# Performance Models for LAS-based Scheduling Disciplines in a Packet Switched Network *

Idris A. Rai
Institut Eurecom
BP 193
06904 Sophia-Antipolis, France
rai@eurecom.fr

Guillaume Urvoy-Keller
Institut Eurecom
BP 193
06904 Sophia-Antipolis, France
urvoy@eurecom.fr

Mary K. Vernon
Department of Computer Sciences
University of Wisconsin Madison
Wisconsin 53706
vernon@cs.wisc.edu

Ernst W. Biersack
Institut Eurecom
BP 193
06904 Sophia-Antipolis, France
erbi@eurecom.fr

## ABSTRACT

The Least Attained Service (LAS) scheduling policy, when used for scheduling packets over the bottleneck link of an Internet path, can greatly reduce the average flow time for short flows, while not significantly increasing the average flow time for the long flows that share the same bottleneck. No modification of the packet headers is required to implement the simple LAS policy. However, previous work has also shown that a drawback of the LAS scheduler is that, when link utilization is greater than 70%, long flows experience large jitter in their packet transfer times as compared to the conventional First-Come-First-Serve (FCFS) link scheduling. This paper proposes and evaluates new differentiated LAS scheduling policies that reduce the jitter for long flows that are identified as "priority" flows.

To evaluate the new policies, we develop analytic models to estimate average flow transfer time as a function of flow size, and average packet transmission time as a function of position in the flow, for the single-bottleneck "dumbbell topology" used in many ns simulation studies. Models are developed for FCFS scheduling, LAS scheduling, and each of the new differentiated LAS scheduling policies at the bottleneck link. Over a wide range of configurations, the analytic estimates agree very closely with the ns estimates. Thus, the analytic models can be used instead of simulation for comparing the policies with respect to mean flow transfer time (as a function of flow size) and mean packet transfer time. Furthermore, an initial discrepancy between the analytic and simulation estimates revealed errors in the parameter values that are often specified in the widely used ns Web workload generator. We develop an improved Web workload specification, which is used to estimate the packet jitter for long flows (more accurately than with previous simulation workloads).

Results for the scheduling policies show that a particular policy, LAS-log, greatly improves the mean flow transfer time for priority long flows while providing performance similar to LAS for the ordinary short flows. Simulations show that the LAS-log policy also greatly reduces the jitter in packet delivery times for the priority flows.

## Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Performance attributes.; I.6 [**Simulation and Modeling**]: Model Validation and Analysis.

## General Terms

Performance, Design.

## Keywords

Scheduling, FCFS and LAS models, LAS-based scheduling and models, models validation, simulations, service differentiation.

## 1. INTRODUCTION

### 1.1 Background

Previous results for timesharing systems show that, for certain workloads, some scheduling policies offer low mean response times to short jobs without starving large (batch) jobs. One such policy is *Least Attained Service* (LAS), which is also called *Foreground-Background* (FB) [15] or *Shortest Elapsed Time* (SET) first [5]. LAS is a preemptive scheduling policy that gives service to the job in the system that has so far received the least service. In the event of ties, the multiple jobs that have received the least service share the processor in a processor-sharing mode. Thus, LAS favors short jobs *without prior knowledge* of job sizes.

The impact of short jobs on the mean response time of large jobs under LAS highly depends on the job size distribution [20]. In

*Institut Eurécom's research is partially supported by its industrial members: Bouygues Télécom, Fondation d'entreprise Groupe Cegetel, Fondation Hasler, France Télécom , Hitachi, ST Microelectronics, Swisscom, Texas Instruments, Thales.


*SIGMETRICS/Performance'04*, June 12–16, 2004, New York, NY, USA.

particular, for job size distributions with a high coefficient of variation[1], the small mean response times seen by the short jobs comes at a cost of only a very small increase in mean response times for the largest jobs. Internet flow sizes have a large coefficient of variation. Most of the flows are short, while more than half of the bytes are carried by a small percentage of flows that are very large [4].

A network flow is identified by it's source and destination addresses and ports. To implement LAS scheduling, a network router can easily (a) identify the first packet and subsequent packets in a flow, (b) store the sequence number of the latest packet that has arrived in the flow, and (c) use that sequence number to insert the packet into the priority queue[2]. Note that the priority queue insertion is feasible even at high data rate[2].

Packets are served in order of earliest sequence number first, since sequence number corresponds to attained service. Packets having the same sequence number, which are from flows that have so far received the same amount of service, are served in first-come-first serve order. These rules imply that packets from flows that have received the same amount of service are served in an approximate round-robin fashion, which is an approximation of the processor-sharing service of equal priority jobs in the idealized LAS policy.

Keeping per-flow state is a challenging task for routers in the Internet backbone, due to the large number of simultaneously active flows. However, congestion and consequent packet losses occur most often at the edge and the access links of the Internet [17], where the number of active flows is more moderate. Furthermore, since bottleneck queues have a significant impact on end-to-end performance, deploying LAS in the access and edge routers that transmit packets over bottleneck links should reap the benefits of LAS in terms of reducing the response time of short TCP flows.

## 1.2 Modified LAS Policies: Motivation

Recent work [19] has shown that although LAS scheduling on a bottleneck link can greatly improve the average flow transfer time for short flows without appreciably degrading the average total transfer time of long flows, the individual packets in a long flow can experience significantly higher variance in their transfer time as compared with the case that the router uses first-come first serve scheduling for all packets. For download applications, and for streaming applications that buffer data at the destination to mask the variability in packet transfer times, the jitter in packet delivery does not hinder application performance. On the other hand, we observe in this paper that since a LAS scheduler already maintains a small amount of state for each flow, extensions to the LAS scheduler can provide a simple differentiated service for flows that require lower jitter, as well as for other flows that may be willing to pay for improved service. To illustrate this point, we consider the following two-class policies, in which ordinary (class 2) flows have packet priority equal to the sequence number, $x$, while priority (class 1) flows have modified packet priority values that are a function of the sequence number, $P(x)$, as defined for each policy:

- *LAS-fixed (k)*: Packets in class 1 flows have a constant priority value $P(x) = k$, for a specified integer $k > 0$.

- *LAS-linear (k)*: Packets in class 1 have priority value $P(x) = x/k$ for a specified integer $k > 0$.

- *LAS-log (k)*: Packets in class 1 have priority value $P(x) = (\log_2(x))^{\frac{1}{k}}$, for a specified integer $k > 0$.

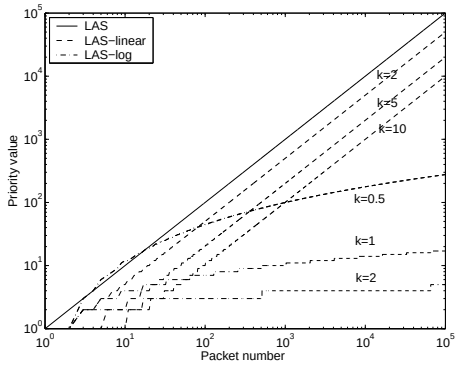



**Figure 1:** ***Priority value as a function of attained service*** $x$

Note that smaller priority values have higher priority, with priority value 1 representing the *highest* priority. In each policy, the function $P$ determines the relative priority value of each packet in a class 1 flow. If the priority values of the consecutive packets in a class 1 flow increase more slowly than linearly, a class 1 flow will have a lower expected transfer time than a similar size class 2 flow. Refinements in the priority assignment function to more than two classes of flows are also possible. We use the above two-class policies to obtain some initial insights into the relative advantages and disadvantages of the different types of priority value functions.

A flow of a given size is said to be *penalized* by a policy if it has higher mean transfer time under the policy than with a standard FCFS router. A recent hybrid scheduling policy called FLIPS, which uses LAS only for flows shorter than a specified threshold value, was proposed to reduce the penalty for large network flows [8]. In contrast, the differentiated policies proposed in this paper reduce the penalty for differentiated (class 1) flows that require or pay for better service. If the differentiated large flows are a small fraction of all large flows, the penalty for the differentiated flows can be eliminated.

## 1.3 Modeling Approach

The performance of size-based scheduling policies for Internet routers has generally been analyzed using ns simulations of flows that share a common bottleneck (e.g., [23, 10, 8]). In other prior work, analytic models have been developed and applied to compare the theoretical properties of LAS and related scheduling policies. Examples of these analytic models include *shortest remaining processing time* (SRPT) [11, 1], *least attained service* (LAS) [20, 8, 15], a study of the asymptotic performance of scheduling policies [12], and a new scheduling policy known as *Fair Sojourn Protocol* (FSP) [9]. In most cases, these papers investigate unfairness or optimality properties of the policies. For example, Wierman and Harchol-Balter [22] classify scheduling policies based on their unfairness.

In this paper, we take a different approach. That is, we derive analytic models that estimate, for each router scheduling policy we consider, the average flow transfer time as a function of flow size for TCP flows that share a given bottleneck link. The models also estimate average packet transfer time as a function of sequence number (i.e., position in the flow). One goal is to produce the same estimates analytically as are obtained in an ns simulation of the dumbbell topology with the router scheduling policy, and

[1] Coefficient of variation is the standard deviation divided by the mean of a distribution.

[2] We use the term sequence number in the sense of **counter for the amount of service attained** by a flow.

TCP flows generated by the ns Web workload model. At first this may seem like a daunting task, since the TCP flow control policy and the router scheduling policy interact in complex ways that may be difficult to capture analytically. For example, previous analytic models of router scheduling policies assume that a job arrives to the server all at once, whereas the packets belonging to a single flow arrive at a router at disparate points in time that are partially determined by the TCP congestion avoidance protocol.

We make three key observations that enable the development of the analytic models. First, due to the nature of a system bottleneck at link utilizations (e.g., 70%) where differences in scheduling policy performance are observable, each flow is likely to have at least one packet queued at the bottleneck link a high fraction of the time. Second, for LAS-based scheduling of two flows that have so far received the same amount of service, if the next several packets from one of the flows arrives after the corresponding packets from the other flow have been served, the late packets will have priority for service after they arrive and thus the flow with late packets will "catch up" to the flow whose packets arrived earlier. Third, for FCFS scheduling of the packets waiting for the bottleneck link, moderate to large flows that have moderate propagation delays and share the common bottleneck will have their packets transmitted in an interleaved fashion; the "quantum" for the interleaving varies due to the dynamic TCP window size and unpredictable packet arrival times, but is small compared to the flow size. Due to these observations, we conjecture that the flow transfer time for each router scheduling policy can be analyzed assuming each flow approximately always has at least one packet in the queue for the bottleneck link. Using this conjecture, we show how previous theoretical models of scheduling policies, with customizations to model connection establishment, can be used to obtain the average flow transfer time as a function of flow size and average packet transmission time as a function of sequence number for the LAS and FCFS routers, respectively. We also use the conjecture to derive new models of the average flow and packet transfer times for each of the new differentiated service policies at the bottleneck link. We validate the conjectures by implementing the LAS and differentiated service policies in the ns simulator, and comparing the analytic results against the simulation results for the "dumbbell" bottleneck topology used in many previous ns simulation studies.

A key result in the paper is that the analytic results for the widely implemented first-come first-serve routers and single-class LAS scheduling, as well as the proposed differentiated service policies, are in excellent agreement with the ns simulation estimates. This result could motivate development of analytic models for many other types of design questions that have commonly been evaluated using ns simulations. Key advantages of the analytic models include very quick solution time, and key insight into the system features that have a direct impact on the computed performance. Furthermore, a detailed analysis of initial discrepancies between the analytic results and simulation results, for FCFS and LAS router scheduling, revealed an error in the ns workload parameters often specified for a popular ns workload generator [3]. The error in these workload parameters was not obvious from the simulation results alone, because simulation estimates of flow transfer times appeared plausible for the heavy tailed flow sizes that are found in the Internet. However, the analytic models predicted lower average transfer time as a function of flow size for the same heavy tailed distribution of flow sizes. As shown in section 4, a detailed examination of the simulation flow times revealed that the ns2 workload parameter values were unrealistic. This illustrates a common experience in analytic modeling, which is that analytic models can also be very useful in validating and correcting the simulation.

This paper uses the validated analytical models to compare the performance of the LAS-based differentiated service policies, the single-class LAS policy, and FCFS router scheduling. The validated simulator is also used to evaluate jitter, in more detail than possible analytically, for the best of the differentiated service policies and for the LAS policy. The results show that, compared to a FCFS router, LAS-log simultaneously achieves low mean transfer time for short flows, a mean transfer time for the ordinary long flows similar to LAS, and similar jitter in the packet transfer times for long flows.

We validate the new policies in Section 4 and compare their performances in terms of mean transfer time, loss, jitter, and fairness in Section 5. Section 6 summarizes the paper.

## 2. FCFS AND LAS SCHEDULING

In this section we provide two analytic models of the average flow transfer time as a function of flow size, $T(x)$, for flows that (a) share a common bottleneck link and (b) experience negligible contention at other points in their respective transmission paths. One of the two models provides results for the widely deployed FCFS scheduling of packets on the bottleneck link; the other provides results for the single-class LAS scheduling policy. The goal is to estimate average transfer time for such flows as accurately as is estimated by ns2 simulations of the simple dumbbell topology with a given flow arrival rate, flow size distribution, bottleneck link speed, bottleneck link scheduling policy, and range of flow propagation delays.

### 2.1 Notations and Assumptions

We assume that (a) flow sizes are expressed in units of packets of a specified size, and (b) the unit of time for all measures that require a time unit, is the time it takes to transmit one packet on the bottleneck link. The overall distribution of flow size is given by $F(x)$, and the complement of the distribution is given by $F^c(x)$. Key measures for modeling LAS policies include (1) the moments of the flow size distribution, $\overline{x_x^n}$, (2) the link load, $\rho_x = \lambda\overline{x}$, (3) the $n$th moments of the flow size distribution in which the flows are truncated at a given value $x$, which are given by

$$\overline{x_x^n} = \int_0^x y^n dF(y) + x^n F^c(x), \tag{1}$$

and (4) the load due to the truncated flows,

$$\rho_x = \lambda\overline{x_x}.$$

For each scheduling policy, we derive $T(x)$, the *m*ean end-to-end flow transfer time as a function of flow length $x$. For a flow of a given size $x$, the average time to transfer the first $j$ packets of the flow is equal to $T(j)$. Thus, the average time between when packets $j$ and $j+1$ complete transmission is equal to $T(j+1) - T(j)$.

An important goal is to create the simplest models that contain the essential details for producing the same results as a (correct) detailed ns2 simulation, for key workloads and system configurations of interest. One important advantage of such models is that they clarify which system details have a principal impact on observed policy performance. The analytic models can be used in place of simulations to quickly compare policy performance with respect to $T(x)$ for a variety of workloads and system configurations, such as varying relative load of high priority and low priority flows, and for a range of bottleneck link speeds. Simulation can then be used to evaluate the most promising policies using more detailed performance measures and for system configurations that violate the assumptions in the analytic model. To these ends, we make the simplifying assumptions outlined below.

The workloads of interest involve TCP flows, with realistic flow size distributions, that share a common bottleneck link that has utilization in the range of 70% or higher. This is the range of link utilizations for which link scheduling policy has a discernable impact on system performance. We assume that *most* flows through the bottleneck have modest disparity in their respective round trip propagation delay (e.g., propagation delay under 100 milliseconds), and that the queue for the bottleneck link is non-empty most of the time. Unless otherwise stated, we also assume that each flow has a window of packets in flight (i.e. at least one packet in the bottleneck queue) most of the time.

We derive the analytic models for the case of low packet loss rate (i.e., the link buffer is large enough to achieve a loss rate under, for example, 2%). Simulations can be used to study the impact of configurations with higher loss rate on the most promising policies that are identified in the low-loss analytic comparisons. Alternatively, the analytic models could be extended to estimate and account for packet loss; we defer such extensions to future work.

Initially, we derive the mean transfer time assuming zero propagation delay between the flow endpoints and the bottleneck, for each flow. We then add two times the average round trip propagation delay for the flows. This estimates the average delay for connection establishment, which is non-negligible for short flows. We assume that the propagation delay for (most) packets other than the first packet in the flow is fully overlapped with the bottleneck queueing delay and service time of at least one other packet in the same flow.

For simplicity in the presentation of the models, we assume that all packets have uniform size (e.g., one kilobyte). This is also a common assumption in ns2 simulations. We note that the models can easily be extended to have a distribution of average packet size for each flow, where the distribution could depend on the class of the flow as well as the flow size. However, such extensions are beyond the scope of the paper.

## 2.2 FCFS Model

If the packets are scheduled on the bottleneck link in FCFS order, transmissions of packets from different flows will be interleaved since the packets in each flow arrive at variable times. The "quantum" for interleaving will typically be non-uniform because a variable number of consecutive packets in a particular flow may arrive between two consecutive packets in another active flow. However, we assume roughly similar flow rates for moderate to large flows that each have moderate propagation delay, share the common bottleneck, and experience negligible contention in the rest of their transmission path. Thus, we assume that the scheduling of the flows on the link is approximately round-robin with an average quantum size that is small for moderate to large flows. Ignoring the delays for connection establishment and retransmitting lost packets, we estimate the average transfer time for a flow of size $x$ using the processor-sharing approximation for round-robing scheduling, as follows [15]:

$$T(x)_{PS} = \frac{x}{1-\rho} \tag{2}$$

We note that this processor-sharing approximation to round robin is likely to underestimate the mean transfer time for short flows, since the model does not include the impact of the variable and non-negligible quantum size on the average delay seen by such flows. This impact of this approximation will be be evaluated in the model validations.

## 2.3 LAS Model

The LAS link scheduling policy was defined in Section 1. Assuming each flow has at least one packet in the queue for the bottleneck link most of the time, the mean transfer time for a flow of size $x$ under the LAS policy, expressed in units of the time to transmit a packet on the bottleneck link, can be modeled by the mean total time to serve a job of size $x$ at a LAS server, which is given by [15] as:

$$T(x)_{LAS} = \frac{W_o(x) + x}{(1-\rho_x)} \tag{3}$$

where $W_o(x)$ is the average backlog of packets with sequence numbers less than or equal to $x$ at a random point in time,

$$W_o(x) = \frac{\lambda \overline{x_x{}^2}}{2(1-\rho_x)}.$$

We note that although each flow does not actually have a packet in the queue at every point in time, as would be required for the above model to be exact, a packet that arrives a little later than the time at which it would receive service in the exact model will be served earlier than all packets with higher sequence numbers, and thus will cause one unit of interference in the total transfer time of the other flows that have so far received more service. Thus, we anticipate that this model of the LAS router may be more accurate than the model of the FCFS router.

# 3. DIFFERENTIATED LAS SCHEDULING

Under the single-class LAS scheduling policy, packets later in a flow have lower priority and thus longer expected transmission delays. In this section we consider three alternative two-class differentiated service scheduling policies that are each simple extensions of the single-class LAS policy. The goals of these differentiated service scheduling disciplines are to reduce the expected total transfer time of (moderate to long) class 1 flows, and to reduce the differential between the average transfer time for earlier and later packets in the class 1 flows, without appreciably increasing the expected transfer time of short class 2 flows. The policies are LAS-fixed(k), LAS-linear(k) and LAS-log(k), which have packet priorities for class 1 that are *fixed*, *linear* and *logarithmic* in their packet sequence number, respectively. Packets in Class 2 flows have priority equal to their sequence number, as under the single-class LAS policy. Scheduling of the packets on the outgoing link is otherwise the same as in the single-class LAS policy. Thus, packets in a class 1 flow have on average higher priority than the packets in a class 2 flow of the same size, but the long class 1 flows do not significantly interfere with the large number of very short class 2 flows.

Flows in class 1 will be called the *priority* flows and will be denoted by subscript or superscript $p$, while the flows in class 2 will be called the *ordinary* flows and denoted by subscript or superscript $r$. In these differentiated service policies, a fraction $q$ of the flows are class 1 flows. Each class of flows has a possibly different distribution of flow sizes, denoted by $F_i(x)$ for class $i$. The analytical models of the differentiated LAS scheduling policies presented in this section estimate average flow transfer time as a function of flow size. The models adopt the technique used to evaluate the LAS policy, which is to view each flow as a job with service requirement $x$ that arrives to the bottleneck queue at a random point in time.

## 3.1 LAS-fixed(k) Model

In the LAS-fixed(k) policy, each packet in a class 1 (priority) flow has priority equal to the $k$th packet in a class 2 flow, where $k$ is a parameter of the policy. The motivation for this policy is

to provide the same average delay in the bottleneck link queue for each packet in a class 1 flow. Clearly flows that are shorter than $2k$ would receive lower average priority as class 1 flows than as class 2 flows. Thus, we anticipate that only flows longer than $2k$ would select class 1.

We first derive the average flow transfer time for the first $x < k$ packets in an ordinary class 2 flow, $T_r(x)$. These packets do not see any interference from class 1 flows. Thus, the average flow transfer time is given by the time in a single-class LAS system that serves only class 2 flows that are truncated at size $k-1$. Expressed in units of the time to transmit a packet on the bottleneck link,

$$T_r(x) = \frac{W_{x,r} + x}{1 - \rho_{x,r}}, \quad x < k, \tag{4}$$

where, $\rho_{x,r}$ is the load due to class 2 flows with service requirements less than $x$, and $W_{x,r}$, is the corresponding average number of backlogged class 2 packets, is given as

$$W_{x,r} = \frac{\lambda_r \overline{x^2_{x,r}}}{2(1 - \rho_{x,r})} \tag{5}$$

with $x^n_{x,i}$ is defined similarly to Equation (1):

$$\overline{x^n_{x,i}} \triangleq \int_0^x y^n dF_i(y) + x^n F_i^c(x). \tag{6}$$

For a priority class 1 flow, let $R_p(j), j > 1$ denote the average time between when packets $j-1$ and $j$ complete service at the bottleneck link. We can express $T_p(j)$, the average total flow transfer time for the first $j$ packets in a class 1 flow, as follows:

$$T_p(j) = T_p(j-1) + R_p(j). \tag{7}$$

Similarly, the average transfer time for an ordinary class 2 flow of size $k$ is given by

$$T_r(k) = T_r(k-1) + R_r(k). \tag{8}$$

To derive expressions for $T_p(1)$, $R_p(j)$, and $R_r(k)$, we need the following two quantities: $\overline{r_p}$, the average size of the priority flows not including the first packet in the flow:

$$\overline{r_p} = \overline{x_p} - F_p(1), \tag{9}$$

and $\overline{R_p}$, the average service time of class 1 packets that have sequence number greater than 1:

$$\overline{R_p} = \frac{\sum_{j=1}^{\infty} R_p(j+1)[1 - F_p(j)]}{\overline{x_p} - 1}. \tag{10}$$

The average time for the first packet in a class 1 flow to complete service at the bottleneck link, $T_p(1)$, expressed in units of the time to transmit a (fixed size) packet on the bottleneck link and assuming the packet arrives at a random point in time, is given by

$$\begin{aligned} T_p(1) &= W_{k-1,r} + \lambda_r F_r^c(k) R_r(k) + \lambda_p R_p(1) + \\ &\quad \lambda_p \overline{r_p}\overline{R_p} + \lambda_r [T_p(1) - 1]\overline{x_{k-1,r}} + 1. \end{aligned} \tag{11}$$

The first two terms in the above sum are the the average number of backlogged class 2 packets that have sequence number less than $k$ and equal to $k$, respectively. The third and fourth terms are the average number of backlogged class 1 packets that have sequence number equal to 1 and sequence number greater than 1, respectively. (Recall that all of these class 1 packets have equal priority $k$.) The fifth term is the average number of class 2 packets with sequence number less than $k$ that arrive while the first packet in a class 1 flow is waiting in the queue. Finally, the last term represents the transmission time of the first packet in the class 1 flow.

Let the time between when packets $j-1$ and $j$ in a class 1 flow complete service at the bottleneck link be the time during which packet $j$ is the next packet in the flow to be transmitted on the link. This time has average value $R_p(j), j > 1$, given in units of a packet transmission time on the link as follows,

$$\begin{aligned} R_p(j) &= \lambda_p \overline{x}_p R_p(j-1) + \lambda_r F_r^c(k) R_p(j-1) \\ &\quad + \lambda_r [R_p(j) - 1]\overline{x_{k,r}} + 1, \quad j > 1. \end{aligned} \tag{12}$$

The first two terms in the above sum are the average number of class 1 priority packets, and the average number of ordinary packets with sequence number $x = k$, respectively, that arrive after packet $j-2$ is transmitted on the bottleneck link and before the transmission of packet $j-1$ is complete. These packets will be transmitted on the link after packet $j-1$ is transmitted and before packet $j$ is transmitted. The third term in the above sum is the average number of class 2 packets with sequence number less than $k$ that arrive while packet $j$ in a class 1 flow is the next packet in the flow to be transmitted on the bottleneck link. Finally, the fourth term represents the transmission time of packet $j$ in the class 1 flow.

To derive the average time between when packet $k-1$ and packet $k$ in a class 2 flow are transmitted on the bottleneck link, $R_r(k)$, we assume the first packet of the class 2 flow arrives at a random point in time. Note that the priority $k$ packets from class 1 and class 2 that are in the queue when the first packet in the class 2 flow arrives, are still in the queue when the $k-1$st packet completes its transmission, assuming the flow continuously has at least one packet in the queue. The average backlog, estimated by the first three terms in the sum below, is summed together with the priority $k$ packets that arrive from other flows during the time to transmit the first $k-1$ packets in the class 2 flow, which are the next two terms in the sum below, to obtain the total number of priority $k$ packets from other flows that must be transmitted after the class 2 packet with sequence number $k-1$, and before the class 2 packet with sequence number $k$. Thus,

$$\begin{aligned} R_r(k) &= \lambda_p R_p(1) + \lambda_p \overline{r_p}\overline{R_p} + \lambda_r F_r^c(k) R_r(k) + \\ &\quad [\lambda_p \overline{x_p} + \lambda_r F_r^c(k)] T_r(k-1) + \\ &\quad \lambda_r R_r(k) \overline{x_{k,r}} + 1. \end{aligned} \tag{13}$$

Note that the last two terms in the above sum are the average number of class 2 packets with sequence number less than $k$ that arrive while packet $k$ in a class 2 flow is the next packet in the flow to be transmitted on the bottleneck link, and the time to transmit packet $k$ in the class 2 flow, respectively.

Finally, we derive the average transfer time for class 2 flows with size $x > k$,

$$T_r(x) = \widehat{W_x} + \lambda_r T_r(x) \overline{x_{x,r}} + \lambda_p T_r(x) \overline{x}_p + x, \quad x > k, \tag{14}$$

where $\widehat{W_x}$ is the average backlog, at a random point in time, of all packets from class 1 and class 2 flows that have priority less than $x$, given by

$$\widehat{W_x} = \frac{\lambda[q\overline{x_p^2} + (1-q)\overline{x^2_{x,r}}]}{2(1 - \lambda[q\overline{x_p} + (1-q)\overline{x_{x,r}}])}. \tag{15}$$

## 3.2 LAS-linear(k) Model

The LAS-fixed(k) policy has the key property that each packet in a priority class 1 flow has the same expected delay in the bottleneck queue. However, this policy also has the drawback that *all* packets in the priority flows have higher priority than ordinary class 2 packets that have sequence number greater than $k$. Thus, depending on the fraction of flows that are identified as priority class 1 flows, and the distribution of class 1 flow sizes, class 2 flows that

are longer than $k$ may have high expected total transfer time. The LAS-fixed(k) model can be used to explore the quantitative impact of the workload parameters on the average flow and packet transfer times. In this section we develop an alternative differentiated service policy.

LAS and LAS-fixed(k) represent two extremes in the growth rate of the priority function, $P(x)$, for class 1 packets, namely $P(x) = x$, and $P(x) = k$. Here we consider an intermediate policy, namely the LAS-linear(k) policy in which $P(x) = x/k$. Figure 1 illustrates this priority function, which grows more slowly for larger values of $k$.

To compute the mean flow transfer times for class 1 priority flows and class 2 ordinary flows we first need to compute the moments of the flow size distribution for flows that interfere with a class 1 flow of size $x$. An arriving priority flow that requires $x$ units of service in a LAS-linear scheduler will be delayed by other priority flows that have received less than $x$ units of service and by ordinary flows that have received less than $x/k$ units of service. The moments of the flow size distribution for priority flows truncated at $x$ and ordinary flows truncated at $x/k$, are given by

$$\overline{x^n_{x,linear(k)}} = q\overline{x^n_{x,p}} + (1-q)\overline{x^n_{x/k,r}}, \tag{16}$$

where $\overline{x^n_{x,i}}, i = p, r$ is defined in equations 6. The bottleneck link load for these flows, $\rho_{x,linear(k)} = \lambda\overline{x_{x,linear(k)}}$.

To compute the mean transfer time for a priority class 1 flow of size $x$, $T_p(x)$, we note that the flow size defines which packets from other flows will delay the flow, and otherwise the packet queue is a single-class LAS queue. Thus,

$$T_{p,linear}(x) = \frac{W_{x,linear(k)} + x}{(1-\rho_{x,linear(k)})}, \tag{17}$$

where

$$W_{x,linear} = \frac{\lambda\overline{x^2_{x,linear}}}{2(1-\rho_{x,linear})} \tag{18}$$

An ordinary class 2 flow of size $x$ will have a sequence of priorities for each of its packets that corresponds with the sequence of packet priorities in a class 1 flow of size $kx$. Thus,

$$T_{r,linear}(x) = \frac{W_{kx,linear(k)} + x}{(1-\rho_{kx,linear})}. \tag{19}$$

## 3.3 LAS-log Model

The LAS-linear(k) policy provides a slower decrease in consecutive packet priority for class 1 flows than the single-class LAS policy, but the decrease still as the same form, namely linear. Initial results from the model for LAS-linear(k) showed that later packets in the long class 1 flows have high expected delay in the bottleneck queue for workloads of interest. This illustrates the use of the models to quickly assess the properties of policy performance. We thus propose the LAS-log(k) policy, which has a much slower growth of the packet priority function, $P(x)$, as illustrated in Figure 1.

A high priority flow that requires $x$ service units under *LAS-log* is preempted by all high priority flows that have received less than $x$ units of service and all ordinary class 2 flows that have received less than $(log_2 x)^{(1/k)}$ units of service. The moments of the distribution of flow sizes for priority and ordinary flows truncated at $x$ and $(log_2 x)^{(1/k)}$, respectively, are given by

$$\overline{x^n_{x,log}} = q\overline{x^n_{x,p}} + (1-q)\overline{x^n_{(log_2 x)^{1/k},r}} \tag{20}$$

We define the load that these flows place on the bottleneck link , $\rho_{x,log} = \lambda\overline{x}_{x,log}$.

Since this policy is similar to LAS-linear(k) except for the priority function, the equations for $T_{p,log(k)}$ and $W_{x,log(k)}$ are the same as the corresponding equations for the LAS-linear(k) policy, with subscripts "linear" replaced by "log". The equation for $T_{r,log(k)}$ is the same as the corresponding equation for the LAS-linear(k) policy, with the same subscript replacement and with the subscript $kx$ replaced by $2^{x^k}$.

## 3.4 LAS-fixed(k) Model: UDP Priority Flows

The above models were developed assuming that both the priority class 1 flows and the ordinary class 2 flows are TCP flows. A key feature of such flows is that most of the time there is a window of packets in flight from the source. Another key workload of interest is one in which the priority flows are long-lived UDP media streams, in which the sending of packets is paced because the client is displaying the media stream as it arrives. Media streaming can alternatively be performed using TCP, but scalable streaming protocols benefit from the slower pacing under UDP, since more clients can share later portions of the stream if the data is not downloaded too quickly.

In this section, we consider the case that each class 1 priority flow is a UDP stream that sends a packet every $r$ units of time, where the unit of time is the time to transfer one packet on the bottleneck link. In this case, $T_p(x) = x/r$.

We illustrate that the models can be modified to compute the average transfer time as a function of flow size for ordinary class 2 flows, by showing how this is done for the LAS-fixed(k) policy. Specifically, the first two terms in equation (13), estimate the average number of priority packets in the queue at a random point in time. These terms are replaced by $\frac{1}{r}R_{p,UDP}$, where $R_{p,UDP}$ is the mean transmission time for a class 1 UDP packet. Thus,

$$\begin{aligned} R_r(k) &= \frac{1}{r}R_{p,UDP}(k) + \lambda_r F^c_r(k) R_r(k) + \\ &\quad [\frac{1}{r} + \lambda_r F^c_r(k)] T_r(k-1) + \\ &\quad \lambda_r R_r(k)\overline{x_{k,r}} + 1. \end{aligned} \tag{21}$$

To compute $R_{p,UDP}$, we estimate the average number of class 2 and class 1 packets in the queue when the UDP packet arrives (first three terms in the sum below), and the average number of ordinary class 2 packets with sequence number less than $k$ that arrive while the UDP packet is in the queue (the fourth term below), to obtain

$$\begin{aligned} R_{p,\text{UDP}} &= W_{k-1,r} + \lambda_r F^c_r(k) R_r(k) + \frac{1}{r}R_{p,UDP} \\ &\quad + \lambda_r (R_{p,UDP} - 1)\overline{x_{k-1,r}} + 1. \end{aligned} \tag{22}$$

As in the LAS-fixed(k) model for TCP class 1 flows, $T_r(k+1) = T_r(k) + R_r(k)$ and $T_r(x)$ for $x > k$ is as given in equation (14).

# 4. MODEL VALIDATIONS

## 4.1 Methodology

To validate the analytical models we implement the scheduling policies for network links in the network simulator ns2 [14]. In the validation experiments, we use the dumbbell topology shown in Figure 2, where C1-CN are clients that each simulate a series of Web sessions. *Source* is a UDP source sending a priority flow to the destination *Sink*, which is only used in the validation of the LAS-fixed(k) model for UDP priority flows. The bottleneck link in the network topology is R1–R0. Note that the total one-way end-to-end propagation delay between client and server is 50 milliseconds.

During each Web session, the client requests a series of Web pages, each containing several objects from a randomly chosen

| pages/session | objs/page | inter-session time (sec) | inter-page time (sec) | inter-obj time (sec) | obj size (packets) |
|---|---|---|---|---|---|
| *Exp(60)* | *Exp(3)* | *Exp(8)* | *Exp(5)* | *Exp(0.5)* | *P(1,1.2,12)* |

**Table 1: Web traffic profile ($Exp(1/\mu)$ denotes the exponential distribution with mean $1/\mu$)**

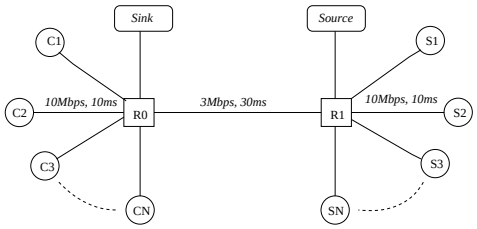



**Figure 2: *Simulated network topology***

server in the pool S1-SN of servers. This Web model was first introduced by Feldmann et al. [7]. The simulated Web traffic is specified by setting the distributions of inter-session, inter-page and inter-object time (all in seconds), session size (in web pages), page size (in objects) and object size (in packets) as illustrated in Table 1. The particular combination of parameters in the table will result in a total load $\rho$ on the bottleneck link of $\rho = 0.73$. To vary the load, we adjust the mean inter-session time. The transfer of each object creates a new TCP flow. For this reason, object size and flow size are the same, and we use both terms interchangeably. The *object sizes* are Pareto distributed, denoted as $P(a, \alpha, \bar{x})$, where $a$ is the minimum possible object size, $\alpha$ is the shape parameter, and $\bar{x}$ is the mean object size. The Pareto distribution has a high variance and a sub-exponential tail, and models flow size distributions that have been observed in the Internet [13, 18, 6].

Using the simulator, we can obtain performance measures such as the mean transfer time as a function of flow size, mean packet transfer time as a function of position in the flow, packet loss rate, and jitter in the inter-packet arrival times at the client. Simulations are run for 6000 seconds and the data is collected after a warm up period of 2000 seconds. We compare the results obtained from simulation with the analytical results for mean flow transfer time as a function of flow size. If the results are in agreement, we consider the analytical model (and the simulator) validated. The validated analytical model for a particular scheduling policy can then be used to evaluate the performance of the policy.

## 4.2 A Note on the Workload Model

The ns Web workload generator we used to validate our analytic models has been used extensively [7]. The validation of our flow level models using this workload generator revealed that the choice of the workload model input parameter must be done carefully. Otherwise, while achieving a target load e.g. $\rho = 0.7$, the generated workload may include some unrealistic long overload periods. The starting point of this finding was an initial discrepancy we observed between analytic and simulation results for average flow transfer times. The discrepancy was originally detected for heavy tailed flow sizes, in which both simulation and analytic results appeared plausible. We then simplified the flow size distribution to the exponential distribution to help determine which results might be in error. The curves for exponential flow sizes immediately indicated that the simulation results would not be expected in practice. (e.g., an *average* flow transfer time of more than 25 sec for flows with less than 100 packets).

A closer examination of the number of timeouts and other measures in the ns2 simulations showed that the ns2 workload generator was generating an unrealistic number of parallel flows during particular periods in the simulation. The reason behind this is that some parameters of the workload model (arrival rate of sessions, number of pages per session, number of object per page and average object size) control the average (macroscopic) load on the bottleneck link while other parameters (inter-page and inter-object times) control the microscopic load of the model. If the latter are set to too large values (e.g., inter-page time of 10 seconds, as has been used in the previous literature), sessions overlap unrealistically, which generates transient yet long high overload periods. The use of a better combination of parameter values for the workload generator, as given in the table, gave results that agree with the analytic model .

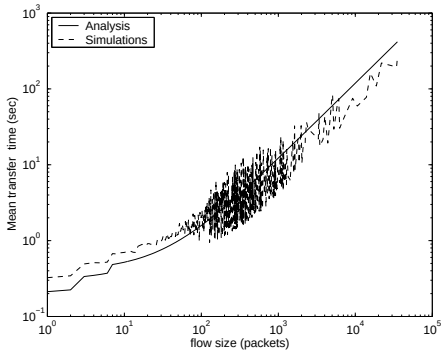



**Figure 3: *Validation of FCFS model, load $\rho = 0.73$, loss rate = 1.2%***

## 4.3 FCFS Model Validation

Figure 3 presents the mean flow transfer time versus flow size for both the analytical model of the dumbbell topology with FCFS link scheduling, and its corresponding ns simulation. The results demonstrate that the analytic model of the FCFS router has very good overall agreement with the ns2 simulation estimates. The mean flow transfer time is slightly underestimated for flow of size less than 100 packets (as anticipated in Section 2.2). The average transfer time is perhaps also slightly overestimated for flows of size larger than 10,000 packets. However, the simulation with Pareto flow sizes needs to be run for a very long time to get accurate measures of mean transfer time for flows larger than 100 packets. The overall agreement between analytic and simulation estimates is excellent, and the small discrepancy for short flows can easily be taken into account when comparing analytic results for FCFS against results for other policies.

## 4.4 LAS Model Validation

To validate the analytical model for the LAS link scheduler, we considered system workloads that have various values of link load, loss rate, and flow size distribution. In each figure we plot the mean transfer time versus flow size obtained from the analytic model as well as from the simulation of the *TCP* flows in the Web workload.

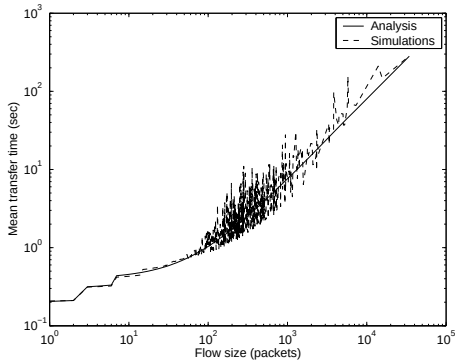



**Figure 4:** ***Validation of LAS model, load*** $\rho = 0.73$***, loss = 1.2%***

Figure 4 shows validation results for load $\rho = 0.73$ and loss rate 1.2%. We observe a near perfect agreement between the analytic model and the simulation.

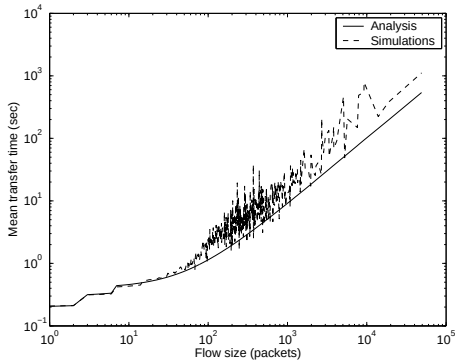


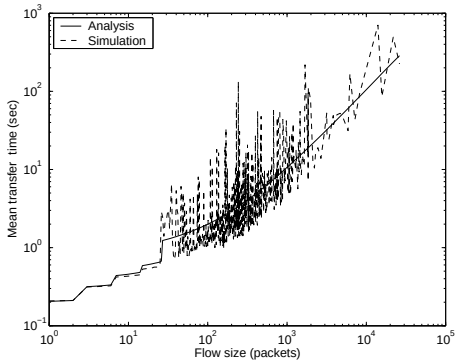

**Figure 5:** ***Mean transfer time validation for load*** $\rho = 0.73$ ***and loss rate*** $= 3.6\%$

It is well known that the throughput of the commonly deployed TCP protocols (e.g., TCP Reno) is inversely proportional to the square root of the loss rate [16]. Thus, an increase in loss rate will reduce the throughput and increase the average flow transfer time. The analytic models presented in Sections 2 and 3 do not consider the impact of packet loss, and thus will underestimate the mean flow transfer time when packet loss is significantly higher than 1%. This is illustrated in Figure 5, which shows the mean transfer time results in a system with LAS scheduling and loss rate of 3.6%. There is good agreement between the analytic and simulation results for small flow sizes, since packets in the short flows have a high priority under LAS and are rarely lost. Most of the loss is experienced by the long flows, which have mean transfer time higher than predicted by the analytic model. The analytic models can easily be extended to include the impact of packet loss on the mean flow transfer time, using a variation on the techniques in [21], but such extensions are beyond the scope of this paper.

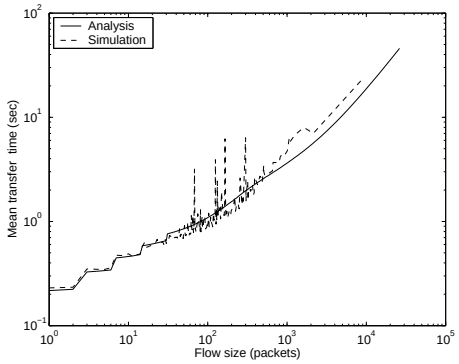
Extending of the analytic models to include the impact of packet loss is is

## 4.5 LAS-fixed(k) Model Validation

In the validations of the differentiated LAS policies, unless otherwise stated, the priority flows are a random 10% of the Web TCP flows (i.e., $q = 0.1$), and thus comprise 10% of the total load on the bottleneck link. For validating the LAS-fixed(k) model for UDP priority flows, the priority flow is a single long-lived UDP flow that




(a) *Ordinary Flows,* $k = 25$




(b) *Priority Flows,* $k = 25$

**Figure 6:** ***Mean transfer time validation for LAS-fixed(k) TCP;*** $\rho = 0.73$***, loss rate*** $= 1.2\%$***, and*** $q = 0.1$

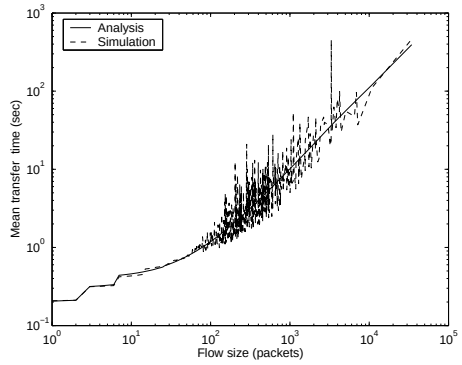
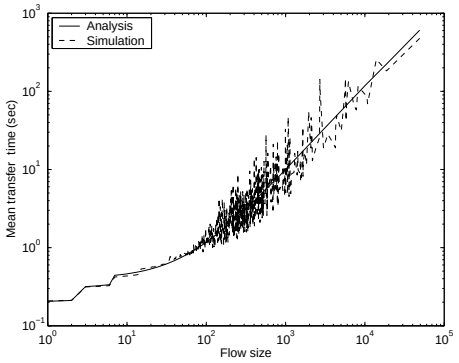


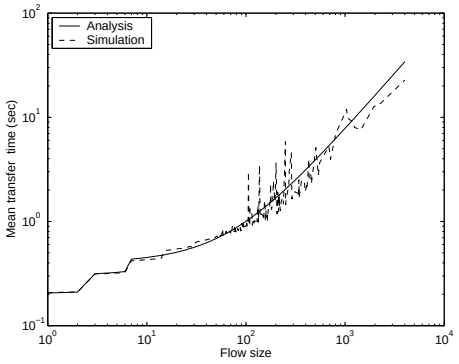

**Figure 7:** ***Mean transfer time validation for LAS-fixed(k) UDP; load*** $\rho = 0.73$***, loss rate*** $= 1.2\%$***, and*** $q = 0.1$

sends packets at a constant bit rate during the entire simulation. The rate of the flow is chosen such that it contributes 10% of the total load.

Figure 6 shows the validation results for LAS-fixed(k), with $k = 25$ and TCP priority flows. We observe that the simulation validates the model. Similarly, Figure 7 shows the results for ordinary TCP flows when $k = 25$ and the priority flow is the UDP flow. Observe again that the analytic mean flow transfer time as a function of flow size for the ordinary flows is in excellent agreement with the simulation results. Note also that the mean total transfer time of the priority flow is not validated, since the priority flow is at fixed rate and is active during the entire simulation.

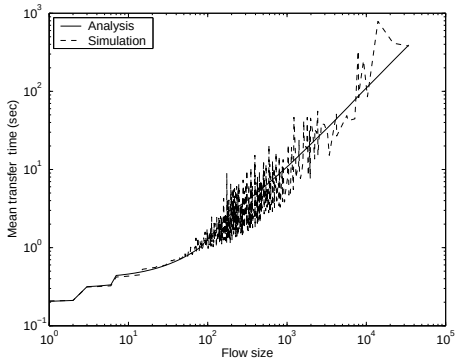



(a) *Ordinary flows,* $k = 5$

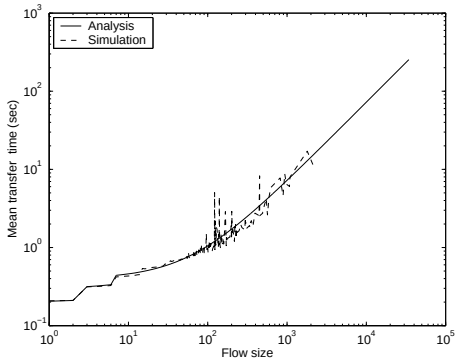



(b) *Priority flows,* $k = 5$

**Figure 8:** ***Mean transfer time validation for LAS-linear; load*** $\rho = 0.73$***, loss rate= 1.2%, and*** $q = 0.1$**.**




(a) *Ordinary flows,* $k = 0.5$




(b) *Priority flows,* $k = 0.5$

**Figure 9:** ***Mean transfer time validation for LAS-log; load*** $\rho = 0.73$***, loss rate*** $= 1.2\%$***, and*** $q = 0.1$**.**

## 4.6 LAS-linear(k) Model Validation

Figures 8(a) and (b) show the mean transfer time as a function of flow size for ordinary and priority flows, respectively. The simulation results and analytical estimates are in excellent agreement. We have observed the same level of agreement for other values of $k$.

## 4.7 LAS-log(k) Model Validation

Figures 9 and 10 show the mean flow transfer time as a function of the flow size under LAS-log(k), for $k = 0.5$ and $k = 2$, respectively. For $k = 0.5$, the simulation results and the analytical estimates are in excellent agreement. For $k = 2$ the mean transfer times of low priority flows differ slightly. However, the analysis and simulation results are in good agreement for $k = 2$; in particular, both results exhibit the significant increase in transfer time for the ordinary flows of size 3 as compared to ordinary flows of size 2. Note that while ordinary flows of size 2 need to share the link with priority flows up to size $2^{2^2} = 16$, ordinary flows of size 3 need to share the link with priority flows up to size $2^{3^2} = 64$.

In summary, in this section we have validated the analytical models of the FCFC, LAS, and differentiated LAS policies for a Web traffic profile with Pareto distributed object sizes. We presented representative results for particular values of $k$ and $q$; we have also obtained similar validation results for other parameter values and for exponentially distributed object sizes, for all policies. Given these validations, the analytical models can be used to evaluate the performance of LAS and differentiated LAS policies for flows that share a bottleneck link (e.g., at the edge of the network) and experience little other contention in the network. We apply the models in this way in the next section.

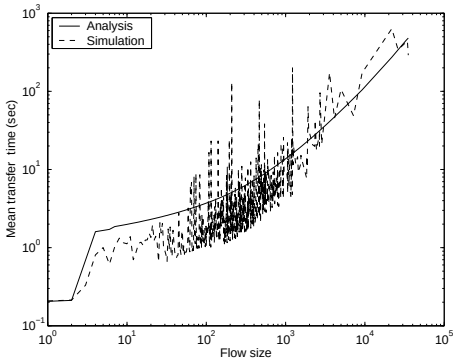
Extensions of the analytic models for multiple points of contention are possible, but are outside the scope of this paper. We note that simulations of topologies that have multiple bottleneck links are quite complex to specify and very time-consuming to run, so analytic models for such networks could be quite advantageous.

# 5. POLICY COMPARISONS

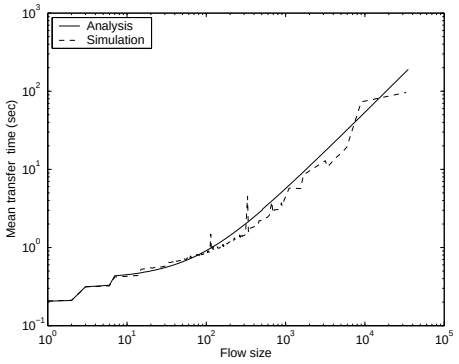
In differentiated LAS scheduling, ordinary flows have packet priorities equal to the position of the packet in the flow, whereas priority flows have packet priorities determined by a priority function $P(x)$, where $x$ is the position in the flow. Since $P(x) < x$, for each of the differentiated service policies, an ordinary flow with attained service $x$ will experience equal interference from other ordinary flows and greater interference from the priority flows than in the single-class LAS system. Conversely, a priority flow with attained service $x$ will experience equal interference due to other priority flows but less interference from ordinary flows in the differentiated LAS policy as compared with the single-class LAS system. In this section we provide results on the quantitative impact of the priority functions on mean flow transfer times, packet loss, and packet jitter.

## 5.1 Mean Flow Transfer Time

Figure 11 plots, for LAS and each differentiated LAS policy, the ratio with the mean flow transfer time under the policy to the average flow transfer time under FCFS, versus flow size. The results are shown for a system with load $\rho = 0.7$ and 30% of the flows having priority ($q = 0.1$). Compared with LAS, the LAS-fixed(k) policy with $k = 100$ provides the best performance for the (large)

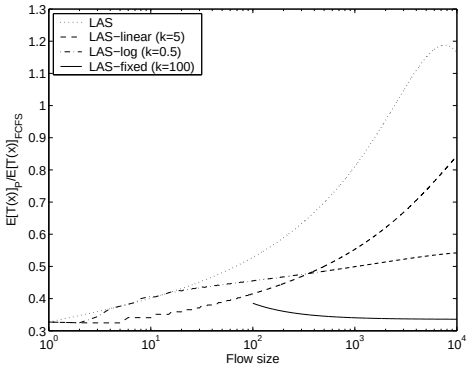



(a) *Ordinary flows,* $k = 2$

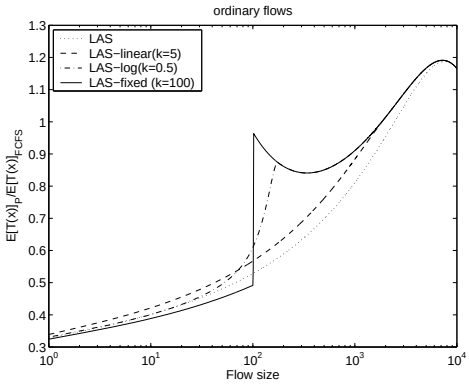



(b) *Priority flows,* $k = 2$

**Figure 10:** ***Mean transfer time validation for LAS-log; load*** $\rho = 0.73$***, loss rate*** $= 1.2\%$***, and*** $q = 0.1$**.**




(a) *Priority flows*




(b) *Ordinary flows*

**Figure 11:** ***Policy Performance Comparisons: Ratio of Mean Flow Transfer Time for Policy P to FCFS (*** $q = 0.1$**,** $\rho = 0.7$***).***

priority flows, but at a significant performance penalty for ordinary flows larger than $k$. The LAS-linear(k) with $k = 5$ achieves better performance than LAS for large priority flows, and similar performance for all other flows. The LAS-log(k) policy with $k = 0.5$ achieves better performance than LAS-linear(k) for large priority flow, with some performance penalty for moderate-size ordinary flows.

Figures 12(a) and (b) show the impact of a varying percentage of priority flows on the average transfer time of ordinary flows of size less than 100 or 200 packets, respectively. For ordinary flows of size less than or equal to 100, the transfer time under LAS-linear $(k = 5)$ and LAS-log $(k = 0.5)$ increases as the percentage of priority flows increases. However, for LAS-fixed$(k = 100)$ the priority flows do not interfere with the ordinary flows of size less than 100. Therefore, as the percentage of priority flows increases, the traffic volume due to ordinary flows of size less than 100 decreases, and thus the mean transfer time for short ordinary flows also decreases. The analytical results in this section are obtained using Bounded Pareto flow size distribution with $BP(a, \alpha, P)$ with $a = 1$, $\alpha = 0.92$, and $P = 10^4$. The mean of the distribution is 12.

For ordinary flows of size less than or equal to 200, the average transfer time increases as the percentage of priority flows increases, for all of the differentiated service policies. For LAS-fixed $(k = 100)$, the increase is most pronounced, since all packets from priority flows will have priority over packets from ordinary flows that have attained a service of at least 100. On the other hand, the increase is modest for the LAS-linear($k = 5$) policy, and moderate for the LAS-log($k = 0.5$) policy if $q < 0.3$.

Figure 13 shows the evolution of the average transmission time for each bin 100 packets in a priority flow of size $10^4$ packets $(J(n))$ for We compute the average transmission time for the $n$th block of 100 packets as:

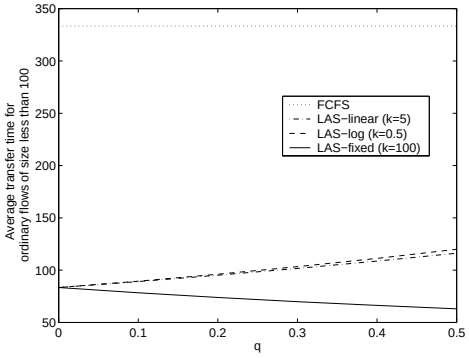
$$J(n) \triangleq T_{p,P}(100n) - T_{p,P}(100(n-1)) \qquad (23)$$

where $T_{p,P}(j)$ denotes the average transfer time of the first $j$ packets in the priority flow under policy $P$, and $T_{p,P}(0) = 0$. For FCFS and LAS-fixed $(k = 100)$, the average transmission time per block is constant, as all packets have the same priority. However, the transfer time under LAS-fixed $(k = 100)$ is much lower than for FCFS. For the other LAS-based policies, the priority of the packets decreases with increasing packet number, and therefore the interference due to packets from ordinary flows will increase. As a consequence, the average transfer time for the block increases as the flow progresses. We note that, for LAS-log $(k = 0.5)$, the decrease in priority occurs very slowly and the average interpacket jitter increases very slowly.

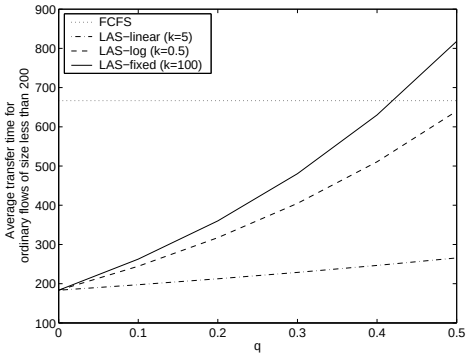
## 5.2 Packet Loss

Figure 14 shows simulation results for the packet loss rate of ordinary and priority TCP flows under LAS-log and LAS-linear as a function of normalized values $k/min(k)$, where $min(k) = 0.5$ for LAS-log and $min(k) = 25$ for LAS-linear.

We observe that priority flows see their loss rate reduced by more than one order of magnitude. For LAS-log the reduction in loss rate for priority flows is higher than for LAS-linear as large flows see their priority decrease much more slowly under LAS-log than




(a) *Flow size* $x \leq 100$




(b) *Flow size* $x \leq 200$

**Figure 12:** ***Average transfer time of ordinary flows as a function of priority jobs*** $q$***; load*** $\rho = 0.7$**.**

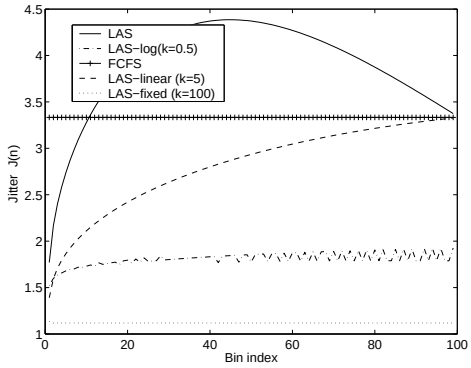



**Figure 13:** ***Average transmission time*** $J(n)$ ***for blocks of 100 packets.***

LAS-linear (c.f. Figure 1). Increasing the $k$ helps for both policies to further reduce the loss rate of priority flows

Figure 15 shows for LAS-fixed(k) the packet loss rate for priority and ordinary Web TCP and UDP constant bit rate flows. As already observed for LAS-log and LAS-linear, the loss rate of priority flows is reduced by more than one order of magnitude. While the loss rate for priority TCP flows is fairly insensitive to the choice of $k$, this is not the case for the UDP flow. This difference is due to the underlying transmission protocol: When experiencing packet loss, a TCP sources reduces its sending rates. The UDP source that transmits packets at a constant bit rate and does not adapt its rate. As the value of $k$ is increased the priority UDP flow will see more

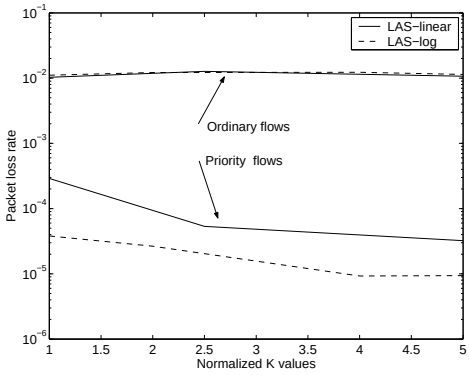



**Figure 14:** ***Packet loss rate as a function of normalized*** $k$ ***values for LAS-linear and LAS-log policies; load*** $\rho = 0.73$**,** $q = 0.1$**.**

interference due to the packets from ordinary flows.

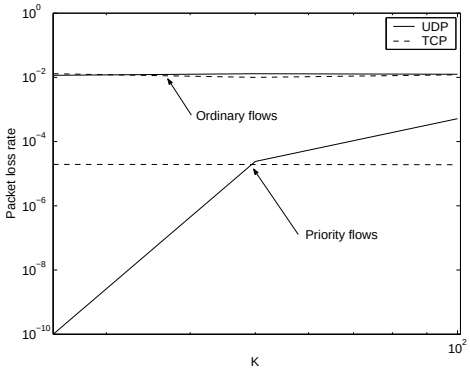



**Figure 15:** ***LAS-fixed(k): Packet loss rate as a function of normalized*** $k$ ***values for UDP and TCP flows ; load*** $\rho = 0.73$**,** $q = 0.1$**.**

## 5.3 Jitter in a Priority UDP Flow

We use simulation to study the jitter experienced by a long-lived UDP under plain LAS and under differentiated LAS scheduling, where the UDP flow is treated as priority flow. The UDP flow sends packets from source to sink at a constant rate of 128 Kbps until the simulation terminates.

Figure 16 shows the average jitter of received packets, which are grouped in non-overlapping windows of 100 sent packets each. We observe that the LAS scheduling has the highest average jitter among the three policies and LAS-log has the lowest average jitter. This is to be expected since for $k = 2$, the priority of the packet at position 64000 in the UDP flow under the LAS-log policy is $(log_2(64000))^{0.5} \approx 4$. The priority value of the UDP flow is constant at 25 for the LAS-fixed(k). Under LAS, simulation results (omitted) show that the priority UDP flow experiences temporary loss rates up to 10% or more, while the flow does not experience any appreciable packet loss under either LAS-log and LAS-fixed(k).

# 6. CONCLUSION AND OPEN ISSUES

LAS scheduling for jobs has been known for over thirty years. This paper is the first to study how variants of LAS scheduling in edge routers can improve the performance of TCP and UDP flows.

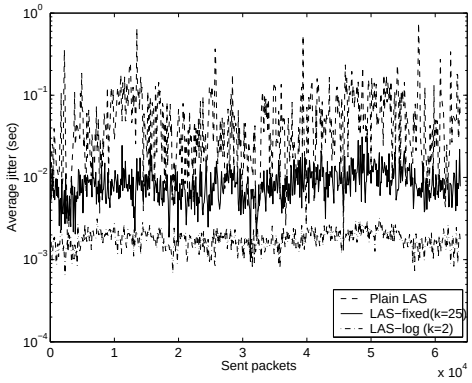



**Figure 16:** ***Average jitter of UDP flow for non-overlapping windows of 100 sent packets, load*** $\rho = 0.7$

While LAS significantly reduces the mean transmission time and loss rate of short flows, it also increases the loss rate and average packet transmission time of the later packets in long flows. To reduce this performance penalty, we have developed several alternative differentiated LAS scheduling scheduling policies that provide service differentiation for flows that require improved performance. For these policies, we have developed analytical models that estimate the mean flow transfer time as accurately as an ns2 simulation, for the dumbbell topology with loss rate below 2%. Advantages of the accurate analytic models include their efficiency for policy performance comparisons over a wide range of system configurations and workloads, the ease of obtaining measures such as the average packet transmission time as a function of position in the flow, and the insights into which system features have principal impact on observed performance.

Among the alternative differentiated LAS scheduling policies, the LAS-log policy significantly improves the performance of priority flows while having a small impact on the performance of the ordinary flows.

Future work includes extending the analytic models to estimate loss rate for a given workload specification, and to accurately estimate the average flow transfer time for moderate loss rates.

Another topic for future work is to explore *where* in the network to deploy LAS. The deployment of a new scheduling policy in routers faces many obstacles, and wide deployment may be an elusive goal. Instead, a limited deployment of LAS in routers at well identified bottlenecks has the potential to reap most of the benefit. Such bottlenecks are often the access links, edge links, or at the transition from the wired to the wireless Internet.

## 7. REFERENCES


[1] N. Bansal and M. Harchol-Balter, "Analysis of SRPT Scheduling: Investigating Unfairness", In *Sigmetrics 2001 / Performance 2001*, pp. 279–290, June 2001.

[2] R. Bhagwan and B. Lin, "Fast and Scalable Priority Queue Architecture for High-Speed Network Switches", In *INFOCOM 2000*, pp. 538–547, 2000.

[3] X. Chen and J. Heidemann, "Preferential Treatment for Short Flows to Reduce Web Latency", *Computer Networks: International Journal of Computer and Telecommunication Networking*, 41(6):779–794, 2003.

[4] M. G. Claffy, K. and K. Thompson, "The nature of the beast: Recent traffic measurements from an Internet backbone", In *Proceedings of INET '98, July 1998*, July 1998.

[5] E. G. Coffman and P. J. Denning, *Operating Systems Theory*, Prentice-Hall Inc., 1973.

[6] M. E. Crovella et al., *A Practical Guide to Heavy Tails*, chapter 3, Chapman and Hall, New-York, 1998.

[7] A. Feldmann, A. Gilbert, P. Huang, and W. Willinger, "Dynamics of IP traffic: A study of the role of variability and the impact of control", In *Proc. of the ACM SIGCOMM*, August 1999.

[8] H. Feng and M. Misra, "Mixed Scheduling Disciplines for Network Flows", In *The Fifth Workshop of Mathematical Performance Modeling and Analysis (MAMA 2003), San Diego, California, USA*, 2003.

[9] E. Friedman and S. G. Henderson, "Fairness and efficiency in Web Servers", In *Proc. ACM SIGMETRICS*, pp. 229–237, June 2003.

[10] L. Guo and I. Matta, "Scheduling Flows with Unknown Sizes: Approximate Analysis", In *Proc. ACM SIGMETRICS*, pp. 276–277, June 2002.

[11] M. Harchol-Balter et al., "Implementation of SRPT Scheduling in Web Servers", In *IPDS 2001*, pp. 1–24, 2001.

[12] M. Harchol-Balter, K. Sigman, and A. Wierman, "Asymptotic Convergence of Scheduling Policies with respect to Slowdown", *Performance Evaluation*, 49:241–256, September 2002.

[13] M. Harchol-Balter, "The Effect of Heavy-Tailed Job Size. Distributions on Computer System Design", In *Proc. of ASA-IMS Conf. on Applications of Heavy Tailed Distributions in Economics*, June 1999.

[14] http://www.isi.edu/nsnam/ns/, "The Network Simulator ns2",

[15] L. Kleinrock, *Queuing Systems, Volume II: Computer Applications*, Wiley, New York, 1976.

[16] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP Throughput: A Simple Model and its Empirical Validation", In *Proc. of ACM SIGCOMM'98*, pp. 303–314, Vancouver, Canada, August 1998.

[17] V. N. Padmanabhan, L. Qiu, and H. J. Wang, "Server-based Inference of Internet Link Lossiness", In *IEEE INFOCOM 2003*, 2003.

[18] V. Paxon and S. Floyd, "Wide-Area Traffic: The failure of Poisson Modelling", *IEEE/ACM Transactions on Networking*, 3:226–244, June 1995.

[19] I. A. Rai, E. W. Biersack, and G. Urvoy-Keller, "Analyzing the Performance of TCP Flows in Packet Networks with LAS Schedulers", RR-03.075, April 2003.

[20] I. A. Rai, G. Urvoy-Keller, and E. W. Biersack, "Analysis of LAS Scheduling for Job Size Distributions with High Variance", In *Proc. ACM SIGMETRICS*, pp. 218–228, June 2003.

[21] H. D. Tan, D. L. Eager, M. K. Vernon, and H. Guo, "Quality of service evaluations of multicast streaming protocols", In *Proc. ACM SIGMETRICS*, pp. 183–194, June 2002.

[22] A. Wierman and M. Harchol-Balter, "Classifying Scheduling Policies with Respect to Unfairness in an M/G/1", In *Proc. ACM SIGMETRICS*, pp. 238–249, June 2003.

[23] S. Yang and G. Veciana, "Size-based Adaptive bandwidth Allocation: Optimizing the Average QoS for Elastic Flows", In *INFOCOM*, 2002.