

Network Working Group
Request for Comments: 2859
Category: Experimental

W. Fang
Princeton University
N. Seddigh
B. Nandy
Nortel Networks
June 2000

A Time Sliding Window Three Colour Marker (TSWTCM)

Status of this Memo

This memo defines an Experimental Protocol for the Internet community. It does not specify an Internet standard of any kind. Discussion and suggestions for improvement are requested. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2000). All Rights Reserved.

Abstract

This memo defines a Time Sliding Window Three Colour Marker (TSWTCM), which can be used as a component in a Diff-Serv traffic conditioner [RFC2475, RFC2474]. The marker is intended to mark packets that will be treated by the Assured Forwarding (AF) Per Hop Behaviour (PHB) [AFPHB] in downstream routers. The TSWTCM meters a traffic stream and marks packets to be either green, yellow or red based on the measured throughput relative to two specified rates: Committed Target Rate (CTR) and Peak Target Rate (PTR).

1.0 Introduction

The Time Sliding Window Three Colour Marker (TSWTCM) is designed to mark packets of an IP traffic stream with colour of red, yellow or green. The marking is performed based on the measured throughput of the traffic stream as compared against the Committed Target Rate (CTR) and the Peak Target Rate (PTR). The TSWTCM is designed to mark packets contributing to sending rate below or equal to the CTR with green colour. Packets contributing to the portion of the rate between the CTR and PTR are marked yellow. Packets causing the rate to exceed PTR are marked with red colour.

The TSWTCM has been primarily designed for traffic streams that will be forwarded based on the AF PHB in core routers.

The TSWTCM operates based on simple control theory principles of proportionally regulated feedback control.

2.0 Overview of TSWTCM

The TSWTCM consists of two independent components: a rate estimator, and a marker to associate a colour (drop precedence) with each packet. The marker uses the algorithm specified in section 4. If the marker is used with the AF PHB, each colour would correspond to a level of drop precedence.

The rate estimator provides an estimate of the running average bandwidth. It takes into account burstiness and smoothes out its estimate to approximate the longer-term measured sending rate of the traffic stream.

The marker uses the estimated rate to probabilistically associate packets with one of the three colours. Using a probabilistic function in the marker is beneficial to TCP flows as it reduces the likelihood of dropping multiple packets within a TCP window. The marker also works correctly with UDP traffic, i.e., it associates the appropriate portion of the UDP packets with yellow or red colour marking if such flows transmit at a sustained level above the contracted rate.

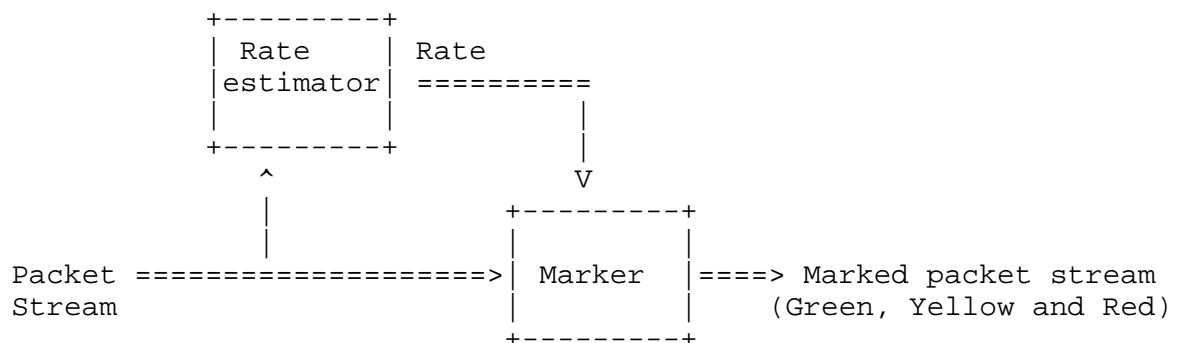


Figure 1. Block diagram for the TSWTCM

The colour of the packet is translated into a DS field packet marking. The colours red, yellow and green translate into DS codepoints representing drop precedence 2, 1 and 0 of a single AF class respectively.

Based on feedback from four different implementations, the TSWTCM is simple and straightforward to implement. The TSWTCM can be implemented in either software or hardware depending on the nature of the forwarding engine.

3.0 Rate Estimator

The Rate Estimator provides an estimate of the traffic stream's arrival rate. This rate should approximate the running average bandwidth of the traffic stream over a specific period of time (AVG_INTERVAL).

This memo does not specify a particular algorithm for the Rate Estimator. However, different Rate Estimators should yield similar results in terms of bandwidth estimation over the same fixed window (AVG_INTERVAL) of time. Examples of Rate Estimation schemes include: exponential weighted moving average (EWMA) and the time-based rate estimation algorithm provided in [TON98].

Preferably, the Rate Estimator SHOULD maintain time-based history for its bandwidth estimation. However, the Rate Estimator MAY utilize weight-based history. In this case, the Estimator used should discuss how the weight translates into a time-window such as AVG_INTERVAL.

Since weight-based Estimators track bandwidth based on packet arrivals, a high-sending traffic stream will decay its past history faster than a low-sending traffic stream. The time-based Estimator is intended to address this problem. The latter Rate Estimator utilizes a low-pass filter decaying function. [FANG99] shows that this Rate Estimator decays past history independently of the traffic stream's packet arrival rate. The algorithm for the Rate Estimator from [TON98] is shown in Figure 2 below.

```
=====
Initially:
```

```
    AVG_INTERVAL = a constant;
    avg-rate      = CTR;
    t-front       = 0;
```

```
Upon each packet's arrival, the rate estimator updates its variables:
```

```
    Bytes_in_win = avg-rate * AVG_INTERVAL;
    New_bytes     = Bytes_in_win + pkt_size;
    avg-rate      = New_bytes / ( now - t-front + AVG_INTERVAL );
    t-front       = now;
```

```
Where:
```

```
    now           = The time of the current packet arrival
    pkt_size      = The packet size in bytes of the arriving packet
    avg-rate      = Measured Arrival Rate of traffic stream
    AVG_INTERVAL  = Time window over which history is kept
```

Figure 2. Example Rate Estimator Algorithm

```
=====

The Rate Estimator MAY operate in the Router Forwarding Path or as a
background function. In the latter case, the implementation MUST
ensure that the Estimator provides a reasonably accurate estimation
of the sending rate over a window of time. The Rate Estimator MAY
sample only certain packets to determine the rate.
```

4.0 Marker

The Marker determines the colour of a packet based on the algorithm presented in Figure 3. The overall effect of the marker on the packets of a traffic stream is to ensure that:

- If the estimated average rate is less than or equal to the CTR, packets of the stream are designated green.
- If the estimated average rate is greater than the CTR but less than or equal to the PTR, packets are designated yellow with probability P0 and designated green with probability (1-P0). P0 is the fraction of packets contributing to the measured rate beyond the CTR.

```

=====
    avg-rate = Estimated Avg Sending Rate of Traffic Stream

    if (avg-rate <= CTR)
        the packet is green;
    else if (avg-rate <= PTR) AND (avg-rate > CTR)
        (avg-rate - CTR)
        calculate P0 = -----
                        avg-rate
        with probability P0 the packet is yellow;
        with probability (1-P0) the packet is green;
    else
        (avg-rate - PTR)
        calculate P1 = -----
                        avg-rate
        (PTR - CTR)
        calculate P2 = -----
                        avg-rate
        with probability P1 the packet is red;
        with probability P2 the packet is yellow;
        with probability (1-(P1+P2)) the packet is green;

    Figure 3. TSWTCM Marking Algorithm
=====

```

- If the estimated average rate is greater than the PTR, packets are designated red with probability P1, designated yellow with probability P2 and designated green with probability (1-(P1+P2)). P1 is the fraction of packets contributing to the measured rate beyond the PTR. P2 is the fraction of packets contributing to that part of the measured rate between CTR and PTR.

The marker MUST operate in the forwarding path of all packets.

5.0 Configuration

5.1 Rate estimator

If the Rate Estimator is time-based, it should base its bandwidth estimate on the last AVG_INTERVAL of time. AVG_INTERVAL is the amount of history (recent time) that should be used by the algorithm in estimating the rate. Essentially it represents the window of time included in the Rate Estimator's most recent result.

The value of AVG_INTERVAL SHOULD be configurable, and MAY be specified in either milliseconds or seconds.

[TON98] recommends that for the case where a single TCP flow constitutes the contracted traffic, AVG_INTERVAL be configured to approximately the same value as the RTT of the TCP flow. Subsequent experimental studies in [GLOBE99] utilized an AVG_INTERVAL value of 1 second for scenarios where the contracted traffic consisted of multiple TCP flows, some with different RTT values. The latter work showed that AVG_INTERVAL values larger than the largest RTT for a TCP flow in an aggregate can be used as long as the long-term bandwidth assurance for TCP aggregates is measured at a granularity of seconds. The AVG_INTERVAL value of 1 second was also used successfully for aggregates with UDP flows.

If the Rate Estimator is weight-based, the factor used in weighting history - WEIGHT - SHOULD be a configurable parameter.

The Rate Estimator measures the average sending rate of the traffic stream based on the bytes in the IP header and IP payload. It does not include link-specific headers in its estimation of the sending rate.

5.2 Marker

The TSWTCM marker is configured by assigning values to its two traffic parameters: Committed Target Rate (CTR) and Peak Target Rate (PTR).

The PTR MUST be equal to or greater than the CTR.

The CTR and PTR MAY be specifiable in bits per second or bytes per second.

The TSWTCM can be configured so that it essentially operates with a single rate. If the PTR is set to the same value as the CTR then all packets will be coloured either green or red. There will be no yellow packets.

If the PTR is set to link speed and the CTR is set below the PTR then all packets will be coloured either green or yellow. There will be no red packets.

6.0 Scaling properties

The TSWTCM can work with both sender-based service level agreements and receiver-based service level agreements.

7.0 Services

There are no restrictions on the type of traffic stream for which the TSWTCM can be utilized. It can be used to meter and mark individual TCP flows, aggregated TCP flows, aggregates with both TCP and UDP flows [UDPTCP] etc.

The TSWTCM can be used in conjunction with the AF PHB to create a service where a service provider can provide decreasing levels of bandwidth assurance for packets originating from customer sites.

With sufficient over-provisioning, customers are assured of mostly achieving their CTR. Sending rates beyond the CTR will have lesser assurance of being achieved. Sending rates beyond the PTR have the least chance of being achieved due to high drop probability of red packets.

Based on the above, the Service Provider can charge a tiered level of service based on the final achieved rate.

8.0 Security Considerations

TSWTCM has no known security concerns.

9.0 Acknowledgements

The authors would like to thank Juha Heinanen, Kenjiro Cho, Ikjun Yeom and Jamal Hadi Salim for their comments on earlier versions of this document. Their suggestions are incorporated in this memo.

10.0 References

- [TON98] D.D. Clark, W. Fang, "Explicit Allocation of Best Effort Packet Delivery Service", IEEE/ACM Transactions on Networking, August 1998, Vol 6. No. 4, pp. 362-373.
- [RFC2474] Nichols, K., Blake, S., Baker, F. and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.
- [RFC2475] Black, D., Blake, S., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.
- [FANG99] Fang, W. "The 'Expected Capacity' Framework: Simulation Results", Princeton University Technical Report, TR-601-99, March, 1999.

- [YEOM99] I. Yeom, N. Reddy, "Impact of Marking Strategy on Aggregated Flows in a Differentiated Services Network", Proceedings of IwQoS, May 1999.
- [AFPHB] Heinanen, J., Baker, F., Weiss, W. and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.
- [UDPTCP] P. Piedad, N. Seddigh, B. Nandy, "The Dynamics of TCP and UDP Interaction in IP-QoS Differentiated Service Networks", Proceedings of the 3rd Canadian Conference on Broadband Research (CCBR), Ottawa, November 1999
- [GLOBE99] N. Seddigh, B. Nandy, P. Piedad, "Bandwidth Assurance Issues for TCP flows in a Differentiated Services Network", Proceedings of Global Internet Symposium, Globecom 99, Rio De Janeiro, December 1999.

11.0 Authors' Addresses

Wenjia Fang
Computer Science Dept.
35 Olden Street,
Princeton, NJ08540

EMail: wfang@cs.princeton.edu

Nabil Seddigh
Nortel Networks,
3500 Carling Ave
Ottawa, ON, K2H 8E9
Canada

EMail: nseddigh@nortelnetworks.com

Biswajit Nandy
Nortel Networks,
3500 Carling Ave
Ottawa, ON, K2H 8E9
Canada

EMail: bnandy@nortelnetworks.com

12. Full Copyright Statement

Copyright (C) The Internet Society (2000). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

