In Proceedings of IEEE INFOCOM 2001

Measuring Service in Multi-Class Networks

Aleksandar Kuzmanovic and Edward W. Knightly
Department of Electrical and Computer Engineering
Rice University

Abstract— **Quality of Service mechanisms and differentiated service classes are increasingly available in networks and servers. While network clients can assess their service by measuring basic performance parameters such as packet loss and delay, such measurements do not expose the network's core QoS functionality. In this paper, we develop a framework and methodology for enabling network clients to assess a system's multi-class mechanisms and parameters. Using hypothesis testing, maximum likelihood estimation, and empirical arrival and service rates measured across multiple time scales, we devise techniques for clients to (1) determine the most likely service discipline among EDF, WFQ, and SP, (2) estimate the server's parameters with high confidence, and (3) detect and parameterize non-work-conserving elements such as rate limiters. We describe the important role of time scales in such a framework and identify the conditions necessary for obtaining accurate and high confidence inferences.**

I. Introduction

Both research and commercial networks are increasingly able to provide minimum quality-of-service levels to traffic classes, e.g., [20]. Example components of such networks include QoS schedulers [10], [18], diffserv-style service level agreements [4], [8], [14], [21], [25], edge-based traffic shaping and prioritizing devices, and novel architectures and algorithms for scalable QoS management [6], [7], [18], [19], [25].

However, even as the network's infrastructure and services become increasingly sophisticated, the network's *clients* lack reciprocal tools for validation and monitoring of the network's QoS capabilities. Clients of Service Level Agreements (SLAs) will have monitoring requirements ranging from basic validation of the SLA's raw bandwidth to more sophisticated inference of multi-class functionalities. For example, is a class rate limited (policed)? If so, what are the rate limiter's parameters and what is necessary to detect this? In a multi-class environment with multiple classes within or among SLAs, what is the inter-class relationship? Fair, weighted fair, strict priority, and with what parameters? Is resource "borrowing" across classes fully allowed or only allowed within certain limits?

Obtaining "off-line" answers to such questions can be quite trivial. In particular, consider a system with an unknown service (suppose the system is a single router for simplicity). To assess whether classes are rate limited, one could probe each class, one at a time, with a high rate test sequence: the output of the system would yield the policing parameters. Similarly, simultaneously probing at a high rate in all classes would yield the inter-class relationships: if one class receives all of the service, the system is strict priority (at least for that class); if weighted service is received, the system performs a variant of weighted fair queueing.

This research was supported by NSF CAREER Award ANI-9733610, NSF grants ANI-9730104 and ANI-0085842, and Texas Instruments. The authors may be reached via http://www.ece.rice.edu/networks.

In contrast, the "on-line" case, in which one cannot force all other traffic classes to remain idle while experiments are performed, is quite different. Even for classes which are under the control of the client, it may be highly undesirable to disrupt the class with experiments such as above. For example, sending at a high rate to detect rate-limiters may cause excessive packet losses for established sessions.

The goal of this paper is to develop a framework for monitoring, validation, and inference of multi-class services for the on-line case in which existing services cannot be disrupted. In particular, we show how passive monitoring of system arrivals and departures can be used to detect if a class has a minimum guaranteed rate and/or a rate limiter. Moreover, if such elements exist, we will show how to compute their maximum likelihood parameters. Beyond a single class, we will also show how inter-class relationships can be assessed. For example, we devise tests which infer not only whether a service discipline is work-conserving or non-work-conserving, but also the relationship among classes, such as weighted fair or strict priority.

Throughout our analysis, it is clear that time scales play a key role. Short time scale measurements are crucial for detecting and analyzing non-work-conserving elements such as rate limiters. In contrast, long time scale measurements best reveal "link sharing" rules and weights. Thus, a key aspect of our contribution is that we develop all such measurement tools using a unifying abstraction of envelopes [5], [9], [16], hypothesis testing, and maximum likelihood estimations. In this way, we treat phenomena occurring at different time scales in a uniform and methodical way.

In addition to network services, our techniques also have applications to other multi-class systems such as quality-of-service web servers [1], [3], [11], [13]. For example, the framework can be applied to allow a client of a web hosting service to infer the mechanisms and parameters by which capacity is allocated to various hosted sights. We therefore consider a simple system model which is sufficiently general to encompass a broad class of multi-service elements ranging from routers to servers, yet we necessarily forgo modeling of many of the intricacies of realistic systems (e.g., we limit our discussion to a single network node).

Thus, our contribution is to develop a basic framework for using passive monitoring to assess a system's core multi-class mechanisms and parameters. Despite the simplified system model, a large set of simulation experiments indicate that the technique has practical implications. For example, in our experiments with the majority-rule hypothesis test performed across multiple time scales, multi-class EDF scheduling was correctly

inferred 100% of the time when the class delay bounds were sufficiently differentiated, and class-based fair queueing was correctly inferred 94% of the time. Once the service discipline is known, the algorithm estimated class WFQ weights within 1.4% of the correct value with 95% confidence.

The remainder of this paper is organized as follows. In Section II we describe the basic system model, define the measurement and inference problem, and describe the measurement methodology. In Section III, we devise the maximum likelihood estimates for the system parameters and hypothesis tests for inference of the service discipline. Next, in Section IV, we present a set of *ns-2* simulations to evaluate the effectiveness of the scheme under a number of different node functionalities. Finally, in Section V, we conclude.

II. Service Measurements

A. Scenario

Figures 1 and 2 depict our two targeted systems. (In both cases, passive measurement modules are depicted by diamonds.) Figure 1 illustrates a distributed web server. QoS functionalities in the server may include prioritized scheduling of incoming requests at the front-end, prioritized distribution of jobs to back-end nodes, and operating-system mechanisms such as prioritized scheduling of CPU, memory, and disk access [11]. In any case, our goal is to provide an application-layer characterization of the system's multi-class QoS mechanisms. For example, if several QoS mechanisms are simultaneously employed with the goal of providing weighted fair service among different classes, our technique will estimate a class' net "guaranteed rate", i.e., it's minimum request throughput. Such inferences have important implications for both performance monitoring and resource management.

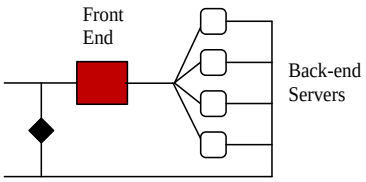

Fig. 1. Web Server

Figure 2 depicts the targeted networking scenario. In this case, measurement modules are placed at the periphery of the network. The goal is to use passive edge-based client measurements to infer the multi-class QoS mechanisms and parameters employed by the network operator. With an improved understanding of the way traffic is internally serviced, clients can better manage their use of multi-class networks. Similarly, operators or third parties can employ the methodology to test and validate the performance and potential performance of multiple service classes.

Below, we describe the system model and problem formulation for multi-class service inference. We then devise a measurement methodology based on empirical arrival and service

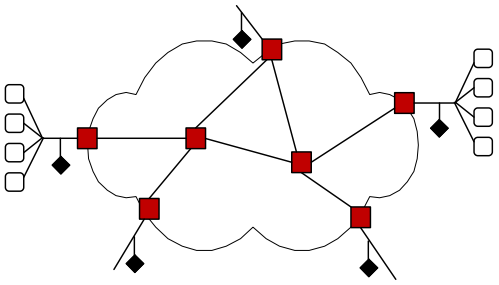

Fig. 2. Network

envelopes which we employ in the inference and hypothesis tests of Section III.

B. System Model and Problem Formulation

The general system model that we consider in this paper is depicted in Figure 3. As in the basic abstraction of service disciplines described in [24], it consists of two stages: non-work-conserving elements which limit a class' rate and a work-conserving packet scheduler. For rate limiters, we consider single-level leaky bucket regulators. For the packet scheduler, we consider Strict Priority (SP), class-based Weighted Fair Queueing (WFQ), and Earliest Deadline First (EDF). In this context, an SP scheduler consists of one queue per traffic class with packets from the highest priority non-empty class serviced first. For example, a packet in level k is serviced only if no packets are backlogged in levels $1, \cdots, k-1$. For WFQ [15] each traffic class k is allocated a guaranteed capacity C_k such that whenever packets from class k are backlogged, the class receives service at a rate of at least C_k. Unused capacity of non-backlogged classes is distributed in a weighted fair manner among backlogged classes. For EDF, each class has an associated delay bound so that packet j of class k arriving at time a_j^k has deadline a_j^k plus its delay bound, and the scheduler selects the packet with the smallest (earliest) deadline for service.

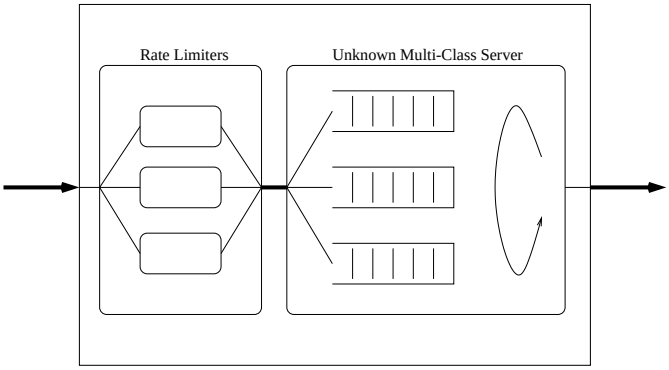

Fig. 3. System Model for Multi-Service Measurement

This formulation covers a broad set of class-based scheduling elements, including minimum guaranteed rates, maximum policed rates, weighted fairness, sorted priority, and strict priority. While necessarily not comprehensive, it incorporates both

work-conserving and non-work-conserving service disciplines and a number of mechanisms for inter-class resource sharing and quality-of-service differentiation.

For inferences of the system's multi-class characteristics, we consider the case where internal system information is *not* available, i.e., neither static configuration information (such as the scheduler's parameters) nor empirical information (such as mean buffer length). Instead, the available information consists of the external observations from passive monitoring of packets, namely packet arrival and departure times along with packet class labels and sequence numbers.[1]

The goal is to provide a framework for inference of the mechanisms and parameters of the key elements in the multi-class system, utilizing only the aforementioned external measurements. In particular, we develop a hypothesis test to identify the basic scheduling algorithm as SP, EDF, or WFQ. We devise techniques to estimate the maximum-likelihood values of the class parameters, such as "guaranteed rate" in WFQ or deadlines in EDF and rate limiters in non-work-conserving servers.

C. Empirical Arrival Model

Here we describe a general arrival characterization which can be applied to the multi-class inference problem. The technique is based on traffic envelopes which provide a unifying abstraction for both arrivals and services and incorporate the system's behavior across time scales [16]. Measurement at multiple time scales is important in this context as different system components are most accurately detected at different time scales.

Focusing on a single class for illustration, Figure 4 depicts an example arrival-departure sequence, with the j^{th} packet having size p_j, arrival time a_j and departure time d_j.

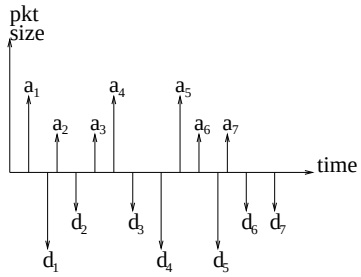

Fig. 4. Example of a Class' Arrival-Departure Sequence

Denoting the total arrivals in the interval $[s, s+t]$ by $A[s, s+t]$, a traffic envelope refers to a time invariant characterization of the arrivals as a function of interval length t (see [22] for examples of deterministic envelopes). For a measurement window $[s, s+T]$ and a particular interval length I_k beginning at time $s+(j-1)I_k$, class i's arrival *rate* is given by

$$R_{k,j}^{i,A} = \frac{A^i[s+(j-1)I_k,\ s+jI_k]}{I_k}$$

for $j = 1, \cdots, N_k$, where $N_k = \lfloor T/I_k \rfloor$ is the number of successive intervals of length I_k in the measurement window $[s, s+T]$.

Using measured rates over different sub-intervals within the window T, the mean and variance of the empirical rate *envelope* of class i for intervals of length k can be computed as

$$\bar{R}_k^{i,A} = \frac{1}{N_k} \sum_{j=1}^{N_k} R_{k,j}^{i,A} \tag{1}$$

and

$$RV_k^{i,A} = \frac{1}{N_k} \sum_{j=1}^{N_k} (R_{k,j}^{i,A} - \bar{R}_k^{i,A})^2 \tag{2}$$

In Section III, we describe how this empirical class-based arrival envelope is incorporated into the above multi-class inference problems, and in Section IV, we experimentally investigate applications of this traffic characterization.

D. Empirical Service Model

Here, we describe a general mechanism for measuring and characterizing a class' service rate. Analogous to the traffic envelope, the service envelope is not simply a single service bandwidth, but a statistical characterization of service across time scales. This multiple-time-scale characterization is critical to inference of diverse service components such as maximum policed bandwidth, minimum service, and analysis of inter-class resource sharing relationships. Moreover, its statistical nature reflects the fact that a class' service can fluctuate according to the varying demands of other classes and the mechanism by which the scheduler arbitrates this demand.

The empirical service envelope characterizes the service rate received by the flow as a function of the interval length over which the class is backlogged, where a flow is said to be backlogged whenever it has at least one packet in the system. A traffic flow is continuously backlogged for k packet transmissions in the interval $[a_j, d_{j+k-1}]$ if

$$d_{j+m} > a_{j+m+1}, \text{ for all } 0 \le m < k-2,$$

for $k \ge 2$. Note that all packet transmissions are backlogged for $k = 1$ in the interval $[a_j, d_j]$.

To illustrate the backlogging condition, consider the example of Figure 4 which depicts an arrival and departure sequence for packets belonging to a particular class. The second packet arrives into the system after the first packet departs. Hence, for the first packet, the backlogging condition is satisfied only for $k = 1$; likewise for the second packet. In contrast, for the third packet, the flow is also backlogged for $k = 2$ consecutive packets as the fourth packet arrives before the service of the third packet. Similarly, a sequence of $k = 3$ packets are backlogged

[1] In the case of web servers (both single node and distributed), both arrivals and departures are directly observable from the system's front end (see [1], [11] for a detailed description of such an architecture). In the case of networks, packet time stamping provides a mechanism to observe both arrival and departure times at the departure node [17]; otherwise, the measurement modules can communicate their collected information off-line.

beginning with the arrival of packet 5 and ending with the departure of packet 7. In other words, the interval $[a_5, d_7]$ is a backlogging interval for $k = 3$. Notice that the sub-intervals $[a_5, d_6]$ and $[a_6, d_7]$ are also backlogged for $k = 2$ packets.

Note that the measured class must be backlogged in order to infer its service rate. However, the measured class does not require *other* classes to be backlogged when monitoring its service, as this information is indirectly revealed by fluctuations in its own measurements.

Thus, denoting $U^i[s, s+t]$ as number of class-i bits serviced in the backlogged interval $[s, s+t]$, the service *rate* received in $[s, s+t]$ is simply

$$R^{i,S}(t) = \frac{U^i[s, s+t]}{t}. \quad (3)$$

Finally, the measurement for each backlogged interval is included in the measurement $\vec{R}_k^{i,S}$ if

$$(k - 0.5)I_1 < t \leq (k + 0.5)I_1. \quad (4)$$

Measured service envelope samples $\vec{R}_k^{i,S}$ are used in inferring scheduling discipline, as explained in detail in the following Section.[2]

III. Multi-Class Inference

As described above, our goal is to infer the elements and parameters of a multi-class system. In Part A we consider the server and in Part B the rate limiters as defined in Figure 3.

A. Inference of the Scheduler and Inter-Class Relationships

In a multi-class system, the packet service discipline defines the inter-class relationships or the service received when different classes compete for resources. For example, with an SP scheduler, the highest priority class receives all demanded service up to the available link capacity, and in that way is completely isolated from other classes' demands. In contrast, lower priority classes utilize only *remaining* capacity from higher priority classes and their performance is strongly dependent on these classes' demands.

Here, we provide a precise theoretical description of such inter-class relationships via statistical service envelopes. Under a particular scheduler hypothesis, we perform Maximum Likelihood Estimations (MLEs) of the scheduler's parameters, such as guaranteed rates in WFQ and deadlines in EDF. Using the envelope's ideal description of a class' service, we then develop hypothesis tests to infer which service discipline is employed by the system via statistical analysis of the empirical inter-class sharing relationships. Finally, we select the MLEs of the unknown parameters under the inferred scheduler.

[2]Notice that for convenience, the arrival envelope is discretized in time and the service envelope is discretized in bits. However, to perform the comparative computations of Section III, both are expressed in discrete time rates with service interpolated.

A.1 Theoretical Service Envelopes

In [16], statistical admission control tests are developed for several multi-class schedulers. The key technique for exploiting inter-class resource sharing is to characterize a class' available service beyond its worst-case allocation. For example, in a WFQ server a class with weight ϕ_i receives service at rate no less than $\frac{\phi_i}{\sum_j \phi_j} C$ whenever it is backlogged. However, due to statistically varying demands of other classes, the service received can be far greater than this lower bound. A statistical service envelope $S^i(t)$ is therefore a general characterization of the service received by class i over intervals of length t for which the class is continually backlogged.

Table I shows the statistical service envelopes (derived in [16]) for SP, WFQ, and EDF. The envelopes are a function of the link capacity C, and as described above, the other class' input traffic, described by the arrival envelope $B^i(t) \sim A^i[s, s+t]$. For SP, observe that class i's service is only a function of the workload in classes $1, 2, \cdots, i-1$. In contrast, for WFQ class i's service is a function of all other classes' traffic but is upper bounded by Ct if all other classes are always idle and lower bounded by $\frac{\phi_i}{\sum_j \phi_j} C$ if all other classes are continuously backlogged. Finally, with EDF class i's service again depends on all other class' inputs as well as the delay bound of class i denoted by d_i.[3]

Observe that the *rate* traffic envelope defined in Section II (e.g., $\bar{R}_k^{i,A}$ and $RV_k^{i,A}$) are simply empirical and normalized versions of the first two moments of $B^i(I_k)$. On the other hand, $\vec{R}_k^{i,S}$ contains normalized (on intervals of length I_k) samples of the service envelope $S^i(I_k)$. The key problem is then to assess the system's scheduler, parameters, and other components by comparing empirical measurements with the idealized relationships.

A.2 Empirical Service Distributions

Here, we describe the expected distributions of service for a given arrival distribution under different service disciplines. For simplicity, we consider a two-class system and aggregate traffic $A^i[s, s+t]$ with a Gaussian distribution.[4] Notice that even under Gaussian arrivals, the service envelopes will be non-Gaussian due to the non-linearities of the multi-class server. With two traffic classes, SP is a special case of WFQ with $\phi_i = 0$ or 1. Thus, we must first infer whether the scheduler is EDF or WFQ, and if the latter, whether the weights are 0 and 1, which would indicate SP.

Denote X_k as a Gaussian random variable with mean $CI_k - \bar{R}_k^{n,A} I_k$, variance $RV_k^{n,A} {I_k}^2$ and probability density function

[3]With abuse of notation, we henceforth denote d_i as the class-i delay bound, as opposed to a departure time as in the previous section.

[4]The Gaussian assumption is not necessary for traffic envelopes; see [5] for example. Regardless, we make the assumption in this paper as it makes our solution more computationally efficient while also retaining a high degree of accuracy, both for our purposes here as well as in other scenarios such as admission control [12].

Scheduler	Service Envelope $S^i(t+\tau_i)$
SP	$\left(C(t+\tau_i) - \sum_{n=1}^{i-1} B^n(t+\tau_i)\right)^+$
WFQ	$\phi_i C(t+\tau_i) + \left((1-\phi_i)C(t+\tau_i) - \sum_{n\neq i} B^n(t+\tau_i)\right)^+$
EDF	$\left(C(t+\tau_i) - \sum_{n\neq i} B^n(t-d_n+d_i)\right)^+$

TABLE I
STATISTICAL SERVICE ENVELOPES FOR DIFFERENT SCHEDULERS

$p_{X_k}(x)$, i.e.,[5]

$$X_k \sim N\left(CI_k - \bar{R}_k^{n,A} I_k, RV_k^{n,A} {I_k}^2\right).$$

From Table I, the probability density function of the service envelope $S_k^i = S^i(I_k)$ under the hypothesis that the server is WFQ is given by

$$\begin{aligned} p_{S_k^i}^{WFQ}(x) &= P(X_k \leq \phi_i C I_k)\delta(x - \phi_i C I_k) \\ &+ p_{X_k}(x) I\left(\phi_i C I_k \leq x \leq C I_k\right) \\ &+ P(X_k \geq C I_k)\delta(x - C I_k) \end{aligned} \quad (5)$$

where $I(\cdot)$ is an indicator function and $\delta(\cdot)$ is a delta function.

Similarly, define the random variable Y_k such that

$$Y_k \sim N\left(CI_k - \bar{R}_l^{n,A} I_l, RV_l^{n,A} I_l^2\right).$$

Further denote the probability density function of Y_k by $p_{Y_k}(y)$, where $l = k - \lfloor \bar{D}_i - d_n + d_i \rfloor$ if $\lfloor \bar{D}_i - d_n + d_i \rfloor > 0$ and $l = k$ otherwise, and $\bar{D}_i$ is empirical *mean* delay. From the EDF service envelope of Table I, we have that the probability density function of S_k^i under the EDF hypothesis is given by

$$\begin{aligned} p_{S_k^i}^{EDF}(y) &= P(Y_k \leq 0)\delta(y) + \\ & p_{Y_k}(y) I(0 \leq y \leq C I_k) + P(Y_k \geq C I_k)\delta(y - C I_k). \end{aligned}$$

Examples of empirical class service rate distributions for WFQ and SP servers are presented in Figures 5 and 6. The interval length I_k is 400 msec and additional parameters such as traffic load and statistical workload characterization are given in Section IV.

We make several observations about the figures. First, the service distribution of WFQ visibly exhibits the truncated behavior defined by Equation (5): this is due to WFQ's guaranteed rate which lower bounds the service. Second, observe that no such "hard" lower border exists for SP without strict rate limiters on all higher priority traffic classes. Finally, notice that upper limits on the density functions are not evident here, as in this case, neither class reached its upper limits due to statistical fluctuations in the demand of the other class.

[5] n indexes the non-measured class.

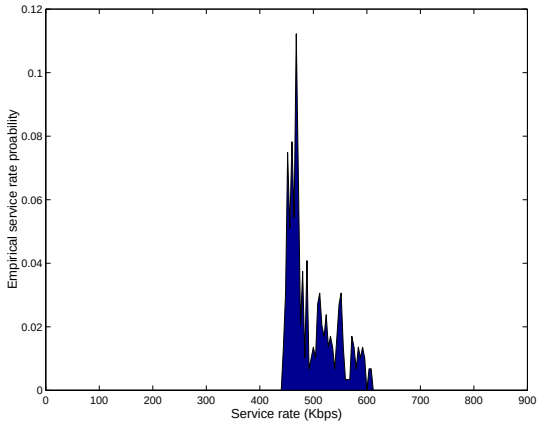

Fig. 5. WFQ Service Rate Histogram

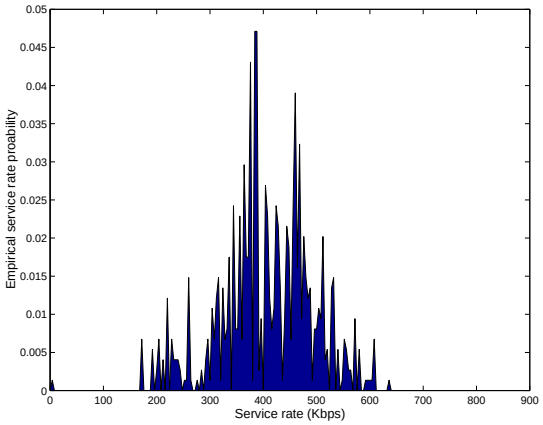

Fig. 6. SP Service Rate Histogram

A.3 Parameter Estimation Under Scheduler Hypothesis

Here, we describe how a scheduler's parameters such as weights in WFQ and deadlines in EDF can be estimated under the hypothesis of a particular scheduler EDF, WFQ, or SP. We employ the Generalized Likelihood Ratio Test by first obtaining Maximum Likelihood Estimates of unknown parameters under each hypothesis, and then using the likelihood ratio test. We then show how the scheduling mechanism itself can be inferred by choosing the more likely hypothesis as the true one. Finally, the MLEs of unknown parameters under the chosen hypothesis become the final estimates.

The first problem is to determine class 1's unknown weight parameter under the hypothesis that the server is WFQ. Given the observations of each class' service in intervals of length I_k , we use the MLE to estimate the unknown parameter ϕ_i as

$$\hat{\phi}_{1,k} = \underset{\phi_{1,k}}{\text{argmax}}\, p^{WFQ}(\vec{R}_k^{1,S} I_k, \vec{R}_k^{2,S} I_k | \phi_{1,k}) \quad (6)$$

where

$$p^{WFQ}(\vec{R}_k^{1,S} I_k, \vec{R}_k^{2,S} I_k | \phi_{1,k}) = \prod_{m=1}^{M} p_{S_k^1}^{WFQ}(x = \vec{R}_{k,m}^{1,S} I_k) \prod_{n=1}^{N} p_{S_k^2}^{WFQ}(y = \vec{R}_{k,n}^{2,S} I_k)$$

and M and N denote the respective sizes of $\vec{R}_k^{1,S}$ and $\vec{R}_k^{2,S}$. Since a closed form expression cannot be found for the MLE in Equation (6), we employ a numerical grid search by maximizing the likelihood function with respect to the unknown parameter ϕ_1 in the interval $[0, 1]$. (Notice that the unknown values have known and closed borders so that the grid numerical search is justified.) The estimate is obtained for each interval I_k independently, and the final estimate of $\hat{\phi}_1$ is is computed by averaging the estimates for different time scales. Finally, in the two-class case, $\hat{\phi}_2$ is simply $1 - \hat{\phi}_1$.

The physical interpretation of Equation (6) is as follows. The relative class weight estimation can be performed only over time intervals when both classes are backlogged since it is only during such intervals that both classes incur their lower bounds in service. Such intervals cause peaks at the lower clipping of the service rate distribution (cf. Figure 5) and also maximize the joint distribution of Equation (6).

For EDF, similar expressions can be derived by applying the same methodology of using the EDF service envelopes to compute the MLE expressions for the class delay bounds, and performing a grid search to estimate $\hat{d}_1$ and $\hat{d}_2$.

A.4 Scheduler Inference

The above technique allows estimation of a scheduler's parameters under the hypothesis of particular scheduler. Here, we show how the scheduling policy itself can be inferred. The key technique is to choose the hypothesis that makes the measured service observation most likely.

To infer which service discipline is the most likely under the observations, for each time scale I_k, we have the scheduler hypothesis test given by

$$\frac{p_k^{EDF}(\vec{R}_k^{1,S} I_k) p_k^{EDF}(\vec{R}_k^{2,S} I_k)}{p_k^{WFQ}(\vec{R}_k^{2,S} I_k) p_k^{WFQ}(\vec{R}_k^{2,S} I_k)} \underset{WFQ}{\overset{EDF}{\gtrless}} 1. \quad (7)$$

As all time scales are not guaranteed to infer the same scheduler, the final decision is made by using majority rule over all time scales. With this technique, the correct inference is attained 94% to 100% of the time in our experiments.

Thus, the service envelopes of Table I describe particular rules for inter-class resource sharing, i.e., the mechanism by which the scheduler allocates capacity when multiple traffic classes are competing for resources. Applying the above hypothesis test determines which scenario is most likely, considering multi-class measurements across all time scales.

B. Rate Limited Class State Inference

Thus far, we have considered work-conserving service disciplines. Here, we develop a measurement methodology applicable to rate-limiters, i.e., non-work-conserving elements which limit a flow's arrivals to within a pre-specified constraint. For a single token bucket with a bucket depth of one packet, the rate limiter for class i is characterized by an unknown rate r^i. The key problem is to distinguish such a limit on class i's service from throughput limits due to the workloads of other traffic classes and other mechanisms in the multi-class scheduler.

Thus, the goal is to find the maximum likelihood estimation of r_i under the hypothesis of a particular scheduler (inferred as above). With rate limiters, the service envelopes of Table I have r^i in place of C as the maximum service rate. Thus considering the EDF hypothesis as an example, the maximum likelihood estimation of r^i can be computed as

$$(\hat{r}_k^i, \hat{d}_1, \hat{d}_2) = \underset{r_k^i, d_1, d_2}{\text{argmax}}\, p^{EDF}(\vec{R}_k^{1,S} I_k, \vec{R}_k^{2,S} I_k | r_k^i, d_1, d_2). \quad (8)$$

Estimation of rate limiter parameters highlights the importance of time scales. This is illustrated in Figure 7 which depicts the probability that a class transmits at the rate limiter's bound as a function of interval length. The scenario is a two-class class-based fair queueing scheduler with class weights of 0.5. The classes have 60 and 40 exponential on-off flows with peak rate 32 kb/sec. The figure shows the empirical probability that the aggregate traffic of class 1 transmits at its rate limit of 1 Mb/sec as a function of interval length. As shown, for short time scales this occurs quite frequently whereas it is increasingly rare over longer time scales. While this property is an inherent characteristic of any variable rate flow, the key point is that inference of rate limiter parameters at long time scales is inhibited by flows becoming less and less likely to send at peak rates for sustained periods. As a consequence, measurement of multi-level leaky buckets, which *require* longer time scale measurements due to traffic constraint functions which shape the traffic differently at different time scales (see [23] for example) will incur higher measurement errors.

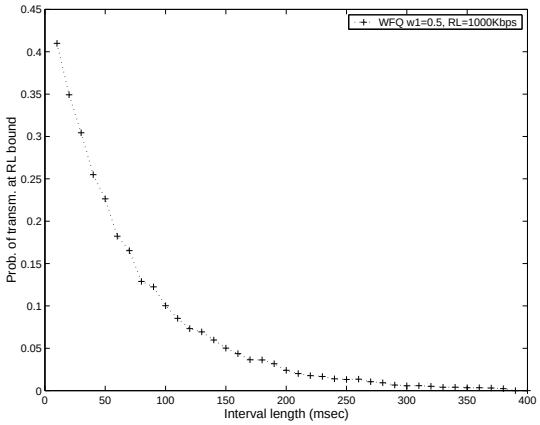

Fig. 7. Probability of Transmitting at Rate Limiter Bound

IV. Experimental Investigations

In this section, we perform a set of simulation experiments to evaluate the effectiveness of the multi-class inference tools described above. We study WFQ weight estimation, inference of the service discipline for EDF, SP, and WFQ , as well as "measurable regions", the conditions necessary to obtain accurate estimates of WFQ weights.

All simulations are performed with the *ns-2* simulator with a single router and various numbers of hosts in the topology of Figure 3. The link capacity is 1.5 Mb/sec and packet sizes are 500 and 100 bytes, as specified in the various experiments. The minimum interval length for measuring arrival and service envelopes is I_1 = 10 msec and the maximum interval-length for measurement is 0.5 sec for a 50-point arrival envelope. The measurement window T is varied in the experiments from 2 to 10 seconds as indicated. We consider two traffic classes and EDF, WFQ, and SP scheduling.

A. WFQ Weight Estimation

Here we experimentally investigate the statistical properties of the WFQ weight estimation algorithm. In this scenario, the system has 65-68 flows exponential on-off sources with on-rate 32 kb/sec with on and off periods of 0.36 sec. Moreover, there are 25-28 sources of the same type for class 2. The true WFQ weights are ϕ_1 = 0.7 and ϕ_2 = 0.3. The packet size is 500 Bytes.

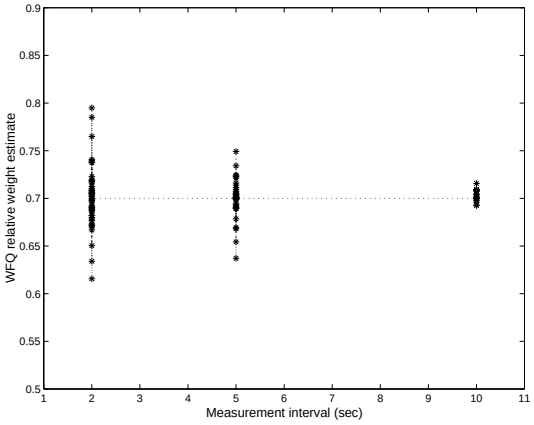

Fig. 8. WFQ Weight Estimation vs. Measurement Interval

In the experiments, 50 simulation runs were performed corresponding to each data point in Figure 8. For a particular simulation, the measurement window T is set to 2, 5, or 10 seconds as reported on the horizontal axis. Each point on the plot indicates the maximum likelihood estimation of ϕ_1, $\hat{\phi}_1$, using the methodology of Section III.

First observe that the variance of the estimator is reduced with increasing T, due simply to the fact that more sample points are available with larger T. For example, with T = 2 sec, 95% of the weight estimations are within 11% of the true value, whereas with T = 10 sec, 95% of the weight estimations are within 1.4% of the true value. However, T should not be set arbitrarily large, as longer-time-scale fluctuations due to flow arrivals and departures may introduce non-stationarities which would bias the tests. While the number of flows in the system did vary in these simulations, as defined above it is within a range of 5 to 10% of the system load.

B. Scheduler Inference

As described in Section III, the above WFQ weight estimations can only be performed under the hypothesis that the server is WFQ. Thus, statistical tests are necessary to infer the scheduling mechanism itself. Here we describe simulation experiments for scheduler inference using the same number of sources for each class and the same packet size as in the previous experiment. Figures 9 and 10 depict the experimental probability of correct decision vs. time scale for the respective correct hypothesis of EDF and WFQ respectively. In both cases, 50 simulations are performed and the probability of correct decision is computed as the number of correct decisions versus total number of tests for each time scale (recall the final decision is performed by majority rule).

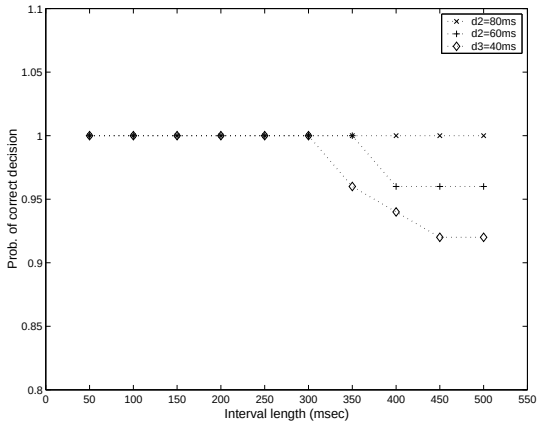

Fig. 9. EDF: Probability of Correct Decision vs. Time Scale

For the experiments of Figure 9, the correct hypothesis is EDF with delay bounds of d_1 = 20 msec for class 1 and d_2 = 40, 60 and 80 msec for the three curves for class 2. As indicated in the figure, EDF is correctly inferred 100% of the time at short time scales (I_k up to 300 msec) while less frequently for longer interval lengths, especially as $d_2 - d_1$ decreases. Yet in all cases the probability of correct decision is no less than 92%. The reason that the probability of correct decision decreases as $d_2 - d_1$ decreases, is that there is less and less differentiation provided by the scheduler, making the service envelopes statistically closer, and the inference problem more difficult. Indeed, if $d_2 = d_1$, the scheduler is actually performing FCFS, as is also evident from the service envelope in Table I.

Regardless, in all cases, the correct *final* decision is made as majority rule is performed over different time scales, and incorrect decisions at a particular time scale are never frequent enough to form a majority.

Figure 10 depicts the experimental results for WFQ. Observe that in this case, the correctness ratio is quite poor on shorter

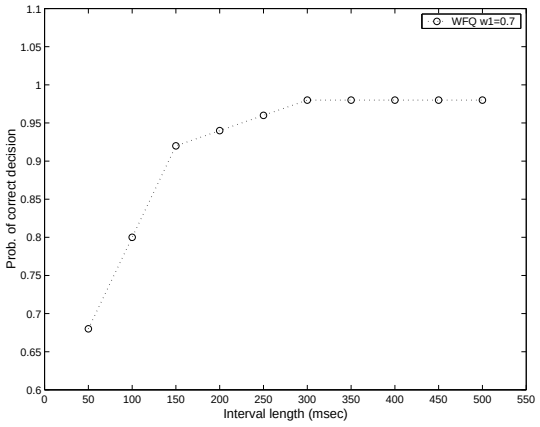

Fig. 10. WFQ: Probability of Correct Decision vs. Time Scale

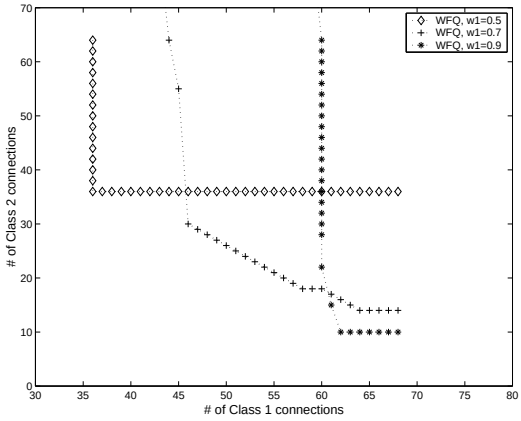

Fig. 11. Measurable Region for Lower Service Bounds

time scales. This is due to the mismatch between the fluid approximation used in the analytical model and the packet-level simulations. In particular, over short time intervals, the fluid approximation does not hold and not every packet gets serviced at rate $\phi_i C$ (indeed, see [2] for a detailed discussion of such short-time-scale unfairness). In this case, such errors impact the final decision and the overall correctness probability is 0.94 (less than the correctness of 1 achieved in the EDF case) as the short-time-scale errors form a majority in 6% of the cases.

Finally notice that the relationship of the probability of correct decision and time scale are reversed for WFQ as compared to EDF. The reason for this is that over longer time scales, WFQ overcomes packet level unfairness and, when flows are backlogged for long durations, it can become quite clear (statistically) that there is a minimum guaranteed service rate clipping the distribution of the service envelope. In contrast, for EDF, the differences are most pronounced for small interval lengths where the shifts in the arrival envelopes (cf. Table I) are more prominent.

C. Measurable Region

The methodology presented in this paper is based on *passive* measurements, i.e., no probing packets are transmitted to modify the system's workload. However, with passive monitoring, it is possible that other classes' particular workloads *prohibit* inference of certain network elements. For example, in the extreme case that all other classes are idle, it is impossible to detect a guaranteed minimum rate. Similarly, the multi-class nature of the scheduler itself would not be measurable, and only rate-limiter parameters could be obtained. We refer to the required workload to measure a particular network behavior as the *measurable region*.

Here, we address the issue of the conditions necessary to infer lower and upper service limits for WFQ schedulers. For the simulations, each flow has on and off periods of 0.36 sec and on-rate 32 kb/sec. The packet size is 100 bytes and C = 1.5 Mb/sec.

Figure 11 depicts the resulting measurable regions for WFQ weight estimates. Each point represents the minimum number of class 1 and class 2 flows needed such that the relative weights can be estimated within 5% of their correct value. That is, if either class has fewer flows then indicated by this measurable region, then estimation is not possible, as the conditions required for weight estimation occur too rarely.

Observe that as the weight of class 1 increases from ϕ_1 of 0.5 to 0.7 and 0.9 (corresponding to the three curves), the curves shift to the lower right indicating that a higher number of class 1 flows and lower number of class 2 flows are needed to infer ϕ_1. The reason for this is that as ϕ_1 becomes larger, a higher traffic load in class 1 is required to backlog class 1 sufficiently to estimate the guaranteed rate.

Finally, observe that a typical point on the curve refers to a relatively modest resource utilization. For example, under ϕ_1 = 0.7, at least 30 class 2 flows are required when 46 class 1 flows are present. This corresponds to an average system utilization of 62%, i.e., the mean utilization must be at least 62% to perform the measurements passively, otherwise active probing is required.

V. Conclusions

Networks and servers are increasingly providing quality-of-service functionalities. The goal of this paper is to provide a framework for clients of multi-class services to assess a system's core QoS mechanisms. We developed a scheme for clients to perform a series of hypothesis tests across multiple time scales in order to infer the packet service discipline among class-based weighted fair queueing, earliest deadline first, and strict priority. For a particular scheduler, we devised techniques for clients to obtain maximum likelihood estimations of the system's class differentiation parameters, such as WFQ weights and EDF delay bounds. Finally, we showed how parameters of non-work-conserving elements such as rate limiters can be estimated. Throughout, we utilized a general multiple-time-scale traffic and service model to characterize a broad set of behaviors within a unified framework.

In future work, we plan to design mechanisms to coordinate the end-point measurements of the modules depicted in Figure 2 and generalize the system model of Figure 3 to consider multiple

nodes and cross traffic.

REFERENCES

[1] M. Aron, P. Druschel, and W. Zwaenepoel. Cluster reserves: A mechanism for resource management in cluster-based network servers. In *Proceedings of ACM SIGMETRICS 2000*, June 2000.

[2] J. Bennett and H. Zhang. WF^2Q: Worst-case Fair Weighted Fair Queueing. In *Proceedings of IEEE INFOCOM '96*, San Francisco, CA, March 1996.

[3] N. Bhatti and R. Friedrich. Web server support for tiered services. *IEEE Network*, 13(5):64–71, September 1999.

[4] S. Blake et al. An architecture for differentiated services, 1998. Internet RFC 2475.

[5] R. Boorstyn, A. Burchard, J. Liebeherr, and C. Oottamakorn. Effective envelopes: Statistical bounds on multiplexed traffic in packet networks. In *Proceedings of IEEE INFOCOM 2000*, Tel Aviv, Israel, March 2000.

[6] L. Breslau, E. Knightly, S. Shenker, I. Stoica, and H. Zhang. Endpoint admission control: Architectural issues and performance. In *Proceedings of ACM SIGCOMM 2000*, Stockholm, Sweden, August 2000.

[7] C. Cetinkaya and E. Knightly. Scalable services via egress admission control. In *Proceedings of IEEE INFOCOM 2000*, Tel Aviv, Israel, March 2000.

[8] C. Chuah, L. Subramanian, R. Katz, and A. Joseph. QoS provisioning using a clearing house architecture. In *Proceedings of IWQoS 2000*, Pittsburgh, PA, June 2000.

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

[10] C. Dovrolis and P. Ramanathan. A case for relative differentiated services and the proportional differentiation model. *IEEE Network*, 13(5):26–35, September 1999.

[11] V. Kanodia and E. Knightly. Multi-class latency-bounded web services. In *IEEE/IFIP IWQoS 2000*, Pittsburgh, PA, June 2000.

[12] E. Knightly and N. Shroff. Admission control for statistical QoS: Theory and practice. *IEEE Network*, 13(2):20–29, March 1999.

[13] K. Li and S. Jamin. A measurement-based admission controlled web server. In *Proceedings of IEEE INFOCOM 2000*, Tel Aviv, Israel, March 2000.

[14] K. Nichols, V. Jacobson, and L. Zhang. Two-bit differentiated services architecture for the Internet, 1999. Internet RFC 2638.

[15] A. Parekh and R. Gallager. A generalized processor sharing approach to flow control in integrated services networks: the single-node case. *IEEE/ACM Transactions on Networking*, 1(3):344–357, June 1993.

[16] J. Qiu and E. Knightly. Inter-class resource sharing using statistical service envelopes. In *Proceedings of IEEE INFOCOM '99*, New York, NY, March 1999.

[17] J. Schlembach, A. Skoe, P. Yuan, and E. Knightly. Design and implementation of scalable admission control. In *Proceedings of the International Workshop on QoS in Multiservice IP Networks*, Rome, Italy, January 2001.

[18] I. Stoica, S. Shenker, and H. Zhang. Core-Stateless Fair Queueing: A scalable architecture to approximate fair bandwidth allocations in high speed networks. In *Proceedings of ACM SIGCOMM '98*, Vancouver, British Columbia, September 1998.

[19] I. Stoica and H. Zhang. Providing guaranteed services without per flow management. In *Proceedings of ACM SIGCOMM '99*, Cambridge, MA, August 1999.

[20] B. Teitelbaum et al. Internet2 QBone: Building a testbed for differentiated services. *IEEE Network*, 13(5):8–17, September 1999.

[21] A. Terzis, L. Wang, J. Ogawa, and L. Zhang. A two-tier resource management model for the internet. In *Proceedings of Global Internet Symposium '99*, Rio de Janeiro, Brazil, December 1999.

[22] D. Wrege, E. Knightly, H. Zhang, and J. Liebeherr. Deterministic delay bounds for VBR video in packet-switching networks: Fundamental limits and practical tradeoffs. *IEEE/ACM Transactions on Networking*, 4(3):352–362, June 1996.

[23] D. Wrege and J. Liebeherr. Video trafic characterization for multimedia networks with a deterministic service. In *Proceedings of IEEE INFOCOM '96*, pages 537–544, San Francisco, CA, March 1996.

[24] H. Zhang. Service disciplines for guaranteed performance service in packet-switching networks. *Proceedings of the IEEE*, 83(10):1374–1399, October 1995.

[25] Z. Zhang, Z. Duan, L. Gao, and Y. Hou. Decoupling QoS control from core routers: a novel bandwidth broker architecture for scalable support of guaranteed services. In *Proceedings of ACM SIGCOMM 2000*, Stockholm, Sweden, August 2000.