A Simulation Study on the Relation of DiffServ Packet Level Mechanisms and Flow Level QoS Requirements

Eeva Nyberg, Samuli Aalto, Riikka Susitaival
Networking Laboratory
Helsinki University of Technology
Corresponding author: Eeva.Nyberg@hut.fi

Abstract

We deepen our previous analytical studies to include simulations, with special considerations on the effect that the metering mechanisms at the DiffServ boundary nodes have on differentiation. The two metering and marking mechanisms compared are token bucket and exponential weighted moving average (EWMA). The conjecture made in our analytical models is that the EWMA principle marks packets *per flow*, while the token bucket is a *per packet* metering and marking mechanism.

We consider a single bottleneck link of a DiffServ cloud loaded by a fixed number of elastic TCP sources in congestion avoidance mode and with a similar round trip time (RTT). The sources are divided into L groups according to price paid by flows. At the boundary of the DiffServ cloud, the traffic of these TCP sources is conditioned, i.e. metered and marked to I priorities, i.e. flow aggregates based on price paid. Inside the DiffServ node the packets of the flow aggregates are forwarded or discarded by a scheduling unit that includes a single buffer with I priority levels.

As a result, we give further hindsight to the effect that separate DiffServ traffic-handling mechanisms have on the fairness perceived by the user. We use the division of bandwidth and the division of packets into priority levels as the decisive factors for our results.

Keywords: DiffServ, QoS, metering, marking, token bucket

1 Introduction

There is still not enough research on the coupling of packet level QoS mechanisms of Differentiated Services (DiffServ) [1], e.g. Assured Forwarding (AF) [2], to flow level analysis. On the other hand, flow level bandwidth allocation and fairness research, e.g. Kelly [4], Massoulié [7], and Veciana [11], continue to assume that weighted fair bandwidth allocations among flows in different service classes are somehow achieved and evade the question of how to do so without flow control or per flow scheduling.

We have previously analytically modeled DiffServ packet level mechanisms in order to gain insight into how packet level mechanisms achieve bandwidth division of flows. We have taken a multilevel approach; we model, both on the packet and the flow level, mechanisms that constitute such differentiation proposals as AF and Simple Integrated Media Access (SIMA) [6]. Our analytical models, [8] and [9], have demonstrated the effect that the flow conditioning mechanisms of the boundary nodes, e.g. marking, and aggregate forwarding actions inside the DiffServ nodes, e.g. scheduling and discarding, have on differentiation of flows. In the analytical model, we assume long time averages of the metering results, and based on this assumption study the effect of the DiffServ mechanisms.

In this paper, we deepen the study to include simulations, with special considerations on the effect that the metering mechanisms at the boundary nodes have on differentiation. The two metering and marking mechanisms compared are token bucket and exponential weighted moving average (EWMA). The hypothesis used in our analytical model is that the metering and marking result of EWMA, the so called momentary bit rate, can be used to mark packets *per flow*, while the token bucket is a *per packet* metering and marking mechanism.

In EWMA, the parameter α adjusts the memory of the mechanism. Marking is performed based on predefined thresholds of the momentary bit rate. The token bucket mechanism is implemented as a cascaded system of many token buckets; see [3] for an example. Each bucket has the same capacity c, but a specified rate in accordance to the marking thresholds of the EWMA system. Using the simulation model, we study the effect that the parameter values have on the metering and marking result and compare them to the conjecture of the analytical model.

As a result, we give further hindsight to the effect that separate DiffServ traffic-handling mechanisms have on the fairness perceived by the user. We use the division of bandwidth and the division of packets into priority levels as the decisive factors for our results.

The paper is organized as follows. Chapter 2 presents the overall picture of DiffServ mechanisms in a network. We then consider in Chapter 3, in more depth the conditioner situated at the boundary node. Chapter 4 illustrates the simulation setting, while chapter 5 introduces the analytical model. Chapter 6 presents the numerical results in terms of bandwidth allocation ratios and chapter 7 concludes the paper.

2 DiffServ modeling overview

The main elements of DiffServ are traffic classification and conditioning at the boundary nodes and traffic forwarding through scheduling and discarding at the DiffServ interior nodes. In addition, congestion control mechanisms designed for the Internet, such as TCP, and active queue management algorithms, such as RED, may be used for Quality of Service in the Internet.

2.1 DiffServ network building blocks

The traffic classification and conditioning, i.e. the division of flows into per hop behaviors (PHBs) and drop precedence levels inside PHBs, is done at the boundary nodes. After packets of flows have been classified, the traffic is conditioned, more specifically, traffic is metered and packets are marked into drop precedence levels.

Inside a DiffServ node all traffic-handling functions are performed on aggregates based on the given PHB, thus a PHB constitutes a forwarding class. With a DiffServ PHB group consisting of multiple PHBs the scheduling function is such that the PHBs are serviced by separate buffers with Weighted Fair Queuing (WFQ) scheduling principle [10]. Packets are discarded in the scheduling unit based on

the hard thresholds of the buffers. When the instantaneous buffer content exceeds a precedence threshold, only packets with higher precedence are accepted to the unit.
Figure 1 summarizes the components, of which the conditioner mechanism will be studied in this paper.

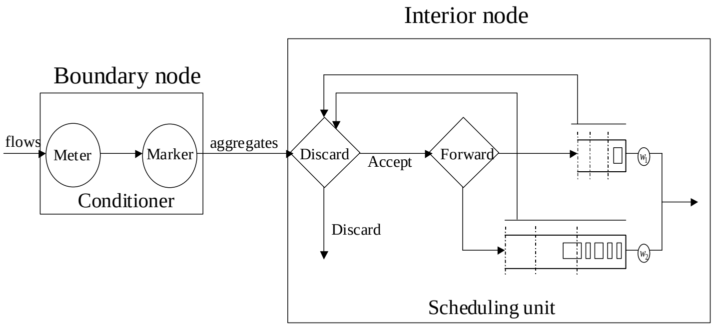

Figure 1: Components of a DiffServ network

2.2 Network and traffic model

In this paper, our purpose is to conduct a preliminary study concerning only elastic TCP traffic. Thus, we have just one forwarding class, or PHB, in our model.

We consider a single bottleneck link of a DiffServ cloud loaded by a fixed number of elastic TCP sources in congestion avoidance mode and with a similar round trip time (RTT). The sources are divided into L groups $l \in \mathcal{L} = \{1, 2, \ldots, L\}$, with all flows inside a group having weight $\varphi(l)$.

At the boundary of the DiffServ cloud, the traffic of these TCP sources is conditioned, i.e. metered and marked to I priorities, i.e. flow aggregates, denoted by $i \in \mathcal{I} = \{1, 2, \ldots, I\}$. The marking is made based on price paid or weights $\varphi(l)$ of the flow group l. Inside the DiffServ node the packets of the flow aggregates are forwarded or discarded by a scheduling unit that includes a single buffer with I priority levels.

3 Conditioner

Traffic is conditioned at the boundary node by first measuring the incoming traffic. Based on the metering result and the weight $\varphi(l)$ the flow or the packets are marked to precedence or priority levels $i \in \mathcal{I}$. In general, if the measured bit rate (*mbr*) of a flow is equal to the weight $\varphi(l)$, the flow is marked to at least middle priority. *The lower the bit rate, the higher the priority of the flow.* Therefore, the higher the weight of the flow, the more traffic it can carry, until marked to a lower priority level.

3.1 Metering and marking mechanisms

The two metering and marking alternatives considered are:

- **Token bucket:** Packets are marked in-profile if the bucket holds enough tokens upon arrival, out-of- profile otherwise.

- **Exponential weighted moving average:** The measured bit rate of previous time instants are exponentially dampened according to a time parameter α and the time interval that has passed since the measurement was done.

The token bucket or leaky bucket principle is a popular metering principle also referred to in the AF specification. For three precedence levels the metering and marking may be performed with two token buckets for each group *l*, with rates *r*(*l*,1) > *r*(*l*,2) and capacities *c*, as shown in Figure 2. If the first bucket does not have enough tokens at the arrival of a packet, the packet is out of profile and is marked to the lowest precedence level; in other cases the state of the second bucket determines the precedence level. If the second bucket does not have enough tokens at the arrival of the packet, the packet is out of profile and is marked to middle precedence, and if there are enough tokens both in the first and second bucket, i.e. the packet is in profile for both buckets, the packet is marked to the highest precedence level.

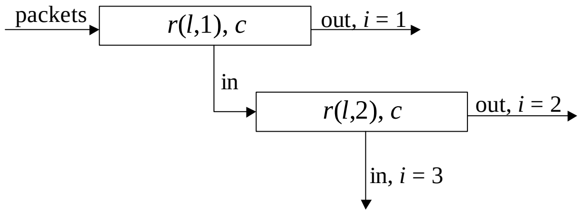

Figure 2: Token bucket scheme for marking packets of group *l* to three priority levels.

Exponential weighted moving average (EWMA) is another traditional metering principle. Its use for metering flow arrival rates was proposed, for example, in the SIMA specification. The measurement results of previous time instants are taken into account, but exponentially dampened according to a time parameter α and the time interval that has passed since the measurement was done. The marking is then performed based on predefined thresholds on the resulting measured arrival rate. Parameter α describes the time scale of the flow's bit rate measurement. According to [6], the measured bit rate of a flow *k* at the moment of transmission of the *j*:th packet is

$$\begin{aligned} mbr(k, j) &= \frac{\ln(1-\alpha)}{\ln(1-\alpha/\rho(k, j))} \\ \rho(k, j) &= \alpha + \rho(k, j)(1-\alpha)^{N_{kj}}, \end{aligned} \tag{1}$$

where N_{kj} is the distance between the *j*:th and *j*-1:th packet in time slots.

The priorities are determined based on thresholds. We allow the thresholds to depend on the group *l*, which the flow *k* belongs to, as follows.

$$\begin{aligned} t(l,0) &= \infty \\ t(l,i) &= \varphi(l)a(i), \quad i = 1,\ldots, I-1 \\ t(l,I) &= 0 \end{aligned} \tag{2}$$

The function *a*(*i*) could, for example, be defined as in SIMA, where

$$a(i) = 2^{I/2-i-0.5} \tag{3}$$

Note that $a(i-1)/a(i) = 2$ for all i. The j:th packet of flow $k \in l$ has priority i, if

$$t(l,i) \leq mbr\ (k,\ j) < t(l,i-1) \tag{4}$$

We will study the effect of the metering principles and their time parameters α and c, as both the memory or time span of the measurement as well as the dampening of bursts affect the way traffic is divided into precedence levels. Note, furthermore that α and c may depend on the forwarding class of the flow, but as we only consider one class, we will study the absolute value of the time parameters.

3.2 Metering and marking model

Assume that a metering principle exists. The measured traffic intensity then determines which precedence level the packets are marked to. We can model the resulting two marking alternatives as follows:

- **Per packet marking:** Only those packets of a flow that exceed the marking threshold are marked to the lower precedence level.
- **Per flow marking:** Once the measured load of a flow exceeds a marking threshold, all packets of the flow are marked to the same precedence level.

As an example, consider the case of three precedence levels and thresholds $0 = t(l,3) < t(l,2) < t(l,1)$. Denote by ν the bit rate of flow $k \in l$. With *per flow* marking, all packets of the flow have the same priority, corresponding to $\arg\min_i [\nu \geq t(l,i)]$. With *per packet* marking, the ratio of packets $(\min[\nu, t(l,i-1)] - \min[\nu, t(l,i)])/\nu$ have priority i. Figure 3 depicts the resulting marks given to the packets of this flow.

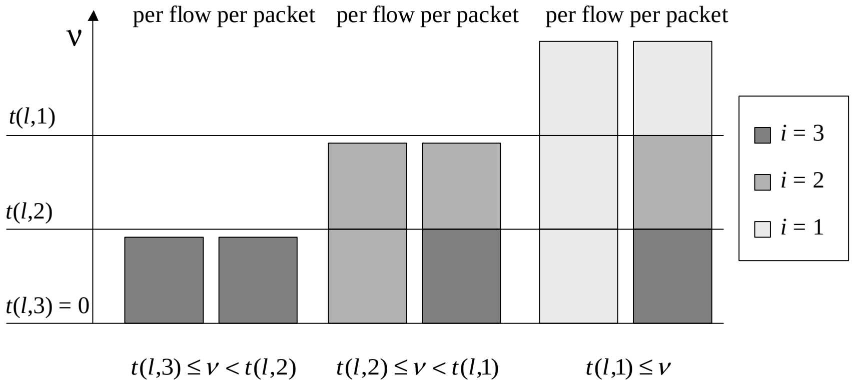

Figure 3: Differences in marking, for three priority levels.

The hypothesis of our model is that those metering and marking functions using the token bucket scheme mark packets to precedence levels *per packet*, while the use of exponential moving average of a flow allows the marker to mark packets *per flow*. One aim of our paper is to validate this conjecture using simulations.

4 Simulation setting

Our simulation setting is such that a single bottleneck link of a DiffServ cloud is loaded by a fixed number n of greedy TCP sources with a similar, but random round trip time (RTT). The TCP flows are modeled in congestion avoidance phase, i.e. whenever a packet is lost the corresponding source halves its window size; otherwise it is linearly increased. All flows are classified into the same forwarding class, and thus serviced by a single buffer. The buffer space is further divided by I discarding thresholds.

We wish to study the effect of the metering mechanisms and the time parameters on the system. The evaluation is made based on the resulting bandwidth allocation and division into priority classes. Note that we consider differentiation as a function of the number of sources, opposed to as function of the load of the network. This is due to the fact that we are considering elastic sources that adjust there sending rate, and thus adjust the load, according to the congestion of the network.

The marking mechanism assigns a drop precedence level to the flow or its packets. Then, depending on the congestion of the link, some packets may be discarded, namely when the priority of the packet is less than the threshold priority of the scheduler. The TCP mechanism adjusts its window size according to the feedback signal from the buffer system. By gathering information on how the stable sending rate and corresponding priority level allocation depends on the number of flows, we can study the differences between the marking schemes. Figure 4 summarizes the simulation setting.

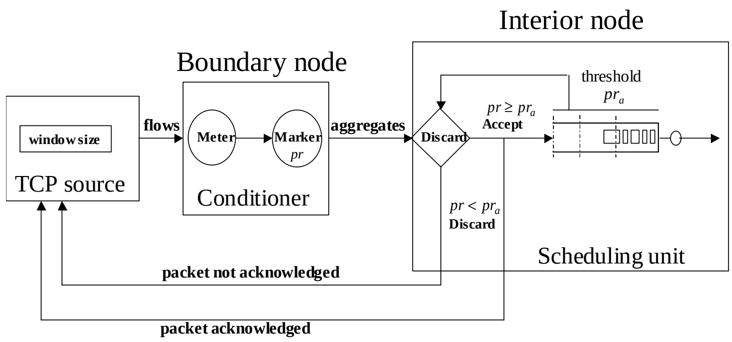

Figure 4: The simulation setting

4.1 Metering flows at the conditioner

Flows are metered at the boundary of the DiffServ cloud by either using token buckets or the EWMA principle. The corresponding time parameters c and α must be in accordance with the RTT of the packets.

For the token bucket metering, we fix the token bucket rate in accordance with the EWMA thresholds so that $r(l,i) = t(l,i)$, where $t(l,i)$ is given by equations (2) and (3) . For the EWMA metering, we calculate the momentary bit rate (mbr) of the flow as given by equation (1), with the following definition

$$\alpha = \frac{5}{K \cdot D},$$

where K is the size of the buffer in the scheduling unit and D is a free parameter. The priorities are determined based on the thresholds $t(l,i)$ as described by equations (2) - (4).

4.2 TCP model

The TCP sources are modeled in congestion avoidance mode, adjusting their sending rate according to the feedback signal from the forwarding unit.

RTT is the time that it takes for an acknowledgment to reach the TCP source, after the packet has been sent. We let RTT be an exponentially distributed random variable, and let the value of RTT denote the mean round trip time.

The window size is initialized to $w = 1$. It refers to the number of unacknowledged packets that can be sent at a time. Once a packet is sent, the counter *unack* is incremented by one. After time RTT/2 the packet reaches the scheduling unit of the interior node. If the packet is accepted to the scheduling unit, i.e. it has high enough priority not to be discarded, the acknowledgement reaches the source after time RTT/2 and the *unack* counter is decremented by one. The window size is updated to

$$\begin{aligned} inc &= 1/w \\ w &= w + inc \end{aligned}.$$

The number of new packets sent after the update is $\lfloor w \rfloor - unack$.

On the other hand, if the priority of the packet is too low it is discarded and not accepted to the scheduling unit. After time RTT/2 the information reaches the source. Then the window size is halved and the *unack* counter is decremented by one. If the counter *unack* is 0 after the halving, a new packet is sent. The window size is thus always at least 1.

4.3 Scheduling unit

The packets of the TCP sources are forwarded inside the DiffServ node by a scheduling unit consisting of one FIFO buffer and exponentially distributed service times. The buffer space is divided by I hard thresholds. Denote the discarding threshold of precedence level $i \in \mathcal{I}$ by K_i, with $K_I = K$, the size of the buffer. If the buffer state is $> K_i$, only packets with priority $> i$ are accepted to the scheduling unit, other packets are discarded.

5 Analytical model

In this section, we outline, at the packet level, an analytical model for the mechanisms inside the DiffServ node. We combine three separate models: a conditioner model, a TCP model and a buffer model. The model appeared first in [8] and [9].

As in the simulation study, let us consider the single link network, with the link capacity scaled to one. The network is loaded by a fixed number of greedy elastic flows belonging to L different flow groups.

Figure 5 shows the modeling approach used for the analytical model. Each part is separately discussed below, starting with the marking at the conditioner.

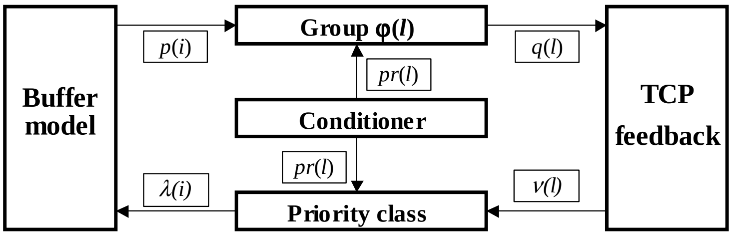

Figure 5: Analytical model setting

5.1 Marking flows at the conditioner

We compare the marking models described earlier, *per flow* and *per packet* marking. Assume that the metering result gives the packet arrival intensity $\nu(l)$ corresponding to a long-term average of the bit rates of individual flows belonging to group l. The flows of group l are assigned precedence $pr(l)$ according to the ratio of the arrival intensity to the weight of the flow, as proposed in [6].

$$pr(l) = \max\left[\min\left[\left\lfloor I/2 + 0.5 - \frac{\ln \nu(l)/\varphi(l)}{\ln 2}\right\rfloor, 1\right], I\right]. \tag{5}$$

The corresponding thresholds are as given in equations (2) and (3).

Marking all packets of a flow to the same precedence level gives the aggregate arrival intensities $\lambda(i)$ of precedence level i as

$$\lambda(i) = \sum_{l \in \mathcal{L} : pr(l) = i} n_l \nu(l). \tag{6}$$

From the buffer model of section 5.3 we get the loss probabilities $p(i)$ for each precedence level i. Under the *per flow* marking mechanism, the packet loss probability, $q(l)$, for a flow belonging to group l is then

$$q(l) = p(pr(l)),\ l \in \mathcal{L}. \tag{7}$$

Marking only overflow packets to the lower precedence level, i.e. with *per packet* marking, gives the aggregate arrival intensity $\lambda(i)$ of precedence level i as

$$\lambda(i) = \sum_{l \in \mathcal{L} : pr(l) = i} n_l \left(\min\left[\nu(l), t(l, i-1)\right] - \min\left[\nu(l), t(l,i)\right]\right). \tag{8}$$

The corresponding packet loss probability, $q(l)$, for a flow belonging to group l is

$$q(l) = \sum_{i=1}^{I} p(i) \frac{\min\left[\nu(l), t(l, i-1)\right] - \min\left[\nu(l), t(l,i)\right]}{\nu(l)},\ l \in \mathcal{L}. \tag{9}$$

5.2 The TCP feedback mechanism

For the analytical model, we have a basic TCP model where the TCP mechanism is assumed to be in congestion avoidance mode, resulting in the differential equations discussed in [5] for aggregates of flows. Furthermore, we assume that the dynamics of the buffer is faster than that of TCP resulting in the equilibrium relation between arrival intensity, round trip time (RTT) and loss probability of the

TCP mechanism, namely the stable point [5]. In the case of many flow groups and precedence levels, and thus differing loss probabilities, the stable point is formulated for each group separately,

$$\nu(l) = \frac{1}{RTT}\sqrt{2\frac{1-q(l)}{q(l)}}, \quad l \in \mathcal{L}, \tag{10}$$

where $q(l)$ is defined by equations (7) or (9).

5.3 Buffer model for one forwarding class

The analytical model for the one buffer system servicing elastic TCP traffic has discarding thresholds K_i, for $i \in \mathcal{I}$. The packet transmission time with full link rate is assumed to be exponentially distributed with mean $1/\mu$. Let $\lambda(i)$ denote the arrival rate of packets in precedence level i. Define the cumulative sum of arrival intensities of those precedence levels with index i or greater as λ_i. The buffer can then be modeled as an M/M/1/K queue with state dependent arrival intensities as depicted in Figure 6.

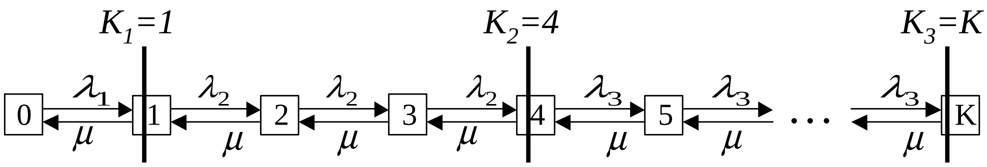

Figure 6: State transition diagram for one buffer modeled as an M/M/1/K queue.

The stationary distribution of the system can easily be solved analytically, with the resulting stationary probabilities π_j depending on the arrival intensities $\lambda(i)$. The probability $p(i)$ that packets belonging to precedence level i will be lost is

$$p(i) = \sum_{j=K_i}^{K} \pi_j \text{ , for } i = 1,\ldots,I \text{ .} \tag{11}$$

The equilibrium bandwidth allocation $\nu(l)$ for each group is solved from the set of equations (6)/ (8), (7)/(9), (10) and (11).

6 Numerical results

The basic scenario is as follows: two flow groups, $L = 2$, three precedence levels, $I = 3$, $\mu = 1$ packet/time unit and mean RTT = $50/\mu$, $100/\mu$ or $1000/\mu$. Set $K_1 = 22$, $K_2 = 33$ and $K = 39$ packets. Define $k = \varphi(2)/\varphi(1)$.

We study the equilibrium bandwidth allocation in terms of the ratio of throughputs and ratio of offered traffic between flows in the two groups. In the simulations, we consider the two metering mechanisms: token bucket and EWMA. In the analytical model, we have the token bucket mechanism modeled as *per packet* marking and EWMA modeled as *per flow* marking.

We wish to validate the assumptions of the analytical model and study the effect of the time parameters c and α using simulations.

6.1 Simulation results

We study the effect of the time parameters α and c in parallel to the round trip time RTT.

6.1.1 EWMA parameter α

The set of figures depicted in Figure 7 summarizes the relationship between α and RTT. Note that we vary the free parameter D of α. The relationship is shown in terms of the ratio of throughputs for flows with weights ($\varphi(1)$, $\varphi(2)$) = (0.04,0.08), i.e. $k = 2$, as a function of number of flows n_1 and n_2. The black areas correspond to equal bandwidth allocation, the white areas correspond to bandwidth allocation at least equal to k, and the gray areas to bandwidth allocation between 1 and k.

Table 1 gives the corresponding numerical values for ratios of throughput and Table 2 the numerical values for ratios of offered traffic.

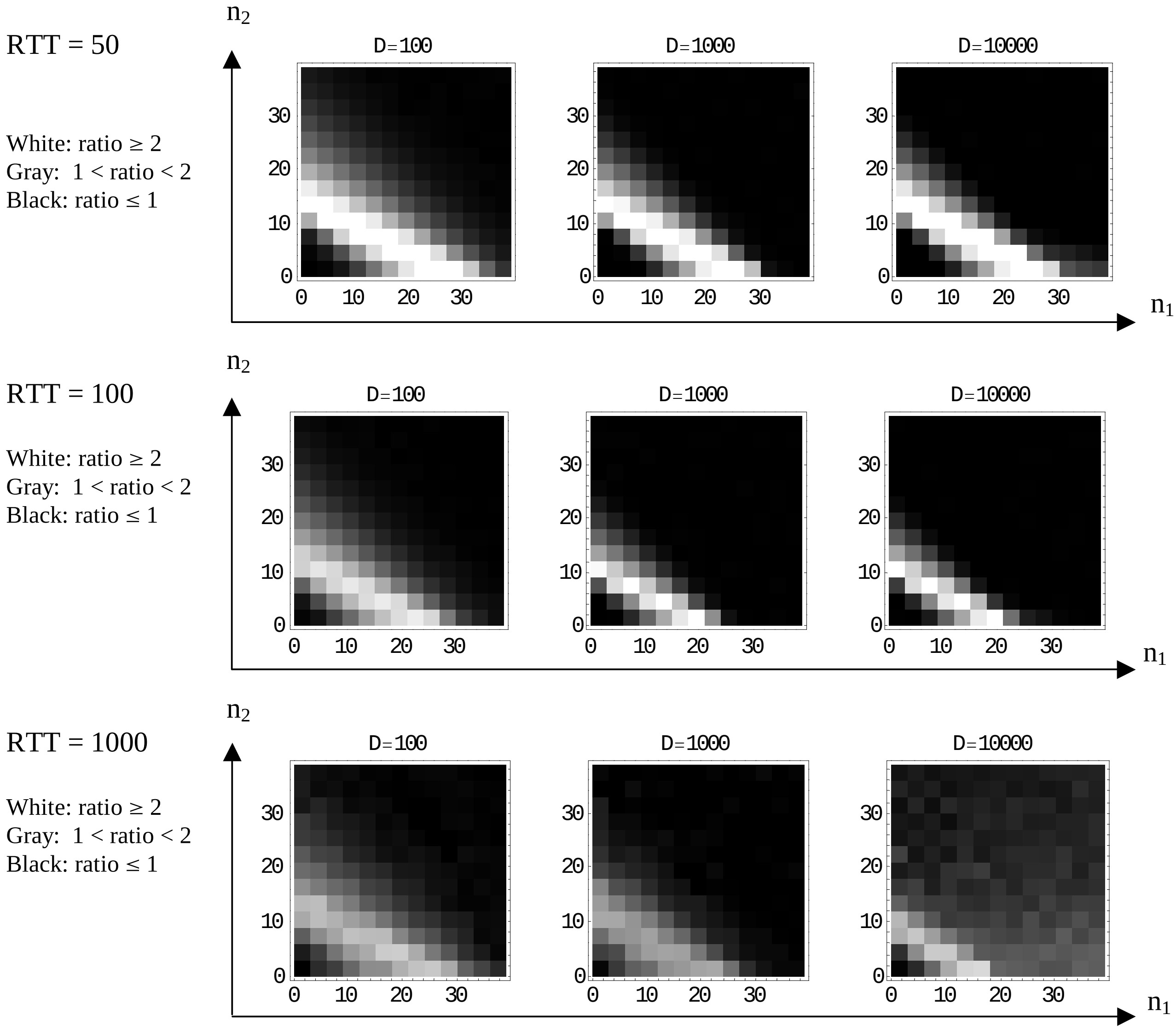

Figure 7: Ratios of throughputs for I = 3, RTT = 50, 100 and 1000. The time parameter has values D = 100, 1000 and 10000.

RTT		D=100	D=1000	D=10000
50	Min Max	0.978437 2.31008	0.978525 2.48056	0.988121 2.58277
100	Min Max	0.982882 1.92585	0.983543 2.10297	0.988843 2.16211
1000	Min Max	0.951966 1.80139	0.95911 1.65922	1.01888 1.84981

Table 1: The minimum and maximum values of the ratios of throughputs in Figure 7.

RTT		D=100	D=1000	D=10000
50	Min Max	0.988076 1.88536	0.987168 1.98159	0.993539 2.01127
100	Min Max	0.985907 1.79279	0.986997 1.95817	0.990926 2.01054
1000	Min Max	0.95285 1.79738	0.959991 1.65611	1.01887 1.84697

Table 2: The minimum and maximum values of the ratios of offered traffic in Figure 7.

From Figure 7 we make the following observations on the EWMA principle and the parameter D.

1. The minimum ratio of throughputs and offered traffic is always approximately 1.
2. The maximum ratio of throughputs and offered traffic is always of the same order as k.

Thus we can deduce that a flow inside group l receives maximum bandwidth in proportion to price paid, i.e. in proportion to the weight $\varphi(l)$.

3. There is a clear dependency between the two time parameters D and RTT.

As RTT decreases, the area of differentiation widens and deepens, i.e. the maximum ratio increases. The free parameter D is also sensitive to the time scale of the simulation, the RTT. If we wish that the maximum differentiation ratio is equal to the ratio k of the weights and the minimum differentiation is not less than 1, we could then choose $D = 100$ when RTT = 100. On the other hand, when RTT = 100, the loss probabilities are not negligible and if we want the offered traffic to have maximum ratio of k, we could choose, based on Table 2, $D = 1000$ or $D = 10000$ when RTT = 100.
Furthermore, we also made simulations with $D = 1$ and practically no area of differentiation resulted. Therefore, it is essential that the parameter D is not too small compared to RTT.

4. With three precedence levels, $I = 3$, only one area of maximum differentiation is obtained.

As will be shown later, in the analytical model there are two areas of differentiation when $I = 3$. However, by increasing the number of priority levels we are able to obtain a larger area of differentiation and more than one area of maximum differentiation. Figure 8 depicts the cases $I = 2, 3, 6$ and 8, with RTT = 100 and $D = 100$.

Table 3 gives the corresponding values for the minimum and maximum ratios of throughputs, and we note that for $I = 6$, the minimum value of the ratio of throughputs is 1,45. Therefore, the flow group with higher weight always receives more bandwidth than the group with lower weight.

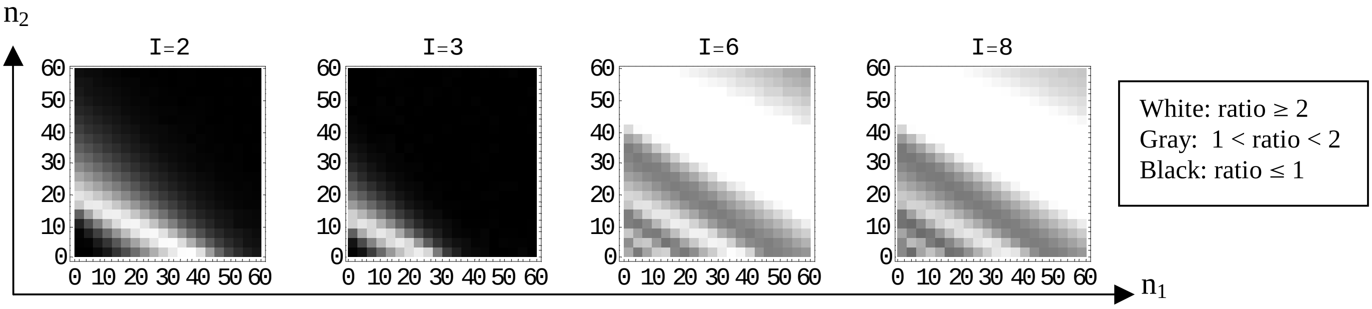

Figure 8: The ratio of throughputs, RTT = 100, *D* = 100 and *I* = 2, 3, 6 and 8.

I =	Min	Max
2	0.992562	1.98859
3	0.982882	1.92585
6	1.45289	2.39094
8	1.41926	2.39341

Table 3: The minimum and maximum values of the ratios of throughputs in Figure 8.

6.1.2 Token bucket parameter *c*

The set of figures depicted in Figure 9 summarizes the relationship between *c* and RTT. As with the EWMA principle, we study the ratio of throughputs for flows with weights (φ(1), φ(2)) = (0.04,0.08), i.e. *k* = 2, as a function of number of flows n_1 and n_2. The black areas correspond to equal bandwidth allocation, the white areas correspond to bandwidth allocation at least equal to *k*, and the gray areas to bandwidth allocation between 1 and *k*.

Table 4 gives the corresponding numerical values for ratios of throughput and Table 5 the numerical values for ratios of offered traffic.

RTT		C=100	C=1000	C=10000
50	Min Max	0,991143 1,59379	0,990008 1,59379	0,982195 1,60033
100	Min Max	0,961958 1,60769	0,966481 1,61122	0,976683 1,60857
1000	Min Max	0,926936 1,59202	0,904089 1,59522	0,937499 1,82334

Table 4: The minimum and maximum values of the ratios of throughputs in Figure 9.

RTT		C=100	C=1000	C=10000
50	Min Max	0,993397 1,47559	0,993311 1,47559	0,988377 1,48043
100	Min Max	0,967552 1,55598	0,970519 1,56103	0,979241 1,55799
1000	Min Max	0,927288 1,5914	0,904419 1,59459	0,93769 1,82228

Table 5: The minimum and maximum values of the ratios of offered traffic in Figure 9.

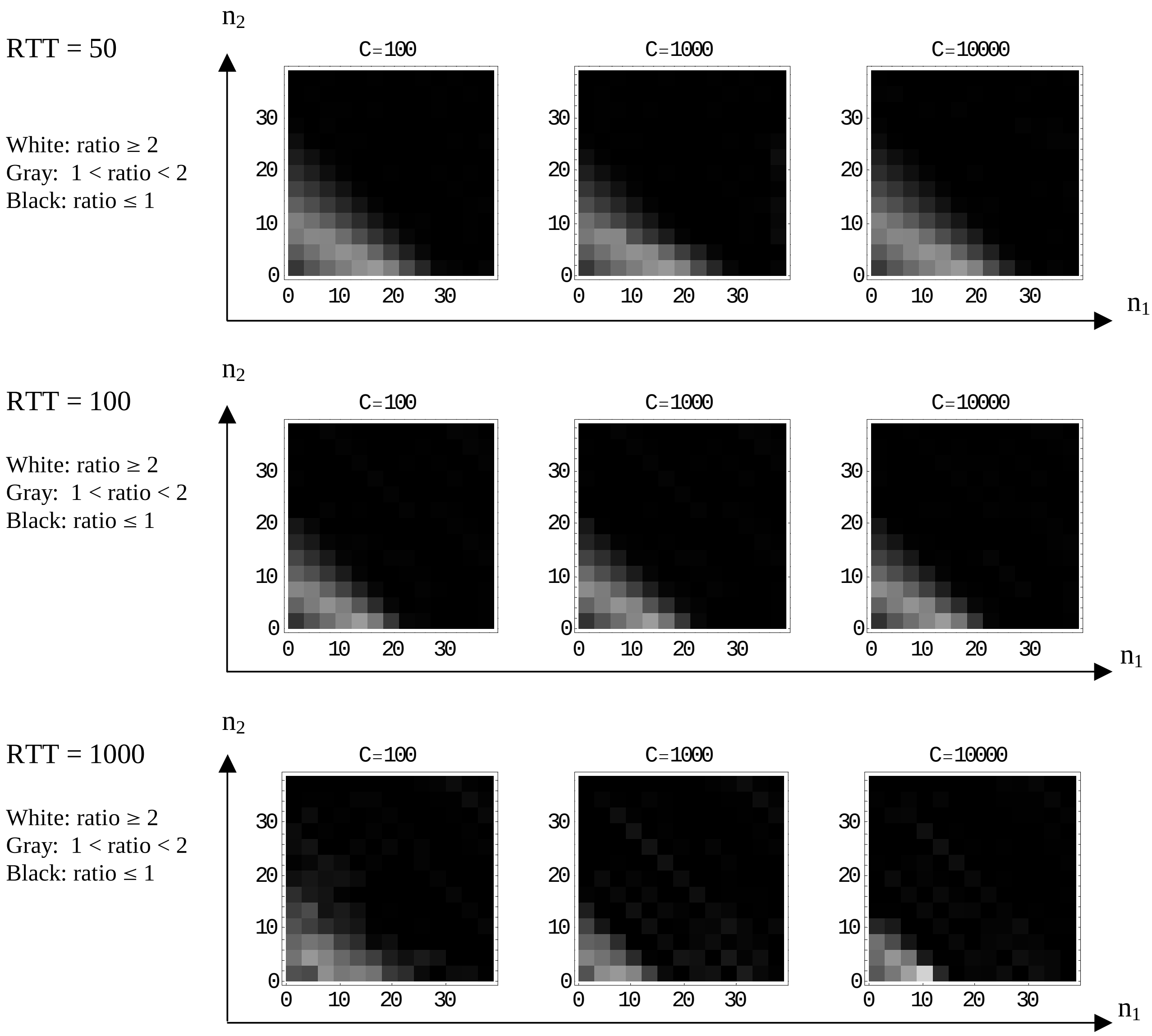

Figure 9: Ratios of throughputs for I = 3, RTT = 50, 100 and 1000. The time parameter has values c = 100, 1000 and 10000.

From Figure 9 we make the following observations on the token bucket principle and the parameter c.

1. The minimum ratio of throughputs and offered traffic is approximately 1, though not as clearly as with the EWMA principle.
2. The maximum ratio of throughputs and offered traffic is clearly less than k.

A flow inside group l does not receive maximum bandwidth in proportion to price paid, i.e. in proportion to the weight $\varphi(l)$, to the same extent as with the EWMA principle.

3. There is dependency between the time parameters c and RTT.

The main difference occurs, when RTT = 1000 and c changes from 1000 to 10000.
Furthermore, simulations made with c = 5 and RTT = 100 showed that the area of differentiation reduces as c decreases. Therefore, it is essential that the parameter c is not too small compared to RTT.

Though corresponding figures are not included in this paper, we have simulated the token bucket principle with more than three precedence levels. As a result, the areas of differentiation increased as *I* increased in the same way as in Figure 8.

6.2 Analytical model

The analytical model gives us the equilibrium offered traffic $\nu(l)$ and the corresponding throughput for each flow group. Using the ratio of throughput, we can study the effect that marking has in dividing bandwidth between elastic flows and compare the results given by the simulations.

In Figure 10 marking is *per flow*, while in Figure 11 marking is *per packet*. Both figures show the ratio, $\nu(2)/\nu(1)$, of offered traffic and the corresponding ratio of throughputs for flows with weights $(\varphi(1), \varphi(2)) = (0.04, 0.08)$, i.e. $k = 2$, as a function of number of flows n_1 and n_2. The round trip time is 100. The black areas correspond to equal bandwidth allocation, the white areas correspond to bandwidth allocation at least equal to k and the gray areas to bandwidth allocation between 1 and k.

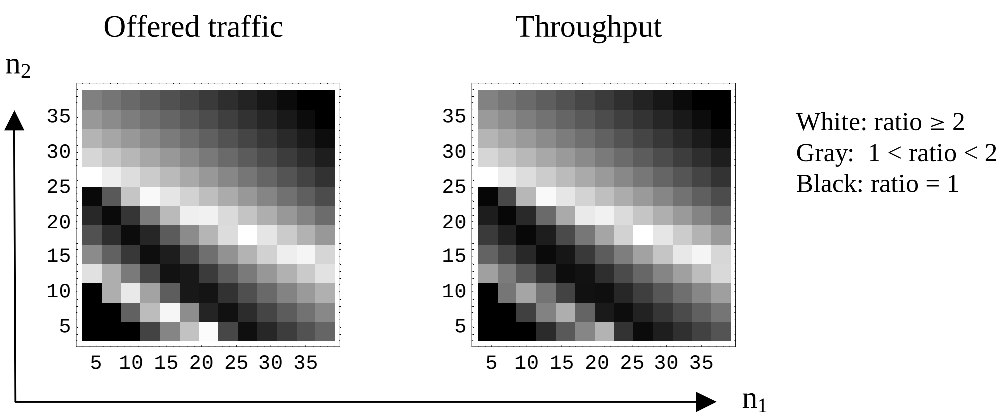

Figure 10: *Per flow* marking: Ratio of offered traffic (max=2) and ratio of throughput (max=2.67) for RTT = 100, *I* = 3.

The following observations can be made on Figure 10 and Figure 11:

1. When the marking is *per flow* a maximum ratio of *k* is achieved. On the other hand, when the marking is *per packet* the maximum bandwidth ratio $\nu(2)/\nu(1)$ is less than *k*.

The difference stems from the fact that in *per packet* marking some packets are always also marked to highest precedence.

2. Bandwidth is divided in equal proportions more often with the *per flow* marking scheme than with the *per packet* marking scheme.

When only the overflow packets are marked to lower precedence levels, the rest of the packets have highest or medium precedence and the ratio is higher than 1.

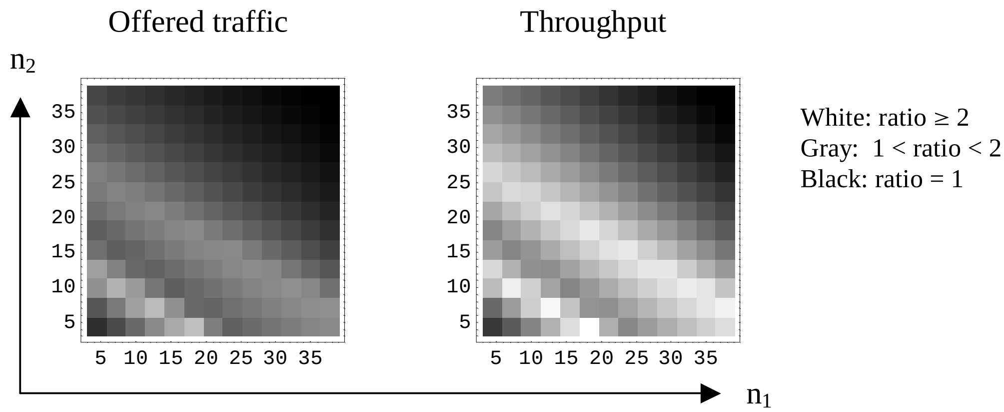

Figure 11: *Per packet* marking: Ratio of offered traffic (max=1.75) and ratio of throughput (max=1.86) for RTT = 100, I = 3.

Though the *per packet* scheme does not attain a maximum differentiation of k, its minimum differentiation is always larger than that of the *per flow* marking as long as enough precedence levels are employed.

6.3 Comparisons between simulation and analytical results

We note that the simulation and analytical results are the same qualitatively, though they differ a bit quantitatively. The following similarities were observed:

1. Both models result in bandwidth divisions with similar minimum and maximum values, as long as the parameters D and c used in the simulations are large enough compared to RTT.
2. Furthermore, simulations done using the token bucket mechanism give, in terms of minimum and maximum bandwidth division, the same results as analytical results for *per packet* marking. The same is observed for simulations using the EWMA principle and analytical results using *per flow* marking.
3. The number of differentiation areas increases as I increases as shown by simulations in Figure 8 and by analytical results in [8].

The following disparity was observed

1. There are a different number of differentiation areas. For the analytical model there are always I-1 differentiation areas, while the simulations resulted in only one area of differentiation for I= 3.

The results thus confirm our hypothesis that the token bucket and EWMA methods can be modeled as *per packet* and *per flow* metering and marking mechanisms, respectively. Furthermore, the analytical results are qualitatively consistent with the simulation outcome and can be used to gain insight to how differentiation mechanisms should be designed.

7 Conclusions

We have demonstrated several aspects of importance to the study of differentiation mechanisms.

1. We showed that marking and metering flows to I priorities with I-1 cascaded token buckets can be modeled as *per packet* marking, where only those packets exceeding a predefined threshold are

marked to lower priority. We also showed that the use of the EWMA principle in measuring the bit rate is able to capture the flow rate and the resulting marking is *per flow*, where all packets of the flow are marked to the same precedence level when the measured bit rate of a flow exceeds the predefined threshold.

2. We studied the effects that the time parameters α and c have on the metering and resulting differentiation of flows compared to the round trip time. We showed that the metering has to be done on the same time scale as the RTT is, for differentiation to occur.
3. Finally, we showed the applicability of our simple analytical models in gaining qualitative results on the various differentiation mechanisms needed for fair bandwidth division among flows.

Further research has to be done in explaining the differences between the simulations and the analytical results, namely the difference in the number and position of the differentiation areas. We will also simulate the case of multiple forwarding classes and the effect that the presence of non-TCP traffic has on differentiation as has been done analytically in [8] and [9].

8 References

[1] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, An Architecture for Differentiated Service, Dec 1998, RFC 2475

[2] J. Heinänen, F. Baker, W. Weiss and J. Wroclawski, Assured Forwarding PHB Group, Jun 1999, RFC 2597

[3] J. Heinänen and R. Guerin, A Two Rate Three Color Marker, Sep 1999, RFC 2698.

[4] F. Kelly, "Charging and rate control for elastic traffic", *Eur. Trans. Telecommun*, vol. 8, pp. 33-37, 1997.

[5] F. Kelly, "Mathematical modelling of the Internet", in *Proceedings of Fourth International Congress on Industrial and Applied Mathematics*, pp.105-116, 1999.

[6] K. Kilkki, Simple Integrated Media Access, 1997, available at http://www-nrc.nokia.com/sima.

[7] L. Massouliè and J. Roberts, " Bandwidth Sharing: Objectives and algorithms", in *Proceedings of IEEE INFOCOM* , 1999, pp. 1395-1403.

[8] E. Nyberg, S. Aalto and J. Virtamo, "Relating flow level requirements to DiffServ packet level mechanisms", *COST 279*, TD(01)04, 2001, available at *http://tct.hut.fi/tutkimus/cost279/*.

[9] E. Nyberg, S. Aalto, "How to achieve fair differentiation", *Submitted for publication*, 2001, available at *http://tct.hut.fi/tutkimus/cost279/*.

[10] A.K. Parekh and R.G. Gallagher, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case", *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 344-357,1993.

[11] G. de Veciana, T.-J. Lee, T. Kontantopoulos , "Stability and Performance Analysis of Networks Supporting Elastic Services", *IEEE/ACM Transactions on Networking*, vol. 9, no. 1, pp. 2-14, 2001.