BUFFER AND BANDWIDTH ALLOCATION FOR DIFFSERV CLASSES

Seyyed M-R Mahdavian and Alberto Leon-Garcia
University of Toronto
Dept. of Electrical and Computer Engineering
Toronto, Ontario, Canada M5S 3G4
{ seyyed,alg } @nal.utoronto.ca

Abstract This paper proposes an optimal method for allocating buffer and bandwidth to different classes of a forwarding engine within a Differentiated Services domain in the Internet. The optimality criterion is based on the cost of the buffer and bandwidth. Based on this criterion, the best class that matches a certain traffic with a certain statistical characteristics and the maximum packet delay corresponding to this class is found. The results are general and can be applied to other networks such as ATM.

Keywords: Diffserv classes, Resource allocation, Pareto distribution

1. INTRODUCTION

The current Internet 'best effort' service does not provide any guarantee for packet loss or delay and is therefore not suitable for demanding applications such as high quality voice and video. The 'Differentiated Services' model [1], or diffserv, is a recent proposal to solve the above problem and with its simple architecture is particularly appealing for high speed routers carrying a large number of connections.

In diffserv, packets are differentiated based on the contents of the 'DS field' [2]. The DS field of a packet indicates the treatment or the so-called 'per hop behavior' (PHB) that this packet will receive at each diffserv node. Like ATM networks, it is desirable to define different classes of traffic and different loss priorities within each class. Each class and priority can then be associated with a certain PHB. A possible classification of PHB's is described in [3].

This paper focuses on matching different traffic to different classes. Using a simple mathematical model, we obtain a relationship between the characteristics of the traffic and the diffserv class that is optimum for serving this traffic.

D. H. K. Tsang et al. (eds.), *Broadband Communications*

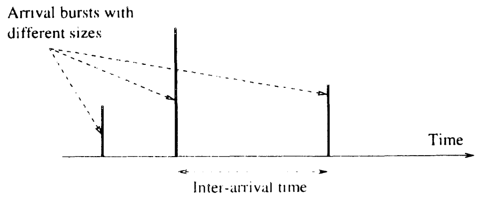

Figure 1 Illustration of the model

The optimality criterion is based on the cost of the buffer and the bandwidth available at a diffserv node. To model the traffic, we use both exponential and Pareto distributions for the burst size and the inter-arrival time. The Pareto distribution is a heavy-tail distribution which can simulate the fractal behavior present in the Internet traffic [4].

The rest of the paper is organized as follows: In section 2 we describe our mathematical model and assumptions. The method used for the analysis of the model is then explained in section 3. In section 4 we discuss and interpret numerical results. Finally in section 5 we summarize our findings.

2. DESCRIPTION OF THE MODEL

We consider a diffserv node allocating a certain amount of bandwidth and buffer to a diffserv class C. It is assumed that each diffserv class has a separate queue for incoming packets. The incoming traffic of class C is modelled as a sequence of bursts which arrive at the node as shown in figure 1. It is assumed that when the first bit of a burst arrives at the node, the whole burst is instantly loaded into a buffer with capacity X, if there is enough space available in the buffer, otherwise the fitting part of the burst is buffered and the excess part is lost. In practice, bursts are not transferred into the buffer instantly; the transfer time is equal to the burst size divided by the arrival rate of the burst. However, given that the average incoming rate of a single class is a small percentage of the whole traffic rate, the peak rate of a burst can be very high as compared to the average incoming rate. Therefore bursts can be considered to arrive almost instantly. This assumption is especially true when the input port of the node is connected to a high speed LAN.

It is furthermore assumed that the size of the burst is a continuous random variable with probability density function $u(x)$, independent of other bursts and the inter-arrival time between bursts. The inter-arrival time is also assumed to be a continuous random variable independent of other random variables. The buffer is depleted at constant rate during the inter-arrival time according to the bandwidth allocated to the class C. One can imagine an equivalent model in which there is an infinite sequence of 'pump up' and 'pump down'

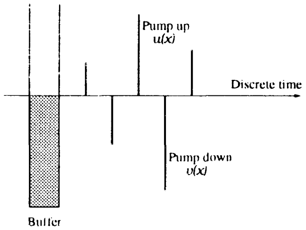

Figure 2 An equivalent representation of the model

cycles taking place at the buffer, as shown in Figure 2. The buffer content is increased or pumped up by a random number with probability density function $u(x)$ (to not more than X, the buffer size,) immediately followed by a pump down or decrease by a random number with probability density function $v(x)$ (to not less than zero) associated with the inter-arrival time and the allocated bandwidth for the class. Our model is therefore similar to the fluid flow model in [5] in that it does not capture the discrete nature of packets. This however has been shown to have negligible effect when the input buffer is large.

Two different distributions for $u(x)$ and $v(x)$ have been studied: The exponential and the Pareto distributions. It has been shown that using the Pareto distribution for the burst size [6] or the packet inter-arrival time [7][8] effectively simulates a self-similar behaviour which has been observed in many practical networks.

3. ANALYSIS METHOD

Given $u(x)$ and $v(x)$ in the model described in section 2, we wish to find the loss rate of incoming bursts. The loss rate is defined as the average of that part of a burst which does not fit into the buffer, divided by the average of a burst size:

$$\text{Loss rate} = \frac{\text{E(loss)}}{\text{E(burst)}} \tag{1}$$

E(loss) and E(burst) are given by:

$$\text{E(burst)} = \int_0^\infty xu(x)dx \tag{2}$$

$$\text{E(loss)} = \int_0^X f(x)E_c(X-x)dx + qE_c(X) \tag{3}$$

where:

$$\begin{aligned} f(x) &= \text{Equilibrium buffer state pdf} \\ E_c(x) &\stackrel{\text{def}}{=} \int_x^\infty (y-x)u(y)dy \\ q &= \text{Equilibrium probability of buffer being empty} \\ X &= \text{Buffer size} \end{aligned}$$

with $f(x)$ and q being observed at the time just prior to an incoming burst. Finding $f(x)$ and q is not trivial. They satisfy the following integral equation:

$$f(x) = \int_0^X K(x,y)f(y)dy + qK(x,0) \tag{4}$$

where $K(x,y)$ is the equilibrium pdf of the buffer state just prior to an incoming burst, given that the state of buffer just prior to the previous burst is y. $K(x,y)$ is given by the following formula:

$$K(x,y) = \int_{\max(x,y)}^X u(z-y)v(z-x)dz + v(X-x)\int_{X-y}^\infty u(z)dz \tag{5}$$

Equation 4 is a Fredholm integral equation of the second kind [9]. As an approximation, one can assume that the buffer is infinite and that loss happens when the buffer state is greater than X. In this case, Equation 4 becomes similar to the Wiener-Hopf equation [10] for which a closed form solution exits when $u(x)$ and $v(x)$ are both exponential. However, since we are interested in the Pareto distribution as well, we shall focus on a numerical solution to Equation 4.

By sampling $f(x)$ at N discrete points, Equation 4 can be transformed into a set of linear equations:

$$f_i = \sum_{j=1}^N A_{ij} f_j + qD_i \tag{6}$$

$$1 = q + \frac{X}{N}\sum_{i=1}^N f_i \tag{7}$$

Equations 6 and 7 can then be solved using any linear algebra software package. The coefficients A_{ij} and D_i can be computed using a numerical integration package such as [11]. E(loss) can therefore be approximated using discrete samples of $f(x)$ obtained from Equations 6 and 7. Simulation results showed that for $N = 50$, a fair approximation can be obtained.

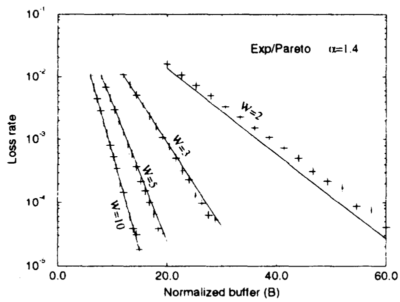

Figure 3 Comparison of the analysis with simulation, Exp/Pareto case

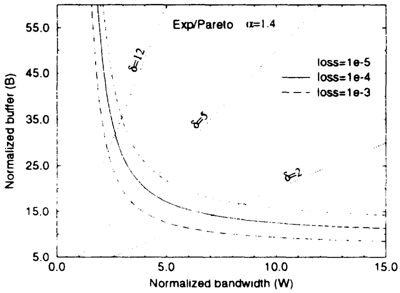

Figure 4 Buffer/Bandwidth tradeoff for different loss rates

4. RESULTS AND DISCUSSION

Consider a class C with a certain allocated bandwidth and buffer. Let B be the *normalized* allocated buffer, i.e. the ratio of the buffer size to the average input burst. Let W be the *normalized* allocated bandwidth, i.e. the ratio of the bandwidth to the average input rate. Given B and W, $u(x)$ and $v(x)$ can be determined for the exponential or Pareto distributions and the loss rate can be computed using the analysis of Section 3. Figure 3 shows the result of this analysis for $u(x)/v(x)$ being Exponential/Pareto and $N = 50$. Marks correspond to simulation results and lines correspond to the analysis.

Given a certain required loss rate, there is a tradeoff between buffer and bandwidth. The tradeoff between buffer and bandwidth has been previously examined in other contexts in the literature [12] [13] [14] [15]. In this paper, we are interested in the effect of this tradeoff on the optimal configuration of a diffserv node.

Figure 4 explicitly shows the tradeoff between buffer and bandwidth for the Exp/Pareto distribution case and for different loss rates. It is not economical

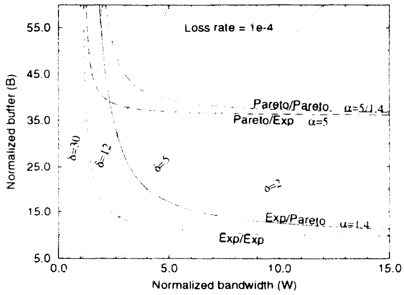

Figure 5 Buffer/Bandwidth tradeoff for different distributions

to push the (W, B) operating point toward the limits. If the buffer is too small, the cost of the bandwidth becomes very high without much gain in the buffer cost. Conversely, if the bandwidth is too low, the cost of the buffer will be too high without gaining much on the cost of the bandwidth. Therefore, the optimal operating point must be around the knee of the $B(W)$ tradeoff curve.

Let d_c be the maximum queueing delay of a diffserv class defined as:

$$d_c \stackrel{\text{def}}{=} \frac{\text{buffer size}}{\text{output rate}}$$

Let d_i be the 'input delay' defined as:

$$d_i \stackrel{\text{def}}{=} \frac{\text{average input burst}}{\text{average input rate}}$$

Then the 'normalized maximum delay' δ is defined as:

$$\delta \stackrel{\text{def}}{=} \frac{d_c}{d_i} = \frac{B}{W} \tag{8}$$

The dotted lines in Figure 4 correspond to different values of the normalized maximum delay δ. This figure shows that for the Exp/Pareto case, the knee of the curve corresponds to the value of $\delta \approx 5$, regardless of the desired loss rate value. Figure 5 compares different distributions with the same loss rate.

The Buffer/Bandwidth optimization can be made more precise. Let $\mathcal{C}$ be the normalized cost of allocating a normalized bandwidth W and a normalized buffer B. i.e.

$$\mathcal{C} = \kappa W + B \tag{9}$$

where κ is the factor by which bandwidth costs more than buffer, i.e.:

$$\kappa = \frac{\text{cost of unit } W}{\text{cost of unit } B} = \frac{\text{cost of unit bandwidth}}{\text{cost of unit buffer}} \times \frac{1}{d_i} \tag{10}$$

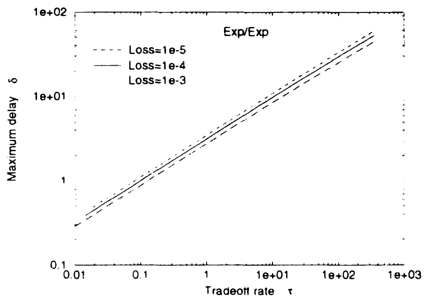

Figure 6 Maximum delay for different loss rates

To find the optimal point, we set the derivative of the cost function to zero:

$$\frac{dC}{dW} = 0 \Rightarrow \kappa + \frac{\partial B}{\partial W} = 0 \Rightarrow \left|\frac{\partial B}{\partial W}\right| = \kappa \tag{11}$$

If the absolute value of the slope of the $B(W)$ tradeoff curve in Figure 4 or 5 is called the 'tradeoff rate' τ, i.e.

$$\tau \stackrel{\text{def}}{=} \left|\frac{\partial B}{\partial W}\right| \tag{12}$$

then Equation 11 shows that the optimal point on the $B(W)$ tradeoff curve is the point where $\tau = \kappa$.

Figure 6 shows τ vs. δ for the Exp/Exp case with different loss rates. It can be seen that different loss rates will cause almost the same delay at a diffserv node. It can also be seen that the $\delta(\tau)$ curve is a straight line in the log/log scale. This has some important implications which we will now investigate.

Suppose that the slope of the line in Figure 6 is equal to $1/(1+a)$ (the slope in this figure is actually 0.5, corresponding to $a = 1$.) Also suppose that the line passes through the point $(\tau = b, \delta = 1)$. Then one can write the equation for this line as follows:

$$\log(\tau) = (1+a)\log(\delta) + \log(b)$$

Taking out the log and replacing δ and τ from Equations 8 and 12 yields the following differential equation for B as a function of W:

$$\frac{dB}{dW} = -b\left(\frac{B}{W}\right)^{1+a}$$

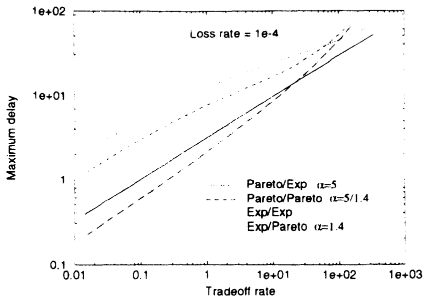

Figure 7 Maximum delay for different distributions

Solving the above differential equation, one obtains the following equation for the $B(W)$ tradeoff curve:

$$B = \frac{\mu W}{(W^a - \lambda^a)^{\frac{1}{a}}} \quad \text{with} \quad \left(\frac{\lambda}{\mu}\right)^a = b \tag{13}$$

λ and μ correspond to the asymptotes of the curves shown in Figure 5. Since the slope of the lines in Figure 6 is 0.5, Equation 13 in this case reduces to the well know bilinear function. λ is always equal to 1 and μ corresponds to the buffer size that gives the desired loss rate assuming that the buffer is completely empty. The $B(W)$ tradeoff curve is therefore readily available for the Exp/Exp case.

The linearity of the $\delta(\tau)$ curve has another important implication with respect to the delay of bursty traffic. Supposed that the burstiness of the traffic is increased by a factor of σ. Then given that the slope of the $\delta(\tau)$ line is $1/(1+a)$, it turns out that the optimal class delay d_c becomes $\sigma^{\frac{a}{1+a}}$ times larger. For the Exp/Exp case where $a = 1$, the optimal class delay grows as the square root of the traffic burstiness.

The function $\delta(\tau)$ is more complicated in cases other than the Exp/Exp case. Figure 7 compares different distributions for the same loss rate of 10^{-4}. According to this figure, all curves corresponding to different distributions are almost linear with the same slope at low tradeoff rates (when there is plenty of bandwidth available at low cost.) These lines however have a considerable difference in offset. This means that the Buffer/Bandwidth tradeoff curve equation is the same bilinear equation for all distributions near the $B = \mu$ asymptote, with the value of μ itself being considerably different for different distributions.

Different diffserv classes are expected to have different maximum delays and loss probabilities in order to cover a wide range of traffic characteristics and quality of service requirements. Given a certain traffic burstiness, one can use the above analysis to match the diffserv class (and the corresponding set of PHB's) which is 'best' for that traffic from an economical point of view. Assuming that d_i is know for the aggregate traffic as a measure of its burstiness, κ can be determined from Equation 10. Therefore the optimal normalized delay δ can be obtained from a $\delta(\tau)$ curve such as in Figure 7 and the optimal class delay d_c can be computed from Equation 8. Then a class having a maximum delay close to this number can be assigned to that traffic.

The converse problem is a bit more complicated. Here the class delay d_c is known and we are interested in finding the traffic burstiness that best matches this class from an economical point of view. The easiest way to solve this problem is probably by trial and error. Make an initial guess for d_i and insert it in Equation 8. Then use the obtained δ to read the corresponding tradeoff rate τ which can be compared against the optimal value κ obtained from Equation 10 to make an adjustment for the initial guess.

5. CONCLUSION

A simple model for the aggregated traffic of a diffserv class was used to match the traffic with the best buffer/bandwidth configuration allocated to the class. The following results were obtained:

- The value of the desired loss rate has significant impact on the amount of the required buffer and bandwidth, but has little effect on the resulting maximum delay if buffer and bandwidth are configured cost effectively.

- A simple bilinear formula for the buffer/bandwidth tradeoff curve can be derived in the Exp/Exp distribution case. For other cases simple intuitive comparison with the bilinear formula can be made.

- The maximum delay corresponding to the optimal buffer/bandwidth configuration grows as the square root of the burstiness of the traffic if both the burst and the inter-arrival time are assumed to have exponential distribution. For other distributions, the exponent of the growth can be obtained from the slope of the $\delta(\tau)$ curve.

- Matching the traffic with the best class can be done using the $\delta(\tau)$ curve.

References

[1] S. Blake *et al.*, "An Architecture for Differentiated Services," RFC 2475

[2] K. Nichols *et al.*, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," RFC 2474

[3] Juha Heinanen *et al.*, "Assured Forwarding PHB Group," RFC 2597

[4] W. Willinger and V. Paxson, "Where mathematics meets the Internet," *Notices of the American Mathematical Society*, Vol. 45, No. 8, pp. 961-970, Sept. 1998.

[5] D. Anick *et al.*, "Stochastic theory of a data-handling system with multiple sources," *The Bell System Technical Journal*, Vol. 61, No. 8, Oct. 1982, pp. 1871-1894

[6] B. Tsybakov and N. Georganas, "On self-similar traffic in ATM queues: definitions, overflow probability bound, and cell delay distribution," *IEEE ACM Transactions on Networking*, Vol. 5 June '97, pp. 397-409

[7] J. Gordon, "Pareto process as a model of self-similar packet traffic," *IEEE GLOBECOM '95 Proceedings*, Vol. 3, pp. 2232-2236, Singapore 1995.

[8] R. Alexander *et al.*, "Modelling self-similar network traffic," Technical report, Department of Statistics, University of Auckland, http://www.stat.auckland.ac.nz/reports/report95/stat9514.html

[9] G.B. Arfken and H.J. Weber, "Mathematical methods for physicists," San Diego: Academic Press, ©1995

[10] A. Leon-Garcia, "Probability and random processes for electrical engineering," Reading, Mass.: Addison-Wesley, ©1989

[11] Clenshaw-Curtis-Quadrature automatic numerical integration package, http://momonga.t.u-tokyo.ac.jp/~ooura/intcc.html

[12] L. Kleinrock, "The latency/bandwidth tradeoff in gigabit networks," *IEEE Communications Magazine*, Vol. 30, Apr. 1992, pp. 36-40.

[13] S.H. Low, "Equilibrium allocation of variable resources for elastic traffics," *IEEE INFOCOM '98 Proceedings*, Vol. 2, pp. 858-864, San Francisco, USA, 1998.

[14] A. Elwalid *et al.*, "A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node," *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 6, August 1995, pp. 1115-1127.

[15] F. Lo Presti *et al.*, "Source time scale and optimal buffer/bandwidth tradeoff for regulated traffic in an ATM node," *IEEE INFOCOM '97 Proceedings*, Vol. 2, pp. 675-682, Kobe, Japan, 1997.