

一种面向文件片段选择策略的对等网络污染方法

孔劼, 蔡皖东, 丁军平

(西北工业大学计算机学院, 710129, 西安)

摘要: 为了提高对等网络(P2P)文件共享过程的控制效果,提出了一种面向 P2P 文件共享系统中文件片段选择策略的 P2P 污染方法,并通过数学建模对污染效果进行了分析. 该方法利用 P2P 文件共享系统中最少优先的文件片段选择策略,通过干扰文件片段的副本数达到控制文件传播的目的. 实验结果表明,污染是否成功由初始资源提供节点退出 P2P 文件共享系统的概率和系统内污染节点的数量相互作用来决定,退出率和污染节点数量越高,文件片段污染越容易成功. 当污染失效时,节点完成整个文件下载所用时间近似服从 Frechet 分布. 当污染成功时,该方法可以在污染者不发送文件数据的情况下,使整个文件共享系统中每个资源请求节点都无法完成文件下载,因此具有污染效果好、对网络与计算资源要求低的优点.

关键词: 对等网络;污染;文件片段选择策略

中图分类号: TP393.08 **文献标志码:** A **文章编号:** 0253-987X(2012)03-0058-06

Pollution for Selection Strategy of File Segment in Peer-to-Peer System with Modeling and Analyzing

KONG Jie, CAI Wandong, DING Junping

(Department of Computer, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract: To promote the effect of peer-to-peer (P2P) file sharing control, a pollution method pointing to the selection strategy of file segment in P2P file-sharing system is proposed, and the mathematical model is established to analyze the pollution effect. The rarest first policy of file segment selection in P2P system is used to pollute the quantity of the file segment copies. The simulation demonstrates that the success of pollution depends on the mutual action of the exit rate of the initial data provider and the quantity of the polluter. The higher the exit rate of the initial data provider and the quantity of the polluter are, the higher the success rate of pollution is. If the pollution fails, the downloading time of the file is close to Frechet distribution. While the pollution succeed, without sending any data of a file, the polluter makes every downloading node in the P2P system fail to accomplish the file downloading. The proposed pollution method is effective requiring less computing resources.

收稿日期: 2012-02-08. 作者简介: 孔劼(1981—),男,博士生;蔡皖东(通信作者),男,教授、博士生导师. 基金项目: 高新技术研究发展计划资助项目(2009AA01Z424);西北工业大学基础研究基金资助项目(JC201149).

网络出版时间: 2012-02-08

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20120208.1122.0>

对等网络(P2P)污染是一种用于控制特定文件在 P2P 文件共享系统中传播的文件下载控制技术,最早被美国音像出版界用于控制盗版音像制品通过 P2P 文件共享系统的非法传播。2005 年,Liang 等人^[1]对当时已发现的 P2P 污染给出了详细的描述,与此同时又派生出多种 P2P 污染技术。Dhungel 等人^[2]介绍了被称为 Fake-Block Attack 和 Uncooperative Peer Attack 的内容污染方法。Liang 等人^[3]介绍了非结构化的 FastTrack 系统与结构化的 Over-net 系统中的索引污染,并对污染效果进行了测量分析。Yoshida 等人^[4]采用索引污染对 Winny 系统中的文件共享进行了控制。

由于建立大规模 P2P 污染实验环境的难度较大,而在真实的 P2P 网络中进行污染实验又会涉及到法律和道德问题,因此目前学术界对于 P2P 污染的常见研究方法进行数学建模。Dumirtiu 等人^[5]通过建立数学模型对 P2P 文件污染进行了分析,结果表明,影响污染的因素包括 P2P 文件共享协议特性、网络拓扑、节点的抗污染策略、用户行为等。胡辛遥^[6]提出了污染统一模型,将内容污染与索引污染结合起来进行建模分析。方群等人^[7]基于古典概率模型创建了文件污染模型。

本文面向 P2P 文件共享系统中的文件片段选择策略,提出了一种被称为文件片段污染的 P2P 文件污染方法,它不同于现有的 P2P 污染方法,通过对文件片段副本数进行干扰,文件片段污染可以在污染节点不发送文件数据的情况下,使整个文件共享系统中每个节点都无法完成文件下载。此外,本文还通过建立的数学模型对该方法的污染效果进行了分析和评估。

1 面向文件片段选择策略的 P2P^污

1.1 P2P 文件共享系统的文件片段选择策略

P2P 文件共享系统的一个特点是在进行数据传输时,将文件分割成若干个小的文件片段进行传输,使参与 P2P 文件共享任务的节点在下

载文件时,需要解决文件片断的选择问题。以 BitTorrent 系统为例,基本的文件片段选择策略为最少优先策略。下载者在选择下一个将要开始下载的文件片断时,通常选择当前文件共享系统中副本数量最少的文件片断。最少优先策略确保了每个结点都拥有其他节点最希望得到的那些文件,也确保了将那些副本数量较多的片断放在最后下载,从而减少了某个节点当前提供下载,而随后却没有任何被别人感兴趣的文件片断这种事件发生的概率^[8]。在 eMule 系统中,也优先采用类似的文件片段选择策略^[9]。

1.2 文件片段污染的原理

在 P2P 文件共享系统中,各文件片段的数据来自若干个初始资源提供节点和大量的资源请求节点。文件由文件片段组成,当系统内某个文件片段的副本数为 0 时,资源请求节点将无法完成整个文件的下载。文件片段污染的基本思想正是利用了最少优先策略来影响文件的充足率,使资源请求节点无法完成文件下载。节点在加入 P2P 文件共享系统时,会向系统声明自己拥有的文件片段,同理,在文件片段污染中,污染节点在加入文件共享任务时,会向 P2P 文件共享系统声明自己只拥有一个文件片段。重复加入若干个拥有相同文件片段的污染节点,将使该文件片段的副本数大于其他文件片段,这个过程称为对该文件片段的副本数进行干扰。令文件片段 i 的副本数为

$$C_i = \begin{cases} S + M_i & i \neq j \\ S + M_i + A & i = 1, 2, 3, 4, 5, \dots, n \end{cases}$$

式中: S 为拥有全部文件片段的资源提供的节点数; M_i 为系统中拥有文件片段 i 的资源请求的节点数; A 为干扰文件片段 j 的污染节点数。

根据最少优先策略,加入文件共享任务的资源请求节点,将优先下载未受到污染节点干扰的文件片段。因此,在一段时间内,整个文件共享任务的数据传播都集中在未受到干扰的文件片段上。P2P 文件共享系统中存在着一种被称为搭便车(Free-Riding)的现象。由于 P2P 系统中不存在

对节点行为的监控或统计机制,因此随着用户数的增加,在人类自私和利己心理的作用下,存在某些只下载不上传,或在下载完成后很快退出文件共享任务,不继续作为资源提供者向其他节点分享数据的节点.学术界对搭便车现象的研究表明,该现象在 P2P 文件共享系统中广泛存在.因此,在实际的 P2P 文件共享任务中,资源提供节点都存在一定的退出率.由于污染节点对文件片段副本数的干扰,在一段时间内数据的复制集中在未受干扰的文件片段上,而受到干扰的文件片段的副本数并未增加,即 $M_j=0$.在这段时间内,若资源提供节点按照一定的退出率全部退出文件共享任务,即 $S=0$,由于污染节点的副本数是虚假的,对于文件片段 j ,可用的副本数 $C_j=S+M_j=0$,因此文件共享系统内的其他资源请求节点将无法完成文件下载.

2 文件片段污染的数学模型

模型通过对 P2P 文件共享系统内各节点状态在不同时刻的模拟来反映污染的变化过程,在某一时间单位内,各节点以顺序方式执行模拟.被共享的文件 F 由 M 个文件片段组成,令 F_i 表示文件 F 的第 i 个文件片段,则有 $F=F_1 \cup F_2 \dots \cup F_M$. 模型内的节点类型分为资源提供节点、污染节点和资源请求节点.资源请求节点在完成所有的文件片段下载后将转为资源提供节点,为了便于描述,将这种资源提供节点称为后继资源提供节点.资源提供节点会以某一退出率退出系统,而污染节点加入系统后则一直存在.为了便于模拟,模型假设所有节点具有相同的网络带宽,下载一个文件片段所需的时间也相同.模型中节点的下载事件和退出系统事件用概率事件发生器进行模拟,概率事件发生器是一个函数,执行若干次的概率事件发生器,其运行成功的次数与指定的概率值相对应.模型的模拟过程分为若干轮,每一轮的模拟周期中对系统内所有的节点依次进行模拟.每个模拟周期相当于真实的 P2P 系统中一段相同的时间长度,模拟过程中轮次的增加反映了真实系统中时间的增加.

当第 t 轮的第 j 个资源请求节点执行模拟

时,系统内所有节点的总数满足以下关系

$$N_j(t) = M_j(t) + S(t) + F_j(t) + A \quad (1)$$

$$S(t) = \begin{cases} 10 & t = 0 \\ S(t-1) - \sum_{k=1}^{S(t-1)} (1^k \times H(P_1)) & t > 0 \end{cases} \quad (2)$$

式中: $S(t)$ 为初始资源提供节点的数量; $M_j(t)$ 、 $F_j(t)$ 分别为资源请求节点及后继资源提供节点的数量; H 为概率事件发生函数; P_1 为初始资源提供节点退出系统的退出率,若初始资源提供节点退出系统的事件发生,则 $H(P_1)$ 的值为 1,事件未发生,则函数值为 0.第 t 轮的第 j 个资源请求节点执行模拟时,文件片段 i 的副本数为

$$C_{i,j}(t) = S(t) + F_j(t) + C_{R,i,j}(t) + A \quad (3)$$

$$C_{R,i,j}(t) = \sum_{j=B}^{J-1} C_{N,i,j}(t) + \sum_{j=J+1}^E C_{N,i,j}(t-1) \quad (4)$$

式中: $C_{R,i,j}(t)$ 为资源请求节点拥有的文件片段副本总数; $C_{N,i,j}(t)$ 为资源请求节点 j 在第 t 轮是否已完成文件片段 i 的下载,默认值为 0,完成下载后置为 1; B 为第 t 轮的第 1 个资源请求节点编号; E 为第 t 轮的最后一个资源请求节点编号; J 为当前正在进行模拟的资源请求节点编号.

若第 t 轮的第 j 个资源请求节点处于请求 i 的状态,则该节点与能够提供文件数据的节点建立连接的概率为

$$P_{S,i,j}(t) = (C_{i,j}(t) - A) / N_j(t) \quad (5)$$

模型中产生变量 $M_j(t)$ 、 $F_j(t)$ 、 $C_{N,i,j}(t)$ 的离散事件的算法伪代码如下.

```

 $M_T = F_T = 0;$ 
for( $t=0; t < \text{模拟时间}; t++$ )
{
 $M_T = M_T + \lambda;$ 
for( $j=0; j < M_T; j++$ )
{
if(已完成全部文件片段下载)
{
 $F_T = F_T + 1; F_j(t) = F_T;$ 
 $M_T = M_T - 1; M_j(t) = M_T;$ 
if( $H(P_T)$ )
{
 $F_T = F_T - 1; F_j(t) = F_T;$ 
}
}
}
}

```

```

else if(节点处于下载状态)
{
    下载计时器自减 1;
    if(计时器为 0)
    {
        将节点置于空闲状态;  $C_{N_{i,j}}(t) = 1$ ;
    }
}
else
{
    选择未完成下载的文件片段中副本数最少的片段;
    if( $H(P_{S_{i,j}}(t))$ )
    {
        将节点置于下载状态; 将下载计时器置为  $T$ ;
    }
}
}

```

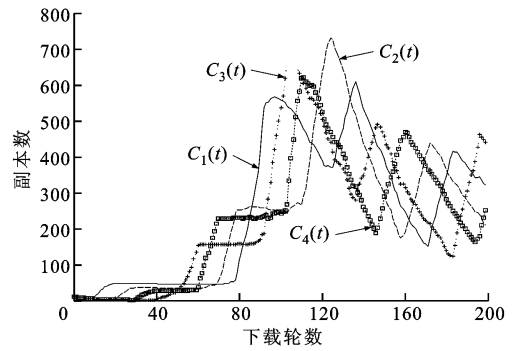
其中 M_T, F_T 为存放 $M_j(t), F_j(t)$ 值的临时变量, P_T 为后继资源提供节点退出系统的概率, λ 为每轮模拟开始时新加入系统的资源请求节点的数量, 即资源请求节点到达系统的到达率服从参数为 λ 的泊松分布, T 为下载一个文件片段所需下载的轮数。

3 仿真实验

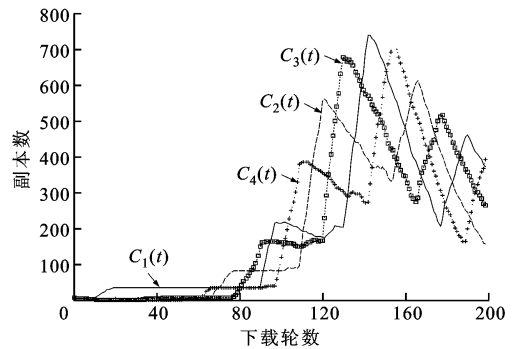
实验的硬件采用 AMD 公司的双核速龙 3800+ 处理器, 内存为 2 GB, 硬盘为 1 TB. 软件采用 Windows XP SP3, Matlab 7.0, Visual C++ 6.0. 默认设置为: 实验迭代次数 200 次, 即对模型的模拟为 200 轮; 考虑缩短实验时间, 令共享文件由 4 个文件片段数组成, 污染节点对文件片段的干扰集中在第 1 个文件片段; $\lambda = 10, S(t) = 10, A = 100, T = 10, P_1 = 10\%, P_T = 10\%$.

当第 t 轮的最后一个资源请求节点的模拟完成后, 文件片段 i 的副本数为 $C_i(t)$. 当 P2P 文件共享系统中不存在污染时, P_1 的取值对 $C_i(t)$ 的影响如图 1 所示. 由图 1 可知, 在最少优先策略的作用下, 4 个文件片段的副本数随着时间的推移交替增加, 当一个文件片段的副本数明显超过其他文件片段的副本数时, 在接下来一段时间内下载将集中于其他文件片段, 使

这些文件片段的副本数快速增加, 系统内各文件片段的副本数保持基本均衡。



(a) $P_1 = 10\%$



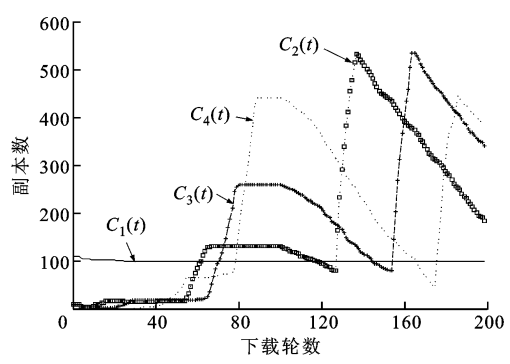
(b) $P_1 = 1\%$

图 1 无污染时 P_1 对 $C_i(t)$ 的影响

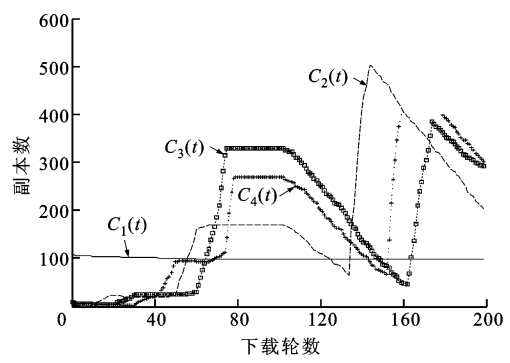
当 P_1 的取值分别为 10%、1% 时, 整个实验过程中分别有 1 093、1 368 个资源请求节点完成了整个文件的下载. 这些资源请求节点的平均下载轮数分别为 69.52、66.03. 由此可知, 在无污染状况下, 不同 P_1 值对系统内资源请求节点的下载状况影响不显著. 在 P2P 文件共享系统引入文件片段污染后, P_1 的取值对 $C_i(t)$ 的影响如图 2 所示.

图 2a、2b 显示了在污染节点的作用下, 系统内文件片段 1 的副本数受到干扰, 使系统中资源请求节点无法正常完成下载的过程. 图 2c、2d 反映了污染失效的情况. 在实验开始时刻, $C_1(t)$ 为 110, 其中属于污染节点的副本数为 100, 属于初始资源提供节点的副本数为 10. 初始资源提供节点以概率 P_1 退出系统, 使实验开始后 $C_1(t)$ 逐步减少. 由于受到污染节点的

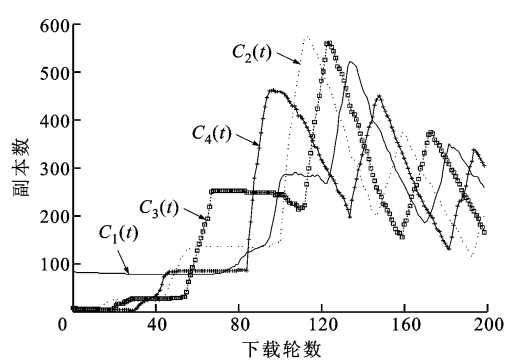
干扰,在 $C_2(t)$ 、 $C_3(t)$ 、 $C_4(t)$ 与 $C_1(t)$ 持平之前,系统内的资源请求节点在最少优先策略的作用下,下载请求都集中在文件片段 2、3、4 上。如图 2a、2b 所示,当 $C_2(t)$ 、 $C_3(t)$ 、 $C_4(t)$ 都超过 $C_1(t)$ 时,资源请求节点已经全部退出文件共享系统,即 $C_1(t)$ 为 100,则文件片段污染成功。此



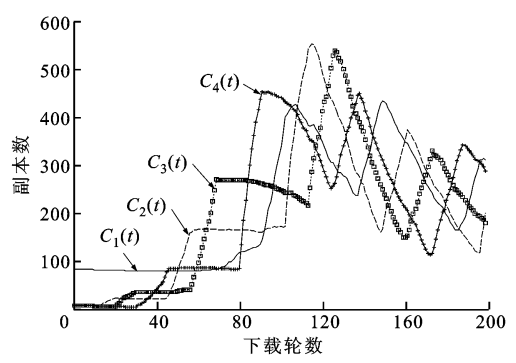
(a) $P_1=10\%$



(b) $P_1=5\%$



(c) $P_1=2\%$



(d) $P_1=1\%$

图 2 文件片段污染时 P_1 对 $C_i(t)$ 的影响

资源请求节点无法完成整个文件的下载。如图 2c、2d 所示,当 $C_2(t)$ 、 $C_3(t)$ 、 $C_4(t)$ 都超过 $C_1(t)$ 时,资源请求节点尚未全部退出文件共享系统,根据最少优先策略,下载请求将集中于文件片段 1 上,使 $C_1(t)$ 迅速提高,文件片段污染失效。此外,由图 2 还可发现,随着 P_1 的减少,初始资源提供节点全部退出系统的时间也随之推迟。如图 2a、2b 所示,在实验过程中完成整个文件下载的节点数为 0,在与图 2c、2d 相对应的实验中,完成整个文件下载的节点数分别为 1 347 和 1 370,平均下载轮数分别为 66.58 和 65.89。

图 3a、3b 反映了提高污染节点数对污染干扰效果的影响。在与图 3a、3b 相对应的实验中,污染节点的数提高至 200,在实验过程中完成整个文件下载的节点数分别为 0、1 223,图 3b 中节点的平均下载轮数为 72.10。对比图 3 与图 2c、2d 可知,提高污染节点的数量可使 $C_3(t)$ 超过 $C_1(t)$ 发生的时间推迟,例如在图 3a 中,大约推迟了 20 轮,资源提供节点在模拟时间内全部退出系统,因此污染成功。在图 3b 中,初始资源提供节点在推迟的时间段内仍然未能全部退出系统,因此污染失效。

综合上述实验结果可得出以下结论,在其他参数一定的情况下,文件片段污染成功与否,是由 P_1 和 A 相互作用决定的。 P_1 和 A 的取值越高,文件片段污染越容易成功。当污染失效时,不同 P_1 和 A 的取值对于系统内完成整个文件下载的节点数、平均下载时间的影响不显

时,系统内文件片段 1 不存在可用的真实副本,

著.

在文件片段污染失效的情况下,当 $P_1 = 2\%$ 时,节点完成整个文件下载所用时间的直方图如图 4 所示. 根据直方图的形状,可判断资源请求节点完成全部文件片段下载所用时间近似服从 II 型极值分布,即 Frechet 分布. 对图 4 的实验结果进行参数估计的概率密度函数为

$$f(t) = \frac{0.3181}{10.7469} \left(\frac{10.7469}{t} \right)^{1.3181} \cdot \exp \left(- \left(\frac{10.7469}{t} \right)^{0.3181} \right) \quad (6)$$

其中形状参数值为 0.3181, 尺度参数值为 10.7469, 位置参数值为 56.7017.

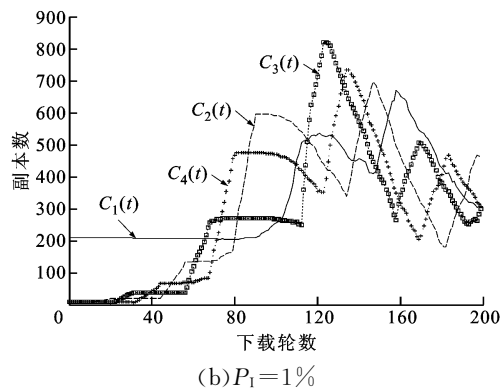
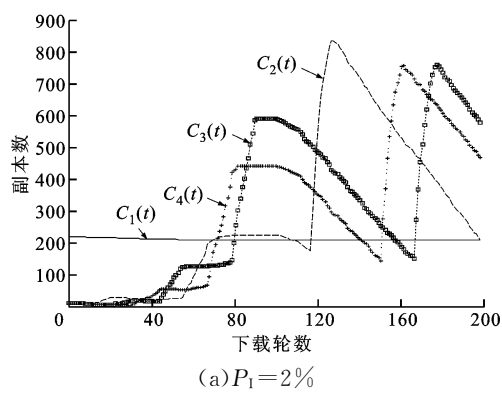


图 3 P_1 对 $C_i(t)$ 的影响 ($A=200$)

4 结 论

本文提出了一种针对文件片段选择策略的 P2P 污染方法,该方法利用 P2P 系统中文件片段选择策略的漏洞对 P2P 文件传播过程进行

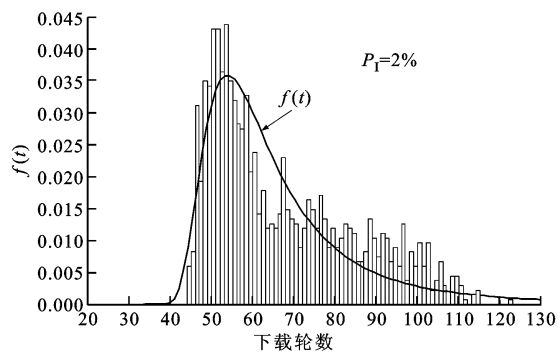


图 4 节点完成整个文件下载所用时间的直方图

控制. 利用该方法建立了数学模型,通过仿真实验对污染的效果进行分析,结果表明,文件片段污染成功与否,是由初始资源提供节点退出 P2P 文件共享系统的概率和系统内污染节点的数量相互作用决定的,退出率和污染节点数越高,文件片段污染越容易成功. 当污染成功时,本文方法可以在污染者不发送文件数据的情况下,使整个文件共享系统中每个资源请求节点都无法完成文件下载,因此具有污染效果好、对网络与计算资源要求低的优点. 当污染失效时,节点完成整个文件下载所用时间近似服从 Frechet 分布,并且所有资源请求节点都能完成文件下载.

参考文献:

- [1] LIANG J, KUMAR R, XI Y, et al. Pollution in P2P file sharing systems [C] // Proceedings of IEEE INFOCOM 2005. Piscataway, NJ, USA: IEEE, 2005:1174-1185.
- [2] DHUNGEL P, WU D, SCHONHORST B, et al. A measurement study of attacks on BitTorrent leechers[EB/OL]. [2011-01-20]. <http://www.iptps.org/papers-2008/47.pdf>.
- [3] LIANG J, NAOUMOV N, ROSS K W. The index poisoning attack in P2P file sharing systems [C] // Proceedings of IEEE INFOCOM 2006. Piscataway, NJ, USA: IEEE, 2006: 1-12.
- [4] YOSHIDA M, OHZAHATA S, NAKAO A, et al. Controlling file distribution in Winny network through index poisoning [C] // Proceedings of ICOIN 2009 International Conference on Information Networking. Piscataway, NJ, USA: IEEE,

2009;1-5.

- [5] DUMITRIU D, KNIGHTLY E, KUZMANOV-IC A, et al. Denial-of-service resilience in peer-to-peer file-sharing systems [C]//Proceedings of ACM SIGMETRICS 2005. New York, USA: ACM, 2005;38-49.
 - [6] 胡辛遥. 对等网络污染的建模 [D]. 上海: 上海交通大学, 2008.
 - [7] 方群. P2P 文件污染随机模型 [J]. 小型微型计算机系统, 2009, 30(10):1980-1984.
- FANG Qun. Stochastic modeling of file pollution in P2P system [J]. Journal of Chinese Computer

Systems, 2009, 30(10):1980-1984.

- [8] 王珏,周莉. BitTorrent 模型原理分析 [J]. 华东交通大学学报, 2009, 26(1):82-83.
- WANG Jue, ZHOU Li. An analysis on model principle of BitTorrent [J]. Journal of East China Jiaotong University, 2009, 26(1):82-83.
- [9] KULBA Y, BICKSON D. The eMule protocol specification [EB/OL]. [2011-07-10]. <http://code.google.com/p/emule-xtreme/downloads/>.

(编辑 管咏梅)