Bandwidth Assurance Issues for TCP flows in a Differentiated Services Network *

Nabil Seddigh, Biswajit Nandy, Peter Pieda
Computing Technology Labs, Nortel Networks
Ottawa, Canada
{nseddigh, bnandy, ppieda}@nortelnetworks.com

Abstract

Much industry attention has recently been focused on providing differentiated levels of service to users on IP networks. One such proposal is the RIO scheme proposed by Clark [4]. RIO is an extension of the RED algorithm that relies on a differentiated drop treatment during congestion to cause different levels of service. The end result of differentiated dropping of packets during congestion is differentiated throughput rates for end-users. The IETF's Diffserv Working Group is currently discussing a similar differentiated drop mechanism called the AF (Assured Forwarding) PHB (Per Hop Behaviour).

This paper raises issues with providing bandwidth assurance for TCP flows in a RIO-enabled Differentiated Services network. The main contribution is a detailed experimental study of five different factors that impact throughput assurances for TCP and UDP flows in such a network. Our study demonstrates that these factors can cause different throughput rates for end-users in spite of having contracted identical service agreements.

1. INTRODUCTION

In the traditional IP network model, all user packets compete equally for network resources. The rise in usage and popularity of the Internet coupled with new applications such as voice, video and www has fuelled research to improve the Quality of Service delivered by today's best-effort networks. The underlying concept in IP Quality of Service (IP-QoS) is the ability of network operators to offer differing levels of treatment to user traffic based on their requirements.

The Differentiated Services (Diffserv) [3] approach proposes a scalable means to deliver IP QoS based on handling of traffic aggregates. It operates on the premise that complicated functionality should be moved toward the edge of the network with very simple functionality at the core. Edge devices in this architecture are responsible for ensuring that individual user traffic conforms to traffic profiles specified by the network operator and for grouping flows in an aggregated fashion into a small number of classes. Core devices perform differentiated aggregate treatment of these classes based on the marking performed by the edge devices.

RIO-based [4][5] schemes have been proposed as a simple means of providing Differentiated Services. The basis of the RIO mechanism is RED-based [2] differentiated dropping of packets during congestion at the router. In RIO, traffic profiles for end-users are maintained at the edge of the network. When user traffic exceeds the contracted target rate, their packets are marked out-of-profile. Otherwise, packets are marked in-profile. The RIO scheme utilizes a single queue. All user packets are directed to and serviced from the same queue.

Two sets of RED thresholds are maintained, one each for in-profile and out-of-profile. Two separate average buffer occupancy calculations are tracked, one for in packets and one for in and out packets. The possibility of dropping in packets depends only on the buffer occupancy of in packets while the possibility of dropping out packets depends on the buffer occupancy of in plus out packets. This scheme gives the appearance of two coupled virtual queues within a physical queue.

The RIO scheme is particularly appealing because it uses a single FIFO queue and relies only on a remarking policer at the edge of the network. As a result, it promises to deliver packet differentiation based on incremental upgrades to best-effort routing devices. Further, it may enable service offerings at a lesser cost than services based on mechanisms such as the Expedited Forwarding (EF) PHB [8]. RIO-like schemes allow service providers to offer one-to-anywhere services. It should facilitate significant statistical multiplexing gains in terms of network resource usage. The end result should translate into a lower cost per service offering with an acceptable QoS for the end customer.

Though a RIO-based scheme is compelling, there are some open issues that need to be understood. The most important question relates to the kind of end-to-end service that can be realized by the end-user. Since RIO does not focus on delay, it is probable that end-users will use throughput assurance as a measure of good or bad network performance. There is concern that a RIO-like scheme should show some measure of predictability for service providers who use it to create a service for their paying customers.

* To be submitted to Globecom 99

In this paper, we study five different factors to understand their impact on offering predictable bandwidth assurance services to customers. We focus primarily on TCP flows because they are most affected by the five factors and constitute majority of the Internet traffic today. The studies examine cases for under-provisioned and over-provisioned networks. Our work is motivated by concern for fairness among customers who pay equal amounts for access to the network as well as lower paying customers with Best Effort access. The study was carried out using VxWorks-based prototypes developed at the Computing Technology Lab, Nortel Networks.

The organization of the paper is as follows: Section 2 reviews related work in this area. Section 3 provides the experimental background. In sections 4 and 5, we describe the experiments, present results, and perform analysis. Finally, we point to future work based on our analysis.

2. RELATED WORK

There have been a number of recent simulation studies that focused on a RED-based Differentiated Services scheme.

Clark and Fang in [5], reported one of the early simulation studies on a RIO-like scheme. The paper introduced RIO (RED with In/Out) and a remarking policer that utilized an average sliding window rate estimator and intelligent marker. The main contribution of that work was to show that source target rates could be assured in a simple capacity allocated network that relies on statistical multiplexing. The paper showed that the in-profile portion of TCP flows are protected from issues such as RTT (Round-Trip Time) and non-responsive UDP flows.

Feng et al [9][11] examine the use of adaptive priority marking similar to that of the profile meter in the RIO scheme. They present results showing that the compliant part of the TCP flow throughput is largely independent of the RTT of the individual flows. However, the non-compliant part and best-effort flows see their throughput affected by the RTT. Finally, the authors report that in an over-provisioned network, target rates are ignored and all RIO TCP flows have an equal share of the Assured Bandwidth.

Ibanez and Nichols [6], via simulation studies, confirm that RTT is a key factor in the throughput of flows that obtain an Assured Service using a RIO-like scheme. Their conclusions are that such an Assured Service cannot provide “clearly defined and consistent rate guarantees”.

Using simulation, Kim and Thomson [9] study different issues associated with active queue management techniques for providing differentiated levels of service. The authors indicate that bandwidth service is immune from RTT variation. Further, they report that there is no differentiation between RIO and Best Effort flows in an under-engineered network. In contrast to Ibanez and Nichol’s findings, they conclude that “bandwidth assurance service using RED yields predictable network behaviour under all conceivable realistic conditions.”

More recently, Ikjun [12] has shown that excess network bandwidth is not distributed proportional to target rates. The paper proposes and evaluates new schemes to address the issues discovered with TCP flows in a RIO-capable network.

The IETF Diffserv working group [3] is evaluating a proposal to define a router PHB (per-hop-behaviour) that uses RED-like differentiation of packets as its basis. AF derives its operating principles on RIO but attempts to provide greater flexibility by having priority classes and another level of drop preference. The AF PHB draft [7] proposes 4 classes of service with 3 drop preferences per class. To date, there has been no work justifying the choice of 4 classes and 3 drop preferences. It is unclear as to why a smaller number of classes and drop preferences would not have sufficed.

3. EXPERIMENTAL BACKGROUND

The studies were performed using an experimental testbed that included networking elements with edge and core device functionality as specified in [5]. The devices, running on a Pentium platform with VxWorks as the RTOS (Real Time Operating System), were designed, developed and implemented at the CTL in 1998 [13].

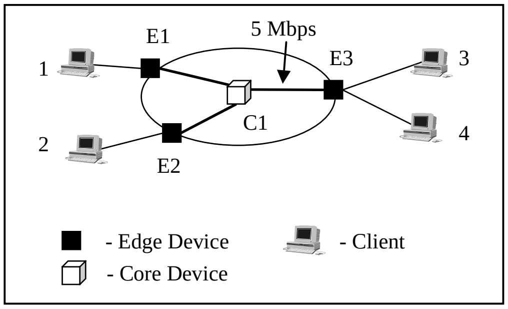

Figure 1: Experimental Testbed

The experimental testbed can be seen in Figure 1. The setup consisted of four network elements: E1, E2, E3 and C1. The end-systems consisted of Pentium PCs running Linux 2.0.34 as the operating systems. The Netperf [14] tool was used as the TCP traffic generator. The TCP flows generated were all long lasting. The network topology consisted of a number of end-hosts with 10-BaseT connections interconnected through a bottleneck link of 5Mpbs. A Link Delay emulator device was developed and installed in the network to assist with the study on RTT.

In carrying out the experiments, efforts were made to eliminate TCP/IP stack implementation, Ethernet cards, and receiver window size as contributing factors to the results though we realize that they do play a role in the observed service.

The Edge devices in the testbed classify, police and mark packets based on source/destination IP address. The traffic conditioning scheme is a remarking policer that utilizes the ARE (Average Rate Estimator) [5] window scheme proposed as part of RIO. Best Effort traffic shares the same virtual queue as the out-of-profile packets.

Careful consideration needs to be given to the setting of RED parameters for the two different virtual queues. [13] reports results of studies with varied RED parameter settings for a RIO-like scheme. Most of our experiments unless otherwise specified utilize the RED parameter settings in Table 1.

	In-profile	**Out-of-profile/Best-Effort**
Min_{th}	20 pkts	10 pkts
Max_{th}	40 pkts	20 pkts
Max_p	0.02	0.1
w_q	0.002	0.002

Table 1: Red Parameter Settings

4. EXPERIMENTAL DESCRIPTION AND RESULTS

The experiments focus on five factors to determine how they affect end-to-end bandwidth service for TCP flows in over-provisioned and under-provisioned network scenarios. These factors include RTT, number of micro-flows in a target aggregate, size of the target rate, packet size and existence of non-responsive flows. We study the under-provisioned case because we expect that service provider tendency to maximize profit and minimize over-provisioning will cause hotspots in the network.

4.1 Experiment 1: Round Trip Time

The goal of the first experiment is to study the effect of RTT on throughput. This is repeated for both over and under-provisioned scenarios. A link delay emulator is used to emulate varied link transmission delay.

In the over-provisioned scenario, there are four sets of traffic. One set of flows between Clients 1 and 3 has a target rate of 1Mbps while there is another set of best effort flows without a target rate. The same is true for Clients 2 and 4. The RTT for Client 1-3 is fixed at 20ms while the RTT for Client 2-4 is varied from 40ms to 160ms. The results can be seen in Figure 2.

In this scenario, the low RTT flows with traffic profile of 1Mbps, clearly achieve their target rates and consume a large portion of the excess bandwidth. The flows with larger RTT achieve their targets but get a lesser share of the excess bandwidth as the RTT ratio increases. Best Effort flows with low RTT get improved bandwidth as the RTT of clients 2-4 is increased.

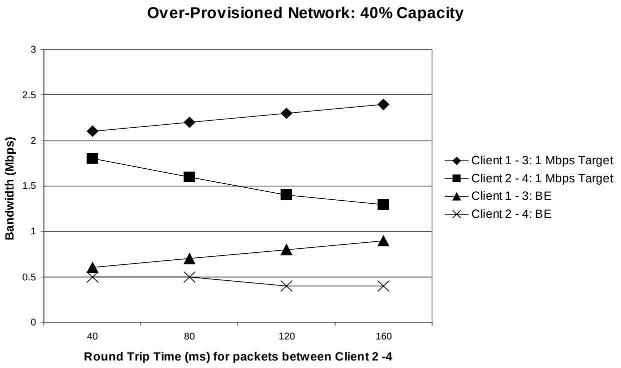

Figure 2: RTT: Over-provisioned

The above testcase is repeated in the over-provisioned scenario. The only change in this case is that the 2 sets of target rates are changed from 1Mbps each to 3Mbps each, which results in a cumulative total larger than the 5Mbps bottleneck link BW.

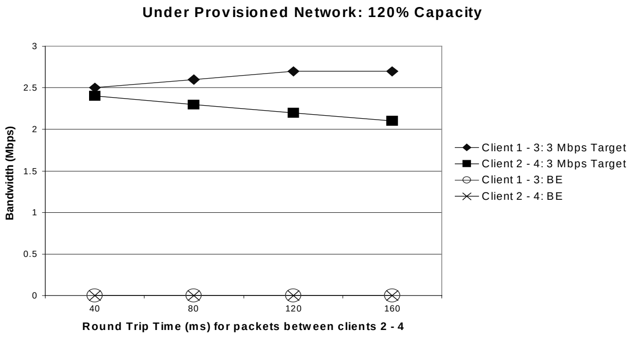

Figure 3: RTT: Under-Provisioned

The results of the under-provisioned case are presented in Figure 3. In this case, neither the low RTT nor the high RTT flows achieve their target rates. However, the size of the RTT clearly affects how close the set of flows get to their respective targets. This difference increases as the ratio of RTTs increases.The best effort flows do not get any bandwidth. At this point, most out and Best-Effort packets are being dropped.

4.2 Experiment 2: Impact of Number of microflows

The goal of the second experiment is to study the effects of aggregation. In a Differentiated Services Internet, service contracts will not necessarily be on a per end-user basis. One common scenario will be the case where a company contracts a target rate with a service provider. Source flows originating from the company would then compete for the aggregate target rate.

This experiment investigates what happens when the number of microflows competing for a particular target rate are different. I.e. Do the throughput levels for a target aggregate differ if we have 2 microflows versus 8 microflows striving to share the policy target rate. This is observed for both over and under-provisioned cases.

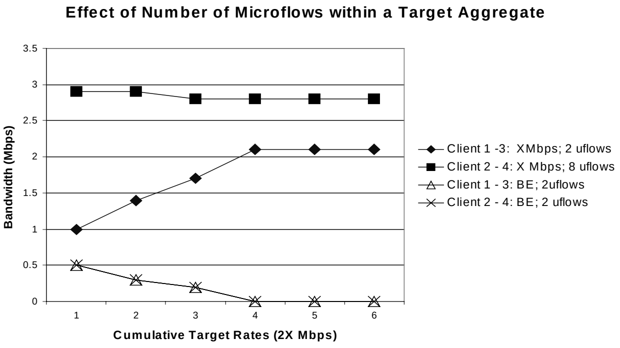

Figure 4: Number of microflows

In this scenario, there are again four sets of flows. Between Clients 1 and 3, there are two microflows with target rates and two best effort flows. Between Clients 2 and 4, there are eight microflows with target rates and two best effort flows. The cumulative target rate is varied from 1 to 6Mbps (both source-destination pairs have equal target rates), thus emulating a range of provisioning levels on the network.

The results of this experiment are captured in Figure 4. In this Figure, we see that the aggregate containing 8 microflows consistently outperforms the aggregate that contains 2 microflows. In the over-provisioned scenario, the difference is quite extreme as the 8-microflow set simply has more flows to compete for the excess bandwidth. There is less of a difference in the capacity allocated and under-provisioned cases but it is noticeable. More studies are needed in this area, as it will have tremendous implications on a large Internet. Some organizations will have thousands of flows in a target aggregate while others will have just hundreds of flows.

4.3 Experiment 3: Size of Target Rate

The goal of the third experiment is to study whether or not the size of the target rate for an AF flow has any bearing in the type of service received. In an over-provisioned network, we study how the excess bandwidth is distributed. In an under-provisioned network, paying customers will naturally assume that they will achieve the same ratio of their target rate as all other customers paying in the same class.

In this scenario, there are again four sets of flows. Between Clients 1 and 3, there are a set of flows with a target rate of 1Mbps and an equivalent set of best effort flows. Between Clients 2 and 4, there are a set of flows with varying target rates and an equivalent set of best effort flows.

The results of this experiment are in Figure 5. In the over-provisioned scenario, both targets are achieved. However, the excess bandwidth is not proportionally distributed as would be desired by a customer. Instead, we see an almost even distribution of the excess bandwidth between the four sets of competing flows. In the under-provisioned scenarios, we see that neither target is achieved. However, there is fair degradation of service for both sets of target rates.

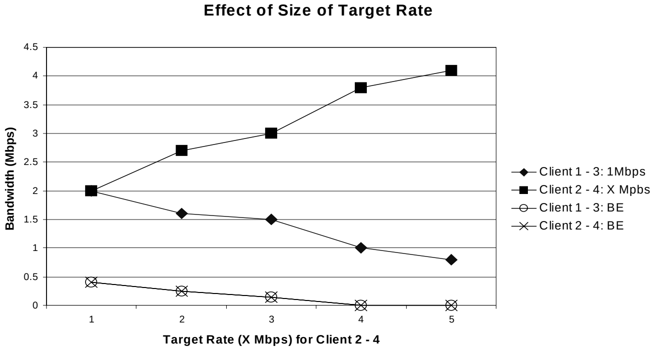

Figure 5: Target Size

4.4 Experiment 4: Packet Size

This experiment studies the impact of packet size on the type of throughput observed. A recent study of an Internet backbone [15] indicates that 11% of packets are of size between 552-576bytes and 10% of the packets are of size 1500 bytes. Here, we present results for the over-provisioned scenario.

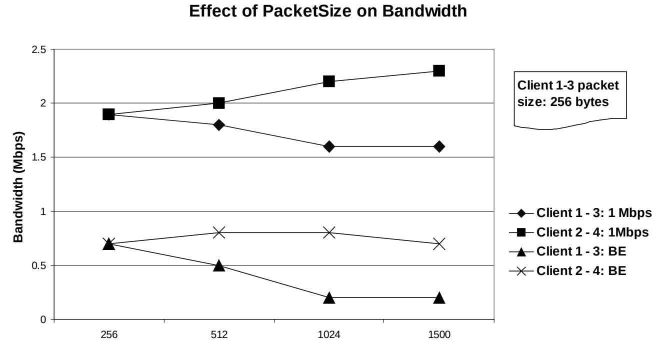

Figure 6: Packet Size: Over-provisioned Scenario

In this scenario, there are again four sets of flows. Between Clients 1 and 3, there is a set of flows with a target rate of 1Mbps and an equivalent set of best effort flows. The packet size for this set of flows is fixed at 256 bytes. Between Clients 2 and 4, there are a set of flows with target rates of 1Mbps and an equivalent set of best effort flows. The packet sizes for this set of flows is varied from 256 bytes to 1500 bytes.

Figure 6 depicts the results of the test. In this Figure, we see that both targets are achieved. However, as the difference in packet sizes increases, there is a significant unfairness in terms of the amount of excess bandwidth that is seized. Repeating the testcase for the under-provisioned network yields similar results.

4.5 Experiment 5: Non-Responsive flows

This experiment studies the impact of non-responsive UDP flows on the throughput attained by the TCP flows. We are interested in two cases. In the first case, we study the interaction between UDP best effort flows and TCP flows whose target rate is varied between 1 and 6Mpbs, thus studying a range of congestion conditions on the network. The second scenario studies the case where 1Mbps target rate is set for UDP traffic and the TCP flow target rates are varied from 1 to 6 Mbps. In both cases, TCP Best Effort flows exist to compete for excess bandwidth where available.

Case 1: UDP Best Effort

In this scenario, there are 6 sets of aggregate flows. Between Clients 1 and 3, there is a 0.5 Mbps UDP best effort flow and some TCP best effort flows. There are also an equal number of TCP flows with target rates. The same traffic mix exists between Clients 2 and 4. The total target rate for TCP flows is varied between 1Mbps and 6Mbps and the bandwidth observed.

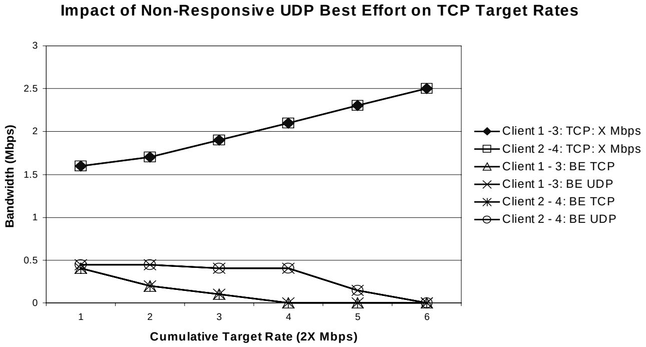

Figure 7: UDP Best Effort vs TCP with Target

The results of this experiment are depicted in Figure 7. The Figure shows that when the cumulative target rates for the RIO flows is increased, they take as much of the bandwidth as they need to achieve their target and compete for the excess bandwidth where it is available. In the over-provisioned scenario, the UDP flow obtains its 0.5Mbps. However, in the under-provisioned case, the UDP flow is starved and the TCP flows compete to achieve as much of their target as possible. One clear conclusion from this testcase is that the UDP best effort flow does not hinder TCP flows from achieving their target rates.

Case 2: UDP with Target Rate of 1Mpbs

In this scenario, there are again four sets of flows. Between Clients 2 and 4, there is a 1Mbps UDP flow with a target rate of 1Mbps and some TCP best effort flows. Between Clients 1 and 3, there are an equal number of TCP best effort flows and TCP flows with a target rate that is varied between 1 and 6Mbps.

The results of this experiment are captured in Figure 8. We observe that the UDP flow achieves its target rate of 1Mbps during over-provisioning. The TCP flows also achieve their targets when the network is over-provisioned. However, as the network approaches capacity-allocation and under-provisioned state, the TCP flows do not achieve their target and are unfairly degraded in favour of the UDP flow that is only minimally degraded. The best effort TCP flows compete for the excess bandwidth where available.

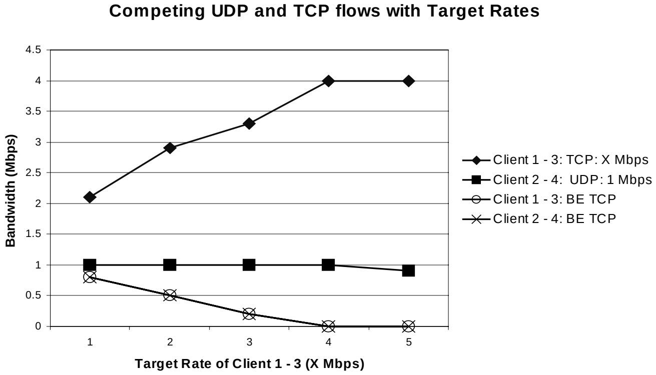

Figure 8: UDP vs TCP with Target Rates

5. Analysis and Discussion

Examining the results of the previous section yields the following observations:

1. All five factors in some way or another affect the throughput rates achieved by TCP flows of end users.
2. In an over-provisioned network, all target rates are achieved regardless of the five factors. However, the degree to which excess bandwidth is fairly distributed depends a great deal on the five factors. End-users who pay the same amounts for target rates will not be satisfied with obtaining unfair shares of the excess bandwidth.
3. As the network approaches an under-provisioned state, none of the target rates may be achieved. Since not all target rates can be achieved, it is hoped that there would be some form of fair degradation amongst the flows with target rates. However, this is not the case. Four of the five factors all play a role in biasing the degradation for or against particular flows.

The role that the RTT, target rate size and packet size play in determining throughput rates can be explained via the following equation captured by Mathis et al [16]:

$$BW < \frac{MSS}{RTT}\frac{1}{\sqrt{p}}$$

The equation shows the relationship of bandwidth to packet size, RTT and packet loss rate. Thus, uneven values for any of the above factors will cause uneven distribution of excess bandwidth in the over-provisioned network and non-fair degradation in the case of the under-provisioned network.

One solution this problem is to perform intelligent packet marking on out-of-profile packets. The marking should take into account the RTT, packet size and target rate in order to mitigate the effect of these factors. Such an approach is possible for flows that pass through a single edge device. However, it is problematic for flows that pass through two or more different edge devices since the marking depends on knowledge of the relativity of these factors between differing flows. The first edge device has no notion of the second device's information on RTT, packet size or target rate. One alternative is to consider some form of communication between edge devices. While this scheme may be scalable for a carrier network with hundred's of edge nodes, there are concerns it may not be scalable for a global network such as the Internet.

The issue of differing numbers of microflows in a target aggregate is also of concern as it has a number of implications. In a heteregenous network, there will be a variety of aggregates competing for network bandwidth. Some of the aggregates will be composed of small numbers of flows while others will consist of larger numbers of flows.

Non-responsive UDP flows are another source of concern. Due to their non-responsive nature, they are more likely to achieve their target rates than TCP flows. There are a number of possibilities to address this problem. One possibility is to perform intelligent marking. Another possibility that has been proposed is to give UDP its own drop preference. This imposes policy and classification requirements. Yet another possibility is to put UDP in a separate queue.

One other important factor that was observed during the experiments is that RED parameter settings play a key role in the type of results obtained. Further analysis and study is required to determine clear guidelines for RED parameter settings based on these five factors.

6. CONCLUSIONS

In this paper, we have used an experimental testbed to study the effects of five different factors on the throughput rates of TCP flows in a RIO-based Differentiated Services network. We present results showing that in an over-provisioned network, all target rates are achieved regardless of the five factors. However, the factors clearly cause uneven distribution of excess bandwidth amongst TCP flows. The results also indicate that as the network approaches an under-provisioned state, these factors play an important role in determining the extent to which aggregates of TCP flows achieve or don't achieve their target rates.

Partial solutions to alleviating the effect of the factors may lie in intelligent packet marking schemes. There are scalability and inter-device communication issues that need to be resolved if such marking schemes are implemented. These issues need to be resolved, as they are required for predictable throughput assurance. In the absence of such guarantees, it will be difficult to use a RIO-like scheme for anything other than a differentiated congestion control mechanism between domains. Further discussion on the value of RIO-like schemes can be found in [17].

7. ACKNOWLEDGEMENTS

We are grateful to Alan Chapman, Jamal Hadi Salim and Steve Jaworski for numerous technical discussions on Differentiated Services related issues.

8. REFERENCES

[1] Jacobson V., *"Congestion Avoidance and Control"*, In Proceedings of SIGCOMM '88, Stanford, CA, August 1988, ACM

[2] Floyd, S., and Jacobson, V., "*Random Early Detection gateways for Congestion Avoidance* ", IEEE/ACM Transactions on Networking, V.1 N.4, August 1993, p. 397-413.

[3] Blake, S. Et al, *"An Architecture for Differentiated Services"*, RFC 2475, December 1998

[4] Clark D., and Wroclawski J.,*"An Approach to Service Allocation in the Internet"*, Internet Draft, draft-clark-diff-svc-alloc-00.txt, July 1997

[5] Clark D. and Fang W., *"Explicit Allocation of Best Effort Packet Delivery Service"*, http://diffserv.lcs.mit.edu/exp-alloc-ddc-wf.ps, 1998

[6] Ibanez J, Nichols K., *"Preliminary Simulation Evaluation of an Assured Service"*, Internet Draft, draft-ibanez-diffserv-assured-eval-00.txt>, August 1998

[7] Heinanen J., Baker F., Weiss W., and Wroclawski J., *"Assured Forwarding PHB Group"*, Internet Draft, <draft-ietf-diffserv-af-05.txt>, February 1999.

[8] Jacobson V, Nichols K, Poduri K, *"An Expedited Forwarding PHB"*, Internet Draft, <draft-ietf-diffserv-phb-ef-00.txt>

[9] Kim H., and Thomson, S., *"Evaluation of Bandwidth Assurance Service using RED for Internet Service Differentiation*", Submitted for publication, August 1998.

[10] Feng, W, Kandlur, D, Saha, D, Shin, K, *"Understanding TCP dynamics in a Differentiated Services Internet"*,NOSSDAV '97, May 1997

[11] Feng, W, Kandlur D, Saha D, Shin K, *"Adaptive Packet Marking for Providing Differentiated Services in the Internet,"* ICNP '98, October 1998

[12] Yeom, I and Reddy N, *"Realizing throughput guarantees in a differentiated services network"*, http://dropzone.tamu.edu/~ikjun /papers.html, December 1998.

[13] Seddigh, N, Nandy B, Pieda P, Hadi Salim J, and Chapman A., *"An Experimental Study of Assured Services in a Diffserv IP QoS Network"*, SPIE symposium on QoS Issues Related to the Internet, Boston, November 1998.

[14] http://www.netperf.org/netperf/NetperfPage.html

[15] Thompson K, Miller G, Wilder R, *"Wide-Area Internet Traffic Patterns and Characteristics"*, IEEE Network Magazine, November/December 1997.

[16] Mathis M, Semske J, Mahdavi J, Ott J, *"The macroscopic behaviour of the TCP congestion avoidance algorithm."*, Computer Communication Review, 27(3), July 1997

[17] Nandy B, Seddigh N, Pieda P, "Diffserv Assured Forwarding PHB: What can we Assure?", Submitted to IwQoS '99, Feb. 99