ISSN 1000-0054 清华大学学报(自然科学版) 2009年 第49卷 第10期 w24
CN 11-2223/N J Tsinghua Univ(Sci & Tech), 2009, Vol. 49, No. 10 http://qhxbw.chinajournal.net.cn

基于网络坐标的BitTorrent Multi-Tracker机制

林 源, 陈 阳, 张增斌, 邓北星, 李 星

(清华大学 电子工程系, 北京 100084)

摘 要: 针对BitTorrent(BT)的Multi-Tracker机制中,下载节点对Tracker服务器的随机选择策略影响了文件传输性能的问题。该文引入了网络坐标技术,提出了一种基于网络坐标的Multi-Tracker机制(mTracker)。在这种机制下,下载节点将通过网络坐标的计算选择与自己邻近的Tracker服务器进行连接,并且通过Tracker得到邻近的邻居。这种改进优化了整个BT的覆盖网络。仿真实验表明:基于网络坐标的Multi-Tracker机制使得BT的覆盖网络更有效率,有效降低了BT下载中用于下载节点与Tracker间通信以及节点之间数据传输的开销。

关键词: 因特网; BitTorrent; Multi-Tracker; 网络坐标

中图分类号: TP 393.4 文献标识码: A

文章编号: 1000-0054(2009) 10-0099-04

Network coordinate based bit torrent multi-tracker scheme

LIN Yuan, CHEN Yang, ZHANG Zengbin, DENG Beixing, LI Xing

(**Department of Electronic Engineering, Tsinghua University, Beijing 100084, China**)

Abstract: The random selection strategy used to choose Trackers for clients in the BitTorrent (BT) Multi-Tracker Scheme reduces the performance of the file transmission. This paper desorbs a Multi-Tracker scheme for BitTorrent using Network Coordinates (mTracker) to overcome this shortcoming. With this scheme, the overlay of the clients is optimized by selecting the nearest Tracker according to their locations after calculating their network coordinates. Simulations results show that the bandwidth consumed by the overlay as well as the connection cost between the clients and the trackers is significantly reduced.

Key words: Intnet; BitTorrent; Multi-Tracker; network coordinates

近些年来, Multi-Tracker[1]技术作为BitTorrent(BT)[2]的扩展,改变了传统BT中使用单一Tracker的模式,有效提高了Tracker的服务能力,在BT下载中得到普遍的应用。然而BT在为大量用户提供高效而快速的文件资源共享的同时,也占用了大量的互联网带宽[3]。所以研究如何提高Multi-Tracker下的BT性能对提高互联网服务质量和降低网络负载有重要的意义。

一个完整的BT系统主要包括Tracker、下载节点Peer和Torrent文件3个部分。

Multi-Tracker机制是基于原有BT技术的扩展,它允许多个Tracker服务器同时对相同的文件下载任务进行维护。这不仅提高了Tracker服务的可获得性,同时避免了Tracker的单点失效问题[4]。然而,在实际Multi-Tracker机制的BT系统中,下载节点从Tracker列表中选择Tracker进行连接以及Tracker向节点返回的邻居列表信息都是随机的,这种随机选择导致下载节点不能根据自身位置对Tracker和邻居进行有偏选择,影响了BT的下载性能。

网络坐标[5]是通过少量的端到端测量来预测网络距离的工具。它的理论基础是网络中的节点间的距离基本满足三角不等式的原则,所以可以通过少量的测量将网络中的节点位置映射到欧式空间中去。这样通过两个节点的网络坐标就可以预测出他们的网络距离。

Vivaldi[6]是具有代表性的分布式网络坐标系统,在互联网P2P文件共享[7]和应用层路由[8]中有着广泛的应用。为了解决BT系统中节点随机选择Tracker和邻居带来的问题,本文将网络坐标技术应用于BT系统,将Vivaldi与BT系统进行结合,提出了基于网络坐标的mTracker机制。

收稿日期: 2008-09-17

基金项目: 国家"九七三"重点基础研究项目(2007CB310806);
国家自然科学基金资助项目(60473087, 60703052)

作者简介: 林源(1986—), 男(汉), 海南, 硕士研究生。

通讯联系人: 邓北星, 副教授,
E-mail: dengbx@mail.tsinghua.edu.cn

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1 mTracker 机制的BT 系统设计

1.1 系统总体结构

本文设计的具有mTracker 机制的BT 系统,主要在原有的Multi-Tracker 机制下的BT 系统的基础上,增加了网络坐标系统和开放式数据库系统2 个模块。系统的总体结构图如图1 所示。2 个模块的主要功能分别如下:

1) 网络坐标系统模块。负责为所有的Tracker 和Peer 计算网络坐标,使得Peer 可以通过网络坐标的计算选择最近的Tracker 进行连接。

2) 分布式数据库模块。作为Tracker 发布网络坐标以及Peer 获得Tracker 坐标的平台,使得Peer 不需连接Tracker 也能通过网络坐标的计算预测到Tracker 的距离。

通过添加这2 个功能模块,Peer 对Tracker 的选择不再是随机的,而是根据自身和Tracker 的位置信息对Tracker 进行有偏的选择。相当于所有Peer 分别以离自身最近的Tracker 为中心进行聚类,每个Tracker 维护的节点列表都是同一聚类下的节点。

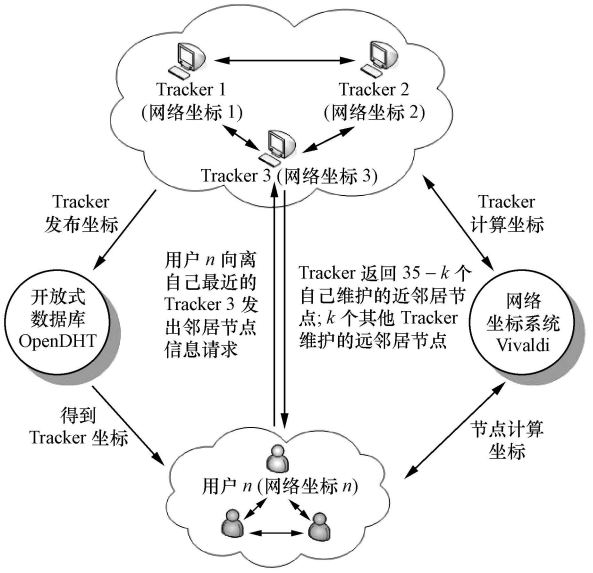

图1 系统总体结构图

1.2 Vivaldi 算法

为了将网络坐标系统加入BT 系统中,首先介绍Vivaldi 的网络坐标算法。

Vivaldi 是分布式的网络坐标系统,它将整个网络看成一个弹簧系统。假设L_{ij}是系统中节点i 和节点j 的实际距离,并且x_i 是节点i 的网络坐标。在理想情况下,所有节点的理想坐标应该使得误差公式(1)取得最小值。该误差公式的值相当于弹簧系统的总能量。

$$E = \sum_i \sum_j (L_{ij} - \| x_i - x_j \|)^2. \quad (1)$$

其中$\| x_i - x_j \|$是通过网络坐标计算出来的节点i 和节点j 之间的距离。

在实际的分布式Vivaldi 系统中,节点i 将对它当前的网络坐标x_i 和本地误差e_i 进行维护,通过多次更新使它们不断接近理想值。在每一次更新中,节点先测量它与某个邻居节点间的实际延时,然后通过测得的延时对当前的网络坐标x_i 和本地误差e_i 的值进行调整。Vivaldi 算法的步骤如下所示:

1) $w = e_i/(e_i + e_j)$,

2) $e_s = \| x_i - x_j \| - l_{ij} / l_{ij}$,

3) $e_i = e_s c_e w + e_i(1 - c_e w)$,

4) $\delta = c_c w$,

5) $x_i = x_i + \delta(l_{ij} - \| x_i - x_j \|) u(x_i - x_j)$.

其中: c_e 和c_c 是可以调整的参数, l_{ij}是系统中节点i 和节点j 的实际距离。

首先,通过当前本地节点i 的误差e_i 和邻居节点j 的误差e_j 计算出权重w; 然后计算出本地节点和邻居节点之间通过网络坐标预测出来的距离$\| x_i - x_j \|$和实际测量距离l_{ij}的相对误差e_s; 第3 步通过加权计算更新本地节点的误差e_i; 根据上面的结果,第4 步和第5 步最终算出更新后的网络坐标。

在Vivaldi 中,所有的节点都具有相同的初始坐标。当2 个节点占据相同坐标时,他们会像弹簧一样向相反的方向将对方推离。

1.3 系统的工作流程

本文设计的具有mTracker 机制的BT 系统的工作流程图如图2 所示,系统说明如下:

1) 在本文的系统中,所有节点(Tracker 和Peer)都加入到分布式的网络坐标系统Vivaldi 中分别计算自己的网络坐标。

2) Tracker 通过开放的分布式数据库OpenDHT[9] 对网络坐标进行发布。同时下载节点可以在不和Tracker 进行连接之前,事先从OpenDHT 获得所有Tracker 的网络坐标。

3) 下载节点将结合自己的网络坐标和所有Tracker 的网络坐标,计算出它到每个Tracker 的距离,然后选择最近的Tracker 进行连接并请求邻居列表。这样每个Tracker 维护的节点都是与自己相临近的下载节点。

4) Tracker 收到连接请求后,将从自己维护的

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

节点列表(近邻居列表)中随机挑选出35－k个节点，连同从其他Tracker维护的节点列表(远邻居列表)中挑选出来的k个邻居节点一同返回给下载节点。最终形成优化后的BT下载覆盖网络。

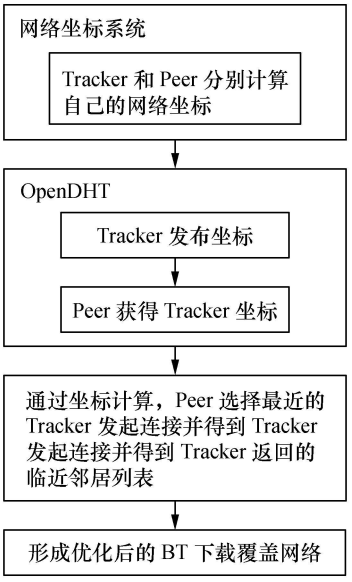

图2　**mTracker**机制的**BT**系统工作流程图

mTracker机制中引入了网络坐标系统，使得每个下载节点都可以选择离自己最近的Tracker进行连接，相当于Tracker维护的节点(向Tracker发起连接的节点)都是以Tracker为中心进行聚类的节点。这样下载节点不仅能和最近的Tracker进行连接，同时能通过Tracker获得35－k个同一聚类的近邻居节点。这样不仅节约了下载节点与Tracker之间的连接开销，同时节约了下载节点之间数据传输的开销，极大地优化了BT的覆盖网络。

2　仿真实验

2.1　仿真环境

本文通过基于实际互联网络数据的仿真方式对系统构建出的覆盖网络的性能进行评价。采用king[10]数据集作为本文的网络拓扑，这个数据集是由互联网上1 740个域名服务器两两之间的延迟所组成的1个1 740×1 740的延迟矩阵。

利用这个数据集，通过3个主要的评价标准对系统性能进行评价。在这3种性能评价标准中，分别独立进行了10次实验，在每次试验中，将从网络中随机选择了8个作为Tracker节点，其他1 732个作为下载节点。最后的数据结果是10次实验的平均结果。

2.2　性能评价标准

本文有以下3个主要的性能评价标准[11]，分别对覆盖网络整体性能，以及节点与节点、节点与Tracker之间这两种BT下载中最主要的连接开销进行评价。

相对延时代价(RDP)：定义为任意两个下载节点之间通过覆盖网络得到的路径上的延时与这两个节点在下层网络中的最短路径延时之比。它反映了覆盖网络提供的路径在延时方面的效率。

传输开销：定义为覆盖网络中所有节点间连接的平均延时。它反映了在不同网络结构中传输相同数据量消耗网络资源的程度。

连接开销：定义为覆盖网络中所有下载节点与Tracker连接的平均延时。它反映了在不同网络结构中连接Tracker消耗网络资源的程度。

2.3　仿真结果

针对k(远距离邻居数)为3、6、9、12、15、18的情况分别对系统进行了仿真。同时将其性能和传统的Multi-Tracker进行了比较。

图3主要对平均相对延时代价进行了比较，发现采用mTracker机制生成的覆盖网络引入的延时相比传统Multi-Tracker减少了35%左右，这说明了mTracker机制能改善网络在路径延时方面的效率。

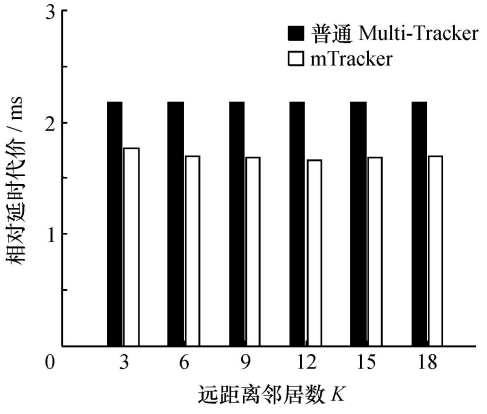

图3　平均相对延时代价对比

图4主要比较了节点间的传输开销，可以看出mTracker机制的引入可以减少节点间的传输延时。这说明传输相同的数据量，采用mTracker机制的BT比传统BT消耗更少的网络资源。

图5主要对节点连接Tracker的平均延时进行了比较。与传统Multi-Tracker机制相比，mTracker机制减少了约35%用于连接Tracker的开销，大大降低了用于连接Tracker的网络资源的

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

消耗。

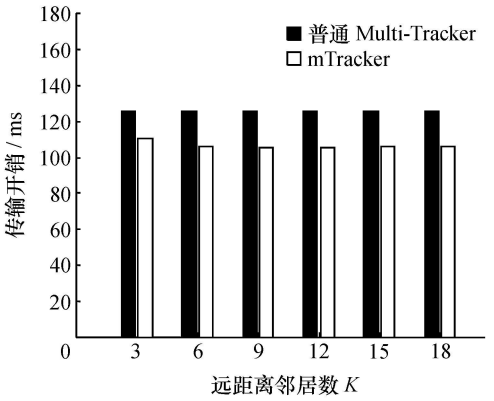

图 4 传输开销对比

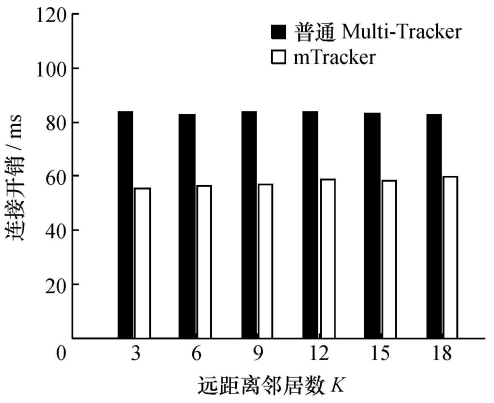

图 5 连接开销对比

通过上面的比较,发现mTracker机制的引入优化了BT下载的覆盖网络,使得节点间的数据传输开销以及节点和Tracker之间的连接开销都大大减少。极大地改善了BT的服务质量同时有效降低了BT下载对互联网带宽的消耗。

由于mTracker机制引入了网络坐标系统,节点在程序启动时首先会通过Vivaldi计算自身坐标,这会带来10s左右[6]的时间延迟,而当坐标收敛后,坐标再次进行修正的时间可以忽略。这些坐标计算都是在本地进行的,不会对网络开销造成影响。所以可以认为,引入网络坐标系统带来的延迟代价可以忽略。

3 结论及今后的工作

针对传统BT系统的不足,本文提出了基于网络坐标的mTracker机制。通过仿真,可以得到以下结论:本文提出的mTracker机制能利用节点本身的计算能力,通过计算网络坐标实现Tracker和邻居的有偏选择,优化了BT的覆盖网络,同时减少了BT下载中用于节点与Tracker间通信以及节点之间数据传输的开销。这种改进不仅提高了BT服务性能,同时降低了网络流量负载,具有很好的实际意义。

下一步的工作准备引入可扩展性更好、准确度更高的网络坐标系统,如Pharos[12],从而进一步提高网络坐标的计算精度,提供更好的BT下载服务。

参考文献 (References)

[1] Hoffman J. Multi-Tracker Website [EB/OL]. (2008-08-5). http://wiki.Depthstrike.com/index.php/P2P:Protocol:Specifications:Multitracker.

[2] Cohen B. BitTorrent Website [EB/OL]. (2008-08-5). http://www.bittorrent.com.

[3] Parker A. The True Picture Of Peer-to-Peer File Sharing [EB/OL]. (2008-08-05). http://www.cachelogic.com.

[4] Neglia G, Reina G, Zhang Honggang, et al. Availability in BitTorrent Systems [C]//IEEE Infocom, USA, May 2007: 2216 - 2224.

[5] Pietzuch P, Ledlie J, Mitzenmacher M, et al. Network-Aware Overlays with Network Coordinates [C]// International Workshop on Dynamic Distributed Systems, Lisbon, Portugal, 2006: 12 - 12.

[6] Dabek F, Cox R, Kaashoek F, et al. Vivaldi: A decentralized network coordinate system [J]. *ACM SIGCOMM Computer Communication Review*, 2004, **34**(4): 15 - 26.

[7] Ledlie J, Gardner P, Seltzer M. Network Coordinates in the Wild [C]// NSDI '07. Cambridge, MA, April, 2007: 299 - 311.

[8] Ledlie J, Mitzenmacher M, Seltzer M, et al. Wired geometric routing [C]//Proceeding of International Work-shop on Peer-to-peer Systems (IPTPS). Bellevue, WA, USA, February, 2007.

[9] Sean Rhea, Brighten Godfrey, Brad Karp, et al. OpenDHT: A Public DHT Service and Its Uses [J]. *ACM SIGCOMM Computer Communication Review*, 2005, **35**(4): 73 - 84.

[10] Gummadi K P, Saroiu S, Gribble S D. King: estimating latency between arbitrary internet end hosts [J]. *ACM SIGCOMM Computer Communication Review*, 2002, **32**(3): 11 - 11.

[11] 张增斌, 陈阳. 基于临近原则的Bittorrent实验研究[J]. 厦门大学学报(自然科学版), 2007, **46**(A0): 213 - 215.
ZHANG Zengbin, CHEN Yang, DENG Beixing, et al. Experimental Study on Network Coordinate Based Locality-aware BitTorrent [J]. *Journal of Xiamen University(Natural Science)*, 2007, **46**(A0): 213 - 215. (in Chinese)

[12] CHEN Y, XIONG Y, SHI X, et al. Pharos: a decentralized and hierarchical network coordinate system for internet distance prediction [C]// GLOBECOM '2007, Washington DC, USA, 2007: 421 - 426.

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net