# Marking for QoS Improvement [1]

I. Yeom [a,2] and A. L. N. Reddy [a,3]

[a] *Department of Electrical Engineering, Texas A & M University, College Station, TX 77843-3128*

**Abstract**

Differentiated services architecture is receiving wide attention as a framework for providing different levels of service in the Internet. Current architecture allows customers to mark their packets and the network provider to check them for conformance to service contracts. This paper looks at the problem of achieving specific QoS goals of individual flows by flexibly managing resources available to an aggregated source. The paper shows that an aggregated source can maintain state of individual flows at the edge of the network and utilize this state effectively in adaptively marking packets of individual flows to meet their QOS goals. The paper shows that the realized bandwidth can be impacted by the interaction between marking strategies employed by different sources. The paper also proposes a simple scheme for improving the service provided to a receiving-intensive application by transferring resources to the edge of the network on the sender's side. The paper studies the impact of these sender's side marking strategy and the receiver's willingness to pay for resources in achieving QOS goals of individual flows.

*Key words:* Differentiated service, Aggregation, Quality of service, Packet marking strategies, sender, receiver QOS.

## 1 Introduction

Differentiated services (diff-serv) architecture is receiving wide attention as a proposal to provide different services over networks in a scalable manner [1,2]. There are currently two per-hop behaviors (PHBs) standardized by Internet Engineering Task Force (IETF). Expedited Forwarding (EF) PHB provides

[1] This work was supported in part by a Texas ATP grant and by an NSF Career Award and by a gift from EMC Corp.

[2] E-mail: ikjun@ee.tamu.edu

[3] E-mail: reddy@ee.tamu.edu

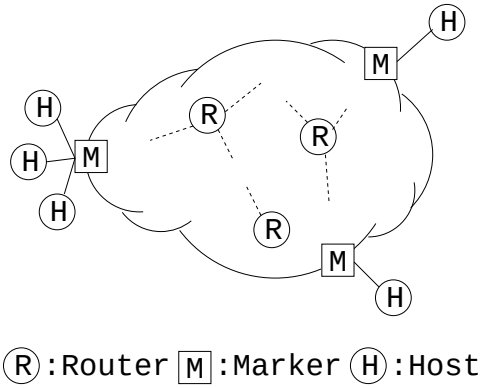



Fig. 1. Network elements

low loss, low latency, low jitter and assured bandwidth [4,5]. Assured Forwarding (AF) PHB allows a service provider to provide different levels of forwarding assurances according to the customer's profile [3]. In this paper, we focus on AF PHB.

Figure 1 shows the different elements of a network for AF service [6]. In AF service framework, the routers at the edge of the network monitor and mark packets of flows (individual or aggregated). The packets of a flow that obey the service profile are marked IN (in profile) and the packets that are beyond the service profile are marked OUT (out-of-profile). The network gives preference to IN packets while dropping OUT packets disproportionately at the time of congestion. The router doesn't distinguish between packets of individual flows and can use FIFO style scheduling mechanisms. This preferential drop mechanism is expected to provide better throughput for IN packets than OUT packets. Diff-serv framework allows aggregated sources as well as individual sources. We will assume that a customer with an aggregated source employs his/her own marker to manage individual flows within the aggregation. The network provider may monitor and remark packets to ensure compliance of the contract.

Recent work on diff-serv networks mostly dealt with individual sources [6–8] and has shown that the service provided depends on the interaction of the actions of the routers/switches inside the network, the sender, the marker and the interaction among the different flows. This paper focuses on aggregated sources and specifically on techniques for achieving specific performance goals of individual flows within an aggregation while adhering to the service contracts. The paper looks at how the edge devices or aggregated sources could maintain state about individual flows and how this state could be utilized in improving the QOS goals of individual flows. Strategies for sending and receiving data are studied. The current diff-serv framework mostly deals with QOS issues in sending data and we propose a simple scheme to extend this to providing QOS on the receiving side as well.

This paper makes the following significant contributions: (1) presents an extensive study of different policies for managing the contracted bandwidth among the individual flows of an aggregation, (2) shows that the marking policies of one source can impact not only its performance but the performance of other sources and can change network dynamics, (3) proposes a strategy for improving the QoS for receiving-intensive applications and (4) studies the interaction of sender and receiver strategies.

The rest of the paper is organized as follows. Section 2 presents simple simulations with aggregation to motivate the rest of the paper. In Section 3, we study aggregate marking schemes to manage contract rate effectively within an aggregation through simulations. In Section 4, we propose an adaptive marking scheme to meet performance goals of individual flows and, at the same time, to avoid oversubscribed network. In Section 5, we address the problem of providing sufficient bandwidth for data reception and present simulation results based on the proposed technique. In Section 6, we discuss the simulation results and present related work. Section 7 concludes the paper and points to future work.

## 2 Motivation

The immediate motivation for this work came out of an observation that there may exist serious unfairness within aggregated flows, while the total throughput of the aggregation reaches its target rate. The unfairness can be caused by different round-trip times (RTTs), different link bandwidths, or different levels of congestion experienced by individual flows within the network.

Fig. 2 shows a simple network topology used in simulations to illustrate the impact of different RTTs within an aggregated source. There are five aggregated sources, each aggregated source consists of ten individual sources. Bandwidth of every link except the link between the two routers is 10 Mbps, and bandwidth of the link between two routers is limited to 6 Mbps. Each aggregated source reserves 1 Mbps. We assign $RTT_j^i$, RTT of $j^{th}$ individual source in $i^{th}$ aggregated source as;

$$RTT_j^i \;=\; 130 \;+\; 4 \times (i-1) \times (j-5.5) \text{ (ms)} \qquad (1)$$

This results in five aggregated sources with varying differences in RTTs. For example, aggregated source 1 has a (min RTT, max RTT) = (130 ms, 130 ms) compared to that of aggregated source 5 with (58 ms, 202 ms). We use RIO (RED with IN/OUT) mechanism [6] for the dropper with parameters $(q_{min}/q_{max}/p_{max})$ = 20/40/0.5 for OUT and 40/100/0.02 for IN packets.

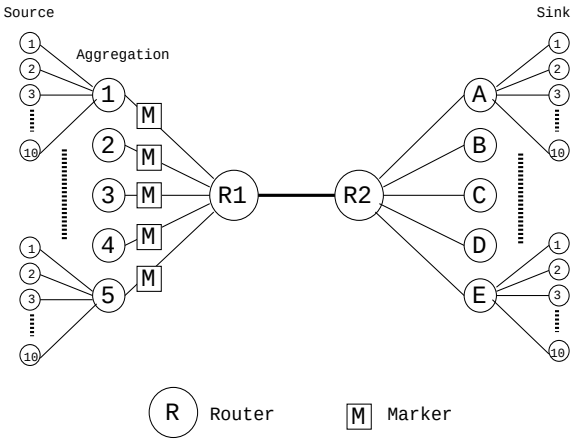



Fig. 2. Aggregated network topology with different RTTs

Fig. 3 shows the achieved rates of individual sources. The horizontal line in the figure shows the individual target rate if the aggregate target bandwidth is equally shared among the 10 individual sources. It is clear that there exists unfair bandwidth sharing within aggregated sources. This unfairness increases as the differences in RTTs increase. For example, a source achieves 4.8 times the bandwidth of another source in the fifth aggregation because of better RTTs. Even though the 5 aggregated sources achieved nearly identical overall throughputs, individual sources can realize widely varying throughputs.

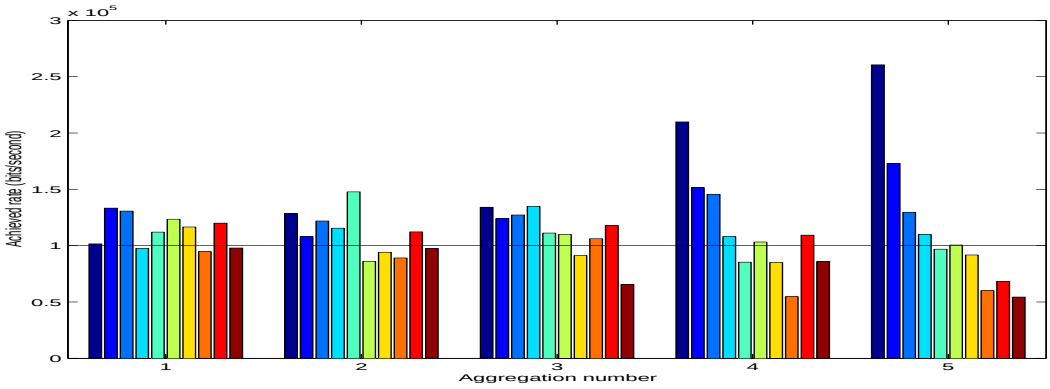



Fig. 3. Impact of different RTTs within aggregated sources

Fair sharing of bandwidth is used here as an example. In general, individual flow requirements will be considered as targets. In the current simulation, the aggregated traffic is marked without any knowledge about the individual flows. We will term this *Proportional Marking* since each flow is likely to get an equal fraction of its traffic marked OUT when the aggregated traffic exceeds contracted bandwidth profile. However, an aggregated source may employ its own marker such that packets of individual flows may be marked differently (based on their QOS goals) while ensuring that the aggregated traffic marking does not violate the contract with the network provider. In the following sections, we present such strategies for marking packets of individual

flows to improve QOS for both sending and receiving data.

## 3 Aggregate Marking for Aggressive Bandwidth Management

We consider maintaining state for each flow within an aggregation at the boundary router. Average sending rate of a flow is maintained as state information for each flow at the marker of the aggregated source. This information is used in balancing resources across the different flows within the aggregation.

If we apply the proportional marking strategy with TSW (Time Sliding Window) proposed by [6] for aggregated sources, the cumulative total sending rate of $n$ aggregated sources $B$ is

$$B = \sum_{i=1}^{n} b_i \tag{2}$$

where $b_i$ is the individual sending rate of $i^{th}$ flow. When $B$ is less than the contract rate $M$ [4], every packet is marked IN. If $B$ is greater than $M$, then a packet is marked IN with a probability of $M/B$. Therefore, we have,

$$M = \sum_{i=1}^{n} m_i = \frac{M}{B} \sum_{i=1}^{n} b_i \tag{3}$$

$$m_i = \frac{M}{B} b_i \tag{4}$$

where $m_i$ is the marking rate of $i^th$ flow. Here note that $M/B$ is the same to every individual flow within the aggregation. Thus, $m_i$ is proportional to $b_i$. The proportional marking has merit in its simple implementation since it does not need to maintain per-flow state. However, it also has two undesirable properties: (1) Contract rate is unfairly distributed within an aggregation, and (2) Marking rate of an individual flow is affected by other flows within the aggregation. Increasing throughput of an individual flow increases $B$ and decreases $M/B$. Although a flow maintains its throughput ($b_i$), its marking rate($m_i$) is reduced causing $b_i$ to decrease, and vice versa.

Now we propose two aggregate marking algorithms, called *IN-fair* and *BW-fair marking*. IN-fair marking scheme distributes contract rate to individual flows within an aggregation equally. The IN-fair marker maintains per-flow

[4] In this paper, *contract rate* means contracted profile rate for AF traffic between users and network provider and thus, it is interchangeable with *marking rate*.

state and marks a packet by its current sending rate and individual marking rate. The individual marking rate is given by

$$m_i = \frac{M}{n} \tag{5}$$

where $n$ is the number of active flows within an aggregation. It is clear that the individual marking rate is not affected by the throughput of other flows.

The BW-fair marking realizes equal throughput of individual flows within an aggregation. The marking rate of an individual flow within an aggregation is given inversely proportional to its current sending rate. The individual marking rate in BW-fair marking is given by

$$m_i = \frac{M}{B(\frac{1}{\overline{b}} - \frac{1}{b_{max}})}(1 - \frac{b_i}{b_{max}}) \tag{6}$$

where $\overline{b}$ is the mean of $b_i$ for all the flows, and $b_{max}$ is the maximum throughput among the flows within the aggregation.

### *3.1 Simulations*

We will show that the proposed bandwidth management results in improved realization of individual target rates. We modified ns-2 [9] to implement the new marking algorithm. In all simulations, we used a TCP-Reno agent in ns-2 as a source and FTP application as a traffic generator.

#### *3.1.1 Dealing with different RTTs*

To show how IN-fair marking deals with different RTTs, we conducted the same simulation as in Section 2 with IN-fair marking. Fig. 4 shows the achieved rates of individual sources. Unlike Fig. 3, it is clear that the goal of fair sharing within the aggregation is better achieved.

To compare the results quantitatively, we present Table 3.1.1. The average throughput of each aggregation is not much different from each other in all the schemes even though RTTs of individual sources are different. The row STD shows the standard deviation among the individual rates within an aggregation. It is observed that STD increases significantly with increased RTT differences within an aggregation. The IN-fair marking algorithm achieves significantly smaller variation compared to proportional marking. The row

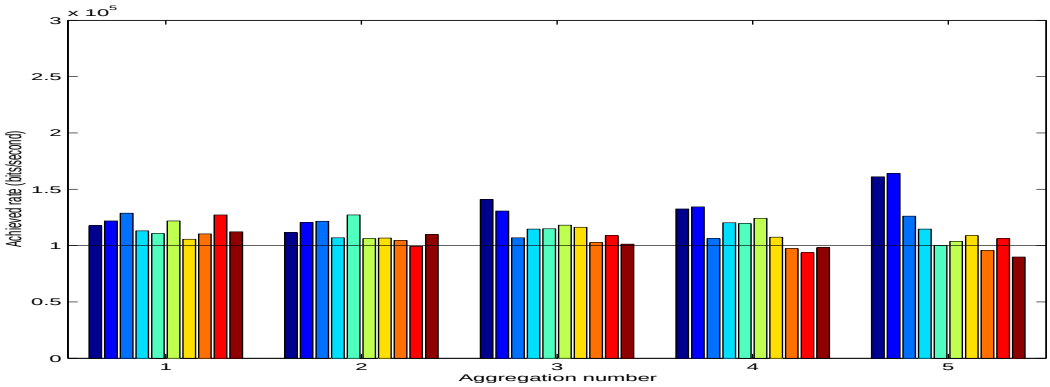



Fig. 4. Dealing with different RTTs within aggregated sources

Max/Min compares the maximum and minimum rates realized within an aggregation. Again, it is observed that fairness is considerably improved with IN-fair and BW-fair marking algorithms.

Table 1
Quantitative comparisons of marking schemes

| Aggregation | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| RTT Max/Min | 1 | 1.32 | 1.77 | 2.42 | 3.48 |
| Proportional marking | | | | | |
| Average(Kbps) | 112.7 | 110.0 | 112.2 | 113.8 | 114.5 |
| STD (Kbps) | 14.3 | 19.3 | 21.2 | 44.4 | 62.1 |
| Max/Min | 1.41 | 1.72 | 2.06 | 3.83 | 4.80 |
| IN-fair marking | | | | | |
| Average(Kbps) | 116.9 | 111.4 | 115.5 | 113.3 | 117.0 |
| STD (Kbps) | 7.7 | 8.8 | 12.3 | 14.8 | 25.3 |
| Max/Min | 1.22 | 1.28 | 1.39 | 1.43 | 1.82 |
| BW-fair marking | | | | | |
| Average(Kbps) | 113.4 | 115.3 | 117.4 | 115.3 | 112.8 |
| STD (Kbps) | 4.3 | 5.2 | 5.6 | 6.2 | 8.8 |
| Max/Min | 1.10 | 1.10 | 1.12 | 1.17 | 1.22 |

We present another simulation using the same topology in Fig. 2. RTTs of all the flows are set to 60 ms. Target rate of $j^{th}$ individual flow in $i^{th}$ aggregation is set to

$$t_j^i \;=\; 100 \;+\; 2 \times (i-1) \times (j-5.5) \text{ (Kbps)} \tag{7}$$

Each flow is assigned IN bandwidth proportional to its target. Fig. 5 shows the results. It is observed that every flow reaches its target rate and that weighted IN-fair marking easily extends to achieve specific performance goals.

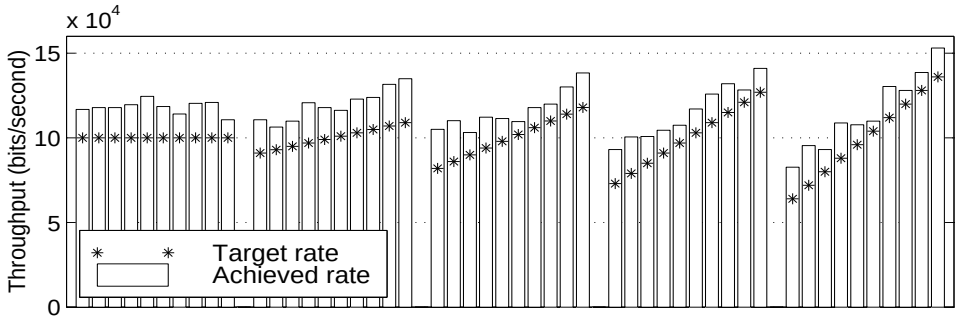



Fig. 5. Dealing with different target rates within an aggregation

### *3.1.2 Interaction of different marking schemes*

Results from the earlier section showed that IN-fair marking and BW-fair marking are effective bandwidth management strategies. How do these different marking strategies interact with each other when different sources employ different marking strategies? This section looks at this issue and points out that a source's realized bandwidth may be impacted by the marking strategy employed by a competing source.

Fig. 6 shows the network topology used for simulations. There are two aggregated sources, and each aggregated source consists of ten individual sources. Each aggregated source reserves 1 Mbps. The network consists of a 1 Mbps link between the router and node 'A' and 2 Mbps link between the router and node 'B'. Individual sources 1∼8 of each aggregated source send packets to node 'A' and individual sources 9 and 10 send packets to node 'B' through the router. In this topology, therefore, the link between the router and node 'A' is 160% subscribed, and the link between the router and node 'B' is 20% subscribed if we assume that each individual source expects to get 0.1 Mbps ($M$/number of individual sources). The network as a total has enough capacity (3 Mbps) to support the two aggregated sources (total reservation of 2 Mbps). Due to the dynamic nature of flows, one of the links may be oversubscribed as in this example. Each individual source has same RTT as 40 msec. With this simulation topology, we conducted two simulation experiments. In the first simulation, we applied proportional marking to the marker for aggregation 1 and the IN-fair marking for aggregation 2. In the second simulation, we applied proportional marking for aggregation 1 and the BW-fair marking for aggregation 2.

Fig. 7(a) shows the results of Simulation 1. Each bar shows the average throughput of individual sources, and the dark portion in each bar indicates the throughput achieved by IN packets. In proportional marking, it is observed that each packet is marked OUT with the same probability even if its source cannot reach its target rate, and IN packets are unfairly distributed to flows achieving higher rates. In IN-fair marking, however, each individual source

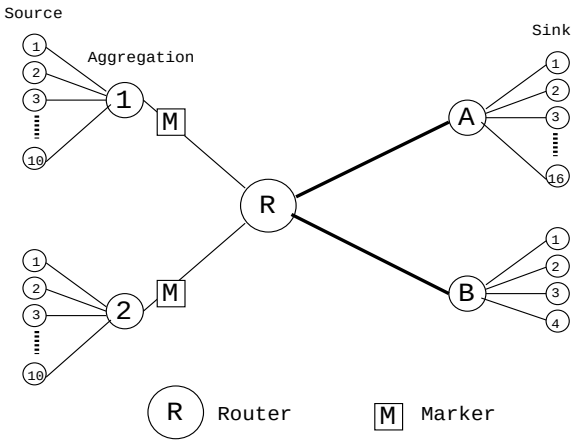
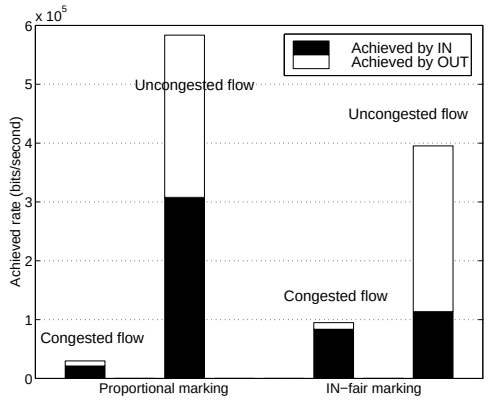



Fig. 6. Network topology with different congestion levels

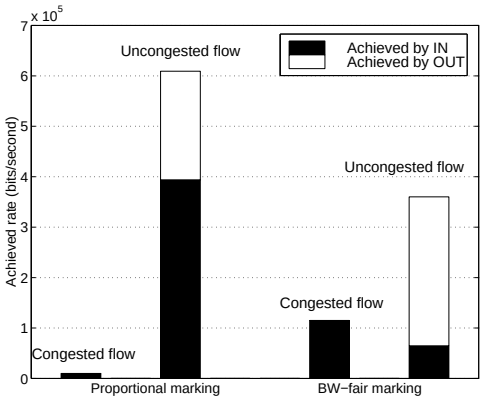
shares IN packet throughput equally even with different levels of congestion. Clearly, fair sharing of IN-profile bandwidth improved the performance of the flows through the congested link. Both the aggregated sources stay within the contract-profile, but the second source achieved a higher bandwidth through the congested link than the first source. This is a result of managing IN-profile bandwidth effectively by distributing it fairly among the individual sources. When source 1 and source 2 compete for bandwidth on the congested link, source 2 achieves higher share due to marking higher number of packets IN.

Fig. 7(b) shows the results of Simulation 2. The BW-fair marking is more aggressive by the fact that it sends more IN packets on congested links than on uncongested links so as to get more bandwidth in congested links. Again, aggregated source 1 used proportional marking and aggregated source 2 used the BW-fair marking. In proportional marking, the flows through congested link in Fig. 7(b) loose more bandwidth than the flows in Fig. 7(a). The flows using the BW-fair marking get more bandwidth than the flows using the IN-fair marking.

The BW-fair marking is more aggressive than the IN-fair marking in trying to meet the performance goals. The goal here is to achieve 0.1 Mbps for each individual source while staying within the contract-profile. As can be seen from Fig. 7(b), the BW-fair marking algorithm allocates more IN-profile bandwidth to sources observing congestion than the ones that are not experiencing congestion. As a result, these sources claim a larger share of the congested link bandwidth, exceeding the individual targets of 0.1 Mbps. The flows within aggregated source 1 achieve significantly less bandwidth due to proportional marking. These two experiments show that individual marking strategies employed by customers can impact each other even when every source stays within the contract-profile.




(a) Proportional and IN-fair marking




(b) Proportional and BW-fair marking

Fig. 7. Bandwidth comparison with different marking schemes

## 4 Adaptive Marking Strategy

In the previous section, we have presented new marking schemes for aggressive bandwidth management. The simulation results show that the new schemes improve throughput when the network resources are enough to meet the individual QoS requirements. However, it is also observed that, if the resources are not enough, the new schemes cause severe congestion resulting in resource wastage due to their aggressive manner in managing the contract rate. In this section, we propose an adaptive marking strategy. The main objectives of the adaptive marking strategy are:

(1) To achieve bandwidth objectives of flows within an aggregation.
(2) To avoid resource wastage if the first objective is not achievable due to the current network conditions.

To achieve the first objective, initially, we set the marking rate of each flow proportional to its bandwidth objective. If every flow gets throughput more than their marking rate without IN packet loss, then the adaptive marker works as a weighted IN-fair marker. On the other hand, if the network path of a flow is oversubscribed [5] and observes IN packet losses (resource wastage), the adaptive marker adjusts marking rates of individual flows in order to avoid IN packet loss (achieving the second objective). However, it is not easy for a marker to find whether a flow observes an oversubscribed network or not unless

[5] In [20], oversubscribed network has been defined as a situation in which a flow does not transmit any OUT packets since every OUT packets are dropped or no OUT packet is sent when the sending rate is less than the contract rate. In a oversubscribed network, a flow usually experiences some number of IN packet losses.

the marker is combined into the sender. For marking of aggregated flows, the marker cannot be combined into an individual sender. Thus, it can be just estimated from the current throughput. To estimate the current condition of a flow, we use throughput model proposed in [20]. From the model, throughput $B$ of a TCP flow experiencing oversubscribed network is given by

$$B = \min\{\frac{3}{4}m, \frac{k}{\mathrm{RTT}}(\sqrt{\frac{1}{9} + \frac{8}{3p_{in}}} - \frac{1}{3})\} \tag{8}$$

where $m$ is the contract rate of the flow (or the IN-marking rate), $k$ is the packet size, and $p_{in}$ is the probability of IN packet loss. From (8), when throughput achieved by a flow is less than $0.75m$, the flow should observe an oversubscribed network. Therefore we classify a flow into one of the following three states and treat these states differently. Here, $t_i$ is the target rate of $i^{th}$ individual flow, $m_i$ is the marking rate, and $b_i$ is the realized throughput. The target rate can be specified by the individual users, and $\sum t_i$ is the contract rate for the aggregation.

- $b_i \leq 0.75m_i$: In this state, the flow observes an oversubscribed network, and some IN packets are lost. Thus, the marker reduces $m_i$ so that $b_i$ is maintained to be higher than $0.75m_i$ to avoid wasting resources.
- $0.75m_i < b_i < t_i$: In this state, the flow does not reach its target. Since the network is not oversubscribed, $b_i$ can be increased by increasing $m_i$. Thus, the marker increases $m_i$ of that flow if resources are available.
- $t_i \leq b_i$: In this state, the flow already achieved its target. Thus, the marker does not need to change $m_i$ of the flow. However, if there is another flow which needs more resources, $m_i$ of this flow can be reduced as long as $b_i$ is maintained to reach its target.

Fig. 8 shows an example algorithm for the adaptive marker. This algorithm adjusts marking rate of each flow within an aggregation so that we can reduce IN-packet drops while increasing the number of flows which reach their targets. In line 1∼5, when individual throughput ($b[i]$) remains below 75% of assigned marking rate ($m[i]$) (due to severe congestion, say), the marking rate is reduced. When flows are below target rates ($b[i] < t[i]$) but above marking rates ($m[i]$), marking rate is increased since additional resources are likely to be effective. In line 7∼10, if needed, we reduce resources from the flows getting more than their targets. If every flow reaches its target rate, the marking rate is not changed. The parameters $\Delta$ and observation period determine the rate at which the flows are observed and adapted to reach their performance goals. Time complexity of this algorithm is $O(n)$ where $n$ is the number of flows, and this is allowable for an edge device marker. In this algorithm, we use TSW [6] to smooth out the individual throughput. It is important to choose $\Delta$, observation period and window size for rate estimator properly. We study impact of $\Delta$, observation period and window size in the following simulations.

At every observation period:
1.**for** $i \leftarrow 1$ to $n$
2. **if** $0.75m[i] < b[i] < t[i]$
3. $\quad m[i] = m[i] + \Delta(r[i] - t[i])$
4. **else if** $b[i] \leq 0.75m[i]$
5. $\quad m[i] = m[i] - \Delta(0.75m[i] - b[i])$
6.$M' = \sum^{i=1,n} m[i]$
7.**if** $M' > M$
8. **for** $i \leftarrow 1$ to $n$
9. $\quad$ **if** $b[i] > t[i]$
10. $\quad\quad m[i] = m[i] - \Delta(b[i] - t[i])$
11.$M' = \sum^{i=1,n} m[i]$
12.**for** $i \leftarrow 1$ to $n$
13$m[i] = M/M' \times m[i]$

$m[i]$: Marking rate of $i^{th}$ flow
$b[i]$: current rate of $i^{th}$ flow
$t[i]$: Target rate of $i^{th}$ flow
$M$: Total marking rate = Aggregate contract rate
$n$: Number of flows

Fig. 8. An algorithm for adaptive marking

Table 2
Expected bandwidth achieved by individual flows

| Destination | Sink 0 | Sink 1 | Sink 2 | Sink 3 | Sink 4 | Sink 5∼9 |
|---|---|---|---|---|---|---|
| Achieved BW (Mbps) | 0.125 | 0.25 | 0.375 | 0.5 | 0.625 | 0.75 |

### *4.1 Simulations*

In this section, we present a number of simulations and discuss the results. We set up the network topology shown in Fig. 9 using ns-2 [9]. There are four aggregations and ten sinks. Each aggregation consists of 10 individual TCP flows. $i^{th}$ individual flow of each aggregation sends packets to $i^{th}$ sink through $R0$ and $R1$. Contract rate of an aggregation is 5 Mbps. Individual target rate is set to 0.5 Mbps for simplicity. Link bandwidth between $R0$ and $R1$ is 22.5 Mbps which is higher than the total contract rate (20 Mbps). Link bandwidth between $R1$ and each sink is set differently so that each flow within an aggregation experiences different network conditions. Link bandwidth between $R1$ and $i^{th}$ sink is set to $0.5 \times (i + 1)$ Mbps. Propagation delay of each link is 5 msec. If the bandwidth is equally shared, individual flows of each aggregation are expected to get the bandwidth as in Table 2.

For droppers, we use RIO presented in [6] with parameters `20/40/0.5` for OUT packets and `40/80/0.02` for IN packets.

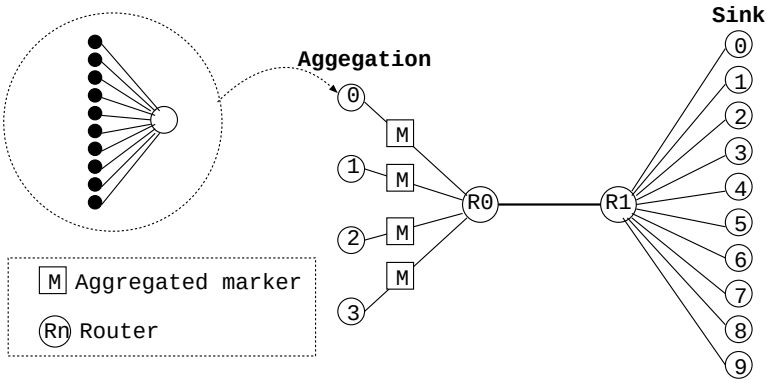



Fig. 9. Simulation topology

*4.1.1 Impact of different* $\Delta$

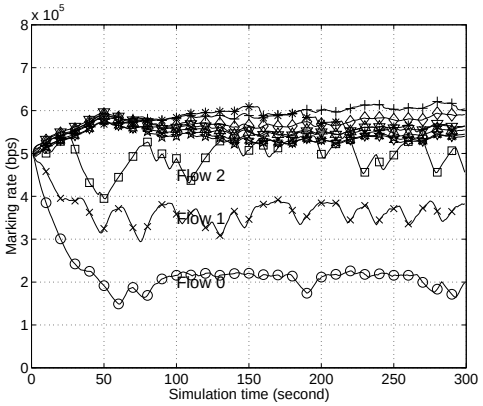
We study impact of different $\Delta$ through simulation. Again, $\Delta$ controls the rate at which the edge marker adapts the marking rate based on observed bandwidths of flows. In this simulation, we set $\Delta$ of each marker to 0.5%, 1%, 5% and 10%. Observation period of each marker is 100 msec. which is greater than RTT (= 40 msec.).

Fig. 10 shows the result. Note that we do not specify the marking rate of Flows 3∼9 in Fig. 10(a)∼10(d) since they already reach their target rates (see Fig. 10(e)), and thus their marking rates are not much different from each other. Initially, marking rate for each individual flow is 0.5 Mbps. Then, Flow 0 of each aggregation shares 0.5 Mbps link between $R1$ and Sink 0 with other three flows and gets about 0.125 Mbps average throughput. Thus, it is observed that the marking rate of Flow 0 in each aggregation is reduced so that the current throughput is maintained to be at least 75% of its marking rate.

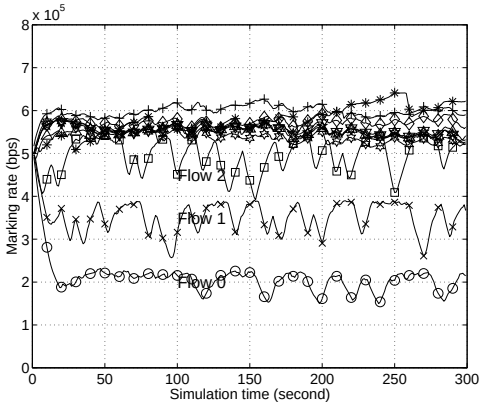
Different $\Delta$ impacts the time taken to reach the steady state. With smaller $\Delta$, it takes a longer time to reach a steady state. With larger $\Delta$, it is more likely to have oscillations. This is very similar to a heavily damped or under damped control in traditional control systems. Under stable network conditions, a flow's marking rate reaches steady state at a rate of $(1 - \Delta)$ per observation period. So over $k$ observation periods, the error is $(1 - \Delta)^k$. If steady state marking error goal is $\delta$, then we can find the suitable transient time ($k$ observation periods) through

$$(1 - \Delta)^k < \delta \tag{9}$$

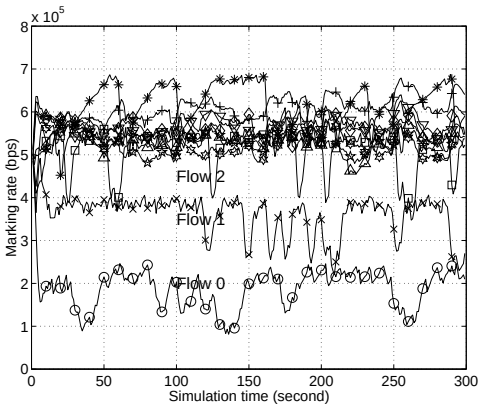
In other words, given a goal of how fast we need to reach a steady state (i.e., given $k$), we can find the rate of adaptation $\Delta$, through (9). However, it is also observed that average throughput of different $\Delta$ is not much different from each other over long period of time in Fig. 10(e).



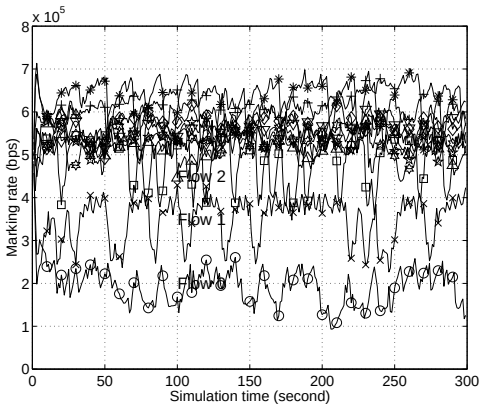

(a) $\Delta = 0.5\%$



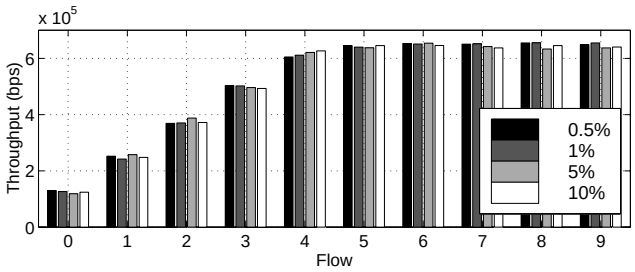

(b) $\Delta = 1\%$




(c) $\Delta = 5\%$




(d) $\Delta = 10\%$




(e) Throughput

Fig. 10. Marking rate and throughput with different $\Delta$

### *4.1.2 Impact of different observation periods*

In this section, we set observation period of each marker differently. First, we change observation period from 0.05 sec. to 1 sec. with a fixed $\Delta$ (1%) to study impact of observation period on time taken for individual marking rate

to reach steady state. In the second simulation, we select $\Delta$ so that the amount of marking rate change in a given time $\tau$ is equal to each marker. Given $o_i$ (observation period) and $\Delta_i$, if we want two flows to converge to their target in the same amount of time, from (9) we have

$$(1 - \Delta_i)^{\tau/o_i} = (1 - \Delta_j)^{\tau/o_j} \tag{10}$$

$$\frac{o_i}{o_j} = \frac{\log(1 - \Delta_i)}{\log(1 - \Delta_j)} \tag{11}$$

From (11), following four ($\Delta$, Obs. period) pairs are selected for each marker: (0.5%, 0.05 sec.), (1%, 0.1 sec.), (5%, 0.5 sec.) and (10%, 1 sec.).

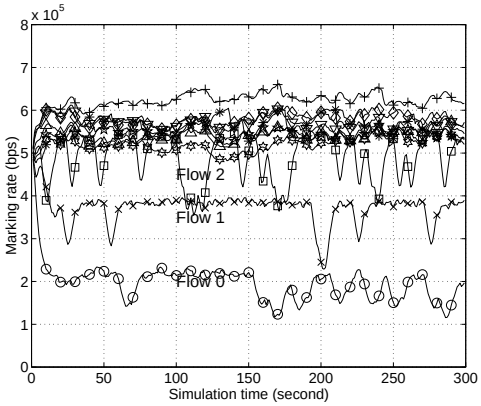
Fig. 11 shows the result of the first simulation. In Fig. 11(a)$\sim$11(d), it is clearly shown that time to steady state is linearly proportional to the observation period when $\Delta$ is the same. Note the marking rate of Flow 0: when obs. period = 1 sec, it takes about 200 sec. to reach 0.2 Mbps. With 0.5 sec. of obs. period, it is reduced to about 100 sec., and so on. In Fig. 11(e), it is also shown that throughput of flows using small observation period is slightly higher than throughput of flows using large period in Flows 1$\sim$6. It is because a marker with small observation period can adjust marking rate quickly according to the change in network conditions.

Fig. 12 shows the result of the second simulation. In Fig. 12(a), it is clearly shown that the converging time in the adaptive marker can be effectively controlled using (11). It is also shown that marking rates with small $\Delta$ and small observation period change smoothly due to small $\Delta$. In Fig. 12(b), it is observed that average throughput of different aggregations using different observation periods is not much different from each other. However, it is also shown that throughput of flows using small observation periods is slightly higher than throughput of flows using large periods in Flows 1$\sim$5.

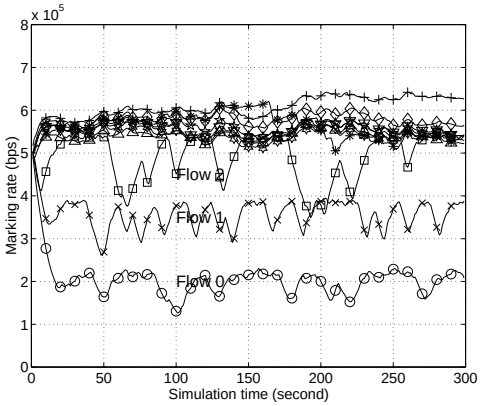
#### *4.1.3 Impact of window size for rate estimator*

In this section, we study impact of window size of TSW rate estimator. TSW rate estimator was proposed to estimate sending rate for packet tagging and shown that it is effective to smooth out TCP burstiness in [6]. We use this rate estimator for adapting the marking rate to the current individual throughput. Generally, there is a trade-off in choosing window size: With a small window, the estimated throughput reflects the changes of throughput quickly but fails to smooth the burstiness. With a large window, throughput is effectively smoothed out but not changed quickly.

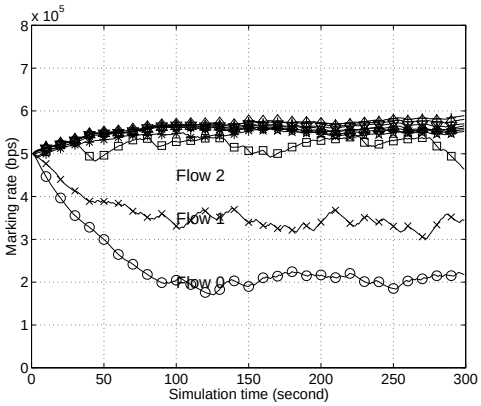
To observe impact of window size, we conducted simulation with different



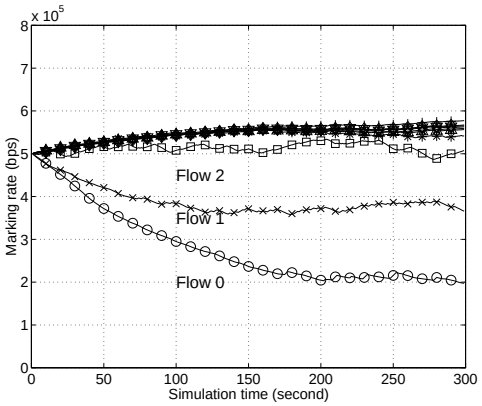

(a) Obs. period = 0.05 sec.



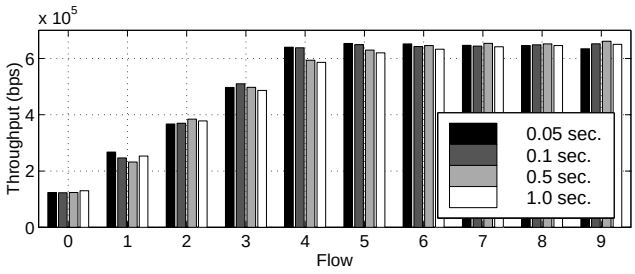

(b) Obs. period = 0.1 sec.




(c) Obs. period = 0.5 sec.




(d) Obs. period = 1 sec.

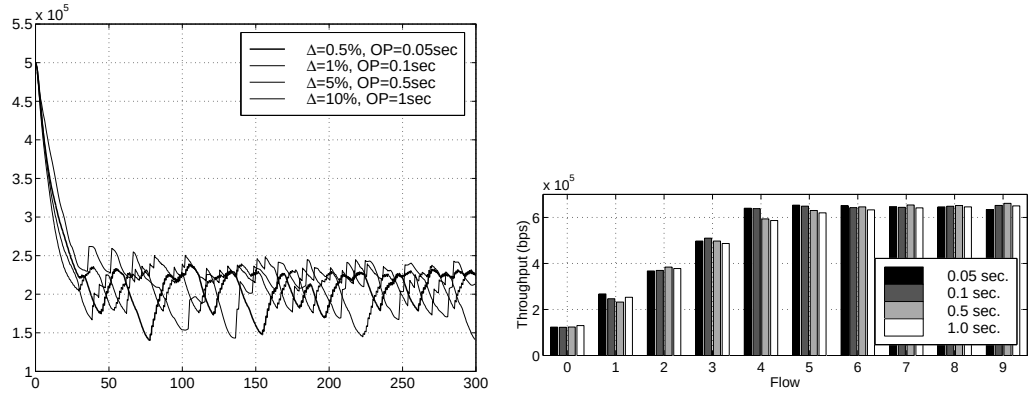



(e) Throughput

Fig. 11. Marking rate and throughput with different observation period and constant $\Delta$

sizes of window for each marker. Observation period is set to the same as its window size, and $\Delta$ is 100% for every marker.

Fig. 13 shows the marking rate and throughput of Flow 0 within each aggrega-




(a) Marking rate of Flow 0 (b) Throughput

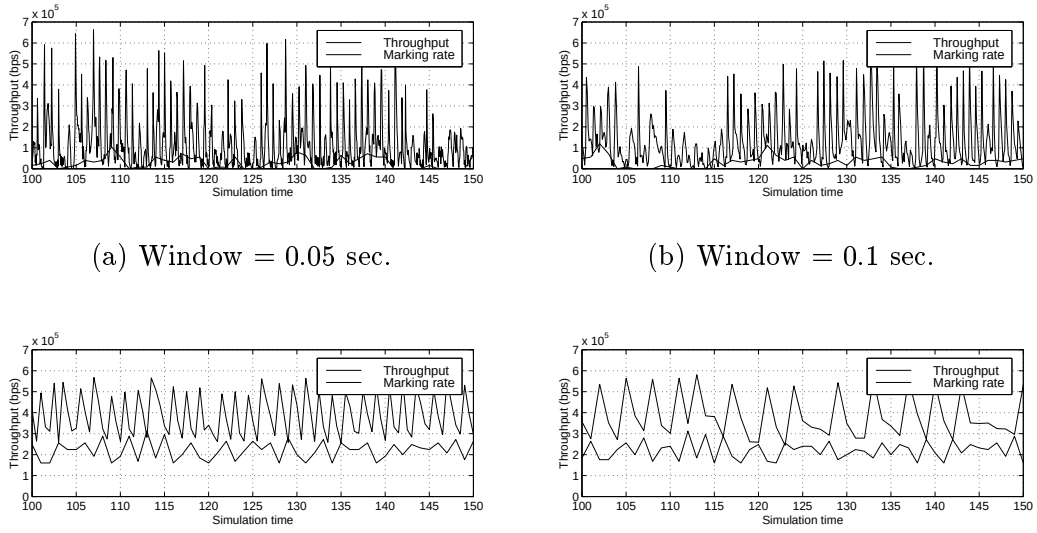
Fig. 12. Marking rate and throughput with different observation period and different $\Delta$

tion. It is observed that the marking rates oscillate due to the large $\Delta$ (100%). In Fig. 13(a) and 13(b), the marking rate oscillates over large range ($0 \sim 0.7$ Mbps) since TCP throughput cannot be smoothed with the small window (0.05/0.1 sec.). Here note that throughput stays at the bottom of the marking rate. When a set of packets arrives at the marker, estimated throughput instantaneously goes up due to the small window. This increases the marking rate. However, this increasing marking rate is not effective since the source stops sending until receiving ACKs (an RTT = 40 msec. delay). At this time, throughput decreases, and the marking rate also decreases. Then, the next set of packets observes the decreasing marking rate. This is due to the fact that the marking rate is changed quickly based on the changes of throughput. With a large window (0.5/1 sec.), the marking rate stays around 4 Mbps even through it oscillates over $0.3 \sim 0.5$ Mbps, and throughput is managed to achieve around 0.22 Mbps in Fig. 13(c) and 13(d).

Fig. 14 shows the average throughput of individual flows. With a small window, the flows observing congested links (Flows $0 \sim 2$) cannot reach their targets while the other flows get much higher than their targets (since resources are shifted to those flows).

### *4.1.4 Dealing with network dynamics*

In this section, we study how the adaptive marker adjusts the marking rate to changes in network conditions over time. To simulate changes in network conditions, we start Aggregation 0, 1, 2 and 3 at time 0 sec., 60 sec., 120 sec. and 180 sec. and stop at time 240 sec., 300 sec., 360 sec. and 420 sec., respectively. ($\Delta$, Obs. period) pair is set to (1%, 0.1 sec.) for all the markers.




(a) Window = 0.05 sec.

(b) Window = 0.1 sec.

(c) Window = 0.5 sec.

(d) Window = 1 sec.

Fig. 13. Marking rate and throughput of Flow 0 with different window size and O.P.

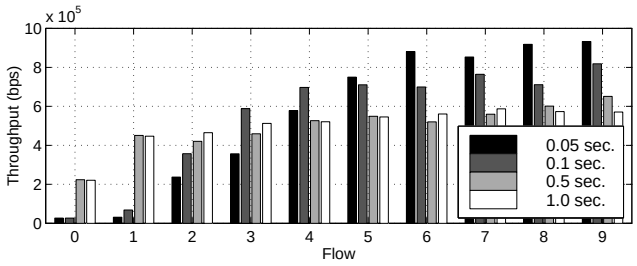
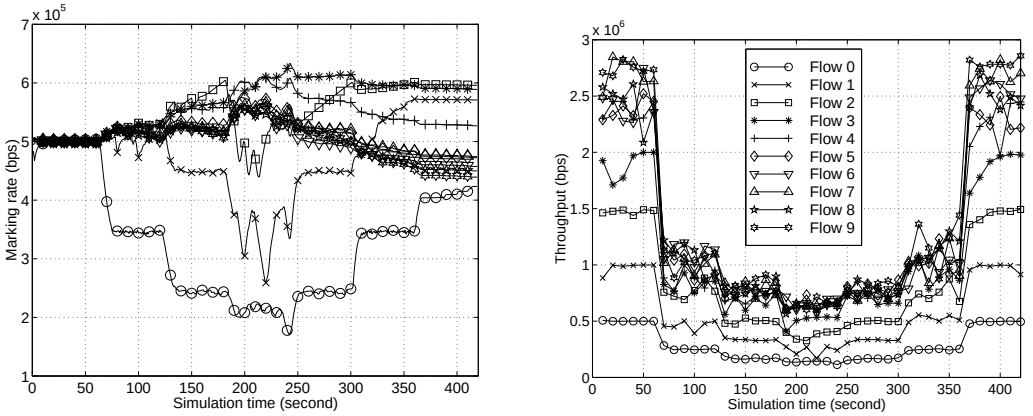



Fig. 14. Average throughput with different window size and O.P.

Fig. 15 shows the marking rate and throughput of individual flows within Aggregation 0 over time. In Fig. 15(b), every flow reaches its target rate until Aggregation 1 starts sending packets at time 60 sec. Thus, it is shown that the marking rate of every flow is equal to each other as 0.5 Mbps. In time duration from 60 sec. to 120 sec., throughput of Flow 0 falls down to 0.25 Mbps since 0.5 Mbps of link bandwidth is shared with Aggregation 1, and the marking rate is also reduced to avoid IN packet loss. Marking rates of other flows increases to utilize the total contract rate. At 120 sec. and 180 sec., Aggregation 2 and 3 start sending packets, respectively. Then, similarly, throughputs of Flows 1 and 2 do not reach their current marking rate, and the marking rate is reduced. At time 240 sec., 300 sec. and 360 sec., throughput increases as other aggregations stop sending. Here, note that the marking rate of Flow 0 does not increase to 0.5 Mbps after time 360 sec. since its throughput already reaches its target rate.




(a) Marking rate (b) Throughput

Fig. 15. Marking rate and throughput in network changed over time

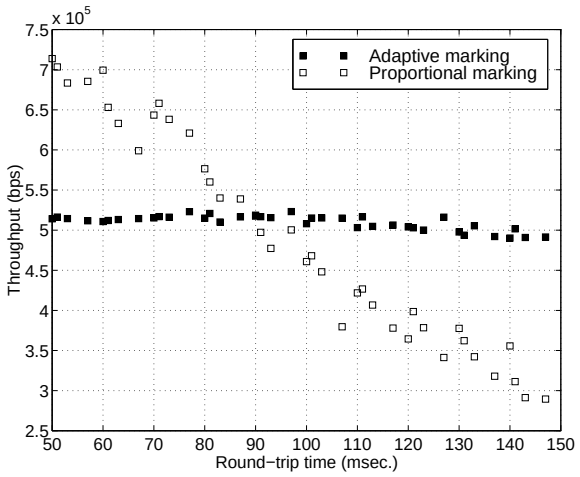



Fig. 16. Throughput of flows with different RTT

### *4.1.5 Dealing with different RTTs*

In this section, we observe how the adaptive marker deals flows with different RTTs. In this simulation, we use the same topology as used in the previous simulations except that propagation delay of link between $R1$ and each sink is randomly selected from 10 msec. to 60 msec. and that bandwidth is set to 3 Mbps for all the links between $R1$ and sinks.

Fig. 16 shows the result. It is clear that the adaptive marking effectively removes RTT-bias of TCP flows and realizes QoS goals of individual flows within aggregations.

#### *4.1.6 Dealing with unresponsive flows*

In this section, we compare the adaptive marking with proportional marking in presence of unresponsive flows. In network topology in Fig. 9, we attach three UDP sources at $R1$ with negligible contract rate. The UDP flows are connected to Sink 7, 8 and 9 and start sending packets at 180 sec, 120 sec. and 60 sec., respectively. The sending rate of each UDP flow is 3 Mbps. With this topology, we conducted simulation two times. In the first simulation, the adaptive marker is employed, and the proportional marker is employed in the second.

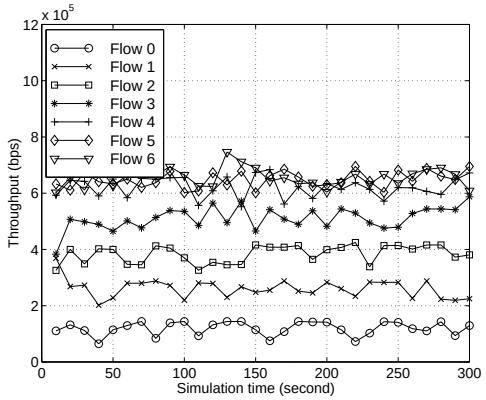
Fig. 17 shows the results. In Fig. 17(c), 17(d) and 17(e), it is observed that throughput with the proportional marking is affected by UDP flows. Throughput with the adaptive marking is maintained at a stable rate. In Fig. 17(a), it is shown that Flows 0∼6 within the aggregation also maintain their throughput. In Fig. 17(b), however, throughputs of Flows 0∼6 are constantly fluctuating over time even though they do not observe UDP flow along their path. It is because marking rate of an individual flow is directly affected by other flow's current throughput in the proportional marking case.

## 5 Receiver-side Marking Strategies

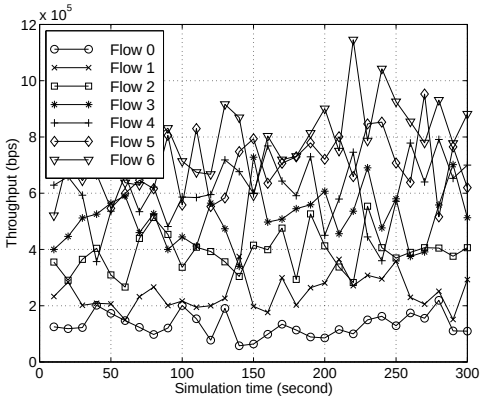
In the above sections, we have discussed marking strategies for achieving bandwidth targets for sending data. However, when the customer wants to receive data from another host (eg., using *get* command in FTP, browsing web-sites or Video-on-demand), it does not provide service differentiation, since the marker in receiver's side can mark only ACK packets. Even when the application is sending-intensive, with asymmetric link bandwidths, it may be necessary to ensure that the ACKs get sufficient bandwidth through the reverse path [12].

A *receiver-controlled scheme* using the explicit congestion notification (ECN) bit was proposed in [6]. ECN bit scheme was originally designed to provide congestion avoidance without packet drops. When congestion starts to occur, a router sets the ECN bit of packets instead of dropping them. The receiver copies the ECN bit in ACK replies to the sender, and then the sender reduces window size (or rate) to avoid congestion. In *receiver-controlled scheme*, a profile meter, installed at the receiver, measures the incoming rate. If the rate is within the profile, the meter resets the ECN bit, so that the sender maintains its transmission rate.

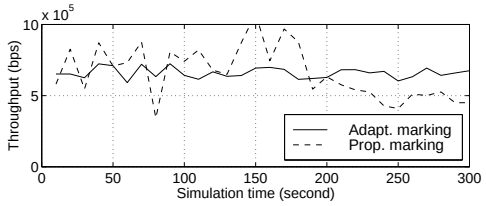
In this section, we propose a strategy for providing better service for receiving-intensive applications in a *sender-controlled* network. The basic idea is to inform the sender's side about the receiver's target rate and to transfer receiver's



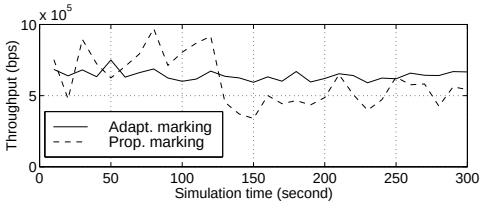

(a) Throughput of Flow 0-6 (Adaptive marking)



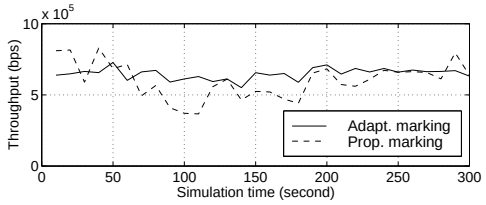

(b) Throughput of Flow 0-6 (Proportional marking)



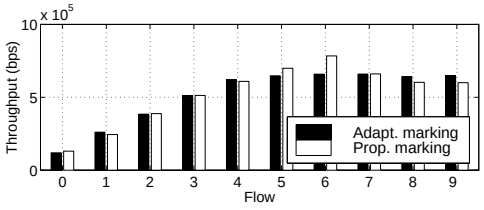

(c) Throughput comparison of Flow 7




(d) Throughput comparison of Flow 8




(e) Throughput comparison of Flow 9




(f) Average throughput

Fig. 17. Throughput comparison with adaptive and proportional marking with unresponsive flows

contracted bandwidth to the sender's edge router. The network marker on the sender side adds the transferred bandwidth to the sender's profile and upgrades OUT packets to IN within the increased profile. A signaling protocol similar to RSVP [11] needs to be developed to enable such transfers of IN-profile bandwidth. In the proposed scheme, however, the signaling protocol messages need to be processed only by the edge routers unlike RSVP which need to setup every router along the flow's path. Thus, it can be implemented without compromising the scalability of the current diff-serv architecture.

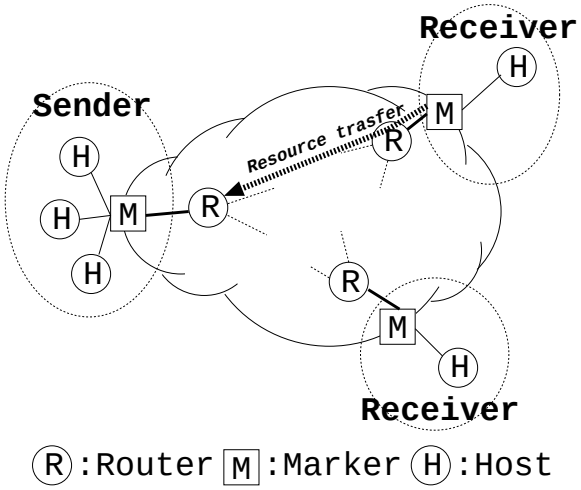



Fig. 18. Reserved bandwidth transfer by receivers in diff-serv network

Fig. 18 shows a simple example of receiver's side marking strategy. It is possible to achieve this transfer through "bandwidth brokers" in different networks. We expect this signaling to result in a transfer of IN-profile bandwidth from the receiver to the edge router connected to the sender. The edge router connected to the sender will maintain state for this flow and use the transferred bandwidth to upgrade OUT packets to IN transparently so as to improve the service delivered to the receiver. The bandwidth is not transferred to the sender in order to ensure that the transfered bandwidth is applied to upgrade service to the requesting receiver (and not to other flows being served by the sender). This is somewhat akin to the "receiver-pay mode" available in telephone networks in the form of collect calls.

How much bandwidth should a receiver transfer to the sender side to achieve a target rate? If the receiver transfers bandwidth aggressively, the sender could exploit this by reducing the number of packets marked IN to this flow. Ideally, the receiver should transfer minimal bandwidth to the sender to reach its target rate. The proposed algorithm for the receiver profile meter is presented in Fig. 19.

In the above algorithm, the receiver profile meter keeps and updates average rates and average OUT packet rate of subscribed flows. When the average rate is less than the target rate requested by the receiver (line 1), and if there is excess bandwidth (achieved by OUT packet) (line 2), the meter transfers some amount of contract rate to the sender's marker (line 3). The amount is limited by the current excess bandwidth in order to avoid resource wastage. If the average rate is higher than the target rate, then the meter takes back the contract rate to reduce payment.

At every observation period:
1.**if** $b[i] < t[i]$
2. **if** $b^{out}[i] > 0$
3. $\quad m^r[i] += \min(b^{out}[i], \Delta(t[i] - b[i]))$
4.**else**
5. **if** $m^r[i] > 0$
6. $\quad m^r[i] -= \min(m^r[i], \Delta(t[i] - b[i]))$

$b[i]$: Average rate of $i^{th}$ flow
$b^{out}[i]$: Average OUT packet rate of $i^{th}$ flow
$t[i]$: Target rate of $i^{th}$ flow
$m^r[i]$: Marking rate of $i^{th}$ flow paid by the receiver

Fig. 19. A simple algorithm for receiver-based bandwidth contract

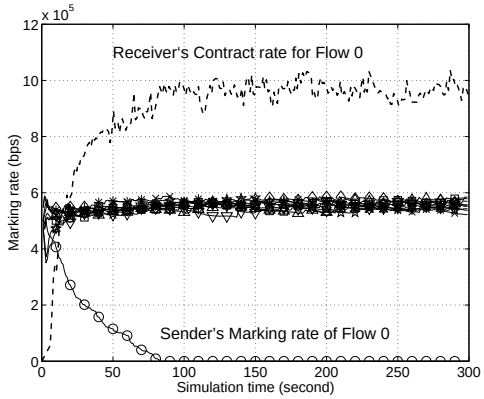
### *5.1 Simulations*

In this section, we present simulations of receiver-based strategy when the adaptive marking is employed by the sender's marker. We use the same topology shown in Fig. 9 and observe throughput realized by the interaction between the sender's strategy and the receiver's strategy. Each aggregated sender has a contract rate of 5 Mbps. Sender's target rate of an individual flow is set to 0.5 Mbps. To observe the interaction with the receiver strategy, the receiver of Flow 0 within Aggregation 0 sets a target rate of 1 Mbps. We set link bandwidth differently to produce the following three scenarios:

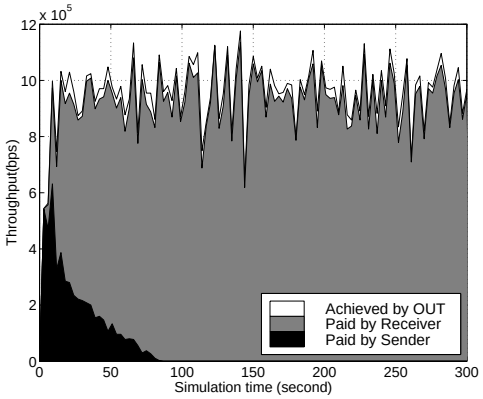
(1) Sufficient bandwidth: Network links have enough bandwidth to satisfy the receiver's target rate of 1 Mbps. Link bandwidth between $R0$ and $R1$ is 22.5 Mbps, and bandwidth between $R1$ and each sink is 3 Mbps.
(2) Insufficient bandwidth: Network links do not have enough bandwidth to reach the receiver's target. We set link bandwidth between $R1$ and Sink 0 to 2.2 Mbps.
(3) Plenty of bandwidth: Network links have plenty of bandwidth to satisfy every flow's target rate. We set link bandwidth between $R0$ and $R1$ to 30 Mbps.

In all the above simulation scenarios, $\Delta$ and the observation period are 5% and 0.1 sec. for both sender's and receiver's side. We present marking rate of individual flows within Aggregation 0 and throughput of Flow 0 in Aggregation 0.

Fig. 20 shows the result of the first scenario. It is observed that Flow 0 can realize the requested throughput 1Mbps. As the receiver increases its contribution, the sender's marker reduces the marking rate of Flow 0 since its throughput exceeds the sender's target rate (0.5 Mbps) and some of the other

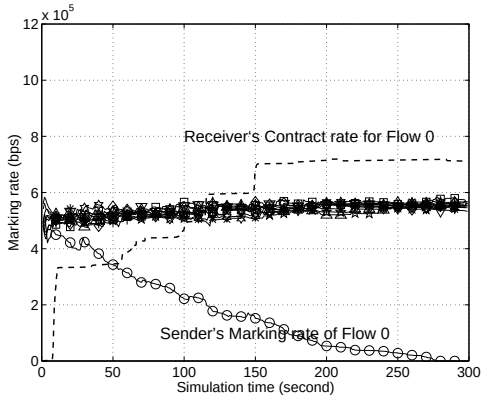



(a) Marking rate

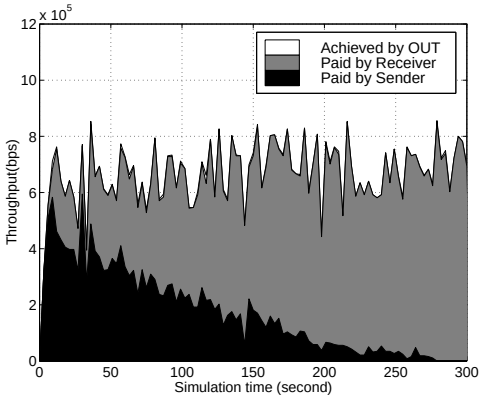



(b) Throughput

Fig. 20. Marking rate and throughput when the bandwidth is sufficient




(a) Marking rate



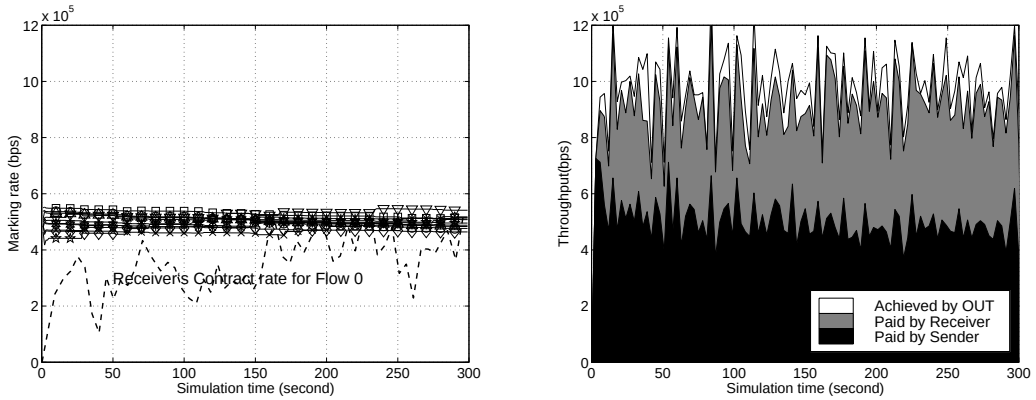

(b) Throughput

Fig. 21. Marking rate and throughput when the bandwidth is not sufficient

flows have not reached their targets.

Fig. 21 shows the result of the second scenario. Compared to Fig. 20(b), it is observed that throughput is achieved by only IN packets in Fig. 21(b) because the network bandwidth is not enough. Hence, in Fig. 21(a), the receiver stops transferring resources at around 0.7 Mbps to avoid resource wastage even though the throughput does not reach the target rate of 1 Mbps. Again, we observed that the sender moves its resources to other flows since other flows have not reached their targets.

Fig. 22 shows the result of the third scenario. In Fig. 22(a), note that the individual sender's marking rate stays around 0.5 Mbps. The sender's marker continues allocating 0.5 Mbps to this flow since all the flows have exceeded their target rates. The resources contributed by the receiver stay around 0.4




(a) Marking rate (b) Throughput

Fig. 22. Marking rate and throughput when the bandwidth is plentiful

Mbps since this amount is enough to reach the target rate.

## 6 Discussion and Related work

Our simulation experiments on sender's marking strategies show that: (a) Aggregate source can effectively manage resources by marking packets of individual sources differently. (b) Proportional marking of flows within an aggregation is shown to be ineffective in competing with the other strategies. (c) The adaptive marker performance is impacted by choosing the adaptation rate and the observation period. (d) The adaptive marking strategy is effective to deal flows with different RTTs and unresponsive flows.

Our results on receiver-based strategies show that: (a) A simple technique of transferring bandwidth from receiver to the sender's side can be effective in improving the performance seen by a receiver, (b) the sender can exploit receiver's willingness to pay by moving its resources to other flows within its aggregation, and (c) if sender employs proportional marking, a receiver willing to pay for improved service can extract a higher amount of the sender's bandwidth for its flow.

Pricing [16,17] will have an important effect on a number of the above observations, specifically on the nature of moving up the service levels. It has been suggested [18] that resources should be priced based on the level of congestion to balance load evenly across the network links. Network providers may employ fair-sharing techniques to balance resource utilization among competing aggregated sources at the time of congestion. In such a case, shifting resources to congested links will likely have less impact than observed in this study. If

out-of-profile packets are dropped at the network edge, end sources cannot exploit the resource management strategies presented here.

Aggregated sources pose interesting new questions. If the aggregated source readjusts its resources among the individual flows, even when an individual flow backs off, it is likely that another flow within the aggregation may send more packets through the congested link. All of these issues point to the need for studying network bandwidth management and network dynamics further when bandwidth can be aggregated and flexibly allocated at the edge of the network.

Recent work on diff-serv networks dealt with individual sources [6–8]. Adaptive marking to achieve throughput targets for single sources is studied in [8]. Unfair resource sharing among responsive and unresponsive flows has been studied, and a fair marking scheme has been proposed in [13]. Aggregation of individual traffic sources and the resulting traffic distributions have been studied [14,15]. Our earlier work studied fairness within an aggregation [19]. This paper considered a more general resource (IN-profile bandwidth) management problem in meeting QOS goals of individual flows.

## 7 Conclusions

In this paper, we have studied how QOS of individual flows could be improved by marking strategies employed by an aggregated source. We have proposed new marking algorithms that enable reaching specific QOS goals of individual flows within the aggregation. We have shown that proportional marking of packets within an aggregation can lead to significant disadvantages when other sources employ different marking strategies. We have presented simulation results to show the impact of the proposed algorithms on realized throughputs, network congestion and the scalability of the proposed approaches. Proposed adaptive marking scheme is shown to provide predictable and robust performance. With the proposed receiver-based scheme, the receivers can achieve improved service even in a sender-controlled network. We have shown that the amount of bandwidth paid by the receiver can be impacted by the sender's marking scheme and observed network conditions.

## 8 Acknowledgments

The authors would like to thank anonymous reviewers for their useful comments on this work.

## References


[1] K. Nichols, S. Blake, F. Baker, and D. L. Black, "Definition of the Differentiated Service Field (DS Field) in the IPv4 and IPv6 Headers," RFC2474, Network Working Group, December, 1998.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, "An Architecture for Differentiated Services," RFC2475, Network Working Group, December, 1998.

[3] J. Heinanen, F. Baker, W. Weiss and J. Wroclawski, "Assured Forwarding PHB Group," RFC2597, Network Working Group, June, 1999.

[4] V. Jacobson, K. Nichols and K. Poduri, "An Expedited Forwarding PHB," RFC2598, Network Working Group, June, 1999.

[5] V. Jacobson, K. Nichols and K. Poduri, "The Virtual Wire Behavior Aggregate," INTERNET DRAFT, IETF, March, 2000.

[6] D. D. Clark and W. Fang, "Explicit Allocation of Best-Effort Packet Delivery Service," November, 1997. http://diffserv.lcs.mit.edu/exp-alloc-ddc-wf.pdf

[7] A. Basu and Z. Wang, "A Comparative Study of Schemes for Differentiated Services," *Bell labs Technical Report*, Aug. 1998.

[8] W. Feng, Dilip D. Kandlur, D. Saha, and Kang G. Shin, "Adaptive Packet Marking for Providing Differentiated Services in the Internet." *Proc. of Int. Conf. on Network Protocols*, October 1998.

[9] Network simulator (Ns), University of California at Berkeley, CA, 1997. Available via http://www-nrg.ee.lbl.gov/ns/.

[10] W. Feng, Dilip D. Kandlur, D. Saha, and Kang G. Shin, "Understanding TCP Dynamics in an Integrated Services Internet," *Proceedings of the International Workshop on NOSSDAV*, May, 1997.

[11] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: A New Resource ReSerVation Protocol," *IEEE Transaction on Networking*, September, 1993.

[12] B. Suter, T. V. Lakshman, D. Stiliadis and A. Choudhury, "Design considerations for supporting TCP with per-flow queueing," *Proc. of INFOCOM'98*, April, 1998.

[13] I. Andrikopoulos, L. Wood and G. Pavlou, "A Fair Traffic Conditioner for the Assured Service in a Differentiated Services Internet," *IEEE ICC 2000*, June, 2000.

[14] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level," *Proc. ACM SIGCOMM '95*, 1995.

[15] M. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *Proc. ACM SIGMETRICS '96*, May 1996.

[16] R. Cocchi, D. Estrin, S. Shenker, and L. Zhang. "Pricing in Computer Networks: Motivation, Formulation and Examples," *ftp://parc.ftp.xerox.com/pub/net-research/policy.ps.Z.*

[17] D. Clark. "A model for cost allocation and pricing in the internet," *MIT workshop on Internet Economics*, 1995.

[18] I. Stoica and H. Zhang. "LIRA: An approach for service differentiation in the internet," *Proc. of NOSSDAV '98*, June 1998.

[19] I. Yeom and A. L. N. Reddy. "Impact of marking strategy on aggregated flows in a differentiated services network," *Proc. of IWQoS 99*, June 1999.

[20] I. Yeom and A. L. N. Reddy. "Modeling TCP Behavior in a differentiated services network," Tech. Report, Texas A&M University, May 1999.