第 33 卷 第 8 期 计 算 机 工 程 2007 年 4 月
Vol.33 No.8 Computer Engineering April 2007
·安全技术· 文章编号：1000—3428(2007)08—0139—03 文献标识码：A 中图分类号：TP393

蠕虫在 P2P 网络中的传播研究

郝向东[1]，王开云[1]，张春瑞[1]，李 佳[2]
(1. 中国工程物理研究院计算机应用研究所，绵阳 621900；2. 中国工程物理研究院总体工程研究所，绵阳 621900)

摘 要：P2P 蠕虫是利用 P2P 机制进行传播的恶意代码。通过 P2P 节点的共享列表，蠕虫很容易获得攻击目标的信息，所以其爆发时传播速度很快，这种大量的快速传播导致的直接后果是网络阻塞。该文分析蠕虫在 P2P 网络中的传播原理，在经典病毒传播模型基础上提出了考虑带宽及治愈响应起始时间因素的蠕虫传播模型，从带宽饱和与阻塞两个方面分析带宽对蠕虫传播的影响，在此基础上分析了蠕虫的防御措施。通过模拟实验，该模型能够较真实地描述蠕虫大规模爆发时引起带宽拥塞的情况。

关键词：蠕虫；P2P 系统；传播模型；防御措施

Research on Propagation of Worms in P2P Networks

HAO Xiangdong[1], WANG Kaiyun[1], ZHANG Chunrui[1], LI Jia[2]
(1. Institute of Computer Application, Chinese Academy of Engineering Physics, Mianyang 621900;
2. Institute of System Engineering, Chinese Academy of Engineering Physics, Mianyang 621900)

【Abstract】 The P2P-based worm is a kind of malicious code that takes advantage of P2P system to propagate. Because hosts in P2P system maintain a lot of neighbors list, infected hosts in P2P systems can easily propagate the worms to its neighbors. It can spread with high speed, which would lead to network congestion. This paper addresses the issue by analyzing the theory of P2P worm's propagation and presents a mathematical model that takes into account bandwidth and the time of treatment response. In particular, it studies bandwidth affecting the worm propagation in the aspects of saturation and congestion. Furthermore, it studies the measure of defense based on the propagating model. From the simulation experiment, it can describe the process of worm's propagation in a P2P system.

【Key words】 Worms; Peer-to-peer (P2P) system; Propagation model; Defense measure

随着P2P应用软件日益流行，产生了利用其共享机制传播的蠕虫。在P2P环境中，每个节点都存有大量的邻居节点列表，只要有一个节点被蠕虫感染，该系统中所有的对等节点都将处于被感染的危险中[1]；同时，有些P2P客户端软件本身就存在着安全漏洞，例如最近著名的P2P应用软件KaZaA被指出存在着缓冲器溢出的漏洞。蠕虫自身传播机制的多变性与网络的复杂性为防御蠕虫增加了难度。因此，设计一个能够较为准确地描述蠕虫扩散的传播模型，对蠕虫的研究以及防御工作有重要意义。本文通过分析蠕虫在P2P网络中的传播原理，用数学模型描述了蠕虫的传播过程，并着重分析了利用P2P通信机制进行扩散的蠕虫传播特点。

1 相关研究工作

蠕虫的传播具有传染病的特点，所以目前主要以传染病传播模型为基础来研究蠕虫的传播。最经典的蠕虫传播模型是SI模型[2]，它将节点分为易感染和被感染两类，利用线性微分方程来描述蠕虫的传播过程，SI模型能反映网络蠕虫初期传播行为，但不能准确描述网络蠕虫中后期的传播状态；邹长春等提出的双因素模型[4]考虑了动态的感染率和蠕虫对抗因素，是SI模型的扩展。在P2P蠕虫传播研究方面也取得了一些进展，Giuseppe等人在文献[3]中提出了一个基于自治域带宽因素的传播模型；文献[5]从P2P系统参数和蠕虫参数两个方面来分析蠕虫的传播，并从随机探测、离线、在线 3 个状态考虑蠕虫的传染情况。但是，这些传播模型基本上是针对某一特定蠕虫的传播特点而建立，考虑带宽因素对蠕虫传播的影响研究较少。本文将从带宽与治愈响应起始时间以及P2P蠕虫自身的传播机制方面研究蠕虫的传播，并重点从带宽饱和阻塞两个方面来研究蠕虫的传播。

2 P2P 原理与蠕虫传播关系

目前流行于Internet中的蠕虫主要传播方式是随机扫描，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，并且它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结点，同时要求用很高的速度探测不同的主机[1]。而所谓的P2P蠕虫就是能够优先利用P2P通信机制传播的蠕虫，目前蠕虫呈现出随机扫描与利用P2P机制混合传播的趋势。P2P网络的一个主要特点是资源共享，每个节点都保存着大量的邻居节点信息。利用P2P的共享机制，蠕虫只需要感染其中的一个节点，就可以通过结点中邻居结点的信息列表迅速探测新的易攻击结点，整个过程重复迭代，蠕虫迅速蔓延整个网络。总结起来这类蠕虫有 3 个主要的特点[1]：(1)传播速度极快；(2)传播成功率很高；(3)攻击不易被察觉。

由于 P2P 网络拓扑结构的特点，为 P2P 蠕虫准确地寻找易攻击节点提供了捷径，因此其攻击不具有网络异常性。

3 P2P 蠕虫传播模型

理想的网络蠕虫传播模型能够充分地反映蠕虫的传播行为，要考虑带宽、节点差异性等因素对网络蠕虫传播链的影

基金项目：中国工程物理研究院面上基金资助项目(20050657，20060651)
作者简介：郝向东(1978－)，男，硕士生，主研方向：计算机网络与信息安全；王开云，研究员；张春瑞、李 佳，硕士生
收稿日期：2006-04-25 **E-mail：**fenghe2166@tom.com

响，同时要能够预测网络蠕虫可能带来的威胁。在蠕虫传播模型的研究中，大多数是以随机扫描为传播方式的前提下建立的，而针对P2P网络中蠕虫的传播模型研究较少。需要说明的是，本文的分析是建立在蠕虫连续传播的前提下，因为有些蠕虫本身具有时间潜伏性(如cord-red)，对于这样的蠕虫仅能对其活动的时段进行描述[4]。

3.1 经典蠕虫传播模型 SI

SI(susceptible-infections)模型是描述蠕虫以随机扫描方式传播的经典模型，从它派生出很多其它的传播模型。SI 模型中将节点分为易感染节点(S)和已感染节点(I)，具体的形式如下：

定义整个网络中的易感染的节点总数为 N；在 t 时刻已经被感染的节点为 I(t)；在 t 时刻易被感染的节点数为 S(t)；s(t)= S(t)/N 代表节点在 t 时刻的易感染率；i(t)=I(t)/N 代表节点在 t 时刻的已感染率；K 代表感染节点的平均传染率。

则有

$$dI = KISdt \tag{1}$$

因为，$\frac{dI}{N} = di$，S+I=N，所以，式(1)变为

$$\frac{dI}{dt} = KI(N - I) \tag{2}$$

对微分方程式（2）求解得

$$i(t) = \frac{e^{KN(t-T)}}{1 + e^{KN(t-T)}} \tag{3}$$

式(3)表示的是蠕虫随时间被感染的比例，其中 T 表示最大感染比例的时刻。尽管该模型没有考虑带宽、拓扑等系统因素，但还是能够描述蠕虫大量爆发初期的情形，详细的模型推导过程见文献[3]。

3.2 基于 SI 模型的 P2P 蠕虫传播模型

由于P2P蠕虫不需要在网络中随机地探测易攻击节点，它只需要根据节点中邻居节点的信息列表便能很容易地获得攻击目标的信息，因此蠕虫爆发具有突然性，并且目前很多蠕虫采用UDP传输协议[6]（如Slammer），蠕虫探测易感染节点会耗费大量的带宽，使得网络的带宽成为影响蠕虫扩散的一个重要参数。另外，在之前的蠕虫传播模型中，均未考虑治疗的响应时间的影响，本文在SI模型基础上，从带宽和治疗响应时间两个方面建立蠕虫的传播模型。

3.2.1 含带宽因素的传播模型

蠕虫的蔓延造成网络拥塞主要来自两部分：一部分是扫描易攻击目标，即使是P2P蠕虫最开始也需要探测到大量的攻击目标才能进行大规模传播，并且探测的过程会耗费很多带宽，如一个Slammer蠕虫在 1s用于探测所产生的流量大约是 13.7Mbps[6]；另外一部分来自蠕虫副本在机器间的传递引起的流量。假设蠕虫的探测速度为R_s，单次探测所产生的流量为S_s，则单个蠕虫探测所产生的流量为R_sS_s，蠕虫在一个区域探测所产生的流量为R_sS_sI/Mbps，传播蠕虫副本产生的流量为sKI/Mbps。

目前的网络多为星型和树型的结构，我们以网络中的交换机为边界将网络分成若干个区域，整个互联网可以看成由众多的这样区域组成。同一区域蠕虫的传播不足以使通信阻塞，但是不同区域之间的通信将成为网络的瓶颈。以一个三级交换网络为例，假设同一层次中每个交换机的性能是相同的，每个交换机所连接的机器数也是相同的。假设该网络中总机器数是N，其中共有M台交换机，则每个区域的机器数是N/M。两个交换机间的允许通过的最大带宽是B，这里B的值是引起大量丢包的最小值。最底层的交换机称为第 3 级交换机，顶层的为第 1 级交换机。按照实际的情况，设各级交换机之间的带宽为 1 000Mbps，交换机模块内的交换带宽为100Mbps。假设每个二层交换机下面连 10 台三级交换机，每个三级交换机连 16 台PC机。则当三级交换机每个模块的吞吐量均达到饱和状态时，它与二级交换机的通信量接近于 80 Mbps～90 Mbps，那么此时二层交换机总的负荷达到 800 Mbps～900 Mbps，这时二级交换机接近于饱和状态；当“供给”超过了交换机的“需求”时将导致二层交换机拥塞，使得三级与二级交换机间无法通信，但是三级交换机的内部模块间通信一般不会受影响。为了不失一般性，用Q（t）表示各区域间允许通过的带宽比例，用Q_1代表带宽阻塞的情况；用Q_2代表带宽达到饱和状态时的通过率。N，S(t)，I(t)，s(t)，i(t)等仍然引用 3.1 节中定义的参数，具体的参数说明见表 1。

表 1 本文模型中定义的参数

N	易感染的总节点数目
S(t)	t 时刻面临感染的节点
I(t)	t 时刻已经被感染的节点数
s(t)	t 时刻易感染率
i(t)	t 时刻被感染率
ρ	被感染节点治愈率
K	单个感染节点的平均传染率
s	感染一台机器耗费的带宽大小
T_a	治疗响应起始时间
B	网络允许的带宽上限
Q（t）	区域间允许通过的带宽比率
R_s	一个蠕虫探测易攻击主机速率
S_s	探测一次所产生的流量
M	网络包含的区域数

定义带宽阻塞时的通过率为

$$Q_1(t) = \frac{1}{1 + e^{(\frac{I(R_sS_s + Ks)}{M} \times \frac{M-1}{M} - B)}}$$

进一步简化得

$$Q_1(t) = \frac{1}{1 + e^{(b-B)}}$$

其中，$b = \frac{I(R_sS_s + Ks)}{M} \times \frac{M-1}{M}$，即为一个区域蠕虫向外传播所产生的总带宽。对于P2P蠕虫来说，探测所产生的开销很小，但其传播速度却很快。容易看出，当b＜＜B时，$Q_1(t)$=1，这时蠕虫传播几乎不受带宽影响；当b＞＞B时，$Q_1(t)$=0，这时网络已经处于阻塞状态。

定义带宽饱和时的通过率为

$$Q_2(t) = \frac{B}{b + (B-b)\frac{1}{1 + e^{(b-B)}}} = \frac{B}{b + (B-b)Q_1}$$

当b＜＜B时，$Q_1(t)$=1，这时蠕虫传播几乎不受带宽影响；当b＞＞B时，$Q_1(t)$=B/b，这时网络处于饱和状态，仅有B/b的流量可以通过。

$$\begin{cases} Q(t) = xQ_1 + (1-x)Q_2 \\ x \in [0,1] \end{cases} \tag{4}$$

式(4)中，x 是网络产生阻塞情况的概率，1-x 是网络处于饱和时的概率。根据以上的分析以及 SI 模型，得到在一个区域内的感染数变化为

$$\frac{dI}{dt} = IK(N-I)[\frac{1}{M} + \frac{M-1}{M}Q(t)] \tag{5}$$

其中 1/M 代表蠕虫在区域内部传播信，不受带宽的影响；(M-1)/MQ(t)代表在其它 M-1 个区域内的通信。当蠕虫传播引起的流量很小时，几乎不受带宽影响；当蠕虫传播发生的流

量超过了额定带宽时，Q（t）开始起作用，蠕虫的传播效率开始下降。

3.2.2 含治疗响应时间的传播模型

一种新的蠕虫在爆发初期，一般是很难迅速采取措施杀掉或打补丁进行预防的。为了反映真实的传播情况，我们设定一个时间响应起始点，当蠕虫爆发时间达到该起始点时，开始对蠕虫进行处理。根据这个思想得出如下微分方程：

$$n = NIK(N-I)dt - \Phi(t)\rho NIdt \tag{6}$$

该式表示在 dt 时间内被感染的节点数是 n，其中ϕ（t）是时间响应函数，ρ为感染节点的平均治愈率。假设 N 是连续的，则有 n=d(Ni)=Ndi，代入式(6)化简整理为

$$\begin{cases} \dfrac{dI}{dt} = IK(N-I) - \Phi(t)\rho I \\ \Phi(t) = \begin{cases} 1, t \ge T_a \\ 0, t < T_a \end{cases} \end{cases} \tag{7}$$

其中 Ta 是时间响应起始，根据蠕虫的种类与具体的网络不同而变化，当时间 t<Ta 时，式(7)与 SI 模型相同；当时间 t>Ta 时，治愈因子开始起作用。

如果将式(5)、式(7)综合，即得到了同时考虑带宽和治疗响应时间的传播模型：

$$\begin{cases} \dfrac{dI}{dt} = IK(N-I)[\dfrac{1}{M} + \dfrac{M-1}{M}Q(t)] - \Phi(t)\rho I \\ \Phi(t) = \begin{cases} 1, t \ge T_a \\ 0, t < T_a \end{cases} \end{cases}$$

4 模拟与结果分析

为检验模型的正确性，建立模拟程序分别对 SI 模型、加入响应时间的模型以及加入带宽因素的模型进行了模拟，并将模拟结果相互比较。为了便于比较，基本参数使用文献[3,6]中的数据。

4.1 带宽对传播的影响

由于蠕虫利用 P2P 机制进行传播，不用扫描探测易攻击目标，因此导致其传播速率很快。图 1 是不同传播速率下的模拟曲线。通过曲线可以看出，当速率越大时，在同一时刻被感染的机器数量越多，达到带宽饱和的时间也越短。

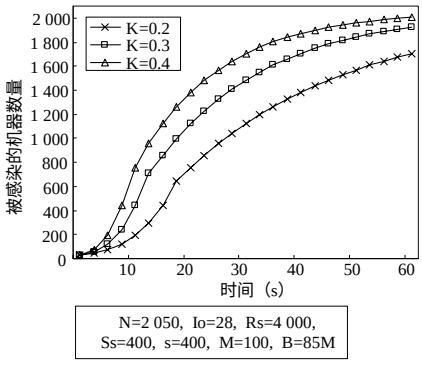

图 1 P2P 环境下蠕虫传播曲线

图 2 是考虑带宽因素的模型模拟曲线。从图中可以看出除了阻塞的情况外，另外的 3 条曲线最后都将趋于易感染节点，因为，虽然带宽处于饱和状态，但是还没有阻塞，这时蠕虫仍然可以传播，只是这时的传播效率下降。

如果网络完全阻塞，区域间无法进行通信，从图中可以看出这时蠕虫几乎停止传播，但是，在一个区域内部仍然有可能通信，所以曲线的趋势是缓慢增长的。和文献[6]中的 slammer 真实观测数据作比较，本文的带宽阻塞情况接近于实际的观测结果。

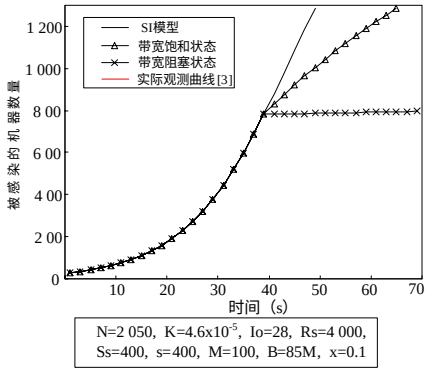

图 2 考虑带宽因素的传播模型与观测数据比较曲线

4.2 治愈响应时间的影响

图 3 是加了治愈因子的模拟曲线，从图 3 可以看出响应时间起始 Ta=60 时，当达到响应起始后，感染率迅速下降；当响应时间起始 Ta=40 时，感染率平滑地下降。产生该现象的原因是由于在蠕虫爆发初期，其传播速度很快，在这个时刻开始治疗能够牵制蠕虫的迅速增长趋势，虽然从曲线上看整体的治愈效果没有在 Ta=60 时明显，但整体被感染的百分比控制在较低的范围内，因此这也说明在蠕虫爆发初期开始采取治疗措施是最有效的。

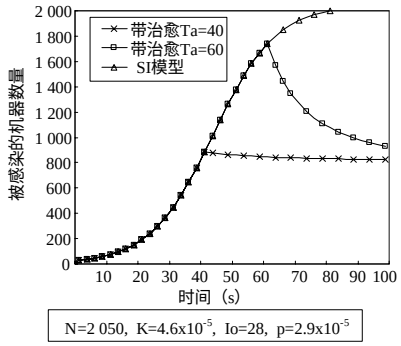

图 3 考虑治疗响应起始时间传播模型

4.3 带宽和治愈的综合影响

同时考虑带宽与治愈的因素将会更接近于实际的蠕虫传播情况，为此我们进行了模拟，实验的结果如图 4 所示。

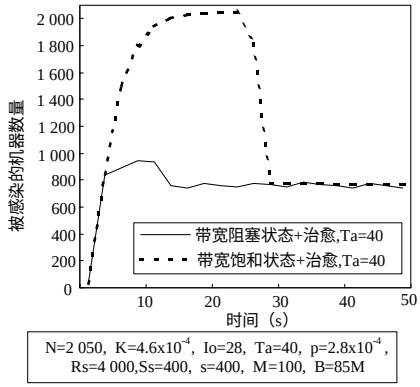

图 4 考虑带宽与治疗响应时间因素的传播模型曲线

从图中可以看出当达到一定的时间后，被感染的机器数量开始上下波动，这是因为超过治疗响应时间后，有一部分蠕虫被治愈，Q(t)开始变大，带宽拥塞的状态被缓解，蠕虫可以继续进行向外传播；如果带宽饱和后没有完全阻塞，则被感染的机器数量会缓慢增加，这个过程中整个链路很容易

(下转第 147 页)