

即时通信蠕虫研究与发展*

卿斯汉^{1,2,3+}, 王超^{1,2,3}, 何建波^{1,3}, 李大治²

¹(中国科学院 软件研究所 信息安全技术工程研究中心,北京 100080)

²(北京中科安胜信息技术有限公司,北京 100086)

³(中国科学院 研究生院,北京 100049)

Research and Development of Instant Messaging Worms

QING Si-Han^{1,2,3+}, WANG Chao^{1,2,3}, HE Jian-Bo^{1,3}, LI Da-Zhi²

¹(Engineering Research Center for Information Security Technology, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

²(Beijing ZhongkeAnsheng Corporation of Information Technology, Beijing 100086, China)

³(Graduate School, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: Phn: +86-10-62635150, E-mail: qsihan@ercist.iscas.ac.cn

Qing SH, Wang C, He JB, Li DZ. Research and development of instant messaging worms. *Journal of Software*, 2006,17(10):2118–2130. <http://www.jos.org.cn/1000-9825/17/2118.htm>

Abstract: With the abroad application of instant messaging and explosive growth of users in current years, IM worms have higher outbreak frequency, wider coverage and deeper hurt, which have been a serious threat to network security. In this paper, the concept and research situation of IM (instant messaging) worms, function component, execution mechanism and comparison with other worms are presented, then the network topology and propagation models are discussed, and finally the critical techniques of IM worm prevention are given. Some major problems and research trends in this area are also addressed.

Key words: network security; IM (instant messaging) worm; Internet worm; network topology; propagation model

摘要: 随着即时通信(instant messaging)应用的日益广泛和用户数量的迅速增加,即时通信蠕虫(IM 蠕虫)的发生频率也相应提高,传播范围变广以及危害程度加深,其真正成为网络安全的重要威胁.首先综合论述IM蠕虫的研究概况;然后剖析IM蠕虫的基本定义、功能结构和工作机理以及IM蠕虫与其他网络蠕虫的区别与联系;讨论IM蠕虫的网络拓扑和传播模型;归纳目前防范IM蠕虫的最新技术;最后给出IM蠕虫研究的若干热点问题与展望.

关键词: 网络安全;即时通信蠕虫;网络蠕虫;网络拓扑;传播模型

中图法分类号: TP393 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant No.60573042 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant No.G1999035802 (国家重点基础研究发展规划(973)); the Beijing Natural Science Foundation of China under Grant No.4052016 (北京市自然科学基金)

Received 2006-04-14; Accepted 2006-06-30

随着互联网应用的深入,网络蠕虫对信息系统安全的威胁日益增加.其中以 CodeRed,Blaster 和 Slammer 等为代表的主动探测蠕虫和以 Melissa,LoveLetter 和 MyDoom 等为代表的 E-mail 蠕虫流行时间长,覆盖面积广,它们给信息系统造成了巨大的危害.因此,针对主动探测蠕虫和 E-mail 蠕虫的研究成为恶意代码领域长时间研究的重点.例如:Nicholas 等人对主动探测蠕虫的快速扫描策略进行了研究^[1],并实现了 Warhol 实验蠕虫,理论推测该蠕虫能在 30 分钟内感染整个互联网^[2].Kephart,White 和 Chess 等人对病毒传播模型进行了研究^[3],在此基础上,邹长春等人以 CodeRed 为例讨论了基于微分方程的双因素传播模型^[4],并构建了 E-mail 蠕虫的传播模型^[5].Dug Song 等人对蠕虫引起的网络流量统计特征进行了研究,力图通过对网络流量异常检测实现对网络蠕虫的防范^[6].

相对而言,作为网络蠕虫一员的 IM 蠕虫由于诞生时间较晚(2001 年 8 月)、数量较少,表现出的危害不如前两类网络蠕虫,因此,IM 蠕虫未得到足够的重视.然而,进入 2005 年之后,IM 蠕虫呈现出迅速增长势头,发生频率增强,攻击范围扩大,其正逐渐引起安全研究人员的关注.2005 年 3 月,几家主要即时通信提供商(包括微软,AOL,Yahoo!)和安全公司(如 Symantec)共同成立了 IM 威胁中心,加强对即时通信漏洞和 IM 蠕虫的跟踪和研究^[7].HP 公司的 Parry 等人提出了一种通过限制通信速率来扼杀 IM 蠕虫的技术,并在实验中证明其有效性^[8].

本文第 1 节阐述 IM 蠕虫的定义、历程和现状.第 2 节将分析 IM 蠕虫的功能结构及其工作机理.第 3 节分析 IM 蠕虫与其他网络蠕虫的区别.第 4 节重点讨论 IM 蠕虫的网络拓扑和传播模型.第 5 节介绍当前 IM 蠕虫的主要对抗技术.最后对 IM 蠕虫的发展趋势进行展望.

1 IM 蠕虫概述

1.1 IM 蠕虫定义

Kienzle 和 Elder 曾对网络蠕虫进行了简单的定义:网络蠕虫是通过网络传播、可以无须用户干预、能够独立或者利用文件进行攻击和传播的恶意代码^[9].IM 蠕虫是网络蠕虫的一种,它与主动探测蠕虫和 E-mail 蠕虫等相并列.Mannan 和 Oorschot 对 IM 蠕虫与其他网络蠕虫进行了区分,将 IM 蠕虫定义为:IM 蠕虫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^[10].根据定义可以看出,即时通信自身的技术特征对 IM 蠕虫有着直接的影响.

即时通信是一种基于 Internet 的网络应用,任意两个即时通信客户端在即时通信服务器的帮助下能够实现方便而快捷的信息通信.其主要技术特征和发展趋势如下所示:

- 即时通信主要包括客户端(用户)和服务器两种实体.客户端通过联系人列表维护与其他客户端的连接关系.图 1 表明即时通信的 3 种主要通信模式:客户端与服务器的通信,主要涉及用户登录和联系人信息获取等; 客户端之间在服务器中介下的通信,主要涉及文件传输、语音视频通信以及 NAT 内部的用户通信等; 客户端之间无须服务器干预下的 P2P 通信,主要涉及文本通信等.
- 即时通信近几年来得到了迅猛发展,用户数量日益庞大.根据行业研究机构 iResearch 的研究报告,全球即时通信帐户在 2004 年为 7.4 亿,2005 年增长到 8.67 亿,预计 2006 年将达到 12 亿.中国即时通信用户在 2004 年为 7 000 万,2005 年增长到 9 300 万,预计 2006 年将增长 29%,达到 1.2 亿人^[11].
- 即时通信数据量巨大.2004 年日平均 72.2 亿条个人信息和 1.12 亿条企业信息被发送,而 2005 年日平均信息总量达到 139 亿条,据预测,在 2006 年即时通信信息总量将首次超过 E-mail^[11].

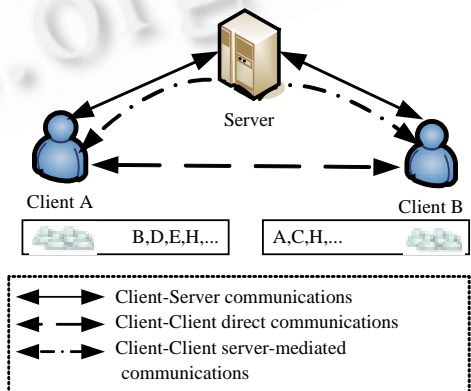


Fig.1 IM communication modes

图1 即时通信主要通信模式

- 即时通信软件的类型集中在:国外应用最为广泛的即时通信软件包括 MSN Messenger, Yahoo! Instant Messenger(YIM)和 AOL Instant Messenger(AIM)这三者;国内则是腾讯 QQ 占据主导,用户份额为 78.8%, MSN 次之,用户份额为 13%^[11].
- 即时通信与其他一些网络应用结合紧密.例如,MSN 的部分主要功能被集成到 Outlook Express 和 Live Communication Server 中,用户容易在不知道的情况下安装并使用了即时通信^[12].

即时通信自身灵活而且多样的工作模式,庞大的用户群、海量的信息量对 IM 蠕虫的工作机制、传播机理和破坏能力等产生直接的影响,使得 IM 蠕虫成为一种新型的与众不同的网络蠕虫.

1.2 IM蠕虫历程与现状

第一个 IM 蠕虫(W32.Funnyfiles.Worm)诞生于 2001 年 8 月,由于当时黑客的注意力主要集中于 CodeRed, Blaster 和 Slammer 等主动探测蠕虫上,IM 蠕虫在 2002 年和 2003 年并没有突出的表现.随后两年,尤其是在进入 2005 年之后,IM 蠕虫开始呈现出迅速增长的趋势.IMLogic 统计数字表明:2005 年,即时通信威胁事件约为 2004 年的 17 倍,其中被确认的 IM 蠕虫攻击为 2 403 个.图 2 表明,IM 蠕虫占威胁事件总数的 90%.图 3 表明了 IM 蠕虫针对不同即时通信软件的攻击比例^[13].另外,CNCERT/CC 的 2005 年网络安全工作报告也表明,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一位^[14].

与此同时,在 2005 年,随着防病毒产品的应用,E-mail 蠕虫开始呈现下降趋势;而主动探测蠕虫虽然继续增长,但也因为 WinXP SP2 中应用了 DEP/NX 技术导致许多常规缓冲区溢出攻击无法实现,主动探测蠕虫已无法达到 Slammer 和 Blaster 等早期蠕虫的传播规模^[14].

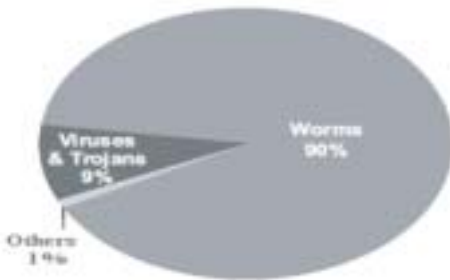


Fig.2 Ratio of IM worms in all threats

图 2 IM 蠕虫在威胁事件中的比例

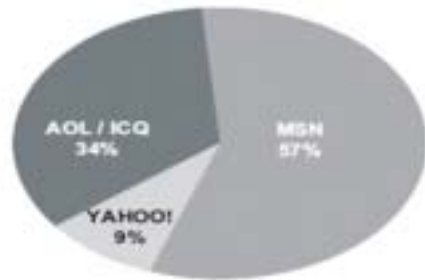


Fig.3 Ratio of IM software attacked

图 3 被攻击的即时通信软件比例

在 2005 年,IM 蠕虫还呈现出种类迅速增多的趋势,尤以表 1 所列出的 IM 蠕虫变种最为明显^[11].

Table 1 Stat of IM worm variants in 2005

表 1 2005 年 IM 蠕虫变种统计

Group	Latest variant	Latest posting	First reposted	Variants	Attack targets
Kelvir	Kelvir-BJ	12/21/2005	2/25/2005	140	MSN, AIM, YIM
Bropia	Bropia-K	12/22/2005	1/19/2005	29	MSN
Opanki	Opanki-W	12/8/2005	5/6/2005	26	AIM
Chode	Chode-Q	12/27/2005	3/18/2005	16	AIM, MSN
Rbot	Rbot-BDV	1/6/2006	1/1/2004	16	AIM

2 IM 蠕虫工作机理

2.1 IM蠕虫的功能结构

Jose Nazario 等人将网络蠕虫按照功能划分成 6 个模块:搜索模块(reconnaissance capabilities)、特殊攻击模块(specific attack capabilities)、命令操作界面模块(command interface)、通信模块(communication capabilities)、智能模块(intelligence capabilities)和非攻击使用模块(unused attack capabilities)^[15].该框架主要是对网络蠕虫的

预测,难以准确表达当前 IM 蠕虫的功能结构.在文献[6,15,16]的基础上,我们结合 IM 蠕虫实例,归纳分析出 IM 蠕虫分为主体功能模块和辅助功能模块.主体功能模块负责完成攻击和传播的主体流程,而辅助功能模块则赋予 IM 蠕虫更强的生存力和破坏力.IM 蠕虫功能结构如图 4 所示.

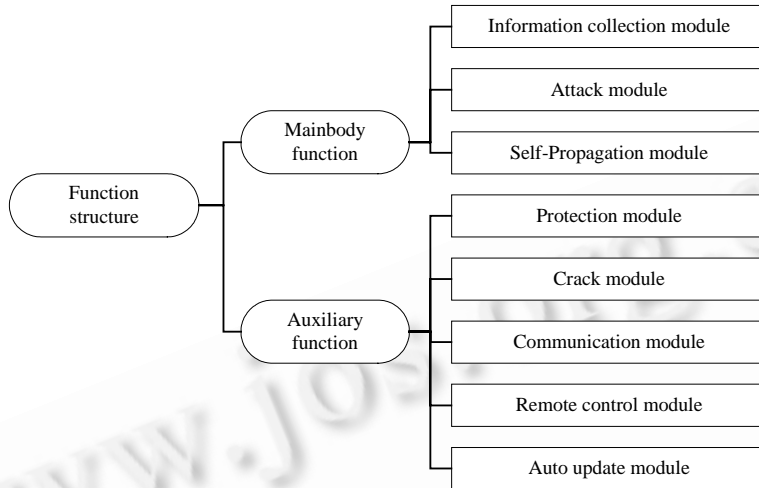


Fig.4 Function structure of IM worms

图 4 IM 蠕虫功能结构

2.1.1 主体功能模块

IM 蠕虫的主体功能模块由 3 部分构成: 信息搜集模块,该模块决定采用何种机制搜集即时通信中的信息,包括本机系统信息和联系人信息等,这些信息可以单独使用或被其他蠕虫个体共享; 攻击渗透模块,该模块决定采用何种机制向联系人发起攻击并建立传播途径,该模块在攻击方法是开放的、可扩充的; 传播推进模块,该模块决定生成何种形态的 IM 蠕虫副本,以及采用何种方式向联系人传递蠕虫副本.相关工作机理将在第 2.2 节中详细分析给出.

2.1.2 辅助功能模块

辅助功能模块是对除主体功能模块以外的其他模块的归纳和预测,主要由 5 部分构成: 实体保护模块,包括 IM 蠕虫对实体组成部分的隐藏、变形和加密等,以及关闭防病毒软件、阻塞系统访问安全站点等自我保护机制,主要目的是提高 IM 蠕虫的生存能力; 宿主破坏模块,该模块用于构建僵尸网络、安装后门、窃取信息以及执行 DOS 攻击等; 信息通信模块,该模块能使 IM 蠕虫之间、IM 蠕虫与黑客之间实现信息交流.利用通信模块,IM 蠕虫之间可以共享信息,IM 蠕虫编写者能够更好地了解 IM 蠕虫的工作状态; 远程控制模块,利用控制模块编写者能够下达新的指令,调整 IM 蠕虫的行为,控制被感染主机; 自动升级模块,该模块可以使编写者随时更新 IM 蠕虫其他模块的功能,从而实现不同的攻击目的.

2.2 IM蠕虫的工作机理

IM 蠕虫利用即时通信网络进行传播,首先要收集感染主机用户的联系人信息,然后利用即时通信的协议或者漏洞向联系人实施攻击,并将生成的 IM 蠕虫副本传播给被攻击的联系人.IM 蠕虫在信息搜集、攻击渗透和传播推进 3 个阶段中所主要采用的工作机制归纳如下.

2.2.1 信息搜集机制

(1) 基于即时通信 API 的联系人搜集

基于即时通信 API 的联系人搜集机制是指 IM 蠕虫通过调用即时通信软件制造商发布的 API 函数接口,直接获取被感染主机上用户的联系人列表.例如:MSN 和 ICQ 均提供了详细的编程接口和文档^[17,18],以便能够更好地被集成到其他网络应用中去.IM 蠕虫可以利用这些 API 直接获取用户的联系人信息,Choke 蠕虫^[19]就是

典型的利用 MSN 的 MessengerContacts.Item 接口搜集联系人信息的 IM 蠕虫.

(2) 模拟用户操作行为的联系人搜集

模拟用户操作行为的联系人搜集机制是指 IM 蠕虫通过调用操作系统的底层接口获取被感染主机上即时通信软件的窗体句柄,然后模拟用户的鼠标或者键盘操作,实现联系人列表的定位和信息读取.这种机制主要针对没有提供即时通信 API 的即时通信软件,例如,YIM 就因为安全保护等方面的因素没有提供 API 编程接口.Goner 蠕虫^[20]是这种 IM 蠕虫的一个代表,它利用 Windows 的 FindWindow 接口获取 YIM 窗体的句柄,然后使用 Windows 窗体消息定位到 YIM 的联系人列表上.

(3) 基于钩子的联系人搜集

基于钩子的联系人搜集机制是指 IM 蠕虫通过篡改或者修补被感染主机上即时通信软件中的某些动态链接库,在其中嵌入钩子函数,从而在用户与服务器通信的过程中获取联系人的信息.AimVen 蠕虫^[21]就是一种典型的案例,该 IM 蠕虫能够对 AIM 的 Icbmft.ocm 文件打补丁来嵌入钩子函数,从而在通信过程中获取其他联系人的信息.

(4) 基于注册表的通信语言搜集

基于注册表的通信语言搜集机制是指 IM 蠕虫通过访问被感染主机的注册表获取操作系统的语言类型,并以此确定发送攻击信息时所使用的语言.以用户习惯的语言发送攻击信息,能够降低联系人对接收信息的安全警惕意识,从而提高攻击的成功比率.Kelvir.H 蠕虫^[22]就能够利用注册表获取 Windows 操作系统的语言类型,并以此从预备的 10 种语言中选择一种语言发送攻击信息.

2.2.2 攻击渗透机制

(1) 主动攻击

主动攻击机制是指 IM 蠕虫在搜集到被感染主机上的联系人信息之后,无论当前用户是否正与其他联系人进行通信,都主动向其他联系人发送攻击信息.这种攻击机制具有攻击速度快的特点,但是其存在易产生异常数据报文而被检测的不足,该机制正在被大多数 IM 蠕虫所采用.例如,Choke 蠕虫^[19]、SoFunny 蠕虫^[23]、Kelvir 蠕虫^[22]等均采用主动攻击机制实施攻击渗透.

(2) 被动攻击

被动攻击机制下,IM 蠕虫并不主动向联系人实施攻击,而是依赖用户的正常活动向正与之通信的联系人发动攻击.该攻击机制需要用户行为的触发,因此,其速度比主动攻击机制要慢,但该机制下 IM 蠕虫的攻击不容易引起联系人的怀疑,攻击成功率要高;而且不会产生异常通信,IM 蠕虫自身的安全性要高.AimVen 蠕虫^[21]就是在用户与联系人之间传递含有 .exe 后缀的可执行文件时,篡改或者替换该可执行文件实施攻击.

2.2.3 传播推进机制

(1) URL 推进

URL 推进机制是指 IM 蠕虫以文本格式向联系人发送包含链接 IM 蠕虫副本的 URL 信息,利用社会信任关系诱骗联系人下载并执行该 URL 所链接的 IM 蠕虫副本伪装文件,从而达到向远程主机传播推进的目的.发送 URL 文本具有传递信息体积小的优点,而且能够躲避多数防病毒软件实时监测二进制文件的防范机制.URL 可以指向某个具体的站点,也可以指向 IM 蠕虫在感染主机上建立的 Web 服务的地址.Aplore 蠕虫^[24]在感染主机的 8180 端口开启 Web 服务,然后向所有的联系人发送指向该感染主机 Web 服务的 URL 信息,诱使联系人从该 URL 下载伪装成 Web 浏览器插件的蠕虫副本,联系人一旦执行将立刻被该 IM 蠕虫感染.

(2) 用户干预下的文件推进

用户干预下的文件推进机制是指 IM 蠕虫直接向联系人发送包含蠕虫副本的文件,同样利用社会信任关系诱骗联系人接收并执行该文件,从而达到传播推进的目的.该推进机制需要被攻击联系人的参与,如果联系人拥有足够的安全防范意识而不执行接收到的蠕虫副本文件,此次传播就以失败结束.IM 蠕虫多数会利用加密或者捆绑等技术,对蠕虫副本文件进行伪装,提高其躲避防病毒软件实时监测的能力.Serflog 蠕虫^[25]的若干变种就能够利用 MSN 直接向联系人传输经过加壳伪装的蠕虫副本文件.

(3) 非用户干预下的文件推进

非用户干预下的文件推进机制是指 IM 蠕虫通过攻击即时通信软件客户端的某些缓冲区溢出漏洞,在远程主机上执行 Shellcode,并以此建立传输通道来直接传递蠕虫副本,从而在没有联系人干预的情况下在远端顺利执行该蠕虫副本。2005 年,ICQ 的 sound scheme file location 漏洞、MSN 的 PNG 漏洞以及 GIF 漏洞均被证实可以让 IM 蠕虫在没有用户干预的情况下传输和执行文件。Bizex 蠕虫^[26]就是一种典型的不依赖用户干预而实施传播的 IM 蠕虫。

(4) 基于操作系统漏洞的文件推进

基于操作系统漏洞的文件推进机制同样不需要用户的干预,它是指 IM 蠕虫通过攻击目标联系人所在主机的操作系统漏洞,然后执行 Shellcode 并建立通道传递蠕虫副本,从而达到在没有联系人干预的情况下在远端顺利执行该蠕虫副本的目的。Bropia.F 蠕虫^[27]就可以利用 Windows 操作系统一些已知的漏洞(包括 SQL、IIS 和 WebDav 等著名的缓冲区溢出漏洞)实施攻击并传输蠕虫副本。

3 IM 蠕虫与其他网络蠕虫的对比

在对 IM 蠕虫功能结构和工作机制分析的基础上,我们归纳总结出 IM 蠕虫与主动探测蠕虫和 E-mail 蠕虫之间的主要联系与区别。

3.1 联系

IM 蠕虫与主动探测蠕虫和 E-mail 蠕虫都属于网络蠕虫的范畴,IM 蠕虫与这两种网络蠕虫之间存在一些相同和相似之处,主要联系有以下一些:

- 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^[1]相类似,两者均拥有明确的攻击目标,而且传播速度都很快;
- 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方面、宿主的破坏功能方面以及对漏洞进行缓冲区溢出攻击等方面;
- IM 蠕虫和 E-mail 蠕虫两者均不必发送探测数据报来实施对攻击目标的探测,均直接采用联系人列表为明确的攻击目标;
- IM 蠕虫和 E-mail 蠕虫两者均以发送文件和 URL 文本为实施传播的主要方式,感染成功率都会受到用户安全意识和社会信任关系等方面因素的影响。

3.2 区别

即时通信与普通网络通信以及 E-mail 通信之间存在着巨大的技术差异,导致 IM 蠕虫与主动探测蠕虫和 E-mail 蠕虫之间存在一些显著的区别:

- 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。
- 主动探测蠕虫需要事先进行漏洞扫描以便收集网络中漏洞主机的信息,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,从而会影响主动探测蠕虫的传播^[1]。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,两者所面临的网络拓扑不同、背景数据噪声不同,相应的检测机制也会不同。
- E-mail 蠕虫同样拥有地址簿中的联系人信息,并以此作为 hit-list,但是该 hit-list 无法保障每个联系人均处于在线状态,因此,该 hit-list 也被称为 offline hit-list^[5]。相对而言,IM 蠕虫的 hit-list 被称为 online hit-list^[10]。
- E-mail 协议的开放性使得不同 E-mail 软件能够进行有效的通信,操作系统的差异并不影响电子邮件的

传递,因此,必然存在适于 Windows 的 E-mail 蠕虫传递到 Linux 系统上产生失效的情况;而即时通信协议间存在壁垒,运行同一种即时通信软件的主机的操作系统基本相同,例如,使用 MSN 的用户通常都安装 Windows 操作系统,因此,IM 蠕虫攻击成功率比 E-mail 蠕虫要高^[10].

- E-mail 蠕虫的传播速度受到被攻击者收取邮件时间的影响,而 IM 蠕虫攻击的联系人一直处于在线状态,在被攻击者安全意识相同的情况下,IM 蠕虫被接收和执行的时间间隔要小于 E-mail 蠕虫^[10].
- E-mail 通信多采用传统的 C/S 架构,而即时通信则融入了 P2P 技术,其通信协议复杂度进一步提升,因此,IM 蠕虫的复杂度要大于 E-mail 蠕虫,抑制 IM 蠕虫的难度要大于抑制 E-mail 蠕虫的难度.

4 IM 蠕虫的若干理论研究

针对主动探测蠕虫和 E-mail 蠕虫的理论研究时间较长,在传播模型等方面已经取得了若干成果,例如,SEM 模型^[28]、KM 模型^[29]、SIS 模型^[30]以及 Two-Factor 模型^[4]等.然而,IM 蠕虫的理论研究相对较晚,至今仍未形成系统化的研究.本文将最近若干重要的理论研究归纳整理如下.

4.1 网络拓扑

即时通信软件之间通过联系人列表,在实际的 Internet 网络拓扑之上构建一层虚拟的网络拓扑,对该虚拟网络拓扑结构的研究有助于对 IM 蠕虫传播速度和趋势的分析.

Erdos 和 Renyi 在 1959 年构建了 ER 随机图模型,并以此对通信和生命科学中常见的复杂网络进行分析.在该模型中,节点间通过随机生成的连接建立 ER 网络,并且大多数节点所拥有的连接数是基本相同的.在 20 世纪 90 年代,更多的网络实验数据表明:复杂现实网络中的节点连接数并不是随机分布的,而是遵循 power-law 规则^[31].Barabasi 和 Bonabeau 提出了无尺度(scale-free,简称 SF)模型描述复杂网络的拓扑结构.在 SF 模型中,随着新节点的不断加入,网络能够持续地增长;而且新加入的节点优先与拥有高连接数的节点相连接.SF 模型一方面揭示了复杂网络对随机异常情况具有较好的容错能力,另一方面也表明复杂网络中病毒和蠕虫能够得到持续而广泛的传播^[32].

在现实社会中的某个人拥有一些朋友,而这些朋友之间很有可能相互也是朋友,因此,即时通信用户的联系人列表同样存在这种社会因素的影响.例如:用户 A 的联系人列表中含有用户 B、C 和 D,而 B 的联系列表中很有可能也含有 C 和/或 D.在此基础上,Smith 对一个实际的 Jabber 即时通信网络进行研究与统计,并将统计结果与随机图模型和 SF 模型进行对比分析,提出即时通信所构建的虚拟网络是一个 SF 网络,同时还证明该虚拟网络遵循 power-law 规则.实验中部分数据见表 2,详细的推理过程请参见文献[33].

Table 2 Statistics of jabber instant messaging network

表 2 Jabber 即时通信网络统计数据

Types	In degree	Out degree	Diameter	Un-Strong connection (%)	Strong connection (%)
Jabber IM network	9.1	8.2	4.35	99	89
Random graph	9.6	9.6	4.79	-	-

即时通信网络被证明是 SF 网络表明:IM 蠕虫在该网络拓扑下能够实现持续而广泛的传播,检测和对抗 IM 蠕虫的难度将会很大.

4.2 SF 网络下的 SIS 传播模型

相对于为主动探测蠕虫所构建的传播模型(例如,SEM 模型、KM 模型以及 Two-Factor 模型等),有效的 IM 蠕虫传播模型尚未完整建立.Pastor-Satorras 和 Vespignani 曾经利用 SIS 模型对 SF 网络下蠕虫的传播趋势进行分析^[34];Mannan 和 Ooorschot,Williamson 和 Parry 等人认为该研究能够近似反映 IM 蠕虫的传播趋势^[10,35],并在此基础上各自开展 IM 蠕虫的相关研究.

在标准 SIS 模型中,主机节点保持两种状态:易感染(susceptible)和被感染(infectious).模型假定状态转变过程为:易感染→被感染→易感染^[30].SIS 模型的微分方程表达式为

$$\begin{cases} dI(t)/dt = \beta I(t)S(t) - \gamma I(t) \\ dS(t)/dt = \beta I(t)S(t) - \gamma I(t) \end{cases} \quad (1)$$

在公式(1)中: $I(t)$ 表示时刻 t 已被感染的主机数; $S(t)$ 表示时刻 t 易感染的主机数;而 β 表示主机感染率; γ 表示主机修补率.公式(1)的详细介绍请参见文献[30].SIS 模型设定有效传播率 $\lambda=\beta/\gamma$.在本地连接或者随机图连接下,存在着一个非零的传播阈值 λ_c .当 $\lambda \geq \lambda_c$ 时,被感染主机会产生扩散并能够最终保持在一个平稳的状态;而当 $\lambda < \lambda_c$ 时,被感染主机将呈指数形式减少并最终趋于 0.

Pastor-Satorras 和 Vespignani 对 1996 年 2 月~2000 年 3 月 Virus Bulletin 公布的病毒和蠕虫信息进行统计,推导出 SF 网络下 SIS 模型中传播阈值 λ_c 的表达式为^[31]

$$\lambda_c = \langle k \rangle / \langle k^2 \rangle; \lambda_c \rightarrow 0 \quad (2)$$

在公式(2)中: $\langle * \rangle$ 表示取*的平均值; k 表示节点的拓扑度(连接数).由于 SF 网络规模可以无限增长,可以得到 $\langle k^2 \rangle \rightarrow \infty$,因此有 $\lambda_c \rightarrow 0$.传播阈值 λ_c 趋向于 0 表明:在 SF 网络中,即使蠕虫的感染率很低,其仍然能够得到广泛传播.Pastor-Satorras 和 Vespignani 还就 SF 网络下 SIS 模型得出如下推论:

$$P_k = k \lambda \theta(\lambda) / (1 + k \lambda \theta(\lambda)) \quad (3)$$

在公式(3)中: P_k 表示一个含有 k 个连接的节点被感染的概率; $\theta(\lambda)$ 表示任意一条连接指向被感染节点的概率.公式(3)表明:节点的连接数越高,其被感染的几率也就越高.Pastor-Satorras 和 Vespignani 结合 SF 网络中节点含有 k 个连接的概率 $P(k)=2m^2/k^{-3}$,进一步得出节点被感染平均概率表达式为

$$P \approx 2e^{-1/m\lambda} \quad (4)$$

Pastor-Satorras 和 Vespignani 的研究对 SF 网络下蠕虫传播得出一些重要的结论.但是,该研究也存在一些不足: SIS 模型本身并没有考虑被感染主机对蠕虫免疫的情况,因此,Cliff Zou 等人认为,SIS 模型难以准确反映蠕虫的传播行为^[5]; 该研究所统计的蠕虫数据中没有包括 IM 蠕虫,由于 IM 蠕虫与其他网络蠕虫之间存在显著的技术差异,因此不能证明该研究成果是否适用于对 IM 蠕虫传播行为的描述.

4.3 IMWDP(IM worms discrete propagation)模型

IMWDP 模型没有采用连续微分方程,而是采用离散递归方程描述 IM 蠕虫在即时通信网络内的传播趋势.我们在建模的过程中,首先确认即时通信网络为 SF 网络,然后重点考虑以下两个方面的因素: 即时通信网络下用户的联系人数量(即节点的连接拓扑度)对 IM 蠕虫传播的影响; 用户干预(安全意识以及社会信任关系)对 IM 蠕虫感染几率的影响.为了研究问题方便,IMWDP 模型假设: IM 蠕虫在感染节点上实施对所有联系人的攻击和传播,以及联系人接收到蠕虫副本的处理工作都在一个标准时间内完成; 已感染 IM 蠕虫的主机不会再被同样的 IM 蠕虫再次感染.

在标准时刻内,新感染 IM 蠕虫的主机将依据联系人列表向联系人发动攻击,每个 IM 蠕虫攻击目标的数量是由感染主机上用户所拥有的联系人数量决定的.IM 蠕虫攻击的目标将会存在两种情况:一种是尚未感染 IM 蠕虫的主机;另一种为已经感染 IM 蠕虫的主机.而接收到 IM 蠕虫副本的用户依据社会信任关系以及安全防范意识决定是否打开执行 IM 蠕虫副本.IMWDP 模型的离散递归方程如下所示:

$$\begin{cases} R(i+1) = (N - A(i))(1 - (1 - 1/N)^{\Phi(E(i))}) \\ \Phi(E(i)) = \sum P(n_j), n_j \in E(i) \\ E(i+1) = \Gamma(R(i+1)) = \sum T(n_j), n_j \in R(i+1) \\ A(i+1) = A(i) + E(i+1) \\ E(0) = A(0) = A_0 \end{cases} \quad (5)$$

在公式(5)中: N 表示在线用户(主机节点)总数; $E(i)$ 表示在时刻 i 新增加的 IM 蠕虫感染主机数; $A(i)$ 为时刻 i 已感染 IM 蠕虫的主机总数; $P(n)$ 表示节点 n 的联系人数量,节点 n 拥有 d 个联系人的概率遵循 power-law 规则,即 $F(d) \propto d^{-\alpha}$; $\Phi(*)$ 表示*中的所有节点的联系人总和,因此, $R(i+1)$ 表示时刻 i 新接收到 IM 蠕虫的主机总数; $T(n)$ 表示节点 n 的用户打开 IM 蠕虫副本导致该节点被感染的概率,该概率遵循高斯分布 $P \sim N(\mu_p, \delta_p^2)$; $\Gamma(*)$ 表示*中

的节点因打开蠕虫副本而被感染的主机总数.

IMWDP 模型下 IM 蠕虫的仿真传播趋势如图 5 所示.仿真实验取节点数 $N=100000$,每个用户的平均联系人数量为 8,power-law 规则中指数 $a=1.7$,用户打开执行 IM 蠕虫概率的高斯分布为 $P\sim(0.5,0.32)$,初始发布 IM 蠕虫数量为 1.图 5 表明,IM 蠕虫在即时通信网络中能够快速增长.

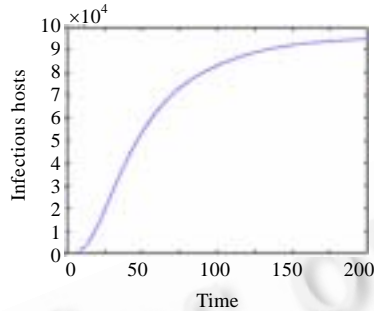


Fig.5 IM worms propagation in IMWDP model

图 5 IMWDP 模型 IM 蠕虫传播趋势

IMWDP 模型的缺点在于其尚未考虑用户打补丁修补漏洞以及安全软件查杀 IM 蠕虫的行为,因此它只能反映出 IM 蠕虫爆发初级阶段的传播趋势.

4.4 传播速度

Symantec 公司的研究人员通过仿真实验对 IM 蠕虫的传播速度进行了初步分析.仿真实验对 IM 蠕虫感染 500 000 台主机的时间进行采样,实验表明,IM 蠕虫感染时间随着平均联系人数量的增加和发送时间间隔的缩短而不断减少^[12].具体采样数据见表 3.

Table 3 Time of IM worm to Infect 500 000 machines

表 3 IM 蠕虫感染 500 000 台主机时间统计

Unique buddies	Time to send (s)					
	0.5	0.75	1	1.5	2	2.5
1.2	31	47	63	94	126	157
1.5	15	22	30	45	60	75
2	9	13	18	27	36	45
2.5	7	10	14	21	28	35
3	6	9	12	18	24	30

实验显示,IM 蠕虫通过在线联系人进行传播的速度非常快.相对于 CodeRed II 通过选择性随机探测在 14 小时内感染 350 000 台主机、Slammer 蠕虫在 15 分钟内感染了 80 000 主机的速度,IM 蠕虫能够在 30s 内感染 500 000 台主机的速度显然要快很多.该仿真实验从另一个方面证明了 IM 蠕虫是一种 Flash 蠕虫^[1].

5 IM 蠕虫对抗技术

IM 蠕虫自身所具有的快速性、隐蔽性和复杂性,不仅使得防火墙和防病毒软件等传统的安全防范技术难以应对 IM 蠕虫的挑战,而且直接利用主动探测蠕虫和 E-mail 蠕虫的对抗技术对抗 IM 蠕虫也无法保证其有效性.对抗 IM 蠕虫需要多种技术的综合应用,包括 IM 蠕虫监测与预警、IM 蠕虫传播抑制和阻断、漏洞自动修复等.本节重点讨论主要的 IM 蠕虫传播抑制技术.

5.1 关闭服务器

服务器过滤技术是一种能够有效抑制 E-mail 蠕虫传播的技术,而 IM 蠕虫所传递的 URL 文本和蠕虫副本文件通常并不经过服务器,因此,服务器过滤技术并不适合抑制 IM 蠕虫的传播.虽然即时通信客户端可以直接向联系人发送 URL 文本,但是客户端在很多重要过程(包括登录和获取联系人信息等)都无法离开服务器,因

此,IM 蠕虫同样无法脱离服务器单独运行.典型的关闭服务器抑制 IM 蠕虫传播的技术为:即时通信的服务提供商暂时关闭服务器,在关闭服务器的时间内对 IM 蠕虫进行分析,生成客户端补丁程序,然后开启服务器,并在用户再次登录的时候强制用户升级客户端程序.2005 年 4 月,Reuters Messaging 就曾经采用该技术关闭服务器 3 个小时以上来对抗 IM 蠕虫的传播^[36].

关闭服务器能够有效阻断 IM 蠕虫的传播,但是这样做最大的缺陷在于会导致正常即时通信的中断.

5.2 切断高连接用户

基于对即时通信网络 SF 拓扑的分析,Smith 提出一种通过切断高连接用户抑制 IM 蠕虫传播的技术:切断联系人数量高的用户与服务器之间的连接,使其无法进行即时通信,以此增加即时通信网络的网络直径,从而减缓 IM 蠕虫的传播速度,为 IM 蠕虫分析和发布补丁赢得时间.Smith 通过研究发现:切断联系人数量最高的前 10%的用户,能够将余下用户所构成的即时通信网络直径增加两倍,从而有效减慢 IM 蠕虫的传播步伐^[33].

切断高连接用户技术能够起到延缓 IM 蠕虫传播的作用,但是该技术仍会给一些用户带来严重的影响,而且它要求服务器对每个用户的联系人数量进行排序,因而增加了服务器的工作负荷.

5.3 IM蠕虫扼杀技术

Williamson 和 Parry 提出了一种针对快速传播蠕虫的通用扼杀技术^[37],并将其应用到抑制 IM 蠕虫的传播中.扼杀技术主要通过区分正常用户和 IM 蠕虫在通信速率上的差异,对异常的 IM 蠕虫通信进行限制,从而抑制 IM 蠕虫的传播,其主要组成和流程如图 6 所示.首先由服务器为用户维护一张联系人的历史记录(working set).当用户试图向某个联系人发送信息时,当服务器判断该联系人在联系人历史记录中时,该消息可以被立即发送;相反情况下,该消息将被放置到延迟队列(delay queue)中,留待以后发送.服务器按照设定的速率处理延迟队列中尚未发送的信息,并更新历史记录和延迟队列.当延迟队列超过设定的警戒阈值时,系统认为存在 IM 蠕虫的传播企图,所有该用户发出的通信请求都将被阻塞;而且系统还将询问该用户是否了解这些信息的传输^[38].

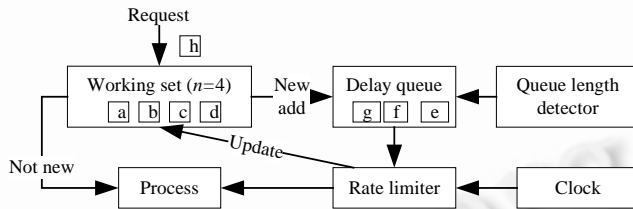


Fig.6 Throttling mechanism for IM worms

图 6 IM 蠕虫扼杀机制

Williamson 和 Parry 在 HP 公司的一个拥有 710 个用户、日发送信息约为 552 条的 Jabber 即时通信系统中对扼杀技术进行了实验.实验设定的历史记录长度为 5,延迟队列长度为 2,更新速率为每分钟一个新联系人,在实验中,IM 蠕虫的传播得到有效抑制^[38].但是,IM 蠕虫扼杀技术存在若干缺陷: 不同用户的通信频率和方式有所不同,进行统一的设定势必影响部分用户的正常通信;而为不同用户进行不同的设定将会增加服务器的工作负荷; 对于依赖用户通信行为发起攻击的 IM 蠕虫,扼杀技术将失效; IM 蠕虫扼杀技术对用户间正常的群组即时通信产生影响,因为群组通信时联系人数量将会超过常规通信时联系人数量,这对历史记录的长度设定构成新的挑战; 扼杀实验中,即时通信用户规模和信息数量相对于现实社会中主要的公用即时通信系统来讲要小很多,其有效性尚不能完全得到验证.

5.4 改进的IM蠕虫扼杀技术

Mannan 和 Oorschot 对 Eyeball Chat 即时通信系统从 2001 年 9 月~2005 年 4 月的统计数据进行分析,发现总计有 17 840 000 个文件传播请求和 3 200 000 000 个文本信息,人均每天的文件传输为 1.84 个,而相应的文本传输为 334.03 个,文件传输只占文本传输的 0.55%.针对即时通信中文件和文本传输数量上的巨大差异,Mannan

和 Oorschot 对 IM 蠕虫扼杀技术进行了改进:将扼杀点专注于文件和 URL 文本之上,而不是对所有通信的传输请求进行抑制.在改进后的 IM 蠕虫扼杀技术中,普通的文本传输将不再被延迟而直接放行,而 URL 文本和文件传输请求将按照相同的扼杀机理判断是否需要延迟发送、是否属于 IM 蠕虫的传播行为、以及是否需要向用户发出询问^[10].

改进后的 IM 蠕虫扼杀技术进一步提高了原有 IM 蠕虫扼杀技术的准确性和运行效率,然而它同样面临着参数设定以及被动攻击等方面的问题.

5.5 基于CAPTCHA的IM蠕虫扼杀技术

CAPTCHA(completely automated public turing test to tell computers and humans apart)技术^[35]是一项采用挑战应答机制验证请求者是自然人还是电脑的技术,它已经被广泛应用到网站(例如 Yahoo!)和电子邮箱(例如 163)的注册环节.基于 CAPTCHA 的 IM 蠕虫扼杀技术描述如下:当客户端试图发送文件或者 URL 文本时,服务器将产生一个如图 7 所示的挑战信息,并将其发给客户端;对于用户而言,正确回答挑战信息十分容易,而对于 IM 蠕虫则十分困难,从而达到抑制 IM 蠕虫通过文件和 URL 自动传播的目的.

Answer the following challenge:



Fig.7 CAPTCHA's challenge

图7 CAPTCHA的挑战信息

对于用户而言,正确回答挑战信息十分容易,而对于 IM 蠕虫则十分困难,从而达到抑制 IM 蠕虫通过文件和 URL 自动传播的目的.

基于 CAPTCHA 的 IM 蠕虫扼杀技术对主动攻击 IM 蠕虫的传播能够起到有效抑制的作用,然而它也增加了频繁发送文件用户的负担,而且对于依附用户行为攻击的 IM 蠕虫仍然难以起到抑制作用.

5.6 其他

除此之外,Symantec 公司提出一套安全强化的即时通信协议,能够防止即时通信软件被某些 IM 蠕虫滥用,详情请参考文献[38].目前针对 IM 蠕虫的对抗技术研究仍然很少,还需要更多的投入.

6 结束语

通过上述分析,我们可以对 IM 蠕虫得出如下结论: IM 蠕虫正在步入快速发展阶段,它已经对网络安全构成新的严重威胁,正在成为网络蠕虫的一个重要分支; IM 蠕虫依托于即时通信应用进行传播,它在传播机理、攻击方法和推进机制等方面与其他类型的网络蠕虫存在较大差别; 设计良好的 IM 蠕虫具有传播速度快和感染范围广的基本技术特点; 现有的网络蠕虫防范技术难以有效抑制 IM 蠕虫的传播.

除此之外,IM 蠕虫还呈现出如下发展趋势: IM 蠕虫攻击的次数迅速增加,攻击频率加快,变种也在不断增加; IM 蠕虫的攻击手法日趋完善和丰富,一方面其利用社会工程知识进行攻击的方法变得更加复杂,另一方面它正逐渐利用系统漏洞进行无用户干预的攻击.因此有理由相信:IM 蠕虫将成为未来几年内网络蠕虫的主要发展方向之一,势必对网络安全提出新的挑战.

鉴于 IM 蠕虫所呈现出的技术特点和发展趋势,恶意代码研究人员必须加强对 IM 蠕虫技术的研究,才能尽早地将 IM 蠕虫扼杀在发展阶段,避免其造成如主动探测蠕虫和 E-mail 蠕虫那样严重的破坏结果.在未来一段时间内,IM 蠕虫研究的重点将主要包括: 加强对 IM 蠕虫自身传播机理的研究,弥补现有 IM 蠕虫传播模型的不足,模拟出 IM 蠕虫的仿真传播趋势; 加强对 IM 蠕虫攻击和推进机理的跟踪,探讨新型的攻击方法和推进技术; 增强对 IM 蠕虫预警和对抗技术的研究,以及讨论如何对 IM 蠕虫进行追踪和取证.总之,对抗 IM 蠕虫将是一个长期的过程,既要掌握当前 IM 蠕虫的工作机理,又要加强对以后 IM 蠕虫发展趋势的研究,做到防患于未然.

致谢 中科安胜公司的蒙杨和李晓东为本项研究提供了诸多 IM 蠕虫样本,在此表示感谢.

References:

- [1] Staniford S, Paxson V, Weaver N. How to own the Internet in your spare time. In: Boneh D, ed. Proc. of the 11th Usenix Security Symp. San Francisco, 2002. <http://www.icir.org/vern/papers/cdc-usenix-sec02/cdc.pdf>
- [2] Weaver N. Warhol worms: The potential for very fast Internet plagues. 2002. <http://www.cs.berkeley.edu/~nweaver/warhol.html>

- [3] Kephart JO, Chess DM, White SR. Computers and epidemiology. <http://www.research.ibm.com/antivirus/SciPapers/Kephart/Spectrum/Spectrum.html>
- [4] Zou CC, Gong W, Towsley D. Code red worm propagation modeling and analysis. In: Proc. of the 9th ACM Symp. on Computer and Communication Security. Washington, 2002. 138–147. <http://tennis.ecs.umass.edu/~czou/research/codered.pdf>
- [5] Zou CC, Towsley D, Gong W. Email worm modeling and defense. 2004. <http://www-unix.ecs.umass.edu/~gong/papers/emailModel-ICCCN04.pdf>
- [6] Song D, Malan R, Stone R. A snapshot of global Internet worm activity. Technical Report, Arbor Networks, 2001. <http://www.first.org/events/progconf/2002/d5-02-song-slides.pdf>
- [7] Imlogic. IM management, security and compliance solutions. 2006. <http://www.imlogic.com>
- [8] Williamson M, Parry A. Virus throttling for instant messaging. In: Proc. of the Virus Bulletin Conf. (vb2004). 2004. <http://www.hpl.hp.com/techreports/2004/HPL-2004-81.pdf>
- [9] Kienzle DM, Elder MC. Recent worms: A survey and trends. In: Staniford S, ed. Proc. of the ACM CCS Workshop on Rapid Malcode (WORM 2003). Washington, 2003. 1–10.
- [10] Mannan M, van Oorschot PC. On instant messaging worms, analysis and countermeasures. In: Proc. of the ACM CCS Workshop on Rapid Malcode (WORM 2005). Fairfax, 2005. <http://www.scs.carleton.ca/~paulv/papers/imworms.pdf>
- [11] iResearch Consulting Group. China Instant Messaging Research Report. 2005. <http://wwwh.iresearch.com.cn/>
- [12] Hindocha N, Chien E. Malicious threats and vulnerabilities in instant messaging. 2005. <http://www.symantec.com/avcenter/reference/malicious.threats.instant.messaging.pdf>
- [13] IMlogic Threat Center. 2005 real-time communication security: The year in review. 2005. http://www.imlogic.com/pdf/2005ThreatCenter_report.pdf
- [14] CN/CERT. CN/CERT China Network Security Research Annual Report. 2005. http://www.hais.org.cn/doc/2005CNCERTCCAnnualReport_Chinese.pdf
- [15] Nazario J, Anderson J, Wash R, Connelly C. The future of Internet worms. Blackhat Briefings, 2001. <http://www.crimelabs.net/docs/worm.html>
- [16] Zheng H. Internet worm research [Ph.D. Thesis]. Tianjin: Nankai University, 2003 (in Chinese with English abstract).
- [17] Messenger APIs. <http://msdn.microsoft.com/downloads/list/messengerapi.asp>
- [18] ICQ APIs. <http://www.icq.com/api/>
- [19] W32.Choke.Worm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.choke.worm.html>
- [20] W32.Goner.A@mm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.goner.a@mm.html>
- [21] W32.AimVen@mm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.aimven.worm.html>
- [22] W32.Kelvir.HI@mm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.kelvir.hi.html>
- [23] W32.SoFunny.Worm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.sofunny.html>
- [24] W32.Aplore@mm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.aplore@mm.html>
- [25] W32.Serflog.A@mm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.serflog.a@mm.html>
- [26] W32.Bizex.worm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.bizex.worm.html>
- [27] W32.Bropia.F.worm. <http://securityresponse.symantec.com/avcenter/venc/data/w32.bropia.f.worm.html>
- [28] Streftaris G, Gibson GJ. Statistical inference for stochastic epidemic models. In: Proc. of the 17th Int'l Workshop on Statistical Modelling. Chania, 2002. 609–616. http://www.ma.hw.ac.uk/~georges/research/SG_iwsm02.pdf
- [29] Frauenthal JC. Mathematical Modeling in Epidemiology. New York: Springer-Verlag, 1980.
- [30] Wang Y, Wang CX. Modeling the effects of timing parameters on virus propagation. In: Staniford S, ed. Proc. of the ACM CCS Workshop on Rapid Malcode (WORM 2003). Washington, 2003. <http://www.ece.cmu.edu/~chenxi/pub/worm.pdf>
- [31] Adamic LA, Lukose RM, Puniyani AR, Huberman BA. Search in power-law networks. Technical Report, 1063-651X, American Physical Society, 2001. 64–71. <http://www.hpl.hp.com/research/idl/papers/plsearch/pre46135.pdf>
- [32] Goh K, Oh E, Jeong H, Kahng B, Kim D. Classification of scale-free networks. PNAS, 2002,99(20):12583–12588.
- [33] Smith RD. Instant messaging as a scale-free network. 2006. <http://arxiv.org/abs/cond-mat/0206378>
- [34] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. The American Physical Society, 2001,86(14):3200–3203.
- [35] Carnegie Mellon University. The project of completely automated public turing test to tell computers and humans apart. 2005. <http://www.captcha.net/>
- [36] News.com Staff. Yahoo fills in messenger hole. 2005. <http://news.com.com/2100-1023-923638.html>

