、9 所示。

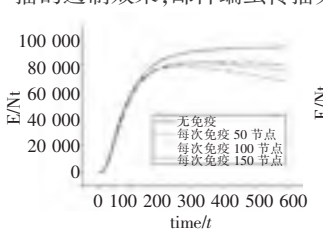

图 8 重复感染情况下不同数量的节点免疫

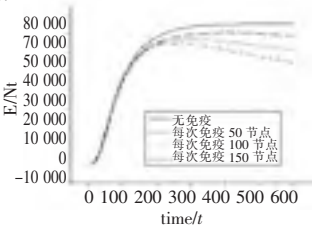

图 9 非重复感染情况下不同数量的节点免疫

从图 8、9 可以发现,随机性的节点免疫即使数量不同,对于蠕虫的传播依然不能起到良好的遏制效果,邮件蠕虫基本会在 t 为 200 时达到传播高峰状态,随后才会随着免疫节点的增多,感染节点逐渐减少。

3.6 选择性节点优先免疫

在幂律网络拓扑中,具有大量连接的节点称为超级节点,它的稳定性和安全性对整个拓扑网络具有至关重要的影响。实验中在初始时刻分别对 5 000 个超级节点、一般节点和随机节点进行优先免疫,邮件蠕虫传播如图 10、11 所示。

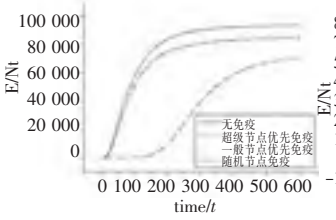

图 10 重复感染情况下不同节点优先免疫

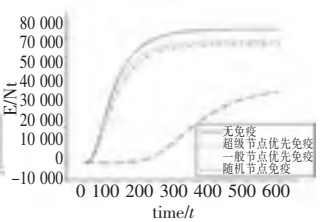

图 11 非重复感染情况下不同节点优先免疫

从图 10、11 可以看出,超级节点的优先免疫可以极大地减缓蠕虫的传播,而一般节点优先和随机节点免疫区别不大,基本不能遏制蠕虫的传播。同时在非重复感染情况下,超级节点优先免疫取得的遏制效果非常明显。

3.7 不同时间对超级节点优先免疫的影响

实验中在 t 分别为 1、25、50、100 时对 5 000 个超级节点优先免疫,观察超级节点的免疫时间对邮件蠕虫传播的影响情况,邮件蠕虫传播如图 12、13 所示。

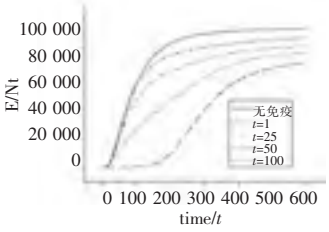

图 12 重复感染情况下超级节点优先免疫的时间影响

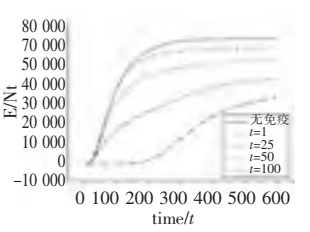

图 13 非重复感染情况下超级节点优先免疫的时间影响

从图 12、13 可以看出,超级节点优先免疫策略的时间敏感性很强。在蠕虫爆发初期进行免疫能有效地减缓蠕虫的传播速度;时间稍微延后,蠕虫即呈现线性传播甚至于指数传播,超级节点的免疫效果迅速降低。但在两种不同的感染情况比较下,超级节点优先免疫还是可以对非重复感染性的邮件蠕虫起到较好的效果,这缘于蠕虫传播链中缺少超级节点的参与,蠕虫邮件数量大幅减少,导致其他用户的感染率下降;而对于重复感染性的邮件蠕虫,一般节点会反复发送蠕虫邮件,一定程度上弥补了超级节点缺席传播的不足。

4 结束语

本文针对重复感染和非重复感染两种类型的邮件蠕虫,利用仿真实验模拟其在邮件服务器的过滤干预和动态免疫情况下的传播过程,研究了这两种防御策略对抑制邮件蠕虫传播的真实效果,对进一步研究邮件蠕虫的防御策略起到了很好的启示作用。

参考文献:

[1] STANIFORD S, PAXSON V, WEAVER N. How to own the Internet in your spare time[C]//Proc of the 11th Usenix Security Symp. Berkeley: USENIX Association, 2002: 149-167.

[2] CHEN Z, GAO L, KWIAT K. Modeling the spread of active worms[C]// Proc of IEEE INFOCOM '03. 2003: 1890-1900.

(下转第 1537 页)

密同态加密算法更具有实用性。加密过程中即使经过 $E_{k1}(x)$ 加密转换后得到相同的数据,由于第二次同态加密素数的随机选取和加密数据的随机分割,这样得到的加密数据也是不一样的。浮点数同态加密即在外层加密中保留了原始秘密同态加密的安全性,同时也对原数据进行了双重同态变换,在安全性上只有过之而无不及。由 Dominggo-Ferrer 等人在文献[2]中的讨论可知,在浮点数上的同态加密机制在安全性方面同样能抵抗仅知密文攻击和已知明文攻击。

3　字符串数据的同态加密

与整数不同的是,整数加密后能够实现的在密态下的加、减、乘、除运算对于字符串是没有意义的,所以本文通过中国剩余定理将字符串进行转换,然后利用秘密同态算法进行加密处理。具体算法如下:

a)设一字符串 B,将字符串中的每一个字符取其 ASCII 码,分别表示为 $b_1, b_2, \cdots, b_k$。其中 k 是字符串中字符的个数。

b)对应取 k 个两两互素的正整数,设为 $m_1, m_2, \cdots, m_k$ $(m_i \geqslant 121)$。

c)由中国剩余定理得同余式组

$$\begin{cases} x \equiv b_1 \pmod{m_1} \\ x \equiv b_2 \pmod{m_2} \\ \cdots \\ x \equiv b_k \pmod{m_k} \end{cases}$$

d)同余式组的解 $x \equiv \sum_{i=1}^{k} M'_i M_i b_i \pmod{m}$ (令 $m = m_1 m_2 \cdots m_k$; $M_i = m/m_i$; $M'_i = M_i^{-1} \bmod m_i$; $i = 1, 2, \cdots, k$) 就是将要加密的整数值。

e)根据秘密同态算法解密后可得加密数据 x,则由 $b_i = x \bmod m_i$ 可得字符串中每个字符的对应 ASCII 码值。

利用中国剩余定理的证明方法可对字符串与所得加密整数值的对应关系进行证明。在具体的实现过程中,素数的选取和存储也是需要考虑的问题。在安全性方面,它仍然保留了原整数秘密同态加密的特点。

4　结束语

秘密同态加密技术是一种能对密文数据库进行数学运算和常规的数据库操作的加密机制。可以对数据库信息进行操作但又不泄露其中的内容,但到目前为止最好的成果仅能用在整数的范围内。本文在原加密算法的理论基础上,对同态算法进行了扩展,实现了在浮点数和字符串数据上秘密同态算法的应用。目前本算法已成功应用于某煤码头皮带秤管理系统中,较好地解决了系统的安全问题。

参考文献:

[1] DOMINGO F J. A new privacy homomorphism and applications[J]. Information Processing Letters,1996,60(5):277-282.

[2] BRICKELL E F, YACOBI Y. On privacy homomorphisms[C]//Proc of Advances in Cryptology-Eurocrypt'87, LNCS 304. Berlin: Springer-Verlag,1988:117-125.

[3] RIVEST R L, ADLEMAN L, DETROUZOS M L. On data banks and privacy homomorphism[C]//DeMILLO R A. Proc of Foundations of Secure Computation. New York: Academic Press,1978:169-179.

[4] 杨勇,方勇,周安民. 秘密同态技术研究及其算法实现[J]. 计算机工程,2005,31(2):157-159.

[5] 向广利,陈苹萌,马捷,等. 实数范围上的同态加密机制[J]. 计算机工程与应用,2005,41(20):12-14.

[6] 王晓峰,王尚平. 秘密同态技术在数据库安全中的应用[J]. 计算机工程与应用,2003,39(14):194-196.

(上接第 1534 页)

[3] KESIDIS G, HAMADEH I, JIWASURAT S. Coupled Kermack-McKendrick models for randomly scanning and bandwidth-saturating Internet worms[C]//Proc of the 3rd Int'l Workshop on QoS in Multiservice IP Networks (QoS-IP). 2005:101-109.

[4] NICOL D, LILJENSTAM M. Models of Internet worm defense[EB/OL]. (2004-01). http://www.ima.umn.edu/talks/workshops/1-12-16.2004/nicol/talk.pdf.

[5] NOJIRI D, ROWE J, LEVITT K. Cooperative response strategies for large scale attack mitigation[C]//Proc of the 3rd DARPA Information Survivability Conf and Exhibition. 2003.

[6] WU J, VANGALA S, GAO L, *et al*. An efficient architecture and algorithm for detecting worms with various scan techniques[C]//Proc of the 11th Ann Network and Distributed System Security Symp (NDSS'04). 2004.

[7] ZOU C, GAO L, GONG W, *et al*. Monitoring and early warning for Internet worms[C]//Proc of the 10th ACM Conf on Computer and Comm Security (CCS '03). 2003:190-199.

[8] ZOU C, GONG W, TOWSLEY D. Code red worm propagation modeling and analysis[C]//Proc of the 9th ACM Conf on Computer and Comm Security (CCS '02). 2002:138-147.

[9] PASTOR-SATORRAS R, VESPIGNANI A. Epidemic spreading in scale-free networks[J]. Physical Rev Letters,2001,86(14):3200-3203.

[10] MORENO Y, GOMEZ J, PACHECO A F. Epidemic incidence in correlated complex networks[J]. Physical Rev E, 2003, 68(3): 035103.

[11] MORENO Y, SATORRAS R P, VEPIGNANI A. Epidemic outbreaks in complex heterogeneous networks[J]. European Physical J B, 2002,26(4):275-288.

[12] BOGUNA M, PASTOR SATORRAS R, VESPIGNANI A. Epidemic spreading in complex networks with degree correlations[C]//Proc of Lecture Notes in Physics: Statistical Mechanics of Complex Networks. 2003.

[13] WANG C, KNIGHT J C, ELDER M C. On viral propagation and the effect of immunization[C]//Proc of the 16th ACM Ann Computer Applications Conf. [S.l.]: ACM,2000.

[14] WONG C, BIELSKI S, McCUNE J M, *et al*. A study of massmailing worms[C]//Proc of ACM Conf on Computer and Comm Security Workshop Rapid Malcode (WORM '04). 2004.

[15] 王长广,王方伟,张运凯,等. 一种无尺度网络上垃圾邮件蠕虫的传播模型[J]. 计算机科学,2007,34(2):68-70.

[16] NEWMAN M, FORREST S, BALTHROP J. E-mail networks and the spread of computer viruses[J]. Physical Rev E, 2002, 66(3): 035101.

[17] ZOU C C, TOWSLEY D, GONG Wei-bo. Modeling and simulation study of the propagation and defense of internet e-mail worms[J]. IEEE Trans on Dependable and Secure Computing, 2007, 4(4):105-118.