End-to-End Statistical Delay Service under GPS and EDF Scheduling: A Comparison Study

Vijay Sivaraman
Atoga Systems
49026 Milmont Drive
Fremont, CA 94538, USA
email: vijay@atoga.com

Fabio M. Chiussi
Data Networking Research
Bell Labs, Lucent Technologies
Holmdel, NJ 07733, USA
email: fabio@bell-labs.com

Mario Gerla
Dept. of Computer Science
University of California
Los Angeles, CA 90095, USA
email: gerla@cs.ucla.edu

***Abstract*— Generalized Processor Sharing (GPS) has gained much popularity in recent years as a simple and effective scheduling mechanism for the provisioning of Quality of Service (QoS) in emerging high-speed networks. For supporting *deterministic* end-to-end delay guarantees, GPS is known to be sub-optimal in comparison to the Earliest Deadline First (EDF) scheduling discipline; nevertheless it is often prefered over EDF due to its simplicity. In this paper, using analytical frameworks developed recently in the literature, we reassess the merits of GPS as compared to EDF in the setting of *statistical* delay service.**

Our contributions are threefold. The statistical frameworks in the literature enable the *aggregate* losses (i.e., delay bound violations) at an EDF scheduler to be estimated – our first contribution, therefore, is to develop a mechanism that allows the aggregate losses to translate to *per-flow* guarantees. This is achieved by means of a simple packet *discard* scheme that drops packets fairly when delay violations are imminent at the EDF scheduler. The discard mechanism has a constant complexity and is feasible for implementation in current packet switches.

The ability to derive the per-flow guarantees from the agregate allows a direct comparison between EDF and GPS – our next contribution, therefore, is to show for various traffic mixes with given per-flow loss constraints that EDF offers consistently larger schedulable regions than GPS, both in the single-hop and multi-hop setting.

As our final contribution, we argue that the use of GPS for statistical delay support is *inherently* problematic. We demonstrate that achieving the maximal schedulable regions under GPS could necessitate *dynamic resynchronization* of the GPS weights, an operation considered infeasible for practical implementation.

I. Introduction

Generalized Processor Sharing (GPS) [14], [15] (also known as Weighted Fair Queueing (WFQ) [4]) and Earliest Deadline First (EDF) [6], [20] have emerged as among the most popular packet scheduling schemes for the provision of Quality of Service (QoS) guarantees to real-time communication services in emerging broadband packet-switched networks. Though the exact nature of the QoS guarantees is still under debate, it is generally accepted that real-time services such as voice and video typically require some form of performance predictability in terms of *end-to-end transfer delays*.

GPS is an idealized fluid discipline with a number of very desirable properties, such as the provision of minimum service guarantees to each flow and fair resource sharing among the flows. Additionally, in networks supporting *deterministic* end-to-end delay bounds, a simple Call Admission Control (CAC) procedure [15] can be derived by directly mapping the end-to-end delay requirements to bandwidth guarantees. Due to its powerful properties, GPS has become the reference for an entire class of GPS-related packet-scheduling disciplines, and relatively low cost implementations have started reaching the market. Nevertheless, it has been shown that the tight coupling between rate and delay under GPS in the deterministic setting leads to sub-optimal performance and reduced network utilizations [10].

EDF has long been known in the context of processor scheduling and has more recently been applied to broadband packet switches. In a single-node setting, EDF is known to be the optimal scheduling policy [8], [12] in terms of the schedulable region for a set of flows with given deterministic delay requirements. EDF scheduling in conjunction with per-hop traffic shaping (together referred to as Rate Controlled EDF or RC-EDF) permits the provision of end-to-end delay guarantees [23], and work in [10], [1] has shown that in the *deterministic* setting, RC-EDF can offer substantial performance gains over GPS. However, CAC procedures for EDF in the deterministic regime are considerably more complex than those for GPS, necessitating the use of approximation techniques [7].

Frameworks based on deterministic QoS guarantees are generally accepted to be overly conservative, and of limited practical value as they result in extremely poor network utilizations. Moreover, most real-time applications are typically resilient to infrequent packet losses (i.e., are not unduly hindered if a small fraction, say 10^{-5}, of their packets are excessively delayed or dropped within the network). This necessitates a reassessment of the relative merits and demerits of GPS and EDF in the *statistical* setting, wherein the end-to-end delay guarantees are probabilistic rather than worst-case. In particular, two crucial questions need to be addressed: 1) Does EDF offer any performance gains over GPS in the statistical setting, and if so, how much?, and 2) Does GPS still offer a simple and efficient CAC mechanism as compared to EDF in the statistical setting? To the best of our knowledge, these important questions have not been addressed in the literature.

The difficulty in tackling these questions arises from the fact that the *exact* schedulability criteria for the statistical setting are non-trivial to ascertain (this is in contrast to the deterministic setting where necessary and sufficient conditions are well-established). This necessitates the use of analytical frameworks that employ assumptions and approximations to *estimate* the actual schedulable regions. (Note that simulation methods do not suffice, since the performance of GPS is very sensitive to the choice of weights, and ascertaining the appropriate weights that allow maximal GPS schedulable regions to be realized necessitates an *analytical* understanding of GPS.) The validity of the comparison study of GPS and EDF, therefore, relies heavily upon the *accuracy* of the analytical frameworks in estimating the schedulable regions of the associated scheduling disciplines. In our work, we have employed *multiple* analytical frameworks for each scheduling discipline and found the resulting schedulable regions to be quite consistent; we take this as an indication that the analytical frameworks provide reasonably accurate estimates of the real schedulable regions.

Analytical frameworks have already been developed in the literature for estimating the losses (throughout this work "losses" refers to delay bound violations) at GPS and EDF schedulers. For GPS, we use the analytical frameworks of [5], [11] (based on the Chernoff approximation) and [16] (based on the central limit approximation), while for EDF we employ [19] (based on the Beneš approach) and [16] (based on the central limit approximation). In contrast to GPS, where the losses are computed on a per-flow basis, the statistical frameworks for EDF estimate the loss probabilities over the *aggregate* (i.e., over the *entire* set of flows multiplexed at the scheduler). To make the comparison between GPS and EDF meanigful, therefore, we first address the issue of how the aggregate loss metric relates to the per-flow metrics.

Note that EDF, unlike GPS, inherently lacks the "isolation" mechanism to protect flows from one another. Thus the aggregate EDF losses could be distributed arbitrarily among the flows, making the provision of per-flow guarantees problematic. To overcome this problem, we propose a solution that overlays onto EDF a simple packet *discard* (alternatively known as *pushout*) mechanism that drops packets fairly when delay violations are imminent at the EDF scheduler. We present simulation results to show that our discard policy allows the aggregate losses to be spread fairly among the flows, thereby enabling QoS guarantees on a per-flow basis. Moreover, our discard policy is shown to have a small constant complexity (independent of the number of flows at the switch), and is hence feasible to incorporate into current packet switches.

The use of the above discard mechanism allows us to use the existing analytical methods in the literature to compare the performance of GPS and EDF schedulers for given per-flow statistical QoS requirements. We consider dual-leaky-bucket regulated traffic flows, and show that for given per-flow delay and loss requirements, EDF consistently offers larger schedulable regions than GPS, much like as in the deterministic regime. We also present results demonstrating that the benefits of EDF extend to the end-to-end multinode setting in the presence of appropriate per-hop traffic reshaping.

We then consider the design of the GPS scheduler and its associated CAC mechanism for supporting statistical delay guarantees. The choice of the GPS weights is very central to realizing large schedulable regions, and the analytical frameworks indicate that optimizing the performance of the GPS scheduler might require the flow weights to be resynchronized dynamically as the traffic mix changes. We show that the need for dynamic weight synchronizations is not merely an artifact of the analytical frameworks, but an *inherent* requirement of GPS, in the absence of which its performance is sub-optimal. This requirement imposes a considerable implementation burden and is considered impractical in the packet switches of today, making the use of GPS for supporting statistical delay guarantees problematic.

By showing that EDF allows per-flow statistical QoS guarantees to be realized (by being coupled with a simple discard policy), yields larger schedulable regions than GPS, and does not have the weight resynchronization overheads of GPS, we think that EDF scheduling offers a simple and efficient mechanism for end-to-end statistical delay service support in packet-switched networks.

The rest of the paper is organized as follows: section II gives the requisite background on GPS and EDF scheduling and associated analytical frameworks. The discard policy which allows aggregate QoS under EDF to translate to per-flow guarantees is presented in section III. In section IV we consider various traffic mixes and present numerical results that quantify the performance gains offered by EDF over GPS, both in the single-node and multi-node setting. This is followed by a discussion on the optimal choice of GPS weights and their resynchronization in section V. The concluding remarks are presented in section VI.

II. Background

A. GPS

Every flow i multiplexed at a GPS server serving K flows and operating at rate C is characterized by a positive real number ϕ_i such that for any interval $(\tau, t]$ in which the flow is continuously backlogged

$$\frac{S_i(\tau,t)}{S_j(\tau,t)} \geq \frac{\phi_i}{\phi_j}, \qquad j = 1, 2, \ldots, K$$

where $S_j(\tau, t)$ is the amount of flow j traffic served by the server during the interval $(\tau, t]$. The seminal work on GPS by Parekh and Gallager in [14], [15] (and its extensions by Cruz [2]) establish tight *deterministic* delay bounds for dual-leaky-bucket regulated flows. A number of frameworks for *statistical* delay guarantees have also been developed recently in the literature. Many of these [22], [3], [24] consider *stochastic* traffic models; we, on the other hand, choose the *deterministic* dual-leaky-bucket regulated traffic model for two reasons: 1) it is easy to enforce conformance with the deterministic model (by shaping or policing); in contrast, the statistical models are difficult to enforce, 2) the dual-leaky-bucket description provides a uniform basis for the description of real-time traffic, whereas there is a large heterogeneity in accepted stochastic models for real-time traffic.

In the following, we summarize the two frameworks that we are aware of that permit computation of the schedulable regions under GPS in the presence of dual-leaky-bucket regulated sources with heterogeneous statistical delay requirements: the first by Elwalid and Mitra [5] (and its extension [11]) that is based on Chernoff approximations, and the other by Qiu and Knightly [16] that employs gaussian approximations.

1) Elwalid-Mitra (EM-GPS) Framework: The EM framework [5] considers two heterogeneous QoS classes (the extension to multiple classes is addressed in [11]) multiplexed at a GPS server operating at rate C. Class j $(j = 1, 2)$ flows, k_j in number, offer (p_j, σ_j, ρ_j) dual-leaky-bucket[1] regulated traffic, and have QoS parameters d_j, the delay bound, and L_j, the loss (recall that loss refers to delay bound violations) probability. The traffic model is the fluid rate process which is adversarial while compliant with dual-leaky-bucket regulation, i.e., an on-off process that transmits at peak rate from the instant the token bucket is full till it is empty, and then turns off and remains so till the token bucket is full again. Further, the flows are non-colluding, and have uniformly distributed random phases. The framework determines the maximal schedulable region, i.e., the set of all feasible flow combinations $\mathbf{k} = (k_1, k_2)$ such that the statistical QoS of each flow is satisfied, and gives the design of the GPS weights (ϕ_1, ϕ_2) (or, equivalently, the ratio $\phi = \phi_1/\phi_2$), which helps realize the maximal schedulable region.

The development of the single node analysis is in two phases. In the first phase, each class j flow is characterized by its *effective bandwidth* $e_0^{(j)}$, which corresponds to the minimum rate required by the flow to meet its *lossless* (i.e., without delay violations) delay requirements:

$$e_0^{(j)} = \frac{p_j}{1 + d_j(p_j - \rho_j)/\sigma_j} \tag{1}$$

where it is tacitly assumed that $d_j < \sigma_j/\rho_j$ (otherwise mean-rate allocation $e_0^{(j)} = \rho_j$ suffices). In the second phase, the probability that a flow is unable to obtain its effective bandwidth at a given instant is estimated. The admissible set $\mathcal{A}(\phi) = \{(k_1, k_2) : \text{QoS of all flows is satisfied}\}$ for given GPS weight $\phi = \phi_1/\phi_2$ is characterized by the simultaneous constraints

$$\begin{aligned} P\left[\phi_1 \sum_{i=1}^{k_1-1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{k_2} \xi_i^{(2)} > \phi_1(C/e_0^{(1)} - 1)\right] &\leq L_1, \\ P\left[\phi_1 \sum_{i=1}^{k_1} \xi_i^{(1)} + \phi_2 \sum_{i=1}^{k_2-1} \xi_i^{(2)} > \phi_2(C/e_0^{(1)} - 1)\right] &\leq L_2 \end{aligned}$$

where $\xi_i^{(j)}$ is the activity indicator for flow i of class j, and is a binomial random variable with $P[\xi_i^{(j)} = 1] = 1 - P[\xi_i^{(j)} = 0] = \omega_i^{(j)} = \rho_j/e_0^{(j)}$. The first constraint is derived by tagging a random class 1 flow and ensuring its QoS, while the second concerns itself with QoS for class 2. Since the boundaries of the admissible regions $\mathcal{A}(\phi)$ are typically non-linear, the authors propose linear approximations, based upon a small number (2 to 4) of corner points. The "pinned" corner points $\bar{k}_1^{(1)}$ and $\bar{k}_2^{(2)}$, where $\bar{k}_j^{(j)}$ is defined as the maximum number of class j flows whose QoS is satisfied when there are no flows of the complementary class, are independent of the GPS weight ϕ and are such that

$$\begin{aligned} P\left[\sum_{i=1}^{\bar{k}_1^{(1)}} \xi_i^{(1)} > (C/e_0^{(1)} - 1)\right] &\approx L_1, \text{ and} \\ P\left[\sum_{i=1}^{\bar{k}_2^{(2)}} \xi_i^{(2)} > (C/e_0^{(2)} - 1)\right] &\approx L_2 \end{aligned}$$

The "design" corner points $\bar{k}_2^{(1)}(\phi)$ and $\bar{k}_1^{(2)}(\phi)$ for class 1 and 2 respectively, where $\bar{k}_i^{(j)}(\phi)$ is the maximum number of class i flows such that, for fixed ϕ, a single class j $(j \neq i)$ flow receives its QoS, are such that

$$\begin{aligned} P\left[\sum_{i=1}^{\bar{k}_2^{(1)}(\phi)} \xi_i^{(2)} > \phi(C/e_0^{(1)} - 1)\right] &\approx L_1, \text{ and} \\ P\left[\sum_{i=1}^{\bar{k}_1^{(2)}(\phi)} \xi_i^{(1)} > \frac{1}{\phi}(C/e_0^{(2)} - 1)\right] &\approx L_2 \end{aligned}$$

The "pinned" corner points $(\bar{k}_1^{(1)}, 0)$ and $(0, \bar{k}_2^{(2)})$, and the "design" corner points $(1, \bar{k}_2^{(1)})$ and $(\bar{k}_1^{(2)}, 1)$ are computed using Chernoff approximations, and provide a linear approximation for the boundary of $\mathcal{A}(\phi)$. To characterize the schedulable region $\mathcal{R} = \{(k_1, k_2) : \exists\phi \text{ such that } (k_1, k_2) \in \mathcal{A}(\phi)\}$, the *critical weights* $\phi_c^{(1)}$ and $\phi_c^{(2)}$ are defined such that

$$\bar{k}_2^{(1)}(\phi_c^{(1)}) = \bar{k}_2^{(2)} \quad \text{and} \quad \bar{k}_1^{(2)}(\phi_c^{(2)}) = \bar{k}_1^{(1)} \tag{2}$$

In the "effectively homogeneous" case, i.e., when $\phi_c^{(1)} \leq \phi_c^{(2)}$, the conservative linear approximation $\mathcal{L}_H$ to the schedulable region $\mathcal{R}$ comprises of the triangle formed by the corner points $(\bar{k}_1^{(1)}, 0)$, $(0, \bar{k}_2^{(2)})$ and the origin $(0, 0)$. Moreover, any choice of the weight ϕ in the interval $[\phi_c^{(1)}, \phi_c^{(2)}]$ realizes this entire region. For the "effectively non-homogeneous" case, i.e., when $\phi_c^{(2)} < \phi_c^{(1)}$, the design procedure is more involved. The linear approximation $\mathcal{L}_{NH}$ to the schedulable region $\mathcal{R}$ in this case consists of the concave simplex with the four corner points $(0, 0)$, $(\bar{k}_1^{(1)}, 0)$, $(0, \bar{k}_2^{(2)})$, and the fourth point being the intersection of the line joining $(\bar{k}_1^{(1)}, 0)$ to $(1, \bar{k}_2^{(1)})$ with the line joining $(0, \bar{k}_2^{(2)})$ to $(\bar{k}_1^{(2)}, 1)$. No single GPS weight realizes the entire region $\mathcal{L}_{NH}$; however, two values of the weight suffice, namely $\phi_c^{(1)}$ and $\phi_c^{(2)}$. Therefore as the desired operating point moves, it might be necessary to switch between the two critical weights in order to realize the entire schedulable region.

In the presence of output rate regulation at rate $e_0^{(j)}$ for each class j flow (in other words, the flow is never served at a rate larger than $e_0^{(j)}$ at the GPS server, even if spare capacity is available) the framework can be extended to the multi-node setting. By allocating the flow's entire delay budget to the first node on its path, the CAC and GPS design at all intermediate nodes is identical, and follows the single-node procedure outlined above. The case of general allocation of end-to-end delay budget among the nodes on the path of the flow is computationally undesirable, as it results in a proliferation of classes within the network, leading to an explosion in computational complexity.

[1]The (p_j, σ_j, ρ_j) dual-leaky-bucket descriptor corresponds to the traffic envelope $A_j(t) = \min\{p_j t, \sigma_j + \rho_j t\}$.

The EM-GPS framework (and its extension to multiple-classes in [11]) is therefore an effective mechanism for end-to-end QoS provisioning under GPS. However, note that the framework focuses on computing the probabilities of flows not receiving their lossless effective bandwidths, and this could in general be *higher* than the delay violation probabilities. This is because flows could clear their backlog in the GPS scheduler by receiving service in excess of their effective bandwidth, so that at a later time they do not incur delay violations even if they receive a service rate below their effective bandwidth - such situations are not accounted for within the EM-GPS framework, and as we shall see in our numerical results, this leads to conservative estimates of the GPS schedulable region.

2) Statistical Service Curve (SC-GPS) Framework: This approach, developed by Qui and Knightly [16], provides a reasonably general framework for the statistical analysis of a variety of scheduling schemes, including GPS and EDF. It is based on a stochastic extension of the idea of "service curves" developed by Cruz [2]. Let $B_i(t)$ denote the *statistical traffic envelope* for flow i, defined as a sequence of random variables such that in any interval $[u, u+t)$ the input traffic $A_i[u, u+t]$ satisfies

$$A_i[u, u+t] \leq_{st} B_i(t)$$

where $A_i[u, u+t] \leq_{st} B_i(t)$ (stochastic inequality) denotes $P[A_i[u, u+t] > z] \leq P[B_i(t) > z]$ for all z. Further, let $S_i(t)$ denote the *statistical service envelope*, defined as a sequence of random variables such that in any interval $[u, u+t)$ the *available service* $\tilde{Y}_i[u, u+t]$ (corresponding to the amount of service received by the flow if it were continuously backlogged in the specified interval) satisfies

$$\tilde{Y}_i[u, u+t] \geq_{st} S_i(t)$$

Then the loss probability L_i that the flow i delay exceeds its requirement d_i is upper bounded as

$$L_i \leq P[\max_{t\geq 0}\{B_i(t) - S_i(t+d_i)\} > 0] \tag{3}$$

This framework does not give the design of the GPS weights which maximize the schedulable region; instead, the authors analyze the system assuming the weight assignments are given. They further assume that the flows are partitioned into two subsets $\mathcal{S}$ (sharing) and $\mathcal{I}$ (isolation). Though the partitioning could be arbitrary, the service classes requiring less aggressive statistical services, i.e., which do not wish to exploit spare capacity from other classes, are typically assigned to the isolation set, while those which exploit inter-class resource sharing using their statistical service envelope to admit an increased number of flows into the traffic class are assigned to the sharing class. The authors make the further assumption that by virtue of over-provisioning of resources to the isolation class, the output traffic envelope of the isolation class is almost identical to the input envelope (this is in general an optimistic estimate and could underestimate the loss probabilities). Thus a flow i belonging to the isolation class and assigned a guaranteed rate g_i experiences losses bounded by

$$P[\max_{t\geq 0}\{B_i(t) - g_i(t+d_i)\} > 0] \tag{4}$$

A flow i belonging to the sharing class, on the other hand, experiences losses bounded as in (3), where

$$S_i(t) = \frac{\phi_i}{\sum_{m\in\mathcal{S}}\phi_m}[Ct - \sum_{i\in\mathcal{I}} B_i(t)] \tag{5}$$

The quantity in (3) is computed using the "maximum variance" approximation under gaussian assumptions. Letting

$$\sigma_t^2 = \text{var}\{B_i(t) - S_i(t+d_i)\} \tag{6}$$

$$\alpha_t = \frac{0 - E\{B_i(t) - S_i(t+d_i)\}}{\sigma_t} \tag{7}$$

$$\alpha = \inf_t \alpha_t, \tag{8}$$

the Gaussian approximation for $P[\max_{t\geq 0}\{B_i(t) - S_i(t+d_i)\} > 0]$ yields

$$P[\max_{t\geq 0}\{B_i(t) - S_i(t+d_i)\} > 0] \leq e^{\frac{\alpha^2}{2}} \tag{9}$$

The variance of $B_i(t)$ in (6) can be computed by assuming the adversarial leaky bucket regulated traffic pattern as in the EM-GPS framework.

The SC-GPS framework has numerous drawbacks - 1) it does not give the design of the GPS weights, which is crucial for realizing maximal GPS schedulable regions, 2) the classification of the flows into isolation and statistical classes is arbitrary, especially when all flows desire the advantages of statistical multiplexing; moreover the assumption that the isolation class output traffic envelope is identical to its input envelope could be over-optimistic and grossly underestimate the loss probabilities, 3) it is not clear how this framework can be extended to the multi-node setting. In spite of these shortcomings, the SC-GPS framework provides a useful cross-check to validate the EM-GPS framework.

B. EDF

The EDF scheduling discipline [6], [20] works as follows: each flow i at the switch is associated with a *local* delay bound d_i; then, a flow i packet arriving to the scheduler at time t is stamped with a deadline $t + d_i$, and packets in the scheduler are served by increasing order of their deadline.

In the *deterministic* setting, EDF is known to be the *optimal* scheduling policy at a single switch [8]. The authors in [8], [12] have shown that EDF has the largest schedulable region of all scheduling disciplines. Given K flows, where flow i ($i = 1, 2, \ldots, K$) has traffic envelope $A_i(t)$ and worst-case delay requirement d_i, the EDF schedulability check is given by:

$$\sum_{i=1}^{K} A_i(t - d_i) \leq Ct, \qquad \forall t > 0 \tag{10}$$

where traffic is assumed to be fluid, C denotes the link rate, and $A_i(t) = 0$ for $t < 0$. To extend the advantages of EDF scheduling to the end-to-end setting, the authors in [23] propose the reshaping of traffic at each hop. EDF in conjunction with per-hop traffic shaping (referred to as Rate Controlled EDF or RC-EDF) has been studied in [10] and expressions are derived for the deterministic end-to-end delay bounds in terms of the flow i shaper envelope $E_i(t)$:

$$d_i = d_i^{sh} + \sum_{m=1}^{M} d_i^m \tag{11}$$

where $d_i^{sh} = D(A_i \| E_i)$ denotes the maximum shaper delay and d_i^m is the local scheduler delay bound at the m-th switch for flow i. The maximum shaper delay is incurred only *once*, and is independent of the number of nodes on the path. Equation (11), in conjunction with the single-node schedulability criteria (10) readily leads to an end-to-end CAC framework [10] that guarantees deterministic delay bounds.

Two statistical frameworks for the analysis of EDF in the setting of dual-leaky-bucket traffic models have been developed recently in the literature - one based on the Beneš approach in [19] and the other based on statistical service curves and the gaussian approximation in [16]. We briefly summarize each in turn.

1) Beneš (Beneš-EDF) Framework: The framework of [19], based on the Beneš approach, considers J QoS classes. As before, class j $(j = 1, \ldots, J)$ flows, k_j in number, offer (p_j, σ_j, ρ_j) dual-leaky-bucket regulated traffic and have QoS parameters d_j, the delay bound, and L, the loss probability. (Note that the EDF loss probability is computed over the *aggregate* traffic at the scheduler, not on an individual flow or class basis. In section III we show how the aggregate losses can be made to yield the desired per-flow metrics.) The traffic model is again the adversarial dual-leaky-bucket regulated fluid process, where the flows are non-colluding and have independent random phases.

Under stationarity conditions and the assumption that packets are not discarded (even if they have expired deadline), the following theorem facilitates the computation of the loss probability L at the EDF scheduler:

Theorem 1: [19] Consider the EDF server at a random time 0. Construct a hypothetical system H which discards all class j $(j = 1, \ldots, J)$ traffic arriving in interval $[-d_j, 0)$. Then

$$L = \frac{1}{\rho} P\{Q^H(0) > 0\} \tag{12}$$

where L denotes the stationary probability of delay violations at the EDF server, ρ the server utilization, and $Q^H(0)$ the queue length, at time 0, in the hypothetical system H.
The probability measure in (12) is estimated using the Beneš approach [13], [17]. Let $\nu(x) = P\{Q^H(0) > x\}$ denote the complementary distribution of the queue length at time 0 in the hypothetical system H. Then the following bound is directly obtained from the Beneš method:

$$\nu(x) \leq \int_{u>0} \sum_{0 \leq \lambda < C} (C - \lambda) \phi_u(x + Cu, \lambda) du \tag{13}$$

where

$$\phi_t(w, \lambda) = \frac{d}{dw} P\{A(t) \leq w, \Lambda_t = \lambda\} \tag{14}$$

denotes the joint density of the rate process Λ_t (denoting the arrival rate at time $-t$) and $A(t) = \int_{-t}^{0} \Lambda_t dt$ (denoting the total amount of work arriving in interval $[-t, 0)$). The quantity $\phi_t(w, \lambda)$ is evaluated by first distinguishing the contribution of the work $A(t)$ arriving in $[-t, 0)$ of flows which are on at $-t$ and of flows which are off at $-t$, and then employing the shifted normal approximation (for brevity, we do not present the expressions here). Using the value of $\phi_t(w, \lambda)$ thus computed, the integral in (13) is evaluated numerically. Finally, the estimate of the loss probability is obtained from $L = \frac{1}{\rho} \lim_{x \to 0} \nu(x)$. It has been shown that the analysis provides a very accurate estimate of the losses over a broad range of parameters - this is verified by comparison with simulations for $L > 10^{-6}$ and with the deterministic analysis for $L \to 0$. The drawback of this framework, however, is its computational complexity, which could become unmanageable when the number of classes is large.

The extension to the multi-node case is by means of per-hop traffic reshaping, much like the deterministic setting described earlier. Once the shaper envelope has been decided and the (worst-case) shaping delay computed, the remaining delay budget is split among the schedulers on the flow's path, and a single-node analysis at each hop determines if the flow is admissible into the network. If the losses are infrequent enough at each switch, they are additive over the path, and allow the end-to-end delay and loss requirements of the flow to be met. The choice of appropriate reshaping parameters is crucial to realizing large schedulable regions - for a (p, σ, ρ) dual-leaky-bucket regulated flow the following choice of the dual-leaky-bucket shaper (p', σ', ρ) has been argued to be simple and effective [19]:

$$p' = \frac{p}{1 + d^{sh}(p - \rho)/\sigma}, \quad \sigma' = \sigma - \rho d^{sh} \tag{15}$$

where the shaping delay d^{sh} is chosen to be

$$d^{sh} = \min\left\{ d\left(1 - \frac{1}{h}\right), \frac{\sigma}{\rho} \right\} \tag{16}$$

where d denotes the flow's end-to-end delay requirement and h its hop-length.

2) Statistical Service Curve (SC-EDF) Framework: The statistical QoS framework of [16] based on service envelopes applies to EDF scheduling as well. Letting $B_i(t)$ denote the statistical arrival envelope for flow i, the loss probability L at the EDF scheduler serving K flows is upper bounded by

$$L \leq P[\max_t \{ \sum_{i=1}^{K} B_i(t - d_i) - Ct \} > 0] \tag{17}$$

The quantity on the right is computed by employing Gaussian approximations similar to the ones employed for GPS in (6)-(9). As before, the variance of $B_i(t)$ is computed assuming an adversarial dual-leaky-bucket regulated traffic pattern.

The SC-EDF framework has a reasonably low computational complexity; however, comparison with simulations in [19] show that it is not very accurate, especially at low loss probabilities (in contrast to the Beneš-EDF framework, which is very accurate over the entire range of values). The extension of the SC-EDF framework to the multi-node case is not explicitly discussed in [16], but can be achieved by the per-hop reshaping mechanism as in the Beneš-EDF framework.

TABLE I

FLOW PARAMETERS FOR SCENARIO 2

Class	p (Mbps)	σ (Kbits)	ρ (Mbps)	d (msec)	h
class-0	30	1188	0.3	20	2
class-1	15	250	2.5	10	2

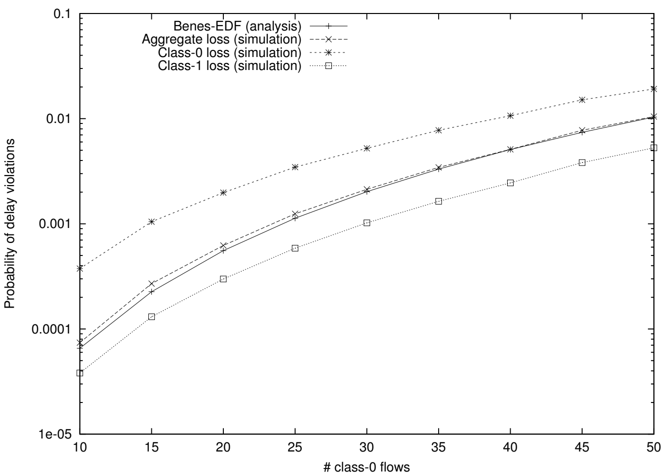

Fig. 1. Class-specific losses under EDF for scenario 2

III. DISCARD POLICY

The analytical frameworks for EDF above focus on the *aggregate* losses at the scheduler – in practice, the losses seen by the individual flows could be quite different from the aggregate. As an example, consider a 100 Mbps link multiplexing traffic from two classes of flows with dual-leaky-bucket parameters as shown in table I. When the number of class-1 flows is fixed at 10 and the number of class-0 flows is varied between 5 and 50, and each flow is assumed to generate extremal on-off traffic, the loss probability for the flows in each class as obtained from simulation is depicted in figure 1. (For validation the aggregate losses are also plotted, and the simulation values are found to be in excellent agreement with those obtained from the Beneš-EDF analysis.) We observe that there is a considerable disparity in the losses experienced by the two classes; class-0 flows, by virtue of being very bursty, experience significantly higher losses than class-1 flows. This makes the provision of per-flow QoS problematic, as the analytical frameworks for EDF in the literature capture only the aggregate QoS behavior at the EDF scheduler.

To overcome the above problem, we propose the use of packet *discard* (alternatively known as *pushout*) mechanisms that allows the per-flow loss metrics to be realized by selectively discarding packets at the EDF scheduler when delay violations are imminent. Numerous discard mechanisms have been proposed and analyzed under various contexts the literature; of particular interest is the study on discard policies that support per-flow delay and loss requirements in the context of EDF scheduling in [21]. A discard policy called G-QoS was presented which operates as follows: it tracks the normalized loss performance $\Delta_f = \delta_f / L_f$ for each flow f, where L_f denotes the desired and δ_f the measured loss probability for flow f. Upon each packet arrival, a test is performed to determine if the set of packets at the EDF scheduler is schedulable (i.e., all packets meet their delay bounds), and if not, a packet from the backlogged flow with the lowest normalized loss performance is discarded.

The authors in [21] show that the G-QoS discard policy is *optimal* among the set of *space-conserving* discard policies (a discard policy is space-conserving if it discards a packet if and only if the set of packets at the EDF scheduler becomes non-schedulable). Moreover, when the traffic flows are equally demanding, G-QoS is shown to be optimal among *all* discard policies, and hence allows the per-flow delay and loss metrics to be realized when used in conjunction with EDF scheduling. However, the need for performing a schedulability check on the *entire* set of packets at the EDF scheduler makes the scheme computationally too complex and infeasible for implementation in high-speed packet switches. In what follows, we propose a discard policy that approximates the performance of the optimal G-QoS scheme, but has a low constant complexity, making it feasible for implementation in practical switches.

BPF-ℓ
/* denote by p_i the packet with the i-th lowest timestamp
  at the EDF scheduler, and by *flow*(p_i) its flowid */
for i from 1 to ℓ
  if p_1 through p_i are not schedulable
    determine flowid *flow*(p_k) $(1 \leq k \leq i)$ having
      minimum normalized performance δ_k
    discard p_k
  end if
end for
end BPF-ℓ

Fig. 2. The BPF-ℓ discard algorithm

Our scheme, called Best Performance First with look-ahead, or simply BPF-ℓ, is shown in figure 2, and is invoked at each scheduling decision instant of the EDF scheduler. It is a generalized version of G-QoS, and restricts the schedulability check to at most ℓ packets, corresponding to the ones with the lowest timestamps in the system. BPF-0 thus corresponds to a scheme which never discards any packet, while BPF-∞ is equivalent to the G-QoS scheme. Obviously the larger ℓ is chosen to be, the more closely BPF-ℓ approximates the behavior of the optimal G-QoS discard scheme.

Observe that BPF-ℓ detects and eliminates all imminent delay violation "bursts" of size ℓ or less (thus G-QoS, which is equivalent to BPF-∞, eliminates delay violations *altogether* by discarding packets whenever delay violations are imminent). If the aggregate delay violation probability at the EDF scheduler is by design small, the *sizes* of the delay violations bursts can in general also be expected to be small. This leads us to expect that even for reasonably small values of the look-ahead ℓ, the BPF-ℓ discard algorithm can detect and eliminate most delay violations, thereby achieving loss performance (note that loss now includes both discards and delay violations) very close to that of the optimal G-QoS scheme. Moreover, the above argument applies independent of the number of flows multiplexed at the scheduler, and hence the performance of BPF-ℓ does not degrade as the number of flows or flow classes

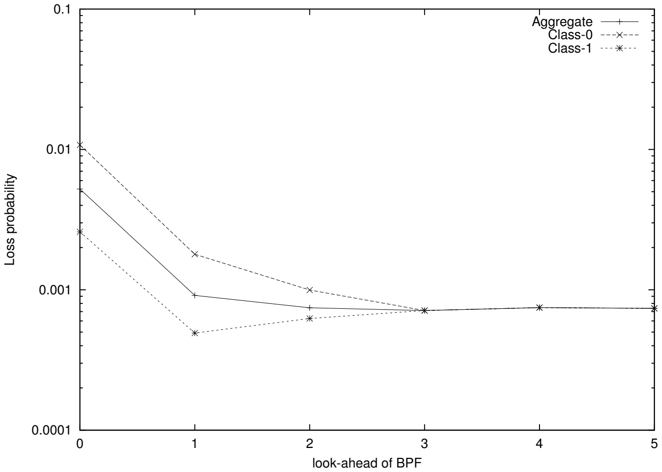

Fig. 3. Class-specific losses for scenario 2 as look-ahead of BPF-ℓ is varied

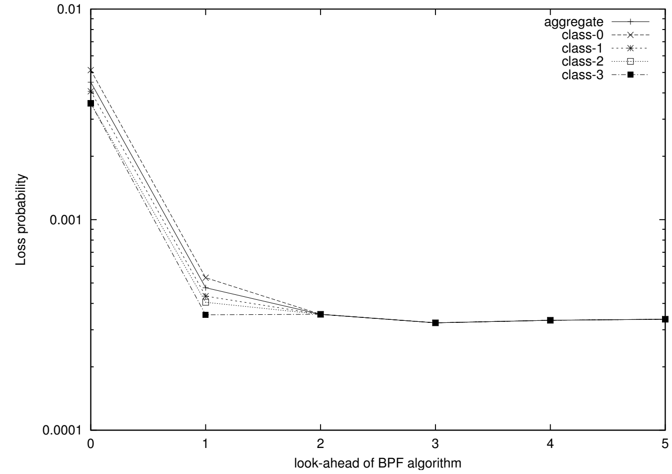

Fig. 4. Class-specific losses for scenario 3 as look-ahead of BPF-ℓ is varied

TABLE II

FLOW PARAMETERS FOR SCENARIO 3

Class	k	p (Mb/s)	σ (Kbits)	ρ (Mb/s)	d (ms)
0 (Soccer)	40	5.0	2133.3	1.0666	80
1 (Terminator)	40	3.4	800.0	0.3666	60
2 (Video Conf)	40	10	80.0	0.5	40
3 (Audio)	40	0.064	1	0.064	20

multiplexed at the switch increases.

To study the impact of the look-ahead on the performance of BPF-ℓ, we consider two traffic mixes. The first is identical to the multi-node setting of scenario 2 considered in table I, for which the disparity in loss probabilities experienced by the two classes was shown in figure 1. We fix the number of class-0 and class-1 flows at 10 and 40 respectively, and the desired loss requirements for both class are set to be identical. Figure 3 plots the aggregate and the class-specific losses as the look-ahead of the BPF-ℓ discard algorithm is increased. Note first that the aggregate losses are a *non-increasing* function of the look-ahead ℓ; this is because the discarding of packets which are doomed to violate their deadlines frees up bandwidth that can be utilized by other packets to meet their delay requirements which could otherwise possibly not have been met. Thus the aggregate losses decrease with increasing look-ahead; when the look-ahead reaches a value large enough to eliminate delay violations altogether, a further increase does not yield any benefits. Now observe the effect of increasing the look-ahead on the class-specific losses. Even though the two classes have very different traffic characteristics and experience quite disparate losses in the absence of discards, a look-ahead as low as $\ell = 3$ suffices for BPF-ℓ to optimally equalize the losses across both classes.

A more realistic traffic scenario consisting of four traffic classes is considered in table II. The Soccer and Terminator video streams have parameters derived from the four-segment characterizations considered in [7] (in turn derived from empirical envelopes in [18]). The audio flows are constant bit rate. All flows are assumed to traverse a single hop (since our object is to study the efficacy of the BPF-ℓ discard policy), and the link rate as before is fixed at 100 Mbps. Figure 4 shows the effect of the look-ahead of the BPF-ℓ discard policy on the aggregate and per-class losses. Again we observe that even in the presence of a larger number of classes a reasonably low look-ahead $\ell = 3$ optimally equalizes the losses across all classes.

We believe that BPF-ℓ is feasible for implementation in current-day packet switches. Identifying the ℓ packets with the lowest timestamps is typically easy since EDF schedulers support fast and efficient structures for sorting timestamps. For fixed ℓ, the BPF-ℓ discard algorithm thus has constant complexity independent of the number of flows at the switch, making it feasible for implementation in high-speed packet switches supporting a large number of flows.

Note that the G-QoS discard scheme, in spite of being optimal, does not guarantee that the losses can be discarded fairly among the traffic classes. In fact, there is no known method of computing whether a given set of per-flow loss guarantees is feasible to achieve given the aggregate loss metric at the EDF scheduler. Nevertheless, for most realistic traffic scenarios, we expect that the BPF-ℓ discard mechanism, even for reasonably low values of ℓ, should fairly distribute the losses across the flows, thereby yielding the desired per-flow loss metrics. This allows the computation of the schedulable regions under EDF as described in the previous section to extend to the setting of per-flow loss guarantees, and enables a direct comparison between EDF and GPS.

IV. COMPARISON OF SCHEDULABLE REGIONS

Using the analytical frameworks for GPS and EDF described in section II, coupled with the discrd mechanism described in the previous section for EDF, we can compare the schedulable regions of the two schedulers for various traffic mixes with given per-flow end-to-end delay and loss requirements. For simplicity of exposition and ease of depicting the schedulable regions, all our scenarios consider only two traffic classes being multiplexed. All traffic is assumed to be fluid (to facilitate the use of the analytical frameworks described above), and link speeds are fixed at 100 Mbps. Unless stated otherwise, the loss probability L for each flow is fixed at 10^{-5}.

TABLE III
FLOW PARAMETERS FOR SCENARIO 1

class number	p (Mbps)	σ (Kbits)	ρ (Mbps)	d (ms)	h
0 (video conf.)	10	80	0.5	80	4
1 (stored video)	10	800	3	180	3

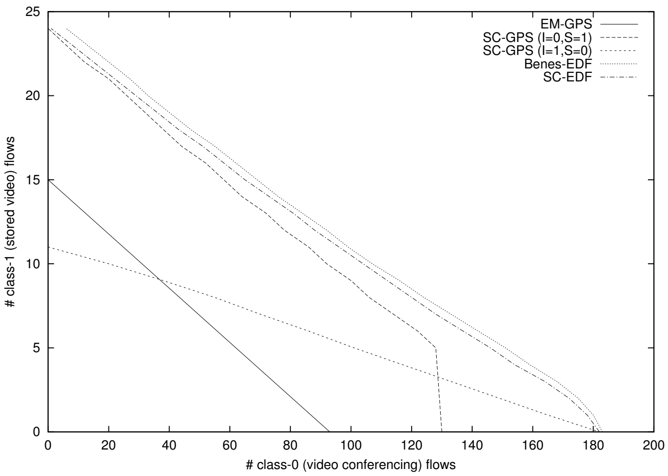

Fig. 5. Single-node GPS and EDF schedulable regions for scenario 1

For the first scenario, we consider a switch multiplexing traffic with parameters shown in table III. The class-0 and class-1 flows are representative of video conferencing and stored video respectively, and have parameters consistent with the ones selected in [9]. The video conferencing flows have a burst size of 10 Kbytes and an average rate of 0.5 Mbps. For stored video, the values are typical from an MPEG trace (of the Star Wars movie), with an average rate of about 3 Mbps and a burst size of 100 Kbytes. The peak rate for both is limited to 10 Mbps (ethernet rate). The video conferencing flows have an end-to-end delay requirement of 80 msec and traverse 4 hops, while the end-to-end delay requirement for the stored video flows is 180 msec and they traverse 3 hops.

We begin by comparing the schedulable regions under for the *single-node* setting, i.e., using the traffic envelopes given in table III and delay bounds of $80/4 = 20$ and $180/3 = 60$ msec respectively for the two classes. Figure 5 plots the schedulable regions of GPS and EDF as computed by the various frameworks for $L = 10^{-5}$.

We first observe that the schedulable region under EM-GPS is significantly smaller than the other frameworks. In fact, it may seem surprising that even in the presence of flows from only *one* class, the number of admissible flows under this framework is significantly lower. This is not a consequence of the GPS scheduler itself, but as noted earlier, arises because the EM-GPS framework computes the probability of flows not receiving their lossless effective bandwidths, which could in general be higher than the delay violation probability.

Under the SC-GPS framework, both classes should ideally be placed in the sharing set $\mathcal{S}$, since both desire the benefits of statistical resource sharing. However, such a partition leaves the isolation set $\mathcal{I}$ empty and yields very loose bounds. This forces us to choose partitions with one class each in $\mathcal{S}$ and $\mathcal{I}$. Further, since this framework does not give the design of the GPS weights, we make the optimistic assumption that the *entire* link capacity is always available to the isolation class, the unused being consumed by the sharing class. Schedulable regions for the two choices (class-0 in $\mathcal{S}$ vs. class-1 in $\mathcal{S}$) are plotted in figure 5, and show that flows when placed in the sharing set extract larger multiplexing gains.

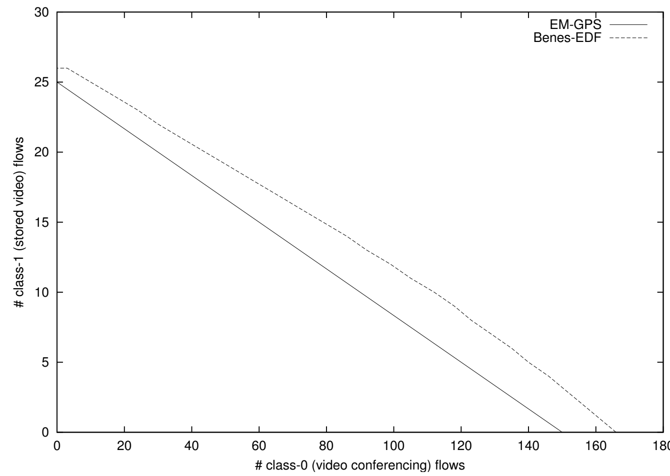

Fig. 6. EDF and GPS schedulable regions for end-to-end QoS scenario 1

The EDF frameworks (Beneš-EDF and SC-EDF) are quite consistent with each other, and yield similar characterizations of the EDF schedulable region. Moreover, this schedulable region is significantly larger than those obtained from the EM-GPS or SC-GPS frameworks (in spite of the excessively optimistic approximations employed under SC-GPS). Also note that the analytical frameworks for EDF do not account for packet discards, and hence provide conservative bounds on aggregate losses at the EDF scheduler. In the presence of discards, even larger schedulable regions than predicted by the conservative EDF analytical frameworks are thus achievable.

It is also worth noting that the shape of the schedulable region under SC-GPS suggests that realizing the maximal GPS schedulable region might require dynamic reassignment of classes to the isolation and sharing sets, or in other words, a dynamic realignment of the class weights. This will be discussed in more detail in the next section.

We now extend the results to the multi-node setting. We consider a switch within the network multiplexing the traffic mix of table III. It is assumed that this switch is the bottleneck and determines the number of flows of each class that can be admitted into the network. For GPS, we consider only the EM-GPS framework; we drop the SC-GPS framework since it is unclear how it extends to the multi-node setting, and moreover does not give the design of the GPS weights that maximize the schedulable regions. As recommended by the EM-GPS framework, we assume that a flow's entire delay budget is assigned to the first hop; the flow is thus "smoothed out" at the ingress and has a zero delay bound at each hop, making the CAC identical at all nodes. For EDF, we consider only the Beneš-EDF framework, since it has been shown to be more accurate than the SC-EDF framework [19]. Further, the per-hop shaping parameters are chosen as per (15) and (16), and

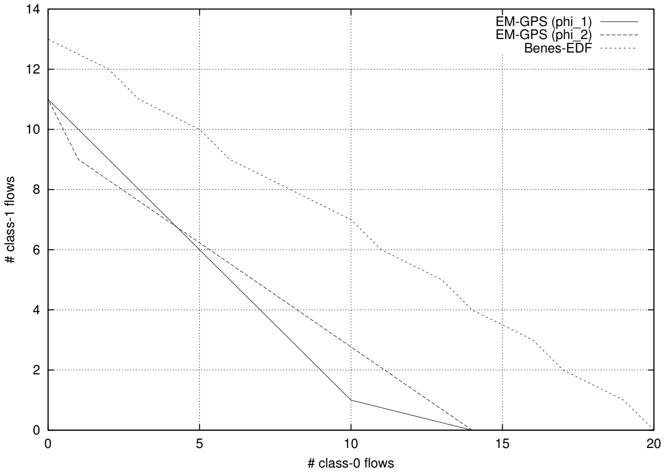

Fig. 7. EDF and GPS schedulable regions for end-to-end QoS scenario 2

the scheduling delay $d - d^{sh}$ is split equally among the hops.

Figure 6 shows the schedulable region at the switch of interest under GPS and EDF from the EM-GPS and Beneš-EDF frameworks respectively. The GPS critical weights are very close to each other and the system is thus almost effectively homogeneous. Nevertheless, the EDF schedulable region is significantly larger than that of GPS, and can realize link utilizations which are around 10% larger than under GPS.

Consider next scenario 2 consisting of the traffic mix in table I. Here class-0 is very bursty and has a high peak-to-mean ratio when compared to class-1. The EM-GPS critical weights satisfy $\phi_c^{(1)} = 2.4 > \phi_c^{(2)} = 2.2$, making the classes effectively non-homogeneous. Figure 7 shows the GPS schedulable regions obtained using each of the two critical weights. We observe that not all traffic mixes can be supported using the same relative GPS weights - for example, the traffic mixes $(1, 10)$, $(2, 9)$, $(3, 8)$ can be realized by the weight ratio $\phi_c^{(1)}$ but not by $\phi_c^{(2)}$, while $(12, 1)$, $(11, 2)$, $(9, 3)$, $(8, 4)$ can be realized using $\phi_c^{(2)}$ but not $\phi_c^{(1)}$. This means that GPS weight assignments *independent* of the traffic mix at the scheduler may not realize maximal schedulable regions, and the weights may have to be *dynamically resynchronized* as the traffic mix varies. Similar observations are reported for GPS schedulers supporting a larger number of traffic classes [11]. The issue of dynamic weight resynchronizations is discussed in greater detail in the next section.

V. Optimal GPS Weights and Resynchronizations

The performance of GPS is very sensitive to the choice of scheduler weights; it therefore becomes very critical to choose appropriate weights that *maximize* the GPS schedulable region. The EM-GPS framework (and its extension to multiple classes in [11]) do give a methodology for identifying "appropriate" GPS weights. However, realizing the maximal schedulable regions under this framework may require the GPS weights to be *dynamically readjusted* as the traffic mix changes. This was observed, for example, in scenario 2 considered in figure 7, where switching between the two critical weights was necessary for the EM-GPS framework to realize the entire concave schedulable region. As the number of traffic classes increase, the problem is found to worsen, and more weight resynchronizations are required [11]. The problem is not confined just to the EM-GPS framework; even the SC-GPS framework tacitly suggests that flows might have to be dynamically swapped between the *isolation* and *sharing* classes as the traffic mix changes – this was observed, for example, for scenario 1 considered in figure 5.

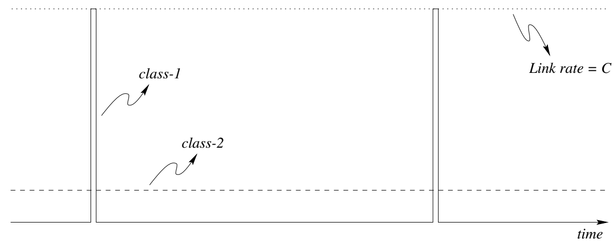

Fig. 8. 2-class traffic scenario demonstrating the need for GPS weight resynchronization

We believe that the dynamic weight resynchronization requirement is not merely a byproduct of the analytical frameworks, but *inherent* to GPS itself. We establish this by presenting an argument below that does not depend on the specific framework employed for analyzing GPS.

Claim: GPS inherently requires dynamic weight resynchronizations in order to realizing maximal schedulable regions.

Proof: We establish the above claim by exhibiting a traffic scenario for which weight resynchronizations are imperative. Consider a GPS scheduler multiplexing two flow classes with traffic profiles as shown in figure 8. Each class-1 flow generates short traffic bursts interspersed with long idle periods, such that the ratio of the average burst duration to the average length of the interval between the start of two consecutive bursts is a small fraction δ. Thus, a class-1 flow is "on" with probability δ and "off" with probability $1 - \delta$. Further, the traffic generation rate during each burst is close to the link rate C. Each class-2 flow, on the other hand, generates traffic at a constant rate. The average rate of a flow of either class is $\rho \ll C$. The loss tolerance L for every flow of either class is chosen such that $\delta < L < 2\delta(1 - \delta)$. Further, the delay requirement of each class-2 flow is very tight ($d_2 \approx 0$), while a class-1 flow has a relatively looser delay requirement d_1, which is such that a class-1 burst meets its delay requirement only if the burst receives service at rate no less than $(C - \rho)$.

Let (k_1, k_2) denote a feasible traffic mix at the GPS scheduler, such that the GPS scheduler can simultaneously support the delay and loss requirements of k_1 class-1 and k_2 class-2 flows. We first claim that the traffic mixes $(1, 2)$ and $(2, 1)$ are feasible. For the mix $(1, 2)$, a relative weight setting $\phi = \phi_1/\phi_2 = \infty$ provides complete isolation to the class-1 flow, allowing it to meet its QoS requirements. Each of the two class-2 flows, meanwhile, experiences losses with probability no greater than δ (the *on* probability of the class-1 flow), and this, by design of L, suffices for it to meet its delay and loss requirements. When the traffic mix is $(2, 1)$, the weight setting $\phi = 0$ provides the class-2 flow the requisite isolation. Using the remaining bandwidth $(C - \rho)$, each of the class-1 flows experiences losses with a probability no more than δ (corresponding to the probability that the other class-1 flow is also generating a burst concurrently).

We now claim that though the traffic mixes $(1, 2)$ and $(2, 1)$ are feasible, no single weight setting $\phi' = \phi_1/\phi_2$ can realize both. To see why, consider first the mix $(1, 2)$. When the class-1 flow is generating its burst, it receives service at rate $\frac{\phi'}{\phi'+2}C$. Since this service rate has to be no less than $(C - \rho)$ in order for the class-1 flow to meet its delay and loss requirements, we require

$$\phi' \geq 2(C/\rho - 1) \tag{18}$$

Now say the weight setting ϕ' remains unchanged while the traffic mix changes to $(2, 1)$. Whenever *any* of the two class-1 flows is generating a burst, the service rate received by the class-2 flow is no more than $\frac{1}{\phi'+1}C$. This, from (18), can be seen to be less than ρ, implying that the class-2 flow experiences losses whenever *any* of the two class-1 flows generates a burst. Since the latter occurs with probability $2\delta(1-\delta)$, and we chose $L < 2\delta(1-\delta)$, the QoS requirements of the class-2 flow are not met. This shows that no *single* weight setting ϕ' can realize both the traffic mixes $(1, 2)$ and $(2, 1)$ though each is feasible at the GPS scheduler. $\triangle$

The above claim shows that dynamic weight resynchronizations are *inherently* required by GPS schedulers in order to realize maximal schedulable regions. This poses a problem since dynamic resynchronizations are too expensive to implement, and are considered impractical in the packet switches of today. In their absence, however, the performance of GPS could be sub-optimal. (The precise quantification of the resulting degradations depends strongly on the analytical framework employed, and is left for future study.) This poses a significant concern in the use of GPS schedulers for provisioning end-to-end statistical delay guarantees.

VI. Conclusions

GPS has gained much popularity in recent years as a scheduling mechanism of choice for end-to-end delay support in emerging high-speed packet-switched networks. In this paper, we have reassessed the merits of GPS when compared to EDF in the setting of *statistical* end-to-end delay service. We have shown that:

- The use of the EDF scheduling (which enables low aggregate losses) in conjunction with a fair discard mechanism (which allows the aggregate losses to translate to per-flow metrics) provides an effective way of realizing larger network utilizations than GPS while still guaranteeing per-flow QoS.
- Realizing maximal GPS schedulable regions is inherently problematic as it requires the dynamic resynchronization of weights, an operation infeasible in practice.

Based on these, we propose EDF as an efficient and simple mechanism for the provision of end-to-end statistical delay service in emerging high-speed packet networks.

References

[1] F. M. Chiussi and V. Sivaraman. Achieving High Utilization in Guaranteed Services Networks Using Early-Deadline-First Scheduling. In *IWQoS '98*, pages 209–217, Napa, CA, May 1998.

[2] R. L. Cruz. Quality of service guarantees in virtual circuit switched networks. *IEEE Journal on Selected Areas in Communication*, 13(6):1048–1056, August 1995.

[3] G. de Veciana and G. Kesidis. Bandwidth allocation for multiple qualities of service using generalized processor sharing. *IEEE Transactions on Information Theory*, 42:268–272, 1996.

[4] A. Demers, S. Keshav, and S. Shenker. Analysis and simulation of a fair queueing algorithm. In *Proceedings of ACM SIGCOMM*, pages 3–12, 1989.

[5] A. I. Elwalid and D. Mitra. Design of generalized processor sharing schedulers which statistically multiplex heterogeneous qos classes. In *Proceedings of INFOCOM '99*, New York, NY, March 1999.

[6] D. Ferrari and D. Verma. A Scheme for Real-Time Channel Establishment in Wide-Area Networks. *IEEE Jrnl. Selected Areas in Comm.*, 8(3):368–379, Apr 1990.

[7] V. Firoiu, J. Kurose, and D. Towsley. Efficient Admission Control of Piecewise Linear Traffic Envelopes at EDF Schedulers. *IEEE/ACM Trans. Netw.*, 6(5):558–570, Oct 1998.

[8] L. Georgiadis, R. Guérin, and A. Parekh. Optimal Multiplexing on a Single Link: Delay and Buffer Requirements. *IEEE Trans. Inf. Theory*, 43(5):1518–1535, Sep 1997.

[9] L. Georgiadis, R. Guérin, V. Peris, and R. Rajan. Efficient support of delay and rate guarantees in an internet. In *Proceedings of ACM SIGCOMM*, Stanford University, CA, August 1996.

[10] L. Georgiadis, R. Guérin, V. Peris, and K. Sivarajan. Efficient Network QoS Provisioning Based on per Node Traffic Shaping. *IEEE/ACM Trans. Netw.*, 4(4):482–501, Aug 1996.

[11] K. Kumaran, G. E. Margrave, D. Mitra, and K. R. Stanley. Novel techniques for the design and control of generalized processor sharing schedulers for multiple qos classes. In *Proceedings of INFOCOM 2000*, Tel Aviv, Israel, March 2000.

[12] J. Liebeherr, D. Wrege, and D. Ferrari. Exact Admission Control for Networks with a Bounded Delay Service. *IEEE/ACM Trans. Netw.*, 4(6):885–901, Dec 1996.

[13] I. Norros, J. W. Roberts, A. Simonian, and J. T. Virtamo. The superposition of variable bit rate sources in an atm multiplexor. *IEEE Journal on Selected Areas in Communications*, 9(3):378–387, April 1991.

[14] A. Parekh and R. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case. *IEEE/ACM Trans. Netw.*, 1(3):344–357, Jun 1993.

[15] A. Parekh and R. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case. *IEEE/ACM Trans. Netw.*, 2(2):137–150, Apr 1994.

[16] J. Qiu and E. D. Knightly. Inter-class resource sharing using statistical service envelopes. Technical report, Rice University, January 1999. An earlier version appeared in INFOCOM '99.

[17] J. Roberts, U. Mocci, and J. Virtamo. *Broadband Network Teletraffic: Performance Evaluation and Design of Broadband Multiservice Networks: Final Report of Action COST 242*. Springer, 1996.

[18] O. Rose. Statistical Properties of MPEG Video Traffic and their Impact on Traffic Modeling in ATM Systems. Technical Report TR 101, University of Würzburg, Würzburg, Germany, Feb 1995. The MPEG coded video traces are available at ftp://ftp-info3.informatik.uni-wuerzburg.de/pub/MPEG/.

[19] V. Sivaraman and F. M. Chiussi. Providing end-to-end statistical delay guarantees with earliest deadline first scheduling and per-hop traffic shaping. In *Proceedings of INFOCOM 2000*, Tel Aviv, Israel, March 2000.

[20] D. Verma, H. Zhang, and D. Ferrari. Guaranteeing delay jitter bounds in packet switching networks. In *Proc. TRICOMM*, pages 35–46, Chapel Hill, NC, Apr 1991.

[21] Y. Xie and T. Yang. Cell discarding policies supporting multiple delay and loss requirements in atm networks. In *Proceedings of GLOBECOM '97*, pages 1075–1080, Phoenix, AZ, September 1997.

[22] O. Yaron and M. Sidi. Generalized processor sharing networks with exponentially bounded burstiness arrivals. *Journal of High Speed Networks*, 3(4):375–387, 1994.

[23] H. Zhang and D. Ferrari. Rate-Controlled Service Disciplines. *Jrnl. High Speed Netw.*, 3(4):389–412, 1994.

[24] Z.-L. Zhang, D. Towsley, and J. Kurose. Statistical analysis of the generalized processor sharing scheduling discipline. *IEEE Journal on Selected Areas in Communications*, 13(6):1071–1080, August 1995.