高技术通讯 2012 年 第22卷 第8期:786 ~ 790

doi:10.3772/j.issn.1002-0470.2012.08.002

基于拓扑感知的时延优先数据流量优化分配算法①

黄泳翔②* 钱德沛③** 伍卫国* 曹仰杰*

(*西安交通大学电子与信息工程学院 西安 710049)

(**北京航空航天大学计算机学院 北京 100191)

摘 要 为解决P2P流媒体的流量优化问题,提出了一种新的数据流量优化算法。该算法根据获得的有关底层网络拓扑信息,将节点分为流量路径冲突节点和非流量路径冲突节点,并分别采用加权时间片轮转和流量枚举优化算法对流量进行时延优先的优化分配。仿真实验表明,该算法可降低进入主干网的重复数据达70% ~90%,并降低时延30%。

关键词 P2P流媒体,拓扑感知,流量优化,加权时间片,枚举优化

0 引 言

对等(P2P)网络可以在内容分发上高效地组织Internet端用户进行协作[1],该技术在近几年里得到了迅猛发展。有关资料显示,中国几大网络服务供应商(ISP)的骨干网的P2P流量都占总流量的50%以上。P2P应用的流量之所以会占用如此高比例的网络带宽,一方面是由于P2P系统中的节点(peer)在选择邻居时的随机性造成系统逻辑拓扑与物理拓扑失配,另一方面是由于系统中的peer在下载数据时的随机性和无序性。这直接导致了ISP不同程度地对P2P应用加以限制。

针对该问题,近年来研究者提出了一系列解决方案[2-7]。Skype[3]采用给中继节点设置一个最大并发中继数来限制节点的中继流量。文献[4]提出了一种在异构的网络节点间根据节点的受欢迎程度按比例分配网络带宽的方法。文献[5]提出了一种基于无限竞拍博弈的方法,每个节点贡献出来的上行带宽按照几个目标节点的竞拍结果来分配。上述两个方法可以解决覆盖网层面数据流的合理分配问题,但不能解决IP层链路的共享问题。文献[6]针对覆盖网的潜在冲突,采用了均分瓶颈带宽的方式。文献[7]在文献[6]的基础上,提出了以降低传输时延为目标的流量分配算法。上述两种算法缓解了节点对共享路径的竞争,但没有消除共享路径的瓶颈问题。综上所述,已有的流量分配算法未能有效解决对等网络的流量优化问题。基于此,本文提出了一种基于拓扑感知的P2P应用层数据流量优化分配算法,该算法根据对共享链路的竞争和对共享中继节点的竞争两种不同方式,将节点分为流量路径冲突节点和非流量路径冲突节点两类,根据数学模型,采用相应的近似优化算法对网络资源进行分配。仿真实验表明,此算法可以有效降低进入骨干网的媒体数据,并能够较大幅度地降低媒体数据到达节点的时延。

1 问题描述

这一节将针对竞争共享路径和竞争中继节点两种不同模式分别建模。首先定义以下概念:

覆盖网的顶层逻辑连接可以由图 $G=(S,V,E)$ 表示,其中 S 为流媒体源数据服务器,V 表示边集,E 表示节点集。

引入概念单播流(conceptual unicast flow)[8]来构建线性方程,以更好地描述并发多数据流的传输模式。

设 $i,j \in E$ 为图 G 的任意两点,用 u_{ij} 表示连接 (i,j) 的传输容量,x_{ij} 表示连接 (i,j) 的传输速率,c_{ij} 表示数据传输在连接 (i,j) 上的延迟。f^t 表示从源节点 S 到接收节点 t 的概念单播流,$|f^t|$ 表示该概

① 863计划(2009AA01A131,2009AA01Z108),国家自然科学基金(61073011)和国际合作(2009DF12110)资助项目。

② 男,1978年生,博士生;研究方向:高性能网络,对等网络,流媒体;E-mail:lendhuang@gmail.com

③ 通讯作者,E-mail:depeiq@263.net

(收稿日期:2011-06-22)

念单播流的速率值,f_{ij} 表示该概念单播流在通过连接 (i,j) 时的速率。用 r 表示流媒体的固定码率。

1.1 竞争共享路径模式的数学模型

节点在竞争共享路径时,数据分配的最佳模式是使数据传输的总延迟最小,即 $\min \sum_{(i,j)\in A} c_{ij}x_{ij}$,约束于:

$$\sum_{j:(i,j)\in V} f_{ij}^t - \sum_{j:(j,i)\in V} f_{ji}^t = b_i^t, \forall i \in E, \forall t \in E, \quad (1)$$

$$f_{ij}^t \geqslant 0, \forall (i,j) \in V, \forall t \in E, \quad (2)$$

$$f_{ij}^t \leqslant x_{ij}, \forall (i,j) \in V, \forall t \in E, \quad (3)$$

$$0 \leqslant x_{ij} \leqslant \mu_{ij}, \ \forall (i,j) \in V, \quad (4)$$

其中 $b_i^t = \begin{cases} r, & i = S \\ -r, & i = t \\ 0, & \text{其它。} \end{cases}$

上述方程组可以通过放松约束条件(3),由拉格朗日松弛 (Lagrangian relaxation) 及次梯度方法 (subgradient algorithm) 求解。通过引入拉格朗日系数 μ_{ij},可以得到上述方程组的拉格朗日对偶方程:

$$\max_{\mu\geqslant 0} L(\mu) \quad (5)$$

其中:

$$L(\mu) = \min_{P} \sum_{(i,j)\in V} x_{ij}(c_{ij} - \sum_{t\in E} \mu_{ij}^t) + \sum_{t\in E}\sum_{(i,j)\in V} \mu_{ij}^t f_{ij}^t \quad (6)$$

P 为约束条件(1)(2)(4)。

式(6)所代表的拉格朗日子问题又可以分为前半部分的共享路径延迟最小值问题,和后半部分的节点集 E 的最短路径问题。

1.2 竞争中继节点模式的数学模型

节点在竞争中继节点时的数据分配最优模式数学模型为 $\min \sum_{(i,j)\in A} c_{ij}x_{ij}$,其约束于:

$$\sum_{j:(i,j)\in V} f_{ij}^t - \sum_{j:(j,i)\in V} f_{ji}^t = b_i^t, \forall i \in S \cup E, \forall t \in E,$$

$$f_{ij}^t \geqslant 0, \forall (i,j) \in V, \forall t \in E,$$

$$f_{ij}^t \leqslant x_{ij}, \forall (i,j) \in V, \forall t \in E,$$

$$\sum_{j:(i,j)\in V} x_{ij} \leqslant O_i, \ \forall i \in S \cup E,$$

$$\sum_{j:(j,i)\in V} x_{ji} \leqslant I_i, \ \forall i \in S \cup E,$$

$$x_{ij} \geqslant 0, \ \forall (i,j) \in V \quad (7)$$

其中 $b_i^t = \begin{cases} r, & i = S \\ -r, & i = t, \\ 0, & \text{其它。} \end{cases}$

O_i 和 I_i 分别为节点 i 的最大输出带宽和最大输入带宽。

该方程组同样可以通过放松约束条件(7),由拉格朗日松弛法求解,其拉格朗日对偶方程为

$$\min \sum_{(i,j)\in V} x_{ij}(c_{ij} - \sum_{t\in E} \mu_{ij}^t) \quad (8)$$

约束于:

$$\sum_{j:(i,j)\in V} x_{ij} \leqslant O_i, \forall i \in S \cup E,$$

$$\sum_{j:(j,i)\in V} x_{ji} \leqslant I_i, \forall i \in S \cup E,$$

$$x_{ij} \geqslant 0, \quad \forall (i,j) \in V。$$

(8)式所代表的是不等式约束的运输问题[6]。

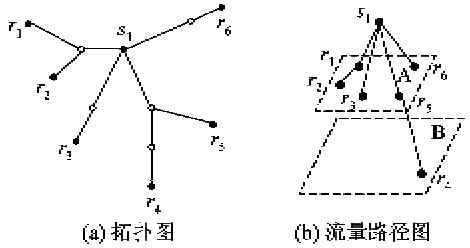
2 时延优先的数据流量优化分配算法

2.1 网络拓扑感知技术

我们用拓扑感知技术划分竞争共享路径节点和竞争中继节点。此外,拓扑感知技术可以有效求解节点集 E 的最短路径问题。本文采用文献[9]提出的一种利用传统 Traceroute 技术和 NT[10] 技术相结合的拓扑重构算法。相比其它拓扑重构算法,该算法在时间复杂度和占用流量方面都有较大的降幅,详细内容参见文献[9]。

2.2 流量路径

为了表示邻居节点间的路径相关性,以节点到邻居节点的路由器跳数相同的节点组成图层,不同图层以字母顺序进行标注,同层节点中的不同路由器以字母后加自然数进行标注,不同层间的路由器构成了数据发送节点到数据接收节点的流量传输路径,如图 1 所示。



(a) 拓扑图　　(b) 流量路径图

图 1　拓扑图与流量路径示意图

图 1(a) 为数据源节点 s_1 与其接收节点 r_1,r_2,r_3,r_4,r_5,r_6 的拓扑结构图。实心点表示终端节点,空心点表示与终端节点相连的路由器。根据流量(传输)路径定义,r_1,r_2,r_3,r_5,r_6 到 s_1 的路由器跳数为 1,可以用字母 A 表示其所在的层;r_1,r_2 共用一个路由器,用 $A0$ 表示 r_1,r_2 到 s_1 的流量路径。需要指出的是,图层后的序数只用于定位节点路径,不具实

际意义,可以由节点自由分配,即在不影响流量路径唯一性的前提下,r_1/r_2,r_3,r_4,r_5,r_6 的流量路径可以互换。因此可以设定 r_3、r_5、r_6 的流量路径分别为 $A1$,$A2$,$A3$。

节点 r_4 到 s_1 的路由器跳数为 2,且其和 r_5 共享路径 $A2$,用 $A2B0$ 标识其流量路径。由此得到节点与其邻居节点的流量路径图,如图 1(b)所示。

流量路径图可在聚类算法生成拓扑树的同时进行构建。算法步骤如下:

(1)在邻居节点第一次聚类时,为每组聚类节点添加以“A”为首的标识,并根据聚类出现的顺序,标注每个聚类的流量路径,如 $A0$,$A1$ 等。

(2)对聚类进行进一步划分时,次级聚类节点的流量标识在上一级流量标识基础上递增字母顺序,如上一级为“A”,则次级聚类标识为“B”。依此类推,直到所有聚类都被标识。

根据流量路径图,可以清晰地看出节点间的流量冲突。如 r_1,r_2,二者的流量路径都为 $A0$,二者为竞争共享路径节点;r_1,r_3,二者的流量路径首字母相同,次级数字不同,二者为竞争中继节点;r_4,r_5 的流量路径有一段完全相同,二者为竞争共享路径节点。

2.3 共享路径流量冲突优化算法

求解式(5)和式(6)的拉格朗日规划问题需要全局信息,但是在大规模网络中,让每个数据源节点得到网络全局的瞬态信息是非常困难的,受文献[7]启发,我们提出一种分布式的启发式流量分配算法。算法思想是根据前一次数据接收情况,如带宽使用效率不足,带宽资源过于紧缺等,对各路径的带宽进行动态微调。算法描述如下:

设 $x_{li}^{(n)}$ 为节点 i 在第 n 步时从路径 l 收到的流量,r 为流媒体播放速率,C_{li} 为该共享路径的最大物理链路流量,I_s 为节点 i 的最大输入带宽,令

$$\lambda_i^{(n)} = \max\left(0, r - \sum_{l\in L} x_{li}^{(n)}\right)$$

(L 为节点 i 的所有输入路径;)

$$e_i^{(n)} = \begin{cases} 1, & \sum_{l\in L} x_{li}^{(n)} > I_s \text{ 或 } x_{li}^{(n)} > C_{li} \\ 0, & \text{其他} \end{cases}$$

其中 $\lambda_i^{(n)}$ 和 $e_i^{(n)}$ 均为方程的调节系数,用于实时改变节点 i 从各竞争链路获得的实时流量。在 $n+1$ 步时,节点 i 根据反馈信息更新 $x_{li}^{(n+1)}$:

$$x_{li}^{(n+1)} = \begin{cases} x_{li}^{(n)}, & \text{如果 } \lambda_i^{(n)} = 0, e_i^{(n)} = 0 \\ x_{li}^{(n)} + \alpha_i^{(n)}\lambda_i^{(n)}, & \text{如果 } \lambda_i^{(n)} > 0, e_i^{(n)} = 0 \\ \min(I_s, C_{li}), & \text{如果 } \lambda_i^{(n)} = 0, e_i^{(n)} > 0 \\ \text{需添加上游节点}, & \text{如果 } \lambda_i^{(n)} > 0, e_i^{(n)} > 0 \end{cases}$$

其中 $\alpha_i^{(n)} = x_{li}^{(n)}/r$,表示通过路径 l 接收到的数据在节点 i 的整个输入带宽中所占的比重。算法对流量的调整一直持续到流媒体会话结束或有一方节点离开系统。

2.4 非共享路径流量冲突优化算法

分布式竞价算法是求解不等式约束的运输问题的有效方法[11]。考虑到最小化数据传输时延,本文采用了加权的时间片轮转算法。算法核心思想是由发送节点根据各接收节点反馈的信息,使性能较高的节点具有较高的时间片权重。由于性能较低的节点在竞争共享中继节点过程中始终处于劣势(其数据请求得不到满足),会促使该节点向其它数据源,如与其竞争中继节点而又具有较高权值的邻居节点申请数据。这样就会降低中继节点重复发送媒体数据的次数。

算法描述如下:

设发送节点 S 数据缓存中的数据大小为 D_s,输出带宽为 O_S。

步骤 1:初始化所有竞争节点 S 的时间片权值为 $\varpi = D_s/O_S$,其中 $i = 1,\cdots,m$,m 为流量冲突节点的个数。

步骤 2:节点 S 从所有节点中选出具有最高权值的节点发送数据,如果具有相同权值的节点不只一个,则从中随机选择一个进行传输。

步骤 3:时间片结束,所有节点的权值递增 τ。不失一般性,设在上个时间片里 S 选择的节点为 p,则此时节点 p 的权值递增 τ 的同时,再减去 $\tau^2 O_s/D_p$,其中,D_p 为 S 向 p 传输的数据量。

步骤 4:流媒体会话结束,算法停止,否则,返回步骤 2。

算法说明:D_p/τ 为在时间片内,S 向 p 传输数据的实际带宽。$\tau O_s/D_p$ 为在时间片内,S 向 p 传输数据的实际带宽在整体输出带宽中的比例,该值越小,表示 S 向该节点传输数据的时延越小。由于节点的权重是以传输时延来衡量,因此递减的权重需在该比例的基础上再乘以 τ。

3 模拟实验

模拟实验从网络整体平均时延和进入主干网的重复网络数据两个方面评价本文提出的数据调度算法。实验在 NS2 环境下进行仿真,底层拓扑由 Brite[12]生成。

3.1 实验的环境

网络由7个子网构成,流媒体服务器 S 位于其中一个子网内,其下行带宽为50Mbit/s。创建了节点连接度符合幂律分布的物理网络拓扑,在其中建立了1000个节点的覆盖网络,采用莱斯大学的 DS^2[13] 生成节点间的时延矩阵,其中扩展因子(scaling factor)设为10,邻居节点数量上限设为16。每个数据源分别向100个目标节点传送数据。数据分成片,流媒体的码率为800kbit/s,每个数据片与上个数据片的最大时差时限为1s。每个数据片通过800个IP数据包进行发送。骨干网的可用带宽为150Mbit/s,子网带宽为2Mbit/s,丢包率2%。

3.2 实验结果及分析

以文献[6]提出的冲突带宽平均分配算法和文献[7]提出的路径加权平均时延优先算法作为对比组,两种算法都具有一定的网络共享竞争感知能力。

(1)网络整体时延

如图2所示,在网络带宽发生竞争的情况下,以拓扑感知的时延优先算法取得的网络整体加权平均时延,比简单的平均分配算法下降了35% ~50%,比路径加权平均时延优先算法降低了15% ~30%。

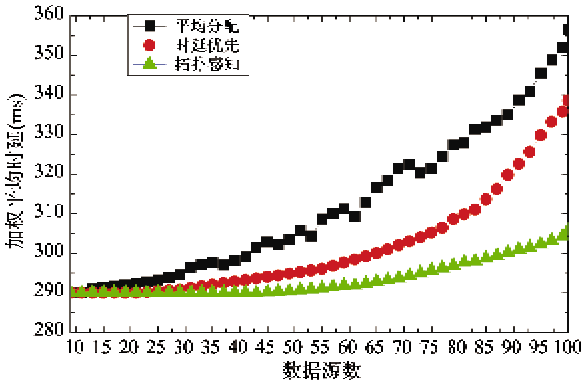


图2 三种算法下网络整体加权时延对比

原因在于:平均分配算法的核心是瓶颈路径的分时共享,算法降低了共享路径瓶颈的冲突,但并没有解决共享路径的瓶颈问题,分时共享的叠加效果仍然体现在了整体平均时延当中。时延优先使共享路径的利用率有了较大提升,但算法并没有让节点选择冗余路径进行数据传输,共享路径仍是流量分配的瓶颈。本文提出的算法,由于降低了高时延节点占用共享路径的时间,在带宽不足时,节点将开启新的冗余路径进行数据传输,极大缓解了共享路径的瓶颈问题,因此数据传输的平均时延得到了显著降低。

(2)骨干网重复数据

如图3所示,拓扑感知的时延优先算法在保证节点获得足够播放带宽的同时,大幅降低了进入骨干网的重复数据,降幅在70% ~90%。

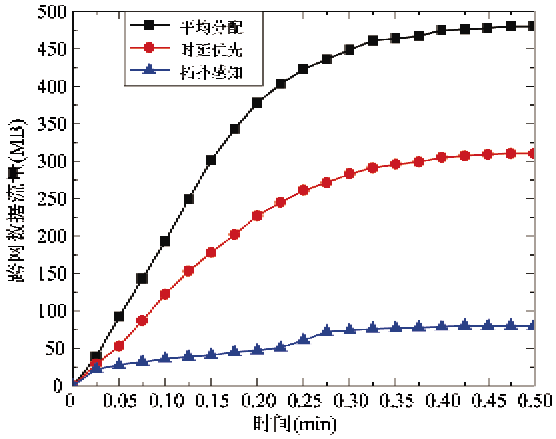


图3 三种算法下跨网数据流量对比

平均分配算法的着眼点在于解决节点带宽冲突,但P2P流媒体的带宽冲突,通常是因拓扑失配引起的,因此要解决带宽冲突需要先解决拓扑失配。平均分配算法没有解决拓扑失配问题,因此其跨网流量居高不下。时延优先算法由于引入时延因素,而物理相近的节点链路时延通常较低,因此其跨网数据流量相对于平均算法有较大降幅。但这种节点选择是被动式的,其对于底层物理拓扑的了解仅限于时延一个变量,因此其跨网流量相对于本文提出的算法仍较高。

4 结 论

本文提出了一种基于拓扑感知的时延优先数据流量分配算法,以解决P2P流媒体系统的流量分配问题。算法在路径流量冲突节点采用加权的时间片轮转算法,而在非路径流量冲突节点采用枚举优化算法。仿真时延表明,相对于简单的流量平均分配算法和路径加权平均时延优先算法,本文提出的算法性能都有较大幅度提高。下一步,我们将把算法应用到实际网络中。

参考文献

[1] Sentinelli A, Marfia G, Gerla M, et al. Will IPTV ride the peer-to-peer stream? *IEEE Communications Magazine*, 2007,45(6):86 - 92

[2] Liu J S, Wei J J, Yue G X, et al. Application layer mul-

ticast technology of streaming media. *Journal of Networks*, 6(8):1122-1128

[3] Lee S J, Banerjee S, Sharma P, et al. Bandwidth -aware routing in overlay networks. In: Proceedings of the IEEE International Conference on Computer Communications, Phoenix, USA, 2008. 1732-1740

[4] Huang X N, Daniel R F, Matthias G, et al. Balanced relay allocation on heterogeneous unstructured overlays. In: Proceedings of the IEEE International Conference on Computer Communications, Phoenix, USA, 2008. 126-130

[5] Wu C, Li B C. Strategies of conflict in coexisting streaming overlays. In: Proceedings of the IEEE International Conference on Computer Communications, Piscataway, USA, 2007. 481-489

[6] Zhu Y, Li B C. Overlay networks with linear capacity constraints. *IEEE Transactions on Parallel and Distributed Systems*, 2008, 19(2):159-173

[7] Wang R, Qian D P, Zhu D F, et al. Tuning performance of P2P mesh streaming system using an network evolution approach. In: Proceedings of the 4th International ICST Conference on Scalable Information Systems, Hong Kong, China, 2009. 135-151

[8] Li Z P, Li B C. Efficient and distributed computation of maximum multicast rates. In: Proceedings of IEEE International Conference on Computer Communications, Miami, USA. 1618-1628

[9] Xing J, Yiu, Gary C, et al. Network topology inference based on end-to-end measurements. *IEEE Journal on Selected Area in Communications*, 2006, 24(12):2182-2195

[10] Coates M, Hero A, Nowak R, et al. Internet tomography. *IEEE Signal Process*, 2002, 19(3): 47-65

[11] Kar K, Sarkar S, Tassiulas L. A simple rate control algorithm for maximizing total user utility. In: Proceedings of IEEE International Conference on Computer Communications, Anchorage, USA, 2001. 376-382

[12] Alberto M, Anukool L, Ibrahim M, et al. Brite: Universal topology generation from a user's perspective. In: Proceedings of the 9th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, Cincinnati, USA, 2001. 346-356

[13] Zhang B, Eugene N, Animesh N, et al. Measurement-based analysis, modeling, and synthesis of the Internet delay space. *IEEE/ACM Transactions on Networking*, 2010, 18(1): 229-242

Topology-awareness based latency minimizing data traffic optimized allocation algorithm

Huang Yongxiang*, Qian Depei**, Wu Weiguo*, Cao Yangjie*

(* School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049)

(** School of Computer Science, Beijing University of Aeronautics and Astronautics, Beijing 100191)

Abstract

To solve the network bandwidth optimization problem of P2P media streaming, a topology-awareness based latency minimizing data traffic optimized allocation algorithm was proposed. By utilizing the knowledge of underlying topology, the algorithm divides the member peers of a P2P streaming system into bottleneck shared peers and non-bottleneck shared ones, and uses the weighted round-robin algorithm and the enumerative optimization algorithm separately to optimize the transmission performance. The simulation results show that the algorithm can reduce the duplicated data about to 70% ~90%, and reduce the network delay of about 30%.

Key words: P2P streaming, topology-awareness, data traffic optimization, weighted round-robin, enumerative optimization