Adapting Fair Queueing to Improve Loss, Delay, and Admission Control Performance

N.G. Duffield† T. V. Lakshman* D. Stiliadis*

† AT&T Laboratories
Rm A175, 180 Park Avenue
Florham Park,
NJ 07932-0971 USA

* Bell Labs
Lucent Technologies
101 Crawfords Corner Road
Holmdel, NJ 07733, USA

Abstract

Fair queueing provides minimum rate guarantees and is fair by being rate proportional, i.e, excess bandwidth due to non-backlogged connections is distributed in proportion to rate guarantees. State dependence of the scheduler is restricted to whether specific connections are backlogged or not. In this paper, we propose three alternate scheduling mechanisms that can provide better service, in terms of losses and delay variation, than fair queueing. The mechanisms preserve the minimum rate guarantees of fair queueing but distribute the excess bandwidth in a manner much more adaptive to instantaneous needs than fair queueing. We compare the performance of the proposed schemes to that of fair queueing using different traffic sources such as voice and video, as well as sources which have aggregate long-range dependent behavior. We find that the proposed schemes, in comparison to packet generalized processor sharing (PGPS), reduce packet losses and curtail the tails of delay distributions for real-time traffic and hence permit the use of significantly smaller playout buffers for the same network load. We argue that these schemes add robustness in measurement-based admission control by providing adaptation for connections whose demands have been underestimated during the measurement process.

I. Introduction

Fair queueing has been the focus of much recent research attention [6], [10], [20], [1] and fair-queueing variants are being increasingly being implemented in switches and routers. Two reasons for use of fair queueing are its ability to provide rate guarantees and its rate-proportionality which ensures fairness. The constraint of fairness and so rate-proportionality limits fair-queueing schedulers in the degree to which they can be state-dependent. The only state dependence in fair-queueing schedulers is on the information about connections being backlogged or not. The price for fairness is reduced state dependence which can translate into increased losses and delay variation. This paper is based on the premise that fairness is not a relevant metric for many applications and instead the relevant metrics are delay bounds, losses, and delay variation. With this in mind, we propose alternate scheduling mechanisms that while preserving the rate guarantees of fair queueing (and therefore preserving its ability to provide worst-case delay bounds to traffic with bounded burstiness) are far more state-dependent. This increased state-dependence causes these schemes to have lower losses and delay variation than fair queueing. The schemes proposed adaptively redistribute only the excess bandwidth, i.e., each connection receives its guaranteed bandwidth even if it is in no danger of violating delay bounds. Three schemes are for excess bandwidth redistribution are proposed in the next section.

The adaptive redistribution gives more of the available bandwidth to connections with the most instantaneous need. Unexpectedly high demands can be made by connections whose bandwidth has been underestimated by measurement during admission control. This can occur because of underlying variability of the demand, or due to sampling error in the measurement process itself. We argue that adaptive redistribution effectively identifies and directs resources to connections whose badly characterized demands would otherwise lead to buffer overflow under purely rate proportional scheduling.

II. Redistributing Excess Bandwidth

Simple Longest Queue First has been shown to minimize the overall packet losses in a system with finite buffers [19]. Here, the scheduler always picks a packet from the longest queue for transmission. However, if the arrival rates and guaranteed shares are such that the expected delays are are not the same for all connections then the longest queue first scheme will always favor high delay (and generally high bandwidth connections). Also, if buffer allocations are not the same, serving the longest queue first may lead to buffer overflow for connections with a low buffer or low relative bandwidth guarantee.

To account for these differences in buffer sizes and bandwidth requirements, we can simply weight the queue size by the allocated bandwidth to make the scheduler a Longest Delay First scheduler. The scheduler then uses the excess bandwidth to try and equalize the delays of all connections in the system and minimize the packet losses if buffer sizes have been allocated proportional to bandwidth requirements. We will refer to this algorithm as FQ-LDF.

Note that this still does not take into account the maximum allowed delay for each connection. An enhancement to FQ-LDF is to explicitly take into account the different delay requirements that applications may have. We would like to be able to allow low maximum delays to voice sources, somewhat higher maximum delays to video sources, and much higher maximum delays for data sources. Also, constant bit rate sources may require extremely low delays, even when their bandwidth allocation is very low. To account for these differences, a maximum delay, that is usually a function of the buffer allocated to a connection, is associated with each connection. The scheduler serves connections that are most likely to exceed their allocated maximum delay. Let D_i be the maximum delay allocated to connection i. Let d_i be the delay that the last packet of the queue associated with that connection at time t may see. The connection that should be served is the one that is more likely to overflow and exceed its delay bound. Thus, connections should be served in increasing order of the $(D_i - d_i)$ values. We call this fair-queuing with least-time to overflow (FQ-LTO) redistribution of excess bandwidth.

The above schemes have not taken into account other characteristics of the source that may be known during admission control. Characteristics could be the maximum burst that can be transmitted by a connection or its peak arrival rate. For example, if sources are leaky bucket shaped, as is the case for sources in the Controlled Load traffic class, we could use this information to predict future arrivals from sources and calculate more precisely the connections that are likely to overflow or see delays higher than the allocated. In systems, where a dual, or multiple leaky bucket is used to shape the input traffic, a similar approach can be used. In fact, we do not require that a connection is shaped by a leaky bucket. All we need is that there be a function specified that provides information as to the maximum burst that can arrive for a connection, as well as the worst-case time that this burst may arrive.

The main idea consists of tracking the state of the connection based on its packet arrivals. This tracking is similar to the policing function that is used to drop packets from connections when they exceed their traffic specification. Based on the state of the connection we can estimate the worst-case arrivals for a given connection and use this to serve connections which are most likely to suffer losses.

An argument that can be made against the proposed adaptive policies, is that a greedy user will always receive more bandwidth. In such a case, users may request less bandwidth and try to receive more by sending more bursty traffic. Limiting such behavior can be explicitly done in many ways. The simplest is policing. If policing is not a viable option, the adaptive policies can themselves be tailored to counter greedy behavior. Instead of picking the user with maximum queue length or the minimum time to overflow, we can, for example, pick the second such candidate. In this case, users that consistently overflow the buffers cannot depend on receiving any excess bandwidth; it will not be in their best interests to be excessively bursty.

III. Performance Studies

We compared by simulation a Packet Generalized Processor Sharing (PGPS) scheduler with our adaptive schemes using a variety of sources to generate traffic. For each connection, the scheduler maintains a separate buffer whose drain time at the guaranteed rate is equal to the connection's maximum allowed delay D_i. The sources used are:

1. Two-state exponential On-Off sources: We used these to emulate compressed voice sources and so used small maximum allowed delays for these sources. The mean load generated per connection is 1/32 of the link capacity.
2. Video teleconference sources: We used the DAR(1) model used in [9] with a mean rate of 1.5 Mbps, a peak-to-mean ratio of 5, and .98 as the one-lag correlation.
3. CBR sources.
4. Data sources. These were modeled as On-Off sources where the on-time is Pareto distributed, and the off-time is exponential. The density of the Pareto distribution is given by

$$f(x) = a^{k-1}(k-1)/x^k, x \geq a = (k-2)/(k-1)\mu \tag{1}$$

We set $k = 5/2$, resulting in infinite variance of the on-times and long-range dependence in the aggregate traffic. The mean on-time was set to 6 packet transmission times and the mean off-time to 200.

We first compared the performance of FQ-LDF with PGPS (which is simulated exactly) using only one type of source. Table I shows packet loss rates for different small buffer sizes from a simulation with 32 voice sources. All connections were served with equal weight and the utilization is about .98. At this high load, PGPS typically has twice the losses of FQ-LDF and redistribution of excess bandwidth to the longest delay queue reduces the aggregate loss rate. We find similar results of lower loss rates for FQ-LDF with video sources as well. With Pareto sources, the loss rates are mostly similar for FQ-LDF and PGPS. This can be accounted for by the heavy-tailed distributions of the on periods.

Apart from packet losses, we also compared the delay distribution resulting from FQ-LDF to that of PGPS. For simulations with 32 voice sources with large enough buffers to avoid packet losses, we find that FQ-LDF needs significantly smaller playout buffers than PGPS. We see the same effect for video sources (plots are not shown for conciseness).

Next, we compare the performance of FQ-LTO with PGPS. FQ-LDF uses the excess bandwidth to equalize delays and it works very well when all sources have the same allowed delays. When different sources have different maximum allowed delays, the FQ-LTO scheduler which takes delays into account is expected to perform better than FQ-LDF

Buffer size in pkts/conn	Cell Loss Ratio:PGPS	Cell Loss Ratio:FQ-LDF
75	13.18e-3	8.3e-3
100	8.198e-3	4.938e-3
125	5.61e-3	3.61e-3
150	4.245e-3	2.849e-3
250	2.061e-3	1.025e-3
350	9.614e-4	1.054e-4
425	4.876e-4	$< 10^{-6}$
500	2.8478e-4	$< 10^{-6}$

TABLE I

CELL LOSS RATIOS FOR 32 EXPONENTIAL ON-OFF SOURCES WITH SHORT BUFFERS

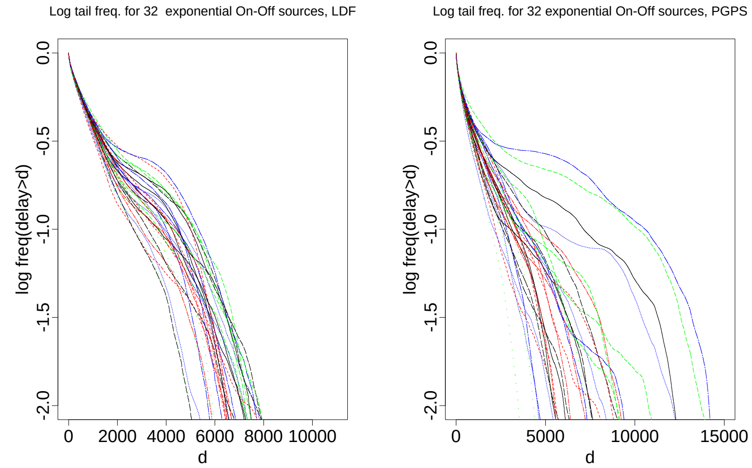

Fig. 1. Log tail frequencies for 32 exponential On-Off sources, according to service discipline (FQ-LDF or PGPS).

and PGPS. We use 32 sources with 8 sources from each of the types described before. CBR sources are guaranteed their mean rate. They generate at most 2 back-to-back packets and have a small buffer allocation of 5 packets. Voice sources are also guaranteed their mean rate and have a maximum buffer allocation of 1200 packets (about 10 ms delay at their guaranteed rates). Video sources are guaranteed two times their mean rates and have a buffer of 12000 packets. Data sources are given a large buffer of 40000 packets. Since they are not delay sensitive, they are guaranteed a rate below their mean rate. Data sources are expected to use some the excess bandwidth from the other classes to make up for their allocation below their means.

The PGPS scheduler gives excess bandwidths to the data class in proportion to the weights. Since the data class has a very large delay allowance, the FQ-LTO scheduler gives priority to the low delay classes in the distribution of excess bandwidth. Consequently, the lower delay classes have very compact delay distributions whereas the very high delay data class (which also has underallocated guaranteed bandwidth) has better tail behavior with PGPS. The FQ-LTO scheduler achieves what one would want in this mixed scenario. Worst case delay bounds are always met even for data (if it is given appropriate guaranteed rates). The lower delay classes have very small tails and hence need drastically lower playout buffers than when using PGPS.

IV. ADAPTATION TO ERRORS IN ADMISSION CONTROL

The predominant approach to admission control has focused on characterizing future bandwidth demands of a connection. If the source of a flow and the network can agree on the characterization, resources can be allocated by the network in order that the performance obtained for the source meets its expectations. Characterization can be done on the basis of declared parameters, or by measurement, or by some combination of these. Many applications, variations and combinations of measurement-based characterization are possible. These include measurement of a single connection, or an aggregate; for admission to a dedicated resource, or a shared one; measurement of admitted connections, or those seeking admission. We refer the reader to [2], [4], [5], [11], [14], [15] for examples of work in this area.

Measurement-based characterization of the bandwidth can fail to be accurate for three reasons: systematic errors, errors due to high variability, and measurement errors. Systematic error occurs when an underlying nonstationarity causes traffic characteristics to depart from their (initially) measured values. High variability, as manifest by stationary long-range dependent processes (e.g. fractional Brownian motion can lead to local trends away from measured values; [18]. Such LRD traffic models have been proposed on the basis of empirical studies; see e.g. [17]. In practice it may be difficult to distinguish between non-stationarity and long-range dependence. The consequences for admission control are, however, much the same: long periods for which the offered load exceeds that expected from measurements. (Here "long" is with respect to the drain-time of a full buffer). Measurement errors occur because measurement entails sampling and is hence subject to statistical error. Such errors can lead to overadmission through underestimation of the required bandwidth, resulting in an attained quality lower than the target. For bufferless resources, Grossglauser and Tse [12] have recently established the extent to which admission controls must be conservative in order that measurement errors do not compromise the target quality with greater than the target frequency. However, such errors will still occur. An analysis of the buffered case is presented in [7].

Whatever the source and frequency of errors in bandwidth characterization, it is desirable to have bandwidth allocation mechanisms which can adapt to provide or increase robustness. Scheduling can provide this, and we believe in particular that adaptive redistribution of excess bandwidth can be an effective means of adapting to badly characterized traffic. Consider, for example, when individually measured sources are each to be provided a guaranteed rate out of individual or shared buffers. If some true source rates are greater than expected, then adaptive excess redistribution policies described above will selectively direct resources to these connections. Thus we expect them to perform better than FQ in preventing the buildup of large occupancies and delays, in the buffers used or shared by these connections. We shall report detailed results elsewhere.

V. Concluding Remarks

Unless the fairness requirement of GPS is required for policy reasons, fair queueing systems can improve their performance by adaptive redistribution of excess bandwidth without losing their worst case fairness property or their ability to guarantee worst case delay bounds. For many applications, it is sufficient that schedulers provide minimum rate guarantees and use state-dependent excess bandwidth allocation so as to curtail delay-distribution tails or offset measurement-errors in admission control.

References

[1] J.C.R. Bennett, H. Zhang, "WF^{2}Q: Worst-case Fair Weighted Fair Queueing", *Proceedings of INFOCOM'96*, pp. 120-128, 1996.

[2] C. Casetti, J. Kurose, D. Towsley, "An Adaptive Algorithm for Measurement-based Admission Control in Integrated Services Packet Networks," Int. Workshop on Protocols for High Speed Networks, (Sophia Antipolis, Oct. 1996).

[3] D. Clark, S. Shenker, L. Zhang, "Supporting Real-Time Applications in an Integrated Services Packet Network: Architecture and Mechanism", *Proceedings of ACM SIGCOMM 1992*, pp. 14-23.

[4] C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand and R. Webber, "Call Aceeptance and Routing using Inferences from Measured Buffer Occupancy," IEEE Trans. Comm., Vol. 43, No. 2/3/4, pp. 1778-1784, 1995.

[5] S. Crosby, I. Leslie, J. Lewis, R. Russell, F. Toomey & B. McGurk, "Practical Connection Admission Control for ATM Networks Based on On-line Measurements", Proceedings IEEE ATM '97, June 1997, Lisbon.

[6] A. Demers, S. Keshav, S. Shenker, "Analysis and Simulation of a Fair Queueing Algorithm" *Internetworking: Research and Experience*, pp, 3-26, vol. 1, 1990.

[7] N.G. Duffield, "A large deviation analysis of errors in measurement based admission control to buffered and bufferless resources", AT&T Laboratories, 1998.

[8] N.G. Duffield, T.V. Lakshman, D. Stiliadis "On Adaptive Bandwidth Sharing with Rate Guarantees," *Proceeding of Infocom 1998*

[9] A. Elwalid, D. Heyman, T. V. Lakshman, D. Mitra, A. Weiss, "Fundamental Bounds and Approximations for ATM Multiplexers with Applications to Video Teleconferencing", *IEEE Journal on Selected Areas in Communications: Special Issue on Fundamental Advances in Networking*, 13:1004–1016, 1995.

[10] S. J. Golestani, "A Self-Clocked Fair Queueing Scheme for Broadband Applications", *Proceedings of INFOCOM'94*, pp. 636-646., 1994.

[11] R.J. Gibbens, F.P. Kelly & P.B. Key, "A decision-theoretic approach to call admission control in ATM networks" IEEE Journal on Selected Areas of Communications: Special Issue on Fundamental Advances in Networking, 13:1101–1114, 1995.

[12] M. Grossglauser and D.N.C. Tse "A Framework for robust measurement-based admission control", In Proc. ACM SIGCOMM 97, Cannes, France, September 1997.

[13] P. Goyal, H.M. Vin and H. Chen, "Start-time Fair Queing: A Scheduling Algorithm for Integrated Services Packet Switching Networks", *Proceedings ACM SIGCOMM'96*, pp. 157-169.

[14] S. Jamin, P.B. Danzig, S. Shenker & L. Zhang, "A measurement based admission control algorithm for integrated services packet networks", in Proc. ACM SIGCOMM '95, 1995.

[15] E. Knightly, "Second Moment Resource Allocation in Multi-Service Netw orks," in: Proceedings of ACM SIGMETRICS '97, Seattle, WA, June 1997.

[16] D. S. Lee, "Weighted Longest Queue First: An Adaptive Scheduling Discipline for ATM Networks", *Proceedings INFOCOM'97*, 1997.

[17] Leland, W. E., Taqqu, M. S., Willinger, W. and Wilson, D. V. (1994) On the self-similar nature of Ethernet traffic, *IEEE/ACM Trans. Networking*, 2, 1–15.

[18] B.B. Mandelbrot and J.R. Wallis (1968). "Noah, Joseph, and Operational Hydrology". *Water Resour. Res.,* **4** 909–918

[19] S. Panwar, D. Towsley, J. Wolf, "Optimal scheduling policies for a class of queues with customer deadlines to the beginning of services", *Journal of ACM*, vol. 35, no. 4, pp. 832—844, 1988.

[20] D. Stiliadis and A. Varma, "Design and analysis of Frame-based Fair Queueing: A New Traffic Scheduling Algorithm for Packet-Switched Networks", *Proceedings of ACM SIGMETRICS '96*, pp. 104-115, May 1996.