P2P 干预式蠕虫传播仿真分析 *

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

摘 要: 对 P2P 干预式主动型蠕虫的传播机制进行了研究,指出其传播主要包括四个阶段:信息收集,攻击渗透、自我推进与干预激活。研究发现,P2P 干预式蠕虫实际是一种拓扑蠕虫,能利用邻居节点信息准确地确定攻击目标,而且攻击非常隐蔽。采用仿真的方法研究了 P2P 相关参数对 P2P 干预式蠕虫传播的影响。仿真实验表明,潜伏主机激活率对干预式蠕虫传播的影响最大,而攻击率对干预式蠕虫传播的影响较小。

关键词: 主动型蠕虫; 干预式蠕虫; P2P 网络; 仿真
中图分类号: TP393 **文献标志码**: A **文章编号**: 1001-3695(2012)01-0297-04
doi:10.3969/j.issn.1001-3695.2012.01.082

Simulation and analysis on propagation of man-activated worms in P2P network

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

Abstract: This paper studied the propagation mechanism of the man-activating worm, which was a kind of active worms in the P2P network. The propagation procedure of this kind of worm consists of four stages: information collection, penetration, self propulsion and activating. It was found that the worm was a kind of topology-awareness worms. It could properly identify the patent target by exploiting the neighbor information, which was stored in its cache. This may made it very difficult to detect it. For the purpose of examining the effect of P2P-related parameters on propagation of the worm, developed simulation program. Simulations experiments show that amongst all the P2P-related parameters, the probability of worm being activated rather than the attacking rate has the most effect on worm propagation.

Key words: active worms; man-activating worms; P2P networks; simulation

0 引言

随着 P2P 网络的迅速发展,P2P 网络的安全形势日益严峻。在 2006、2007 连续两年的《CNCER/CC 网络安全工作报告》[1] 和 Symatic 公司 2007 年度 Internet 安全报告[2] 中,P2P 网络应用均被列为互联网安全的主要威胁。

互联网的最大安全威胁来自于蠕虫的传播。在互联网上,流行的蠕虫采用的主要传播方式是随机扫描(如 Code Red、Slammer 、WS32. Blaster 等)。这类蠕虫没有特定区域的易攻击目标信息,而是随机地来寻找攻击目标。并且其传播的条件必须是在整个 IP 地址范围内存在一定密度的易攻击节点,同时要求用很高的速度探测不同的主机。在这样一种情况下,攻击受到极大的限制,但即使在这样的情况下,蠕虫也能在极短时间内让整个互联网瘫痪。

P2P 网络中的情况更糟糕,因为在 P2P 网络中, 蠕虫通过邻居表就可确切地知道哪些主机在线,然后就可以进行攻击[3]。在 P2P 网络中,每个节点都保存着大量的邻居节点信息。这样,蠕虫只需要感染其中的一个节点,就可以通过其邻居节点信息迅速探测新的易攻击节点,整个过程重复迭代,迅速蔓延整个网络。总结起来这类蠕虫有三个主要特点:a)传播速度极快;b)传播成功率很高;c)攻击不易被发现。与互联网蠕虫相比,P2P 蠕虫由于不需要进行耗费时间的扫描,并且 P2P 网络的规模通常比互联网小得多,所以给 P2P 网络造成的安全威胁更大,进而对 Internet 造成严重威胁。这种利用对等网络拓扑信息及 P2P 软件漏洞进行传播的蠕虫被称做 P2P 主动型蠕虫。

根据激活方式不同,P2P 主动型蠕虫可分为自启式和干预式两种。自启式蠕虫将蠕虫代码传送到目标主机后立即就能将自己激活,而干预式蠕虫在将蠕虫代码传送到目标主机后需

收稿日期: 2011-06-07; **修回日期**: 2011-07-18 **基金项目**: 国家自然科学基金资助项目(60873075);可视化计算与虚拟现实四川省重点实验室课题(J2010N05);四川省科技厅应用基础项目(2010JY0125);四川省教育厅重点课题(10ZA007,11ZB069);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(2010-05)

作者简介:冯朝胜(1971-),男,四川广元人,副教授,硕导,博士(后),CCF 高级会员,主要研究方向为网络与信息安全的研究(csfenggy@126.com);杨军(1978-),男,副教授,硕士,主要研究方向为模式识别、系统仿真;卿昱(1970-),女,研究员,硕士,主要研究方向为网络安全;秦志光(1956-),男,教授,博导,博士,主要研究方向为信息安全.

要用户打开蠕虫代码文件才能将蠕虫激活。本文采用仿真方法对未受到关注的 P2P 干预式主动型蠕虫传播能力以及相关参数对蠕虫传播的影响进行了研究,主要贡献如下:

a)分析并明确了 P2P 干预式主动型蠕虫的攻击机制和传播原理。

b)仿真了 P2P 干预式主动型蠕虫在 P2P 网络上的攻击和传播过程。

c)考查了 P2P 相关参数对 P2P 干预式主动型蠕虫传播的影响。

1　P2P 蠕虫软件仿真

网络软件仿真包括高度抽象仿真、数据包级仿真和大规模分布式软件仿真。高度抽象仿真有着较大的仿真规模,但准确性欠佳;数据包仿真准确性较高,但仿真规模较小;大规模分布式软件仿真在仿真规模和准确度方面效果都不错,但实现难度很大。在数据包级网络仿真平台中,NS2 因其源代码公开、可扩展性强、速度和效率优势明显等特点普及率最高。它支持局域网、广域网、无线移动网及卫星网络仿真,是一种较理想的网络仿真工具[4]。PeerSim 是典型的高度抽象软件仿真平台,它主要用来仿真 P2P 网络。PeerSim 仿真体系主要由最底层、中间层、顶层三层组成。最底层是实现和运行仿真环境中各节点的参数设置和各节点上运行的协议。中间层是构成仿真平台的各组件部分,该层主要由三部分构成。其中,Initializers 组件负责仿真运行前的任务(如网络初始化和节点状态的设定);Dynamics 组件执行仿真运行过程中周期性的任务(如节点的加入、离开、设置调整等);Observers 组件负责仿真过程中节点信息或网络拓扑信息的搜集。顶层是 PeerSim 引擎,也是 PeerSim 体系管理运行的核心,该引擎负责仿真各组件的调用。PeerSim 的体系结构如图 1 所示。

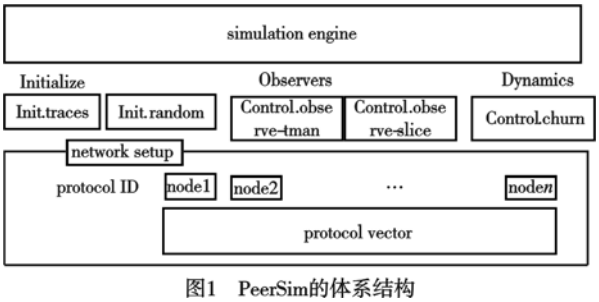

图1　PeerSim的体系结构

P2P 蠕虫仿真包括 P2P 网络仿真和 P2P 蠕虫传播仿真。

在 P2P 网络仿真上,Chen 等人[5]开发了一个 P2P 网络仿真系统。鉴于已有研究表明 P2P 文件的流行度分布为 Zipf,因此,在系统初始化时根据 Zipf 分布来分配文件。在仿真过程中,当有新文件进入系统时,新文件的流行排名由 Zipf 分布确定。当文件加入到系统中后,原来的文件也要根据 Zipf 分布进行调整。文件下载采用多点下载。P2P 网络中,用户进出网络十分频繁,这种动态性使得文件下载并不一定能成功,但有一点是可以确定的,那就是种子越多,成功可能性越大,该系统采用改进后的 Sigmod 函数描述下载成功率。吴堃等人[6]认为底层网络布局对 P2P 系统性能有较大影响,并针对已有的 P2P 仿真系统没有考虑底层网络流量这一不足,基于 NS2 开发了一个通用的、可扩展的 P2P 网络仿真系统。该仿真系统并没有解决 NS2 仿真规模小(1 000 个节点左右)的问题。除此之外,很多研究人员为验证自身研究成果开发了一些专用 P2P 网络模拟器,如 FreeNet Sim、CAN sim、Chord Sim、Gnutella Sim 等。

在 P2P 蠕虫传播仿真上,Chen 等人[5]在其开发的 P2P 网络仿真系统上对被动型蠕虫、激发型蠕虫和主动型蠕虫三种 P2P 蠕虫进行了仿真分析,发现被动型蠕虫传播速度较慢,但隐蔽性很强,很难发现;主动型蠕虫属于利用邻居节点信息传播的拓扑蠕虫, P2P 软件的漏洞为其提供了可乘之机,它的传播速度很快,也较隐蔽;激发型蠕虫则是处于主动性蠕虫和被动型蠕虫之间的蠕虫,它像主动型蠕虫一样传播时要利用软件漏洞,也像被动型蠕虫一样传播时需要利用合法连接。王跃武等人[7]针对已有的仿真模型主要针对主动型蠕虫,无法对 Contagion 蠕虫传播所依赖的业务流量进行动态模拟,提出了一个适用于 Contagion 蠕虫仿真的 Web 和 P2P 业务流量动态仿真模型,并通过选择性抽象,克服了数据包级蠕虫仿真的规模限制瓶颈,在通用网络仿真平台上,实现了一个完整的 Contagion 蠕虫仿真系统,利用该系统对 Contagion 蠕虫传播特性进行了仿真分析。另外,他们还从蠕虫传播时的漏洞主机发现、代码传播以及代码启动三个方面提出了一个完整的数据包级拓扑相关蠕虫仿真模型[8];设计了基于有向图的小世界拓扑结构生成算法,并通过选择性抽象在 NS2 上实现了一个完整的数据包级拓扑相关蠕虫仿真系统;最后,通过该仿真系统实验分析了逻辑拓扑结构和蠕虫代码启动方式(重复感染和非重复感染)对蠕虫传播的影响。在以上仿真分析中,P2P 主动型蠕虫和拓扑蠕虫都采用了自动激活方式,即这些主动型蠕虫都为自启式蠕虫。

从已有研究看,P2P 自启式主动型蠕虫广受关注,而 P2P 干预式主动型蠕虫未受到重视。

2　P2P 干预式蠕虫攻击机制

P2P 干预式蠕虫攻击行为可以分为四个阶段。

2.1　信息收集

同 P2P 自启式蠕虫一样,P2P 干预式蠕虫属于拓扑蠕虫,通过邻居主机信息来确定潜在攻击目标。由于邻居主机通常都在线,所以该蠕虫能很快发现攻击目标,而无须像其他扫描型蠕虫那样通过随机扫描来发现攻击对象。干预式蠕虫的这种攻击特点使得其在攻击准确性上比扫描性蠕虫高得多而且很难被发现。

2.2　攻击渗透

利用漏洞对目标主机进行攻击,为蠕虫的传播留下后门。目前,绝大部分漏洞是缓冲溢出漏洞。利用缓冲溢出漏洞攻击时,攻击主机向目标主机发送包含有攻击代码(为 Shell 代码)的数据,目标主机溢出后会执行攻击代码,这样目标主机为攻击代码控制。通常,攻击代码的主要功能是为蠕虫代码的自我推进开启后门(如打开端口、主动和感染主机建立连接等)。

2.3　自我推进

将蠕虫代码传播到目标主机。干预式蠕虫在发现漏洞主机后需要复制自身代码,并将其传播到目标主机。蠕虫在发现

漏洞主机后一般存在三种方式将蠕虫代码传播到漏洞主机[8,9],分别为:

a)自传播方式,将蠕虫代码作为漏洞主机发现时产生的蠕虫感染数据包的一部分,随着这些感染数据包一起发送到漏洞主机,多数拓扑相关蠕虫都采用该传播方式。如 Serflog 蠕虫[10]的若干变种为利用 MSN 直接向联系人传输经过加壳伪装的蠕虫副本文件,多数邮件病毒也都是将蠕虫作为附件进行传输。

b)第二通道传播方式。即蠕虫感染数据包不包含蠕虫代码,当其传播到漏洞主机后开启另一个传播通道,用于蠕虫代码传播。如 Aplore 蠕虫[11]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。

c)嵌入式传播方式,将蠕虫代码嵌入到正常的网络数据包中,或者替换正常的网络数据包,以正常网络流量的形式传播蠕虫代码,因而具有较高的隐蔽性。如 AimVent 蠕虫[12]等,通过修改用户和联系人之间传输的可执行文件传送蠕虫代码。

当蠕虫代码成功传送到目标主机上后,目标主机进入到潜伏状态,这意味着它属于病毒携带者,还不具有攻击感染能力。

2.4　干预激活

为了提高蠕虫代码被激活的概率,蠕虫代码文件会像许多木马一样以流行文件的文件名给自己命名。当用户误将其当做正常文件打开时,干预式蠕虫被激活,其所在的宿主机也由潜伏状态转变为感染状态。此时,宿主机就具有了攻击感染能力。

3　干预式主动型蠕虫传播仿真实验

3.1　实验参数

对等机的状态根据感染情况分成三种:易感的、潜伏的和感染的。

易感的(S):未被感染但存在被感染风险的主机。

潜伏的(E):被感染但蠕虫尚未被激活的主机。

感染的(I):被感染且蠕虫已被激活的主机。

根据对等机是否与对等网络相连,将网络的主机分为两种,即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。所有的在线主机就构成了对等网络。

为了便于下面的蠕虫建模分析,将仿真时要用到的参数和变量列举在表 1 中。

3.2　P2P 网络仿真

要仿真 P2P 干预式主动型蠕虫,先要仿真 P2P 网络。为此,基于 P2P 网络仿真平台 PeerSim,设计和实现了一个文件共享对等网仿真程序,该程序能被像 eMule 这样的实际的 P2P 协议所驱动。主动型蠕虫传播仿真实验在仿真的对等网上进行。仿真实验包括两个先后相随的阶段,即初始化阶段和仿真阶段。根据文献[5],文件共享对等网中对等机和文件的流行度都服从 Zipf 分布,所以在初始化阶段先按照 Zipf 分布给所有的文件分配文件名,继而根据 Zipf 分布将文件分配到对等机上。在初始化阶段结束时,事件调度器被启动,这意味着仿真实验进入仿真阶段。在仿真阶段,像下载、上线和下线这样典型的对等网操作被执行,被调度的事件决定了究竟会执行什么操作。仿真程序会定期地输出实验结果。P2P 网络的仿真步骤如下:

a)节点的初始化。对应用节点上的数据结构进行初始化,这些数据结构主要包括文件列表和邻居列表。

b)文件的初始化。生成指定数量的文件和文件名,根据 Zipf 分布给文件分配文件名,将文件分配到节点上。

c)添加 P2P 操作事件到事件调度器。P2P 事件主要包括文件查询、下载、节点上线、节点下线等。

d)添加蠕虫攻击事件到事件调度器。本步骤在仿真蠕虫攻击时才会用到。

e)启动事件调度器,开始进行仿真。

表 1　仿真时用到的变量和参数

符号	说明	初值
N	网络中主机总数	2000
$S_{on}(t)$	t 时刻在线易感主机数	999
$S_{off}(t)$	t 时刻离线主机数	1000
$I_{on}(t)$	在线感染主机数	1
$I_{off}(t)$	离线感染主机数	0
$E_{on}(t)$	在线潜伏主机数	0
$E_{off}(t)$	离线潜伏主机数	0
a	蠕虫的攻击率,即单位时间内平均能攻击的目标主机数	1
λ_r	恢复率	0.00001
λ_{on}	上线率(离线时间的倒数)	0.003
λ_{off}	下线率(在线时间的倒数)	0.003
λ_e	潜伏主机激活率,即潜伏主机上的蠕虫文件被用户打开的概率	0.001
λ_d	下载率	0.003

3.3　干预式蠕虫传播仿真分析

根据干预式主动型蠕虫的攻击机制,仿真 P2P 干预式主动型蠕虫。P2P 干预式主动型蠕虫的一次攻击过程如下:

a)从用户邻居主机表中选取攻击目标。如果邻居主机都被感染,攻击暂时停止,直到有新的易感邻居主机。

b)检测攻击目标是否在线。如果没有在线,回到步骤 a);否则,进入下一步。

c)检测攻击目标是否已被感染。如果被感染,回到步骤 a);如果未被感染,转下一步。

d)将蠕虫文件传送到目标主机。

e)激活潜伏主机。用户打开伪装成流行文件的蠕虫文件,干预式蠕虫被激活,目标主机进入感染状态(仿真时,将目标主机状态置为感染状态,总感染主机数和在线感染主机数加 1,为新感染主机建立第一次攻击事件,为老感染主机建立下一次攻击事件,将事件放入事件调度器的事件队列中)。

考虑到仿真本身的随机性,将同样仿真条件下 20 次仿真实验结果的平均值作为该条件下实际的仿真结果。如无特别说明,实验时变量和参数取表 1 中给出的默认值。

图 2~11 分别考查了攻击率、恢复率、潜伏主机激活率、上线率和下线率对干预式蠕虫传播的影响。实验表明,潜伏主机激活率对干预式蠕虫传播的影响最大,而对自启式蠕虫传播影

响最大的攻击率(图 12)对干预式蠕虫传播的影响较小。一般情况下,蠕虫在 1 s 内就能发起成百上千次攻击,这足以使其在 1 s 内将其邻居主机感染,所以自启式蠕虫有着惊人的传播速度。对于干预式蠕虫而言,情况完全不同。当主机被成功攻击感染后,该主机处于潜伏状态。潜伏主机并不具有攻击感染能力。与蠕虫攻击率相比,潜伏主机激活率要小很多(不超过 1)。因此,无论攻击率多大,小得多的潜伏主机激活率都会严重迟滞蠕虫的传播。潜伏主机激活率的大小主要由用户的病毒知识掌握程度和蠕虫文件的欺骗性决定。用户掌握的病毒知识越多,越有能力识别一个文件是否是蠕虫文件,越不可能去启动来历不明的文件。另一方面,蠕虫文件越具欺骗性(如蠕虫利用流行文件名给自身命名等),用户打开蠕虫文件的概率越大。因此,提高用户反病毒能力,特别是尽早发现和识别 P2P 干预式蠕虫的特征的能力,能有效降低潜伏主机激活率,进而能有效控制干预式蠕虫的传播。

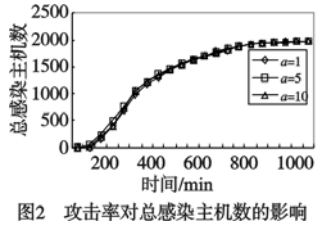

图2　攻击率对总感染主机数的影响

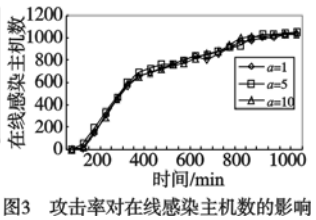

图3　攻击率对在线感染主机数的影响

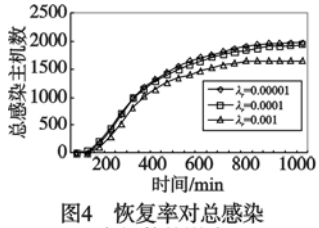

图4　恢复率对总感染主机数的影响

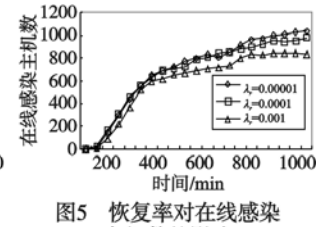

图5　恢复率对在线感染主机数的影响

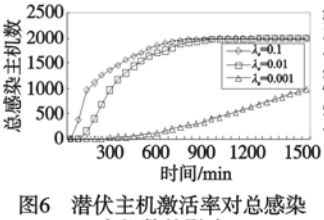

图6　潜伏主机激活率对总感染主机数的影响

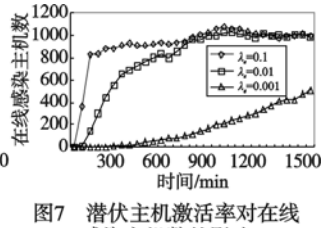

图7　潜伏主机激活率对在线感染主机数的影响

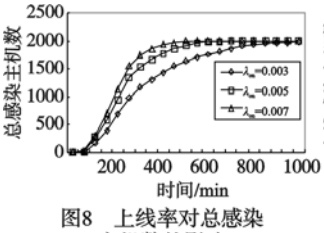

图8　上线率对总感染主机数的影响

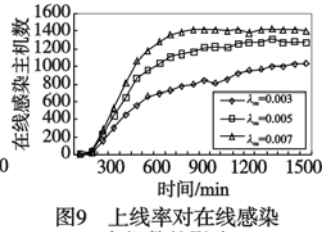

图9　上线率对在线感染主机数的影响

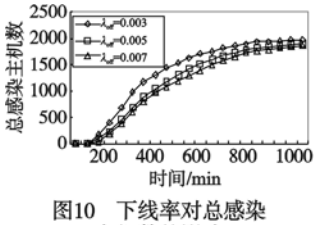

图10　下线率对总感染主机数的影响

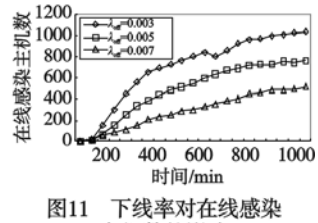

图11　下线率对在线感染主机数的影响

4　结束语

作为 P2P 主动型蠕虫的一种,P2P 干预式主动型蠕虫并未受到重视。本文对其传播机制进行了研究,发现:与自启式蠕虫不同,干预式在传播到目标主机上后并不会自动激活,这时目标主机处于潜伏状态;蠕虫需要用户干预才能被激活,蠕虫被激活后,目标主机就由潜伏状态转变为感染状态。干预式蠕虫的传播主要包括四个阶段:信息收集、攻击渗透、自我推进与干预激活。P2P 干预式蠕虫实际是一种利用邻居信息进行传播的拓扑蠕虫。邻居信息不仅大大增加了攻击的准确性,还使蠕虫在攻击时少了扫描探测这个环节,干预式蠕虫由此变成了隐蔽性极强的非扫描蠕虫。根据 P2P 网络协议和 P2P 干预式主动型蠕虫的传播原理和攻击机制,本文编写了 P2P 网络和 P2P 干预式蠕虫仿真软件,仿真了 P2P 干预式蠕虫在 P2P 网络上的传播过程。仿真实验表明,潜伏主机激活率对干预式蠕虫传播的影响最大,而对自启式蠕虫传播影响最大的攻击率对干预式蠕虫传播的影响较小,因为潜伏主机激活率会严重迟滞蠕虫的传播,而不论攻击力有多强。因此,控制潜伏主机激活率是抑制干预式蠕虫传播的关键。未来将重点研究 P2P 干预式主动型蠕虫的检测和控制方法。

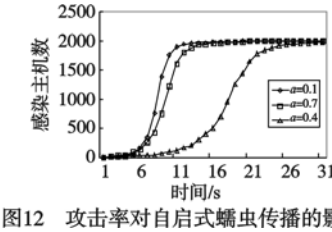

图12　攻击率对自启式蠕虫传播的影响

参考文献:

[1] CNCERT/CC[EB/OL]. [2010-01-19]. http://www.cert.org.cn/UserFiles/File/CNCERTCC2007AnnualReport_Chinese.pdf.

[2] Symantec Inc. Symantec Internet security threat report[EB/OL]. [2007-3-26]. http://www.symantec.com/threatreport.

[3] STANIFORD S, PAXSON V, WEAVER N. How to own the Internet in your spare time[C]//Proc of the 11th USENIX Security Symposium. New York:ACM, 2002:149-167.

[4] 汤碧玉,杨光松. 面向对象技术在 NS2 网络仿真环境中的应用[J]. 信息技术,2004,28(9):45-48.

[5] CHEN G, GRAY R S. Simulating non-scanning worms on peer-to-peer networks[C]//Proc of the 1st International Conference on Scalable Information Systems. New York:ACM,2006.

[6] 吴堃,戴菡,叶保留,等. 基于 NS2 的 P2P 网络模拟平台研究[J]. 系统仿真学报,2006,18(8):2152-2157.

[7] 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.

[8] 王跃武, 荆继武, 向继,等. 拓扑相关蠕虫仿真分析[J]. 软件学报,2008,19(6):1508-1518.

[9] WEAVER N, PAXSON V, STANIFORD S, *et al*. A taxonomy of computer worms[C]//Proc of the 2003 ACM Workshop on Rapid Malcode. New York:ACM,2003.

[10] W32.serflog A@mm[EB/OL]. [2010-03-05]. http://sccurityresponse.symantcc.com/avccntcr/vcnc/data/w32.serflog.a@mm.html.

[11] W32.aplore@ram[EB/OL]. [2010-03-05]. http://securityresponse.symantec.com/avcenter/venc/data/w32.aplore@mm.html.

[12] W32.imVen@mm[EB/OL]. [2010-03-05]. http://securityresponse.symantec.com/avcenter/venc/data/w32.almven.worn1.html.