蠕虫病毒的传播机制研究*

张运凯[1,2]，王方伟[2]，张玉清[3]，马建峰[1]

(1. 西安电子科技大学 计算机学院，陕西 西安 710071；2. 河北师范大学 网络信息中心，河北 石家庄 050016；
3. 中国科学院 研究生院 国家计算机网络入侵防范中心，北京 100039)

摘　要：分析了网络蠕虫的特征,从扫描方法入手研究了蠕虫的传播机制,对其性能进行了比较,并提出了蠕虫的控制方法。

关键词：蠕虫；网络安全；病毒；扫描

中图法分类号：TP309.05　　**文献标识码**：A　　**文章编号**：1001-3695(2005)04-0137-03

Research of Worm Virus s Propagation Mechanism

ZHANG Yun-kai[1,2], WANG Fang-wei[2], ZHANG Yu-qing[3], MA Jian-feng[1]

(1. School of Computer Science & Technology, Xidian University, Xi'an Shanxi 710071, China; 2. Network Center, Hebei Normal University, Shijiazhuang Hebei 050016, China; 3. National Computer Network Intrusion Protection Center, Graduate School, Chinese Academy of Sciences, Beijing 100039, China)

Abstract: Analyses the network worm's characteristic, studies worm's propagation mechanism obtained from the scanning methods, and carries on the comparison to its performance. Some worm's control methods are proposed.

Key words: Worm; Network Security; Virus; Scanning

"蠕虫"这个生物学名词在 1982 年由 Xerox PARC 的 John F. Shoch 等人[1]最早引入了计算机领域,他们给出了蠕虫的两个最基本特征:可以从一台计算机移动到另一台计算机和可以自我复制。最早的网络蠕虫病毒作者是美国的莫里斯[2],他于 1988 年编写的蠕虫病毒是在美国军方的局域网内活动。世界性的第一次大规模在 Internet 网上传播的蠕虫病毒是 1998 年底的 Happy99 蠕虫病毒,它随着邮件一起被发出,一旦被执行,便会在屏幕上不断爆发出绚丽多彩的礼花。1999 年的"美丽杀"蠕虫宏病毒,是世界上最大的一次病毒浩劫,也是最大的一次蠕虫大泛滥,造成巨大经济损失。2000 年,欧美爆发的"爱虫"蠕虫病毒,造成了更大的经济损失。在 2001 年 7 月的"Code Red V2"蠕虫至少感染了 100 多万台计算机,造成至少 26 亿美元的经济损失。同年 9 月的"Nimda"的破坏力远远超过"红色代码",到现在已无法估计。2003 年 1 月的"Slammer"是有史以来传播最快的蠕虫,在 10 分钟内就感染了 90% 多有漏洞的计算机[3,4]。同年 8 月 11 日,又爆发了"Blaster"蠕虫,即"冲击波"蠕虫,其破坏力巨大,在国内,两天时间里就使数千个局域网陷于瘫痪状态。在 8 月 15 日。又出现了"Nachi"蠕虫,即"冲击波杀手"蠕虫,它使"Blaster"蠕虫的影响大大降低,但比"Blaster"影响更大,造成了大规模的网络拥塞。

随着网络应用的普及和深入,蠕虫对网络安全的威胁日益增加。特别是在高速网络环境下,多样化的传播途径和应用环境更使得蠕虫爆发率大为提高,覆盖面也更加广泛,造成的损失越来越大。现在对蠕虫的分析和防治已成为计算机安全领域的研究热点。

收稿日期：2004-05-21；修返日期：2004-07-18

基金项目：国家"863"计划资助项目(2002AA142151)；国家信息关防与网络安全保障持续发展计划资助项目；中国科学院知识创新工程方向项目(KGCX2-106)；北京市科技计划项目资助(H020120090530)

1　蠕虫病毒传播的数学模型

1.1　蠕虫和病毒的区别

病毒是一段可执行代码,它具有独特的复制能力,可以把自身附着在各种类型的文件并在一定条件下激活,这一特点很像生物病毒。而宏病毒是一种新形态的病毒,它寄生在一些数据文档中以便在不同的操作系统中使用,已具备了跨平台传播的能力。蠕虫则完全不同,它是通过计算机网络连接进行传播,通常在计算机内存中复制自己的一段程序;蠕虫不需要宿主程序,它在结构、攻击内容以及检测方法上和普通病毒有着根本上的不同;蠕虫更接近于黑客工具,而不是病毒。"狮子"是一种蠕虫与黑客工具杂糅后的复合体[5],但蠕虫以病毒作为传播媒介,这是一种新的趋势。N. Weaver[6] 从目标发现、传播载体、激活、攻击者的动机等多方面对目前的蠕虫进行了分类。现在已经有人把蠕虫、病毒、木马统称为恶意移动代码(Malicious Mobile Code, MMC),它是一段计算机程序,能够在计算机或网络之间传播,未经授权、故意修改计算机系统。如果 MMC 能入侵到计算机且不需要人工干预就能直接自动执行,那它就是蠕虫;如果 MMC 附着在别的内容里面,在其他计算机上运行取决于何时、是否需要人工处理这些内容,那它就是病毒。简单地说,判断依据是看 MMC 是否自主激活。表 1 从几个主要方面给出了蠕虫与病毒的区别。

1.2　蠕虫传播的数学模型

网络蠕虫和生物病毒在自复制和传播等方面有很多相似之处,所以也可以用数学技术来研究网络蠕虫。在流行病学领域,有两种模型来模拟传染疾病的传播:随机模型和确定模型。

前者适合有简单病毒的小规模系统,后者适合有复杂病毒的大规模系统。因为网络蠕虫的传播是全球性的,所以采用确定模型。目前描述网络蠕虫的传播有很多数学模型,我们主要介绍简单流行病模型[7~9]。为了便于描述,先解释一下模型中用到的符号,如表 2 所示。

表 1 蠕虫和病毒的区别

	蠕虫	病毒
传播速度	极快	快
存在形式	独立程序	寄存文件
主要危害	侵占资源	破坏数据完整性、系统完整性
传染机制	主动攻击	宿主程序运行
传染目标	网络计算机	本地文件
检测方法	复杂	简单

表 2 符号参数说明

符号	定义
t	时刻
$I(t)$	在时刻 t 被感染的主机数
N	主机总数
β	感染率
Ω	扫描地址空间
η	扫描速率

在此模型中,每一台主机只有两种状态:易感或感染,而且一旦被传染就永远保持感染状态,其状态转移只能表示为:易感→感染,其模型表示如下:

$$\frac{dI(t)}{dt} = \beta(t)[N - I(t)] \quad (1)$$

式(1)中,β为感染率,$\beta = \eta/\Omega$ η为单位时间的探测率,即扫描速率。当 $t=0$ 时,$I(0)$ 表示初始感染的主机数。

简单流行病模型没有考虑感染主机的恢复,在 Kermack-Mckendrick 模型中假设每个主机具有三种状态:易感、感染或恢复,其状态转移可表示为易感→感染→恢复,或永远保持易感状态。Zou 等人考虑人们采用了诸如清除蠕虫、打补丁、升级系统、地址过滤、内容过滤等方法和 β不是常数等因素,提出了二因素蠕虫传播模型[8]。本文主要研究蠕虫的传播机制,主要涉及蠕虫传播的开始阶段和快速传播阶段,而人为控制要滞后得多,所以本文主要采用简单模型。

2 网络蠕虫的传播机制

2.1 蠕虫的基本结构和传播过程

网络蠕虫由四个模块组成:攻击模块、感染模块、传播模块和功能模块,其中攻击模块和感染模块最为关键,决定了蠕虫的影响范围和传播速度,关系到能否对目标主机造成巨大破坏。蠕虫的一般传播过程为:①扫描。由蠕虫的扫描功能负责探测存在漏洞的主机。当程序向某个主机发送探测漏洞信息并收到成功的反馈信息后,就得到一个可传播的对象。②攻击。攻击模块自动攻击在第①步中找到的对象,取得该主机的权限(一般为管理员权限)。③现场处理。使计算机在被感染后保留一个后门以便发动分布式拒绝服务攻击。④复制。复制模块通过原主机和新主机的交互将蠕虫程序复制到新主机并激活。蠕虫病毒的工作流程如图 1 所示。

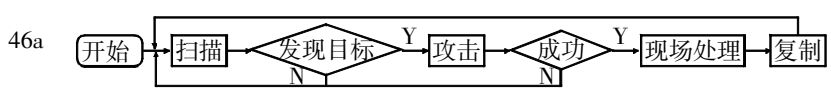

图 1 蠕虫病毒的工作流程

2.2 蠕虫的扫描方法

扫描主机漏洞是蠕虫传播的前提,蠕虫传播经常用 ICMP Ping 包、TCP SYN、FIN、RST 和 ACK 包来进行探测[13]。影响蠕虫传播的主要因素是如何能快速找到新的目标主机,所以扫描方法的性能直接决定着蠕虫传播的速度。下面比较几种常用的扫描方法,我们主要研究了在一定的地址空间内均匀地扫描情况,所以可以使用简单流行病传播模型,其中没有考虑感染时的延时。

2.2.1 随机扫描法(Random Scanning)

所谓随机扫描是感染蠕虫的计算机在寻找新的攻击目标时不知道哪些计算机系统有漏洞,随机地选择网上计算机进行扫描,所以扫描的目标为 IPv4 所有地址空间,即 $\Omega = 2^{32}$。Code Red 蠕虫和 Slammer 蠕虫都是采用随机扫描的方法[3,15],一般情况下,随机扫描由于扫描地址空间大,传播速度较慢。但是 Slammer 蠕虫传播非常快,主要因为它采用 UDP 1434 端口的非连接的扫描,而且采用了大量线程的扫描方式,使得其扫描主要受带宽的限制。

设 $N = 36\ 000$, $\Omega = 2^{32}$, $I(0) = 10$。由于式(1)不能直接求解,我们采用 Mathlab 中的 Simulate 来模拟蠕虫传播。图 2 给出了 $\eta = 100, 200, 358$ 时,随机扫描法受感染率影响的性能曲线比较。由图 2 可以看出,感染率 β越大,传播速度越快,曲线越陡,达到平稳期的时间越短,最后成功率都在 95% 以上。

2.2.2 Hit-list 扫描

Staniford 等人在文献[15]中介绍了一种 Hit-list 扫描方法。这种蠕虫在传播之前先搜集一些有漏洞的计算机地址列表,传播时先感染这些地址列表中的计算机,然后这些感染的计算机再随机扫描互联网上的其他计算机。目前产生 Hit-list 的技术主要有:秘密扫描(Stealthy Scans)、分布式扫描(Distributed Scanning)、DNS 搜索(DNS Searches)、仅监听(Just Listen)、公开调查(Public Surveys)等。扫描过程可分两个阶段:①蠕虫感染 Hit-list 中的计算机的过程。该阶段由于蠕虫作者已经事先取得了存在漏洞的计算机,而且这些机器都具有良好的网络连接,所以该过程速度极快。②随后的随机传播过程中,扫描地址空间为 2^{32},初始感染主机的数量 $I(0)$ 为 Hit-list 表中的数量。采用 Hit-list 扫描的蠕虫要比采用随机扫描的蠕虫传播的速度快得多。

采用同样的模拟方法和参数,得出不同 Hit-list 情况下的蠕虫传播如图 3 所示。由图 3 可以看出,蠕虫传播受 Hit-list 表的影响较大,在最后达到 96.6% 成功率的情况下,Hit-list 表越大,传播速度也就越快。

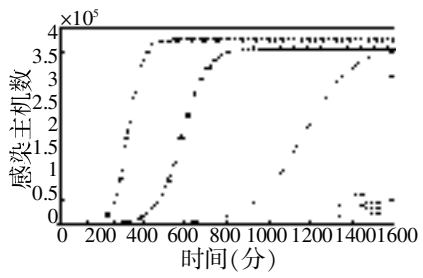

图 2 随机扫描性能比较

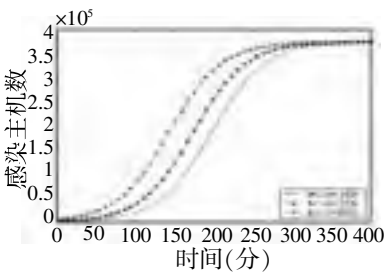

图 3 Hit list 扫描性能比较

2.2.3 路由扫描法(Routing Scanning)

Zou 等人在文献[16]中介绍了一种“路由蠕虫”,它采用 BGP 路由表中的地址来减小要扫描的地址空间 Ω。由于在 BGP 路由表中,只包含了 28.6% 的 IPv4 地址空间[16],所以,若采用路由扫描法,其探测地址空间将比随机扫描减小,β值将增大 1/0.286 = 3.5 倍,所以传播速度会增加。但是,“路由蠕虫”由于要携带地址信息,使得自身的代码变长,从这个角度讲会降低蠕虫的传播速度,但影响不大。图 4 是路由扫描法传播模型模拟结果,其中感染率 β为 0.000 000 08。

2.2.4 分而治之扫描(Divide-Conquer Scanning)

分而治之扫描的主要思想是:在释放蠕虫之前,作者需要收集一个有 10 000 ~50 000 个漏洞且网络连接良好主机的列表,在传播时,蠕虫给每个受害主机发送一个子列表,受害主机就按照子列表进行扫描。例如,主机 A 感染主机 B 后,主机 A 就把列表分成两部分,一半给主机 B,自己保留另一部分。即

使在列表中只有 10% ～20% 的主机有漏洞,这个快速分解方法也能很快通过列表且在一分钟之内蠕虫能在全部有漏洞的主机上复制自身[14]。初始的列表不必很完美,只要有一个简单的列表就可以,常用的技术包括秘密扫描、分布式扫描、DNS 搜索、Spiders 等。此扫描的特点是:隐蔽性强,通过逐渐分解列表,蠕虫越来越小;扫描造成的流量也进一步减小,更不容易检测。但是,如果分得列表的主机被关掉或死机,那这部分列表就会丢失,这种情况发生得越早,对蠕虫传播的影响越大。

2.2.5　DNS 扫描法(DNS Scanning)

蠕虫作者利用从 DNS 服务器得到的 IP 地址建立目标主机地址列表,它的优点是这些地址都是可用的,提高了扫描的准确度。但缺点也很多:①这个地址列表不易得到;②列表涉及范围太小,因为都是来自同一个 DNS 服务器;③蠕虫需要携带一个大的地址列表,传播速度受限。

2.2.6　扫描方法的性能比较

这几种扫描方法的根本不同之处在于目的主机列表的选择上,当列表大小相同时,蠕虫感染其他主机的时间主要取决于这个列表的有效率,因为这个列表可能包括未分配或保留的地址。地址列表的粒度越小,蠕虫扫描就越隐蔽,但是蠕虫代码也就越大,所以需要在地址列表大小和蠕虫传播速度间找一个合理的平衡。图 5 给出了上述几种扫描方法的性能比较。

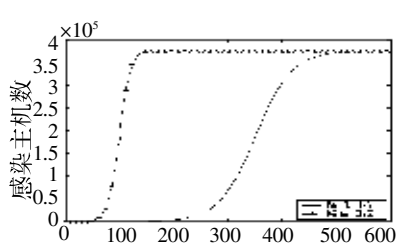

图 4　路由扫描法传播模型

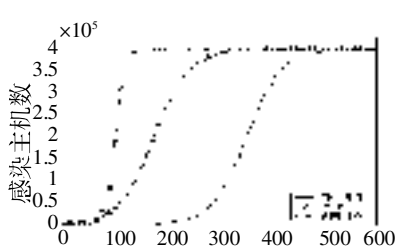

图 5　几种扫描方法的性能比较

比较三种扫描方法所采用的实验环境为:360 000 个有漏洞主机,感染率 β为 0.000 000 08, Hit-list 为 2 000,初始感染主机数为 10。由图 5 比较可知,路由扫描法性能最好,传播速度最快; Hit-list 扫描法在初始阶段传播较快;随机扫描法性能最差。其他扫描方法受实验条件所限,在此没作比较。

3　蠕虫的控制

目前网络蠕虫的种类越来越多,破坏力也越来越大,而且更加隐蔽。多个例子也说明:从发现漏洞到蠕虫爆发的时间越来越短,但从蠕虫爆发到蠕虫被消灭的时间却越来越长,网络蠕虫的防范和控制越来越困难。目前,网络蠕虫的防范和控制还主要是采用人工手段,针对主机主要采用手工检查、清除,利用软件检查、清除,给系统打补丁、升级系统,采用个人防火墙,断开感染蠕虫的机器等方法;针对网络主要采用在防火墙或边缘路由器上关闭与蠕虫相关的端口、设置访问控制列表和设置内容过滤等方法。

通过前面的分析可知,蠕虫的传播速度越来越快,采用人工方法控制蠕虫传播是不现实的,控制蠕虫的快速传播需要蠕虫监测和预警系统[17]。从蠕虫传播时扫描地址空间的角度分析,如果能够大大地增加地址空间,可以有效地控制蠕虫传播。我们知道, IPv6 的地址空间为 128 位,是 IPv4 地址空间的 296 倍,所以升级到 IPv6 可以有效地防止蠕虫传播。

4　结束语

我们利用简单流行病传播模型分析了网络蠕虫的传播机制,研究了几种均匀扫描方法对蠕虫传播性能的影响,对一些本地性扫描(Local Preference Scanning)、顺序扫描(Sequential Sanning)、选择攻击扫描(Selective Attack Scanning) 未作研究。网络蠕虫的技术发展很快,各种新的地址扫描方法不断出现,使得网络蠕虫传播越来越快,造成的影响也越来越严重。我们下一步的工作包括:进一步完善蠕虫的传播模型,研究各参数对模型的影响,使之更接近实际情况;研究蠕虫的网络流量特性;建立蠕虫自动预警系统;目前慢速蠕虫还没有有效的方法[12],我们将研究慢速蠕虫的传播模型及其防治方法。

参考文献:

[1] Shoch, J Hupp. The "Worm" Programs-early Experiments with a Distributed Computation[J]. Communications of the ACM, 1982, 22(3): 172-180.

[2] D Seeley. A Tour of the Worm[C]. CA: Proceedings of the Winter Usenix Conference, 1989. 287.

[3] D Moore, V Paxson, S Savage, *et al*. Inside the Slammer Worm[J]. IEEE Magazine on Security and Privacy, 2003, 1(4): 33-39.

[4] D Moore, V Paxson, C Shannon, *et al*. The Spread of the Sapphire/Slammer Worm[R]. Proceedings of the CAIDA Technical Report, 2003. 1-9.

[5] 左晓栋,戴英侠. "狮子"蠕虫分析及相关讨论[J]. 计算机工程, 2002, 28(1): 16-17.

[6] N Weaver, V Paxson, S Staniford, *et al*. A Taxonomy of Computer Worms[C]. Proc. ACM CCS Workshop on Rapid Malcode, 2003. http://www.silicondefense.com/research/.

[7] Cliff Changchun Zou, Weibo Gong, Donald F. Towsley: Code Red Worm Propagation Modeling and Analysis under Dynamic Quarantine Defense[C]. ACM Conference on Computer and Communications Security, 2003. 51-60.

[8] Cliff Changchun Zou, Weibo Gong, Donald F. Towsley: Code Red Worm Propagation Modeling and Analysis[C]. ACM Conference on Computer and Communications Security, 2002. 138-147.

[9] J Daley, J Gani. Epidemic Modeling, an Introduction[M]. Cambridge Univ. Press, 1999.

[10] D Moore, C Shannon, Geoffrey, *et al*. Internet Quarantine: Requirements for Containing Self-propagating Code[C]. INFOCOM 2003. http://www.cs.ucf.edu/～jglenn/research.html.

[11] D Moore, C Shannon. Code-red: A Case Study on the Spread and Victims of an Internet Worm[C]. Proceedings of the 2002 ACM SICGOMM Internet Measurement Workshop, 2002. 273-284.

[12] D Nojiri, J Rowe, K Levitt. Cooperative Response Strategies for Large Scale Attack Mitigation[C]. DISCEX III, 2003. 22-24.

[13] J O Kephart, S R White. Measuring and Modeling Computer Virus Prevalence[C]. Proc. of the 1993 IEEE Computer Society Symposium on Research in Security and Privacy, 1993. 2-15.

[14] Nicholas Weaver. Potential Strategies for High Speed Active Worms: A Worst Case Analysis[R]. U. C. Berkeley BRASS Group, 2002. 1-9.

[15] S Staniford, V Paxson, N Weaver. How to Own the Internet in Your Spare Time[C]. San Francisco: 11th Usenix Security Symposium, 2002.

[16] C C Zou, D Towsley, *et al*. Routing Worm: A Fast, Selective Attack Worm Based on IP Address Information[R]. Univ. Massachusetts Technical Report TRCSE-03-06, 2003.

[17] C C Zou, L Gao, W Gong, *et al*. Monitoring and Early Warning for Internet Worms[C]. Washington D C: 10th ACM Symposium on Computer and Communication Security, 2003.

作者简介:

张运凯,男,副教授,博士研究生,主要研究方向为网络安全;王方伟(1976-),硕士,主要研究方向为网络安全;张玉清(1967-),副研究员,博士后,主要研究方向为信息安全;马建峰(1963-),教授,博士生导师,主要研究方向为信息安全。