Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT®

Journal of Algorithms 46 (2003) 79–95

Journal of Algorithms

www.elsevier.com/locate/jalgor

# Loss-bounded analysis for differentiated services ☆

Alexander Kesselman * and Yishay Mansour

*School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel*

Received 25 January 2002

**Abstract**

We consider a network providing Differentiated Services (DiffServ) which allow network service providers to offer different levels of Quality of Service (QoS) to different traffic streams. We focus on loss and first show that only trivial bounds could be obtained by means of traditional competitive analysis. Then we introduce a new approach for estimating loss of an online policy called *loss-bounded* analysis. In loss-bounded analysis the loss of an online policy is bounded by the loss of an optimal offline policy plus a constant fraction of the benefit of the online policy. We relate the loss-bounded analysis to the throughput-competitive analysis. We derive tight upper and lower bounds for various settings of DiffServ parameters using the new loss-bounded model. We believe that loss-bounded analysis is an important technique that can complement traditional competitive analysis providing new insight and interesting results.


*Keywords:* Competitive analysis; Online algorithms; Loss-bounded analysis; Buffer scheduling; DiffServ

## 1. Introduction

The prevalent today's Internet service model is the best-effort model (also known as the so-called send and pray model). This model does not permit users to obtain better service, no matter how critical their requirements are, and no matter how much they may be willing to pay for better service. Clearly, with the increased use of the Internet for commercial purposes, such a model is not satisfactory any more.

In IP-based networks providing any form of streams differentiation requires the network to keep some per stream state information. This translates to increased memory

☆ A preliminary version of this work appeared in SODA 2001.

* Corresponding author.
*E-mail addresses:* alx@math.tau.ac.il (A. Kesselman), mansour@math.tau.ac.il (Y. Mansour).


PII: S0196-6774(02)00270-5

requirements and processing power. The majority of the installed routers use architectures that will experience a dramatic decrease in performance if mechanisms to provide sophisticated flow differentiating features are added. Thus, with existing infrastructure, implementing any complicated functionality inside the network is either impossible or would be prohibitively expensive.

Differentiated Services were proposed as a compromise solution for the Quality of Service (QoS) problem in Internet networks. Differentiated Services assigns each packet a predetermined QoS and aggregates traffic to a small number of classes [2]. Each class is forwarded using the same per-hop behavior at the routers, thereby simplifying the processing and storage requirements. In addition, there is no overhead of signaling, other than the class type in the header of each packet. This is in contrast with ATM, a network designed to support QoS, where a special setup procedure is required in order to establish the QoS guarantees.

From the router perspective, the tools for providing Differentiated Services are based on the following operations that should be done at high speeds: packet classification, buffer management and packet scheduling. In this work we investigate the second aspect.

Over the past few years Differentiated Services has attracted a great deal of research interest in the networking community [5,12,15,17]. Two basic paradigms were proposed: the “Premium” service [11] and the “Assured” service [4]. The Premium service model provides to the user the same QoS guarantee as a dedicated line with a predefined bit rate. A Premium service traffic flow is shaped at the entry to the network and hard-limited to its provisioned peak rate. On the other hand, an Assured service traffic flow may exceed its provisioned rate, but the excess traffic is not given the same assurance level (Assured service may be viewed as “pay more—get more”). One can relate a high QoS class packets to the “in-profile” packets (packets that agree with the provisioned rate) and a low QoS class to the “out-profile” packets (packets that surpass the provisioned rate). Another interpretation is to classify the traffic to a few different QoS classes.

We abstract the Differentiated Services priority model as follows. Packets of different QoS priority have distinct benefit values starting from the lowest benefit of 1 and up to the highest benefit of $\alpha \geqslant 1$. For example, in the Assured service model we have two levels of priority, $\alpha$ and 1. For more advanced traffic models there may be a need in more than two distinct benefits.

Most today’s Internet routers deploy FIFO buffering policy, i.e., packets are sent in the same order as they arrive. One of the advantages of this policy is its amenability to simple and efficient hardware implementation. In addition, and even more important, FIFO buffering scheme reflects the nature of the network. FIFO order is critical for many applications, for example, multimedia applications where audio/video frames must be played in order. Moreover, the main Internet transport protocol TCP is optimized to receive packets in FIFO order. In TCP, packets arriving out of order lead to significant overhead and even retransmissions, which can result in a drastic drop of the performance. For this reasons the FIFO scheme is the most natural approach for network buffering today.

In our model each link of a network is serviced by a single FIFO queue. A queuing policy is presented with a sequence of packet arrivals and has to serve each packet online, i.e., without knowledge of future packets. It performs two functions: stores and selectively rejects/preempts packets subject to the buffer capacity constraints. When the queue is not empty each time unit the first packet in the FIFO order is sent to the output link. A policy obtains the benefit of packets it delivers. The goal is to maximize the policy's benefit, that is the sum of the benefits of delivered packets, or alternatively minimize the benefit of the packets it drops.

In competitive analysis the online policy is compared with an optimal offline policy, that knows the entire input sequence in advance. An online policy is said to be $c$-throughput-competitive if for any input sequence its benefit constitutes at least $c$-fraction of the benefit of an optimal offline policy. Conversely, an online policy is said to be $c$-loss-competitive if the loss of an optimal offline policy constitute at least $c$-fraction of its loss. Note that a throughput-competitive guarantee does not translate to a loss-competitive guarantee. For example, an online policy might lose a constant fraction of the total benefit of the input sequence while an optimal offline policy may have no loss at all. In such a case the online policy is throughput-competitive but not loss-competitive.

Loss-competitive guarantee is very desirable in communication networks, which are designed to minimize loss. Unfortunately, only trivial bounds can be obtained by means of loss-competitive analysis. To highlight the difference between loss-competitive and throughput-competitive analysis we note that in this paper we demonstrate that the loss-competitive ratio of the simple Greedy Policy is $1/\alpha$ while in [7] it was proved that its throughput-competitive ratio is $1/2$.

Motivated by this, we propose a new model called *loss-bounded* analysis for estimating loss of an online policy. In traditional competitive analysis loss of an online policy is compared directly with the loss of an optimal offline policy. In $c_{lb}$-loss-bounded analysis the loss of an online policy is upper bounded by the loss incurred by an optimal offline policy plus a constant fraction $c_{lb}$ of the benefit of the online policy. We let this fraction $c_{lb}$ be the *loss-bounded ratio* of the online policy. We expect loss-bounded ratio to be a small constant, and in such a case the results are especially interesting. Observe that a trivial loss-bounded ratio of $c_{lb} = \infty$ is achieved by any online policy and the smaller $c_{lb}$ the better the performance of the online policy. An optimal policy has $c_{lb} = 0$.

We demonstrate that the loss-bounded approach is dual to the throughput-competitive analysis, i.e., $c_{lb}$-loss-bounded guarantee translates to $1/(1 + c_{lb})$-throughput-competitive guarantee and, conversely, $c_{tc}$-throughput-competitive guarantee translates to $(1 - c_{tc})/c_{tc}$-loss-bounded guarantee. For this reason, the main difference between the loss-bounded and the throughput-competitive analysis is rather conceptual. Note that there is also a notational difference: 0-throughput-competitive ratio corresponds to $\infty$-loss-bounded ratio.

We are not the first to try and analyze Differentiated Services. Initial works have focused on simple probabilistic traffic models [10,14]. Unfortunately, giving a realistic model for Internet traffic is a major problem by itself. Network arrivals were often modeled as Poisson processes for analytic simplicity, however a number of studies have demonstrated that packet inter-arrivals are not exponentially distributed [13]. Moreover, Internet was shown to exhibit chaotic behavior [18]. This highlights the advantage of competitive analysis [3, 16], where a uniform performance guarantee is provided over all input instances.

Competitive analysis of queuing policies for Differentiated Services focused on throughput-competitiveness. In [1] different non-preemptive policies are analyzed for the two distinct benefit values model. We extend the model of [1] by allowing preemptions and considering multiple benefit values. In [9] preemptive queuing policies for arbitrary benefit values are studied in context of smoothing video streams. They establish an impossibility result showing that no online policy can have a better throughput-competitive ratio than $4/5$ and demonstrate that the greedy policy is at least $1/4$-throughput-competitive. Recently in [7] the greedy policy has been shown to achieve the throughput-competitive ratio of $1/2$. In fact, it has been demonstrated that the competitive ratio of the greedy policy is $(\alpha + 1)/(2\alpha + 1)$. They also considered a new model in which packet delay is bounded and packets may be reordered. In this work, in contrast with [7,9], we concentrate on the loss of a policy. This dramatically changes both the analysis and the results.

Following this research, Hahne et al. [6] extend single buffer analysis to the shared memory switches. They consider preemptive buffer management policies and show that the well-known Longest Queue Drop policy is at least $1/2$-competitive and at most $1/\sqrt{2}$-competitive for the case of fixed size and value packets. They also present a $3/4$ general lower bound on the competitive ratio of any online policy. Non-preemptive switch management policies are studied in [8], where they propose a new scheduling policy called Harmonic whose throughput competitive ratio is almost optimal.

The rest of the paper is organized as follows. Section 2 contains the summary of our results. In Section 3 we formally define our model. Description of queuing policies appears in Section 4. Loss-bounded analysis of two benefit values is presented in Section 5. Section 6 contains loss-bounded analysis of restricted benefit values and impossibility result for general benefit values. Traditional loss-competitive analysis appears in Section 7. In Section 8 we compare the loss-bounded analysis with the traditional competitive analysis. Our concluding remarks appear in Section 9.

## 2. Summary of results

In this section we give a brief overview of our main results while the formal definitions and proofs are deferred to the following sections. We present upper and lower bounds for various benefit setting models with regards to the traditional loss-competitive and the new loss-bounded models. We note that all the obtained upper bounds are almost tight. We analyze the Greedy Policy that always accepts to the buffer high benefit packets and the $\beta$-Preemptive Greedy Policy that does exactly the same, but when it accepts a high benefit packet may additionally preempt low benefit packets whose total value is bounded by $1/\beta$ times the value of the accepted high benefit packet. We also study so-called Combined Policy which simulates either the Greedy Policy or the $\beta$-Preemptive Greedy Policy, whichever gives the best competitive ratio, and derive a loss-bounded ratio which is independent on $\alpha$.

The first set of results appearing in Table 1 deals with the two benefit values model, where we have packets with either a high benefit of $\alpha$ or a low benefit of 1. We show that the Greedy Policy achieves a non-interesting $1/\alpha$ loss-competitive ratio, which turns out to be the tight upper bound. The loss-bounded ratio of the Greedy Policy follows from its throughput-competitive ratio of $(\alpha + 1)/(2\alpha + 1)$ established in [7] by applying Theorem 10. We prove that the loss-bounded ratio of the $\sqrt{\alpha}$-Preemptive Greedy Policy is $1/\sqrt{\alpha}$, which is approximately tight. Observe that the provided guarantee is much

Table 1
Two benefit values setting results

| Policy/model | Loss-competitive | Loss-bounded |
|---|---|---|
| Greedy | $1/\alpha$ | $\frac{\alpha}{\alpha+1}$ |
| $\sqrt{\alpha}$-preemptive greedy | 0 | $1/\sqrt{\alpha}$ |
| Combined | 0 | 0.68 |
| Impossibility results | $1/\alpha$ | $1/\sqrt{\alpha} - \frac{3}{\alpha+1}$ |

Table 2
Restricted benefit setting results

| Policy/model | Loss-competitive | Loss-bounded |
|---|---|---|
| Greedy | $1/\alpha$ | $\frac{\alpha}{\alpha+1}$ |
| $\sqrt{\alpha^{1/n}}$-preemptive greedy | 0 | $\frac{\sqrt{\alpha^{1/n}}+1}{\alpha^{1/n}-\sqrt{\alpha^{1/n}}-1} + \frac{\sqrt{\alpha^{1/n}}}{\alpha^{1/n}-1}$ |
| Impossibility results | $1/\alpha$ | $1/\sqrt{\alpha^{1/n}} - \frac{3}{\alpha^{1/n}+1}$ |

Table 3
Arbitrary benefit setting results

| Policy/model | Loss-competitive | Loss-bounded |
|---|---|---|
| Greedy | $1/\alpha$ | $\frac{\alpha}{\alpha+1}$ |
| Impossibility results | $1/\alpha$ | 0.2 |

stronger than that of the Greedy Policy. The throughput-competitive guarantee of the $\sqrt{\alpha}$-Preemptive Greedy Policy approaches 1 when $\alpha$ is large.

Next we extend the two benefit model and study the case of $n+1$ different benefit values $\{\alpha^{i/n}\colon 0 \leqslant i \leqslant n\}$. Table 2 summarizes these results. One can see that as the number of values increases, the guarantee is weaker. Both the competitive ratio of the $\sqrt{\alpha^{1/n}}$-preemptive greedy and the impossibility results are of the order of $1/\sqrt{\alpha^{1/n}}$, for large values of $\alpha^{1/n}$.

The last set of results that corresponds to arbitrary benefit values between 1 and $\alpha$ appears in Table 3. We prove that no online policy can have less than a constant loss-bounded ratio.

## 3. Model description

We consider a FIFO buffer that can hold $B$ packets. We assume that packets may arrive to the queue at any time and send events are synchronized with time. Each packet $p$ has corresponding *benefit*, $b(p)$. The system obtains the benefit of the packets it sends, and its aim is to maximize the benefit of the transmitted packets.

Now we define the system more formally. We denote the $i$th packet in the FIFO buffer as $F[i]$ and let the index of the last packet in this order be *last* $\leqslant B$. When a packet arrives, a queuing policy can either reject or accept the packet. At any time the policy can also preempt packets that are currently in the buffer. Each time unit a send operation is executed if the buffer is not empty, i.e., $last > 0$, and the first packet in the buffer is sent. In addition, we require that at the start and at the finish times of a schedule the buffer is empty.

**Definition 1.** For a sequence of packets $S$ and an online policy $A$ we denote:

- The subsequence of packets with benefit $b$ by $S_b$.
- The benefit of $A$ on $S$ by $V_A(S)$ and the loss of $A$ on $S$ by $L_A(S)$ (note that $V(S) = V_A(S) + L_A(S)$).

We denote an optimal offline policy by OPT and for an input sequence $S$, $V_{\mathrm{OPT}}(S)$ and $L_{\mathrm{OPT}}(S)$ are the optimal benefit and the optimal loss, respectively. Next we define a new model for analyzing loss of an online policy.

**Definition 2.** We say that a policy $A$ is $c_{lb}$-loss-bounded iff for every sequence of packets $S$, $L_A(S) \leqslant L_{\mathrm{OPT}}(S) + c_{lb} \times V_A(S)$.

For completeness, we also present the traditional definitions of competitiveness.

**Definition 3.** We say that a policy $A$ is $c_{tc}$-throughput-competitive iff for every sequence of packets $S$, $V_A(S) \geqslant c_{tc} \times V_{\mathrm{OPT}}(S)$.

**Definition 4.** We say that a policy $A$ is $c_{lc}$-loss-competitive iff for every sequence of packets $S$, $c_{lc} \times L_A(S) \leqslant L_{\mathrm{OPT}}(S)$.

## 4. Scheduling policies

The basic buffer management operations are defined in the following way.

```
accept(p) { last = last + 1;  p.index = last;  F[last] = p; }
reject(p) { free(p); }
remove(p) {
   for i = p.index to last − 1 do
      F[i] = F[i + 1];
   last = last − 1;
   free(p);
}
preempt(p) { remove(p); }
send() { transmit(F[1]); remove(F[1]); }
```

Now we describe a natural Greedy Policy that always retains in the buffer a set of packets with highest benefit. An arriving packet is *accepted* if either the buffer is not full or the buffer is full and the minimal benefit among the accepted packets in the buffer is less than the benefit of the arriving packet. In the latter case a packet with the minimal benefit is preempted from the buffer before acceptance of the arriving packet. An arriving packet is *rejected* otherwise.

**Greedy Policy**

*last* $= 0$;

*GreedyPacketArrivalHandler*$(p)$ {

  **if** $(last < B)$ **then**

    *accept*$(p)$;

  **else**

    $min = F[1]$; /* get a packet with the minimal benefit */

    **for** $i = 2$ to *last* **do**

      **if** $(F[i].benefit < min.benefit)$ **then** $min = F[i]$;

    **if** $(min.benefit < p.benefit)$ **then**

      *preempt*$(min)$;

      *accept*$(p)$;

    **else**

      *reject*$(p)$;

}

Next we introduce the $\beta$-Preemptive Greedy Policy that behaves like the Greedy Policy except that upon acceptance of a packet, additional packets may be preempted. The preempted packets are the low benefit packets *closest to the transmitting end of the FIFO* whose total benefit is bounded by $1/\beta$ times the benefit of the accepted packet. Note that these are *additional* preemptions.

The intuition behind the additional preemptions is as follows. Consider the following scenario. At the beginning the buffer is full of low benefit packets. For the next $B$ time units a single high benefit packet arrives. If no low benefit packet is preempted, after $B$ time units the buffer is full of high benefit packets. Assume now that a burst of $B$ high benefit packets arrives. In this case they are all lost. The $\beta$-Preemptive Greedy Policy solves this problem by preempting $\alpha/\beta$ low benefit packets upon arrival of each packet of benefit $\alpha$. For sufficiently large value of $\beta$, the $\beta$-Preemptive Greedy Policy performs like the Greedy Policy, which was shown in [7] to be 1/2-throughput-competitive. However, small values of $\beta$ may cause excessive preemptions and large loss of low benefit packets. Thus, there is a need to optimize the value of $\beta$ in order to achieve a balance between maximizing of the current throughput and minimizing of potential future loss.

**$\beta$-Preemptive Greedy Policy**

*last* $= 0$;

*PreemptiveGreedyPacketArrivalHandler*$(p)$ {

  /∗ apply the Greedy packet arrival handling routine ∗/

  *GreedyPacketArrivalHandler*$(p)$;

  /∗ check whether $p$ was accepted ∗/

  **if** $(F[last] == p)$ **then**

    $val = p.benefit/\beta$; /∗ calculate how much to preempt ∗/

    **for** $i = 1$ to $last - 1$ **do**

      **if** $(F[i].benefit \leqslant val)$ **then**

        $val = val - F[i].benefit$;

        $preempt(F[i])$;

}

Finally, we introduce the Combined Policy which simulates the policy with the best competitive ratio between the Greedy Policy and the $\beta$-Preemptive Greedy Policy.

**Combined Policy with input parameter $\alpha$**

If $\alpha/(\alpha + 1) > 1/\sqrt{\alpha}$ then simulate the Greedy Policy. Otherwise, simulate the $\sqrt{\alpha}$-Preemptive Greedy Policy.

## 5. Two benefit values

In this section we consider packets having two possible benefit values, that is a high benefit of $\alpha \geqslant 1$ or a low benefit of 1. Next we prove that the $\sqrt{\alpha}$-Preemptive Greedy Policy achieves a $1/\sqrt{\alpha}$ loss-bounded ratio, which is nearly the best competitive ratio possible for an online policy. First we need some auxiliary lemmas.

**Lemma 1.** *When packets are scheduled according to the $\sqrt{\alpha}$-Preemptive Greedy Policy, and there are at least $B/\sqrt{\alpha}$ high benefit packets in the buffer then at the next time step the policy schedules a high benefit packet.*

**Proof.** Note that the number of low benefit packets initially preceding to the first high benefit packet in the FIFO order is at most $B$. Each high benefit packet preempts $\sqrt{\alpha}$ low benefit packets from the beginning of the FIFO order, if any, and there are at least $B/\sqrt{\alpha}$ high benefit packets. Thus, the first packet in the FIFO order should be a high benefit packet. □

**Definition 5.** A scheduling interval $[t_s, t_f]$ is *overloaded interval* if:

(1) Some high benefit packet was rejected during this interval and only high benefit packets were scheduled between $t_s$ and $t_f$.
(2) At time $t_s - 1$ either a low benefit packet was scheduled or the buffer was empty and at time $t_f + 1$ either a low benefit packet was scheduled or the buffer was empty.
(3) The interval is maximal, i.e., it is not contained in any other overloaded interval.

**Definition 6.** A scheduling interval $[t_s, t_f]$ is *active interval* if:

(1) At each time slot between $t_s$ and $t_f$ some packet was scheduled.
(2) The interval is maximal.

**Definition 7.** A packet *belongs* to a scheduling interval iff it was either scheduled or lost during this interval.

**Claim 1.** *When packets are scheduled according to the $\sqrt{\alpha}$-Preemptive Greedy Policy the number of high benefit packets in the buffer at the time unit preceding the beginning of an overloaded interval $[t_s, t_f]$ is at most $B/\sqrt{\alpha}$.*

**Proof.** Either at time $t_s - 1$ a low benefit packet was scheduled and the claim follows by Lemma 1, or the buffer was empty and the claim holds trivially. □

**Claim 2.** *When packets are scheduled according to the $\sqrt{\alpha}$-Preemptive Greedy Policy the length of an overloaded interval is at least $B$.*

**Proof.** When a high benefit packet is lost, the buffer is full of high benefit packets. Since high benefit packets are never preempted it takes at least $B$ time units to schedule all of them. This yields the claim. □

In the following lemma we bound from below the number of low benefit packets dropped by the optimal policy.

**Lemma 2.** *For any input sequence, the optimal policy rejects at least the same number of low benefit packets as the number of low benefit packets rejected by the $\sqrt{\alpha}$-Preemptive Greedy Policy plus the number of low benefit packets preempted by some of the first $B$ high benefit packets scheduled during an overloaded interval.*

**Proof.** The proof is by way of contradiction. Assume that the statement of the lemma does not hold for a particular schedule. Let us consider the active intervals of the $\sqrt{\alpha}$-Preemptive Greedy Policy. It must be the case that the statement is violated for at least one interval. Otherwise, it is fulfilled for the whole schedule.

Let $\mathcal{P} = [t_s, t_f]$ be such an interval, $S$ be the set of packets that belong to $\mathcal{P}$ and $t$ be the last time at which a low benefit packet was either rejected or preempted by some of the first $B$ high benefit packets scheduled during an overloaded interval. Observe that there is a time moment $t' \geqslant t$ at which the buffer of the $\sqrt{\alpha}$-Preemptive Greedy Policy is full. We consider the set of packets $S'$ ($S' \subseteq S$) that belong to time interval $[t_s, t']$. Notice that any policy can schedule at most $t' - t_s + B$ packets from $S'$. The $\sqrt{\alpha}$-Preemptive Greedy Policy up to time $t'$ had scheduled $t' - t_s$ and buffered $B$ packets from $S'$. Since the optimal policy sends at least the same number of high benefit packets from $S'$ as the $\sqrt{\alpha}$-Preemptive Greedy Policy does and it cannot accept more than $t' - t_s + B$ packets from $S'$, it should have rejected at least the same number of low benefit packets as the $\sqrt{\alpha}$-Preemptive Greedy Policy did prior to time $t'$. That contradicts to our assumption. □

At this point we are able to prove the main theorem.

**Theorem 1.** *The loss-bounded ratio of the $\sqrt{\alpha}$-Preemptive Greedy Policy is at most $1/\sqrt{\alpha}$.*

**Proof.** We process loss of low and high benefit packets separately. We denote by $S_\alpha^o$ the union of the first $B$ packets scheduled during all overloaded intervals. The rest of the high value packets scheduled is denoted by $S_\alpha^u$. We show that

$$L_A(S_1) \leqslant \frac{1}{\sqrt{\alpha}} V_A\big(S_\alpha^u\big) + L_{\mathrm{OPT}}(S_1) \quad \text{and} \quad L_A(S_\alpha) \leqslant \frac{1}{\sqrt{\alpha}} V_A\big(S_\alpha^o\big) + L_{\mathrm{OPT}}(S_\alpha),$$

establishing the theorem, since $S_\alpha = S_\alpha^u \cup S_\alpha^o$ and $L_A(S) = L_A(S_1) + L_A(S_\alpha)$.

First we bound the loss of low benefit packets. There are two cases of loss. The first one is due to additional preemptions (preemption of a low benefit packet by the policy when a high benefit packet arrives and the buffer is full) and the second one is due to buffer overflow (low benefit packets that are rejected when they arrive). We denote the former case by $L_A^{\mathrm{Extra}}$ and the loss of the latter case by $L_A^{\mathrm{Rej}}$. The loss of the first kind we further divide to the loss of low benefit packets preempted by a high benefit packet scheduled among the first $B$ packets during an overloaded interval, $L_A^{\mathrm{Ovfl}}$, and the rest, $L_A^{\mathrm{Prm}}$. By Lemma 2,

$$L_A^{\mathrm{Rej}}(S_1) + L_A^{\mathrm{Ovfl}}(S_1) \leqslant L_{\mathrm{OPT}}(S_1).$$

Now let us bound the value of $L_A^{\mathrm{Prm}}$. Since each high benefit packet can preempt low benefit packets with cumulative benefit at most $\sqrt{\alpha}$ and high benefit packets themselves are never preempted, we can charge the high benefit packet for the preempted low benefit packets, obtaining that

$$L_A^{\mathrm{Prm}}(S_1) \leqslant \frac{1}{\sqrt{\alpha}} V_A\big(S_\alpha^u\big).$$

By adding up both of these inequalities we obtain:

$$L_A(S_1) \leqslant \frac{1}{\sqrt{\alpha}} V_A\big(S_\alpha^u\big) + L_{\mathrm{OPT}}(S_1).$$

A slightly more complicated task is to bound the loss of high benefit packets. We divide the schedule into underloaded and overloaded intervals. No high benefit packet is lost during underloaded interval. According to Claim 1 at the very beginning of an overloaded interval there are at most $B/\sqrt{\alpha}$ high benefit packets in the buffer. All the high benefit packets that were lost during the interval arrived throughout this interval. Clearly, an optimal offline policy could have sent additionally at most $B/\sqrt{\alpha}$ high benefit packets since the $\sqrt{\alpha}$-Preemptive Greedy Policy throughout an overloaded interval behaves like the Greedy Policy with respect to high benefit packets.

Thus, the loss of high benefit packets by the $\sqrt{\alpha}$-Preemptive Greedy Policy is bounded by the loss of an optimal offline policy plus at most extra $B/\sqrt{\alpha}$ packets per each overloaded interval. We denote the loss of the first kind by $L_A^{\mathrm{Ncs}}$ and the loss of the second kind by $L_A^{\mathrm{Add}}$. By definition,

$$L_A^{\mathrm{Ncs}}(S_\alpha) \leqslant L_{\mathrm{OPT}}(S_\alpha).$$

To bound the loss of $L_A^{\mathrm{Add}}$ note that, by Claim 2, the length of an overloaded interval is at least $B$. Hence, the ratio between the loss and the cumulative benefit of the first $B$ packets scheduled during an overloaded interval is at most $B\sqrt{\alpha}/B\alpha = 1/\sqrt{\alpha}$. Therefore, the additional loss of high benefit packets is bounded by

$$L_A^{\mathrm{Add}}(S_\alpha) \leqslant \frac{1}{\sqrt{\alpha}} V_A\bigl(S_\alpha^o\bigr) \quad \text{and} \quad L_A(S_\alpha) \leqslant \frac{1}{\sqrt{\alpha}} V_A\bigl(S_\alpha^o\bigr) + L_{\mathrm{OPT}}(S_\alpha),$$

which completes the proof. □

Now we consider the Combined Policy and show that its loss-bounded ratio is at most 0.68.

**Theorem 2.** *The loss-bounded ratio of the Combined Policy is at most* 0.68.

**Proof.** The loss-bonded ratio of the Combined Policy is the minimum of the loss-bonded ratio of the Greedy Policy and the loss-bonded ratio of the $\sqrt{\alpha}$-Preemptive Greedy Policy, that is $\min(\alpha/(\alpha+1), 1/\sqrt{\alpha})$. To optimize the lower bound, we take the worst-case value of $\alpha$ that equates the two ratios. Using numerical solution we find that the loss-bonded ratio of the Combined Policy, is maximized when $\alpha \approx 2.15$. Substituting this value of $\alpha$, we obtain that the loss-bounded ratio of the Combined Policy is at most 0.68. □

In the next theorem we prove a lower bound on loss-bounded ratio showing that the bound of Theorem 1 is approximately tight.

**Theorem 3.** *The loss-bounded ratio of any online policy is at least* $1/\sqrt{\alpha} - 3/(\alpha+1)$ *for* $B \geqslant \sqrt{\alpha}$.

**Proof.** Suppose that packets are scheduled according to the online policy $A$. We construct a bad sequence of packets similar to the one used in [9]. At time $t_s = 0$ the buffer is empty and $B-1$ low benefit packets arrive. Each following time unit a high benefit packet arrives. Let $t_f$ be the first time when the buffer contains no low benefit packets and let $k = t_f - t_s$. Note that the number of the scheduled low benefit packets is $k$ and the number of the lost low benefit packets is $B-k-1$ respectively. Now there may be two cases.

(1) In case $0 \leqslant k < B/\sqrt{\alpha}$, the input sequence is terminated. The loss-bounded ratio of the online policy $c_{lb}$ must fulfill the following:

$$B - k - 1 \leqslant c_{lb} \cdot k(\alpha+1).$$

Therefore, $c_{lb} > (B-k-1)/(k(\alpha+1))$. This ratio is minimized when $k$ is maximal, thus substituting the upper bound for $k$ we get

$$c_{lb} > \frac{B - B/\sqrt{\alpha} - 1}{B\sqrt{\alpha} + B/\sqrt{\alpha}} \geqslant \frac{1}{\sqrt{\alpha}} - \frac{1/\sqrt{\alpha} + 1/B + 1/\alpha}{\sqrt{\alpha} + 1/\sqrt{\alpha}} \geqslant \frac{1}{\sqrt{\alpha}} - \frac{3}{\alpha+1}, \tag{1}$$

which holds for $B \geqslant \sqrt{\alpha}$.

(2) In case $B/\sqrt{\alpha} \leqslant k < B$, at time $t_f$ a burst of $B$ high benefit packets arrives. Note that at least $k$ of them are lost by the online policy. The loss-bounded ratio $c_{lb}$ of the online policy over the whole sequence should satisfy:

$$B - k - 1 + k\alpha \leqslant B - 1 + c_{lb}(k + B\alpha),$$

since an optimal offline policy would have lost $B-1$ low benefit packets. Thus,

$$k(\alpha-1) \leqslant c_{lb}(k + B\alpha), \qquad c_{lb} \geqslant \frac{k(\alpha-1)}{k + B\alpha}.$$

After substituting the lower bound for $k$ to the numerator and the upper bound for $k$ to the denominator of the fraction, we get

$$c_{lb} \geqslant \frac{\frac{B}{\sqrt{\alpha}}(\alpha-1)}{B + B\alpha} > \frac{1 - 1/\alpha}{1 + \sqrt{\alpha}} > \frac{1}{\sqrt{\alpha}} - \frac{2}{\alpha+1}, \tag{2}$$

and the theorem follows. □

## 6. Restricted and general benefit values

In this section we first consider packets that have a restricted set of $n+1$ benefit values, that is $\{\alpha^{i/n}\colon\ 0 \leqslant i \leqslant n\}$. We obtain results that are generalization of the corresponding results for the two benefit values model. In particular, we prove that the loss-bounded ratio of the $\sqrt{\alpha^{1/n}}$-Preemptive Greedy Policy is $(\sqrt{\alpha^{1/n}}+1)/(\alpha^{1/n} - \sqrt{\alpha^{1/n}} - 1) + \sqrt{\alpha^{1/n}}/(\alpha^{1/n} - 1)$ and that this bound is almost the best achievable by any online policy. Then we present a general lower bound for loss-bounded analysis.

**Lemma 3.** *When packets are scheduled according to the $\sqrt{\alpha^{1/n}}$-Preemptive Greedy Policy, and there are at least $B/\sqrt{\alpha^{1/n}}$ packets of benefit greater than $\alpha^{(i-1)/n}$ in the buffer at the next time step the policy schedules a packet with benefit of at least $\alpha^{i/n}$.*

The proof of Lemma 3 is identical to that of Lemma 1.

**Definition 8.** A scheduling interval $[t_s, t_f]$ is *$i$-overloaded* if:

(1) All the following holds regarding the interval:
   (a) Some packet of benefit $\alpha^{i/n}$ was rejected during this interval, and
   (b) no packet of benefit greater than $\alpha^{i/n}$ was rejected during this interval, and
   (c) only packets with benefit at least $\alpha^{i/n}$ were scheduled between $t_s$ and $t_f$.

(2) One of the following holds at time $t_s - 1$:
   (a) A packet with benefit less than $\alpha^{i/n}$ was scheduled, or
   (b) it is the finish time of a $j$-overloaded interval with $j > i$, or
   (c) the buffer was empty.
(3) One of the following holds at time $t_f + 1$:
   (a) a packet with benefit less than $\alpha^{i/n}$ was scheduled, or
   (b) it is the start time of a $j$-overloaded interval with $j > i$, or
   (c) the buffer was empty.
(4) The interval is maximal, i.e., it is not contained in any other $i$-overloaded interval.

**Claim 3.** *When packets are scheduled according to the $\sqrt{\alpha^{1/n}}$-Preemptive Greedy Policy the last interval in a sequence of consecutive overloaded intervals with increasing index has length at least $B$ time units.*

**Proof.** Suppose that the last interval in the sequence is $i$-overloaded. An $i$-overloaded interval may be interrupted either by losing of a packet with benefit greater than $\alpha^{i/n}$ or by scheduling a packet of benefit less than $\alpha^{i/n}$. Since the overloading index of the next interval is less than $i$ it must be the case that transmitting of a lower benefit packet took place.

By the definition of $i$-overloaded interval some packet of benefit $\alpha^{i/n}$ was rejected during this interval. At this time moment the buffer was full of packets with benefit not less than $\alpha^{i/n}$ and it requires at least $B$ time units to schedule all these packets to completion. □

Next we prove the loss-bounded ratio of the $\sqrt{\alpha^{1/n}}$-Preemptive Greedy Policy.

**Theorem 4.** *The loss-bounded ratio of the $\sqrt{\alpha^{1/n}}$-Preemptive Greedy Policy is at most* $(\sqrt{\alpha^{1/n}}+1)/(\alpha^{1/n}-\sqrt{\alpha^{1/n}}-1)+\sqrt{\alpha^{1/n}}/(\alpha^{1/n}-1)$.

**Proof.** The packets may be lost either due to preemptions or due to rejections. We denote the loss of the first case by $L_A^{\mathrm{Prm}}$ and the loss of the second case by $L_A^{\mathrm{Rej}}$. We show that

$$L_A^{\mathrm{Prm}}(S) \leqslant \frac{\sqrt{\alpha^{1/n}}+1}{\alpha^{1/n}-\sqrt{\alpha^{1/n}}-1} V_A(S)$$

and

$$L_A^{\mathrm{Rej}}(S) \leqslant L_{\mathrm{OPT}}(S) + \frac{\sqrt{\alpha^{1/n}}}{\alpha^{1/n}-1} V_A(S),$$

establishing the theorem.

First we bound $L_A^{\mathrm{Prm}}$. An arriving packet $p$ may first preempt a packet of lower benefit in time of buffer overflow, which constitutes at most $1/\alpha^{1/n}$ fraction of the benefit of $p$. In addition $p$'s extra preemptions constitute at most $1/\sqrt{\alpha^{1/n}}$ fraction of its benefit. We should also take in account that each preempted packet might have itself preempted other packets. Notice that such a chain is always finite because the benefit increases very fast and it is bounded by $\alpha$. Moreover, the total benefit of a chain is dominated by that of the last packet, which is necessarily sent to the output link.

We define the *recursive preemptions set* of a packet to be the union of the directly preempted packets and their recursive preemptions sets. We charge the last packet in the preemptions chain of all the lost packets in its recursive preemptions set. Let $y = 1/\sqrt{\alpha^{1/n}} + 1/\alpha^{1/n}$ and assume that $\alpha$ is sufficiently large so that $y < 1$. If the last packet in the chain of preemptions has benefit $\alpha^{i/n}$ then the cumulative benefit of its recursive preemptions set is bounded by the following fraction of its benefit:

$$y + y^2 + \cdots < \frac{y}{1-y} = \frac{\sqrt{\alpha^{1/n}} + 1}{\alpha^{1/n} - \sqrt{\alpha^{1/n}} - 1}.$$

We charge each packet the loss due to its recursive preemptions set obtaining that

$$L_A^{\mathrm{Prm}}(S) \leqslant \frac{\sqrt{\alpha^{1/n}} + 1}{\alpha^{1/n} - \sqrt{\alpha^{1/n}} - 1} V_A(S).$$

It remains to bound $L_A^{\mathrm{Rej}}$. We start by determining overloaded intervals in order of decreasing index. When we are done with a particular $i$ we mark the corresponding intervals and continue with the remained parts of the schedule.

Notice that by definition of $i$-overloaded interval no packet with benefit greater than $\alpha^{i/n}$ is lost during such an interval. Thus, we have to concentrate on the loss of packets with benefit less than or equal to $\alpha^{i/n}$.

Let $\mathcal{I} = [t_s, t_f]$ be an $i$-overloaded interval and let $t_l$ be the last time moment before $t_s$ at which a packet with benefit less than $\alpha^{i/n}$ was scheduled. If there is no such $t_l$ then an optimal offline policy would have also lost packets belonging to the interval with cumulative benefit that is greater than or equal to the benefit of the interval's lost set. So assume that such time exists.

According to Lemma 3 at time $t_l$ there are at most $B/\sqrt{\alpha^{1/n}}$ packets of benefit $\alpha^{i/n}$ in the buffer. Hence, in the worse case an optimal offline policy could have scheduled additionally at most $B/\sqrt{\alpha^{1/n}}$ of the lost packets with benefit at most $\alpha^{i/n}$. There may be two cases regarding the length of the interval, that is $l(\mathcal{I}) = t_s - t_f + 1$.

(1) In case $l(\mathcal{I}) \geqslant B$, the ratio between the additional loss and the cumulative benefit of the packets that were scheduled during $\mathcal{I}$ is bounded by

$$\frac{\alpha^{i/n} B/\sqrt{\alpha^{1/n}}}{B\alpha^{i/n}} = \frac{1}{\sqrt{\alpha^{1/n}}}.$$

(2) In case $l(\mathcal{I}) < B$, according to Claim 3 the last interval in the sequence of consecutive intervals with increasing index has length at least $B$. We charge this interval of loss of $B/\sqrt{\alpha^{1/n}}$ additional packets. Note that the last interval in a sequence maybe charged at most once by each interval in the sequence. The total value charged for a sequence of intervals in which the last interval has index $k$ is bounded by

$$\sum_{i=0}^{k} \frac{B}{\sqrt{\alpha^{1/n}}} \alpha^{i/n} = \frac{\alpha^{(k+1)/n} - 1}{\alpha^{1/n} - 1} \cdot \frac{B}{\sqrt{\alpha^{1/n}}} < \frac{\sqrt{\alpha^{1/n}}}{\alpha^{1/n} - 1} \cdot B\alpha^{k/n}.$$

Therefore, the cumulative loss owing to rejections is bounded from above by

$$L_A^{\mathrm{Rej}}(S) \leqslant L_{\mathrm{OPT}}(S) + \frac{\sqrt{\alpha^{1/n}}}{\alpha^{1/n} - 1} V_A(S),$$

which completes the proof. □

The next theorem presents a lower bound on loss-bounded ratio demonstrating that the bound of Theorem 4 is nearly tight. The proof is similar to that of Theorem 3 and is omitted.

**Theorem 5.** *The loss-bounded ratio of any online policy is at least* $1/\sqrt{\alpha^{1/n}} - 3/(\alpha^{1/n} + 1)$ *for* $B \geqslant \sqrt{\alpha^{1/n}}$.

Next we present a lower bound for loss-bounded analysis in case of general benefit values. The pathological case for general packet values is still obtained by the interaction between two packet values. The following theorem is proved by maximizing the bounds of Eqs. (1) and (2) from Theorem 3 for $\alpha \approx 5$.

**Theorem 6.** *The loss-bounded ratio of any online policy is at least* $0.2 - C/B$*, for some constant* $C \ll B$.

## 7. Traditional loss-competitive analysis

In this section we consider packets having arbitrary benefit values so that for any packet its benefit is between 1 and $\alpha$. We deal with traditional loss-competitive analysis. The next theorem shows that the Greedy Policy accomplishes a non-zero loss-competitive ratio.

**Theorem 7.** *The Greedy Policy is* $1/\alpha$ *loss-competitive.*

**Proof.** Clearly, the Greedy Policy maximizes the number of scheduled packets. This implies that the cumulative benefit of the lost packets is at most a factor of $\alpha$ far from the optimal. □

This loss-competitive ratio is not appealing since as $\alpha$ increases the loss-competitive ratio decreases to zero. The following theorem shows that this ratio is the best that an online policy could achieve.

**Theorem 8.** *The loss-competitive ratio of any online policy is at most* $1/\alpha$.

**Proof.** Consider the following scenario. At time $t = 0$ the buffer is empty and a low benefit packet followed by a high benefit arrive. If the policy drops the packet of low benefit then it is 0-loss-competitive since there exists a feasible schedule of these packets. In case the low benefit packet is not dropped, it is sent at time $t$. Then at time $t + 1$, $B$ high benefit packets arrive. Thus, the online policy necessarily drops one high benefit packet, since an optimal offline policy could have dropped the low benefit packet instead. The theorem follows. □

## 8. Comparative evaluation of loss-bounded analysis

In this section we evaluate the loss-bounded analysis by comparing it with the traditional competitive analysis. We show that the loss-bounded approach is dual to the throughput-competitive analysis. The following theorem demonstrates translation from the loss-bounded to the throughput-competitive guarantee.

**Theorem 9.** *A $c_{lb}$-loss-bounded policy $A_{lb}$ provides $1/(1+c_{lb})$-throughput-competitive guarantee.*

**Proof.** The loss of $A_{lb}$ by definition are bounded by

$$L_{A_{lb}}(S) \leqslant L_{\mathrm{OPT}}(S) + c_{lb} V_{A_{lb}}(S).$$

After substituting in place of loss the difference between the total benefit of the sequence and the benefit gained by the policy for $A_{lb}$ and OPT we get

$$V(S) - V_{A_{lb}}(S) \leqslant V(S) - V_{\mathrm{OPT}}(S) + c_{lb} V_{A_{lb}}(S),$$

$$V_{A_{lb}} \geqslant \frac{1}{1+c_{lb}} V_{\mathrm{OPT}}(S),$$

which yields the claim. □

Now we show the reverse translation. The proof of the following theorem is analogous to that of Theorem 9 and is omitted.

**Theorem 10.** *A $c_{tc}$-throughput-competitive policy $A_{tc}$ provides $(1-c_{tc})/c_{tc}$-loss-bounded guarantee.*

## 9. Concluding remarks

In this work we investigated a framework of Differentiated Services. We have shown importance of analysis of loss of an online policy and obtained impossibility results for traditional competitive analysis. Then we introduced a new model for loss evaluation—loss-bounded analysis in which the loss of an online policy is upper bounded by the loss of an optimal offline policy plus a constant fraction of the benefit of the online policy. For various QoS parameters settings we presented tight lower and upper bounds for FIFO buffer management.

The proposed policies may be used for managing current Internet routers that wish to provide Differentiated Services. Due to their simplicity they may operate at very high speeds. Moreover, their implementation does not require installing additional costly equipment.

By choosing the appropriate benefit setting and value of $\alpha$ the network operator could manage traffic streams in the best way. For instance, to give high priority packets an absolute preference over low priority packets $\alpha$ may be made very large. Conversely, for $\alpha$ near one, we are basically use the "best-effort" approach optimizing the total throughput and ignoring different priorities. It is worth to note that the loss-bounded ratio of our policies is improved with increasing of $\alpha$.

## References


[1] W.A. Aiello, Y. Mansour, S. Rajagopolan, A. Rosen, Competitive queue policies for Differentiated Services, in: Proceedings of INFOCOM, 2000, pp. 431–440.

[2] Y. Bernet, A. Smith, S. Blake, D. Grossman, A conceptual model for DiffServ routers, Internet Draft, (2000).

[3] A. Borodin, R. El-Yaniv, Online Computation and Competitive Analysis, Cambridge Univ. Press, 1998.

[4] D. Clark, J. Wroclawski, An approach to service allocation in the Internet, Internet Draft, (1997).

[5] C. Dovrolis, D. Stiliadis, P. Ramanathan, Proportional Differentiated Services: Delay differentiation and packet scheduling, in: Proc. ACM SIGCOMM'99.

[6] E.L. Hahne, A. Kesselman, Y. Mansour, Competitive buffer management for shared-memory switches, in: Proceedings of SPAA, 2001, pp. 53–58.

[7] A. Kesselman, Z. Lotker, Y. Mansour, B. Patt-Shamir, B. Schieber, M. Sviridenko, Buffer overflow management in QoS switches, in: Proceedings of STOC, 2001, pp. 520–529.

[8] A. Kesselman, Y. Mansour, Harmonic buffer management policy for shared memory switches, in: INFOCOM, 2002, to appear.

[9] Y. Mansour, B. Patt-Shamir, O. Lapid, Optimal smoothing schedules for real-time streams, in: Proccedings of PODC, 2000.

[10] M. May, J. Bolot, A. Jean-Marie, C. Diot, Simple performance models of Differentiated Services schemes for the Internet, in: Proceedings of IEEE INFOCOM, March, 1999.

[11] K. Nichols, V. Jacobson, L. Zhang, A two-bit Differentiated Services architecture for the Internet, Internet Draft, (1999).

[12] T. Nandagopal, N. Venkitaraman, R. Sivakumar, V. Bharghavan, Relative delay differentation and delay class adaptation in core-stateless networks, in: Proceedings of IEEE INFOCOM, March, 2000, pp. 421–430.

[13] V. Paxson, S. Floyd, Wide-area traffic: The failure of Poisson modeling, IEEE/ACM Trans. Network. 3 (3) (1995) 226–244.

[14] S. Sahu, D. Towsley, J. Kurose, A quantitative study of Differentiated Services for the Internet, in: Proceedings of IEEE Global Internet'99, December, 1999.

[15] N. Semret, R. Liao, A. Campbell, A. Lazar, Peering and provisioning of Differentiated Internet Services, in: Proceedings of INFOCOM, March, 2000, pp. 414–420.

[16] D. Sleator, R. Tarjan, Amortized efficiency of list update and paging rules, Commun. ACM 28 (1985) 202–208.

[17] I. Stoica, H. Zhang, Providing guaranteed services without per flow management, in: Proceedings of SIGCOM, 1999, pp. 81–94.

[18] A. Veres, M. Boda, The Chaotic Nature of TCP congestion control, in: Proceedings of INFOCOM, March, 2000, pp. 1715–1723.