A Simulation Study of Multi-Color Marking of TCP Aggregates

Miriam Allalouf, Yuval Shavitt*

***Abstract*— Service Level Agreements (SLAs) are contracts signed between a provider and a customer to govern the amount of traffic that will be serviced. This work pinpoints an important problem faced by the Internet service provider (ISP) which is to be able to differentiate between the services given to aggregates of multiple TCP connections. The Metro-Ethernet access network, the Differentiated Services (DiffServ) architecture and the ATM reference model are three architectural models where edge routers perform traffic metering and coloring of aggregated flows according to the SLA.**

Finer color marking was suggested to improve differentiation quality. We observe that increasing the number of colors indeed provides a good differentiation between the aggregates according to the committed and the excess rates. We also show that the token bucket coloring policies, which are widely used for this purpose, prefer short packets and mark them with higher priority colors. The differentiation process is more difficult for the short TCP connections that remain in the slow start phase, than for the long connections that are usually in the congestion avoidance phase.

I. Introduction

During the last two decades three important architectural models were designed and standardized: the ATM reference model, the Differentiated Services (DiffServ) architecture [9] and recently the Metro-Ethernet, the evolving Ethernet-based access network [1]. Although these architectures provide substantially different networking models, they all assume inter-AS SLAs, where edge routers perform traffic metering or policing according to the SLA traffic parameters over an aggregate stream and label each packet as it arrives according to its conformance. An aggregate is a group of connections, for example all the connections of a small company, and the agreement controls an aggregate. The core routers, using e.g. active queue management mechanisms, identify the packet and react accordingly. The different packet marking differentiates between service aggregates.

The first DiffServ standard [9] suggested a profile-based packet marking mechanism using one token bucket where the packets are colored by "conforming" (*green*) and "non-conforming" (*red*) labels. Another DiffServ standard [6] suggested coloring with three colors using two cascading token buckets. The Metro-Ethernet Forum (MEF) also suggests three color marking by using another algorithm over the cascading token buckets. Due to its simplicity and inexpensive hardware implementation, most of the vendors of these architectures utilize the token bucket mechanism for rate estimation, which translates to packet tagging.

The bandwidth resources of a network can be split into the committed allocation portion and the excess bandwidth. Whenever two or three colors are used for packet marking, the differentiation process can be achieved mainly among the committed rates of the aggregates. Though, it still lacks the differentiation capabilities within the excess rates of each aggregate. Cao *et al.* [4] developed the well-known Rainbow Fair Queuing (RFQ) algorithm where the packets are marked using a finer multi-color marking, up to a few hundreds colors, in order to increase the level of service differentiation at the core routers and still maintain fairness between aggregates.

Multi-color marking emphasizes the differentiation capabilities among the SLAs within the excess rates without the need to quantify explicitly the demands. Still, the real traffic mix of the Internet,(public traffic traces analysis []) specifically the TCP flows and its close loop control, require deeper study of its capabilities. Previous analytical models [10], [12], [3] expressed the TCP sending rate as a function of the committed rate, assuming two or three color marking per one TCP flow. They assumed a fixed packet drop probability, without fully modeling the TCP rate adjustment feedback to the loss events.

There are no analytical models that examine the multi-color marking and the active queue management policy interaction with the TCP close loop control. Moreever, despite the many efforts that are done in the industry in this direction, we are not aware of simulation studies that examine the multi-coloring differentiation quality of TCP aggregates that are composed of a variety of data volumes and packet sizes.

The goal of this paper is to test the effectiveness of multi-coloring differentiation capabilities of complex TCP aggregates, by comparing it to the three-color marking scheme according to several criteria. First, the rate estimation and marking, and consequently drop decisions are done per aggregate [11] of TCPs. When several TCP flows are aggregated, the impact of an individual TCP sawtooth behavior is reduced, and the aggregated sending rate and its marking is different.

Second, we aim to check whether a single TCP connection receives its fair share within the aggregate by observing the renewal process of each TCP connection and its fitness to the multi-coloring-queueing system. The aggregates we consider are composed of multiple TCP connections that have different data volumes to transfer and various packet sizes; in addition the number of the connections per an aggregate changes over time. We concentrate on the single TCP connection behavior within an aggregate, given its marking and dropping according

[0]The authors are with the School of Electrical Engineering, Department of Electrical Engineering-Systems, Tel-Aviv University. e-mail: miriama@eng.tau.ac.il, shavitt@eng.tau.ac.il.

to the per-aggregate SLA.

We observe that simulation results have substantial differences between cases where aggregates are just a collection of long connections like done in most simulation studies versus cases where aggregates are comprised of TCP connections with variable length. Our simulations, which mimic better the reality, show that the multi color marking and per-color dropping result in a fairer bandwidth allocation and service differentiation according to the contracts. The multi-coloring enables us to predict the TCP performance more accurately, despite the fact that TCP connection with different duration and packet size can reside in a different TCP congestion control stage and be preferred over the others.

The major observations of our study are: (i) the multi-color marking policy provides a good differentiation between aggregates according to their committed and excess rates. (ii)The token bucket coloring policy prefers short packets and mark them with higher-prioritized colors. (iii) The differentiation process is more difficult for the shorter TCP connections that remain in the slow start phase, comparing with the long connections that are usually in the congestion avoidance phase. (iv) Current analytical models fit the behavior of long connections where the number of the connections per aggregate is fixed over all the simulation period, but are not adequate for estimating the performance of a single TCP connection in aggregates of connections with multiple length and packet sizes. These understandings can lead to the development of an extended analytical model that considers the relationships between packet drops, queue length, and TCP average sending rate.

Section II provides the background and describes former analytical models done in the field. Section III describes the token bucket coloring policies that are used in these simulations. Section IV describes the simulation process and results. Section V is the discussion.

II. Multi-Color and Queuing Drop Discipline Background

A coloring method slices the traffic rate to layers, each is represented by a color and requires a different treatment, e.g., different dropping probability. We will adhere to the three color convention [6] and call the packets that arrive within the committed rate *green*; packets that are within the excess rate, *yellow*; and packets which are outside of the peak contract rate (the SLA) *red*. If the rate region that is represented by the *yellow* is divided into multiple subregions, each will be represented by a different yellow hue, such that the ones representing the lower rates (i.e., closer to *green*) will be called light hues, while the ones closer to the peak rate will be called dark hues.

In a fair per-aggregate treatment, the color distribution in an aggregate is proportional to its contract. For example, assume that an aggregate A is permitted to send a 10Mbps excess rate (averaged), while aggregate B can send 40Mbps excess rate. If we use two *yellow* hues for the excess traffic and equally divide the excess rate region between the two layers that are represented by these hues, aggregate A will have up to 5Mbps of its excess traffic colored with light *yellow*, while B will be able to use this color for up to 20Mbps of its excess traffic.

It is not enough to achieve a proportional color distribution. Because of the complex interaction with the TCP close-loop control, a ”good” coloring method depends on the queue management scheme at the router and TCP reactions. We distinguish between two combinations of coloring and queue management approaches. In the first, the coloring mechanism colors the packet deterministically trying to fit the packet in the lowest possible rate layer, by using, for instance, token buckets as markers. The queue management scheme determines the drop probability for each color, e.g., by using color aware dropping policy, such as the Random-Early-Detection (RED) management. In the second approach, the flow rate is estimated and the coloring mechanism randomly assigns the packet a color with a probability that is drawn from the ratio of this color rate to the estimated rate. The drop mechanism in this case will be deterministic, dropping all packet above some threshold color. The Rainbow Fair Queuing (RFQ) [4] algorithm is a good example of this approach.

In this paper we use the first approach, that is implemented with multiple token buckets and multi-GRED. GRED (Gentle RED) was found to be superior to RED [5], thus we use here a variant of GRED that is used for multi-priority dropping[1].

Next we survey papers that provide analytical models for both service differentiation quality and a prediction of the achieved TCP sending rate when two and three-coloring marking is used. Sahu *et al.*[10] modeled the TCP Reno flow renewal process where its derived sending rate follows two cases: under subscription and over subscription. They assume that two-color marking is performed for each individual TCP and pre-determined loss probabilities for the two priority aggregates deployed by the multi-RED dropping policy. The sending rate is derived as a function of the loss probabilities, with no explicit model of how the loss probabilities are affected by the sending rate. They expressed the TCP achieved sending rate as a function of the committed rate, which depends greatly on the token bucket size, RTT and loss probabilities. They obtained the following main results: (i) the achieved rate is not proportional to the committed rate, (ii) it is not always feasible to achieve the committed rate and, (iii) there exist ranges of values of the achieved rate for which token bucket parameters have no influence.

Yeom and Reddy [12] give a better model for committed services, using the substantial construction that is also based on the under and over subscription states. The results are quite the same as given by Sahu et al. Specifically, they developed a throughput model for an individual flow within an aggregated reservation, where the marker marks TCP packets from an aggregation, using a per-aggregate SLA. They developed the following equation that relates the realized bandwidth, B_i of an individual flow i to the aggregated committed rate, R_A and

[1]The extended version of this paper [2] contains more detailed description of the multi-priority dropping queue management

the network conditions observed by various flows within the aggregate, the round trip time RTT_j and the packet loss p_j:

$$B_i = \frac{m_i}{\sum_{j=1}^{n} m_j} \cdot \frac{3R_A}{4} + \frac{3k}{4} \cdot m_i \quad (1)$$

Where

$$m_i = \frac{1}{RTT_i} \cdot \sqrt{\frac{2}{p_i}}, k = packet\ size \quad (2)$$

The manipulations that were done prior to this equation contained a few assumptions. Let us outline some of the given assumptions and point out an opposite behavior that can harm the differentiation process as follows:

- **Assumption:** All the flows transmit packets of the same size. **Refutation:** The token bucket mechanism prefers short packets.
- **Assumption:**This specific equation calculates the TCP sending rate during the congestion avoidance phase. **Refutation:** Most of the connections in Internet today are short and thus stay in their slow start phase.
- **Assumption:** Each TCP connection is assumed to have a fixed RTT_i. **Refutation:** Due to queue size oscillation and varying number of parallel TCP connections, RTT_i changes over time.
- **Assumption:** A fixed packet drop probability, without fully modeling the TCP rate adjustment feedback to the loss events. **Refutation:** It is difficult to provision correctly the queuing thresholds in order to achieve the required dropping probability, a fact that can result in high deviation from this equation
- **Assumption:** There is fairness in the coloring of all the connections within an aggregate such that if $x = \frac{aggregate\_contract}{sum\_of\_arrivals}$ they concluded that x is the ratio of the IN packets for the single connection, as well. **Refutation:** different parameters such as packet size and duration affect the ratio.

The mathematical model that is presented in these papers cannot predict the TCP sending rate within an aggregate, given a complex traffic mix.

III. Marking using Token Buckets

An (r, b)-token bucket is a classical model that regulates the traffic envelope using two parameters: r, the fill rate of the tokens, that dictates the average traffic rate, and b, the bucket size, that determines the allowed burstiness. In a metering and a policing system, the token bucket acts as a rate estimator (or a meter) and a marker. Any packet within these limits is considered to be conforming to the bucket allocation, otherwise, the packet is non-conforming. We say that the conforming packets are within the rate r.

In a marking system with NC colors, $NC-1$ cascading buckets are used rather than one and it colors packets as they arrive using NC colors: *green for the committed traffic*, $NC-2$ yellow hues for the excess traffic and *red* for the non-conforming traffic, according to the corresponding bucket allocation. The traffic demand that composes the SLA is expressed by two rates: CIR the average Committed Information Rate and PIR average Peak Rate; and two burstiness parameters CBS the Committed Burst Size and PBS Peak Burst Size. $NC-1$ buckets are used: the '*green*' (CIR, CBS)-token bucket; $NC-2$ excess (r_i, bs_i)-token buckets that are associated with the *yellow hue* y_i where r_{NC-1} equals the PIR and bs_{NC-1} equals PBS.

Heinanen and Guerin [6] suggested a three color marking, termed trTCM (two rate three color marking), that was adopted by the IETF DiffServ working group. It uses two cascading token buckets: for committed and for excess traffic. The packets are colored in three colors: *green* (within the CIR), *yellow* (above the CIR but within the PIR), and *red* (above the PIR), according to the allocation of the buckets.

The MEF standard [1] proposed a different two bucket implementation for three color marking: A (CIR,CBS)-token bucket, as before, and an excess (EIR,EBS)-token bucket. Here, EIR refers to the excess rate and equals PIR minus CIR and, EBS, the size of the *yellow* bucket and its goal is to get bursts within the EIR range.

Our simulations compare three-color vs. six-color marking and use, specifically, the MEF token bucket setting, metering, and marking. This code was implemented and added by us to the ns-2 code. The colors in the six-color marking are: *green*, red and 4 yellow hues. More colors enable better service differentiation, but require more resources. It is out of the scope of this paper to find the optimal number of colors, what we do instead is point out what makes a color separation work. We claim, that to achieve good service differentiation there should be a token bucket that works at a rate close to the system fair rate. Since in practice the fair share keeps changing by the load on the system, a good separation must allocate colors in a way that will optimize all possible cases doing better in the more "important" system regimes.

The bucket size parameter is difficult to tune. Too large bucket size enables high burstiness, namely many packets will be colored as conforming. Too small bucket size may lead to a state where delay jitter will cause some packets to be marked as out-of-profile and eventually may be dropped. Even assuming that the bucket size is tuned well to the contract burstiness, we show that the packet size can determine its marking. Specifically, smaller packets are significantly more likely to be colored as conforming, and in the case of multi-coloring, are more likely to be colored lightly. In the extreme case only small packets will be colored as conforming. Any packet that is larger than the bucket size will be marked as out-of-profile.

This becomes a problem when one is using many token buckets to implement multiple coloring. There are various recommendation regarding the choice of a committed and excess burst sizes when two buckets are used. Kim *et al.*[7] state that the trTCM marker performs best when PBS equals CBS. Other recommend on a very small peak burst size (a few max length packets), comparatively to the committed burst size, since the intent is to strictly limit the peak rate,

while the committed rate to be exceeded for fairly long time periods, meaning that the committed burst size should be reasonably large (hundreds of packets). In our simulation we tried different combinations of bucket sizes and finally decided that allowing too high burstiness for the darker yellows results in a lot of drops and less differentiation.

IV. Simulation

Our simulations were designed to measure the quality of the differentiation mechanism by examining whether the CIR parameters are respected, and whether the excess bottleneck link bandwidth is shared proportionally to the EIR values of the participated aggregates. For this end, we examine the per-aggregate coloring distribution, as well as, the coloring distribution of the packets in the queue. In addition, we will check whether each TCP connection obtains its fair share of bandwidth within the aggregate.

A. Simulation Setup

We assume a number of aggregates traversing wide links towards a bottleneck link. Each aggregate is metered using a different SLA profile at the coloring gateway, at which the colors are assigned without distinguishing the different TCP connections within this aggregate (Figure 1). There is only one queue at the bottleneck link that absorbs the colored packets of all the aggregates[2].

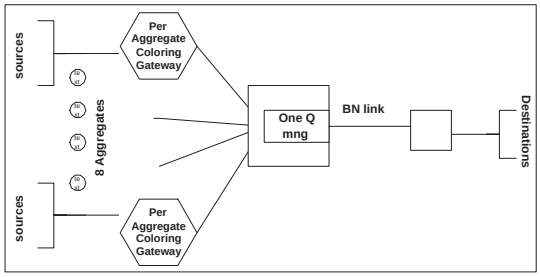

Fig. 1: The simulation Topology. The bottleneck (BN) link rate is 4Mbps. The rate of all other links is 100Mbps. The propagation delay of each link is 5ms. There are 8 aggregates, each enters the network via a dedicated policing gateway

We will make the following definitions with respect to R, the bottleneck link, and A, the group of the aggregates that flow over R, as follows: $comm(A) = \sum_{i \in A} CIR_i$ is the committed rate of R; $ex(R) = bw(R) - comm(A)$ is the excess link rate of R. $AGG_{CIR} = comm(A)/bw(R)$ is the CIR aggregation level. The EIR aggregation-level is the entire EIR allocated on a bottleneck link divided by its excess rate, $AGG_{EIR} = (\sum_{i \in A} EIR_i)/ex(R)$. The fair throughput of an aggregate i is composed of its CIR_i and its fair share of the excess bandwidth $EIR_{agg_i} = (EIR_i/AGG_{EIR})$.

Table I present two sets of multiple SLA combinations which differ in the SLA parameters and the aggregation-level. Each scenario consists of 8 aggregates. Scenario B

[2]Usually multiple priority classes that require the same delay are handled by one queue, and we assume that all the metered aggregates, in the following scenarios belong to the same delay class [8].

Class (aggr.)	CIR	PIR	EIR	CBS	EBS	fair thruput
	Scenario A					
1	300K	2M	1.7M	18K	12K	817K
2	300K	2M	1.7M	18K	18K	817K
3	300K	1M	0.7M	18K	12K	514K
4	300K	1M	0.7M	18K	18K	514K
5	150K	1M	0.85M	18K	12K	410K
6	150K	1M	0.85M	18K	18K	410K
7	150K	0.5M	0.35M	18K	12K	257K
8	150K	0.5M	0.35M	18K	18K	257K
Total	1.8M	9M	7.2M			4M
$bn = 4M, ex(bn) = 2.2, AGG_{CIR} = 0.45, AGG_{EIR} = 3.27$						
	Scenario B					
1	200K	4M	3.8M	18K	12K	833K
2	200K	4M	3.8M	18K	18K	833K
3	200K	2M	1.8M	18K	12K	500K
4	200K	2M	1.8M	18K	18K	500K
5	100K	2M	1.9M	18K	12K	416K
6	100K	2M	1.9M	18K	18K	416K
7	100K	1M	0.9M	18K	12K	250K
8	100K	1M	0.9M	18K	18K	250K
Total	1.2M	18M	16.8M			4M
$bn = 4M, ex(bn) = 2.8, AGG_{CIR} = 0.3, AGG_{EIR} = 6$						

TABLE I: The SLA parameters of the eight aggregates that compose scenario A and B. For each set we present the excess bottleneck link and the CIR and EIR aggregation levels.

was designed to have a higher EIR aggregation rate. The size of the committed bucket (CBS) is the same for all the aggregates. Each scenario contains four pairs of aggregates: (1,2),(3,4),(5,6), and (7,8). Their TB values are the same, except for the excess burst size parameter, EBS. When using six-color marking, the EBS is divided into four TBs (for example: 18K EBS is translated into 4 TBs with the sizes of 9000,4500,2250,2250).

The capacity of the bottleneck link is 4Mbps[3]. The CIR aggregation-levels are 0.45 and 0.3 in scenarios A and B, respectively, and determine an under-subscribed states: meaning that the *green* packets will be guaranteed. The *red* packets will always be dropped.

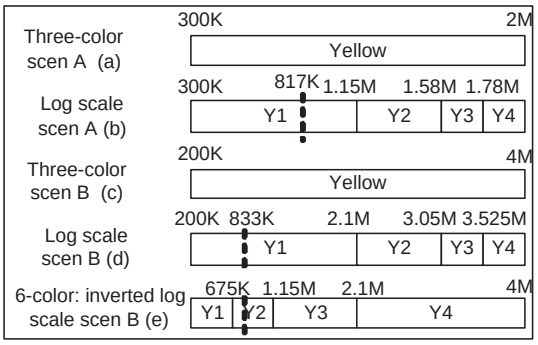

Fig. 2: Rate assignment example for aggregate 1 in both scenarios and both coloring schemes

We initially assigned a logarithmic scale for the rates of

[3]The bottleneck link speed and the SLA rate parameters were chosen in the appropriate proportions, scaled down from real Internet speeds.

the yellow hues. Namely each yellow strip is half the width of its previous lighter yellow (Fig. 2(b)& (d)). Note that the figure shows the rates for aggregates 1 (for both scenarios) but all the aggregates have the same picture only scales to their EIR range and shifted by their CIR value. In scenario (d) it is clear that the multi coloring may have little effect since even if an aggregate will consume twice as much as its fair share it will still have all its packets colored with the lightest yellow hue. Thus we also simulated the inverted log scale ((Fig. 2(e)) where crossing the fair share in both ways results in different color distribution.

Another important set of parameters is the queue length and the thresholds parameters. The maximum possible queueing delay affects the RTT and the number of packets that can be sent in a burst by a single TCP connection. The multi-GRED thresholds should be set to balance between the need to lower the average RTT and allow burstiness at all congestion levels (remember that different congestion levels cause the queue to balance at different yellow hue).

Number of TCPs	File Size	Packet Size
4	1.5MB	1500B
8	75KB	1500B
32	15KB	1500B
128	1600B	400B
500	400B	400B

TABLE II: The traffic mix of each aggregate.

For our simulations, we chose a complex traffic mix that is more appropriate for simulating the actual Internet traffic. As a result we could observe TCP behaviors that were missed by previous more simplified studies. Each aggregate has a mix of TCP connection length, which represent a ”deterministic Pareto distribution: long (elephants), medium, and short (mice). The connection length determine at which TCP congestion control phase the connection spends most of its time. Short connections stay in the slow start phase, while long connections tend to be mostly in congestion avoidance. We also varied the number of active connections during the simulation life time. An important aspect of the simulation was to use a mix of packet length, which showed that the coloring is sensitive to this parameter. The traffic mix is presented in Table II. To achieve different number of parallel TCP connections, we distribute the start of the connection per aggregate using the Poisson distribution, i.e., the times between connection initiation times are distributed exponentially (with an average of 0.045 sec). The simulation terminates when all the connections are done (roughly around 160 seconds). There are two distinctive periods during the simulation: In the first 40 seconds there are many short and long TCP connections in parallel; in the latter period only a few long connections remain. In our results presentation we distinguish between those two periods. In order to measure the effect of each of the above parameters we performed an extensive simulation study using ns-2 network simulator.

B. Simulation results

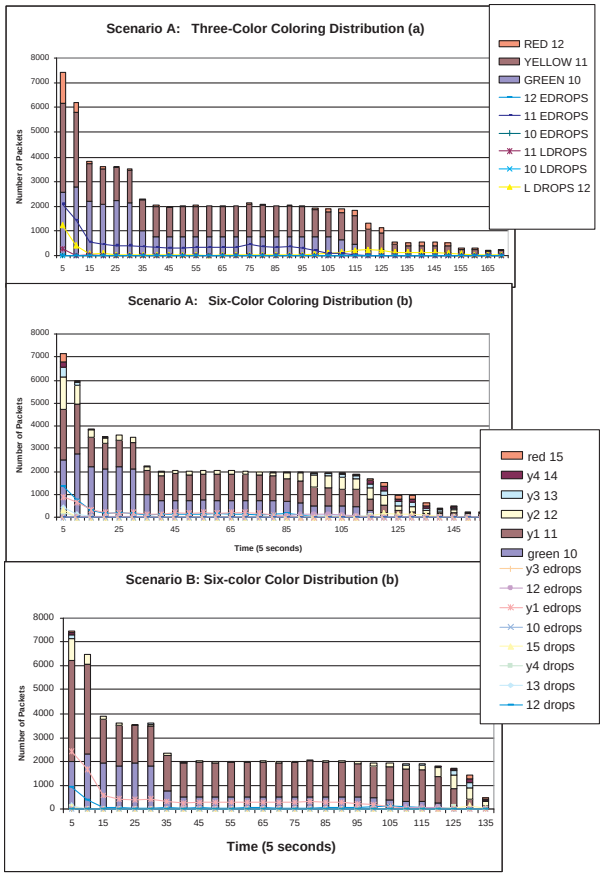

Fig. 3: The coloring distribution over time where each column is a cascading representation of the number of the packets per each color, starting with *green* and climbing to the darkest color that was used. The three and six-color marking are show in (a) and (b) for scenario A, and in (c) for scenario B

1) Color Distribution and Throughput Comparison: Figure 3 presents the coloring distribution of all the packets that arrive at the bottleneck link queue for scenario A and B. The first period of around 40 seconds has connections with short packets and hence the large number of packets in the first columns. As presented in figure 2 for the logarithmic rate assignment, the colors used for Scenario A are darker than for B: The three-color marks more packets with *red* since the marking start using GRED on the yellow at lower thresholds; the six-color marks with all the colors. The coloring distribution in first column in figure 3(b), obeys the logarithmic assignment until later the queue stabilizes around the fair point which is at yellow 1 and the darkest two yellows and the red almost vanish. In scenario B (Figure 3(c)) where the fair share is deep inside the yellow 1 region we see mostly *green* and *yellow 1*.

An aggregate throughput is the effective number of the successfully transmitted bytes per second. The differentiation is achieved when the throughput per aggregate is proportional to its SLA parameters.

Figure 4 presents the throughput for scenarios A and B. The rates of all the aggregates for both scenarios, as shown in the graphs, are higher than their CIR values, which means

that their committed rates are achieved. Next we will check whether the excess bottleneck link rate is shared among the aggregates in proportion to their EIR values.

The fair share of Scenario A is closer to the yellow 1 maximum rate than in scenario B and thus scenario A uses more colors and achieves better differentiation. The better throughput differentiation for scenario A for the three-color marking is explained by its higher EIR aggregation rate and its over-subscribed state [12]. The six-color provides a very good differentiation regarding the CIR and the PIR (Figure 4 (b)). The resulting averaged throughput vector for the six-color marking is ($870K$ $620K$ $550K$ $550K$ $420K$ $420K$ $279K$ $279K$), which is very close to the fair vector, which is presented in Table I, though the aggregates with the highest PIR values (aggregates 1 and 2) fluctuate around the 817K fair value. The per-aggregate throughput, as demonstrated in 4(a) for three-color marking, are more condensed, but the SLA order is kept.

The differentiation for scenario B by the three-color marking is very poor. The six-color marking demonstrate better differentiation according to the CIR and to the PIR values, though worse than what was achieved for scenario A.

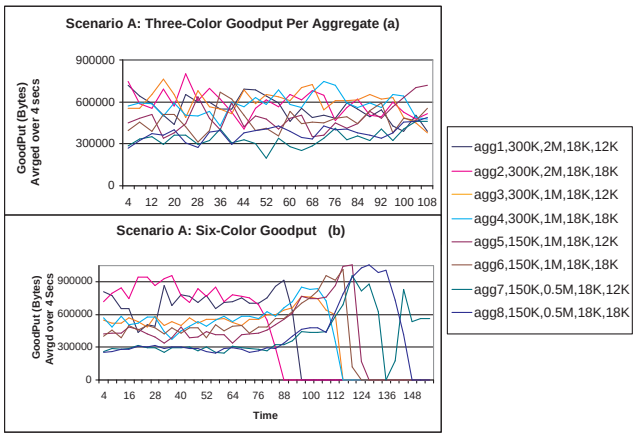

Fig. 4: The effective throughput of all the 8 aggregates over time (in resolution of 4 seconds) in scenario A for three (a) and 6ix-color (b) policies, respectively

Figure 5 presents the coloring distribution and the throughput of another simulation where we used the same parameters as in scenario B except that the logarithmic yellow hue assignment was inverted (see Fig. 2(e)). The demonstrated differentiation quality is absolutely better.

In general, the coloring distribution and the differentiation process is significantly affected by the load and activity of other marked aggregates over the bottleneck link and can vary according to its excess bandwidth. For instance, in figure 4(b) for the six-color marking at 100 seconds, we can see that when aggregates 1 and 2 terminate, packets of aggregates 3, 4, 5, and 6 increase their throughput to 750Kbs. It is also reflected in the last columns of the "coloring distribution" graph in Figure 3(b) where the colors are darker when the total number of packets is smaller.

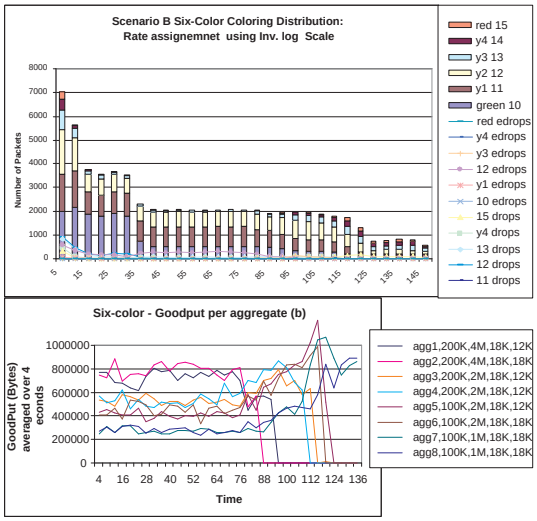

Fig. 5: The effective throughput scenario B for three (a) and 6ix-color (b) policies, respectively when an inverted logarithmic scale is used to the rate assignment

2) Packet Size and Color Distribution Comparison: We would like that the coloring within an aggregate will be distributed uniformly over all the packets. The coloring is done per packet because the queue management drops or accepts packets. However, since the token bucket counts bytes, an increase in the packet size reduces the number of packets that are colored *green*. We demonstrate this by the following results.

We use three possible packet sizes that can arrive at the coloring gateway: (1) 40 byte TCP SYN packets, (2) 440 byte data packets used for transfer of 440B and 1600B byte file sizes, and (3) 1500 byte data packets used for transfer of 15K, 75K, and 1.5M byte file sizes. Denote by S,M, and L the number of the packets of sizes 40, 440, and 1500B, respectively. Further, denote by S_g,M_g and L_g the number of the green packets per each size. Upon a uniform coloring distribution, we expect that: $(S:M)=(S_g:M_g)$, $(S:L)=(S_g:L_g)$, $(S:L)=(S_g:L_g)$ and $(M:L)=(M_g:L_g)$. Table III presents the coloring results of three and six coloring for scenario A. For both coloring schemes it is clear that the ratio of short green packets to longer green packets is higher than the ratio between the total number of corresponding packets. In addition for the same packet size, the CIR ratios among the aggregates are not kept.

Following the above finding regarding shorter packets and files preference, we will compare the duration of file transmission and dropping ratios, per file size.

3) The Packet Loss and the transmission Duration Comparison: This section will check whether the TCP connection share is proportional to the CIR and PIR values of the aggregate it belongs to. We treat the file transmission duration and the packet loss ratio as the metrics to compare and measure whether a single connection obeys the contract. The following results show that it depends on the connection length, packets

Scenario A						
three-color marking	S=40B	M=440	L=1500	S:M	S:L	M:L
Total	5381	10563	14533	0.51	0.37	0.73
Colored Green	5367	6330	3328	0.85	1.61	1.9
Green / Total	0.997	0.599	0.228			
six-color marking	S=40B	M=440	L=1500	S:M	S:L	M:L
Total	5380	10143	14433	0.53	0.37	0.70
Colored Green	5369	6184	3302	0.86	1.62	1.87
Green / Total	0.997	0.61	0.228			

TABLE III: The number of total packets and the number of the green packets per each size and the ratios within each size group.

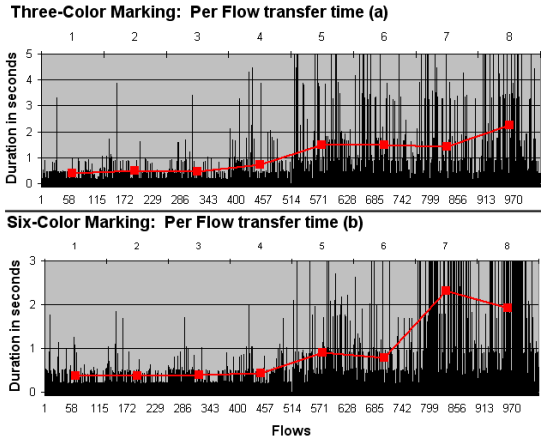

Fig. 6: Termination times for 1600 byte files (128 connections X 8 aggregates): starting from the left with those that belong to aggregate 1 and progressing to the right to aggregate 8. The upper axis shows the aggregates. The lower axis shows the connections number. Figure (a) show the termination times for the three-color marking, that are coupled by the CIR: aggregates 1-4 and 4-8 have similar results. Figure (b) show that results are coupled by the CIR and the PIR.

size, and RTT and that not always the differentiation is possible. Whenever a differentiation is achieved the six-color marking provides a better differentiation.

All the connections start with sending a short SYN packet. A drop of a SYN packet causes a long "connection establishment" timeout of 5.5 seconds to delay the transmission in case of congestion. In a coloring system, very few SYN packets are dropped because they are colored *green* with probability close to 1, as was shown in the previous section. Such a scheme ignores the TCP connection setup congestion control. Furthermore, there is no difference in the number of the established connection in the different aggregates.

The transfer time of a 440B file should take two RTTs: one for the SYN packet and one for the data packet. The drop probability of a data packet consists of its probability to be colored by other color than *green* according to its SLA. When comparing the duration time and the drop percentage for these files, we found that there is no differentiation for both scenarios because the small data packet drop probability is low enough to make the drop event too sporadic to cause meaningful differentiation between aggregates.

The transmission of an 1600B file, with a packet size of 440B is composed of a SYN packet and four data packets that are sent in the slow start phase. In case of a drop, the connection is delayed by the initially set timeout (i.e, the timeout and the RTT values are not tuned because of the small number of packets that were sent). This is the reason that the transfer times we found are in a multiplication of 200 ms, which is the RTT estimation. Figure 6 presents the transmission times of these TCP connections for scenario A. The transfer times are around 0.3 seconds when there are no losses, 0.4 seconds when one data packet is lost and causes a timeout, 0.8 seconds when there are 2 drops, and fewer connections can take 1.6, 3.2, 6.4 and 12.8 seconds transfer time. The higher numbers are usually a result from the rare dropping of the SYN packets. The results[4] show that the differentiation quality is much better than for the former file size. The six-color marker differentiate the transfer times according to the CIR and the PIR values whereas the three color provides a per CIR differentiation only. When comparing the percentage of the loss for this file size between the 8 aggregates we got the same results . The bucket size has no effect since these connection are not bursty.

The transmission of a 15KB file with a packet of 1500B is composed of ten packets and thus remains in the slow start phase. The drop ratios for this file are very high (around 47%) because of its short length and its long packet and they result in a very bad differentiation. In any case the six-rate coloring policies provide definitely better differentiation than the three-color schemes. The 75KB file is long enough to reach the congestion avoidance phase, although it iterates a lot among the two phases because of the load. Relatively to its size, its drop ratios are smaller than the former file size, a fact that causes better differentiation. The longest file size transmits 1.5MB and its packet size is 1500B. There, TCP stays in the congestion avoidance stage most of its duration. All the policies demonstrate a nice differentiation regarding the transfer time and the packet loss ratio.

V. Discussion

A. Per-Color Rate assignment and Queue Parameters Setting

Our results show that multi coloring can improve the differentiation quality with respect to the committed and the excess rates. This improvement depends on the SLA parameters, per-TB rate assignment and queue length thresholds provisioning.

The six-color marking policy achieves good differentiation between the aggregates according to the CIR and the PIR rates when the TB rate level is below (or closely above) the EIR_{AGG_i}. It provides good excess differentiation whereas the three-color marking differentiates only with respect to the CIR values. Further more, six-color marking results show that: (1) The sending rate of the long TCP connections in aggregates with a lower PIR values, are more stable and experience less drops; (2) when the PIR values are higher, a larger bucket size (by comparing aggregate 1 to aggregate 2)

[4]In order to map the quality of the differentiation, we used a density diagram (histogram) of the transfer times.

can improve the throughput of the aggregates. We used in our initial simulation a logarithmic scale that set the *yellow 1 hue* rate too far above the target throughput. It proved to be less effective than an equal division or inverted logarithmic scale of the rate range. Since the EIR aggregation level depends on the bottleneck link rate, a network administrator should assign the rates only by estimating this ratio. Finer color division of the excess range leads to smaller rate subregions and increases the accuracy of this estimation.

The other network design issue that the network administrator should deal with is tuning the multi-GRED queue thresholds to enable a differentiated dropping according to the *yellow* hues and still achieve a stable average queue size (AQS). Indeed, a higher number of dropping priorities result in a better differentiation quality. However, with no a priori knowledge about the traffic mix it is impossible to tell at what AQS the system will stabilize. In a highly congested situation when it is likely that due to competition most aggregates will not be able to transmit dark *yellow* hues, the AQS is expected to be high since only then the queue reaches occupancy levels where packets are discarded. On the other hand, if many flows are inactive, and there is little competition for capacity, flows can reach the darkest yellow hues. In this situation the queue cannot grow much since it will quickly hit an occupancy when dark *yellow* packets are dropped and TCP will react accordingly by halving the transmission rate. The differentiation capabilities are kept, though the AQS and the queuing delay is unknown.

Another traffic engineering inconsistency can happen, when the rate estimator and marker assigns colors by bytes, as the token bucket does, and the queuing policy handles the arriving traffic by packets, as the multi-GRED queuing management. Since the coloring policy prefers short packets and tags them with a lighter color, such as *green*, it can happen that there are much more *green* packets than what was intended by the network administrator. The queue get larger and higher colored packets get dropped although the delay of the queue is not so high.[5]

B. The Differentiation of Shorter Packets and Flows

The differentiation quality is different for various TCP file length, TCP congestion control phase, packet size and parallel number of connections. Typically very short TCP connections use small packet sizes. Those connections are favored over the connection with the larger packets because their colors are lighter and they do not achieve burstiness, a fact that reduce the dropping probability of their packets. These files achieve a weak differentiation quality.

The best differentiation is achieved when the traffic consists only of long connections (1.5Mbps) with long packets (1500B) (was demonstrated in the second period of the throughput graphs). But also in a mix of file lengths, the longer files contribute to the overall per-aggregate throughput differentiation. Previous analytical models mainly considered such files and are inadequate for mix length scenario. In addition, such models of aggregates differentiation by marking, cannot assume a constant drop probability, since the drop depends on a variety of factors such as flow size, burstiness, TCP congestion phase, and packet size.

The situation where shorter packets are favored in coloring may cause users to artificially send shorter packet to improve their performance through markers and global network efficiency will be lost. Thus, it is important to devise coloring implementations that do not exhibit such behavior.

[5]The queue delay is composed of the sum of the transmission delays of the packets within it and shorter packets have shorter transmission delay.

VI. Concluding remarks

The research of multi-coloring marking is an urgent need given the industrial trends. The token bucket is the most popular tool in today industry for SLA management. However, most of its users are not aware of the impacts of different settings on marking. In particular, people are not aware of how this translates to TCP performance. Previous analytical models and papers already showed the difficulties in tuning these parameters and generalizing this problem. In this work we highlight these difficulties and confusion by considering a realistic Internet traffic when two and more colors are used. We show that in a complicated environment that consists of plenty of parameters, an addition of even a few more colors can significantly improve the differentiation quality among aggregates.

References

[1] O. Aboul-Magd. MEF traffic management specification. Metro Ethernet Forum, May 2004.

[2] M. Allalouf and Y. Shavitt. A simulation study of multi-color marking of tcp aggregates. Technical report, Dept. of Electrical Engineering – Systems, Tel Aviv University, 2007. Technical Report EES2007-139.

[3] C. Barakat and E. Altman. A markovian model for tcp analysis in a differentiated services network. *Telecommunication Systems*, 25:129–155, 2004.

[4] Z. Cao, Z. Wang, and E. Zegura. Rainbow fair queuing: Fair bandwidth sharing without per-flow state. In *INFOCOM*, March 2000.

[5] M. Christiansen, K. Jaffay, D. Ott, and F. D. Smith. Tuning RED for web traffic. In *SIGCOMM*, pages 139–150, 2000.

[6] J. Heinanen and R. Guerin. A two rate three color marker. Internet Engineering Task Force, September 1999.

[7] H. Kim, C. Yoo, and W. Y. Jung. Simulation study on the effect of the trtcm parameters. In *Telecommunications'03*, pages 568–578, 2003.

[8] M. May, J. Bolot, A. Jean-Marie, and C. Diot. Simple performance models of differentiated services schemes for the internet. In *INFOCOM (3)*, pages 1385–1394, 1999.

[9] K. Nichols, S. Blake, F. Baker, and D. Black. Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers. Technical Report RFC No. 2474, Internet Engineering Task Force, December 1998.

[10] Sambit Sahu, Philippe Nain, Christophe Diot, Victor Firoiu, and Donald F. Towsley. On achievable service differentiation with token bucket marking for TCP. In *Measurement and Modeling of Computer Systems*, pages 23–33, 2000.

[11] Y. Xu and R. Guerin. Individual qos versus aggregate qos: a loss performance study. *IEEE/ACM Transactions on Networking*, 13:370–383, 2005.

[12] I. Yeom and A. L. Narasimha Reddy. Modeling tcp behavior in a differentiated services network. *IEEE/ACM Transactions on Networking*, 9(1):31–46, February 2001.