P2P文件共享网络中被动蠕虫传播建模与分析

冯朝胜[1,3]，秦志光[1]，劳伦斯·库珀特[2]，罗瑞莎·托卡库克[2]

(1. 电子科技大学计算机科学与工程学院 成都 610054; 2. 伦敦大学玛丽皇后学院 伦敦 E1 4NS;
3. 四川师范大学计算机科学学院 成都 610066)

【摘要】为了抑制P2P被动型蠕虫在大规模P2P文件共享网中的传播和攻击，对P2P网络中文件共享的特点和被动型蠕虫传播的特点进行了深入的分析，并在此基础上提出了3个分别适用于蠕虫传播不同阶段的传播模型。为了证明模型的正确性，进行了大规模仿真实验。使用专门的数值分析工具求出相应模型的理论值。为了仿真大规模P2P文件共享网，专门基于P2P仿真平台开发出能真实模拟流行P2P文件共享网的软件并基于该软件进行了仿真。理论值与仿真值匹配的事实表明，该模型基本上反映了蠕虫的传播情况，可以用来预测蠕虫的传播趋势和行为。

关 键 词 文件共享； 被动型蠕虫； P2P网络； 传播模型； 仿真
中图分类号 TP274.1 **文献标识码** A doi: 10. 3969/j. issn. 1001-0548. 2009. 02. 25

Propagation Modeling and Analysis of Passive Worms in Peer-to-Peer File-Sharing Networks

FENG Chao-sheng[1,3], QIN Zhi-guang[1], LAURENCE CUTHBET[2], and LAURISSA TOKARCHUK[2]

(1. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 610054;
2. Queen Mary College, University of London London E1 4NS; 3. School of Computer Science, Sichuan Normal University Chengdu 610066)

Abstract To counter the attack and propagation of passive worms in large-scale peer-to-peer(P2P) file-sharing networks, three propagating models of passive worms suitable for different stages of worm propagation, are proposed based on deep analysis on the features of file sharing and passive worm propagation. In order to verify the validity of the three modes, large scale simulating experiments are carried out. To simulate the true popular P2P file-sharing networks, a simulating software is developed based on the P2P simulating platform PeerSim. Theory analysis and simulation show that these models are valid and can predict the tendency and behaviors of worm propagation.

Key words file sharing; passive worms; P2P networks; propagating models; simulation

P2P网络中每个节点都具有均衡负载能力，能充分挖掘网络的计算能力，解决C/S结构的服务瓶颈问题。如今，数以百万计的因特网用户通常都使用大规模文件共享P2P网络(如BitTorrent等)上传和下载文件。据统计，eDonkey2000网络在任意时间都有200万用户同时在线，而使用过BitTorrent的网络用户早已超过千万，成为最受欢迎的文件传输网络之一。

伴随着P2P文件共享网络的迅猛发展，该网络的安全问题日益突出。P2P文件共享网络中的文件可能被置入恶意代码，用户一旦下载并执行含有恶意代码的文件必然会被感染。在P2P文件共享网络中，用户一旦下载并执行感染文件，在该用户的共享文件夹中就会增加大量的感染文件，而这些感染文件通常都被冠以最受欢迎文件的文件名，从而大大增加了感染的概率[1-2]。

本文主要对P2P文件共享网络以及其上的被动型蠕虫进行研究，构造出能较准确反映和预测被动型蠕虫传播行为模式的数学模型，为验证将来提出的蠕虫遏制和免疫方法的有效性提供方法。

1 背景和相关工作

1.1 P2P文件共享网络

为了更准确地提出被动型蠕虫的传播模型，先讨论P2P文件共享型网络的特点。在BitTorrent 和eDonkey2000这样的网络中，每个用户都有一个共享文件夹，用户将所有可共享的文件存放到共享文件夹以便其他用户共享，网络中的任何用户都可以从其他任意一个用户的共享文件夹中下载文件。当用

收稿日期：2008－01－14；修回日期：2008－12－02
基金项目：国家自然科学基金(60473090)，国家自然科学基金与英国皇家科学会合作项目(60711130232)
作者简介：冯朝胜(1971－)，男，博士生，主要从事信息安全方面的研究.

户想要下载某个文件时，他会发出搜索文件请求。在BitTorrent中通过Tracker服务器处理搜索文件请求；在Gnutella中，通过邻居不断转发搜索请求的形式搜索文件。无论哪种P2P文件共享网络，请求文件用户最终都会收到与请求相匹配的文件列表。尽管不同的网络生成文件列表的方式有所不同，但生成的文件列表都是满足用户文件请求的所有P2P主机的一部分。获取了文件列表后，用户可以从列表中选择一个或多个主机下载该文件。从多个主机上下载文件，被称为多点下载，意味着每个主机都提供文件的一部分。文件下载后被存放于共享文件夹，可供网络中其他主机下载[3]。

P2P文件共享网络中，文件一经下载马上就可共享这一特点为用户共享文件带来了极大方便，但也为被动型蠕虫的传播留下了可乘之机，一些利用该特点进行传播的蠕虫已经出现，如Duload和Gotorm这样的被动型蠕虫，在其依附的文件被执行时会在共享文件夹中生成固定个数的感染文件，这些感染文件在所有被感染的主机上有着相同的文件名称。攻击力更强的被动型蠕虫如Darby和Benjiamin，则从一个很大的命名域中选取名称。

1.2　相关研究工作及进展

文献[4]指出P2P网络具有非常适合蠕虫传播的特性。文献[5]进行了利用漏洞感染P2P网络中逻辑邻居的蠕虫传播仿真实验，揭示了P2P蠕虫的主动攻击性和强大的感染能力。文献[6]提出P2P蠕虫概念，并对P2P蠕虫的未来发展做了预测，认为P2P蠕虫可以采用更加主动的方式进行传播，利用P2P网络拓扑可以令P2P蠕虫的传播更隐蔽、攻击更精确。

关于P2P蠕虫传播模型的建立与分析，文献[7]提出了利用传统计算机病毒传播模型研究P2P蠕虫传播的方法；文献[1]对非扫描型P2P蠕虫进行仿真分析；文献[8]对各种扫描策略下蠕虫的传播性能进行仿真分析；文献[9]利用数字模拟的方法分析P2P系统参数对被动型P2P蠕虫传播的影响；文献[10]提出主动型P2P蠕虫的传播模型及Matlab仿真分析方法；文献[11]基于结构化对等网路由表构造方法，建立P2P蠕虫在Chord、CAN、Pastry这3种典型结构化对等网中的传播模型，给出刻画P2P蠕虫传播能力的函数，并揭示了覆盖网拓扑对蠕虫传播的负面影响。文献[12]对BT网络中主动型蠕虫的传播进行仿真分析。到目前为止，大量研究工作都是针对P2P主动型蠕虫的，而针对被动型蠕虫的研究很少。

2　P2P被动型蠕虫传播建模

由于蠕虫在不同的传播阶段有不同的传播特点，本文将分阶段建立蠕虫传播模型，建模基于平均场法。

2.1　建模参数和假设

建模的目的是为了预测P2P被动型蠕虫在P2P网络中的传播趋势和行为。为了简化建模，作如下假设：(1) 网络中在线的用户数量没有发生变化；(2) 文件一旦下载马上被执行；(3) 主机被感染(包括搜索、连接、下载和执行)的时间是不变的，该时间被称作一个时间单元(time unit)。主机由感染状态恢复为易感状态或免疫状态花费一个时间单元时间；(4) 一台主机一旦被感染，将在共享文件夹中生成c个文件，所有的感染主机共享同样的c个文件名称；(5) 建模时考虑的文件都是可执行文件，包括被压缩的可执行文件，但非可执行文件(如媒体文件)不考虑。为了便于蠕虫的建模分析，建模时需要的参数设置如下：

$N(t)$：t个时间单元后网络中的主机台数，在本文中该值不随t变化，$N(0)$=10 000。

$S(t)$：t个时间单元后的易感主机数，$S(0)$=9 950。

$I(t)$：t个时间单元后的感染主机数，$I(0)$=50。

$R(t)$：t个时间单元后的免疫主机数，$R(0)$=0。

$K(t)$：t个时间单元后的感染文件数，$K(0)$=500。

$M(t)$：t个时间单元后未感染的文件数，$M(0)$=47 300。

$h(t)$：t个时间单元后下载感染文件的概率为$h(t)=\alpha\dfrac{K(t)}{M(t)+K(t)}$。

λ_{d}：每个时间单元内每台主机下载文件的平均个数，$\lambda_{\mathrm{d}}=0.02$。

λ_{is}：每个时间单元内恢复为易感主机的感染主机数，$\lambda_{\mathrm{is}}=0.001$。

λ_{sr}：每个时间单元内被免疫的易感主机数，$\lambda_{\mathrm{sr}}=0.002$。

λ_{ir}：每个时间单元内被免疫的感染主机数，$\lambda_{\mathrm{ir}}=0.001$。

c：执行了下载的感染文件后在共享文件中增加的感染文件数，c=10。

2.2　SI模型

在SI模型中，P2P网络中的主机被分成易感主机和感染主机。易感主机又被称为健康主机，在其共享文件夹中没有任何感染文件。但是，易感主机一

旦下载并执行感染文件就会被感染，并在其共享文件夹中生成c个感染文件，主机状态转移为$S \to I$。

在一个存在感染主机的P2P文件共享型网络中，当易感主机下载文件时，可能下载的是感染文件。从直觉上推断，下载感染文件的概率应该与网络中感染文件数成正比。在t个时间单元后，网络中的文件总数为未感染文件和感染文件之和为$M(t)+K(t)$，则下载感染文件的概率为：

$$h(t)=\alpha\frac{K(t)}{M(t)+K(t)}$$

式中 参数$\alpha\in[1,0]$与用户掌握的病毒知识相关，用户掌握的病毒知识越多，该值越小。在本文的实验中，该参数的值取为1，表示用户缺乏病毒知识。在一个时间单元内，一台易感主机下载的文件数为λ_{d}，而下载感染文件的概率为$h(t)$，该易感主机被感染的概率为$\lambda_{\mathrm{d}}h(t)$、变化率为$-\lambda_{\mathrm{d}}h(t)S(t)$。显然，感染主机的变化率为易感主机变化率的相反数。由于一台易感主机被感染后将增加c个感染文件，所以感染文件的变化率为$c\lambda_{\mathrm{d}}h(t)S(t)$。根据以上分析，得到的SI传播模型为：

$$\begin{cases}\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}=-\lambda_{\mathrm{d}}h(t)S(t)\\ \dfrac{\mathrm{d}I(t)}{\mathrm{d}t}=\lambda_{\mathrm{d}}h(t)S(t)\\ \dfrac{\mathrm{d}K(t)}{\mathrm{d}t}=\lambda_{\mathrm{d}}h(t)S(t)c\\ \dfrac{\mathrm{d}M(t)}{\mathrm{d}t}=\lambda_{\mathrm{d}}N(1-h(t))\end{cases}$$

2.3 SIS传播模型

在SIS传播模型中，主机的状态转移过程为$S\to I\to S$。当一台主机上的所有的染毒文件都被用户删除后，该主机恢复为易感主机。λ_{is}为一个时间单元内易感主机的恢复率，网络中共有$\lambda_{\mathrm{is}}I(t)$台主机被恢复。结合SI模型的分析，得到的SIS模型为：

$$\begin{cases}\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}=-\lambda_{\mathrm{d}}h(t)S(t)+\lambda_{\mathrm{is}}I(t)\\ \dfrac{\mathrm{d}I(t)}{\mathrm{d}t}=\lambda_{\mathrm{d}}h(t)S(t)-\lambda_{\mathrm{is}}I(t)\\ \dfrac{\mathrm{d}K(t)}{\mathrm{d}t}=c\lambda_{\mathrm{d}}h(t)S(t)-c\lambda_{\mathrm{is}}I(t)\\ \dfrac{\mathrm{d}M(t)}{\mathrm{d}t}=\lambda_{\mathrm{d}}N(1-h(t))\end{cases}$$

2.4 SIR传播模型

与SI模型和SIS模型不同，在SIR模型中，每个主机处于易感的、感染的和免疫的3种状态中的一种。一台主机的状态转移过程为$S\to I\to R$。与SIS模型不同，在SIR模型中，一部分感染主机被免疫而不是恢复成易感状态。当一台易感主机被免疫，意味着该主机上所有的感染文件被删除，之后再也不会被感染。在每个时间单元，易感主机和感染主机的免疫比例分别为λ_{sr}和λ_{ir}，则免疫主机的变化率为$\lambda_{\mathrm{sr}}S(t)+\lambda_{\mathrm{ir}}I(t)$，感染文件的变化率为$c\lambda_{\mathrm{ir}}I(t)$。于是，SIR模型对应的模型为：

$$\begin{cases}\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}=-\lambda_{\mathrm{d}}h(t)S(t)-\lambda_{\mathrm{sr}}S(t)\\ \dfrac{\mathrm{d}I(t)}{\mathrm{d}t}=\lambda_{\mathrm{d}}h(t)S(t)-\lambda_{\mathrm{ir}}I(t)\\ \dfrac{\mathrm{d}R(t)}{\mathrm{d}t}=\lambda_{\mathrm{sr}}S(t)+\lambda_{\mathrm{ir}}I(t)\\ \dfrac{\mathrm{d}K(t)}{\mathrm{d}t}=c\lambda_{\mathrm{d}}h(t)S(t)-c\lambda_{\mathrm{ir}}I(t)\\ \dfrac{\mathrm{d}M(t)}{\mathrm{d}t}=\lambda_{\mathrm{d}}N(1-h(t))\end{cases}$$

如上所述，上面3种模型分别适用于蠕虫传播的3个不同阶段。在传播初期，P2P网络用户没有意识到感染文件的存在，所以主机状态转移为$S\to I$。随着感染主机的增多和蠕虫对感染主机正常运行造成影响，用户察觉到感染文件的存在并将其从感染主机删除，感染主机恢复为易感主机(SIS模型反映了这种情况)。用户手工清除染毒文件难度很大，在感染发展到一定程度时，更多的用户采用打补丁或升级反病毒软件的方法使主机免疫(SIR模型反映了这种情况)。在SIR模型中，由于恢复率远远小于免疫率，恢复的情况被忽略。

3 仿真实验和结果分析

3.1 实验说明

流行的P2P文件共享网络通常由数万个节点构成，在这样的网络中直接验证本文提出的蠕虫传播模型是否有效是不现实的，只有通过仿真实验验证模型的正确性。为了根据数学模型求出理论值，使用数值分析工具Matlab中的Simulink工具。为了仿真蠕虫在P2P文件共享网络中的传播，专门在P2P仿真平台PeerSim上开发出了具有典型文件共享P2P网络特点的仿真软件。在仿真之前首先要通过配置文件为感染主机数、感染文件数、免疫率等变量和参数赋初值。对于每个模型，仿真实验各做20次，将20次实验的平均值作为仿真实验的值。为了便于比较理论值和仿真值，将实验值和理论值在平面直角坐标系中进行对比。鉴于论文篇幅有限，本文只展示部分实验结果。

3.2 实验结果和分析

Simulink求出的蠕虫传播模型的理论值与仿真软件仿真出的仿真值的对比如图1～图3所示。在3个图中，理论值形成的曲线和仿真值形成的曲线匹配得较好，其他大量仿真实验的结果也如此。

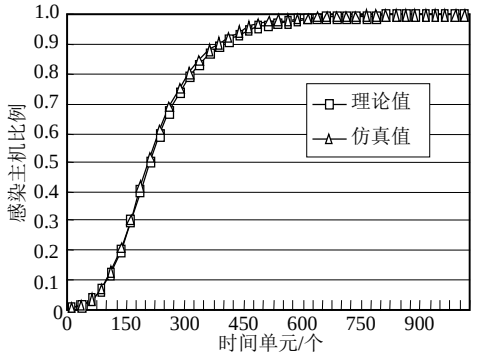

图1　SI模型感染主机的理论值与仿真值的对比

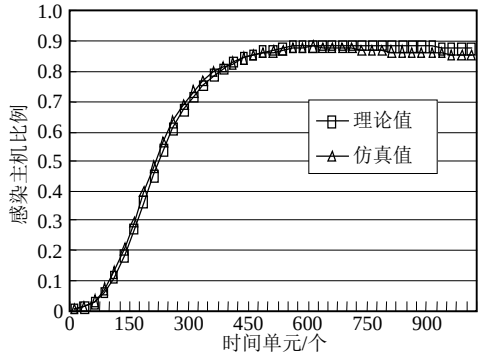

图2　SIS模型感染主机的理论值与仿真值的对比

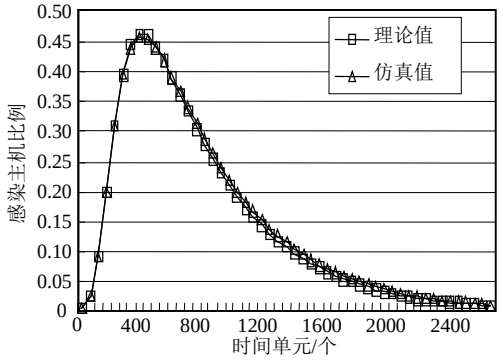

图3　SIR模型感染主机理论值与仿真值的对比

3种模型对应的仿真结果的对比如图4所示。

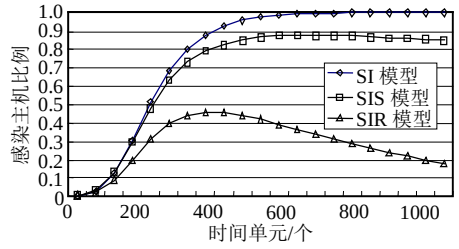

图4　3种模型对应的仿真值的比较

图4中的3条曲线中：(1) SI模型的曲线增长得最快，所有的节点都很快被感染；(2) SIS模型的曲线增长得要慢些，感染节点达到一定数量后，就处于平稳状态，这主要是因为有一定比例的感染节点返回到易感状态，感染节点返回比例越大，曲线增长得越慢。(3) SIR模型曲线是先增长，达到峰值后开始下降。曲线峰值的大小主要由免疫率决定，免疫率越大，峰值越小，被感染的节点越少。

4 总结与展望

本文主要对对P2P文件共享网构成巨大安全威胁的被动型蠕虫进行研究。(1) 对P2P文件共享网络的特点和被动型蠕虫传播的研究情况进行介绍；(2) 在对被动型蠕虫进行深入分析的基础上，提出了分别适用于蠕虫传播的初期、中期和后期的3个被动型蠕虫传播模型；(3) 为验证所提出模型的正确性，进行了大规模的仿真实验。实验值与理论值相当接近的事实表明，提出的模型是有效的，可以用于预测被动型蠕虫在P2P网络中的传播趋势和行为。今后，还要对模型进行改进，使其也能适用于P2P网络规模动态变化的情况。

本文研究工作得到四川师范大学重点项目(07ZD018和08KYL03)的资助，在此表示感谢。

参 考 文 献

[1] CHEN Guan-ling, ROBERT S G. Simulating non-scanning worms on Peer-to-Peer networks[C]//Proceedings of the 1st International Conference on Scalable Information Systems. Hong Kong, China: ACM, 2006.

[2] STUTZBACH D, REJAIE R, SEN S. Characterizing unstructured overlay topologies in modern P2P file-sharing systems[C]//Proceedings of the Fifth ACM Internet Measurement Conference. Berkeley, CA: ACM, 2005: 49-62.

[3] STUART S, VERN P, NICHOLAS W. How to own the internet in your spare time[C]//Proceedings of the 11th USENIX Security Symposium. San Francisco, CA: USENIX, 2002.

[4] KANNAN J, LAKSHMINARAYANAN K. Implications of Peer-to-Peer networks on worm attacks and defense[R]. 2003.

[5] ZHOU Li-dong, ZHANG Lin-tao, MCSHEERY F. A first look at Peer-to-Peer worms threats and defenses[C]// Proceedings of Peer-to-Peer Systems 4th International Workshop (IPTPS). New York: Springer, 2005: 24-35.

[6] WEI Yu. Analyze the worm-based attack in large scale P2P networks[C]//Proceedings of the 8th IEEE International Symposium on High Assurance Systems Engineering. Tampa, Florida: IEEE, 2004: 308-309.

(下转第273页)