一种基于加权公平队列调度的改进型算法

田 冲,周井泉
(南京邮电大学 电子科学与工程学院,江苏 南京 210003)

摘 要:随着网络业务的不断增多,比如IP电话、视频会议、远程教学等应用的不断出现,需要Internet提供良好的QoS支持,传统的队列调度算法无法满足网络质量要求,文中提出了一种基于加权公平队列调度的改进型算法。首先对GPS模型进行详细分析,在此基础上深入研究WFQ算法,针对WFQ队列调度算法无法保证实时性业务的QoS,提出了改进型L_CBWFQ算法。该算法在带宽不足的情况下,区分实时性会话业务,进行权值调整策略,保证实时性业务的QoS。仿真分析表明,L_CBWFQ算法不仅能够提高实时性业务的QoS,而且在延时、抖动方面也有很大的改善。

关键词:带宽;实时;延时
中图分类号:TP301.6 **文献标识码**:A **文章编号**:1673-629X(2013)06-0071-03
doi:10.3969/j.issn.1673-629X.2013.06.018

An Improved Scheduling Algorithm Based on Weighted Fair Queue

TIAN Chong, ZHOU Jing-quan
(College of Electronic Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: With the growing number of network services, such as the increasing of real-time applications of IP telephone, video conference, distance learning, the Internet is required to provide a good QoS support. The basic scheduling algorithm cannot satisfy the quality of service. Proposed an improved scheduling algorithm based on weighted fair queue. Firstly, analyzed the basic principle of GPS, and studied the WFQ algorithm on the foundation of GPS. Propose L_CBWFQ algorithm for WFQ algorithm cannot ensure the quality of real-time service. Under the condition of limited bandwidth the L_CBWFQ algorithm distinguishes between real-time service and non real-time service, adjusts weighted value to ensure the quality of real-time service. Simulation shows the L_CBWFQ algorithm not only can satisfy the quality of real-time service, but also can improve delay and jitter problem in real-time service.

Key words: bandwidth; real-time; delay

0 引 言

在带宽一定的情况下网络通信中如何确保实时数据的有效发送,如何提高实时性数据信息的QoS性能是当今互联网通信领域研究[1]的热点之一。论文将对基于GPS模型的WFQ算法无法保证实时业务QoS的原因,WFQ算法的公平调度原则[2]存在的不足进行深入分析。

研究加权公平的改进型的L_CBWFQ算法,这种算法能根据业务的实时性进行权值调整[3]。在链路带宽不足的条件下,能够可以有效地保证实时业务的QoS要求。

1 GPS模型

通用处理器共享GPS(Generalized Processor Sharing)是公平类调度算法研究[4]的鼻祖,它是一种理想化的数据流模型,各队列是不区分优先级,类似于等优先级的,它根据各个队列的共享比例带宽对所有的进入调度器活动队列同时进行服务[5],它将所有队列中的会话数据包细分为无穷小的传输单位[6]来调度发送。假设GPS调度器要对N个会话进行服务,其模型如图1所示,连接会话1,2,…,N分别以权$\varphi_1,\varphi_2,\cdots,\varphi_N$进入GPS调度器。

服务器连续以固定速率γ进行工作,设$W_i(\tau,t)$为调度器在时间间隔(τ,t)内提供的服务量,GPS服

收稿日期:2012-09-02 **修回日期**:2012-12-06 **网络出版时间**:2013-05-14
基金项目:国家"863"高技术发展计划项目(2009AA01Z202)
作者简介:田 冲(1987-),男,硕士,研究方向为复杂网络与系统;周井泉,博士,硕士生导师,研究方向为通信网络中的路由选择、流量分配、接纳控制等技术。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20130514.1707.002.html

万方数据

务器提供连接会话的业务流[7]定义如下：

$$\frac{W_i(\tau,t)}{W_j(\tau,t)} \geqslant \frac{\varphi_i}{\varphi_j} \quad j=1,2,\cdots,N \tag{1}$$

其中会话 i 在时间段 (τ,t) 有正数量的数据量在队列中等待被服务，对(1)式求导之后对 j 求和得到(2)式：

$$W_i(\tau,t)\sum_j \varphi_j \geqslant (t-\tau)\gamma\varphi_i \tag{2}$$

对(2)式变形得(3)式：

$$\frac{W_i(\tau,t)}{(t-\tau)} \geqslant \frac{\varphi_i}{\sum_j \varphi_j}\gamma \tag{3}$$

其中会话 i 确保转发速率：$g_i = \dfrac{\varphi_i}{\sum_j \varphi_j}\gamma$

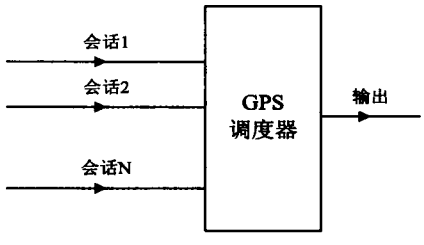

图 1 GPS 模型

GPS 算法的服务原则是绝对公平的[8]，由于它是基于理想流模型把数据包分成无穷小的数据元的，在实践中它是不可能实现的。但是可以模拟 GPS 调度器对会话分组进行调度，通过对 GPS 模型进行改进，形成了在实践中可以实现的 WFQ 算法。

2 WFQ 算法

WFQ(Weighted Fair Queueing，加权公平队列)首先引用了虚拟时间函数和虚拟时钟[9]，它是按照 GPS 处理过程对每一个分组进行处理，是一个基于虚拟完成时间的调度算法。设 t_j 表示第 j 个事件到达的时刻，处于 (t_{j-1},t_j) 服务状态的通路是固定的，用 B_j 来表示这些通路。调度器连续以固定速率 γ 进行工作，会话 j 的参数用权值 φ_j 表示，定义虚拟时间 $V(t)$，$V(t)$ 在调度器处于空闲时被设置为 0，虚拟时间都是从时刻 0 开始，即 $V(0)=0$

$$V(t_{j-1}+\tau) = V(t_{j-1}) + \frac{\tau}{\sum_{i\in B_j}\varphi_i} \tag{4}$$

其中 $\tau \leqslant t_j - t_{j-1}$，$j=2,3,4\cdots$

定义 P_i^k 为会话 i 的第 k 个分组，γ_i 为会话 i 的服务速率，a_i^k 为会话 i 的第 k 个分组的到达时间，L_i^k 为会话 i 的第 k 个分组的长度，S_i^k，F_i^k 分别为会话 i 的第 k 个分组的服务开始时间标签和服务结束时间标签，则这两个虚拟时间函数的定义为：

$$S_i^k = \mathrm{Max}\{F_i^{k-1}, V(a_i^k)\},\ F_i^k = S_i^k + \frac{L_i^k}{\gamma_i} \tag{5}$$

根据公式算法可以计算出各个分组到达时的虚拟结束时间，调度器选择最小的虚拟结束时间，然后进行转发最小的虚拟结束时间的分组，接着更新虚拟时间函数来维护集合 B_j。定义 Next(t) 为时刻 t 的下一次数据分组到达的时间，则 $(t,\mathrm{Next}(t))$ 时间段内没有分组到达，那么下一个更新虚拟时间函数的时刻将是 Next(t)。设 t 时刻分组到达的最小虚拟时间为 $F_{\min}$，由式(4)得 $F_{\min} = V(t) + \dfrac{\mathrm{Next}(t)-t}{\sum_{i\in B_j}\varphi_i}$ 移项，整理得：

$$\mathrm{Next}(t) = t + [F_{\min} - V(t)]\sum_{i\in B_j}\varphi_i \tag{6}$$

根据式(6)更新虚拟时间并且标记分组数据包虚拟结束时间，选择最小的虚拟结束时间的会话分组转发。如果有 N 个连接的会话分组，则复杂度为 $O(N)$[10]。该算法受漏桶 (σ_i,γ_i) 约束，设 $L_{\max}$ 为最大分组长度，C 为调度器链路带宽。其端到端最大抖动为 $\dfrac{\sigma_i}{\gamma_i}$，端到端的时延为 $D_i^* = \dfrac{\sigma_i}{\gamma_i} + \dfrac{L_{\max}}{C}$。

由式(5)得出会话的服务速率越小，虚拟结束时间越大，等待时间就越长，分组的延迟就越大，时延抖动就比较大，就比较容易造成数据包阻塞，容易丢包。当然在链路拥挤的情况下，实时业务得不到有效的转发，无法满足 QoS 的要求。下面主要研究 WFQ 的一种改进型的算法 L_CBWFQ。

3 L_CBWFQ 算法

WFQ 是一种公平类调度算法，不区分业务的实时性[11]，按照各个会话的平均速率分配带宽，当带宽不足时，一些实时业务数据包排队延时就越大，得不到有效的 QoS 的保证。

针对 WFQ 的缺陷，提出改进型的 L_CBWFQ 算法。改进算法的流程图如图 2 所示，其中实时性和权值进行调整是改进算法的核心部分。

由于 WFQ 算法未能对业务流分类，造成了数据包的带宽与权值成正比，对实时性业务流不能满足其性能要求。因此首先对获取的分组进行区分是否是实时业务流，若是实时性业务则对其权值进行调整，通过式(5)可以发现权值越大，虚拟完成时间就越小，分组数据包优先被转发，有效地控制了实时性业务。权值策略调整如下：由式(5) $F_i^k = S_i^k + \dfrac{L_i^k}{\gamma_i}$ 得：

$$\gamma_i' = \frac{L_{\max}}{L_i^k}\gamma_i$$

其中 $L_{\max}$ 表示分组队列的最大长度，L_i^k 表示会话 i 的第 k 个分组的长度，γ_i 表示会话 i 的服务速率。

当带宽充足时，会话满足时延要求，设 C 表示带宽

万方数据

且 $C \geqslant \sum_{i=1}^{N} \gamma_i$,$R$ 表示实时业务。

$$g_i = \frac{\gamma_i}{\sum_{j=1}^{N} \gamma_j} C, i \in R \tag{7}$$

当带宽不充足时,即 $C < \sum_{i=1}^{N} \gamma_i$

$$g_i = (C - \sum_{j \in R}^{N} \gamma_j) \frac{\gamma_i}{\sum_{k \in R^-} \gamma_k}, i \notin R \tag{8}$$

4 仿　真

将利用 NS-2 网络仿真[12]平台仿真 WFQ 算法和 L_CBWFQ 算法。设置链路带宽 C 为 6Mbit/s,三组数据分别为实时的语音数据流 $Source_1$,实时的视频数据流 $Source_2$,非实时的普通数据流 $Source_3$,三组长度分别设定为 216bit、300bit、512bit,分组的最大长度为 512bit,平均速率 γ_i 分别为 0.048Mbit/s、1.5Mbit/s、5.952Mbit/s,$Source_1$ 队列的时延必须小于 100ms; $Source_2$ 的队列时延必须小于 120 ms;$Source_3$ 队列时延必须小于 200ms。漏桶的深度 σ_i 分别为 50Byte、20000Byte、43690Byte。当带宽不充足时,进行仿真,得出的延时仿真图见图 3 ~ 图 5。

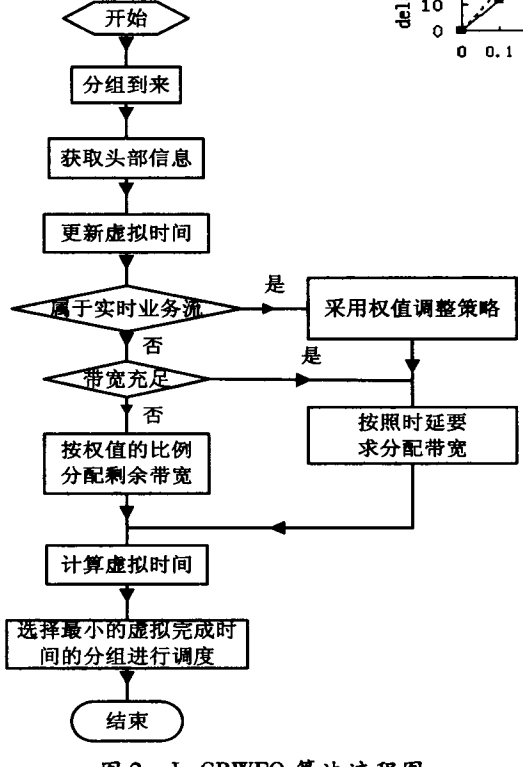

图 2　L_CBWFQ 算法流程图

当带宽不足时,通过仿真图发现实时的语音会话和视频会话的时延值得到了降低,而非实时的普通会话的时延值升高,这说明基于 L_CBWFQ 算法的数据分组中实时性的会话延时范围得到了最大的控制,满足会话延时需求,而且延时的抖动也得到了明显的改善。因此实时性的视频、语音会话带宽得到了很好的满足,提高了 QoS 的性能。

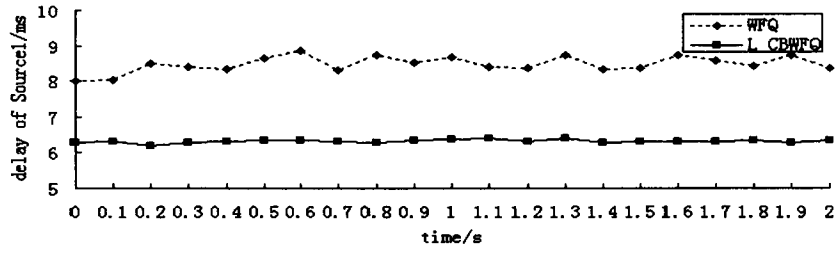

图 3　语音会话的端到端的延时

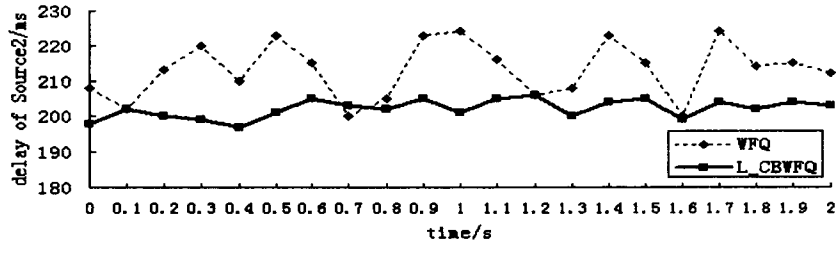

图 4　视频会话的端到端的延时

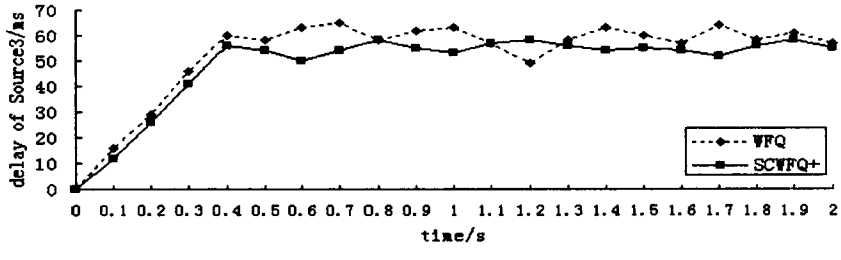

图 5　普通数据会话的端到端的延时

5 结束语

当带宽不足时,在不区分实时性业务时,按会话速率比例分配带宽,不能很好地解决时延问题。区分实时性会话业务,进行权值调整策略,保证实时性业务的 QoS。仿真分析表明,L_CBWFQ 算法在 WFQ 算法的基础上,先保证实时业务传送,再发送非实时性会话业务,不仅能够提高实时性业务的 QoS,而且在延时、抖动方面也有很大的改善,有效地保证了会话的 QoS。

参考文献:

[1] 任丰原,林　闯,刘卫东. IP 网络中的拥塞控制[J]. 计算机学报,2003,26(9):118-129.

[2] 马　原. 基于 GPS 的有线分组调度算法的研究[D]. 天津:天津大学,2007.

(下转第 78 页)

万方数据

情况。从模拟结果可看出，ACO-SS 在用户 QoS 、集群利用率以及平衡节点负载等方面具有更好的优势。

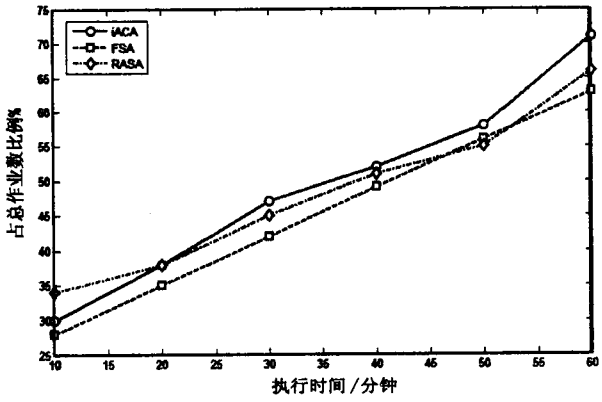

图2 不同执行时间段下的资源利用率

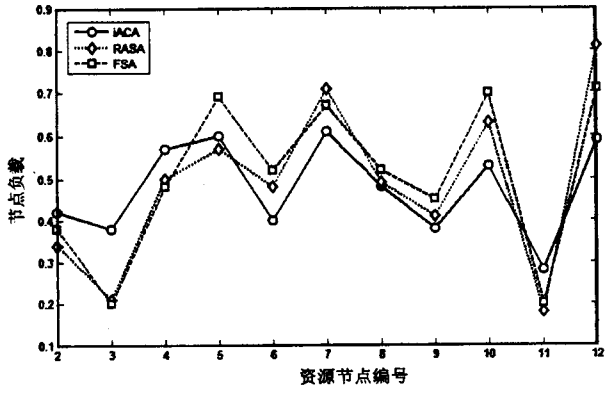

图3 资源节点负载比较

4 结束语

文中针对 MapReduce 集群提出了一种基于蚁群优化算法的调度策略，该策略不但把作业总体完成时间作为了参考指标，而且把集群资源利用率和节点负载也直接作为了整体性能优化的参考指标。该策略能够很好地满足用户 QoS，节点负载均衡以及集群资源利用率。该调度对于 MapReduce 集群来说，是一种十分实用有效的调度策略。

参考文献：

[1] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters[C]//Proc of Sixth Symposium on Operating System Design and Implementation. Berkeley: USENIX Association, 2004: 137-150.

[2] Zaharia M, Borthakur D, Sarma J S. Job scheduling for multi-user Mapreduce clusters[C]//Proceedings of the 5th European Conference. Washington: IEEE Computer Society, 2009: 145-161.

[3] Zaharia M, Borthakur D, Sarma J S. Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling[C]//Proceedings of the 5th European Conference on Computer Systems. New York: ACM, 2010: 265-278.

[4] 顾 宇，周 良，丁秋林. 基于优先级的 Three-Queue 调度算法研究[J]. 计算机科学，2011，38(10)：253-256.

[5] 刘 永，王新华. 云计算环境下基于蚁群优化算法的资源调度策略[J]. 计算机技术与发展，2011，21(9)：19-23.

[6] Jorddal P, Claris C, David C, et al. Resourceaware adaptive scheduling for MapReduce clusters[C]//Middleware 2011-ACM/IFIP/USENIX 12th International Middleware Conference. New York: ACM, 2011: 187-205.

[7] 段海滨. 蚁群算法原理及其应用[M]. 北京：科学出版社，2006：15-20.

[8] Apache Hadoop[EB/OL]. 2012-04-16. http://hadoop.apache.org/.

[9] 郝树魁. Hadoop HDFS 和 MapReduce 架构浅析[J]. 邮电设计技术，2012，21(9)：37-42.

[10] 李振东，谢 立. Web 服务器群的 QoS 确保及其接纳控制研究[J]. 计算机研究与发展，2005，42(4)：662-668.

[11] CloudSim[EB/OL]. 2012-02-11. http://www.cloudbus.org/cloudsim/.

[12] Hadoop 公平调度算法[EB/OL]. 2010-02-19. http://hadoop.apache.org/docs/r0.20.2/fair_scheduler.html.

(上接第73页)

[3] Bennett J, Zhang Hui. WF2Q: Worst-case Fair Weighted Fair Queueing[C]//Proc of IEEE INFOCOM'96. [s. l.]: IEEE Press, 1996: 120-128.

[4] Demers A, Keshav S, Shenkar S. Analysis and simulation of a fair queueing algorithm[C]//Proc of SIGCOMM'89. New York: ACM, 1990.

[5] Amrami B. WFQ and WF2Q[J]. Course: Topics in MultiFrocessing, 2001, 25(8): 122-153.

[6] Wang Song. Hierarchical Qos integration for realtime systems [D]. California: University of California, 2003.

[7] Abhay K, Parekh A. Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case[C]//Proc of IEEE INFOCOMM'92. [s. l.]: IEEE Computer Society Press, 1992: 915-924.

[8] 李方敏，李仁发，欧青立. 路由队列管理机制[J]. 计算机工程，2001，27(8)：222-232.

[9] 王重钢，隆克平，龚向阳. 分组交换网络中队列调度算法的研究及其展望[J]. 电子学报，2001，29(4)：53-59.

[10] Shreedhar M, Varghese G. Efficient fair queuing using deficit roundrobin[J]. IEEE/ACM Transactions on Networking, 1996, 4(3): 375-385.

[11] Bennett J C R, Zhang H. WF2Q: Worst-case fair weighted fair queueing[C]//Proc of IEEE INFOCOMM'96. San Francisco, CA: [s. n.], 1996: 120-128.

[12] NS2 教学手册[EB/OL]. 1996. http://www.isi.edu/nsnam/ns/doc/index.html

万方数据

万方数据 WANFANG DATA 文献链接

一种基于加权公平队列调度的改进型算法

作者： 田冲， 周井泉， TIAN Chong， ZHOU Jing-quan
作者单位： 南京邮电大学电子科学与工程学院,江苏南京,210003
刊名： 计算机技术与发展 ISTIC
英文刊名： Computer Technology and Development
年，卷(期)： 2013,23(6)

本文链接：http://d.g.wanfangdata.com.cn/Periodical_wjfz201306018.aspx