第33卷 第18期 计 算 机 工 程 2007年9月
Vol.33 No.18 Computer Engineering September 2007
·安全技术· 文章编号：1000—3428(2007)18—0167—03 文献标识码：A 中图分类号：TP393.08

对等网络蠕虫建模与分析

王 勇[1,2]，云晓春[1,3]，李奕飞[4]
(1. 中国科学院计算技术研究所，北京 100080；2. 中国科学院研究生院，北京 100049；
3. 哈尔滨工业大学计算机网络与信息安全技术研究中心，哈尔滨 150001；4. 四川大学计算机学院，成都 610064)

摘 要：对等网络蠕虫是威胁对等网络乃至Internet安全的一个重要问题。针对蠕虫扩散过程和对等网络拓扑的特点，构造了对等网络蠕虫传播模型；以Gnutella网络为研究实例，获取其拓扑快照数据，用以模拟蠕虫在Gnutella网络中的传播过程，从而验证模型的有效性，衡量对等网络蠕虫对P2P网络的危害。

关键词：蠕虫；对等网络；Gnutella

Modeling and Analysis of P2P Worms

WANG Yong[1,2], YUN Xiao-chun[1,3], LI Yi-fei[4]
(1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080; 2. Graduate University, Chinese Academy of Sciences, Beijing 100049; 3. Information Security Technology Research Center, Harbin Institute of Technology, Harbin 150001; 4. School of Computer Science, Sichuan University, Chengdu 610064)

【Abstract】 P2P worms are the most serious threats to P2P applications and Internet security. This paper presents a novel P2P worm spreading model, which considers the P2P network topology characteristics. The numerical simulation results based on Gnutella topology snapshots indicate that the spreading model can provide a set of metrics to evaluate the harms of P2P worms.

【Key words】 worms; P2P network; Gnutella

对等网络蠕虫(P2P worms)是一类利用P2P共享程序在网络中传播的恶意代码[1]，根据传播策略可分为：

(1)基于文件共享的传播蠕虫，即：将蠕虫代码伪装后被动地接受下载请求(如通过KaAaA传播的Benjamin蠕虫和通过Gnutella传播的Vbs_GnutellWorm蠕虫[2])；

(2)基于P2P节点(servent)系统应用漏洞扩散的蠕虫，即利用P2P网络拓扑及其交互性质自主扩散，2005年召开的IPTPS会议上，正式提出了这种蠕虫传播策略，其危害的严重性引起了高度关注[3]。Internet网络中，P2P应用种类繁杂，没有统一的安全策略和规范，无法保证这些逻辑覆盖网络应用软件和运行这些软件的操作系统不存在漏洞。因此，利用P2P节点邻居地址信息进行扩散的蠕虫，严重威胁着P2P网络乃至整个互联网的安全。

P2P网络蠕虫传播具有隐蔽性强、传播速度快和危害面广等特点[3]。建立P2P蠕虫传播模型能够充分反映其传播过程，暴露蠕虫传播过程的弱点，预测可能存在的威胁。本文以Gnutella网络为研究实例，针对P2P网络拓扑结构的特性，建立了对等网络蠕虫传播模型；通过实际获取的Gnutella网络拓扑数据模拟验证了传播模型的有效性，该模型也适用于其他的非结构化对等网络系统。

1 相关研究

文献[4]研究了P2P网络蠕虫传播行为，建立对等网络下蠕虫扩散/对抗共存的双因素模型，提出了P2P网络下蠕虫的抑制框架。文中指出：对等网络蠕虫扩散过程中新感染的节点数量 dx/dt 和网络节点的平均邻居数量 D 密切相关，呈指数递增的趋势。文献[5]定义了2种基于P2P网络系统的蠕虫攻击策略，OPHLS(离线态基于攻击列表扫描)和OPS(在线态基于P2P扫描)，文中分析了P2P网络的尺度、度分布和节点脆弱性等因素对蠕虫扩散的影响。

2005年，文献[3]提出了基于非结构化对等网络逻辑拓扑的蠕虫传播模型，刻画了蠕虫传播的非精确上限。文中指出：利用逻辑覆盖网拓扑结构传播的蠕虫更隐蔽、攻击更精确高效，可以使目前绝大多数针对扫描型蠕虫的防御机制失效。作者用基于BA模型的拓扑生成器生成的网络代表实际的对等网络拓扑，模拟了在具有警察节点的P2P网络中，蠕虫的传播过程。

然而，通过实际的对等网络测量和模型化分析[6]，得知：P2P网络中节点的邻居关系并不是一成不变的，邻居的数量受节点在线时间长短、节点使用的客户端软件版本等多种因素的影响，只用平均度数(即节点的平均邻居数)不能完整地刻画P2P网络的拓扑结构，从而也影响了模型的准确性。

另一方面，对等网络的多样性，难以用单一的拓扑模型描述，文献[7]测量分析了Gnutella网络的拓扑特性，指出网络的节点度分布不服从幂律特性，Gnutella网络的成长过程亦不遵循BA模型。

综上所述，对等网络种类繁多，传播特征复杂，需要有针对性地建立蠕虫传播模型。本文建立了纯对等网络下，蠕虫的扩散模型，通过实际获取的Gnutella网络拓扑数据，验证理论模型的有效性。

基金项目：国家自然科学基金资助项目(60403033)；国家网络与信息安全保障持续发展计划基金资助项目(2004-研4-AA-02)
作者简介：王 勇(1976－)，男，博士研究生，主研方向：对等网络测量，信息安全；云晓春，博士、教授；李奕飞，硕士
收稿日期：2006-09-22 **E-mail：**wangyong@software.ict.ac.cn

2 Gnutella 网络蠕虫传播模型

本文讨论的蠕虫传播过程可以描述为：在对等网络中，已感染蠕虫的节点通过访问客户端软件中的访问列表[8,9]，获取被感染节点的邻居信息，攻击邻居节点系统，通过覆盖网络拓扑传播蠕虫体。

根据对等网络逻辑拓扑的特点，为了简化建模过程，本文提出的蠕虫传播模型有如下 4 个假设条件：

(1)感染时间为一个时刻(time tick)；

(2)蠕虫能够同时扫描所有的邻居节点；

(3)对等网络节点总数已知；

(4)蠕虫只感染对等网络内部的节点。

表 1 列出了模型中使用的相关变量及其含义。

表 1 模型中使用的变量及含义

符号	含义	符号	含义
$E(i)$	时刻 i 新增的感染节点数	d	节点平均邻居数
$N(i)$	时刻 i 对等网络中脆弱主机总数	T	对等网络总节点数
$M(i)$	时刻 i 被感染的主机总数	$R(i)$	时刻 i 邻居关系的回边率
P	主机补丁修复率	k	主机宕机率

表中，回边率的定义为：“ 对于一个给定的拓扑图，进行广度优先遍历过程中，重复访问到的节点次数与总节点访问次数的比值 ”。回边率在对等网络信息扩散模型中的物理含义是指：第 i 次感染与第 i-1 次感染之间的节点重复次数在两次感染节点总数中的比率。

设 $E(i)$为 i 时刻新增感染节点数($i\geq 0$)，由前述假设条件，此时刻蠕虫产生的扫描数为 $d\times E(i)$，其中 d 为对等网络的平均邻居数。然而，$E(i)$个扫描对于网络中已经感染了的节点是无效的，因此，引入时刻 i 网络回边率 $R(i)$，将重复感染的节点剔除，得到有效的扫描数量为：$(1- R(i))\times E(i)$。此外，在整个蠕虫扩散过程中，设网络分别以概率 k 和 p 出现节点宕机或预先打补丁的现象，使对等网络中的脆弱主机数量减少。从而，对等网络蠕虫扩散离散模型可以表示成

$$\begin{cases} E(i+1)=(N(i)-M(i))\left[1-(1-\frac{1}{T})^{d\times(1-R(i)\times E(i)}\right] \\ M(i+1)=(1-k-p)\times M(i)+E(i+1) \\ N(i)=(1-k-p)^i\times T \end{cases} \tag{1}$$

其中，在第 i 感染时刻，新增的感染节点数 $E(i+1)$证明如下：

要证明

$$E(i+1)=(N(i)-M(i))[1-(1-\frac{1}{T})^{d\times(1-R(i))\times E(i)}] \tag{2}$$

对所有的 i 成立，只需证明

$$E(i+1|J)=(N(i)-M(i))[1-(1-\frac{1}{T})^{J}] \tag{3}$$

对所有的 J 成立即可。

当 J=1 时，一个扫描命中对等网络中脆弱主机的概率为

$$\frac{N(i)-M(i)}{T}=(N(i)-M(i))\left[1-(1-\frac{1}{T})^{1}\right]$$

设对于 J=K 成立；

当 J=K+1 时， 新增的扫描可能会命中脆弱主机或不命中，于是，对于 $J=K+1$，$E(i+1|K+1)$可以表示为

$$\begin{aligned} &E(i+1|K+1) \\ &=(E(i+1|K)+1)P(Y=1)+E(i+1|K)P(Y=0) \\ &=E(i+1|K)+P(Y=1) \\ &=E(i+1|K)+\frac{(N(i)-M(i)-E(i+1|K))}{T} \\ &=(N(i)-M(i))[1-(1-\frac{1}{T})^{K+1}] \end{aligned}$$

故，对于 $J=K+1$，式(3)也成立。

令 $K=(1-R(i))\times d\times E(i)$，可得式(2)成立，由此式(1)得证。

本文模型的特点在于：

(1)使用离散形式表述蠕虫传播的规律，有利于分步考察对等网络中蠕虫的扩散过程；

(2)考虑了补丁修复和主机宕机等因素，反映了网络中节点下线/关机以及升级打补丁等实际情况；

(3)避免了网络中节点重复感染而导致的感染总数重复计算问题，更准确地反映对等网络中蠕虫的扩散过程。

3 实验验证

验证模型的有效性包括两个方面工作：

(1)获取实际的 Gnutella 网络拓扑数据，生成拓扑图；

(2)在获取了 Gnutella 网络拓扑图后，定义单节点的感染行为，模拟感染过程。本节分别论述这两个方面的实验。

3.1 拓扑图获取

Gnutella网络是一类典型的非结构化P2P网络，拥有百万级在线用户群，用户分布广泛，使用它作为模拟实验的拓扑数据，说服力强。随着Gnutella网络的发展，其网络结构由原来的完全对等形式转变为目前的层次化网络。同时，网络中节点的查询算法、连边策略等影响网络拓扑结构的主要机制也都相应地发生了变化。Gnutella 0.6 协议[8]将网络中的节点化分成两个等级，即超级节点(Ultra-Peers)和叶子节点(Leaf-Peers)。超级节点负责查询消息的路由，是整个Gnutella网络的骨架，实现类似节点集中器的功能。叶子节点以超级节点为代理连接到Gnutella网络，二者之间使用QRP协议完成查询消息的传递转发。新节点加入Gnutella网络时，首先利用GWebCache机制获取Gnutella网络的入口节点IP地址和端口号；接着通过节点握手协议连接到Gnutella网络并表明本节点的身份(Ultra-Peer或Leaf-Peer)；一旦握手过程成功完成，节点就可以和Gnutella网络进行通信了。

本文根据 Gnutella0.6 扩展协议和文献[7]实现了分布式快速网络拓扑爬行器 D-Crawler。D-Crawler 的整体结构如图 1 所示。

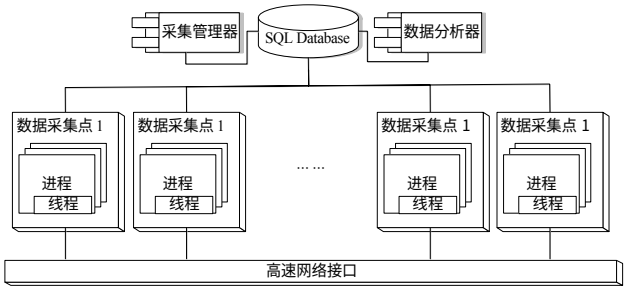

图 1 Gnutella 网络爬行器整体结构

表 2 列出了 6 组 Gnutella 网络快照数据的概要信息。通常情况下，数据获取时大约 40%左右的超级节点不能与 D-Crawler 直接建立连接。其原因有两种：

(1)节点被 D-Crawler 访问时已经下线，不在 Gnutella 网络内，这种节点约占 2%～3%；

(2)节点处于防火墙背后或由于其他原因(例如节点忙碌)拒绝 TCP 连接请求，约占 25%～34%。另外，在数据获取期间，大约有 1%的节点状态发生了改变(由叶子节点转变为超级节点，或相反)。

基于上述原因，在做进一步拓扑特性分析之前，还需要对获取的数据进行后处理，主要包括：

(1)忽略网络中拒绝连接的节点；

(2)获取的拓扑图转化为无向拓扑(节点间 TCP 连接是双向的)；

(3)忽略叶子节点中，声称自己是另一个超级节点的邻居的节点；

(4)忽略叶子节点中，声称自己是另一个叶子节点的父亲的节点。

数据后处理中，第 2 类操作影响 5%的拓扑数据，第 3 和第 4 类操作，共影响的数据不超过 1%。通过对比分析[6,7]，本文的采集拓扑能够反映实际的Gnutella网络拓扑结构。

表 2　Gnutella 网络拓扑数据概要信息

	08-02	08-03	08-04	08-05	08-06	08-07
总节点	1162693	1085870	1449441	1181661	1251851	1226303
上层节点	377551	394115	400175	399710	398452	394000
上层节点/%	32.47	36.29	27.61	33.82	31.83	32.41
建立连接	63779	55785	74884	51342	52651	59688
成功连接	44689	40204	50084	35831	36472	41456
成功连接/%	70.13	72.07	66.88	69.79	69.27	69.63
失败连接/%	29.87	27.93	33.12	30.21	30.73	30.37
叶子节	785142	691755	1049266	781951	853399	832302
节点叶子/%	67.53	63.71	72.39	66.17	68.17	67.59

3.2　蠕虫扩散模拟

基于 3.1 节中获取的 6 组 Gnutella 网络拓扑数据，针对每一个拓扑图模拟了蠕虫传播过程，见图 2；同时，根据 6 组数据的统计特性，设置式 1 的相关参数，得到图 3 所示的理论模型计算曲线。其中 T 为表 2 中的总节点数，p=0.000 01，k=0.000 01，R(i)为对 6 次拓扑快照数据分别计算回边率后的结果(如图 4 所示)。比较图 2 和图 3 可以看出：

(1)模拟实验和理论计算结果都准确地反映了对等网络蠕虫的传播扩散特点，即在对等网络中(本文中的 Gnutella 网络)，蠕虫的扩散速度相当惊人，通常情况下，在 15 个感染步骤后，基本上就能够完成蠕虫在整个网络中的扩散过程；

(2)本文提出的模型较好地描述蠕虫在 Gnutella 网络中的扩散过程，能够体现对等网络拓扑对蠕虫扩散过程的影响；

(3)理论曲线与实际模拟曲线还存在细小差异，主要原因在于：理论模型是从整体性上描述对等网络蠕虫传播行为，是对等网络下具有共性的蠕虫传播过程，而一次具体的蠕虫扩散模拟与网络当时的状态、模拟系统参数设置等因素有关。

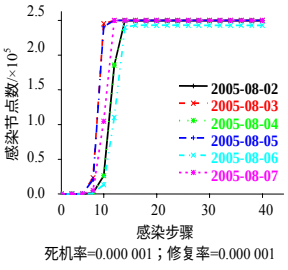

图 2　蠕虫传播过程模拟曲线

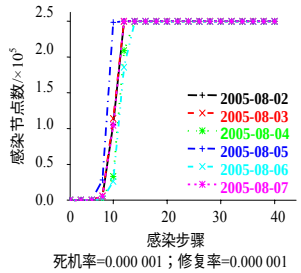

图 3　蠕虫传播过程理论模型曲线

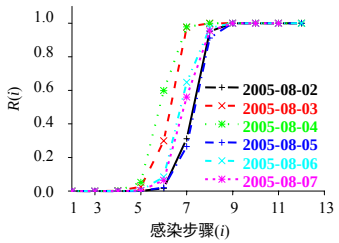

图 4　6 次拓扑快照数据的回边率 *R(i)*

4　结论与展望

本文分析了对等网络蠕虫传播过程的特点，构建了对等网络蠕虫扩散模型，通过实际获取的 Gnutella 网络拓扑数据验证了模型的正确性。该蠕虫扩散模型能够较好地描述蠕虫在 Gnutella 网络中的传播行为，同时也适用于其他非结构化对等网络。对等网络应用纷繁复杂，通过研究不同种类的对等网络系统，建立准确的蠕虫扩散模型，可以为对等网络下蠕虫的检测、防御等研究提供理论支持。

对等网络环境中缺乏有效的蠕虫检测防御机制，对等网络蠕虫传播行为的模型化研究还需要不断完善。今后的主要工作包括：(1)研究 Gnutella 网络中关联特性与蠕虫传播过程的关系；(2)针对连续多个拓扑快照数据，优化蠕虫传播模型，探讨网络中节点的状态对其扩散过程的影响；(3)探索对等网络中有害信息的防御和检测机制，从而抑制蠕虫等有害信息在对等网络中的传播，减少其危害性。

参考文献

1 Joukov N, Chiueh Z. Internet Worms As Internet-wide Threat[R]. Stony Brook University, 2003-09.

2 Virus List[Z]. (2005-10). http://www.viruslist.com: Kaspersky Lab.

3 Zhou L, Zhang L, McSherry F, et al. A First Look at Peer-to-Peer Worms: Threats and Defenses[C]//Proc. of the 4th International Workshop on Peer-to-Peer Systems, Ithaca, NY, 2005.2.

4 Yu Wei. Analyze the Worm-based Attack in Large-scale P2P Networks[C]//Proceedings of the 8th IEEE International Symposium on High Assurance Systems Engineering. 2004: 308-309.

5 Yu Wei, Boyer C, Xuan Dong. Analyzing Impacts of Peer-to-Peer Systems on Propagation of Active Worm Attacks[R]. Dept. of Computer Science and Engineering, the Ohio-State University, 2004-03.

6 刘　刚. 对等网络的测量、模型化与分析[D]. 哈尔滨: 哈尔滨工业大学, 2005.

7 Stutzbach D, Rejaie R. Characterizing the Two-tier Gnutella Topology[C]//Proc. of ACM SIGMETRICS'05. 2005-06.

8 Gnutella RFC Home Pages[Z]. (2003-10). http://rfc-gnutella.sourceforge.net.

9 LimeWire Home Pages[Z]. (2005-10). http://www.limewire.org.

10 文伟平, 卿斯汉, 蒋建春. 网络蠕虫研究与进展[J]. 软件学报, 2004, 15(8): 1208-1218.

11 Watts D J, Strogatz S H. Collective Dynamics of 'Small-World' Networks[J]. Nature, 1998, 393(6684): 440-442.

12 Uoregon University P2P Research Website[Z]. (2005-09). http://mirage.cs.uoregon.edu.

13 Chen Z, Gao L, Kwiat K. Modeling the Spread of Active Worms[C]//Proc. of the 22nd Annual Joint Conference on Computer and Communications Societies. 2003: 1890-1900.