Loss-Bounded Analysis for Differentiated Services.

Alexander Kesselman* Yishay Mansour†

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 are bounded by the loss of an optimal offline policy plus a constant fraction of the benefit of an optimal offline policy. 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 may complement traditional competitive analysis and provide new insight and interesting results.

1 Introduction

Differentiated Services were proposed as a compromise solution for the Quality of Service (QoS) problem in Internet networks. While today there is no QoS, only the best effort model, achieving a QoS per stream (such as in ATM networks) seems to be prohibitively expensive in today's technology. Differentiated Services assigns each packet a predetermined QoS and aggregates traffic to a small number of classes [3]. Each class is forwarded using the same per-hop behavior at the router, 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 last two aspects.

Over the past few years Differentiated Services for Internet traffic have attracted a great deal of research interest in the networking community [14], [5], [11], [9]. Two basic paradigms were proposed: the "Premium" service [8] 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 $\alpha \geq 1$. For example, in the Assured service model we have two levels of priority, α 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 drastically 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

*Check Point Software Technologies and Department of Computer Science, Tel-Aviv University, *email: akesselman@checkpoint.com*.

†Department of Computer Science, Tel-Aviv University, *email: mansour@math.tau.ac.il*.

rejects/preempts packets subject to the buffer capacity constraints. When the queue is not empty 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, i.e., the sum of the benefits of delivered packets, or alternatively minimize the value of the packets it drops.

We use competitive analysis to study our policies. 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 a 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 much more desirable, as we argued. 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 between $1/2$ and $1/4$.

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 are compared with the loss of an optimal offline policy directly. In c_{lb}-loss-bounded analysis the loss of an online policy are upper bounded by the loss incurred by an optimal offline policy plus a constant fraction c_{lb} of the benefit of an optimal offline policy. We let this fraction c_{lb} be the *loss-bounded ratio* of the online policy. We show that c_{lb}-loss-bounded analysis provides $(1 - c_{lb})$-throughput-competitive guarantee. We expect loss-bounded ratio to be a small constant, and in such a case the results may be interesting. Observe that a trivial loss-bounded ratio of $c_{lb} = 1$ is achieved by any online policy and the the smaller c_{lb} is the better performance is guaranteed by the online policy. An optimal online policy has $c_{lb} = 0$.

In our setting one can either maximize the throughput of the policy or minimize the loss of the policy. This dual nature appears in many combinatorial optimization problems. Clearly an optimal solution to one derives also an optimal solution to the other. However, a good approximation of one problem does not necessarily lead to a good approximation of the other. In our setting, an approximation of the throughput does not imply an approximation of the loss, and vise versa. The intuition behind loss-bounded analysis is that we try to optimize both parameters simultaneously by finding an optimal tradeoff between the current gain and potential loss.

To illustrate the difference between the criterias let us consider a sequence of packets where the cost of the loss optimal policy is 20% of the value of the input sequence. In this case a $1/2$-throughput-competitive policy can ensure benefit of 40%, a $1/2$-loss-competitive policy guarantees 60% of the benefit and $1/10$-loss-bounded policy gains at least 72% of the benefit, compared to a maximum benefit of 80% by an optimal offline policy.

We are not the first to try and analyze Differentiated Services. Initial works have focused on simple probabilistic traffic models [6], [13]. 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 [10]. Moreover, Internet was shown to behave chaotically [16]. This highlights the advantage of competitive analysis [12], [2], 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 in the two distinct benefit value model. We extend the model of [1] by allowing preemptions and considering multiple benefit values. The main difference is that we focus on loss rather than on throughput. In [7] 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 show that the greedy policy is constant throughput-competitive. In this work, in contrast with [7], we concentrate on the loss of a policy. This dramatically changes both the analysis and the results.

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. Analysis of our policies is presented in Sections 5, 6 and 7 respectively. In Section 8 we compare loss-bounded analysis with traditional competitive analysis. In Appendixes we have some of the proofs. 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 regard 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 β-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 give a formal definition of these policies later.

The first set of results appearing in Table 1 deals with the binary benefit model, where we have packets with either high benefit of α or low benefit of 1. We show that Greedy Policy achieves a non-interesting $1/\alpha$ loss-competitive ratio, which turns out to be the tight upper bound. For loss-bounded model we prove that the loss-bounded ratio of $\sqrt{\alpha}$-Preemptive Greedy Policy is $2/\sqrt{\alpha}$, which closely matches the lower bound of $1/(2\sqrt{\alpha})$. Observe that the provided guarantee is much stronger than that of the Greedy Policy. The bound for the Greedy Policy approaches 1 (the trivial bound) when α is large, while the bound for $\sqrt{\alpha}$-Preemptive Greedy Policy approaches 0 when α is large.

Policy/Model	Loss-Comp.	Loss-Bnd.
Greedy	$1/\alpha$	$\frac{\alpha-1}{\alpha}$
$\sqrt{\alpha}$-Prmpt. Greedy	0	$2/\sqrt{\alpha}$
Impossibility results	$1/\alpha$	$1/(2\sqrt{\alpha})$

Table 1: Binary benefit setting results.

Next we extend the two benefit model and study the case of n different benefit values $\{\alpha^{\frac{i}{n}} : 0 \leq i \leq n\}$. Table 2 summarizes these results. One can see that as the number of values increases the guarantee is weaker.

Policy/Model	Loss-Comp.	Loss-Bnd.
Greedy	$1/\alpha$	$\frac{\alpha-1}{\alpha+1}$
$\sqrt{\alpha^{\frac{1}{n}}}$-Prmpt. Greedy	0	$\frac{n+2}{\sqrt{\alpha^{\frac{1}{n}}}} + \frac{2}{\alpha^{\frac{1}{n}}}$
Impossibility results	$1/\alpha$	$1/(2\sqrt{\alpha^{\frac{1}{n}}})$

Table 2: Restricted benefit setting results.

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

Policy/Model	Loss-Comp.	Loss-Bnd.
Greedy	$1/\alpha$	$\frac{\alpha-1}{\alpha+1}$
Impossibility results	$1/\alpha$	$\frac{1}{8\log\alpha}$

Table 3: Arbitrary benefit setting results.

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 $last \leq B$. When a packet arrives, a queuing policy can either reject or accept the packet if the buffer is not full, i.e., $last < B$. At any time the policy can also preempt packets that are currently in the buffer. Each time unit 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 starting and at the finishing times of a schedule the buffer is empty.

DEFINITION 3.1. *For a sequence of packets S and an online policy A we denote:*

- *the subsequence of packets with benefit b, denoted S_b*
- *the entire benefit of the sequence, denoted $V(S)$, i.e. $\sum_{p\in S} b(p)$*
- *the benefit of A on S, denoted $V_A(S)$ and the loss of A on S, denoted $L_A(S)$ (note that $V(S) = V_A(S) + L_A(S)$).*

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

DEFINITION 3.2. *We say that a policy A is c_{lb}-loss-bounded iff for every sequence of packets S, $L_A(S) \leq L_{OPT}(S) + c_{lb} \times V_{OPT}(S)$.*

For completeness we also present the traditional definitions of competitiveness.

DEFINITION 3.3. *We say that a policy A is c_{tc}-throughput-competitive iff for every sequence of packets S, $V_A(S) \geq c_{tc} \times V_{OPT}(S)$.*

DEFINITION 3.4. *We say that a policy A is c_{lc}-loss-competitive iff for every sequence of packets S, $c_{lc} \times L_A(S) \leq L_{OPT}(S)$.*

4 Scheduling Policies

First 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 a 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 minimal benefit is preempted from the buffer before acceptance of the arriving packet. An arriving packet is *rejected* otherwise, i.e. when the buffer is full and its benefit is less than the benefit of every packet in the buffer. A pseudo-code of the policy appears in the Appendix A.

Next we introduce β-Preemptive Greedy Policy that behaves like a 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. A pseudo-code of the policy appears in the Appendix A.

The intuition behind additional preemptions is as follows. Consider the following scenario. At the start 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 high benefit packets is released. In this case they are all lost. The β-Preemptive Greedy Policy solves this problem by preempting α/β low benefit packets upon arrival of each packet of benefit α. For sufficiently large values of β the β-Preemptive Greedy Policy performs like a Greedy Policy, which was shown in [7] to be 1/4-throughput-competitive. However, small values of β may cause excessive preemptions and large loss of low benefit packets. Thus there is a need to optimize the value of β in order to achieve a balance between maximizing of the current throughput and minimizing of potential future loss.

5 Binary Benefit Values

In this section we consider packets having binary benefit, that is a high benefit of $\alpha \geq 1$ or a low benefit of 1. We start by analyzing the usual loss-competitive model and then proceed to the new loss-bounded model. The next theorem shows that the Greedy Policy accomplishes a non-zero loss-competitive ratio.

THEOREM 5.1. *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 α far from the optimal. ■

This loss-competitive ratio is not appealing since as α increases the loss-competitive ratio decreases to zero. The following theorem shows that this ratio is the best that an online policy could achieve. The proof can be found in Appendix B.

THEOREM 5.2. *The loss-competitive ratio of any online policy is at most $1/\alpha$.*

The next theorem shows that the Greedy Policy has a non-interesting loss-bounded ratio between $\frac{\alpha-1}{2\alpha}$ and $\frac{\alpha-1}{\alpha}$. The proof appears in Appendix C.

THEOREM 5.3. *The loss-bounded ratio of the Greedy Policy is at most $\frac{\alpha-1}{\alpha}$ and at least $\frac{\alpha-1}{2\alpha}$.*

Next we prove that the $\sqrt{\alpha}$-Preemptive Greedy Policy achieves a $2/\sqrt{\alpha}$ loss-bounded ratio, which is almost tight (up to a constant factor). First we need some auxiliary lemmas.

LEMMA 5.1. *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 moment 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 5.1. *We say that a scheduling interval $[t_s, t_f]$ is* overloaded *iff*

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

CLAIM 5.1. *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 5.1, or the buffer was empty and the claim holds trivially. ∎

CLAIM 5.2. *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. ∎

THEOREM 5.4. *The loss-bounded ratio of the $\sqrt{\alpha}$-Preemptive Greedy Policy is at most $2/\sqrt{\alpha}$.*

Proof. We process loss of low and high benefit packets separately. We show that

$$L_A(S_1) \leq \frac{1}{\sqrt{\alpha}} V_A(S_\alpha) + L_{OPT}(S_1)$$

and

$$L_A(S_\alpha) \leq \frac{1}{\sqrt{\alpha}} V_A(S_\alpha) + L_{OPT}(S_\alpha).$$

establishing the theorem, since $V_A(S_\alpha) \leq V_{OPT}(S)$ and $L_A(S) = L_A(S_1) + L_A(S_\alpha)$.

First we bound the loss of low benefit packets. There are two kinds 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. We denote the loss of the first kind by L_A^{Extra} and the loss of the second kind by L_A^{Ovfl}. Notice that

$$L_A(S_1) = L_A^{Extra}(S_1) + L_A^{Ovfl}(S_1).$$

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 every high benefit packet for the preempted low benefit packets obtaining that

$$L_A^{Extra}(S_1) \leq \frac{1}{\sqrt{\alpha}} V_A(S_\alpha).$$

Observe, that except the additionally preempted packets, the Preemptive Greedy Policy schedules the maximum number of low benefit packets subject to the buffer capacity constraints and preference that is always given to high benefit packets, which is also true for an optimal offline policy. Thus the number of low benefit packets that fall to this category can be bounded by the number of low benefit packets that are lost by an optimal offline policy. Hence, we get that

$$L_A^{Ovfl}(S_1) \leq L_{OPT}(S_1),$$

which implies that,

$$L_A(S_1) \leq \frac{1}{\sqrt{\alpha}} V_A(S_\alpha) + L_{OPT}(S_1).$$

A slightly more complicated task is to bound the loss of high benefit packets. We divide the schedule to underloaded and overloaded intervals. No high benefit packet is lost during underloaded interval. According to Claim 5.1 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 a Greedy Policy with respect to high benefit packets.

Thus the loss of high benefit packets by the $\sqrt{\alpha}$-Preemptive Greedy Policy are 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^{Ncs} and the loss of the second kind by L_A^{Add}. Clearly, $L_A(S_\alpha) = L_A^{Ncs}(S_\alpha) + L_A^{Add}(S_\alpha)$. By definition,

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

To bound the loss of the latter type note that, by Claim 5.2, the length of an overloaded interval is at least B. In this way the ratio between the loss and the cumulative benefit of the packets scheduled during an overloaded interval is at most

$$\frac{B\sqrt{\alpha}}{B\alpha} = \frac{1}{\sqrt{\alpha}}.$$

Therefore, the loss of the latter type are bounded by

$$L_A^{Add}(S_\alpha) \leq \frac{1}{\sqrt{\alpha}} V_A(S_\alpha),$$

and

$$L_A(S_\alpha) \leq \frac{1}{\sqrt{\alpha}} V_A(S_\alpha) + L_{OPT}(S_\alpha),$$

which completes the proof. ∎

In the next theorem we prove a lower bound on loss-bounded ratio showing that the bound of Theorem 5.4 is almost tight. The proof can be found in Appendix D.

THEOREM 5.5. *The loss-bounded ratio of any online policy is at least* $\frac{1}{2\sqrt{\alpha}}$ *when* α *and* B *are large numbers.*

6 Restricted Benefit Values

In this section we consider packets that have a restricted set of n benefit values, that is $\{\alpha^{\frac{i}{n}} : 0 \leq i \leq n\}$. Observe, that all the results of loss-competitive analysis derived in the previous section may be trivially extended to the restricted and arbitrary benefit models. Using loss-bounded analysis we obtain results that are generalization of the corresponding results for the binary benefit model. We prove that the loss-bounded ratio of the $\sqrt{\alpha^{\frac{1}{n}}}$-Preemptive Greedy Policy is $(n+2)/\sqrt{\alpha^{\frac{1}{n}}}+2/\alpha^{\frac{1}{n}}$ and this bound is approximately tight. The proof of the following theorem appears in Appendix E.

THEOREM 6.1. *The loss-bounded ratio of the* $\sqrt{\alpha^{\frac{1}{n}}}$*-Preemptive Greedy Policy is at most* $(n+2)/\sqrt{\alpha^{\frac{1}{n}}}+2/\alpha^{\frac{1}{n}}$.

The next theorem presents a lower bound on loss-bounded ratio proving that the bound of Theorem 6.1 is nearly tight. The proof is similar to that of Theorem 5.5 and is omitted.

THEOREM 6.2. *The loss-bounded ratio of any online policy is at least* $1/(2\sqrt{\alpha^{\frac{1}{n}}})$ *when* α *and* B *are large numbers.*

7 Arbitrary Benefit Values

In this section we consider packets having arbitrary benefit values so that for any packet its benefit is between 1 and α. An interesting question is whether there exists a policy achieving a non-trivial loss-bounded ratio. The following theorem shows that one cannot expect a loss-bounded ratio better than logarithmic on α. The proof can be found in Appendix F.

THEOREM 7.1. *The loss-bounded ratio of any online policy is at least* $\frac{1}{8\log\alpha}$ *when* α *and* B *are large numbers.*

8 Comparative Evaluation of Loss-Bounded Analysis

In this section we evaluate loss-bounded analysis by comparing it with traditional competitive analysis.

CLAIM 8.1. *A* c_{lb}*-loss-bounded policy* A_{lb} *provides* $(1-c_{lb})$*-throughput-competitive guarantee.*

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

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

After substituting instead 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

$$V(S) - V_{A_{lb}} \leq V(S) - V_{OPT}(S) + c_{lb}V_{OPT}(S),$$
$$V_{A_{lb}} \geq (1-c_{lb})V_{OPT}(S),$$

which yields the claim. ∎

To analyze loss-competitive policies we suppose there exists some a-priory information regarding the loss of the input sequence. In this case we are able to precisely determine conditions under which loss-bounded analysis provides better guarantees than loss-competitive analysis. We introduce the following estimation of loss of an input sequence.

DEFINITION 8.1. *We say that a sequence of packets* S *is* x-loss-schedulable *if* $L_{OPT}(S) = x \times V_{OPT}(S)$.

CLAIM 8.2. *A* c_{lb}*-loss-bounded policy* A_{lb} *provides better throughput guarantees than a* c_{lc}*-loss-competitive policy* A_{lc} *while scheduling* x*-loss-schedulable sequences with* $x > \frac{c_{lb}c_{lc}}{c_{lc}-1}$.

Proof. According to Claim 8.1 the benefit of A_{lb} is at least

$$V_{A_{lb}} \geq (1-c_{lb})V_{OPT}(S).$$

At the same time the loss of A_{lc} is bounded by

$$L_{A_{lc}} \leq \frac{1}{c_{lc}}L_{OPT}(S) = \frac{1}{c_{lc}}xV_{OPT}(S).$$

After substituting instead of loss the difference between the total benefit of the sequence and the benefit of the policy for A_{lc} we get

$$V_{A_{lc}} \geq V(S) - \frac{1}{c_{lc}}xV_{OPT}(S),$$
$$V_{A_{lc}} \geq V_{OPT}(S) + xV_{OPT}(S) - \frac{1}{c_{lc}}xV_{OPT}(S),$$

Thus the benefit of A_{lb} is greater than the benefit of A_{tc} provided that

$$1 - c_{lb} > 1 + x - \frac{1}{c_{lc}}x,$$

which yields the claim. ∎

Observe that a loss-bounded policy outperforms a loss-competitive policy on sequences with large values of x.

9 Concluding Remarks

In this work we investigated a framework of Differentiated Services. We shown importance of analysis of loss of an online policy and obtained impossibility results for traditional competitive analysis. Then we introduced a new powerful model for loss evaluation - loss-bounded analysis. For various QoS parameters settings we presented tight lower and upper bounds for FIFO buffer management and packet scheduling.

The main theoretical contribution is a new model in which loss of an online policy are upper bounded by loss of an optimal offline policy plus a constant fraction of the benefit of an optimal offline policy. We shown that loss-bounded analysis can give us much better performance guarantees than traditional competitive analysis.

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 α 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 α may be made very large. Conversely, for α near one, we are basicly 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 α.

Throughout this paper we analyze the Preemptive Greedy Policy and analytically prove its good performance. However, from practical point of view, it could behave pathologically at lower loads, when it does unnecessary preemptions. To overcome this problem we can add a *threshold* of B/β of high benefit packets in the buffer starting from which only high benefit packets are scheduled. This policy may be shown to achieve the same loss-bounded ratio as the β-Preemptive Greedy Policy.

References

[1] W. A. Aiello, Y. Mansour, S. Rajagopolan and A. Rosen, "Competitive Queue Policies for Differentiated Services," *Proceedings of INFOCOM 2000*, pp. 431-440.

[2] Allan Borodin and Ran El-Yaniv, "Online Computation and Competitive Analysis," *Cambridge University Press*, 1998.

[3] Y. Bernet, A. Smith, S. Blake and D. Grossman, "A Conceptual Model for Diffserv Routers," *Internet draft*, March 2000.

[4] D. Clark and J. Wroclawski, "An Approach to Service Allocation in the Internet," *Internet draft*, July, 1997.

[5] C. Dovrolis, D. Stiliadis and P. Ramanathan, "Proportional Differentiated Services: Delay Differentiation and Packet Scheduling", *Proceedings of ACM SIGCOMM'99*.

[6] M. May, J. Bolot, A. Jean-Marie, and C. Diot, "Simple Performance Models of Differentiated Services Schemes for the Internet," *Proceedings of IEEE INFOCOM 1999*, March 1999.

[7] Y. Mansour, B. Patt-Shamir and Ofer Lapid, "Optimal Smoothing Schedules for Real-Time Streams," to appear in *PODC 2000*.

[8] K. Nichols, V. Jacobson and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet," *Internet draft*, July 1999.

[9] T. Nandagopal, N. Venkitaraman, R. Sivakumar and V. Bharghavan, "Relative Delay Differentation and Delay Class Adaptation in Core-Stateless Networks," *Proceedings of IEEE Infocom 2000*, pp. 421-430, March 2000.

[10] V. Paxson and and S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, Vol. 3, No. 3, pp. 226-244, June 1995.

[11] Nemo Semret, Raymond R.F. Liao, Andrew T. Campbell and Aurel A. Lazar, "Peering and Provisioning of Differentiated Internet Services," *Proceedings of INFOCOM 2000*, pp. 414-420, March 2000.

[12] D. Sleator and R. Tarjan, "Amortized Efficiency of List Update and Paging Rules," *CACM 28*, pp. 202-208, 1985.

[13] S. Sahu, D. Towsley and J. Kurose, "A Quantitative Study of Differentiated Services for the Internet," *Proceedings of IEEE Global Internet'99*, December 1999.

[14] Ion Stoica and Hui Zhang, " Providing Guaranteed Services without Per Flow Management," *Proceedings of SIGCOM 1999*, pp. 81-94.

[15] J. Turner, "New Directions in Communications," *Proceedings of the Zurich Seminar on Digital Communication*, pp. 25-32, March 86.

[16] Andras Veres and Miklos Boda, "The Chaotic Nature of TCP Congestion Control," *Proceedings of INFOCOM 2000*, pp. 1715-1723, March 2000.

A Scheduling Policies

Greedy Policy:

Algorithm A.1.

GreedyPacketArrivalHandler(p) :

```
if (last < B) then
  accept(p);
else
  mbp = F[1]; /* get a packet with minimal benefit */
  for i = 2 to last do
    if (F[i].benefit < mbp.benefit) then
      mbp = F[i];
  if (mbp.benefit < p.benefit) then
    preempt(mbp);
    accept(p);
  else
    reject(p);
```

β-Preemptive Greedy Policy:

Algorithm A.2.

PreemptiveGreedyPacketArrivalHandler(p) :

```
/* apply the Greedy packet arrival handling routine */
GreedyPacketArrivalHandler(p);
/* check whether p was accepted */
if (F(last) == p) then
  val = p.benefit/β; /* calc. how much to preempt */
  for i = 1 to last − 1 do
    if (F[i].benefit < val) then
      val = val − F[i].benefit;
      preempt(F[i]);
```

B Proof of Theorem 5.2

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. ∎

C Proof of Theorem 5.3

Proof. First we prove the upper bound. In the worst case scenario the Greedy Policy schedules a packet of benefit 1 and loses a packet of benefit α while an optimal offline policy does exactly the opposite. Thus in addition to the loss of an optimal offline policy the Greedy Policy may lost at most $1 - 1/\alpha$ fraction of the benefit of an optimal offline policy.

Next we prove the lower bound. Consider the following scenario. At time $t = 0$ the buffer is empty and B low benefit packets arrive. For the next B time units a single high benefit packet arrives. Finally, at time $t + B + 1$ a burst of B high benefit packets is released. The Greedy Policy loses all the burst since its buffer at time $t+B+1$ is full of high benefit packets. Notice that an optimal offline policy would have lost all low benefit packets and accept the burst of high benefit packets. By this way the the loss-bounded ratio of the Greedy Policy has to satisfy $L_A(S) \leq L_{OPT}(S) + c_{lb}V_{OPT}(S)$, that is $B\alpha \leq B + c_{lb}(2B\alpha)$, which yields the lower bound. ∎

D Proof of Theorem 5.5

Proof. Suppose that packets are scheduled according to the online policy A. Let at time $t_s = 0$ the buffer is empty and $B - 1$ low benefit packets arrive. Let t_f be the first time when the buffer contains no low benefit packets and let $k = t_f - t_s$. Each time unit in the interval $[t_s, t_f)$ a high benefit packet is released. 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 $0 \leq k < B/\sqrt{\alpha}$, the input sequence is terminated. The loss-bounded ratio of the online policy c_{lb} must fulfill the following

$$B - k - 1 \leq c_{lb}(B - 1 + k\alpha) < c_{lb}(B + k\alpha).$$

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

$$c_{lb} > \frac{B-B/\sqrt{\alpha}-1}{B+B\sqrt{\alpha}} \geq \frac{1-1/\sqrt{\alpha}-1/B}{1+\sqrt{\alpha}} > \frac{1}{2\sqrt{\alpha}}.$$

Note that the ratio falls above $1/(2\sqrt{\alpha})$ for sufficiently large α and B.

(2) In case $B/\sqrt{\alpha} \leq k < B$, at time t_f we have B high benefit packets arrive. 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

$$\begin{aligned} B - k - 1 + k\alpha &\leq B - 1 + c_{lb}(k\alpha + B\alpha), \\ k(\alpha - 1) &\leq c_{lb}(k\alpha + B\alpha), \end{aligned}$$

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

$$c_{lb} \geq \frac{k(\alpha-1)}{k\alpha+B\alpha}.$$

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

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

which holds for sufficiently large α. ∎

E Proof of Theorem 6.1

LEMMA E.1. *When packets are scheduled according to the $\sqrt{\alpha^{\frac{1}{n}}}$-Preemptive Greedy Policy, and there are at least $B/\sqrt{\alpha^{\frac{1}{n}}}$ packets of benefit greater than $\alpha^{\frac{i-1}{n}}$ in the buffer at the next time moment the policy schedules a packet with benefit not less than $\alpha^{\frac{i}{n}}$.*

The proof is similar to that of Lemma 5.1.

DEFINITION E.1. *We say that a scheduling interval $[t_s, t_f]$ is* i-overloaded *iff:*

1. *all the following holds regarding the interval*
   - (a) *some packet of benefit $\alpha^{\frac{i}{n}}$ was rejected during this interval*
   - (b) *no packet of benefit greater than $\alpha^{\frac{i}{n}}$ was rejected during this interval*
   - (c) *only packets with benefit at least $\alpha^{\frac{i}{n}}$ 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 $\alpha^{\frac{i}{n}}$ was scheduled*
   - (b) *it is the finishing time of a j-overloaded interval with $j > i$*
   - (c) *the buffer was empty*
3. *one of the following holds at time $t_f + 1$*
   - (a) *a packet with benefit less than $\alpha^{\frac{i}{n}}$ was scheduled*
   - (b) *it is the starting time of a j-overloaded interval with $j > i$*
   - (c) *the buffer was empty*
4. *the interval is maximal, i.e., it is not contained in any other i-overloaded interval.*

CLAIM E.1. *When packets are scheduled according to the $\sqrt{\alpha^{\frac{1}{n}}}$-Preemptive Greedy Policy the length of a sequence of consecutive overloaded intervals with increasing index is at least B.*

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 $\alpha^{\frac{i}{n}}$ or by scheduling a packet of benefit less than $\alpha^{\frac{i}{n}}$. 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 $\alpha^{\frac{i}{n}}$ was rejected during this interval. At this time moment the buffer was full of packets with benefit not less than $\alpha^{\frac{i}{n}}$ and it took at least B time units to schedule all these packets to completion. ∎

We prove that the loss-bounded ratio of the $\sqrt{\alpha^{\frac{1}{n}}}$-Preemptive Greedy Policy is at most $(n+2)/\sqrt{\alpha^{\frac{1}{n}}} + 2/\alpha^{\frac{1}{n}}$.

Proof. The packets may be lost either due to preemptions (due to buffer overflow and the additional ones) or due to rejections. We denote the loss of the first kind by L_A^{Prm} and the loss of the second kind by L_A^{Rej}. We further divide the loss owing to preemptions into packets that were scheduled and the packets that were lost by an optimal offline policy. These loses are denoted by $L_A^{Prm'}$ and $L_A^{Prm''}$ respectively. We show that

$$L_A^{Prm'}(S) \leq (1/\sqrt{\alpha^{\frac{1}{n}}} + 1/\alpha^{\frac{1}{n}})(V_A(S) + V_{OPT}(S)),$$

and

$$L_A^{Rej}(S) + L_A^{Prm''}(S) \leq L_{OPT}(S) + \frac{n}{\sqrt{\alpha^{\frac{1}{n}}}} V_A(S),$$

establishing the theorem.

First we bound $L_A^{Prm'}$. An arriving packet p may first preempt a packet of lower benefit in time of buffer overflow, which constitutes at most $1/\alpha^{\frac{i}{n}}$ fraction of the benefit of p. In addition p's extra preemptions constitute at most $1/\sqrt{\alpha^{\frac{1}{n}}}$ fraction of its benefit.

We should also take in account that each preempted packet might have itself preempted other packets. However, in a preemptions chain the benefit increases very fast and it is dominated by α. Therefore such a chain is always finite and the last packet 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. If the last packet in the chain of preemptions has benefit $\alpha^{\frac{i}{n}}$ and $y = 1/\sqrt{\alpha^{\frac{1}{n}}} + 1/\alpha^{\frac{1}{n}}$. then the cumulative benefit of its recursive preemptions set is bounded by the following fraction of its benefit (for $y < 1$):

$$y + y^2 + \ldots < \frac{y}{1-y}.$$

In this way the ratio between the benefit of the recursive preemtions set of a packet and their summary benefit is upper bounded by

$$\frac{\alpha^{\frac{i}{n}} y/(1-y)}{\alpha^{\frac{i}{n}} y/(1-y)+\alpha^{\frac{i}{n}}} = y.$$

We charge a packet and the set of packets S from its recursive preemptions set that were scheduled by an optimal offline policy of the loss of S obtaining that

$$L_A^{Prm'}(S) \leq (1/\sqrt{\alpha^{\frac{1}{n}}} + 1/\alpha^{\frac{1}{n}})(V_A(S) + V_{OPT}(S)).$$

It remains to bound L_A^{Rej} and $L_A^{Prm''}$. 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 $\alpha^{\frac{i}{n}}$ is lost in time of such interval. Thus we have to concentrate on the loss of packets with benefit less than or equal to $\alpha^{\frac{i}{n}}$.

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 $\alpha^{\frac{i}{n}}$ was scheduled. If there is no such t_l then an optimal offline policy would have also lost all the packets in the interval's lost set. So assume that such time moment exists.

According to Lemma E.1 at time t_l there are at most $B/\sqrt{\alpha^{\frac{1}{n}}}$ packets of benefit $\alpha^{\frac{i}{n}}$ in the buffer. Thus in the best case an optimal offline policy could have scheduled additionally at most $B/\sqrt{\alpha^{\frac{1}{n}}}$ of the lost packets of benefit no more than $\alpha^{\frac{i}{n}}$. There may be two cases regarding the length of the interval, that is $l(\mathcal{I}) = t_s - t_f + 1$.

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

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

(2) In case $l(\mathcal{I}) < B$ according to Claim E.1 there exists a sequence of consecutive intervals with increasing index, whose length is at least B. We charge this sequence of loss of $B/\sqrt{\alpha^{\frac{1}{n}}}$ additional packets. Note that a sequence of intervals may be charged at most n times because n is the maximum number of intervals in a sequence with increasing index.

Therefore the cumulative loss owing to rejections and the loss of the preempted packets that were rejected by an optimal offline policy is bounded from above by

$$L_A^{Rej}(S) + L_A^{Prm''} \leq L_{OPT}(S) + \frac{n}{\sqrt{\alpha^{\frac{1}{n}}}} V_A(S),$$

which completes the proof. ■

F Proof of Theorem 7.1

Proof. Suppose that packets are scheduled according to the online policy A. In this proof we use "pumping technique", when the weights of packets in the input sequence are increased exponentially. We assume that α is a power of 2.

Let at time $t = 0$ the buffer is empty and B packets of benefit 1 arrive. In the following $B(\log\alpha - 1)$ time units we release each time unit one packet from the following input sequence, where each benefit appears exactly B times:

$$\underbrace{2\cdots 2}_{B}, \underbrace{4\cdots 4}_{B}, \ldots, \underbrace{\alpha/2\cdots\alpha/2}_{B}, \underbrace{\alpha\cdots\alpha}_{B}.$$

The input sequence is terminated if at least $B/(2\log\alpha)$ packets of the same benefit are dropped. Now there may be two cases.

(1) In case the input sequence was terminated at some time $t_d \leq t + B(\log\alpha - 1)$ due to the fact that the online policy had lost at least $B/(2\log\alpha)$ packets of benefit b, we have the following. At time t_d the buffer may contain at most $B-1$ packets of benefit $2b$ and the benefit of the packets arrived before t_d is upper bounded by $2bB - B$. Thus the loss-bounded ratio of the online policy c_{lb} has to satisfy

$$b\frac{B}{2\log\alpha} \leq c_{lb}(2b(B-1) + 2bB - B) < c_{lb}4bB.$$

So,

$$c_{lb} > \frac{1}{8\log\alpha}.$$

(2) In case the online policy had dropped at most $B/(2\log\alpha)$ packets from every benefit, at time $t + B(\log\alpha - 1) + 1$ we release a burst of B packets of benefit α. Observe, that at this time the buffer already contains at least $B/2 - 1$ packets of benefit α since A had dropped at most $B/(2\log\alpha) \times \log\alpha = B/2$ packets. Therefore no more than $B/2 + 1$ packets from the burst could be accepted. Thus the online policy drops at least $B/2 - 1$ packets of benefit α, since an optimal offline policy could have dropped $B - 1$ packets of benefit 1 from the beginning of the input sequence. We take the loss-bounded ratio c_{lb} of the online policy over the entire sequence

$$(B/2 - 1)\alpha \leq B - 1 + c_{lb}(2B\alpha - 2B + B\alpha),$$

obtaining that

$$c_{lb} > \frac{1/2 - 1/B - 1/\alpha - 1/(B\alpha)}{3} > \frac{1}{8\log\alpha},$$

provided that α and B are sufficiently large. ■