Performance enhancement of an AF service using TCP-aware marking and dynamic WRED

Christos Bouras[1,2], Afrodite Sevasti[2,3]
[1]RA Computer Technology Institute-CTI, 61 Riga Feraiou Str, 26221 Patras, Greece
[2]Department of Computer Engineering and Informatics, University of Patras, 26500 Rion, Patras, Greece
[3]Greek Research and Technology Network-GRNET, 56 Mesogion Av., 11527, Athens, Greece
bouras@cti.gr, sevasti@grnet.gr

Abstract

The implementation of successful Assured Forwarding (AF) services according to the DiffServ framework remains a challenging problem today, despite the numerous proposals for AF PHB mechanisms and AF-based service implementations. In this work, we propose two modules, the TCP-Window Aware Marker (TWAM) and the Dynamic WRED (DWRED) mechanism for implementing the DiffServ AF PHB. We provide analytical models and an experimental evaluation in order to demonstrate how they succeed in enhancing the quality, increasing fairness, improving the performance and easing the deployment of a production- level AF-based service.

1. Introduction

The introduction of the Differentiated Services (DiffServ) framework has been a quite recent development in the direction of providing differential treatment to IP packets in backbone networks. In this work we deal with providing to IP traffic a service that is qualitatively better than that of the traditional 'best-effort' model, without deterministic, high-assurance quality guarantees. Such services are built on the Assured Forwarding Per Hop Behaviour (AF PHB) of the DiffServ framework, as defined in [1]. In our previous work ([18]) we have proposed and outlined the basic principles of a service based on the AF PHB, denoted as Relative service.

The effects of AF-based services provisioning to TCP and UDP traffic and a number of relative issues such as fairness among flows, achievable bandwidth guarantees and qualitative performance improvement achieved by AF traffic have been a research topic for some years now ([1], [2], [3], [5], [6]). A common conclusion is that the isolation of TCP flows from UDP traffic is compulsory. If we limit ourselves to TCP traffic, the parameters that affect/obstruct the provision of quantitative guarantees according to [4] are the effects of RTT (Round Trip Time) and the self-clocked sliding window mechanism, the number of concurrent flows in an AF class, the variety in packet sizes, the contracted AF capacity.

In [11] and [17], TCP friendly packet markers for TCP traffic are presented. In [7], the TSW packet marker (or Time Sliding Window Two Colour Marker-TSW2CM) is presented in an effort to achieve fairness in capacity distribution among flows based on the reserved capacity for each flow. The mechanism itself and certain variations appear often in the literature. Taking this into consideration, we have chosen to compare the performance of our proposed marker against the TSW2CM.

Apart from proposed marking mechanisms, a series of AQMs in the framework of 'Assured Services' also exist in the literature. RIO (RED with In/Out bit) is presented in [14], REM (Random Early Management) is introduced in [12] and FRED, an AQM mechanism that maintains statistics for average queue occupancy per flow as well as the current number of packets buffered per flow, is presented in [8]. [13] and [15] deal with how the closed-loop system of TCP traffic and RED, behaves. In [9], the authors propose Adaptive RED according to which the average queue length is constantly observed and updated. In [10], the problem of unfairness of bandwidth allocation among TCP flows under an AF-based service is addressed with the adoption of Core-Stateless Fair Queueing (CSFQ). In [16] a variation of RED is proposed where AQM thresholds apply only to Out packets.

In our work we attempt to address the limitations of the Relative service provisioning. The proposed TCP-

Window Aware Marker (TWAM), applied at the network ingress, addresses the unfairness issues with much less processing overhead than other proposed marking schemes. The proposed Dynamic Weighted Random Early Detection (DWRED) mechanism, applied at the core routers, achieves much higher utilization than other Active Queue Management (AQM) mechanisms, while it successfully adapts to changing network conditions.

In section 2 we give a brief overview of the Relative service provisioning principles. In sections 3 and 4 the proposed mechanisms are presented and approached theoretically while in section 5 a thorough experimental evaluation of the two mechanisms under the Relative services principles in conducted. This paper concludes with implementation issues, our future work and conclusions.

2. Service definition

Relative service (see also [18]) is an AF-based or AF PHB compliant service to which the proposed mechanisms of TWAM and DWRED are applicable.

Each individual customer agrees on a separate instance of Relative service provisioning with the backbone network provider for each ingress to egress router pair between which Relative customer traffic is served. Moreover, the DiffServ-enabled backbone itself adopts traffic engineering to pin the entire path for the set of customers' AF aggregates between a certain ingress and egress router pair. All AF traffic flows from different customers that enter the backbone network in the same ingress (edge) router and exit the backbone network from the same egress router can be assigned to the same MPLS Labeled Switched Path (LSP) and thus comprise a Forward Equivalence class (FEC) according to the MPLS terminology.

For the purposes of the proposed Relative service, we define such a FEC comprised of AF flows as an AF FEC. In a domain providing the Relative service, the maximum number of AF FEC LSPs that have to be set up is equal to all the different combinations of two edge routers, and thus the AF FEC LSP notion preserves scalability as it remains independent of the number of transit flows. By using a destination-aware model and MPLS traffic engineering, unfairness due to unequal RTTs is no longer a concern in our provisioning model.

Within an AF FEC, one AF profile with a guaranteed average service rate is defined for each participating customer. This profile is used by a marker at the AF FEC LSP ingress router to distinguish between AF packets of a customer's flow that fall within and out of the profile. In the core of the AF-enabled network, a dedicated queue Q_{AF}^{j} is configured at each core router along the AF FEC LSP j for serving the aggregated AF FEC traffic with a minimum guaranteed service rate equal to C_{AF}^{j}, despite of the load or congestion of the network.

Apart from the provisioned minimum C_{AF}^{j} service rate for AF packets at the AF FEC LSP core routers, the AF PHB includes also a buffer management or AQM mechanism to manage the size of Q_{AF}^{j}. From this service definition it is evident that the amount of state at each core router depends on the number of AF FEC LSPs served through it and not the number of flows, which makes it coherent with the DiffServ framework.

3. A TCP Window-Aware Marker

In this section we present a TCP Window-Aware Marker, for use at the ingress router (IR) of an AF FEC LSP (LSP_{j}^{AF}). Each TCP flow i of LSP_{j}^{AF} has to be served with a contracted average rate of r_{i}^{j}.

The TWAM operates on each flow or customer aggregate separately at IR. It is self-tuned in intervals of $\alpha * RTT_{LSP_j}$ seconds (with $a \in \{1,2,3,...\}$) for the specific LSP_{j}^{AF}. The α parameter determines how often in consecutive RTT_{LSP_j} intervals does the proposed marker update its configuration parameters for each of the TCP flows.

We denote by $W_{i,j}^{cwnd}$ the current congestion window of each flow in LSP_{j}^{AF}. Using the assumptions and definitions already provided, the following proposition can be formed for the proposed TWAM:

Proposition : *The TWAM distributes the resources available for the total traffic of an AF FEC among the individual AF FEC flows in a fair manner, analogous to the reserved r_i of each such flow.*

We define the bandwidth delay product for the aggregated AF traffic of LSP_{j}^{AF} or the LSP_{j}^{AF} congestion window (thus the number of segments that can be sent over LSP_{j}^{AF} in the upcoming RTT interval), as

$$W_{LSP_j}^{cwnd} = R_{AF}^{j} \times RTT_{LSP_j} \qquad (1)$$

where R_{AF}^{j} is the current aggregate service rate perceived by all of the flows participating in LSP_{j}^{AF}.

For the operation of TWAM, in order to fairly mark packets, and since it is not straightforward to estimate the exact value of R_{AF}^{j} and thus of $W_{LSP_j}^{cwnd}$ for each RTT interval, we make the following approximation

$$W_{LSP_j}^{cwnd} \approx \sum_{i} W_{i,j}^{cwnd} \qquad (2)$$

(2) approximates the value of $W_{LSP_j}^{cwnd}$, which is actually an estimator of the current conditions on the network and more specifically of the resources available for the traffic of LSP_{j}^{AF}.

The aim of the TWAM is to mark the $W_{i,j}^{cwnd}$ packets of each TCP flow i arriving within an interval RTT_{LSP_j} in proportion to r_i^j. Similar approaches for achieving fairness among TCP flows, such as the TSW2CM of [7] have so far only relied on rate approximations per TCP flow, trying to mark packets so that each TCP source would adjust the sending rate to a value proportional to its contracted r_i. However, real-time rate estimation is error prone, estimation parameters differ under different scenarios and the various methodologies are susceptible to deviations under transient network load. Moreover, this approach imposes rate estimation at each packet arrival. Using the TCP congestion window and a time granularity of one or a few RTTs according to our proposal, provides a more accurate and efficient estimation improving the achieved fairness.

Once $W_{LSP_j}^{cwnd}$ is estimated, TWMA marks arriving packets of flow i as In with probability p_i and as Out with probability $1 - p_i$ throughout the current RTT interval, where

$$p_i = \begin{cases} \dfrac{r_i^j * \sum_i W_i^{cwnd}}{\sum_i r_i^j * W_i^{cwnd}}, if\ (r_i^j * \sum_i W_i^{cwnd}) < (\sum_i r_i^j * W_i^{cwnd}) \\ 1 \quad ,\ otherwise \end{cases} \qquad (3)$$

In this way the available LSP_{j}^{AF} congestion window (see (1) and (2)) is 'coloured' in a way that reflects the individual TCP flows' reservations, rather than the current achieved congestion windows by each flow individually.

One of the most valuable achievements of TWAM is that it succeeds in driving TCP flows to continuously operate in Congestion Avoidance at different levels under both congested and un-congested periods and eliminates occurrences of the Slow Start phase.

4. Dynamic WRED

WRED is an AQM mechanism mainly used in core backbone routers. In WRED, packet drop probability for each packet arriving to the router queue is calculated according to a single value of average queue length Q_{avg}, but the values of $\min_{th}$, $\max_{th}$, $\max_{p}$ differ according to the colour of each packet. In related research work RED or WRED parameters are fixed during the operation of a router queue. As we will also experimentally verify, this can lead to inefficient utilization and unfair differentiation of resources.

We propose the use of a dynamically re-configurable WRED AQM mechanism to be used by the Relative service. Dynamic configuration of the WRED parameters ($\min_{th}$, $\max_{th}$ and $\max_{p}$) in the AF queues of the LSP_{j}^{AF} routers takes place (if necessary) at intervals equal or multiple of RTT_{LSP_j}, thus in accordance with the operation intervals of the TWAM. The idea here is to 'co-ordinate' the AQM with the TWAM operation. The proposed WRED configuration mechanism responds to fluctuations in available resources, allowing the use of excessive resources whenever they are available, in a way that achieves a bounded average queuing delay for packets in each Q_{AF} along LSP_{j}^{AF}.

In our previous work ([18]), we have demonstrated how the adaptation of the $\min_{th}, \max_{th}$ parameters of WRED according to the bandwidth-delay ($BW * D$) product of a link, achieves a bounded average queue size of approximately $\{0,2...0,25\} * (BW * D)$, while at the same time ensures a high utilization of available resources by TCP traffic. In this work, we use the TWAM operation principles to guide the DWRED parameters configuration. We propose that for each operating interval RTT_{LSP_j}, the $\max_{th}$ parameter of WRED is set equal to $W_{LSP_j}^{cwnd}$, derived from (2).

$$\max_{th} = \sum_i W_i^{cwnd} = W_{LSP_j}^{cwnd} \quad (4)$$

In a static WRED configuration $\min_{th}$ and $\max_{th}$ have fixed values. If these values are set too low, this results in too many packet drops which do not allow

TCP flows to open up their congestion windows when excess bandwidth resources are available. On the other hand, if they are set too high, fewer Out packets fall within the $\min_{th}$, $\max_{th}$ region and thus dropping is not as effective in providing congestion notification back to the sources.

Under the framework of DWRED, the use of values

$$\min_{th} = b * \max_{th},\ \max_{p} = 0.1 \quad (5)$$

is also proposed, with $b = 0.15$, thus $\min_{th}$ is proposed to also be updated for each operating interval in order to 'follow' the fluctuations of $\max_{th}$.

5. Experimental approach

We have conducted a series of experiments, comparing the performance of our approach against well-known mechanisms for AF-based services provisioning. For our experimentation, we have used the ns-2 simulator ([19]).

5.2. Basic TWAM and DWRED evaluation

A three-hop AF FEC LSP is used for serving Relative traffic entering a DiffServ-enabled domain in router R_1 and exiting from router R_4. For this initial set of experiments (**scenario #1**), the backbone links have a capacity of 30 Mbps and a dedicated router queue (Q_{AF}) is configured to serve Relative traffic at each router with a rate equal to 30Mbps.

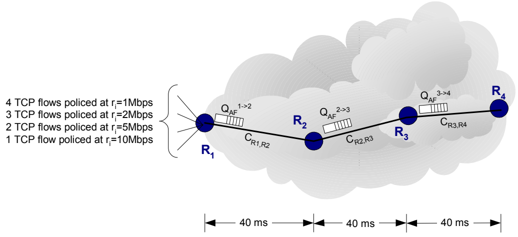
4 TCP flows policed at r_i=1Mbps
3 TCP flows policed at r_i=2Mbps
2 TCP flows policed at r_i=5Mbps
1 TCP flow policed at r_i=10Mbps
$Q_{AF}^{1\text{->}2}$
$Q_{AF}^{2\text{->}3}$
$Q_{AF}^{3\text{->}4}$
R_1
R_2
R_3
R_4
$C_{R1,R2}$
$C_{R2,R3}$
$C_{R3,R4}$
40 ms
40 ms
40 ms

Figure 1. The experimental set-up of scenario #1

Two different configurations are compared:

(i) Use of the TSW2CM marker in R_1 to examine upon each packet arrival the current rate of the TCP flow (marking packets as Out when the flow r_i is exceeded) and use of static WRED with {$\min_{th}$=40, $\max_{th}$=160, $\min_p$=0.1} in each of the Q_{AF} s depicted

(ii) Use of the TWAM marker in R_1 to calculate the probability of marking incoming packets of a flow as In (for intervals according to the RTT measured at the specific LSP) and use of DWRED in each of the Q_{AF} s depicted.

As can be seen from Figure 2, the TWAM+DWRED combination achieves better differentiation at the level of average congestion windows, while at the same time allowing all TCP flows to achieve larger average CWND sizes than those achieved by TSW2CM+static WRED. At the same time, no TCP flow experiences a Slow Start phase when TWAM+DWRED is used and CWND values for TCP flows are 'stabilized' within intervals proportional to each flow's reserved r_i.

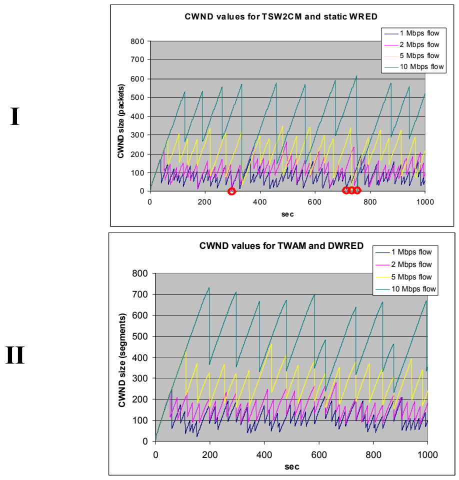

Figure 2. CWND measured for scenario #1

In Figure 3.I, it is shown how the proposed combination of TWAM+DWRED achieves average CWND values that better approximate the theoretical ones. Besides better utilization, TWAM+DWRED achieve fairness by ensuring differentiation among flows that the Relative service requires.

We have also used over-provisioning for the set-up of Figure 1 with $C_{R1,R2} = C_{R2,R3} = C_{R3,R4} = 50Mbps$ and a configuration of Q_{AF} to serve Relative traffic with a rate of 50Mbps, which is higher than $\sum_i r_i$ (**scenario #2**). Over-provisioning has ensured that all Relative flows achieve the contracted capacity under TWAM and DWRED (Figure 3.II). The limitations of TSW2CM and static WRED are once again apparent here since, despite of the over-provisioning, the larger r_i flows fail to reach their contracted r_i.

5.3. Evaluation under realistic operating conditions

Here, two different scenarios are presented:

(i) In **scenario #3,** the setup of Figure 1 is used with $C_{R1,R2} = C_{R2,R3} = C_{R3,R4} = 50Mbps$, $C_{AF} = 50Mbps$ in all Relative traffic queues. However, one flow from each group with a common r_i (Figure 1) starts transmitting 200 seconds after the simulation starts. The purpose here is to compare the efficiency of TWAM and DWRED against that of TSW2CM and static WRED in fluctuations of Relative traffic.

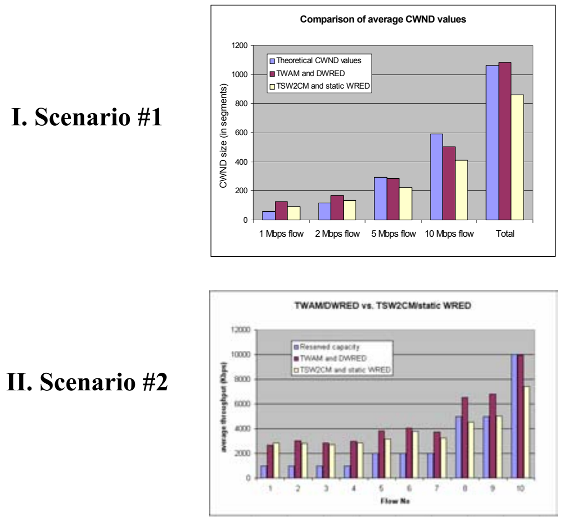

Figure 3. Average CWND and throughput of TCP flows

(ii) In **scenario #4,** the setup of Figure 1 is used with $C_{R1,R2} = C_{R2,R3} = C_{R3,R4} = 50Mbps$. However, in router R$_2$ a separate queue Q_{EF} serving traffic with absolute non-pre-emptive priority and a rate up to 10 Mbps is configured. During the simulation duration, UDP traffic at a rate of 11Mbps (to ensure that Q_{EF} reaches its maximum serving capacity) is injected in R$_2$ and exits the AF FEC LSP from R$_3$ at regular intervals.

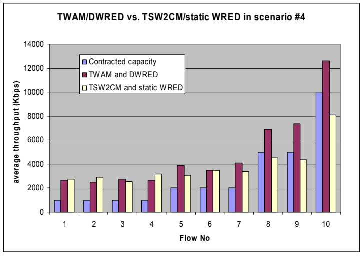

Figure 4. Average throughput of TCP flows in scenario #3

Figure 4 shows the average throughput of all TCP flows measured in the interval from $t_1 = 400$ to $t_2 = 1000$ seconds for scenario #3. Use of TWAM and DWRED ensures that the differentiation between TCP flows is preserved in a fair way. When TSW2CM and static WRED are used Flows No 8, 9 and 10 fail to obtain their contracted rate despite of the over-provisioning conditions. Moreover, differentiation among TCP flows is not fair.

In Figure 5.I, the performance advantages of TWAM and DWRED become clearly evident in the framework of scenario #4. High-priority UDP traffic is only transmitted in specific intervals. Despite the periodic load of 110% in the high-priority class, TCP flows succeed in obtaining and preserving a throughput proportional to their reserved capacity.

On the contrary, the use of TSW2CM and static WRED (Figure 5.II) demonstrates a poor performance. Despite the minimum of 40 Mbps service rate available to Relative traffic along the AF FEC LSP the TCP flows are unable to preserve their contracted capacity. Differentiation among them is also very poor.

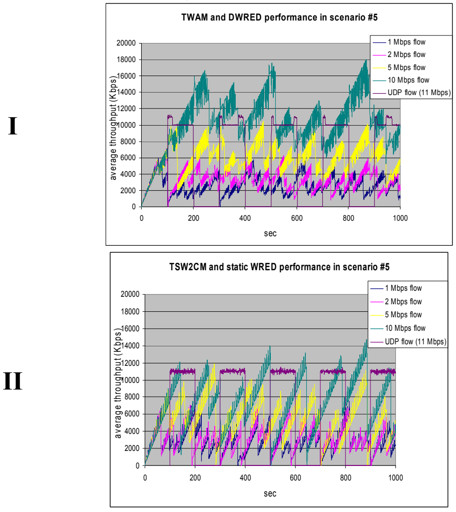

Figure 5. Instantaneous throughput achieved by TCP flows under scenario #4

6. Implementation issues

In the case of TWAM and DWRED a set of computations is required. For each operating interval, thus every $\alpha * RTT_{LSP_j}$ seconds, the TWAM in the ingress router computes (3). Similarly, every $\alpha * RTT_{LSP_j}$ seconds, in each core router, the DWRED parameters for Q_{AF} are calculated according to (4) and (5)

The computational complexity of TWAM is much lower than marking schemes like that of TSW2CM requiring a set of computations to be made upon each packet arrival at the ingress router. DWRED imposes some overhead when compared to static WRED configurations, however this overhead only depends on the update intervals chosen and scales with the number of TCP flows, the topology etc.

By using AF FEC LSPs for TCP traffic and the corresponding acknowledgments between two network edges, we can ensure that all TCP flows belonging to the same LSP_j^{AF} experience the same RTT and thus update their CWND values in the same intervals. Our proposed TWAM and DWRED mechanisms do not allow serving AF LSPs associated with different RTT values through a single router queue of a core router as this can lead to unfairness in throughput distribution. This unfairness can be avoided if a suitable queuing scheme, such as WFQ, is used for AF traffic in order to guarantee a minimum service rate C_{AF}^j for each LSP_j^{AF}. Maintaining per flow or TCP aggregate information such as r_i^j and $W_{i,j}^{cwnd}$ values for the calculation of p_i at the ingress routers where the TWAM operates is completely in accordance with the DiffServ principles. $\sum_i W_{i,j}^{cwnd}$ values can be signaled to the core routers per AF LSP in every operating interval, so that core routers maintain state only per AF FEC LSP and not per flow and thus obtaining scalability and remaining DiffServ obedient.

7. Future work-Conclusions

In this paper we have proposed and evaluated two mechanisms for the AF service PHB implementation that can be used in the framework of the Relative service model, as this was specified in our previous work. The TWAM ensures the distribution of available resources among TCP flows within an AF class in a fair manner, according to their contracted average service rates. Its efficiency is reinforced by the DWRED and in particular by the self-tuning capability of DWRED, since the WRED parameters are dynamically adjusted to the current perceived load at the TWAM. Both mechanisms require much less overhead than existing equivalent mechanisms and achieve improved differentiation, fairness, adaptation to transient network conditions and high utilization of available resources. TCP flows under the TWAM and DWRED operation demonstrate a controlled behavior, with smooth adaptation to network conditions.

As part of our future work, we intend to investigate refinements of the TWAM principles of operation, means for obtaining better differentiation and alternative implementations of DWRED.

8. References

[1] J. Heinanen et al., 'Assured Forwarding PHB Group', RFC 2597, 1999

[2] O. Medina et al., 'Bandwidth sharing under the Assured Forwarding PHB', Internal Report, IRISA No.1478, 2002

[3] M. Goyal et al., 'Performance Analysis of Assured Forwarding', Internet Draft, IETF, 2000

[4] B. Nandy et al. 'Diffserv's Assured Forwarding PHB: What Assurance does the Customer Have?', in Proc. of NOSSDAV'99, USA, 1999

[5] N. Seddigh et al., 'Study of TCP and UDP Interaction for AF PHB', Internet draft IETF, 1999

[6] M. May et al., 'Simple performance models of tagging schemes for service differentiation in the Internet', in Proc. of IEEE INFOCOM 1999, USA, 1999, pp. 1385-94

[7] D. Clark and W. Feng. 'Explicit allocation of best effort packet delivery service', IEEE/ACM Transactions on Networking, 6(4), 1998, pp. 362-374

[8] D. Lin and R. Morris, 'Dynamics of random early detection', in Proc. of ACM SIGCOMM '97, 1997, pp. 127-137

[9] W. Feng et al., 'A Self-Configuring RED Gateway', In Proc. of IEEE INFOCOM 1999, USA, pp. 1320-1328

[10] R. Kapoor et al., 'Core-Stateless Fair Bandwidth Allocation for TCP flows', in Proc. of IEEE ICC2001, Finland, 2001, pp. 146-150

[11] F. Azeem et al., 'TCP-Friendly traffic conditioners for differentiated services', Internet Draft, IETF, 1999

[12] S. Athuraliya et al., 'REM: Active queue management', IEEE Network, vol. 15, 2001, pp. 48-53

[13] C. V. Hollot et al., 'A control theoretic analysis of RED', in Proc. of IEEE INFOCOM 2001, 2001, USA, pp. 1510-1519

[14] M. Mellia et al., 'TCP-aware packet marking in networks with DiffServ support', International Journal of Computer Telecommunications Networking, Vol. 42, Issue 1, 2003, pp. 81-100

[15] G. Iannaccone et al., 'Aggregate Traffic Performance with Active Queue Management and Drop from Tail', Computer Communication Review, Vol. 31(3), 2001, pp. 4-13

[16] W. Feng et al., 'Understanding TCP Dynamics in a Differentiated Services Internet', IEEE/ACM ToN, Vol. 7, 1999, pp. 173-187

[17] F. Azeem et al. 'TCP-friendly traffic marker for IP differentiated services', in Proc. IWQoS'2000, USA, 2000, pp. 35-48

[18] C. Bouras and A. Sevasti, 'Performance Analysis for a DiffServ-enabled network: The case of Relative Service', in Proc. of NCA' 03, 2003, USA, pp. 381-388

[19] S. McCanne and S. Floyd, "ns Network Simulator", available at: http://www.isi.edu/nsnam/ns