Notes on Token Buckets, GPS, and WFQ

Jean Walrand
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
wlr@eecs.berkeley.edu

19th April 2005

Abstract

The objective of this short note is to review the basic ingredients of QoS in networks: traffic regulation by token buckets and guaranteed delays for such regulated traffic by scheduling. The material is from [1].

1 Overview

The Internet is a best-effort network, meaning that the routers and switches transmit packets whenever they can but without any other explicit guarantee on delay or throughput. The success of the Internet results from the great simplifications that are possible when implementing a best effort service and from the clever applications that are designed to operate well with that service. Nevertheless, some applications work much better if delays are bounded and if the throughput is sufficient. (Think of games, voice, video streaming, etc.)

The challenge is to figure out simple modifications of the protocols that provide some form of quality of service (QoS). After many years of research, not much has been implemented, because the slightest modification of the best effort model seem to result in endless complications The complications have to do mostly with management and agreement among providers.

In this note, we review two approaches to QoS: IntServ, which is probably too complicated to fly, and DiffServ which has a better chance and is implemented in some networks. These two ideas rely on the same basic observation: to bound delays, one needs to regulate the traffic and to guarantee some minimum service rate to regulated flows. If the traffic enters a router faster than it can transmit it, the delays build up. Similarly, if bursts of traffic are too large, then the time to transmit those bursts is excessive.

In Section 2 we review the basic idea behind bounding delays by regulating flows and guaranteeing a minimum transmission rate. In Section 3 we explain an idealized scheduling scheme (GPS) to guarantee such a minimum rate and we discuss an implementation (WFQ) of an approximation of that scheme. Finally, in Section 4 we explain that regulated traffic served with WFQ has bounded delays.

2 Bounding Delays

Consider the system shown in Figure 1. Packets move from a packet buffer one by one into the queue. The queue, whenever it is not empty, sends bits at a variable rate $R(t)$. This variable rate is a mathematical fiction that is convenient to model the sharing of a physical link among multiple queues. A token counter controls when packets can move into the queue. A packet of P bits at the head of line in the buffer can move to the queue if the token counter has at least P tokens. In that case, one removes P tokens from the counter and the packet jumps at once into the queue. The token counter gets tokens at the constant rate of r tokens per second and it saturates whenever it has s tokens.

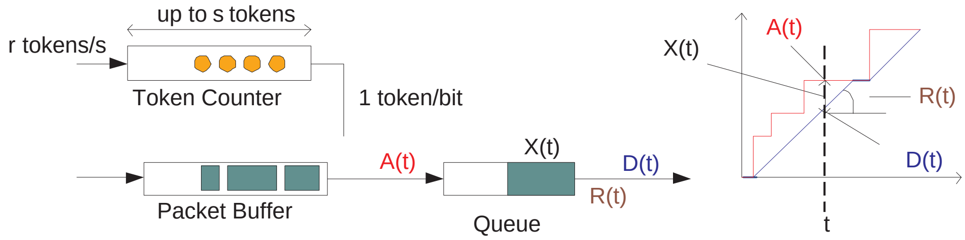

Figure 1: A queue with regulated input.

This token mechanism limits the size of bursts that can enter the queue to s bits and also the long term arrival rate into the queue to r bits per second. More precisely, if $A(t)$ is the number of bits that enter the queue in $[0, t]$, for $t \geq 0$, then we claim that

$$A(t) - A(u) \leq s + r(t - u), \text{ for all } 0 \leq u < t.$$

To see this, note that the token counter has at most s tokens at time u and that it will collect $r(t-u)$ more tokens during $[u, t]$. Since one needs one token per bit that one sends into the queue, the maximum number of bits that can move from the packet buffer to the queue during $[u, t]$ is bounded by $s + r(t-u)$.

Thus, in a very short interval of time, it is possible for a burst of up to s bits to enter the queue. Also, in the long term, the rate of arrivals into the queue cannot exceed r.

The following straightforward fact is one of the building blocks that we need for QoS. It shows that if a router serves a regulated flow at a sufficient minimum rate, then the delays are bounded.

Fact 1 *Assume that $R(t) \geq r_1 \geq r$ for all $t \geq 0$ and $X(0) = 0$. Then $X(t) \leq s$ for all $t \geq 0$. Moreover, the queuing delay of every packet in the queue is at most equal to s/r_1.*

Proof:

Fix some time $t > 0$ such that $X(t) > 0$. Let u be the last time before time t that $X(u) = 0$. That is, $X(u) = 0$ and $X(v) > 0$ for $v \in (u, t)$. Let $D(t)$ be the number of bits that leave the queue in $[0, t]$, for $t \geq 0$. Note that the number of bits $D(t) - D(u)$ that left the queue during $[u, t]$ is at least equal to $r_1(t - u)$ since the service rate is at lest equal to r_1 whenever the queue is nonempty. Also, the number of bits $A(t) - A(u)$ that entered the queue in $[u, t]$ is at most $s + r(t - u)$. Since $X(u) = 0$, we see that

$$X(t) = A(t) - A(u) - [D(t) - D(u)] \leq s + r(t - u) - r_1(t - u) \leq s$$

since $r_1 \geq r$.
The queuing delay of a packet that enters the queue at time t, when the backlog is $X(t)$, is less than $X(t)/r_1$ since the queue clears that backlog at least at rate r_1. We have seen that $X(t) \leq s$. Consequently, the queuing delay is always bounded by s/r_1. □

3 Weighted Fair Queuing

Weighted Fair Queuing (WFQ) is a scheduling mechanism that controls the sharing of one link among packets of different classes. We explain that this mechanism provided delay guarantees to regulated flows. Both the definition of WFQ and its analysis are based on an idealized version of the scheme called Generalized Processor Sharing (GPS). We start by explaining GPS.

3.1 GPS

Figure 2 illustrates a GPS system. The packets are classified into K classes and wait in corresponding first-in-first-out queues until the router can transmit them. Each class k has a weight w_k. The scheduler

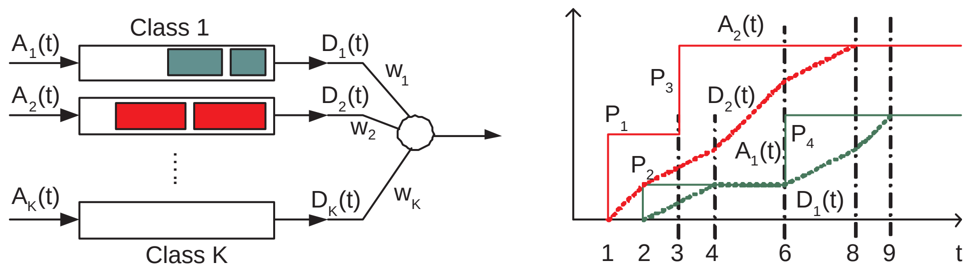

Figure 2: Generalized Processor Sharing.

serves the head of line packets at rates proportional to weight of their class. That is, the instantaneous service rate of class k is $w_k C/W$ where C is the line rate out of the router and W is the sum of the waits of the queues that are backlogged at time t. Note that this model is a mathematical fiction that is not implementable since the scheduler mixes bits from different packets and does not respect packet boundaries.

We draw the timing diagram on the right of the figure assuming that only two classes (1 and 2) have packets and with $w_1 = w_2$. A packet P_1 of class 2 arrives at time $t = 1$ and is served at rate one until time $t = 2$ when packet P_2 of class 1 enters the queue. During the interval of time $[2, 4]$, the scheduler serves the residual bits of P_1 and the bits of P_2 with rate $1/2$ each.

From this definition of GPS, one sees that the scheduler serves class k at a rate that is always at least equal to $w_k C/(\sum_j w_j)$. This minimum rate occurs when all the classes are backlogged. Combining this observation with Fact 1, find the following result.

Fact 2 *Assume that the traffic of class k is regulated with parameters (s_k, r_k) such that $\rho_k := w_k C/(\sum_j w_j) \geq r_k$. Then the backlog of class k never exceeds s_k and its queuing delay never exceeds s_k/ρ_k.*

3.2 WFQ

As we mentioned, GPS is not implementable. Weighted Fair Queuing approximates GPS. WFQ is defined as follows. The packets are classified and queued as in GPS. The scheduler transmits one packet at a time, at the line rate. Whenever it completes a packet transmission, the scheduler starts transmitting the packet that GPS would complete transmitting first among the remaining packets. For instance, in the case of the figure, the WFQ scheduler transmits packet P_1 during $[1, 3.5]$, then starts transmitting packet P_2, the only other packet in the system. The transmission of P_2 completes at time 4.5. At that time, WFQ starts transmitting P_3, and so on.

The figure shows that the completion times of the packets $P_1, \ldots, P_4$ under GPS are $G_1 = 4.5, G_2 = 4, G_3 = 8, G_4 = 9$, respectively. You can check that the completion times of these four packets under WFQ are $F_1 = 3.5, F_2 = 4.5, F_3 = 7$, and $F_4 = 9$. Thus, in this example, the completion of packet P_2 is delayed by 0.5 under WFQ. It seems quite complicated to predict by how much a completion time is delayed. However, we have the following simple result.

Fact 3 *Let F_k and G_k designate the completion times of packet P_k under WFQ and GPS, respectively, for $k \geq 1$. Assume that the transmission times of all the packets are at most equal to T. Then*

$$F_k \leq G_k + T, k \geq 1. \quad (1)$$

Proof:

Note that GPS and WFQ are work-conserving: they serve bits at the same rate whenever they have bits to serve. Consequently, the GPS and WFQ systems always contain the same total number of bits. It follows that they have the same busy periods (intervals of time when they are not empty). Consequently, it suffices to show the result for one busy period.

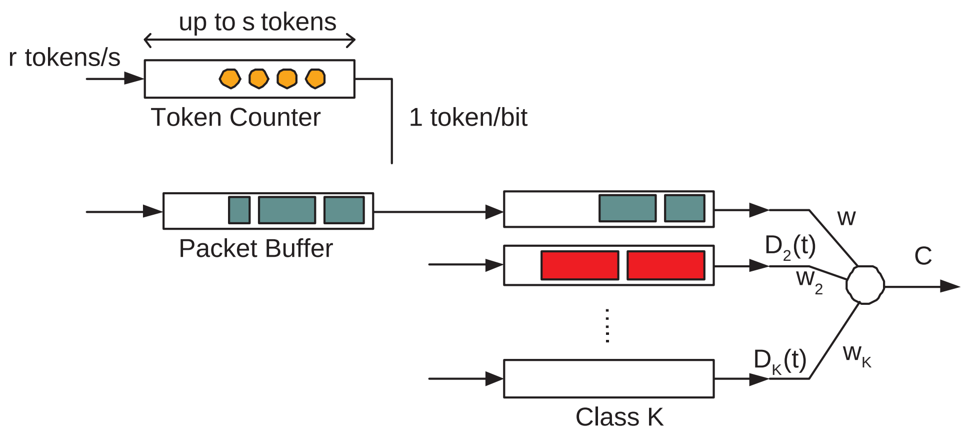

Figure 3: Regulated traffic and WFQ scheduler.

Assume $F_1 < F_2 < \cdots < F_k$ correspond to packets within one given busy period that starts at time 0, say.
If $G_n \leq G_k$ for $n = 1, 2, \ldots, k-1$, then during the interval $[0, G_k]$, the GPS scheduler could serve the packets $P_1, \ldots, P_k$, so that G_k is larger than the sum of the transmission times of these packets under WFQ. Hence $G_k \geq W_k$, so that (1) holds.
Thus assume that $G_n > G_k$ for some $1 \leq n \leq k-1$ and let m be the largest such value of n, so that

$$G_n \leq G_k < G_m, \text{ for } m < n < k.$$

This implies that the packets $\mathcal{P} := \{P_{m+1}, P_{m+2}, \ldots, P_{k-1}\}$ must have arrived after the start of service $S_m = F_m - T_m$ of packet m, where T_m designates the transmission time of that packet. To see this, assume that one such packet, say P_n, arrives before S_m. Let G'_m and G'_n be the service times under GPS assuming no arrivals after time S_m. Since P_m and P_n get served in the same proportions until one of them leaves, it must be that $G'_n < G'_m$, so that P_m could not be scheduled before P_n at time S_m.
Hence, all the packets $\mathcal{P}$ arrive after time S_m and are served before P_k under GPS. Consequently, during the interval $[S_m, G_k]$, GPS serves the packets $\{P_{m+1}, P_{m+2}, \ldots, P_k\}$. This implies that the duration of that interval exceeds the sum of the transmission times of these packets, so that

$$G_k - (F_m - T_m) \geq T_{m+1} + T_{m+2} + \cdots + T_k,$$

and consequently,

$$G_k \geq F_m + T_{m+1} + \cdots + T_k - T_m = F_k - T_m,$$

which implies (1). □

4 Regulated Flows and WFQ

Consider a stream of packets regulated by a token bucket and that arrives at a WFQ scheduler, as shown in Figure 3.
We have the following result.

Fact 4 *Assume that $r < \rho := wC/W$ where W is the sum of the scheduler weights. Then the maximum queueing delay per packet is*

$$\frac{s}{\rho} + \frac{L}{C}$$

where L is the maximum number of bits in a packet/

Proof:

This result is a direct consequence of Facts 1-3. □

References

[1] A. K. Parekh and R. G. Gallager. A generalized processor sharing approach to flow control in integrated service networks : The single node case. IEEE/ACM Transactions on Networking, 1(3), June 1993.