第36卷 第20期 计 算 机 工 程 2010年10月
Vol.36 No.20 Computer Engineering October 2010
·开发研究与设计技术· 文章编号：1000—3428(2010)20—0253—03 文献标识码：A 中图分类号：TP391

基于文件污染的BT文件传播控制技术

赵秋实，蔡皖东，孔 劼
(西北工业大学计算机学院，西安 710129)

摘 要：提出一种基于文件污染的BitTorrent(BT)文件传播控制方法。采用文件索引污染与数据污染相结合的方式，针对特定的下载任务，干扰其BT网络节点之间连接的建立及数据的传输。实验结果表明，该方法可以有效延缓甚至破坏BT网络中特定信息的传播，从而达到控制BT文件传播的目的。

关键词：BitTorrent网络；索引污染；数据污染

BT File Transmission Control Technique Based on File Pollution

ZHAO Qiu-shi, CAI Wan-dong, KONG Jie
(College of Computer, Northwestern Polytechnical University, Xi'an 710129, China)

【Abstract】 This paper presents a BitTorrent(BT) file transmission control method based on file pollution. The method uses combination of file index pollution and data pollution to disturb the establishment connections of the peers and the transmission of data for special download tasks. Experimental results show that this method can efficiently prolong the transmission time or destroy the information exchange and achieve the aim of transmission control.

【Key words】 BitTorrent(BT) network; index pollution; data pollution

1 概述

目前，P2P技术已经成为最重要的网络应用之一，该技术的应用使互联网内容得以快速传播。虽然从P2P技术诞生到现在只有几年时间，但是它已占据了Internet中超过一半的带宽资源[1]。BitTorrent(BT)协议是P2P技术中应用最广泛的协议。在整个P2P流量中，80%以上是BT流量[2]。BT网络已经成为了一个主流的资源共享平台，它的匿名性在保护用户通信隐私的同时也使用户之间可以共享其资源。这种自由性势必使得音乐、电影、游戏和软件等数字产品的版权难以得到有效的保护。同时，用户所共享的资源也有可能涉及机密信息、色情暴力内容和危害国家安全的言论。因此，阻止BT网络中非法信息的传播就显得尤为重要。

目前，对BT网络的控制方法主要采用BT封堵技术。基于内容的BT封堵技术需要识别BT数据内容，因而处理速度受到限制，并且对加密的BT数据包无能为力。与内容无关的BT封堵技术禁止包括合法行为在内的所有BT行为，控制粒度较粗。本文针对以上方法的不足，提出基于文件污染的BT文件传播控制方法，通过索引污染与数据污染相结合，对BT节点之间建立连接和下载数据行为进行控制。

2 BT网络模型及其控制现状

2.1 BT简介

一个BT传输模型由以下5个实体构成：共享的资源文件，种子文件，Web服务器，Tracker服务器，用户BT客户端(抽象为Peer)。其中，资源文件被分割为若干片段(Piece)，一个Piece又被分为若干数据块(Block)，Block是Peer之间信息传输的最小单元；种子文件是根据共享的资源文件，经过BEncoding(又称为B编码)生成后缀名为.torrent的文本文件，包含Tracker服务器信息和共享资源文件信息2部分。一个种子文件对应一个BT任务，Web服务器用来发布种子文件，Tracker服务器与Peer通信，维护BT任务信息和节点信息列表(Peer List)[3]。

加入BT网络的Peer先从Web服务器上下载某一BT任务的种子文件，解析该文件，获得Tracker服务器相关信息。然后，Peer与Tracker服务器通信。Tracker服务器返回给Peer共同参与该任务的节点的子集Peer List。Peer与该表中包括种子节点Seed(拥有完整资源文件的Peer)在内的其他Peer握手，申请他所没有并感兴趣的某个Piece，并共享他所拥有的Piece。通过这种方式完成信息的交互[4]。

2.2 BT网络控制现状

目前对BT网络的控制主要采取封堵的方法。封堵BT行为的方法大多采用封锁BT常用端口、封锁Tracker服务器或者识别BT流信息过滤掉该数据流。这种方法控制粒度较粗，禁止所有BT行为包括合法的BT行为，如在线视频、视频会议和合法的BT文件分发等。基于BT内容的封堵方法可以控制特定信息在BT网络中的传播。但是该方法需要识别BT包内容，处理速度受限，并且对加密的BT数据无法识别，控制范围受到极大的影响。

文件污染是BT网络中对特定信息传播控制的另一种方法。该方法选择特定内容的资源文件作为污染对象，制造“假冒版本”(污染版本)的种子[5]。该版本一般是加入了噪声而不

基金项目：国家“863”计划基金资助项目(2009AA01Z424)
作者简介：赵秋实(1984－)，男，硕士研究生，主研方向：网络与信息安全；蔡皖东，教授、博士生导师；孔 劼，博士研究生
收稿日期：2010-05-26 **E-mail：**qiushizhao@hotmail.com

能正常使用的资源文件或者是无关的垃圾文件。由于用户只有下载完成后才能检验文件的真实性，并且下载同时也在上传，因此最终网络中的污染版本数量大大超过真实版本，使得该内容变得不可用。

3 BT 下载控制技术

3.1 索引污染

索引污染主要用于控制 BT 的 Peer 之间连接建立过程。BT 连接建立控制的基本过程如图 1 所示。

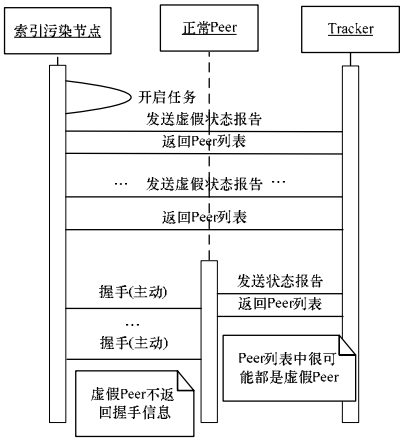

图 1 连接控制基本过程

索引是指 Tracker 服务器维护的 Peer List 中 Peer 信息的索引。在 BT 网络中，Tracker 服务器将 Peer 报告的状态信息态添加到 Peer List 中，而不检查该状态的真实性。索引污染利用 Tracker 服务器的这一特点注册大量的虚假 Peer 信息。这些虚假信息包括不存在的 Peer IP 地址和端口号，以及宣称完整的拥有某资源。当正常 Peer 下载该资源时，首先需要与 Tracker 服务器通信，从服务器得到其他 Peer 信息，而服务器返回的是它所维护的 Peer List 的一个子集，该子集是随机的，很可能不包含正常 Peer 信息而只有虚假的 Peer 信息。如果虚假索引足够多，那么用户很可能连接几次都失败，随后放弃下载，即使连接成功，连接上的 Peer 也会比正常情况下少很多，同时耗费大量的时间建立连接。

3.2 数据污染

数据污染主要用于控制 BT 的文件下载过程。BT 协议规定，资源文件被划分的最小单元是 Block，一个 Block 大小为 16 KB，若干个 Block 组成一个 Piece，大小可以根据用户要求调整但必须为 16 KB 的整数倍，通常为 256 KB。下载资源文件时，下载节点从不同的 Peer 处下载一个 Piece 的不同 Block 部分，当获得一个 Piece 的所有 Block 后，按顺序将他们组装成完整的 Piece，通过 SHA1 计算该 Piece 的哈希值并与种子文件中该 Piece 的哈希值比较，若相同说明得到了正确的 Piece，不同则重新下载该 Piece[6-7]。数据污染原理见图 2。

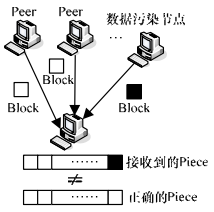

图 2 数据污染原理

采用数据污染方法控制 BT 下载过程时，污染者传送虚假的 Block 数据，下载节点从正常 Peer 和污染节点处都得到 Block，它们所组成的 Piece 必然是错误的，从而被丢弃。数据污染通过这种方法消耗下载节点的带宽，从而延缓资源的下载速度。此外，一些 BT 客户端为提高下载效率规定若从同一个 Peer 获得的 Piece 的哈希值连续出错，则该 Peer 为可疑节点，将该 Peer 放入黑名单，停止从该节点接收数据。而客户端无法分辨错误是哪一个 Block 造成的，因而正常节点也会被放入黑名单，影响下载。

3.3 BT 连接与下载的综合控制

鉴于索引污染与数据污染的特征，本文设计了一种索引与数据污染相结合的 BT 文件传播控制系统，通过综合运用索引污染与数据污染，使其兼具索引污染与数据污染的优点，实现更有效的控制。系统结构如图 3 所示。

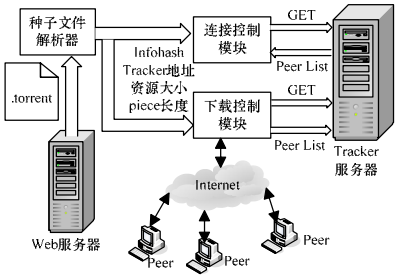

图 3 索引与数据污染综合控制系统结构

系统首先从 Web 服务器上下载与资源文件相关的.torrent 文件，根据 B 编码规则解析种子文件，得到该 BT 任务的 infohash 值、Tracker 服务器地址、资源文件大小和 Piece 长度等信息，将这些与下载任务相关的信息提交给连接控制模块和下载控制模块。索引污染节点在 Tracker 服务器中注册大量的虚假 Peer 信息。在 BT 协议中，Peer 和 Tracker 服务器之间的通信采用 HTTP/HTTPS 协议。Peer 向 Tracker 服务器发 HTTP GET 请求报文，用以注册 Peer 信息。索引污染构造大量虚假 Peer ID、IP 和 Port 的 GET 请求以 HTTP 报文发送给 Tracker 服务器。当接收到 Tracker 返回的 Peer List 后注册成功，索引污染完成，连接控制结束。

与此同时，数据污染也与 Tracker 服务器通信注册节点信息，宣称拥有完整的资源文件，以等待正常 Peer 连接。当连接建立后，双方进行“BT 握手”。握手成功之后 Peer 之间就以循环的消息流进行通信。Peer 向污染节点请求数据，污染节点回送相应的虚假 Block，与真实 Block 组合，得到错误的 Piece，整个 Piece 被丢弃。这样便形成了对 BT 下载的控制。综合控制过程如图 4 所示。

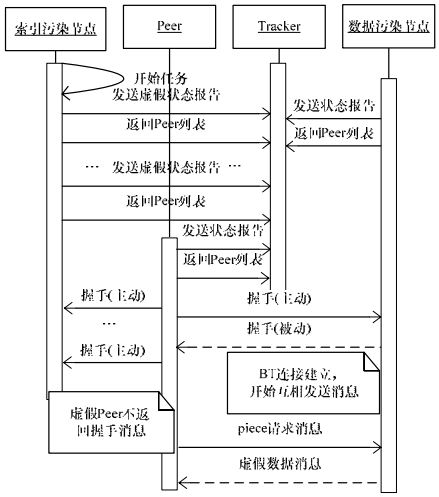

图 4 综合控制过程

4 实验及结果分析

4.1 实验环境

将 Tracker 服务器、污染节点和正常 Peer 都部署在以太局域网内。污染节点与正常 Peer 随机的部署在子网中。Tracker 服务器选用 BitCometTracker_0.5 版，正常 Peer 客户端选用 BitComet1.11 版。选取资源文件 Piece 大小为默认的 256 KB，共 700 个。

4.2 实验结果

(1)实验 1：索引污染分别以正常节点数目的 100 倍、200 倍、400 倍和 800 倍向 Tracker 服务器注册虚假索引，正常节点之间建立连接时长如表 1 所示。

表 1 连接时长对比

污染倍数	污染前/s	污染后/s	延时/s
100	9	47.9	36.9
200	9	246.0	237.0
400	9	1 227.9	1 218.9
800	9	>3 600.0	>3 600.0

由实验 1 可以看出，随着虚假索引倍数的增加连接建立的平均时长也在不断增长。这是由于 Tracker 服务器返回给 Peer 的节点信息是其维护的 Peer List 的 1 个子表，且子表长度不超过 200，随着污染力度的加大，子表中正常节点存在概率大大降低。

(2)实验 2：污染前正常下载完成用时 17.5 min。数据污染 Peer 与正常 Peer 数目分别以 1:1 和 2:1 的比例进行，选取 7 个 Peer 作为采样点，实验结果如表 2 所示。

表 2 数据污染实验结果

N	1:1		2:1	
	完成率/(%)	用时/min	完成率/(%)	用时/min
1	100.0	29.0	100.0	19.8
2	100.0	31.5	52.0	>60.0
3	100.0	41.3	66.3	>60.0
4	96.4	>60.0	23.9	>60.0
5	95.7	>60.0	11.3	>60.0
6	40.5	>60.0	66.6	>60.0
7	37.0	>60.0	70.5	>60.0

由实验 2 可以看出，对于下载完成的节点都有延迟，但延迟效果不均匀，有的节点可以完成下载而有的不能，不能完成下载的节点所完成的比例也不同。下载未完成是由数据污染导致误将正常节点当作可疑节点放入黑名单中所造成的。同时，可由表 2 看出随着污染节点数目与正常 Peer 数目比例的加大，未完成下载的节点数增多。

(3)实验 3：索引污染力度选取 100 倍，数据污染选取污染节点与正常 Peer 数目为 2:1，进行综合控制，选取 7 个 Peer 作为采样点，结果如表 3 所示。由实验 3 可以看出，7 个采样点在相当长的时间内都没有完成下载，而且下载完成率不高于 2%，用户会放弃下载。这种情况是由于索引污染延长了建立连接的时间，同时降低了与正常 Peer 建立连接的概率，而数据污染节点数目远大于正常 Peer 数目的增加。它们之间建立连接概率增大，将正常 Peer 误当作可疑节点放入黑名单的概率增大。由此可见，综合控制 BT 传输的方法是有效的，污染效果好于单一控制效果。

表 3 污染结果

N	完成率/(%)	用时/min
1	1.5	>60
2	1.5	>60
3	0.0	>60
4	0.1	>60
5	0.0	>60
6	0.3	>60
7	0.0	>60

5 结束语

本文讨论了 BT 传输控制的方法，针对 BT 网络的文件下载特点，采用索引污染与数据污染结合的方法，延缓甚至阻止特定资源文件在 BT 网络中的传播；讨论了对 Peer 之间建立连接的控制，以及对已建立连接的 Peer 数据下载的控制，2 种控制方法互为补充、相辅相成。实验结果表明了综合控制方法的可行性与有效性。

参考文献

[1] 陈贵海, 李振华. 对等网络: 结构、应用与设计[M]. 北京: 清华大学出版社, 2007.

[2] 欧阳荣, 雷振明. BitTorrent 带宽模型研究[J]. 计算机科学, 2007, 34(9): 55-57.

[3] 龙柏炜, 阙喜戎, 王文东, 等. IP 组播在 BitTorrent 中的应用研究[J]. 计算机工程, 2010, 36(3): 118-121.

[4] Bittorrent Protocol Specification v1.0[EB/OL]. (2008-03-01). http://wiki.theory.org/BitTorrentSpecification.

[5] 左 敏, 李建华. P2P 中的文件污染与污染防治[J]. 计算机工程, 2007, 33(18): 22-25.

[6] Christin N, Weigend A S, Chuang J. Content Availability, Pollution and Poisoning in Peer to Peer File Sharing Networks[C]//Proc. of the 6th ACM Conference on Electronic Commerce. New York, USA: [s. n.], 2005.

[7] Konrath M A, Barcellos M P, Mansilha R B. Attacking a Swarm with a Band of Liars: Evaluating the Impact of Attacks on Bittorrent[C]//Proc. of the 7th IEEE International Conf. on Peer-to-Peer Computing. Galway, Ireland: [s. n.], 2007.

编辑 顾姣健

~~~~~~

(上接第 252 页)

[7] Fursin G, Cohen A. Building a Practical Iterative Interactive Compiler[C]//Proc. of the 1st Workshop on Statistical and Machine Learning Approaches Applied to Architectures and Compilation. Ghent, Belgium: [s. n.], 2007.

[8] 陈文光, 杨 博, 王紫瑶, 等. 一个交互式的 Fortran77 并行化系统[J]. 软件学报, 1999, 10(12): 1259-1267.

[9] Sinnen O. Task Scheduling for Parallel Systems[M]. New Jersey, USA: John Wiley & Sons, Inc., 2007.

[10] Baldawa S, Sangireddy R. CMP-SIM: An Environment for Simulating Chip Multiprocessor Architectures[EB/OL]. (2009-06-12). http://www.utdallas.edu/~rxs030200/CMP-SIM.

编辑 任吉慧
~~~~~~