Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 EECS Building
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 00-020

Random Packet Marking for Differentiated Services

Ewa Kusmierek and Rajeev Koodli

March 17, 2000

Random Packet Marking for Differentiated Services

Ewa Kuśmierek*
Dept. of Comp. Science & Eng.
University of Minnesota,
Minneapolis, MN 55455
kusmiere@cs.umn.edu

Rajeev Koodli
Nokia Research Center
Burlington, MA 01803
rajeev.koodli@nokia.com

1 Introduction

Internet Differentiated Services (DS) architecture defines packet marker as one of the traffic conditioning components [1]. Markers set the DS field of a packet to a particular codepoint indicating in this way what type of service this packet should receive. Assured Forwarding PHB group [2] defines service differentiation in terms of different levels of forwarding assurance for packets received from a customer DS domain. Following the tendency to push the complexity towards the edge of the network and to keep the core simple, operations such as metering, marking and shaping will be performed at the access node. Access node for a given customer domain classifies each packet as belonging to one of the four AF classes, each class given a different forwarding assurance. Packets belonging to the same class are next marked with some drop precedence level which depends on the traffic conformance with its profile i.e. committed information rate, peak information rate, and burst size.

Committed information rate is the amount of bandwidth the customer is planning to use. If the rate at which packets arrive at the marker does not exceed committed rate, all packets can be considered as sent in-profile and mark with low drop precedence represented by green color. Otherwise packets are given medium or high drop precedence represented by yellow or red color, respectively. Peak information rate indicates the highest rate at which the source is expected to transmit, and may be used to make a decision whether yellow or red color should be used. Burst size specifies the number of packets that can be sent back-to-back at the rate not exceeding peak information rate. In order to determine conformance with the traffic profile, metering has to be performed on all traffic passing through the boundaries of different network domains.

There are two groups of marking methods: per-flow and flow-aggregate. Methods from the first group have to identify a packet as belonging to one of the flows. Sending rate metering and marking is done separately for each flow. It usually requires maintaining per-flow state. Aggregate marking does not require identifying which flow a packet belongs to. Metering and marking is done on the aggregated traffic.With the increasing amount of traffic on the Internet, scalability of any mechanism used becomes a very important issues. A scheme which requires storing and processing a large amount of data such as per-flow state for a large number of flows does not scale very well. Packet marking needs to be performed at the boundary of a single domain where the incoming traffic may consist of many flows and the contract is specified for the aggregate of flows. The

*This work was started and a majority of it was done while the author was at Nokia Research Center, Boston, June-September 1999 "content" of the traffic may not be of interest to the ISP as opposed to traffic conformance with the contract. Random marking is a new packet marking scheme. It is based on the probability function and designed to be performed on the aggregate traffic belonging to a single AF class. Each packet arriving at the access node is marked as green, yellow or red with a probability which is a function of sending rate with respect to committed and peak information rate.

2 Background and Related Work

Two Rate Three Color Marker [4] is one of the existing packet marking schemes which makes its marking decision based on the comparison between sending rate and committed rate (CIR), and between sending rate and peak rate (PIR). It uses two buckets, C and P, which are filled with tokens at the rate corresponding to the committed and peak information rate, respectively.

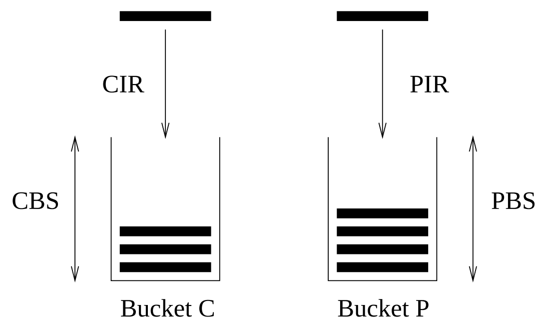

Figure 1: Three Color Marker

TCM operates in one of two modes. In color-blind mode the availability of tokens in the buckets is used to determine which color should be used. A packet is marked as red if there is no token available in either bucket. If there is no token available in bucket C but there is a token available in bucket P, the packet is marked as yellow. Finally, if there are tokens available in both buckets the packet is marked as green.

In color-aware mode packets arrive with some pre-assigned color. In this case the initial assignment is respected, i.e., the drop precedence can be only increased. If the packet has been colored as red, it has to remain red. A yellow packet is marked as red if there are no tokens available in C and P buckets, otherwise it is allowed to remain yellow if there is a token in bucket P. Finally, a green packet can become a yellow packet, if a token is available only in bucket P, or a red packet if there are no tokens available in either bucket. Otherwise a green packet remains green.

Each bucket can also accumulate a number of tokens determined by its depth. Depth of bucket C corresponds to the size of burst of green packets that can be sent back-to-back (CBS). Similarly the depth of bucket P corresponds to the size of the burst of yellow packets (PBS). In both cases tokens are stored only if the sending rate does not exceed CIR for bucket C and PIR for bucket P. Once the bucket becomes full, all additional tokens are discarded.

TCM is designed to perform a per-flow marking. When applied to an aggregate traffic, TCM marking may discriminate some flows. Packets belonging to one flow arriving at the marker just as the needed amount of tokens becomes available, may seize most of the tokens. An example of such behavior is presented in section 5.2. Under the right circumstances then, one flow or a group of flows may receive preferential treatment. In order to assure that none of the flows is discriminated, some control over the consumption of tokens is necessary. Some work has already been done on adjusting token bucket based schemes to perform marking on aggregated traffic. In “Fair Marker” [6] a scheme based on two rate TCM is presented. It adapts fair buffer management algorithms to token allocation. Traces of packets consuming tokens are kept in a complementary queue and are regarded to replace the consumed tokens in the bucket. The state of each flow whose packet traces are queued is maintained. Decision whether a packet trace can be enqueued, in other words, whether it is allowed to consume tokens, is made using fair buffer management algorithms such as Flow RED or Dynamic Threshold (DTQ). Another method of controlling token consumption for each flow is suggested in [7]. A common token bucket is used for all flows but each flow is assigned a logical partition. Each flow uses some fraction of token generation rate and some fraction of bucket depth. Unused tokens in any logical partition are shared equally among other flows. Both methods assure some degree of fairness, however they both require maintaining some per flow state and do not scale as well as a stateless scheme.

3 Random Packet Marker

Traffic originating from a subscriber network entering ISP domain is bound by the subscriber-ISP contract. The contract specifies committed information rate (CIR), peak information rate (PIR) and burst size as the main parameters. Marking performed on the packets should reflect the traffic conformance to the contract. Random packet marking is designed to be performed on the aggregate traffic and makes the decision about drop precedence for each packet based on the estimate of the aggregate sending rate. Colors are assigned randomly with probability which is a function of sending rate and allowed configured rates.

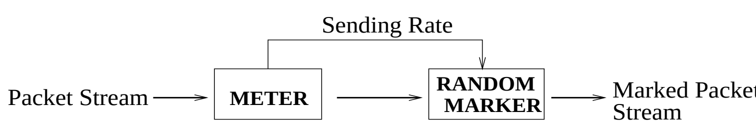

Figure 2: Packet Marker

Green packets should be forwarded at the rate no higher than CIR and yellow packets at the rate no higher than PIR-CIR. The only case in which sending rate of green packets may remain below CIR is when the aggregate sending rate is below CIR. Similarly, the sending rate of yellow packets should be below PIR-CIR only if the aggregate sending rate is below PIR. The rate of red packets is equal to the difference between the aggregate sending rate and PIR. The rate at which the traffic arrives at the marker is measured by the meter (see fig.2) and the result is passed to the marker. The ratio of sending rate to traffic contract rates allows the marker to determine what percentage of all packets sent should be marked as green, yellow and red, respectively. Marking probability function can be defined based on that interpretation of traffic contract as follows:

- probability of marking a packet as green:

$$p_{green}(s) = \begin{cases} 1 & \text{if sending rate } s \leq CIR \\ \frac{CIR}{s} & \text{otherwise} \end{cases} \quad (1)$$

- probability of marking a packet as yellow:

$$p_{yellow}(s) = \begin{cases} 0 & \text{if sending rate } s \leq CIR \\ 1 - \frac{CIR}{s} & \text{if } CIR < \text{sending rate } s \leq PIR \\ \frac{PIR - CIR}{s} & \text{otherwise} \end{cases} \quad (2)$$

- probability of marking a packet as red:

$$p_{red}(s) = \begin{cases} 0 & \text{if sending rate } s \leq PIR \\ \frac{s - PIR}{s} & \text{otherwise} \end{cases} \quad (3)$$

The overall marking probability as a function of sending rate has the shape of a curve. The desirable form of that curve is concave-convex. Using that form allows to mark a small percentage of yellow and red packets initially, increase that percentage more rapidly as the sending rate approaches the value of PIR, and then gradually level off (see fig.3) In figure 3 we can define the overall drop precedence as:

$$d_1 * p_{green}(s) + d_2 * p_{yellow}(s) + d_3 * p_{red}(s)$$

where: d_1, d_2 and d_3 are numerical values chosen to represent three drop precedence levels.

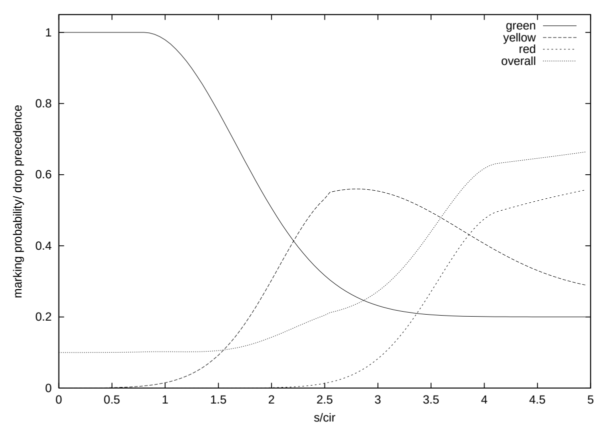

Figure 3: Marking probability and Overall Drop Precedence

The rest of the paper is organized as follows: Section 4 contains a short overview of different metering methods. Since marking probability is a function of the sending rate, the metering method used to obtain the value of its estimate has a significant influence on the marking decision. The key characteristics is how sensitive it is to transients, which is discussed in section 6.2. Section 5 contains evaluation of the random marking scheme by simulations. The aspect of *fairness* among flows is examined in section 5.2. The random marking is performed on the aggregate traffic and it is important to assure that none of the flows is discriminated and experiences lower than the others share of green and yellow bandwidth. A modification to the marking probability function that allows controlling the rate at which burst of packets are sent is proposed in section 5.3. The analysis of RPM in terms of how strictly the contract between subscriber and ISP is, and can be, respected is presented in section 6.1. We end with conclusions in section 7.

4 Metering Methods

Measurement of the sending rate can be performed in several different ways. Usually the estimate of the current rate is obtained first, and then the average value over a period of time is calculated. The current value can be obtained using:

- instantaneous sending rate

$$current\_rate = \frac{pkt\_size}{interpkt\_time}$$

where: interpkt_time is the time between the current and previous packet arrival.
The estimate obtained using this method depends strongly on the inter-packet arrival time. Past history is decayed over packet arrival and not over time [5].

- Time Sliding Window

$$current\_rate = \frac{win\_length * avg\_rate + pkt\_size}{win\_length + interpkt\_time}$$

Time sliding window [5] was designed to eliminate dependency on the inter-packet arrival time. In TSW the estimate is computed upon each packet arrival but TSW maintains a window length of history. It decays the estimated sending rate by a factor of e over window length period of time. The value of window length is a configurable parameter. With the smaller window length the estimate follows the actual value of sending rate more closely.

The estimate of the current sending rate can be next used to calculate an exponentially weighted moving average which gives an option of smoothing the estimated value by filtering out the noise.

$$avg\_rate = (1 - w) * current\_rate + w * avg\_rate$$

Weight w used for that process can be computed either statically or dynamically. If a static value is used, the estimated rate may be sensitive to the packet length distribution. Using a dynamic value which is a function of inter-packet arrival time, allows to eliminate that sensitivity [9]. Thus, when w is dynamically computed

$$avg\_rate = (1 - e^{-\frac{interpkt\_time}{K}}) * current\_rate + e^{-\frac{interpkt\_time}{K}} * avg\_rate$$

where K is a constant (see [9] for more details regarding the choice of its value).

5 Implementation and Simulation Results

5.1 Simulation Topology

In this section we evaluate random marking scheme by simulations. We examine the throughput of green, yellow and red traffic as it leaves the marker to check the conformance with the contract on which the marking is based. Next we analyze the fairness among all flows which constitute aggregate traffic passing through the marker. In the last subsection we present and evaluate a modification to the random marking scheme which allows packet burst control.

RPM is implemented as a module added to the link object in ns-2.1b6. It performs marking on all packets sent on the link. Marking probability function is implemented as a linear approximation of marking probability curve. Metering is performed using time sliding window [5] and exponentially weighted moving average. Figure 4(a) presents simulation topology. Aggregate traffic consists of packets sent by multiple persistent sources Src_1, ..., Src_n sending either UDP or TCP traffic. All packets are assumed to belong to the same AF class and all have the same size. The traffic profile specifies committed information rate set to 0.3 Mbps, peak information rate set to 0.4 Mbps and burst size set to 0 Mb. Router uses GRED, which is an extension of RIO [5] with multiple drop probabilities to support multiple drop priorities, as a queuing policy. A virtual queue within a physical queue is defined for each drop priority with a separate set of parameters equivalent to the set of RED parameters.

TCM was implemented as a set of two token buckets. Bucket C is filled at the rate equal to CIR, bucket P is filled at the rate equal to PIR-CIR. If there is a token present in the bucket C (green token), packet can be marked as green. If there is a token present in bucket P (yellow token) and no green token is available, packet can be marked as yellow. Otherwise it is marked as red.

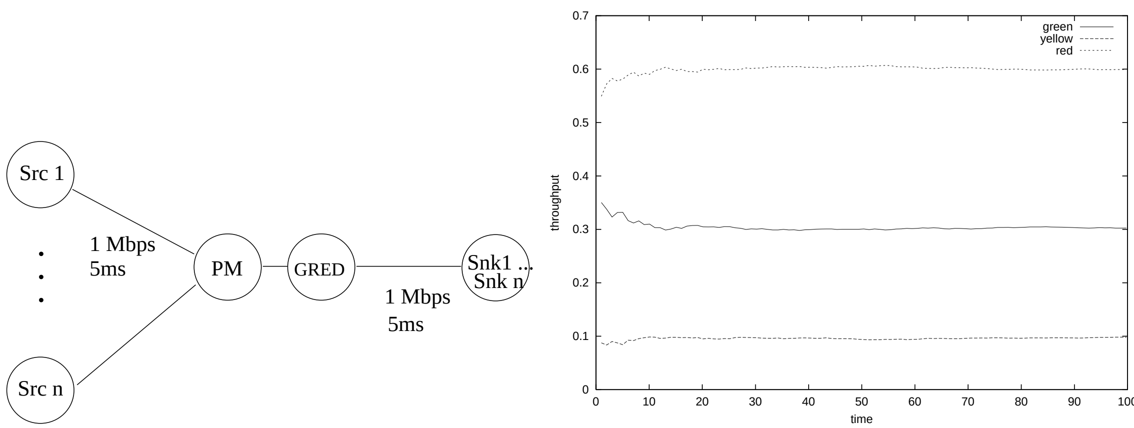

Figure 4: (a) Simulation topology (b) Traffic marking results

5.2 Performance Evaluation

In the first experiment we placed a CBR source at Src_1 (see fig.4(a)) The source is sending traffic over UDP at the rate of 1 Mbps for 100 s. Since $CIR = 0.3$ Mbps and $PIR = 0.4$ Mbps, the value of marking probability for green and yellow packets is 0.3 and 0.4, respectively. We expect 30% of packets to be marked as green and another 10% to be marked as yellow. The remaining 60% of packets should be marked as red. Figure 4(b) presents the results of the simulation. The x-axis represents time and the y-axis represents the throughput measured for each color on the link between packet marker and GRED queue. The throughput is measured as a total amount of data marked with each color since the beginning of the simulation divided by the time over which it was sent. The throughput is approximately 0.3 Mbps for green packets, 0.4 Mbps for yellow packets, and 0.6 Mbps for red packets, as expected.

Since random packet marking is performed on the aggregate traffic, fairness is an important metric that needs to be considered. Each flow which belongs to an aggregate traffic should get a fair share of green ad yellow throughput. By a “fair share” we mean an amount which is proportional to the flow's share of total bandwidth used by the aggregate traffic. Suppose that the percentage of green packets for the aggregate traffic as indicated by marking probability function is x. Since x% of all packets is marked as green at random, each flow will have x% of its packets marked as green. Similarly, the same percentage of packets, say y %, is marked as yellow for each flow. Thus, with random marking, the fraction of packets marked as green, yellow and red for each flow is roughly proportional to that flow's share of bandwidth.

We have verified RPM's fairness by conducting the following experiment. Six FTP sources sending TCP traffic were placed in Src_1, ..., Src_6, respectively (see fig.4). The first set of simulations was conducted with random packet marker, the second set with TCM marker. For both markers parameters were $CIR = 0.3$ Mbps, $PIR = 0.4$ Mbps and burst size 0 Mbps. The number of green, yellow and red packets was counted for each flow after marking was performed and before the packets entered GRED queue.

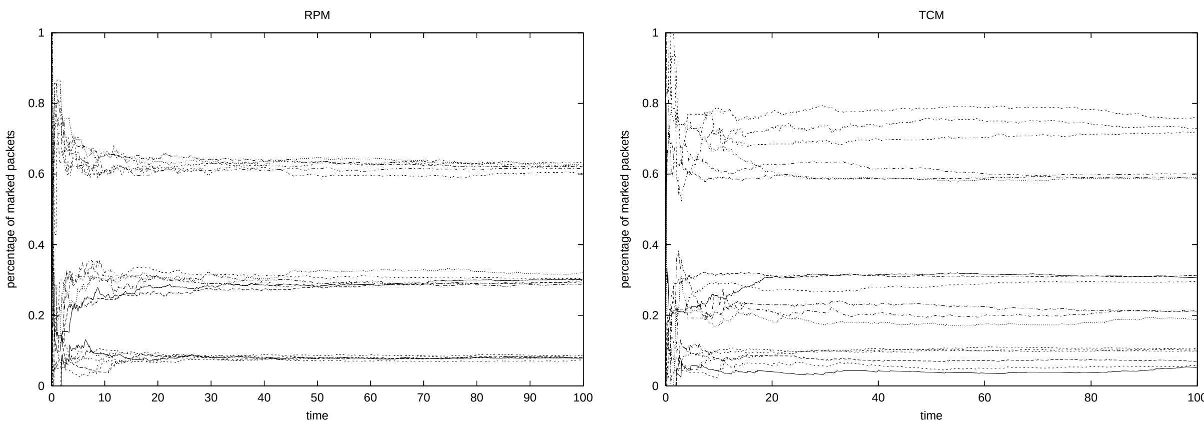

Figure 5: Percentage of packets marked for each of 6 TCP flows

Figure 5. presents results of both sets of simulations. The x-axis represents time and the y-axis represents the percentage of green, yellow and red packets marked for each flow. The percentage was calculated by dividing the number of packets of each color by the number of all packets received for a given flow since the beginning of the simulation. With RPM all flows see similar percentage of packets marked with each drop precedence. TCM marking on the other hand, shows higher differences. The percentage of red packets for example, varies from 60% to almost 80%. Discrimination experienced by some flows under TCM is a result of synchronization between traffic sources and token buckets. Assume that two sources are sending packets during the same token generation interval (time needed to generate an amount of tokens equivalent to the packets size) for bucket C. If packets from one of the sources always arrive first just after the needed amount of tokens becomes available, packets from the second source will never have a chance to seize any token from that bucket. In order to confirm this observation, we have conducted the following experiment.

Two CBR sources were placed at Src_1 and Src_2, respectively. Each source is sending UDP traffic at the rate 0.5 Mbps. Packets generated by both sources have all the same size.

Packets always arrive in a group of two, and in each pair, packets from Src_1 always arrive first. Yellow tokens are generated at 0.1 Mbps which is 3 times slower than green tokens (0.3 Mbps). In other words, yellow token is generated at the same time as every third green token (see fig. 6). All green tokens are acquired by the flow whose packets always arrive first. In every third group of

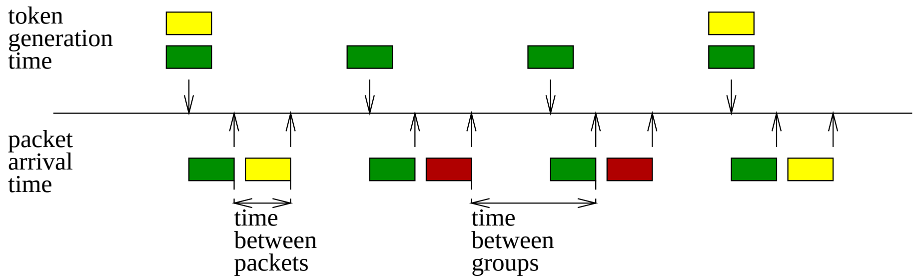

Figure 6: Synchronization of CBR sources

two packets, the second packet acquires yellow token after the green one is consumed by the first packet. The first flow is then deprived of all yellow tokens. As a result all green packets belong to the first flow, while all yellow packets belong to the second one.

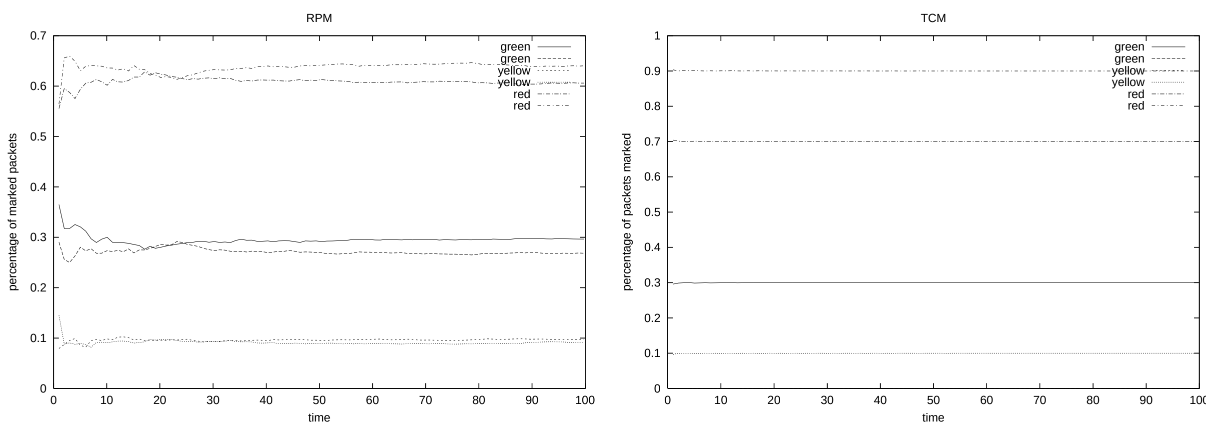

Figure 7: Marking performed for two synchronized CBR sources

Figure 7 presents the percentage of packets marked for each of two flows. 30% of all packets should be marked as green and 10% as yellow. As we can observe 30% of Src_1 flow packets are marked as green with remaining 70% marked as red. Src_2 flow has 10% of packets marked as yellow and the remaining 90% as red.

Random marking gives both flows better chance of acquiring tokens. There are still some differences between the percentage of packets marked with each color for both flows. The estimate of aggregate sending rate s calculated upon each packet arrival is always higher for the packet from Src_2. Src_2 flow experiences smaller than Src_1 flow probability of marking packets as green ($\frac{CIR}{s}$). The longer the interval of time between packet groups compared to the time interval between packets within the group, the higher the difference in marking probability (see fig.6). Using the average value of sending rate measured over long period of time would eliminate these differences.

5.3 Burst Control

Traffic profile may specify, besides CIR and PIR, also a burst size i.e., the number of packets that can be transmitted back-to-back at the peak rate. Since source is allowed to send that amount of data as a burst, all packets within the allowed burst size should be marked as green, as long as the sending rate does not exceed the peak rate. If the burst control is a desirable feature, random marking rules can be modified to allow it.

The enhancement has to take burst size and peak rate into account. Whenever a user is allowed to send a burst of packets, marking probability should be equal to 1 if the sending rate remains below the peak rate. Above the peak rate, its value corresponds to the ratio of peak rate to sending rate. In this way the green marking probability gets increased at the cost of red and possibly yellow marking probability. The overall drop precedence is temporarily decreased as we ignore the committed rate. Once the whole burst of packets is sent, the overall drop precedence is increased back to its original value as defined based on the committed and peak rates.

In the implementation two variables are maintained: the *maximum burst size* as defined in the contract, and the actual *burst size*. The latter varies between 0 and the maximum burst size. As long as the sending rate is below CIR, the customer is not using all bandwidth he or she is entitled to and the actual burst size can be increased until it reaches the maximum value specified. The increase is gradual, the smaller the ratio of sending to committed rate, the faster it is accumulated. If the burst size is larger than 0 and the rate is higher than CIR, the size of each packet sent as green is subtracted from the burst size. Green marking probability is modified as follows:

- if sending rate $s \leq CIR$
  $p_{green}(s) = 1$
  $burst\_size = max(burst\_size + (CIR - s) * interpacket\_time, max\_burst)$
- if $CIR <$ sending rate $s \leq PIR$, $burst\_size > 0$ and $pkt\_size \leq burst\_size$
  $p_{green}(s) = 1$
  $burst\_size = burst\_size - packet\_size$
- if sending rate $s > PIR$, $burst\_size > 0$ and $pkt\_size \leq burst\_size$
  $p_{green}(s) = \frac{PIR}{s}$
  if packet is marked as green
  $burst\_size = burst\_size - packet\_size$

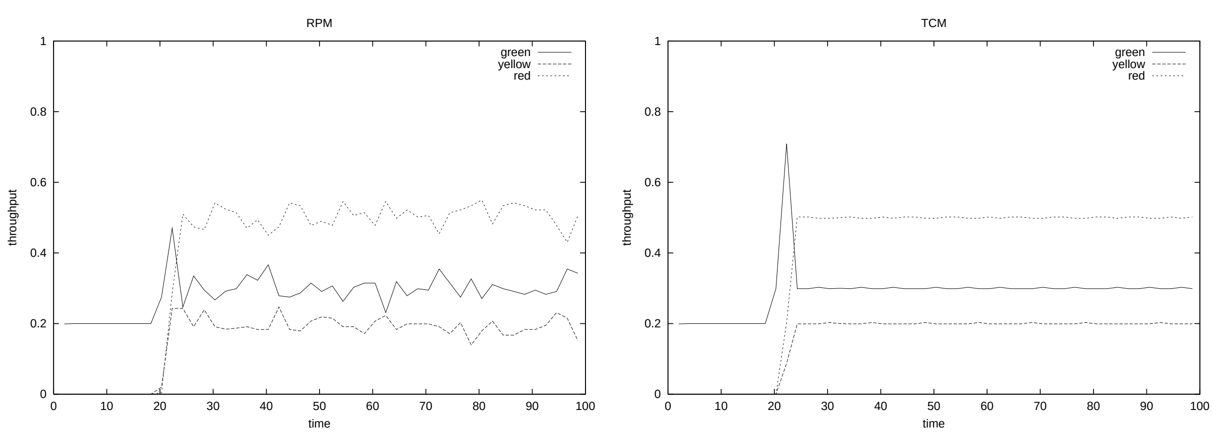

Figure 8: Green throughput with burst control

The following set of simulations was conducted to verify burst control. A CBR source placed at Src_1 (see fig.4) generates UDP traffic at the rate 0.2 Mbps for 20 s. After 20 s the sending rate is increased to 1 Mbps for another 80s. Marking is performed by RPM in the first set of simulations, and by TCM in the second set. The marker parameters for both RPM and TCM are: $CIR = 0.3$ Mbps, $PIR = 0.5$ Mbps, and *burst size* = 1 Mb. The throughput of green, yellow and red packets was measured just before marked traffic enters GRED queue. Figure 8 presents the simulation results. The x-axis represents time and the y-axis represents throughput. Each point represents the value of throughput measured over time period of 1 s. The throughput obtained for green and yellow packets with RPM marking oscillates around the prescribed rates. It is a result of variation in the number of packets marked randomly with each of the colors (for more detailed explanation see section 6.1).

During initial 20 s the sending rate of green packets is 0.2 Mbps. All packets arriving at the marker are marked as green since the sending rate is lower than CIR. The burst size accumulated during that time is max(1Mb, 20s · (0.3 - 0.2)Mbps = 1Mb. After 20 s the sending rate of green packets starts increasing and reaches the value of PIR: 0.5 Mbps. It takes 2s to send the whole burst of packets at that rate ($\frac{1Mb}{0.5Mbps} = 2s$). Once the actual burst size becomes 0, the sending rate for green packets decreases to the prescribed value of CIR: 0.3 Mbps. We can also observe that sending rate of yellow packets starts increasing after 20 s and reaches value of 0.2 Mbps.

The throughput of green packets obtained with TCM marking is similar during the first 20 s. After that time bucket C, whose depth is set to 1Mb, is full of tokens. As the sending rate increases to 1 Mbps, the sending rate of green packets also starts increasing, however this time it reaches value higher than PIR, about 0.7 Mbps. As long as there are tokens available in bucket C, all packets arriving at the marker are marked as green. The sending rate for green packets is equal to total sending rate. It takes about 1 s to empty the bucket at the rate of 1 Mbps and we can observe that the peak of the green sending rate in figure 8 is higher and more narrow for TCM than it is for RPM. The rate at which the burst can be sent is not limited by TCM. RPM, on the other hand, allows us to control that rate.

6 Discussion

6.1 Analysis of RPM

Packet marking at the network edge is based on the estimate of the transmission rate and rates specified in the ISP - customer contract. Regardless of who performs the marking, customer or ISP, the marked traffic should conform to the contract. The ISP and customer should agree on how strict the required conformance is. It may be defined in terms of the time scale over which the transmission rate of green, yellow and red packets is measured and the amount of data marked with each priority level that can be sent during the specified interval of time.

In this section we will derive the expected value of the number of packets marked as green as well as the expected value of the rate of green packets s_g assuming that marking probability function $p_{green}(s)$ according to equation (2) is used by the marker. We will calculate the probability that the expected value of s_g is exceeded or not reached. Based on that analysis two modified marking probability functions, $p_{green\_1}(x)$ and $p_{green\_2}(x)$, are proposed, followed by the similar derivation and the analysis of the expected value of s_g.

Consider traffic arriving at the marker at the constant rate s which is higher than CIR. Assume that n packets arrive during time interval τ so that $s = \frac{n \cdot pkt\_size}{\tau}$. Assume further for simplicity but without loss of generality that all packets have size pkt_size and that they all arrive at the equal time intervals *interpkt_time* (see figure 9). When packets have different sizes and arrive at arbitrary times, the average packet size and the average inter-packet time should be considered.

Since the overall sending rate s exceeds CIR, the sending rate of green packets s_g should ideally be equal to CIR. Let X denote the number of packets marked as green out of n packets arriving at the marker, and $P(X = k, n)$ denote probability that exactly k out of n packets are marked as green.

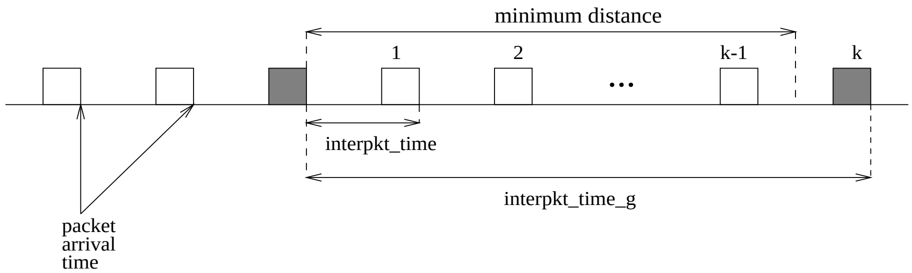

Figure 9: Packet arrival times at the marker.

Since the arrival rate s is constant, the probability of marking a packet as green $p_{green}(s) = \frac{CIR}{s}$ is the same for each packet, and the number of packets marked as green has binomial distribution.

$$P(X = k, n) = p_{green}(s)^k (1 - p_{green}(s))^{n-k} \binom{n}{k}$$

and

$$E(X, n) = \sum_{k=0}^{n} P(X = k, n) \cdot k = n \cdot p_{green}(s)$$

If $E(X, n)$ packets are marked as green, then the expected value of sending rate of green packets is :

$$E(s_g) = \frac{n \cdot p_{green}(s) \cdot pkt\_size}{\tau} = s \cdot p_{green}(s) = CIR$$

The expected value conforms then to the contract. The probability that the number of packets marked as green exceeds the expected value, which means that the sending rate of green packets is higher than CIR, is:

$$P(X > E(X, n)) = \sum_{k=n \cdot p_{green}(s)+1}^{n} p_{green}(s)^k (1 - p_{green(s)})^{n-k} \binom{n}{k}$$

Figure 10 shows how that probability changes for different values of marking probability $p_{green}(s)$ and different number of packets arriving at the marker taken into account.

The probability of exceeding $E(X, n)$ reaches its maximum for $p_{green}(s)$ approximately between 0.6 and 0.7, and gets higher for the larger number of packets arriving at the marker. Probability of not reaching the expected value of the number of green packets has similar distribution. It is possible then, that the actual sending rate of green packets will either exceed or will not reach the expected value $E(s_g)$ over some period of time.

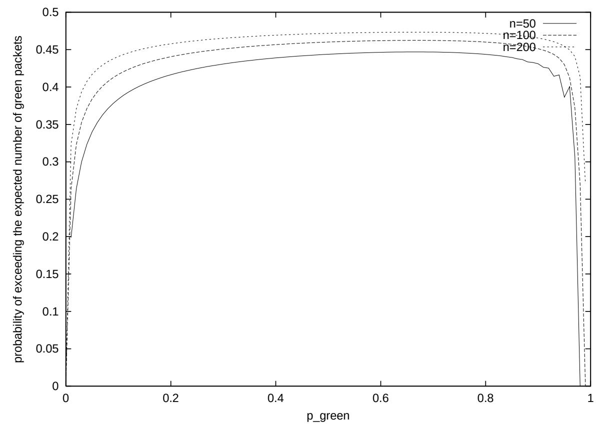

Figure 10: Probability of exceeding $E(X)$ for p_{green_s}

In order to find out how evenly green packets are distributed, let $h(k)$ denote the probability that k^{th} packet that arrives after a previous green packet, is marked as green, assuming that the previous $k-1$ packets were not green (see fig. 9). Let Y denote the number of yellow and red packets separating two green packets.

$$P(Y=k) = h(k+1) \cdot \prod_{i=1}^{k} (1 - h(i))$$

The expected value of Y is:

$$E(Y) = \sum_{k=1}^{\infty} P(Y=k) \cdot k$$

From the definition of $p_{green}(s)$ it follows that:

$$h(k) = p_{green}(s)$$

$$P(Y=k) = (1 - p_{green}(s))^k \cdot p_{green}(s)$$

Figure 11 shows how the probability $P(Y=k)$ changes depending on the value of k. $P(Y=0)$ is the most likely value as k increases. The expected value of Y is:

$$E(Y) = \sum_{k=0}^{\infty} \left((1 - p_{green}(s))^k \cdot p_{green}(s) \cdot k \right) = \frac{1 - p_{green}(s)}{p_{green}(s)} = \frac{s}{CIR} - 1$$

As it appears, even though the expected value of s_g is equal to CIR, the actual value measured over a short period of time may be different. If more strict conformance to the contract is required, the marking probability function has to be modified. One way to control the number of green packets arriving during a short period of time is to take the inter-packet spacing into account.

The minimum time distance between two green packets is indicated by CIR (see fig. 9). In order to respect that requirement, each packet's probability of being marked as green should be an

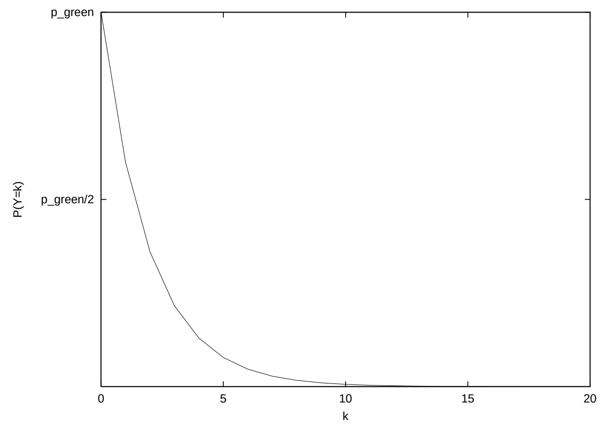

Figure 11: P(Y = k) when $p_{green}(s)$ is used as marking probability.

s	overall traffic arrival rate
s_g	arrival rate of green packets
$interpkt\_time$	time between previous and current packet arrival
$interpkt\_time\_g$	time between the arrival of the last packet marked as green and current packet
X	number of packets marked as green out of n packets arriving at the marker
$P(X = k, n)$	probability that exactly k packets out of n are marked as green
Y	number of packets that arrive after a green packet, until the next packet is marked as green
$h(k)$	probability that k^{th} packet that arrive after a previous green packet is marked as green assuming that the previous $k-1$ packets were not green

Table 1: Notation

increasing function of inter-packet arrival time. The value of the function should be between 0 for a packet arriving exactly at the same time as the previous one, and 1 for any value of $interpkt\_time\_g$ higher than the minimum. In the extreme case only these two values can be used as follows:

$$p_{green\_1}(x) = \begin{cases} 0 & if x < 1 \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

$$x = \frac{CIR}{s_g}$$
$$s_g = \frac{pkt\_size}{interpkt\_time\_g}$$

where s_g is the sending rate of the green packets,
$interpkt\_time\_g$ is the time since the arrival of the last packet marked as green.

This type of marking probability is based on the measurement of the instantaneous sending rate for green packets defined as a ratio of packet size to time since the arrival of last packet which was marked as green. The inter-packet time smaller than the minimum corresponds to the green sending rate being higher than CIR. If the green sending rate is lower than CIR, the next packet can be marked as green without violating the traffic profile.

Consider the expected value of the sending rate of green packets s_g when $p_{green\_1}(x)$ is used as marking probability. From equation (4) it follows that:

$$h(k) = \begin{cases} 0 & \text{if } \frac{pkt\_size}{k \cdot interpkt\_time} > CIR \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

The expected value of number of packets between two green packets is:

$$E(Y) = \sum_{k=0}^{\infty} \left(h(k+1) \cdot \prod_{i=1}^{k} (1 - h(i)) \cdot k \right)$$

From the definition of $h(k)$ given in equation (5) it follows there is only one value of k such that $P(Y = k) > 0$. Assume that k_0 is the largest value of k satisfying the following condition:

$$\frac{pkt\_size}{k \cdot interpkt\_time} > CIR \tag{6}$$

which means that $(k_0 + 1)^{th}$ packet is the first packet that arrives at the distance higher than the minimum from the previous green packet. It follows that $h(k) = 0 \;\; \forall k \leq k_0$ and consequently:

$$\prod_{k=1}^{k_0} (1 - h(k)) = 1$$

At the same time $h(k_0 + 1) = 1$ since

$$\frac{pkt_size}{(k_0 + 1) \cdot interpkt\_time} \leq CIR$$

Therefore $P(Y = k_0) = 1$, and k_0 is the only possible value of Y. From (6)

$$k_0 < \frac{pkt\_size}{interpkt\_time} \cdot \frac{1}{CIR} = \frac{s}{CIR}$$

Since k_0 is the biggest integer satisfying that condition

$$k_0 = \lceil \frac{s}{CIR} - 1 \rceil \tag{7}$$

and

$$E(Y) = \lceil \frac{s}{CIR} - 1 \rceil \tag{8}$$

Now, the probability of exceeding the expected value of Y is:

$$P(Y > E(Y)) = 0$$

Similarly the probability that the value of Y is below the expected value is:

$$P(Y < E(Y)) = 0$$

Since there is only one possible value of Y: k_0, there is also only one possible value of the number of packets marked as green out of n packets:

$$X = \frac{n}{k_0 + 1} = \frac{n}{\lceil \frac{s}{CIR} - 1 \rceil + 1}$$

Probability that the number of packets marked as green is lower or bigger than the above value is equal to 0. The value of the sending rate of green packets is then:

$$s_g = \frac{n \cdot pkt\_size}{(\lceil \frac{s}{CIR} - 1 \rceil + 1) \cdot \tau} = \frac{s}{\lceil \frac{s}{CIR} - 1 \rceil + 1}$$

Since it is the only possible value of s_g, then

$$E(s_g) = \frac{s}{\lceil \frac{s}{CIR} - 1 \rceil + 1}$$

If $\frac{s}{CIR}$ is an integer then $E(s_g) = CIR$, otherwise the expected sending rate of green packets will be lower. Since the probability that number of packets marked as green is higher or lower than the expected value is 0, the actual sending rate of green packets will be equal to its expected value. It will not exceed CIR, however it may remain below it.

The rate at which green packets are sent, should not only conform to the committed rate but also should not stay below it if the aggregate sending rate exceeds CIR. For a packet of given size the value of CIR indicates what the minimum distance from the previous green packet should be. If we mark as green not only packets that arrive at the higher that the minimum distance from the previous green packet, but also packets which arrive after shorter than the minimum period of time and sufficiently close to that minimum, the green sending rate will be closer to CIR. Marking probability can be modified to reflect that change as follows (see fig.12):

$$p_{green\_2}(x) = \begin{cases} e^{\frac{x-1}{a}} + e^{\frac{-1}{a}}(x-1) & \text{if } \frac{pkt\_size}{interpkt\_time\_g} > CIR \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

$$\begin{array}{lll} where: & x = & \frac{CIR}{s_g}, \\ & s_g = & \frac{pkt\_size}{interpkt\_time\_g}, \\ & a & \text{is a configurable parameter} \in (0,1] \end{array}$$

The value of the function is very small for a high sending rate of green packets (small inter-packet time). It increases rapidly as the green sending rate gets close to the CIR and reaches 1 for green sending rate equal or smaller than CIR. By changing the value of parameter a, we can decide how strictly the minimum interval requirement is enforced. The stricter it is, the higher chances that the green sending rate will remain below CIR. On the other hand, by allowing some violation of the minimum spacing, we are allowing the sending rate of green packets to exceed CIR.

From the definition of $p_{green\_@}(x)$ (9) it follows that:

$$h(k) = \begin{cases} e^{\frac{\frac{pkt\_size}{k \cdot interpkt\_time} - 1}{a}} + e^{\frac{-1}{a}}(\frac{pkt\_size}{k \cdot interpkt\_time} - 1) & \text{if } \frac{pkt\_size}{k \cdot interpkt\_time} > CIR \\ 1 & \text{otherwise} \end{cases} \tag{10}$$

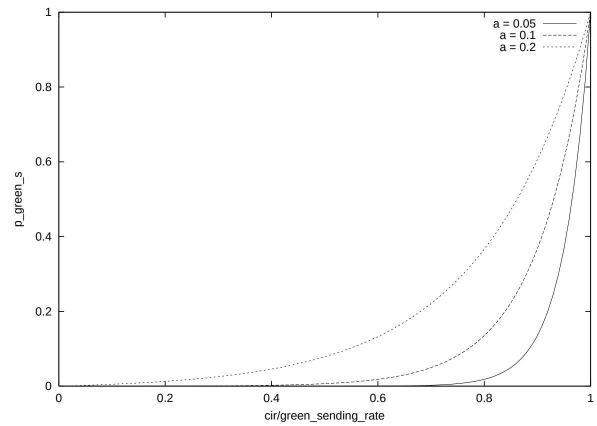

Figure 12: Marking probability function based on instantaneous green sending rate

The marking probability function is now more flexible. As a result there are more possible values for random variable Y. Assume, as previously, that k_0 is the biggest integer satisfying (6), then

$$k_0 = \lceil \frac{s}{CIR} - 1 \rceil$$

From (10)

$$P(Y = k) = 0 \quad for \quad k > k_0 = \lceil \frac{s}{CIR} - 1 \rceil$$

since

$$(1 - h(k)) = 0 \ \ \forall k > \lceil \frac{s}{CIR} - 1 \rceil$$

In other words, the probability that there will be more than k_0 packets separating two green packets (see fig. 9), is 0. On the other hand it is possible that the number of yellow and red packets separating two green packets is smaller or equal to k_0 since the probability of marking a packet arriving as a shorter than the minimum distance as green is higher than 0.

$$P(Y = k) > 0 \quad for \quad k \leq k_0 = \lceil \frac{s}{CIR} - 1 \rceil$$

Therefore $Y \in [0, \lceil \frac{s}{CIR} - 1 \rceil]$.

$$E(Y) = \sum_{k=0}^{\infty} P(Y = k) \cdot k = \sum_{k=0}^{k_0} P(Y = k) \cdot k$$

and

$$\sum_{k=0}^{k_0} P(Y = k) \cdot k < \sum_{k=0}^{k_0} P(Y = k) \cdot k_0 = k_0 \sum_{k=0}^{k_0} P(Y = k) = k_0 = \lceil \frac{s}{CIR} - 1 \rceil$$

The value of $E(Y)$ calculated for p_{green_2} is then lower then $E(Y)$ for p_{green_1} (8). As a result of more "flexible" marking, it is possible that the value of Y will exceed or will not reach $E(Y)$.

Since $E(Y) < \lceil \frac{s}{CIR} - 1 \rceil$ and $P(Y = k) > 0$ for $k \leq \lceil \frac{s}{CIR} - 1 \rceil$:

$P(Y < E(Y)) > 0$ and $P(Y > E(Y)) > 0$.

Since the expected value of Y is lower compared to (8), we can expected $E(X)$ to be higher, and the value of s_g to be higher also. The more flexibility marking probability function $p_{green\_2}(x)$ may result in the sending rate of green packets exceeding CIR.

The degree of flexibility is controlled by the value of parameter a in the definition (9) of $p_{green\_2}(x)$. The higher the value of a, the higher probability that $Y < \lceil \frac{s}{CIR} - 1 \rceil$. Generally, probability of s_g exceeding CIR is proportional to the value of $\frac{1}{a}$.

Consider an example where $\frac{s}{CIR} = 0.3$. Figure 13 shows how value of $E(Y)$ and $P(Y < \lceil \frac{s}{CIR} - 1 \rceil)$ changes depending on the value of a. Marking probability function $p_{green\_1}(x)$ yields $E(Y) = 3$.

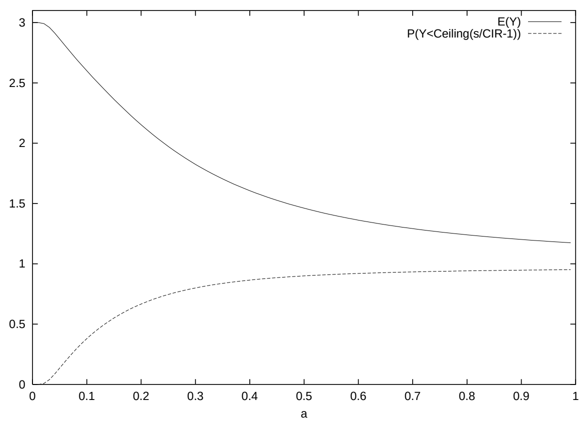

Figure 13: Dependence of $E(Y)$ and $P(Y < \lceil \frac{s}{CIR} - 1 \rceil)$ on a

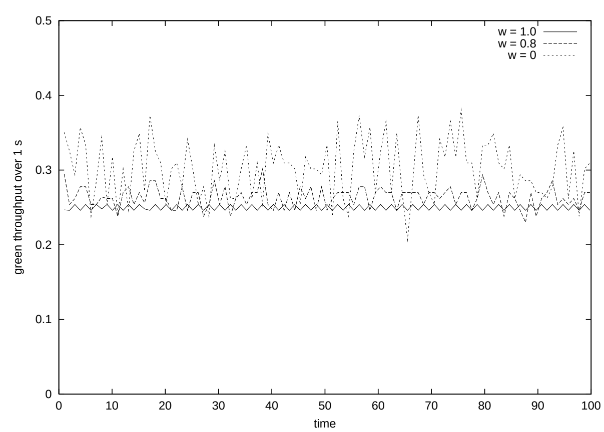

Figure 14: Green throughput obtained for different weight values

The initial marking probability function $p_{green}(x)$ offers fairness among flows in the aggregate traffic, the modified marking probability functions $p_{green\_1}(x)$ and p_{green_2} allow to control inter-packet distance. By combining these two groups of methods we can reach a tradeoff between being fair with possible violation of the minimum distance between two green packets, and respecting the traffic profile specification. For each packet, both the probability based on the value of average sending rate and distance in time from the previous green packet should be taken into account. The final marking probability can be then defined as a weighted average of the probability as defined in equation (2) and probability based on the ratio between CIR and instantaneous green sending rate (equation 4).

$$mprob = w * p_{green\_1}(x) + (1 - w) * p_{green}(s)$$

Changing the value of the weight w gives us an option of moving more towards fairness or more towards respecting the minimum inter-packet distance. Figure 14 presents green throughput obtained using different values of w to determine marking probability for traffic from a CBR source.

6.2 Metering Method Influence

Marking probability is a function of the sending rate and for that reason accuracy of the metering method has a significant impact on the marking process. Consider two methods of obtaining an estimate of the sending rate, one which gives more accurate value reflecting all changes in the sending rate almost immediately, and another which gives an average value over longer period of time.

More accurate estimate of the sending rate is desirable for at least two reasons. It makes precise marking possible, i.e., more accurate number of packets being marked as green and yellow. Smooth estimate (average value) may be higher (or lower) than the actual value of sending rate over short periods of time, e.g., the average value may not reflect changes right away. Thus the percentage of packets that should be marked as green and yellow based on the value of average sending rate, may be too low at times. During network congestion the aggregate traffic may suffer more losses than it should, since red packets are likely to be dropped. In the opposite case, when the sending rate increases rapidly, a smooth estimate can result in a burst of packets marked as green. Again the average value may not reflect the change in sending rate for some time. In this case, the percentage of packets that should be marked as green and yellow will be higher than it is indicated by the value of the instantaneous sending rate. The marking probability function will have higher value and more packets that arrive at the marker will be marked as green and yellow.

The accuracy of the metering method used may also affect fairness. Consider example of two synchronized sources presented in section 5.2. An estimate of the sending rate is calculated upon each packet arrival. If the time between arrival of two packets within one group is longer than the time between the arrival of two groups, the more accurate of two metering methods will give a lower value of the sending rate estimate for the first packet in a group than for the second one (see fig. 6). Consequently, the marking probability for the first packet will be higher for than the second one. The flow whose packets always arrive in the second place in each group will experience lower percentage of its packets marked as green or yellow. With the average estimate of the sending rate this difference will be smaller if present at all.

Finally, also the conformance to the contract can be affected by the choice of the metering method. The average estimate of the sending rate is less sensitive to the inter-packet distance. Marking probability based on the calculation of the average value will allow more packets arriving at the shorter than the minimum distance from the previous green packet, to be marked as green (for discussion see section (6.1). Marking probability based on the more accurate estimate will be different for a packet arriving at the shorter than the minimum distance than for a packet arriving at the longer distance. The probability of marking the latter as green will be higher than the former.

7 Conclusions

Random packet marking is designed to be performed on the aggregate traffic passing through the access node at the domain boundaries. Decision about drop precedence for each packet is based on the estimate of the aggregate sending rate. Colors are assigned randomly with the probability which is a function of sending rate in respect to the committed and peak information rates. RPM does not require maintaining any per-flow state. The fraction of packets marked as green, yellow and red for each flow is roughly proportional to that flow's share of bandwidth. RPM offers also a tradeoff between enforcing the minimum inter-packet time spacing as indicated by the traffic profile and fairness as defined above. Marking probability function can be modified to allow bursts of packets to be sent and to control the rate at which they are sent.

References

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

[2] J. Hainanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999

[3] J. Hainanen, R. Guerin, "A Single Rate Three Color Marker", Internet Draft, May 1999

[4] J. Hainanen, R. Guerin, "A Two Rate Three Color Marker", Internet Draft, May 1999

[5] D. B. Clark, W. Fang, "Explicit Allocation of Best Effort Packet Delivery Service",

[6] H. Kim, "A Fair Marker", Internet Draft, April 1999

[7] A. Mehra, R. Tewari, D. Kandlur, "Design Consideration of Rate Control of Aggregated TCP"

[8] J. Ibanez, K. Nichols, "Preliminary Simulation Evaluation of an Assured Service", Internet Dratf, August 1998

[9] I. Stoica, S. Shenker, H. Zhang, "Core-Stateless Fair Queuing: Achieving Approximately Fair Bandwidth Allocations in High Speed Networks"