

合作环境下 P2P 网络最大吞吐量算法研究

吴 限, 苏德富

(广西大学计算机与电子信息学院, 南宁 530004)

摘 要: 针对 P2P 数据流网络中的文件传输情形进行了分析, 以最大化整个网络的吞吐量为目标, 提出了一个文件传输模型, 并在其上寻找一种可行的多项式时间内可求解的算法对近似最优的网络带宽利用率以及相应的流量路由分配进行计算, 模拟试验表明效果明显。

关键词: P2P; 最大吞吐量; 带宽利用率

Research on Achieving Optimal Throughput of P2P in Cooperative Environment

WU Xian, SU Defu

(School of Computer and Electronic Information, Guangxi University, Nanning 530004)

【Abstract】 This paper analyzes the circumstance of data dissemination in P2P network, aiming at the maximum throughput of the whole network and proposes a model of data dissemination. Then on this model, it develops a practical algorithm performed in polynomial time to approximately compute the optimal usage of network bandwidth and the corresponding routing strategy, and the simulation result turns out to be obviously fine.

【Key words】 P2P; Maximum throughput; Bandwidth usage

P2P是一类应用程序,它们共享分布在因特网边缘节点上的资源,比如存储空间、CPU时钟、内容信息等^[1]。简单地说,P2P就是一个位于应用层的特殊的异步并行分布式存储系统,系统中任一对节点可以通过P2P路由协议进行通信。

目前,随着 P2P 文件交换服务以及 P2P 流媒体服务应用的兴起,P2P 网络数据分发(Data Dissemination)的效率问题被人们关注起来。

1 相关的工作

因特网上一个著名的关于数据分发的协议就是IP多播协议^[2],它的设计初衷是为了减少多个点对点传输中带宽资源的浪费,但由于自治系统的原因没有能够广泛应用。于是出现了构建在应用层上的IP多播协议,它们可以很自然地解决网络层自治系统的问题,但是这种IP多播技术并不能较好地适应P2P的新环境,比如重复的数据分组可能会在节点之间传输、叶子节点不能向网络共享自己的上传带宽等,所以在P2P文件分布效率方面,最大化利用带宽同时避免传输冗余的数据是一个主要面对的问题。Byers等人^[3]提出了一种将数据分块后在节点之间随机传输数据块而提高吞吐量的方法;Bittorrent^[4]采用了一种稀少优先(Rarest First)的数据下载策略来减少节点间数据冗余的情况;还有的通过构建一棵全局多播树^[5],先让各个非根节点之间的数据大致不交,再让节点互相传递数据。以上所述的方法重点均在通过降低节点间数据冗余度寻求提高带宽利用率的方法,但都没有涉及最佳吞吐量以及带宽利用问题。

同时在数据网络中端到端最佳吞吐量的理论研究领域所取得的成果,对我们的问题研究也有很大的帮助。一直以来,端到端(End-to-End)最优吞吐量的计算在数据网络中都是一个非常基本而又计算难的问题,由于众多的网络拓扑结构以及链路带宽的约束,使得其中的理论计算往往是NP完全问

题,比如计算Internet中的从一个源节点到多个汇节点的最大吞吐量等同于求最多Steiner树问题(steiner tree packing)^[6],然而它被证明是一个NP完全问题^[7];利用网络编码(Network Coding)^[8]技术,Z. Li等人提出一种通过线性规划在多项式时间内计算最佳吞吐量的方法^[6],但具体的路由分配需要另外计算。

受图论中计算网络最大流的可增通路的启发^[9],我们构造了一个P2P的传输模型并且在其上给出一个近似最佳吞吐量的算法。该算法可以同时计算出可增流的路由分配,并且计算的规模以及精度可调。

2 模型介绍

2.1 模型假设

该P2P网络模型在合作条件下(不考虑公平等动机因素);暂时考虑节点数目一定的情况;需传递的总的文件数据大小一定,这样才有网络传输的完成时间;忽略线路传输时延以及路由时延;节点处理能力不限制,传输能力取决于线路的带宽,且带宽恒定无变化。

2.2 模型定义

可以将一个P2P网络抽象成为一个连通的边权无向图:一个 n 个节点的边权无向图 $G(V, E)$,边容量 $C(V_i, V_j)$ 代表节点 V_i 与 V_j 之间的带宽,流量 $F(V_i, V_j)$ 代表节点 V_i 与 V_j 之间的流量。其中 $F(V_i, V_j)$ 小于或者等于 $C(V_i, V_j)$;初始状态只有一个源点 S ,其他节点都是汇点且没有文件数据,数据文件长度 Len ,文件数据可以多段同时传递。

3 模型分析

首先定义一些概念:

作者简介: 吴 限(1980—),男,硕士生,主研方向:并行计算与网络安全;苏德富,教授

收稿日期: 2005-11-30 **E-mail:** hnhyul@126.com

定义 1 文件元块(Unit): 最小的数据单元, 不可再分, 用字母 U 表示。这是为了理论上离散化讨论的需要, 在实际应用中可以设定为相应的数值。不同的 Unit 大小可以代表计算的精确程度, 这样可以在计算规模以及精确度之间作出调整; 此外通过 U 来做单位, 可以定义系统中传输文件的大小以及带宽的大小, 并且将其统一化。

定义 2 (1)势: 节点所拥有的文件数据完整性的程度。可以定量的定义为节点拥有的文件元块 U 的数量。(2)势差: 不同节点间势的差异程度, 即节点间所拥有的文件数据的差异程度。很明显, 节点之间有势差才需要进行数据传递。

如果跟踪任何一个文件元块的传输路径, 可以发现如定理 1 的情况。

定理 1 在一个连通图中, 一个源的任何 U 块传输的路径必是一棵支撑树, 即包含图所有节点的树(支撑树详细定义见文献^[9])。

证明 因为图连通且存在势差, 最终这个 U 块能够传递到整个连通图的所有节点, 所以 U 块传递的路径必然连通整个图。对这个 U 块来说, 势相同的两个节点之间不会进行传输, 传输路径不会存在环; 任意 U 块传输的路径为一棵支撑树, 且树根就是源。

既然每一个 U 的路径都是一棵支撑树, 那么我们就只要在同一段时间内找到尽量多的这样的 U 支撑树, 它们的值就是这段时间内的 P2P 网络的流量分配, 同时文件分段的问题也将得到解决。可以确定结果是存在的, 但可能不唯一。

然后, 考虑如何寻找 U 支撑树的算法, 使得这样找到的组合是最优。先看一个例子来说明不同组合方法的区别。初始状态 a 各个边容量(单位 U/s)如图 1、图 2 所示, S 为源点, 文件长度 Len = 60U, 节点 A、B、C 和 D 都没有文件的任何部分。

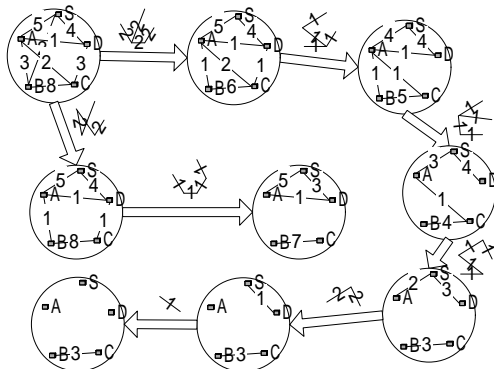


图 1 两个任意划分的情况

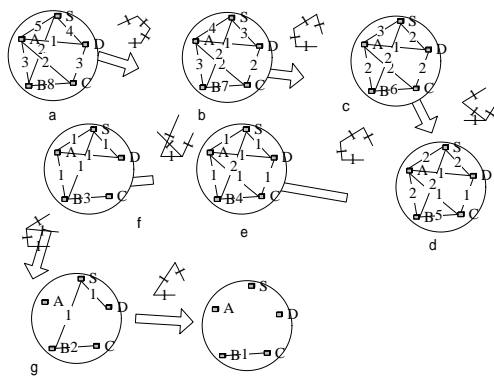


图 2 经过算法划分的情况

可以发现任意划分比不能充分利用网络的带宽, 在图 1 中有两种分配方式, 一种余下 3U 的带宽, 另一种余下 7U 的带宽不能利用, 造成带宽的浪费; 在图 2 中, 经过下面提出的算法过程, 仅余下 1U。

4 吞吐量效率度量

如果需要对不同的流量路由分配算法进行比较, 就需要定义一个度量的标准, 这里定义一个带宽利用率的标准来衡量网络吞吐量的优劣:

$$\rho = \frac{\sum_{i=0, j=0}^{n-1} F(i, j)}{\sum_{i=0, j=0, i < j}^{n-1} C(i, j)} \quad (1)$$

n 为节点的数目, 节点编号从 0 开始。分母为网络的总带宽, 分子为网络的总流量, 它们的比值就是网络的带宽利用率, 相对而言它较具有一般性。

同时, 在数据量一定的情况下, 假如带宽利用率最佳, 则吞吐量就一直是保持最佳状态, 那么就可以得出整个系统完成传输的最小时间 T_{\min} , 其定义为网络中最晚完成的节点的传输时间(T_i 为节点 i 的完成时间):

$$T_{\min} = \text{MAX}(T_i), i \in (0, n-1) \quad (2)$$

5 U 树组产生算法

5.1 算法

(1)从源开始广度搜索所有节点, 将每个节点与源的距离值(hop_to_seed)计算出来, 然后将每条边的 hop_to_seed 以两端节点的 hop_to_seed 值的和赋值。

(2)生成图的最大生成树^[9]。在这个过程中, 边的排序按照以下优先级: 1)前可用带宽(大的优先); 2)hop_to_seed 值(大的优先); 3)边的带宽(小的优先)。

(3)将其树枝的权(当前可用带宽)减去 1(分配了一棵 U 支撑树)。如树枝的权变为 0, 则说明该边的流量已达带宽上限。

(4)是否产生分离图? 如果不产生, 则继续第(2)步; 如果产生且可有多种分离结果, 则回溯一步寻找分离结果中包含源的最大连通子图中节点最多的分配情况, 再判断该最大连通子图是否是单图, 若是则算法结束, 否则将整个连通子图递归第(2)步。

U 树组产生算法(u_tree_generate)伪代码(未回溯)如下:

输入 某一个连通的无向带权图 Graph, 其源为 seed。

输出 一组 U 树。

```
(1) band_fist_search(seed); //初始化 Graph, 计算 hop_to_seed 值
while(!is_single) { //当 Graph 不是单图
(2)   do {sort(edges); //排序所有的边
      opt_tree(&u_tree); //产生一棵 U 树
(3)   transmit(u_tree); //进行传输
(4)   } while(is_connected); //是否产生分离图
}
```

5.2 U 树组产生算法说明

该算法每次分配一棵 U 树, 是为了尽量让所有的节点获得相同的下载带宽, 相当于求尽量多的多播树的问题; 此外, 算法中第(4)步的回溯一步是为了提高算法的完备性而采取的措施, 在对结果精确性要求不是非常高的场合中可以省略; 最后, 该算法逐次分配带宽, 这样有利于增强算法的适应性及灵活性。

5.3 U 树组产生算法的有穷性

因为图的总带宽 B 固定, 且对整个连通图以及图的包含源的某个连通子图总能求得最大生成树, 假设每次求得的生

成树的树枝数目为 K , 那么只要包含源的连通子图不是单图, K 的值总是大于或者等于 1 的, 所以每一次 B 减少 K , 当 $B=0$ 时, 包含源的连通子图必为单图, 则算法结束。

5.4 U 树组产生算法的时间复杂度

该算法的主体是求图的最大生成树, 其时间复杂度为 $O(e^2)$, 该算法时间复杂度为 $O(s * e^2)$, s 为源节点的带宽, e 为图的边数。容易发现该算法的解不是完备的, 但是有效的。

以上是 P2P 网络的第 1 次分配, 分配之后, 源就可以把各个 U 沿着每棵 U 树传递下去。只要不是 U 支撑树, 必然各个节点下载速率不同, 之后, 必然会有一些节点先完成传输, 变成源或者部分源。

5.5 多个源情况

在多个源的情况下, 可以将这些源都合并为一个源, 它们的边变为原来各自连接的边的值和, 然后继续应用以上算法进行运算。在最差的情况下, 图是逐个节点退化合并的, 则算法执行的次数最多不会超过图的节点数 n , 所以整体的算法时间复杂度为 $O(n * s * e^2)$ 。

6 实验结果

本模拟实验在 IBM-PC(256MB RAM、1.5GHz CPU)的 VC 环境下, 采用随机产生的网络拓扑结构, 节点数量为 n , 任两个节点间建立连接的概率为 $p=0.5$, 每一条边带宽不超过 $b=25U/s$, 传输文件的长度为 6000U, 执行的次数 $t=30$ 次。

图 3、图 4、图 5 分别代表不同节点数量情况下带宽利用率随时间变化的曲线, 可以发现在随机生成的网络拓扑结构下, 算法表现出相当好的性能。可以发现如下规律:

(1)如果在网络系统中存在某个瓶颈, 如某个节点的带宽相当窄, 那么整个系统完成的时间就会被它影响, 使得带宽不能充分利用, 这种情况在 n 比较小的情况下表现的最明显。可以发现从 $n=10$ 到 $n=100$, 带宽利用率越来越高, 并且曲线越来越显得稳定、集中。

(2)平均完成传输时间 T_{min} 越来越短, 分别为 273.9s、37.9s、18.8s。其主要原因是随着节点的增多, 虽然总的传输量增加, 但节点间的连接也随着增多, 总的带宽增加; 此外, 随着节点数量的增加, 整个网络的带宽倾向于均匀分布, 这样将会出现更少的瓶颈, 从而使得完成时间缩短, 这种情况在 10 个节点到 50 个节点表现得相当明显, 50~100 个节点就更加表现为带宽增加一倍后的结果。

(3)可以发现在将近完成传输的时候, 带宽利用率会陡然下降, 其原因是随着网络中源数量越来越多, 导致节点间势差迅速减少, 从而使得闲置带宽越来越多, 带宽利用率迅速下降。

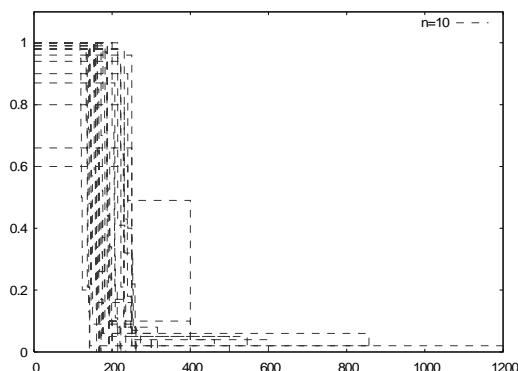


图 3 节点数量为 10 带宽利用率随时间变化的曲线

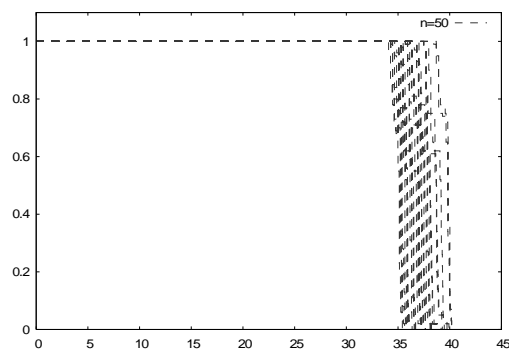


图 4 节点数量为 50 带宽利用率随时间变化的曲线

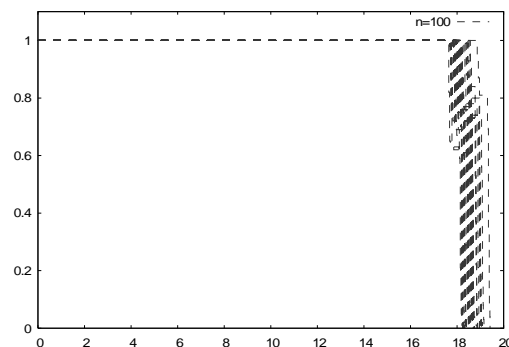


图 5 节点数量为 100 带宽利用率随时间变化的曲线

7 结论以及未来工作

本文在合作的 P2P 环境下提出了一种计算 P2P 网络流量路由分配的增量算法, 通过它可以在多项式时间内求得近似最优的网络吞吐量以及流量分配的方法。然而现在的 P2P 网络具有动态性、分布式等特性, 需要继续研究该算法在这些情况下的表现; 同时如果将网络信息流理论的研究的新成果, 比如网络编码(Network Coding)等理论, 应用到 P2P 网络中来, 将会对未来的研究产生更加积极的影响。

参考文献

- 1 Oram. Peer-to-peer: Harnessing the Power of Disruptive Technologies[Z]. <http://www.oreilly.de/catalog/peertopeer/>, 2001-03.
- 2 Deering S. The Pim Architecture for Wide-area Multicast Routing[J]. IEEE/ACM Transactions on Networking, 1996, 4(2): 153-162.
- 3 Byers J, Considine J, Mitzenmacher M, et al. Informed Content Delivery Across Adaptive Overlay Networks[J]. IEEE/ACM Transactions on Networking, 2004, 12(5).
- 4 Cohen B. Incentives Build Robustness in BitTorrent[Z]. <http://bitconjurer.org/BitTorrent>.
- 5 Kostic D, Rodriguez A, Albrecht J, et al. Bullet: High Bandwidth Data Dissemination Using an Overlay Mesh[C]. Proc. of SOSp, 2003-10.
- 6 Li Z, Li B, Jiang D, et al. On Achieving Optimal End to End Throughput in Data Networks: Theoretical and Empirical Studies[R]. Department of Electrical and Computer Engineering, University of Toronto, 2004-02.
- 7 Jain K, Mahdian M, Salavatipour M R. Packing Steiner Trees[C]. Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms, 2003.
- 8 Ahlswede R, Cai N, Li S R, et al. Network Information Flow[J]. IEEE Transactions on Information Theory, 2000, 46(4): 1204-1216.
- 9 楼世博, 金晓龙, 李鸿祥. 图论及其应用(第 1 版) [M]. 北京: 人民邮电出版社, 1982-07.