Paper title: Evaluating Voice Traffic in a Differentiated Services Environment

Authors: Artur Ziviani — E-mail: `artur@gta.ufrj.br`
José Ferreira de Rezende — E-mail: `rezende@gta.ufrj.br`
Otto Carlos Muniz Bandeira Duarte — E-mail: `otto@gta.ufrj.br`

Contact Author: Artur Ziviani

Affiliation: Universidade Federal do Rio de Janeiro

Address: Grupo de Teleinformática e Automação – GTA
COPPE - Programa de Engenharia Elétrica
Universidade Federal do Rio de Janeiro
P. O. Box 68504 - 21945-970 - Rio de Janeiro - RJ - Brazil

Phone number: +55 (21) 260-5010 ext. 240

Fax number: +55 (21) 290-6626

Abstract

This paper evaluates the performance achieved by voice traffic when using Expedited Forwarding (EF) and Assured Forwarding (AF) schemes in a Differentiated Services (DS) environment. The mutual influence of the two schemes is evaluated when: (i) a controlled amount of voice bursts are allowed into a DS domain; (ii) in the presence of an increasing number of voice flows handled by EF; and (iii) for different best-effort packet sizes. Results show that allowing controlled EF traffic bursts into a DS domain does not significantly increase the delay and jitter of the AF traffic if idle periods from EF traffic do not become scarce. Results also show that different best-effort packet sizes exert a stronger influence on the delay of AF packets than on the delay of EF packets.

Event: General Conference

Topics: Voice Over IP
Multimedia Communications

EVALUATING VOICE TRAFFIC IN A DIFFERENTIATED SERVICES ENVIRONMENT

Artur Ziviani
artur@gta.ufrj.br

José F. de Rezende
rezende@gta.ufrj.br

Otto Carlos M. B. Duarte
otto@gta.ufrj.br

Grupo de Teleinformática e Automação (GTA)
COPPE/EE – Universidade Federal do Rio de Janeiro
P.O. Box 68504 – 21945-970 – Rio de Janeiro – RJ – Brazil

Abstract

This paper evaluates the performance achieved by voice traffic when using Expedited Forwarding (EF) and Assured Forwarding (AF) schemes in a Differentiated Services (DS) environment. The mutual influence of the two schemes is evaluated when: (i) a controlled amount of voice bursts are allowed into a DS domain; (ii) in the presence of an increasing number of voice flows handled by EF; and (iii) for different best-effort packet sizes. Results show that allowing controlled EF traffic bursts into a DS domain does not significantly increase the delay and jitter of the AF traffic if idle periods from EF traffic do not become scarce. Results also show that different best-effort packet sizes exert a stronger influence on the delay of AF packets than on the delay of EF packets.

1 Introduction

The best-effort model of the current Internet architecture does not meet the Quality of Service (QoS) requirements of the new multimedia applications. To address this issue, different approaches are currently being developed, such as adaptive applications [1, 2], QoS-based routing [3, 4], and the Differentiated Services (DS) architecture [5]. The DS proposal provides scalable discriminated services because there is no need of either one state per flow or signaling at every hop. The treatment offered by a DS-compliant node is applied to an aggregation of flows, not to individual flows. Individual flows are classified into aggregations of related flows that will be differentially served within a DS domain. The classification and packet marking take place at the DS domain boundary. Aggregations are forwarded at each node within a DS domain following a Per-Hop Behavior (PHB). The PHB is selected by a mapping between a codepoint that identifies the aggregation and the expected behavior [6]. The service offered to an aggregation within a DS domain results from the combination of all PHBs along its path therein.

This paper was supported by: CAPES, UFRJ, FAPERJ, FUJB, CNPq, REENGE, and COFECUB.

Currently, two PHBs are in discussion in the DS research community: the Expedited Forwarding (EF) PHB [7] and the Assured Forwarding (AF) PHB [8]. Although recent papers [9, 10, 11] analyze distinct aspects of differentiated services, the transmission of voice traffic over a DS domain remains not well explored. In [12], voice traffic is transmitted over the original DS structure, where the concept of PHB was not defined. In a previous work [13] we evaluate the performance of voice traffic forwarded by the EF PHB. Seddigh [14] evaluates the interaction of TCP and UDP using the AF PHB.

This paper evaluates the performance of voice flows transmitted through a DS domain implementing both EF and AF PHBs. We aim at evaluating the interaction between EF and AF PHBs when they are simultaneously implemented on the same DS infrastructure. These PHBs are implemented as a DS extension to the ns-2 simulator [15].

The remaining of the paper is organized as follows. Section 2 describes the models and the topology used in our simulations. Section 3 presents our simulation results. Some final considerations are discussed in Section 4.

2 Network Scenario

Interaction and continuity are the main characteristics of a conversation. Therefore, we evaluate delay, jitter, and loss rate of voice flows crossing a DS domain. The performance evaluation is carried out through simulations at the network scenario depicted in Figure 1. A number of voice flows are forwarded through the DS domain by either the EF PHB or the AF PHB. Along their path, EF traffic is shaped and reshaped to comply with the contracted rate and to not violate the EF definition. Besides the observed voice flows, the bottleneck link carries cross traffic composed by both voice and data flows. This cross traffic is forwarded as best-effort packets and compete for resources with the AF non-conforming packets.

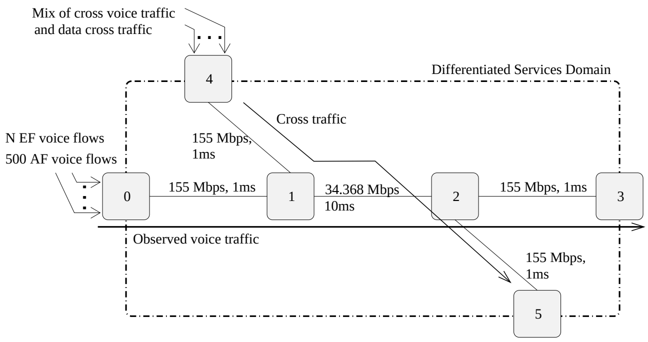


Figure 1: Network scenario.

The voice sources generate voice traffic following the On-Off model. When an On-Off source is in the On state, it generates voice traffic at a peak rate of 64 kbps. The On periods are determined by a random variable with exponential distribution and mean of 400 ms. The Off periods are also exponentially distributed by a random variable with mean of 600 ms. Therefore, the average transmission rate for the On-Off traffic is 25.6 kbps. The packets of the On-Off sources have 20 octets [16], unless otherwise stated. The TCP cross traffic sources use the TCP-Reno implementation. The TCP cross traffic is generated by FTP applications, which are active during the whole simulation, with packets of 1500 octets.

Each node of the DS domain holds two queues served by a WRR (Weighted Round Robin) scheduler. The first is a FIFO queue that receives EF traffic (EF queue). The second is a queue with active queue management implemented by a RIO (RED with IN and OUT) [17] scheme that receives the AF aggregation and the best-effort traffic (AF-BE queue). The WRR scheduler allocates a service share of the output bandwidth to each queue. The EF share must be enough to hold the EF definition at any moment [7].

3 Simulation Results

We assume that to keep a normal interactive conversation, the end-to-end delay of voice traffic in one direction must be below 150 ms [18]. A small jitter is suitable because it reduces the buffer capacity needed by the receivers to compensate these variations in delay. Low loss rates are also desirable to maintain the continuity of voice flows.

3.1 EF Traffic Shaping

In the case of WRR scheduling, traffic conditioning applied to the EF aggregation is needed in the DS domain according to the service share allocated to the EF aggregation. In order to allow the reserved share rate to be less than the peak rate of the EF aggregation, such aggregation must be shaped. Shaping allows to control, or even suppress, bursts from the aggregated traffic. When a token bucket mechanism is used, the bucket depth limits the allowed burst size. On the other hand, If the bucket depth is equivalent to one packet, bursts are completely suppressed. In this case, the service share rate of the WRR allocated to the EF may be equal to the shaper's token rate, i.e. less than the aggregation's peak rate. Nevertheless, shaping implies an additional delay for packets retained in the shaping buffer. If the bucket depth of the shaper is larger than one packet, the reserved service share rate of WRR belonging to EF must be equal to the sum of the individual peak rates. Despite this, because of the work-conserving nature of WRR, when the EF traffic is idle the remaining aggregations use the bandwidth initially allocated to the EF traffic.

The shaping of the EF traffic is implemented by a token bucket mechanism associated to a buffer at the ingress of a DS domain and along the DS-compliant nodes therein. On our simulations the bursts of voice traffic are completely suppressed by setting the bucket depth to be equal to one packet. First, we investigate the delay introduced by a traffic shaper in a flow aggregation. We define the *individual flow shaping rate* as the resulting rate that would be obtained by each component of a flow aggregation if they were equally affected by the shaping of the aggregation. Therefore, the actual token rate of the EF aggregation shaper is the product of the *individual flow shaping rate* and the number of EF active flows. Figure 2 shows the delay of a reference flow that is part of an aggregation of N flows that is shaped to an *individual flow shaping rate* of 30 kbps. The figure presents the average and 99^{th} percentile of the buffer delay. As the number of active voice flows in the aggregation increases, the individual delay drastically decreases because of the increasing statistical gain obtained within larger aggregations.

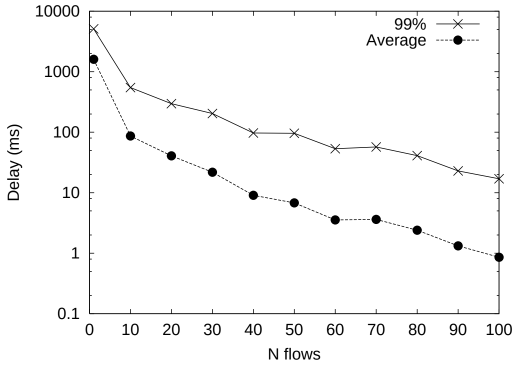


Figure 2: Delay in an aggregation shaper.

3.2 AF Traffic Marking

Packets generated by AF sources are marked in accordance with a contracted traffic profile. Such marking is performed by markers (profilers) located at the sources or at the DS domain boundary. In the first case, a policer for the AF aggregation must be responsible for the remarking of packets if, at any moment, the aggregation exceeds the sum of the individual profilers. In this paper, the markers are implemented as token buckets. Packets that disagree with the contracted individual profile are marked as OUT, otherwise they are marked as IN. These markers are located at the sources and a policer at the DS domain boundary controls the AF aggregation. Non-conforming packets of the AF aggregation are remarked as OUT and forwarded into the DS domain, where they will compete for resources with the best-effort traffic.

Markers are configured in such a way that 90% of the packets generated by the AF sources are marked as IN. The remaining packets are marked as OUT. In order to obtain the marker's bucket depth that performs the described marking, the following simulation has been carried out. The token rate of the bucket is kept at 25.6 kbps, while the bucket depth ranges from the equivalent of 1 to 1000 packets of 20 octets. Figure 3 shows that 90% of packets are marked as IN when the marker has a bucket depth equivalent to 67040 bits. This result is used in the configuration of markers at the AF traffic sources and also of the aggregation policer. The resources allocated within the DS domain take into account that the AF IN packets correspond to 90% of the total incoming AF traffic.

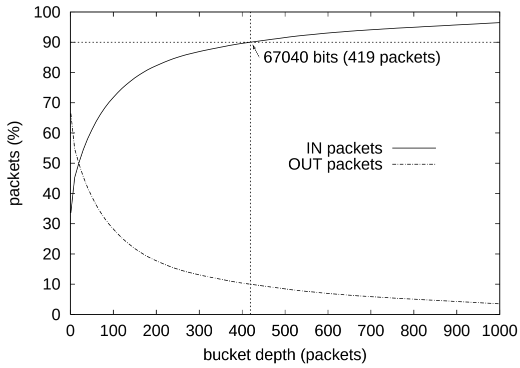


Figure 3: Effect of the bucket depth on packet marking.

3.3 Voice Traffic Evaluation

In this subsection we analyze the performance evaluation of voice flows transported by both the EF and AF PHBs when they are simultaneously implemented on the same DS domain. The active management of the AF-BE (Assured Forwarding - Best-Effort) queue is performed by a RIO algorithm. The discard discrimination of OUT packets in comparison with IN packets is achieved by tuning the RIO parameters. In our simulations, the routers of the DS domain implement the RIO algorithm with [800, 1200, 0.002, 0.02][1] for the IN packets and [400, 800, 0.004, 0.1] for the OUT packets. These parameters discriminate OUT packets, avoiding the discard of IN packets during periods of congestion. The size of the AF-BE queue is equivalent to 1600 packets.

The variation of the *individual flow shaping rate* used on the configuration of the EF aggregation shaper at each simulation implies the modification of the service share rates allocated to each aggregation by the WRR scheduler inside the DS domain. The higher the token rate of the EF shaper, the larger the service share to handle the EF traffic, in order to keep valid the EF definition. As a consequence, the other share is diminished. The goal of the next simulation is to evaluate the delay of voice flows inside the DS domain when they use the EF or the AF service. This delay includes the delay introduced by the EF aggregation (re)shapers located across the DS domain.

Figure 4 shows the average delay of two reference flows, one EF and the other AF. The 99% percentile of each is also shown. There are 250 EF voice flows and 500 AF voice flows besides the cross traffic. The *individual flow shaping rate* of the EF aggregation shaper is varied. The increase in the *individual flow shaping rate*

[1] $[min\_thresh,\ max\_thresh,\ q_{weight},\ P_{max}]$

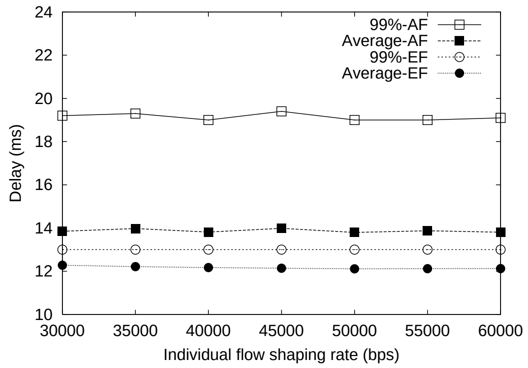
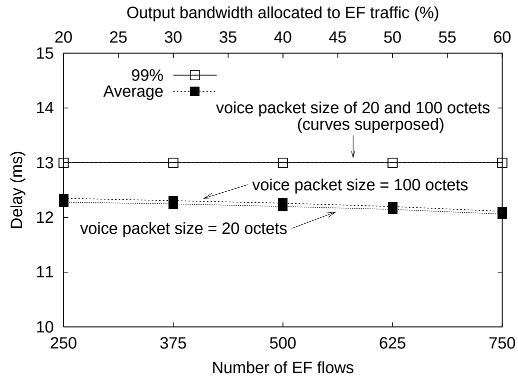
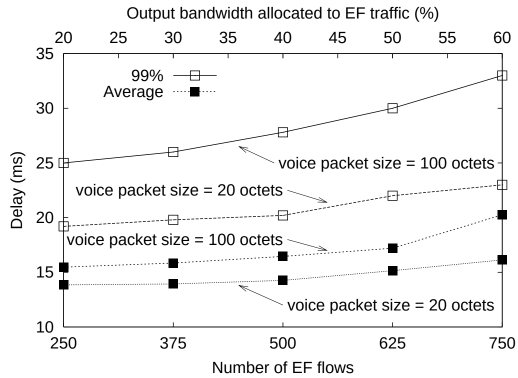


Figure 4: Delay of EF and AF flows versus the variation of the *individual flow shaping rate* of EF aggregation shapers.



(a) EF Traffic



(b) AF Traffic

Figure 5: EF and AF flows versus the variation of the number of active EF flows.

slightly decreases the delay of the EF aggregation. This increase implies the reduction of the WRR share allocated to the AF aggregation. Nevertheless, the delay of the AF aggregation does not significantly increase even for an *individual flow shaping rate* of 60 kbps, close to the individual peak rate. That is because when the EF aggregation takes an *individual flow shaping rate* of 60 kbps, about half the bottleneck, there is still enough bandwidth to carry the AF voice flows. Moreover, when the EF voice traffic is idle, because of the work-conserving nature of WRR, the AF-BE queue is served. There are no losses of EF or AF IN packets, while the AF OUT losses are under 1.5% of the total AF OUT packets. The observed jitter is less than 1 ms for 99% of the EF packets and less than 2 ms for 99% of the AF packets.

If we keep constant the *individual flow shaping rate* of the shaper, but vary the number of EF active voice flows, the service share of the WRR must also be adjusted to guarantee that the EF definition is preserved. In the next simulation, we keep the *individual flow shaping rate* at 30 kbps and vary the number of EF flows from 250 to 750, adjusting the service shares accordingly. We repeat the simulation for voice packets of 100 octets. Figure 5 presents the results. Again the service share allocated to the EF aggregation increases, but now there are more EF voice flows that use this reservation. Therefore, the AF aggregation, composed by the same 500 voice flows, finds less idle periods. As we increase the number of EF active voice flows, the EF reference flow suffers approximately the same delay. The AF flows, however, are penalized and delayed because they have a smaller share of available bandwidth to them. As we use a larger voice packet, the individual penalty is greater due to the greater transmission delay of each voice packet.

The variation of the number of EF flows also significantly influences the loss rate of the AF OUT packets as shown in Figure 6. Note that a loss rate of 10% of the AF OUT packets implies a loss rate of only 1% of the total of AF packets. There are no losses of EF packets or AF IN packets. If the reduction of the AF share causes a congestion in the AF-BE queue, both best-effort and AF non-conforming packets (OUT packets) have precedence in discard.

The next simulation investigates the influence of the TCP cross traffic packet size on the delay of the EF and AF voice traffic. As shown in Figure 7, the delay of the AF packets depends on the size of the TCP packets. That is because they share the AF-BE queue. Whenever the active queue management algorithm accepts a packet (IN or OUT) into the queue, that packet will

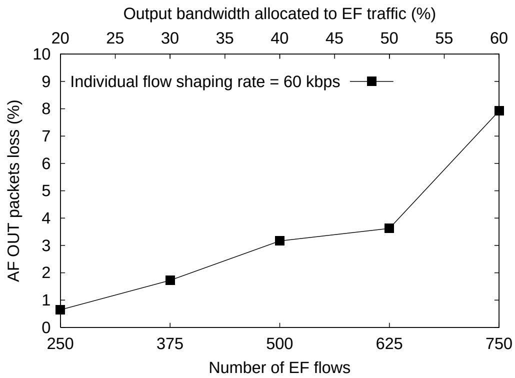


Figure 6: Loss rate of AF OUT packets.

not be discarded and there is no reordering of packets. That is, in order to be sent, packets in the queue must wait the forwarding of all packets ahead of them. The larger the TCP packets, the later the AF voice packets are forwarded. The delay of the EF packets, however, is not significantly affected by TCP packet sizes because the EF aggregation uses an exclusive WRR share. Table 1 shows the average delay of the EF reference flow for different TCP packet sizes.

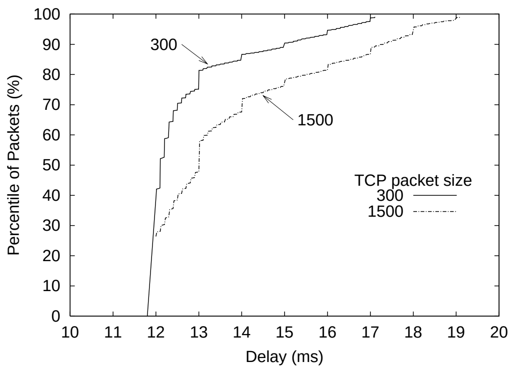


Figure 7: Delay of AF flows for different TCP packet sizes.

Table 1: Average delay of EF flows for different TCP packet sizes.

TCP Packet Size (octets)	Average EF Delay (ms)
300	12.0383
700	12.0585
1100	12.1072
1500	12.1474

4 Conclusion

In this paper, we evaluated the service offered to voice traffic by the Expedited and the Assured Forwarding PHBs when they are simultaneously implemented in the same DS domain. The increase of the token bucket rate at the EF aggregation shaper implies a larger share to this aggregation within the DS domain. Nevertheless, this increase does not significantly affect the AF traffic as long as AF can use the available bandwidth due to idle periods of EF traffic. When these idle periods become scarce by increasing the number of EF flows, the AF traffic is penalized. Due to sharing the same queue with best-effort traffic, the AF aggregation experiences a stronger influence on delay by TCP packet sizes than the EF aggregation. We conclude that the mutual influence of EF and AF PHBs is characterized by the tradeoff between their service share rates in the WRR scheduler. An extra amount of EF voice flows may be allowed into the DS domain as long as the performance of AF voice flows remains in acceptable limits.

References

[1] C. Diot, C. Huitema, and T. Turletti, "Multimedia applications shuold be adaptive," in *HPCS'95*, Aug. 1995.

[2] M. D. de Amorim, O. C. M. B. Duarte, and G. Pujolle, "Multi-criteria arguments for improving the fairness of layered multicast applications," in *Proceedings of Broadband Communications, High Performance Networking, and Performance of Communication Networks - Networking'2000*, no. 1815 in Lecture Notes in Computer Science, pp. 1–10, May 2000.

[3] E. Crawley, R. Nair, B. Rajagopalan, and H. Sandick, "A framework for QoS-based routing in the Internet," *Internet RFC 2386*, Aug. 1998.

[4] L. H. M. K. Costa, S. Fdida, and O. C. M. B. Duarte, "Distance-vector qos-based routing with three metrics," in *Proceedings of Broadband Communications, High Performance Networking, and Performance of Communication Networks - Networking'2000*, no. 1815 in Lecture Notes in Computer Science, pp. 847–858, May 2000.

[5] S. Blake, D. L. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," *Internet RFC 2475*, Dec. 1998.

[6] K. Nichols, S. Blake, F. Baker, and D. L. Black, "Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers," *Internet RFC 2474*, Dec. 1998.

[7] V. Jacobson, K. Nichols, and K. Poduri, "An expedited forwarding PHB," *Internet RFC 2598*, June 1999.

[8] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured forwarding PHB group," *Internet RFC 2597*, June 1999.

[9] J. F. de Rezende, "Assured service evaluation," in *Proceedings of the IEEE GLOBECOM'99*, vol. 1, (Rio de Janeiro, Brazil), pp. 100–104, Dec. 1999.

[10] C. Dovrolis and P. Ramanathan, "Proportional differentiated services, part II: Loss rate differentiation and packet dropping," in *Proceedings of IWQoS'2000*, June 2000.

[11] N. Semret, R. R.-F. Liao, A. T. Campbell, and A. A. Lazar, "Peering and provisioning of differentiated Internet services," in *Proceedings of IEEE INFOCOM'2000*, Mar. 2000.

[12] H. Naser, A. Leon-Garcia, and O. Aboul-Magd, "Voice over differentiated services," *Internet Draft*, Dec. 1998. `draft-naser-voice-diffserv-eval-00`.

[13] A. Ziviani, J. F. de Rezende, and O. C. M. B. Duarte, "Towards a differentiated services support for voice traffic," in *Proceedings of the IEEE GLOBECOM'99*, vol. 1, pp. 59–63, Dec. 1999.

[14] N. Seddigh, B. Nandy, and P. Pieda, "Study of TCP and UDP interaction for the AF PHB," *Internet Draft*, Aug. 1999.

[15] K. Fall and K. Varadhan, "NS – notes and documentation," tech. rep., The VINT Project, Jan. 1999.

[16] Cisco Systems, "Packet voice networking – reference guide," *White Paper*, 1998.

[17] D. D. Clark and W. Fang, "Explicit allocation of best effort packet delivery service," *IEEE/ACM Transactions on Networking*, vol. 6, pp. 362–373, Aug. 1998.

[18] ITU-T, "One-way trasmission time." Recommendation G.114, Mar. 1993.