Adaptive Marking for Aggregated Flows

Ikjun Yeom
Center for Next Generation Information Technology
Kyungpook National University
Taegu, South Korea
ikjun@ee.knu.ac.kr

A. L. Narasimha Reddy
Department of Electrical Engineering
Texas A & M University
College Station, TX 77843-3128
reddy@ee.tamu.edu

Abstract— **The differentiated services architecture is receiving wide attention as a framework for providing different levels of service according to a service profile in the Internet. The current architecture allows aggregated flows sharing a service profile. This paper looks at the problem of achieving specific QoS goals of individual flows by flexibly managing resources available to an aggregated source. We derive a simple analytic model for the relationship between per-session behavior, aggregate packet marking and packet differentiation within a diff-serv network. The paper presents an adaptive marker based on a TCP performance model within a diff-serv network. The paper shows that an aggregated marker 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.**

I. Introduction

The differentiated services (diff-serv) architecture is receiving wide attention as a proposal to provide different services over networks in a scalable manner [1], [2]. Currently, there are two PHBs (Per-Hop Behaviors) standardized by the Internet Engineering Task Force (IETF). Expedited Forwarding (EF) PHB provides guaranteed QoS services such as low delay, low jitter and low loss rate [3]. Assured Forwarding (AF) PHB provides *better service* than best effort according to user's service profile [4]. In this paper, we focus on AF PHB.

In an AF PHB domain, 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. This preferential drop mechanism is expected to provide better throughput for IN packets than OUT packets. The diff-serv architecture allows aggregated sources as well as individual sources.

Recent work on diff-serv networks mostly dealt with individual sources [5], [6], [7] 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. Adaptive marking to achieve throughput targets for single sources is studied in [7]. In [9], we proposed two new marking schemes: *IN-fair* and *BW-fair* marking. IN-fair marking results in fair sharing of contract rate among flows in an aggregation. Fair allocation of contract rate also has been studied recently in [11]. BW-fair marking aims to achieve equal throughput for all the flows within an aggregation.

In this paper, we address a more general problem: given individual target rates for each flow, how to allocate a fixed contract rate among flows within an aggregation under dynamic network conditions? We propose an adaptive marker that relies on the TCP performance model developed by us in [10]. This paper makes the following significant contributions: (1) establishes the relationship between per-session behavior, aggregate packet marking and packet differentiation within a diff-serv network. (2) presents extensive simulations to study behaviors of individual flows followed by an aggregated marker. (3) proposes an adaptive marking scheme for aggregated flows and evaluates it through extensive simulations.

II. Behavior of proportional marking

We first describe the problem of packet marking when a marker does not maintain per-flow state. We call this marker a *proportional marker* since a packet is marked by aggregated sending rate, and consequently, contract rate[1] is distributed to individual flows proportional to their sending rates. Fig. 1 shows the conceptual model which we look at in this section. The marker maintains only the aggregated sending rate.

Suppose that there are flows sharing aggregated contract rate, M, and the current total sending rate of the flows is B.

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

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

[1] In this paper, we use *contract rate* to mean profile rate contracted by aggregated flows.

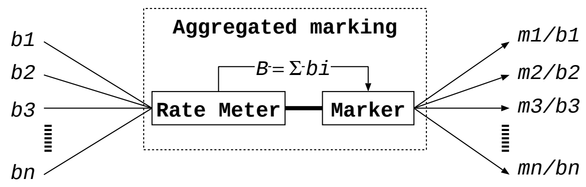


Fig. 1. Aggregated marking

where b_i is the individual sending rate of i^{th} flow, and n is the number of flows. When B is less than M, 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,

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

$$m_i = \frac{M}{B} b_i \tag{3}$$

where m_i is the marking rate of $i^t h$ flow. Here note that M/B is the same to every individual flow within the aggregation. Thus, m_i is proportional to b_i.

Now we present a summary of a throughput model for TCP flows sharing a contract rate originally proposed in [10]. In this model, we assume that a TCP receiver does not employ delayed-ACK and IN-profile packet is not dropped[2]. We also assume that there is no time-out in TCP flows for simplicity. This assumption may cause overestimation in throughput, but it is not derived to a closed form with time-out.

First, we begin with a simple throughput model for an individual TCP flow.

$$b_i = \frac{3}{4} m_i + \frac{3k_i}{4\mathrm{RTT}} \sqrt{\frac{2}{p_i}} \tag{4}$$

where p_i is the drop probability of OUT packet, and k_i is the packet size. Then, the sum of b_i, B is given by

$$B = \sum^{i} b_i = \frac{3}{4} M + \sum^{i} \frac{3k_i}{4\mathrm{RTT}_i} \sqrt{\frac{2}{p_i}} \tag{5}$$

Let $c_i = \frac{k_i}{\mathrm{RTT}_i} \sqrt{\frac{2}{p_i}}$, and using (4) and (3), b_i is expressed by

$$b_i = \frac{3M}{4B} b_i + \frac{3}{4} c_i \tag{6}$$

$$= \frac{3 c_i B}{4B - 3M} \tag{7}$$

Applying (5), finally we have

$$b_i = \frac{3c_i}{4\sum_{i=1}^{n} c_i} M + \frac{3}{4} c_i \tag{8}$$

[2] In a diff-serv domain configured appropriately, IN-profile packet is expected to be transmitted without being discarded.

So far, we have presented a simple model for aggregated flows in a diff-serv network. This model reveals that throughput of an individual TCP flow b_i is affected by other flows (more precisely $\sum c_i$). This property of the proportional marking may cause unstable state of individual TCP flows and degrade QoS consequently.

III. Adaptive Marking for Aggregated Flows

In this section, we discuss how to control individual throughputs with marking rates within an aggregation and propose an adaptive marking strategy. The most desirable situation is clearly to guarantee individual target rates for all the individual flows. However, it is also clear that there exist situations in which some targets cannot be reached: (i) When there is a severe congestion along the path and the current available bandwidth is less than the target. (ii) When the contract rate is not enough to achieve the target. If we try to achieve the target by increasing marking rate of an individual flow observing the case (i), it makes the congestion more severe and results in resource wastage. This is the very undesirable situation for both customers and service providers.

We propose an adaptive marking scheme for aggregated flows which guarantees at least one of the following to all the aggregated flows:

1. Individual target rate when it is reachable.
2. Maximized throughput without IN packet loss when the current available bandwidth is less than the individual target rate.
3. Throughput achieved with M/n marking rate where M and n are the total marking rate and the number of flows within the aggregation, respectively.

These three goals correspond to (1) meeting individual flow's BW needs, (2) maximization of utility of the aggregated contract rate and (3) fairness among the flows within the aggregation.

Initially, we set the marking rate of each flow proportional to its target. 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[3] 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 the

[3] In [10], 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 an oversubscribed network, a flow usually experiences some number of IN packet losses.

At every observation period:
1.**for** $i \leftarrow 1$ to n
2. **if** $0.75m[i] < b[i] < t[i]$
3. $m[i] = m[i] + \Delta(b[i] - t[i])$
4. **else if** $b[i] \leq 0.75m[i]$
5. $m[i] = m[i] - \Delta(0.75m[i] - b[i])$
6. **else if** $b[i] > t[i]$
7. $m[i] = m[i] - \Delta(b[i] - t[i])$
8.**Do** Max-Min fair with $m[i]$ and M
$m[i]$: 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. 2. An algorithm for adaptive marking

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 [10]. From the model, throughput B of a TCP flow experiencing oversubscribed network is given by

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

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 (9), when throughput achieved by a flow is less than $0.75m$, the flow should observe an oversubscribed network.

Hence, 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 $\sum m_i$ 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 reduces m_i to avoid wasting resources.

Fig. 2 shows an example algorithm for the adaptive marker.

Theorem: The adaptive marking algorithm finds m_i for which $b_i \geq \min\{t_i, b_{a,i}, b_{b,i}\}$ for $1 \leq i \leq n$ where t_i is the target rate, $b_{a,i}$ is maximum achievable rate such that $b_i \geq 0.75m_i$, and $b_{b,i}$ is rate achieved with M/n.

Assumption:
1. $\frac{\partial b_i}{\partial m_i} \geq 0$: Throughput of a flow does not decrease as the marking rate of the flow increses when other network condition is not changed.
2. $\frac{\partial^2 b_i}{\partial m_i^2} \leq 0$: There exists only one marking rate m_i at which $b_i = 0.75m_i$.
3. If $b_i < 0.75m_i$, there is IN packet drop.

Proof:
For each flow,

• When $\min\{t_i, b_{a,i}, b_{b,i}\} = t_i$;
If current b_i is less than t_i, b_i should be greater than $0.75m_i$ since t_i is less than $b_{a,i}$. Then, m_i increases from line 3. From the assumption that $\frac{\partial b_i}{\partial m_i} \geq 0$, b_i eventually reaches t_i. Since t_i is less than $b_{b,i}$, $m_{t,i}$, at which b_i is equal to t_i, is less than M/n, and line 8 does not change m_i.

• When $\min\{t_i, b_{a,i}, b_{b,i}\} = b_{a,i}$;
If current b_i is less than $b_{a,i}$, b_i is greater than $0.75m_i$ from the definition of $b_{a,i}$ and, at the same time, less than t_i since t_i is greater than $b_{a,i}$. Then, from line 3, m_i increases until b_i reaches $b_{a,i}$. Here also, since $b_{a,i}$ is less than $b_{b,i}$, $m_{a,i}$ is less than M/n, and line 8 does not change m_i.

• When $\min\{t_i, b_{a,i}, b_{b,i}\} = b_{b,i}$;
If current b_i is less than $b_{b,i}$, b_i is greater than $0.75m_i$ and less than t_i since $b_{b,i}$ is less than $b_{a,i}$ and t_i. Then, from line 3, m_i increases until M/n. Then, b_i reaches $b_{b,i}$.

Time complexity of this algorithm is $O(n \log n)$ where n is the number of flows, and this is allowable for an edge device marker. This algorithm can be invoked at a coarser level than on a packet's arrival at the edge device. In this algorithm, we use TSW [5] to smooth out the individual throughput.

IV. Achieving Target Rates

In this section, we show how the proposed marking scheme realizes achievable individual target rates and finds maximized throughputs when the target rates are not achievable.

We consider a multi-hop path as shown in Fig. 3. There are n routers, and cross traffic is injected to this network at the i^{th} router and exits at the $(i+1)^{th}$ router.

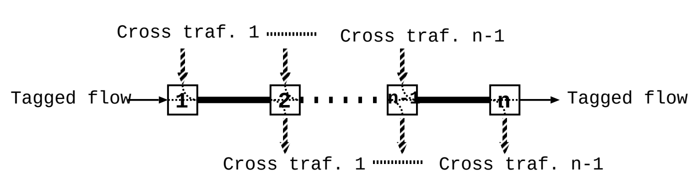


Fig. 3. Multi-hop topology

To observe how the marking rate is adjusted and an individual flow achieve its target rate, we conducted a set of simulations. In the simulation, we set the link capacity 3 Mbps and use 10 TCP flows for cross traffic. The contract rate for each TCP flow is randomly selected from 0 to 1 Mbps, and the total contract of cross traffic is 2.7 Mbps so that the subscription level is 90%. The number of routers (n) is five. For the tagged flow, we use single TCP flow.

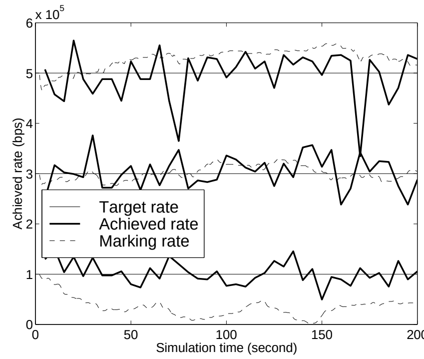
First, to observe path characteristics, we use static contract rate for the tagged flow. We vary the contract rate from 0 to 0.8 Mbps. In Fig. 4, the solid line shows the achieved rate with static marking rate, and the dashed line indicates the achieved rate is equal to the 75% of marking rate. It is clear that the achieved rate increases as the marking rate increases until 0.5 Mbps. It is also observed that after 0.55 Mbps the achieved rate does not incease even if we increase the making rate upto 0.8 Mbps. This observation supports our assumption that if we increase the marking rate more than the point in which the achived rate is the 75% of the marking rate, the flow observes oversubscribed path and wastes the marking rate. In thie example, the maximum achievable rate is about 0.42 Mbps.

Now the tagged flow is an individual flow within an aggregation with aggregated contract rate. The marker for the aggregation employs the adaptive marking. We vary the target rate for the tagged flow from 0.1 to 0.5 Mbps. Fig. 4 shows the results. In each figure, dots indicate instantaneous marking and achieved rate, and a square shows the average.

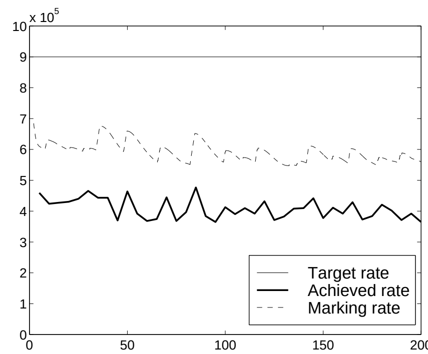
In this path, a flow gets 0.15 Mbps with zero contract rate. When the target rate is 0.1 Mbps (Fig. 4(a)), therefore, the marking rate stays around zero. When the target rate is achievable (less than 4.2 Mbps), it is observed that the adaptive marking scheme finds the minimum marking rate to realize the target (Fig. 4(a) to 4(b)). In Fig. 4(c), it is also observed that the marking rate stays less than 0.55 Mbps to avoid resource wastage when the target is unachievable.

So far, we have looked at throughput of an individual flow within an aggregation. Now we observe aggregated flows. There are nine flows aggregated. The contract rate for the aggregation is 5 Mbps. We set the individual target rates differently and offer different cross traffic at each link between R and a receiver to produce different network conditions.

Fig. 5 shows realized throughputs and marking rates of some individual flows. When the targets are achievable (Fig. 5(a)), the realized throughputs stay around their targets while the marking rates keep changing to adapt to the network conditions. When the target is unachievable (Fig. 5(b)), the marking rate is managed to maintain the achieved rate to be the 75% of the marking rate.



(a) Achievable target



(b) Unachievable target

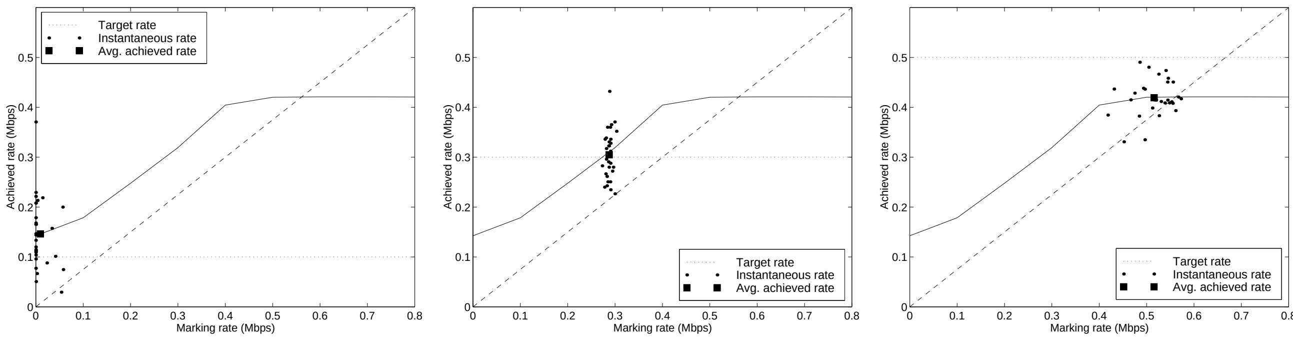
Fig. 5. Realized throughput with the adaptive marking

Table I shows the summary of the results. It is clearly shown that the adaptive marker avoids resource wastage ($p_{in} = 0$ for all the flows). At the same time, it maximizes throughput for the flows (4, 5 and 6 in the table) with unachievable targets ($p_{out} \approx 1$ and utilization by IN packet ≈ 1). The flows with infinite target rates are for simulating FTP kind of applications and consume the residual contract rates.

Now, we observe how the adaptive marker deals flows with different RTTs. To produce different RTTs for each flow, we use a topology in which 40 TCP flows are aggregated and compete a 25 Mbps bottleneck link. The aggregated contract rate is 10 Mbps. RTT (excluding queueing delay) of each flow is randomly selected from 50 to 150 msec. Fig. 6 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.

V. Conclusions

In this paper, we have addressed the important problem of establishing a relationship between per-session behav-



(a) Target = 0.1 Mbps (b) Target = 0.3 Mbps (c) Target = 0.5 Mbps

Fig. 4. Achieved rates with the adaptive marking

TABLE I
SIMULATION RESULTS WITH ADAPTIVE MARKING

Target rate	Marking rate	Achieved rate	p_{in}	p_{out}	Util. by IN
0.1	0.034	0.108	0	0.051	0.321
0.3	0.301	0.302	0	0.102	0.852
0.5	0.529	0.507	0	0.154	0.947
0.7	0.468	0.339	0	0.998	1
0.9	0.591	0.408	0	0.951	0.999
1.1	0.768	0.621	0	0.973	0.999
INF	0.769	0.808	0	0.006	0.794
INF	0.769	0.801	0	0.007	0.799
INF	0.769	0.921	0	0.003	0.611

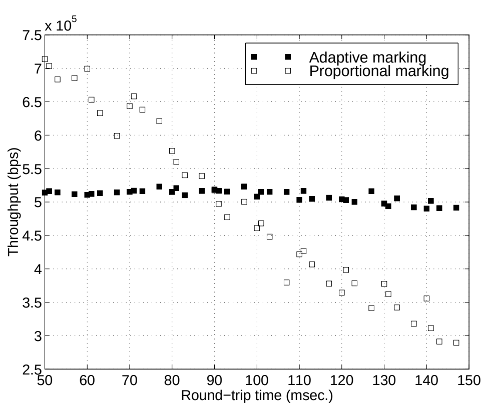


Fig. 6. Throughput of flows with different RTT

ior, aggregate packet marking and packet differentiation within a differentiated services network. We have presented analytical formulation of this relationship and validated this through extensive simulations. Our work has shown that proportional marking is ineffecient in managing aggregate resources.

We proposed a new adaptive marking algorithm based on the TCP performance model within diff-serv networks. The adaptive marker enables reaching specific QoS goals of individual flows while efficiently managing the aggregate resources. We have presented an extensive simulation study of the behavior of the adaptive marker under various parameters. Our simulation experiments on the adaptive marking strategy show that: (a) an aggregate source can effectively manage resources by marking packets of individual sources differently. (b) the adaptive marking strategy is effective in dealing with different network conditions such as different RTTs.

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, IETF, Dec. 1998.

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

[3] V. Jacobson, K. Nichols and K. Poduri, "An Expedited Forwarding PHB," RFC2598, IETF, June, 1999.

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

[5] D. Clark and W. Fang, "Explicit Allocation of Best-Effort Packet Delivery Service," *IEEE/ACM Trans. on Networking*, Aug. 1998.

[6] A. Basu and Z. Wang, "A Comparative Study of Schemes for Differentiated Services," *Bell labs Tech. Rep.*, Aug. 1998.

[7] W. Feng, Dilip D. Kandlur, D. Saha, and Kang G. Shin, "Adaptive Packet Marking for Providing Differentiated Services in the Internet." *Proc. of ICNP*, Oct. 1998.

[8] Network simulator v.2 (ns-2), Univ. of Cal. at Berkeley,1997.

[9] 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.

[10] I. Yeom and A. L. N. Reddy. "Modeling TCP Behavior in a Differentiated Services Network," *IEEE/ACM Trans. on Networking*, Feb. 2001.

[11] 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.