

Per-flow Delay Performance in Traffic Aggregates

Peerapon Siripongwutikorn[†] and Sujata Banerjee^{†,‡}

[†]Telecommunications Program, University of Pittsburgh

[‡]Hewlett-Packard Laboratories

Abstract— Class-based traffic treatment frameworks such as Differentiated Service (DiffServ) have been proposed to resolve the poor scalability problem in the flow-based approach. Although the performance is differentiated in a class-based basis, the performance seen by individual flows in the same class may differ from that seen by the class and has not been well understood. We investigate this issue by simulation in a single node under FIFO, static priority, waiting time priority, and weighted fair queueing scheduling schemes. Our results indicate that such performance discrepancy occurs especially when flows joining the same class are heterogeneous, which is not uncommon considering that the same type of applications can generate traffic having very different statistical behaviors such as video traffic with different activity levels, or voice traffic with different compression schemes. We found that per-flow delay statistics, including the average and the 99th percentile delay, can be very different from the corresponding class delay statistics, depending on flow burstiness, overall traffic load, as well as the queue discipline. We also propose a solution to reduce the mean delay variance experienced by flows in the same class.

Symposium: Internet Performance

I. INTRODUCTION

To overcome poor scalability and high complexity in per-flow traffic handling, per-aggregate or class-based packet scheduling has been suggested (as in the DiffServ architecture [1]), where QoS differentiation is made on a per-class basis. However, the implicit assumption is that the per-flow performance and the aggregate/class performance is close, which may not hold under certain circumstances, as shown in this paper. Equating the per-flow QoS and the class performance when they in fact could *disagree* degrades the perceived application performance, or deteriorates the integrity of service differentiation. For instance, if the packet delay bound or mean delay of class B is supposed to be twice that of class A, it could happen that some flow in class A sees a delay bound 20% worse than the class while some in class B sees a delay bound 20% better than the class. This certainly is unfair to class A users since the ratio of the delays is now 1.33 instead of 2, as expected. This phenomenon calls for the attention to identify quantitative deviations between the per-flow QoS and the class QoS so that applications can be made aware of what QoS levels to expect from the network. More importantly, understanding what dominant factors impact the degree of deviation could help minimize the differences in the perceived performance among flows.

Despite its importance, relatively little emphasis has been put on the understanding of the per-flow QoS under the per-class traffic treatment. Analysis based on Markovian sources allows the computation of per-flow performance [2], [3], [4] but have

severe computational limitation due to state space explosion and is only available for First-in-First-out (FIFO) scheduling. Nevertheless, results for an aggregate of two MMPP flows [3], [4] and an aggregate of many on-off sources [5] (under FIFO scheduling) reveal that the per-flow packet losses can be very different from each other and that of the class, depending on the buffer size, flow burstiness, utilization, and peak rate. For EDF scheduling, the simulation results in [6] show that burstier flows experience higher losses, and their probability of delay violation does not agree with that of the class. For the delay QoS parameter, Trecordi and Verticale develop a mapping function between per-flow *bounding* delay distributions and the overall delay distribution for a multiplex of flows characterized by the arrival curve in a FIFO queue [7]. The simulation with regulated MPEG traces show that individual flows have very close 99th or higher percentile delay, but not necessarily true at lower percentiles. The proposed bounding, however, tends to be conservative compared to the simulation results at the tail of the distribution.

In this paper, we study the performance of the average delay and the 99th percentile delay in a queue with heterogeneous types of flows under FIFO, Static Priority (SP), Waiting Time Priority (WTP) [8], and Weighted Fair Queueing (WFQ) scheduling, which are some important mechanisms in the context of DiffServ. Having heterogeneous flows in the same class is not uncommon because even the same type of applications which normally go into the same class can have very different traffic characteristics. For instance, video conferencing with high activity level or movement has different traffic patterns from those with low activity level, so do voice applications using different compression schemes. Obtaining an exact quantitative mapping between the per-flow and the class delay in a large scale is very difficult, if not impossible, due to the computational intractability mentioned earlier. We instead attempt to identify in general the factors that impact the disagreement between the per-flow and class performance. This hopefully will provide us with more information and insights regarding when we are able to rely on the class performance to represent the per-flow performance. We show by simulation that (1) using the class performance may not always be enough to achieve the per-flow delay requirement because they differ considerably in some cases, (2) one type of flows may see higher delay statistics than the others, depending on its traffic burstiness, and (3) the choice of queue disciplines affects the difference between the per-flow and class performance.

The rest of the paper is organized as follows. We define in §II a performance metric to compare the per-flow and the class delay performance, and describes the system and the traffic models used in this work. We then present the simulation results

*This research is supported in part by the NSF CAREER award ANI-9702389.

and analyze them in §III. We also propose a method to reduce the difference between the per-flow and class performance, followed by the conclusion in §V.

II. PERFORMANCE METRIC AND SYSTEM MODEL

A. Performance Comparison Metric

To quantify as well as visualize the variability among the per-flow and the class delay statistics, we look at the percentage differences and their absolute mean (or normalized mean absolute deviation), which is denoted by δ_i and $\bar{\delta}$ respectively. Let N be the number of flows a given class. Let d_i be the delay statistic of interest, e.g., average or 99th percentile, of the i th flow, $i = 1, 2, \dots, N$, and D be the corresponding class or overall delay statistic taken from all the N flows. We have $\delta_i = \frac{d_i - D}{D} \times 100$ to represent the percentage difference between the per-flow delay statistic d_i and the class delay statistics D . Then, $\bar{\delta}$, the average of $|\delta_i|$, is given by $\bar{\delta} = N^{-1} \sum_{i=1}^N |\delta_i|$. For convenience, we denote respectively D_{avg} and $D_{0.99}$ as the average class delay and the 99th percentile class delay.

B. System Model

The system model under consideration is a single node with the output link capacity of 45 Mbps supporting two traffic classes, where each class is allocated an infinite buffer. The queue disciplines considered are FIFO, SP (with class 1 having higher priority), WTP, and WFQ. For WTP, we use the delay differentiation ratio of $\frac{c_2}{c_1} = 2$. For WFQ, the same weight is assigned to both classes.

Each queue or class is fed by a number of Interrupted Poisson Process (IPP) flows with different characteristics. The IPP source alternates between on and off states, and the time spent in the on (off) state is exponentially distributed with rate α (β). During the on state, packets arrive according to a Poisson process with rate λ and the source stays idle during the off state. The IPP model has been used to model various types of applications including voice and data (e.g., [9], [10]).

C. Traffic Source Parameters

In our study, the individual IPP flows belong to either one of the two types: type A and type B . The parameter values of these two types of IPP sources are adapted from [10], which are derived from realistic voice and data traffic. Obviously, *real voice and data sources have very differing QoS requirements and would normally not be aggregated. We use them here only as examples of traffic flows with specific characteristics.* As mentioned earlier, even the same type of applications which typically go into the same class can have very different traffic characteristics. Further, in a realistic operational network, the traffic class may support flows of different traffic types with different arrival processes but with some common QoS goals. The issue of deciding which types of flows belong to a particular class is out of the scope of this paper, though some of our results will provide some insights on this. We also do not consider in this paper the issue of dimensioning resources to provide a specific level of QoS. We emphasize that voice and

data are used as only two examples of traffic flows with specific characteristics. Our intent is to solely compare the class and the per-flow performance by heterogeneous flows joining the same queue.

The type A source parameters are: $\alpha = 1.538 \text{ sec}^{-1}$, $\beta = 2.841 \text{ sec}^{-1}$, $\lambda = 55.55 \text{ packets/sec}$. Type A packets have a fixed size P of 144 bytes. The average rate of type A source ($\lambda P \cdot \frac{\beta}{\beta + \alpha}$) is thus 22.48 Kbps. The type B source parameters are: $\alpha = 5.0 \text{ sec}^{-1}$, $\beta = 1.25 \text{ sec}^{-1}$, $\lambda = 26.67 \text{ packets/sec}$. Type B packets have a fixed size P of 1,500 bytes. Note that the flow average rate of type B source is approximately three times higher than that of type A source, and also has higher burstiness. We define burstiness as the peak bit rate over the average bit rate ($\frac{\alpha + \beta}{\beta}$), which is equal to 1.54 for type A source and 5.2 for type B source. In our experiments, we assume that each type of aggregates contributes the same amount of the offered load (ρ) when joining the same queue. That is, half of the offered load in the class comes from type A flows while the other half comes from type B flows.

III. SIMULATION RESULTS AND ANALYSIS

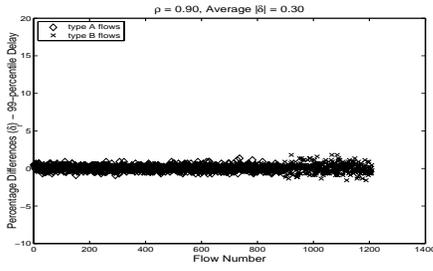
In this section, we look at the the per-flow performance in different scenarios under different scheduling mechanisms. Note that we are not comparing the performance of scheduling disciplines themselves. The simulation is implemented in CSIM. The simulation period lasts until approximately 250,000 packets per type A flow have been collected.

A. Traffic Burstiness

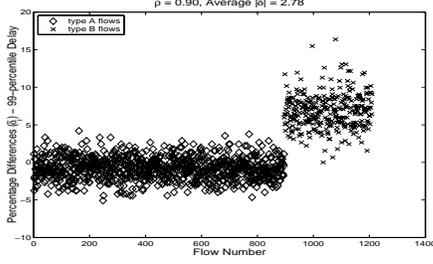
Intuitively, we expect that per-flow delay performance of the smooth traffic flows such as Poisson or constant bit rate will be closer to that of the class than traffic flows exhibiting bursty behavior. We consider two scenarios where a FIFO queue is fed with (1) two types of Poisson flows, and (2) a mix of type A flows and type B flows. In case of Poisson input flows, the first type has the mean packet arrival rate ($\lambda \cdot \frac{\beta}{\beta + \alpha}$) and the packet size equal to those of type A source and equal to those of type B source for flows of the second type. Fig. 1 compares the δ_i of the 99th percentile delay for the above two scenarios. We can see that having Poisson flows with different rates as well as packet sizes in the same queue does not at all affect the per-flow performance, as the individual flows practically experience the same delay bound. In contrast, the per-flow performance for burstier traffic such as IPP depends on the degree of burstiness, as some of the type B flows experience the delay bound as high as 15% larger than the overall, and most of the type B flows tends to see higher delay than the type A flows do.

B. Static Priority

Unless stated otherwise, we assume herein that each class is fed by a mix of type A flows and type B flows, where each type of flows contributes an equal amount of the class load, and the total nodal load (ρ) is equally split between the two classes. The $\bar{\delta}$'s over the range of ρ from 0.15 to 0.9 are well below 5% (not shown here). However, those numbers could be misleading because the actual percentage differences of individual flows (δ_i)



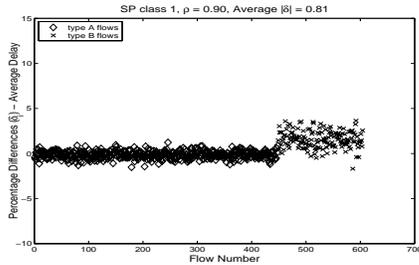
(a) Poisson Traffic, $D_{0.99} = 7010 \mu s$



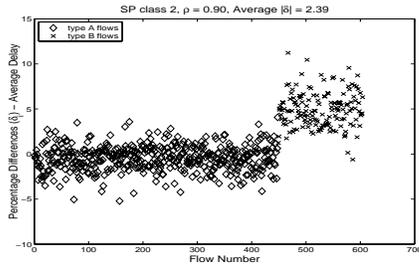
(b) IPP Traffic, $D_{0.99} = 11011 \mu s$

Fig. 1. δ_i of heterogeneous Poisson and IPP flows in FIFO

can be far from their average value and fluctuate highly especially in the high load condition. Fig. 2 and 3 plot the δ_i of the per-flow average delay and the 99th percentile delay at $\rho = 0.9$. Apparently, the per-flow performance in the two classes as well as between two types of traffic within the classes are not the same. The δ_i of flows in the low priority class are higher and have more variation, and within the same class, the burstier traffic, namely type B flows, tend to see worse delay performance than the class. In a low to medium load range, the per-flow performance in the two classes are similar (not shown here due to space limitation) with negligible values of δ_i but still, the type B flows see higher delay.



(a) High Priority, $D_{avg} = 242 \mu s$

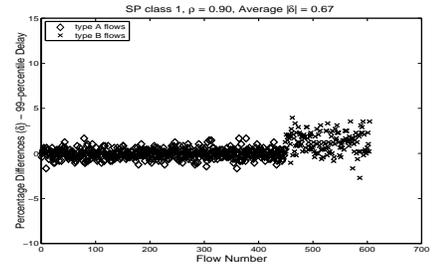


(b) Low Priority, $D_{avg} = 3586 \mu s$

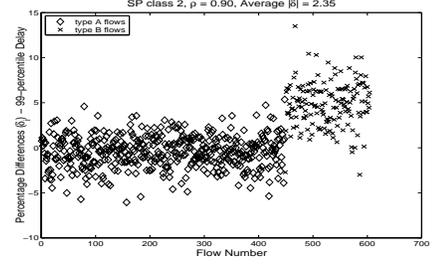
Fig. 2. δ_i of the per-flow average delay under SP

C. WTP

In WTP, the delay performance is of major interest because the main goal of WTP is to achieve the delay differentiation



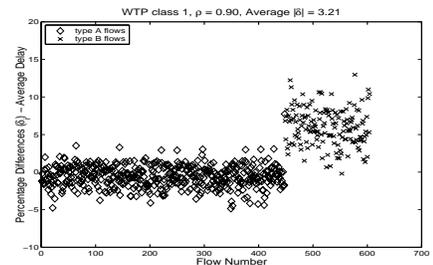
(a) High Priority, $D_{0.99} = 1584 \mu s$



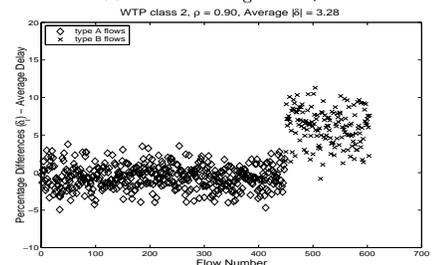
(b) High Priority, $D_{0.99} = 29573 \mu s$

Fig. 3. δ_i of the per-flow 99th percentile delay under SP

among classes. At $\rho = 0.9$, WTP is able to attain a good delay differentiation ratios for both average delay ($\frac{1485}{786} = 1.89$) and 99th percentile delay ($\frac{14042}{7084} = 1.98$). Fig. 4 and 5 plot the δ_i of the per-flow average delay and the 99th percentile delay at $\rho = 0.9$. In contrast to SP where the per-flow performance in the high priority class fluctuates much less from the class, the per-flow delay behaviors of both classes are similar in WTP. However, as in SP, the burstier traffic evidently still experiences higher delay as indicated by the δ_i of the type B flows clustered above those of the type A flows in the same class. Again, at a low to medium load condition, the δ_i 's are well below 5-6% (not shown here) but the type B flows still see higher delay.



(a) Class 1, $D_{avg} = 786 \mu s$



(b) Class 2, $D_{avg} = 1485 \mu s$

Fig. 4. δ_i of the per-flow average delay under WTP

D. WFQ

Since we assume that both classes are assigned the same weight, only the results for class 1 are presented. In contrast to

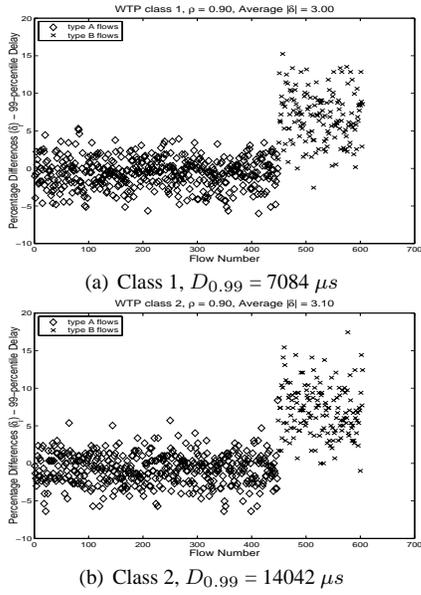


Fig. 5. δ_i of the per-flow 99th percentile delay under WTP

SP and WTP, the per-flow performance discrepancy in WFQ is somewhat less as shown in Fig. 6, noticing that the type B flows perceive the delay performance that are closer to the class.

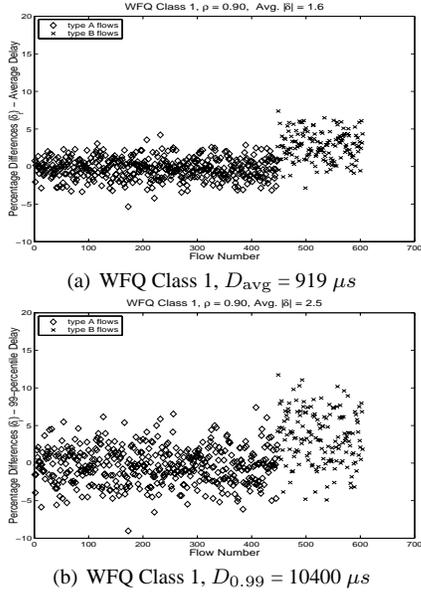


Fig. 6. δ_i of the per-flow average and 99th percentile delay under WFQ

E. Remarks

The simulation study in [11] indicates that when the QoS guarantee is to be satisfied, mixing voice traffic with other kinds of traffic such as TCP into the same queue can appreciably reduce the maximum achievable utilization compared to the case when having voice traffic alone. On the other hand, having voice and real-time video traffic in the same queue still yields acceptable utilization levels. However, the QoS guarantee is made in a class basis, not individual flow. In such a case, our results suggest that the differences between the class and the per-flow delay statistics can be different, potentially leading to the per-flow QoS requirements not being satisfied.

It is interesting to point out that in the various scenarios studied above, up to 50% of overall flows receive a delay QoS that is *better* than the corresponding class QoS (characterized by negative δ_i values). The type A flows receive disproportionately better performance, with up to 80% of the type A flows in some cases receiving better performance than the class QoS. The type B flows on the other hand are penalized, with only at most 20% of the type B flows having negative δ_i . In most cases, less than 5% of the type B flows received better performance than the class.

IV. VARIABILITY REDUCTION

We have previously shown that traffic flows in the same class receive unequal treatment depending on their statistical characteristics. For the average delay performance, we propose applying the Mean Delay Proportional (MDP) scheduling [12] to equalize the mean delay experienced by different types of flows within a class. While the algorithm and implementation given in [12] assumes a fixed packet size and slotted time assumptions, we present them in a more general form to handle a variable packet size. MDP can be combined with an *interclass* scheduling (e.g., SP, WTP, WFQ) as follows. Assume that traffic flows in the same class can be classified into different queues according to their statistical characteristics. Note that this classification can be quite coarse, e.g., based on the peak to mean burstiness ratios, the mean rate, etc. Such classification could be performed by looking up transport port numbers because traffic flows using the same port tend to be generated from the same type of applications and very likely to have similar statistical properties. This kind of classification will only add minor complexity. The interclass scheduling picks the class that is eligible to transmit a packet. Then, intraclass MDP scheduling decides which queue in that class to retrieve an eligible packet from.

For the intraclass MDP scheduling, a separate FIFO queue is assigned to each traffic type, where the Head-of-Line (HOL) packets are the candidates to be eligible. The queueing delay of all the previously departing packets as well as the queueing delay of those waiting in the queue are added together, and then averaged. The HOL packet from the queue with higher average delay is selected for that class. The sum of the queueing delay of the previously departing packets can directly be obtained. However the queueing delay of those waiting in the queue is unknown because it depends on future arrivals. Therefore, this delay is calculated by using the best-case approximation where the packets are serviced back to back. Denote t the current time. Considering the i th packet in the queue counted from the head (including the one being in service), its approximated queueing delay at time t is given by $d_i(t) = (t - a_i) + \text{Time to serve } (i - 1) \text{ packets ahead}$, where a_i is its arrival time to the queue. Since we assume that all the packets are served back to back after time t , it follows that $d_i(t) = (t - a_i) + \frac{Q_i(t)}{C}$, where $Q_i(t)$ is the queue size seen by the i th packet at time t , and C is the link capacity. Therefore the delay sum of $N(t)$ packets waiting in the queue at time t is

$$D(t) = \sum_{i=1}^{N(t)} d_i(t) = Nt - \sum_{i=1}^{N(t)} a_i + C^{-1} * \sum_{i=1}^{N(t)} Q_i(t) \quad (1)$$

To maintain the term $\sum_{i=1}^{N(t)} Q_i(t)$, we can cumulatively add the queue size at every packet arrival instant and decrease it by the packet size of every packet departure instant. Such operations are very simple and thus should not hinder the implementation at high speed. In the special case of fixed packet size, (1) reduces to

$$D(t) = Nt - \sum_{i=1}^{N(t)} a_i + T_p \frac{N(t)(N(t)-1)}{2}. \quad (2)$$

where T_p is one packet transmission time. So $D(t)$ is easily calculated because all the variables are readily obtained. Note that both the terms $\sum_{i=1}^{N(t)} a_i$ and $\frac{N(t)(N(t)-1)}{2}$ can be easily updated. Applying MDP as an intraclass scheduling with WTP equalizes the mean delay of flows in the same class very well compared to WTP alone, as shown in Fig. 7.

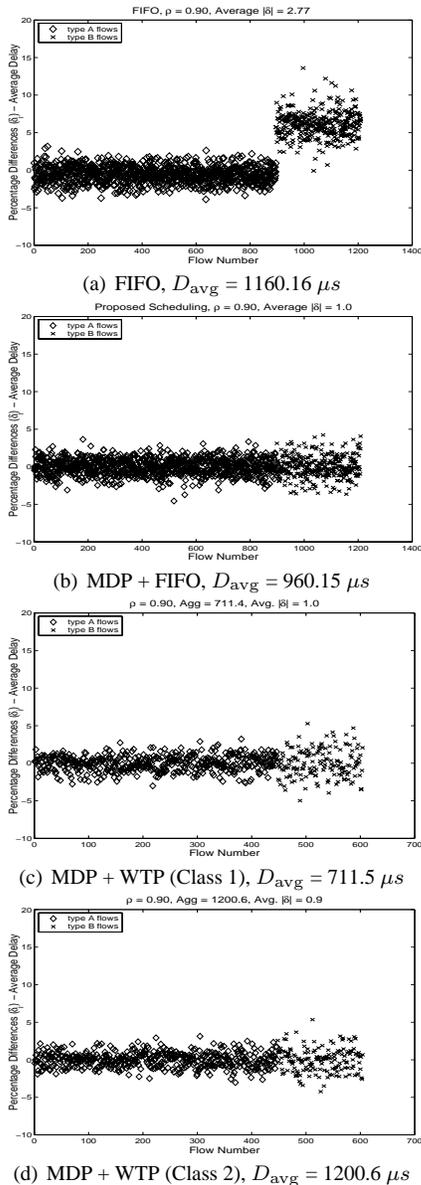


Fig. 7. δ_i of the average delay when applying MDP within the class

V. CONCLUSION

We have investigated the per-flow average and 99th delay performance compared to the class delay performance in a queue with two different types of flows under FIFO, SP, WTP, and WFQ scheduling. Our study indicates that traffic heterogeneity, load condition, and the scheduling discipline affect the per-flow delay performance. When a traffic class supports heterogeneous flows, the main findings are that flows with relatively high burstiness tend to experience higher delay and the performance discrepancy between individual flows and the class increases with the amount of offered load. The per-flow behavior also depends on the scheduling discipline used. For example, the per-flow behaviors in different classes are similar in WTP and WFQ but not in SP. Flow characteristics seem to have less effect to the per-flow performance in WFQ than the others. We also propose using MDP as an intraclass scheduling scheme to equalize the mean delays seen by the diverse flows in the same class. Our study suggests that *we may not be able to achieve delay guarantees for some individual flows based solely on the class delay performance when the flows are heterogeneous in a high load condition*. If we were to admit flows based on the class delay performance, very possibly some differences in the perceived per-flow performance could be expected. This implies we should avoid mixing traffic with different statistical behaviors into the same class, even if they belong to the same application type.

REFERENCES

- [1] Y. Bernet et.al., "A Framework for Differentiated Services," IETF working draft <draft-ietf-diffserv-framework-02.txt>, Feb. 1999.
- [2] W. Fischer and K. Meier-Hellstren, "The Markov-modulated Poisson Process (MMPP) cookbook," *Perf. Eval.*, vol. 18, pp. 149–171, 1992.
- [3] J.J. Bae, T. Suda, and R. Simha, "Analysis of a Finite Buffer Queue with Heterogeneous Markov Modulated Arrival Process: A Study of the Effects of Traffic Burstiness on Individual Packet Loss," in *Proc. IEEE INFOCOM'92*, Florence, Italy, 1992, pp. 219–230.
- [4] E. Dutkiewicz and P. Boustead, "Analysis of Per-flow and Aggregate QoS in Scalable QoS Networks," in *Proc. IEEE Intl. Conf. Networks*, 1999, pp. 289–294.
- [5] Y. Xu and R. Guérin, "Individual QoS versus Aggregate QoS: A Loss Performance Study," in *Proc. IEEE INFOCOM'02*, New York, NY, June 2002, vol. 3, pp. 1170–1179.
- [6] V. Sivaraman, F.M. Chiussi, and M. Gerla, "End-to-End Statistical Delay Service under GPS and EDF Scheduling: A Comparison Study," in *Proc. IEEE INFOCOM'01*, Anchorage, AK, Apr. 2001.
- [7] V. Trecordi and G. Verticale, "Per-flow Delay Performance in a FIFO Scheduler fed by Policed UDP Sources," *Comp. Commun.*, vol. 23, pp. 309–316, 2000.
- [8] C. Dovrolis, D. Stiliadis, and P. Ramanathan, "Proportional Differentiated Services: Delay Differentiation and Packet Scheduling," in *Proc. ACM SIGCOMM*, 1999.
- [9] H. Heffes and D. Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," *IEEE J. Select. Areas Commun.*, vol. SAC-4, no. 6, pp. 856–868, Sept. 1986.
- [10] D. Artiges and P. Nain, "Upper and Lower Bounds for the Multiplexing of Multiclass Markovian on/off Sources," *Perf. Eval.*, vol. 27 and 28, pp. 673–698, 1996.
- [11] M.J. Karam and F.A. Tobagi, "On Traffic Types and Service Classes in the Internet," in *Proc. IEEE GLOBECOM'00*, San Francisco, CA, Dec. 2000.
- [12] T. Nandagopal, N. Venkitaraman, R. Sivakumar, and V. Bharghavan, "Delay Differentiation and Adaptation in Core Stateless Networks," in *Proc. IEEE INFOCOM'00*, Tel Aviv, Israel, Apr. 2000.