第 36 卷 第 5 期 计 算 机 工 程 2010 年 3 月
Vol.36 No.5 Computer Engineering March 2010
·安全技术· 文章编号：1000—3428(2010)05—0143—03 文献标识码：A 中图分类号：TP393

即时通信蠕虫传播建模

冯朝胜[1,2]，邓 婕[3]，秦志光[2]，刘 霞[1]，劳伦斯·库珀特[4]
(1. 四川师范大学计算机科学学院，成都 610066；2. 电子科技大学计算机科学与工程学院，成都 610054；
3. 四川外语学院成都学院，成都 611731；4. 伦敦大学玛丽皇后学院电子工程系，伦敦 E1 4NS)

摘 要：基于对即时通信蠕虫和即时通信网络特点的分析，使用离散时间方法，提出一个即时通信蠕虫离散数学传播模型。开发即时通信蠕虫传播仿真软件用于验证模型的正确性。基于仿真软件进行的大量仿真实验表明，该蠕虫传播模型是正确的，可以用于分析蠕虫的传播行为并预测传播趋势。

关键词：蠕虫；即时通信；建模

Modeling of Instant Message Worms Propagation

FENG Chao-sheng[1,2], DENG Jie[3], QIN Zhi-guang[2], LIU Xia[1], Laurence Cuthbert[4]
(1. School of Computer Science, Sichuan Normal University, Chengdu 610066;
2. School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 610054;
3. Chengdu Institute, Sichuan International Studies University, Chengdu 611731;
4. Department of Electronic Engineering, Queen Mary College, University of London, London E1 4NS)

【Abstract】 By analyzing the properties of instant message worms and instant networks, a discrete-time math model of Instant Message(IM) propagation worms is proposed. To validate this model, a propagation simulation software is developed. Large-scale simulations validate that this model can be used for analyzing behaviors of worms and foreseeing tendency of worm spread.

【Key words】 worms; Instant Message(IM); modeling

1 概述

随着即时通信(Instant Message, IM)软件的广泛使用，一种利用即时通信服务进行传播的 IM 蠕虫出现并迅速发展，特别是 2005 年后，IM 蠕虫呈现出增长迅速、发生频率增强、攻击范围扩大的特点。IMLogic 统计数字表明：2005 年，即时通信威胁事件[1]约为 2004 年的 17 倍，其中，被确认的 IM 蠕虫攻击为 2 403 个，IM 蠕虫占威胁事件总数的 90%，IM 蠕虫针对即时通信软件 MSN, YIM 和 AIM(或 QQ)的攻击比例分别是 57%, 9%和 34%。另外，CNCERT/CC 的 2005 年网络安全工作报告[2]也表明，IM 蠕虫在 2005 年增长迅速，并将其列为 2005 年恶意代码五大趋势的第 1 位。

鉴于即时通信网络如此严峻的安全形势，一些研究人员已经开展了对 IM 蠕虫的研究工作。文献[3]通过实验发现，即时通信网络是无尺度(scale-free)网络。文献[4]对 IM 蠕虫的特点、危害、攻击方法等作了一个较全面的分析。文献[5]则讨论了即时通信网络的安全保护策略。

迄今为止，少有 IM 蠕虫传播数学模型被提出。然而要研究预防和克制 IM 蠕虫的方法，首要的任务是建立 IM 蠕虫的数学传播模型。本文在对即时通信网络和 IM 蠕虫攻击特点深入分析的基础上提出了 IM 蠕虫的离散数学传播模型。

2 IM 蠕虫与即时通信网络

IM 蠕虫[6]是一种利用即时通信系统和即时通信协议的漏洞或者技术特征进行攻击，并在即时通信网络内传播的网络蠕虫；而网络蠕虫是通过网络传播、无须用户干预、能够独立或者利用文件进行攻击和传播的恶意代码。

为了更深入地理解 IM 蠕虫，有必要将其与网络上最流行、危害最大的主机探测蠕虫(如 CodeRed, Slammer 和 Blaster)进行比较。

IM 蠕虫与主动探测蠕虫的相同点是[4]：

(1)IM 蠕虫通过联系人列表进行传播，与主动探测蠕虫所采用的基于 hit-list 的传播机制类似，两者均有明确的攻击目标，而且传播速度都很快。

(2)IM 蠕虫在设计和实现中借鉴了主动探测蠕虫的很多成熟经验，例如，在蠕虫实体的隐匿和保护、宿主的破坏功能以及对漏洞进行缓冲区溢出攻击等方面。

IM 蠕虫与主动探测蠕虫的不同点是[4]：

(1)主动探测蠕虫拥有探测扫描模块，在攻击渗透之前需要通过扫描发现含有系统漏洞的目标主机；而 IM 蠕虫没有探测扫描的功能模块，它拥有即时通信的联系人列表，并直接以此为攻击的目标。

(2)主动探测蠕虫需要事先进行漏洞扫描以便收集网络

基金项目：国家自然科学基金资助项目(60473090)；国家自然科学基金与英国皇家科学会合作基金资助项目(60711130232)；四川师范大学基金资助重点项目(07ZD018)；四川师范大学基金资助项目(08KYL03)

作者简介：冯朝胜(1971－)，男，副教授、博士研究生，主研方向：网络与信息安全；邓 婕，讲师；秦志光，教授、博士生导师；刘 霞，讲师；劳伦斯·库珀特，教授、博士

收稿日期：2009-08-06 **E-mail：**jamesjiangfeng@163.com

中漏洞主机的信息，然后才能生成 hit-list；而且当该 hit-list 被生成时，其中部分主机可能因为发现扫描而及时修补了漏洞，也可能因为其他原因关机，导致 hit-list 的部分内容失效，从而影响主动探测蠕虫的传播。而 IM 蠕虫以即时通信的在线联系人列表为 hit-list，该 hit-list 具有实时和准确的特点，不会遭遇主动探测蠕虫的问题。

主动探测蠕虫使用 IP 地址在底层的 Internet 网络中传播，而 IM 蠕虫在即时通信应用构建的上层虚拟网络中传播。因此，两者面临的网络拓扑不同，相应的检测机制也不同。

IM 通信网络是指正在运行即时通信软件的主机形成的网络。某个在线用户的在线好友主机成为该用户主机的邻居(即有边相连)。显然，IM 网络是实际网络上的一个覆盖网络。该网络具有如下特点：(1)无尺度网络。这点文献[5]已用实验证明。(2)同构网络。IM 客户端软件通常是相同的。

即时通信网络的这些特点使 IM 蠕虫具有很大的危害性。主要包括：

(1)IM 蠕虫为拓扑蠕虫。IM 网络的特点使 IM 蠕虫无须扫描就能通过邻居表准确地找到攻击对象，这表明 IM 蠕虫是拓扑蠕虫，而文献[7]在 2002 年就指出拓扑蠕虫具有很快的传播速度。

(2)IM 蠕虫具有攻破整个 IM 网络的能力。因为 IM 网络为同构网络，所以如果 IM 蠕虫是基于 IM 客户端软件的某个漏洞编制的，那么它很快就能突破整个网络(所有的客户端软件有相同的漏洞)。

(3)IM 蠕虫使用的目标攻击方式会在极短的时间内使网络崩溃。IM 网络是无尺度网络，根据复杂网络理论，如果 IM 蠕虫成功攻击了邻居数很多的节点，整个网络很快就会崩溃。

(4)IM 蠕虫很难检测。已有的网络蠕虫检测方法通常都是基于异常扫描探测的，而 IM 蠕虫在传播时不会进行扫描探测，所以，它很难被已有的检测软件发现。

3 IM蠕虫传播建模

3.1 建模参数和假设

为简化建模，本文作了如下假设：

(1)即时通信网络的拓扑结构不发生变化。由于 IM 蠕虫传播速度很快，因此这样的假设是合理的。

(2)主机状态转移在一个时间单元(time unit)内完成。

在模型中，主机的状态分成易感的(S)和感染的(I)2 种。模型中用到的变量和参数如下：

$I(t)$：t 个时间单元后感染主机在所有主机中所占比例。

$I_k(t)$：t 个时间单元后度为 k 的感染主机在度为 k 的主机中所占比例。

$P(k)$：邻居数为 k 的主机被选中的概率。

$\tilde{k}$：网络的平均度。

β：易感节点被一个感染邻居节点感染的概率。

γ：感染节点恢复为易感节点的概率。

3.2 IM 蠕虫传播模型分析

由于 IM 蠕虫传播速度很快，在免疫方法找到时它可能已经大范围流行，因此本文建模时不考虑免疫的情况，主机的状态要么为易感状态要么为感染状态。易感主机因为邻居节点的感染很可能成为新的感染主机，而感染主机因为蠕虫被用户删除(尽管很难，但可能性存在)而恢复为易感状态，所以，主机的状态转移为 S→I→S。建模将基于离散时间方法进行，具体如下：

(1)易感主机被邻居节点感染的概率

易感主机的一条边与度为 k 的主机相连的概率是：$kp(k)/\bar{k}$，由于度为 k 的主机由度为 k 的易感主机和度为 k 的感染主机两部分组成，它们的比例分别是 $1-I_k(t)$ 和 $I_k(t)$，因此易感主机的一条边与度为 k 的感染主机相连的概率为 $kp(k)I_k(t)/\bar{k}$，于是，易感主机的一条边与感染主机相连的概率为

$$p_{si}=\sum_k kp(k)I_k(t)/\bar{k}\ ,\ \bar{k}=\sum_m mp(m)$$

所以，度为 m 的主机被邻居主机感染的概率为

$$p_{mi}(t)=\begin{cases}\beta m\sum_k kp(k)I_k(t)/\bar{k} & \text{如果}\beta m\sum_k kp(k)I_k(t)/\bar{k}\leqslant 1\\ 1 & \text{否则}\end{cases}$$

(2)度为 k 的感染主机的变化率

因为一台易感主机在时间单元 $t+1$ 时被感染的概率为 $p_{mi}(t)$，而在第 $t+1$ 个时间单元开始时易感主机数为 $(1-I_k(t))$，所以易感主机从第 t 个时间单元结束到第 t+1 个时间单元结束间共有 $\beta k(1-I_k(t))p_{mi}(t)$ 台变成了感染主机；与此同时恢复为易感主机的感染主机比例为 $\gamma I_k(t)$。所以，在度为 k 的主机中，感染主机比例的变化率为

$$I_k(t+1)-I_k(t)=(1-I_k(t))p_{ki}(t)-\gamma I_k(t)$$

(3)感染主机的变化率

因为度为 k 的感染主机的变化率为

$$I_k(t+1)-I_k(t)=(1-I_k(t))p_{ki}(t)-\gamma I_k(t)$$

所以有

$$I_k(t+1)=(1-I_k(t))p_{ki}(t)-\gamma I_k(t)+I_k(t)$$

$$p(k)I_k(t+1)=(1-I_k(t))p(k)p_{ki}(t)+(1-\gamma)p(k)I_k(t)$$

$$\sum_k p(k)I_k(t+1)=\sum_k(1-I_k(t))p(k)p_{ki}(t)+(1-\gamma)p(k)I_k(t)$$

$$\sum_k p(k)I_k(t+1)=\beta(\sum_k kp(k)-\sum_k kp(k)I_k(t))p_{si}+(1-\gamma)\sum_k p(k)I_k(t)$$

$$I(t+1)-I(t)=\beta(\tilde{k}-\tilde{k}p_{si})p_{si}-\gamma I(t)$$

因此，感染主机的变化率为

$$I(t+1)-I(t)=\beta\tilde{k}(1-p_{si})p_{si}-\gamma I(t)$$

4 仿真验证及分析

4.1 实验说明

考虑到已有文献通过实验证明了即时通信网络是无尺度网络，实验时使用被广泛使用的无尺度网络生成器 Brite 来生成无尺度网络拓扑结构。为了进行蠕虫传播仿真，使用 Java 专门开发了一个仿真软件。仿真软件包括 3 个部分：初始化，协议执行和输出结果。在初始化时，除了需要提供周期数、感染率和恢复率外，还需提供由 Brite 生成的无尺度网络拓扑结构文件和 hit-list，即初始感染节点。执行协议以周期(相当于一个时间单元)为单位进行，在每个周期内，每个节点都将执行 1 次协议(如果该节点是易感的，则依次检查邻居节点是否为感染节点，如果是，再以 β 值确定自己是否被感染；如果该节点是感染的，则以 δ 值确定是否恢复)。为了得出较准确的结果，同样初始条件的实验做 50 次，取 50 次结果的平均值作为实验结果。

4.2 实验结果和分析

图 1、图 2 对不同感染率情况下的仿真值和理论值进行

了对比(T 对应理论值，S 对应仿真值)。图 1 的实验结果是在保持恢复率为 $\delta = 0.01$ 和其他参数不变(初始感染节点和网络拓扑结构)的情况下通过改变感染率得到的，感染率 β 的取值分别为 0.5, 0.7 和 0.9。从图 1 可以看出，在不到 10 个时间单元的情况下，网络中所有节点都被感染。

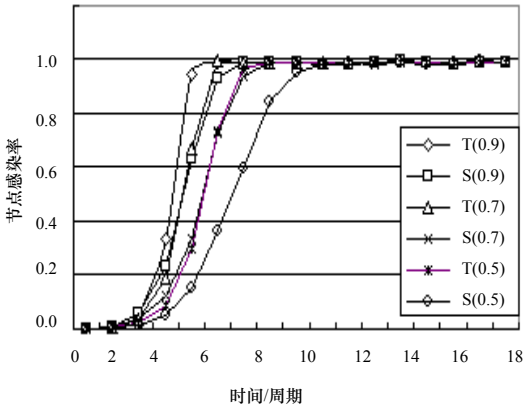

图 1　不同β下的理论值和仿真值(δ=0.01)

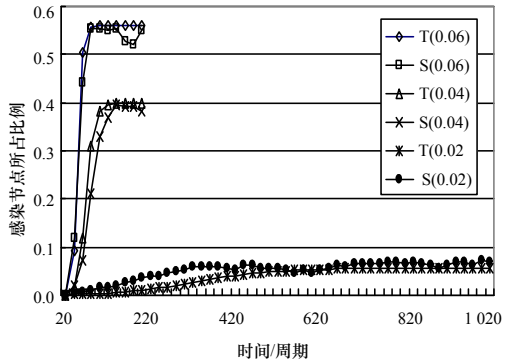

图 2　不同β下的理论值和仿真值(δ=0.01)

图 2 的实验结果是在保持恢复率为 $\delta = 0.1$ 和其他参数不变的情况下通过改变感染率得到的。其中，感染率 β 的取值分别为 0.02, 0.04 和 0.06。实验结果表明，蠕虫同样会像图 1 实验那样流行，但传播速度慢很多，而且在传播处于稳定状态时总有一部分主机处于易感状态(进入稳定状态的时间分别为第 511、第 104 和第 53 个时间单元)，不会出现全都被感染的情况。

与图 1、图 2 相比，图 3 对不同恢复率情况下的理论值与仿真值进行了对比。

实验结果是在保持感染率为 0.1 和其他条件不变情况下通过改变感染率实现的。恢复率的取值分别为 0.05, 0.1 和 0.2。实验结果表明，蠕虫会传播开，最终达到稳定状态。在稳定状态，总有一部分主机处于易感状态。

从图 1~图 3 容易看出，在同样的初始化条件下，基于模型的理论值曲线与仿真值曲线趋势很接近，其他实验有类似结果。理论值与实验值结果匹配的事实充分表明本文反映 IM 蠕虫传播的离散数学模型是正确的。实验还表明，蠕虫的传播能力是由感染率与恢复率的比值决定的，本文称作感染强度，记作 r=感染率/恢复率。

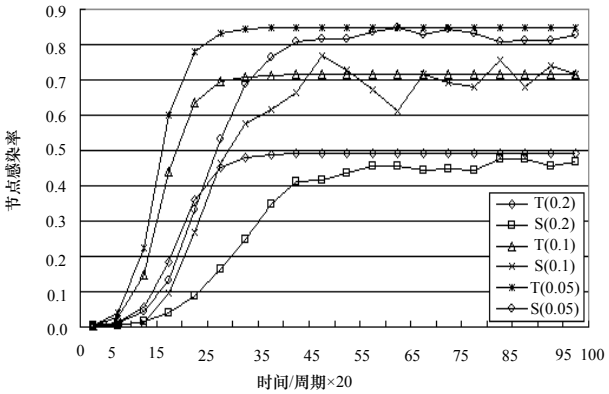

图 3　不同δ下的理论值和仿真值的对比(δ=0.01)

图 1~图 3 实验对应的感染强度如下：

β	0.02	0.04	0.06	0.5	0.7	0.9	δ	0.05	0.10	0.20
r	0.2	0.4	0.6	50.0	70.0	90.0	r	0.5	1.0	2.0

图 1 中蠕虫传播速度比图 2 和图 3 快得多的主要原因是感染强度大得多，其感染强度足以让所有的主机都被感染。

5　结束语

为了有效克制 IM 蠕虫的传播，首先需要研究 IM 蠕虫的传播模型，通过传播模型分析其传播行为并预测其趋势。本文在对 IM 蠕虫和即时通信网络特点研究的基础上提出 IM 蠕虫的离散时间数学传播模型。理论值与仿真的对比结果期表明，提出的模型是正确的，可以用于分析 IM 蠕虫的传播行为和趋势。下一步的研究重点是基于本文模型提出预防和克制 IM 蠕虫的方法和策略。

参考文献

[1] IMlogic Threat Center. 2005 Real-time Communication Security: The Year in Review[Z]. (2005-10-11). http://www.imlogic.com/pdf/2005ThreatCenter_report.pdf.

[2] CN/CERT. CN/CERT China Network Security Research Annual Report[Z]. (2005-09-13). http://www.hais.org.cn/doc/2005CNCERTCCAnnualReport_Chinese.pdf.

[3] Smith R. Instant Messaging as a Scale-free Network[Z]. (2006-05-03). http://arxiv.org/abs/cond-mat/0206378.

[4] 卿斯汉，王　超，何建波，等. 即时通信蠕虫研究与发展[J]. 软件学报, 2006, 17(10): 2118-2130.

[5] 徐向阳，韦昌法. 基于即时通信的安全保护策略[J]. 计算机工程, 2007, 33(21): 125-127.

[6] Mannan M, van Oorschot P C. On Instant Messaging Worms: Analysis and Countermeasures[C]//Proceedings of the ACM CCS Workshop on Rapid Malcode. Fairfax, VA, USA: ACM Press, 2005.

[7] Staniford S, Paxson V, Weaver N. How to Own the Internet in Your Spare Time[C]//Proceedings of the 11th USENIX Security Symposium. San Francisco, CA, USA: [s. n.], 2002: 149-167.

编辑　张正兴