# Effects of Traffic Aggregation on Quality of Service Parameters - Instantaneous Packet Delay Variation from the Viewpoint of Micro Flow Granularity Index

Chandana Watagodakumbura[1] and
Andrew Jennings[2]
*RMIT University, Australia*
*E-mail: S9841378@student.rmit.edu.au[1], ajennings@rmit.edu.au[2]*

Nirmala Shenoy
*Rochester Institute of Technology*
*U.S.A.*
*E-mail: ns@it.rit.edu*

## Abstract

*Differentiated Services (DiffServ) Architecture is based on aggregation of traffic as opposed to per flow traffic. When DiffServ is used in real time traffic environment it is important to study the effects of traffic aggregation on quality of service parameters. One-way delay and instantaneous packet delay variation (IPDV) are the most commonly used quality of service parameters for real time traffic. In this paper we study how one-way delay and IPDV are affected by traffic aggregation. We discover through our simulations and analytical models that IPDV remains unchanged with traffic aggregations as long as the micro flow granularity index (MFGI) is unchanged, while one-way delay is not affected purely due to aggregation. Further the IPDV of a given flow increases when the MFGI decreases for a given number of aggregated micro flows. We used a priority queue scheduler for our simulation where the real time traffic was served with highest priority.*

**Keywords:** Differentiated services architecture, quality of service, real time traffic, IPDV, traffic aggregation.

## 1. Introduction

A scalable solution for supporting quality of service in the rapidly expanding Internet is sought through Differentiated Services (DiffServ) framework [1]. Unlike in Integrated Services (IntServ) where each traffic micro flow is treated separately, the DiffServ treats traffic in aggregation [2]. The aim is to achieve scalability and better bandwidth utilisation. However, one major problem is having to deal with different types of traffic generated from different applications.

Expedited Forwarding (EF) is defined in relation to the DiffServ framework to transmit packets with minimum delay [4]. One main requirement in EF servicing is that packet service rate is always greater than or equal to the arrival rate. The same requirements apply for the transmission of real time packets related to the applications such as voice and video-conferencing. Two major parameters of concern in transmission of real time packets are one-way (end-to-end) delay and instantaneous packet delay variation (IPDV), which is commonly known as jitter [5]. Jitter or IPDV is based on one-way delay of a selected pair of packets while one-way delay is based on the queue delays at the intermediate nodes.

Weighted Fair Queuing (WFQ) [3,13,14] and Priority Queue (PQ) are two contenders for packet scheduling in DiffServ framework. In PQ scheme the real time packets can be serviced with highest priority. As expected, PQ has better one-way delay and IPDV outcomes compared to WFQ for real time data [6]. The aim of this paper is to study the effects of traffic aggregation on quality of service parameters such as one-way delay and IPDV and, the future work will be based on how packet scheduling can be done in order to overcome these effects [10, 11, 12].

Section II of this paper describes the related work carried out by Ferrari et al. [7] and the methodology for measuring one-way delay and IPDV. The basis on which the simulations were carried out was explained in section III while the observations made and the results obtained from the simulations were given in section IV. Section V contains the analytical results and the comparisons made against the results obtained from the simulations. The summary of results is given in section VI under conclusion and we conclude the paper by giving our future direction of work in section VII.

## 2. Related Work

Ferrari et al. [7] have studied the variation of quality of service parameters in the premium service with the number of micro flows. It was found in this study that the packet-loss, one-way delay and IPDV increase with the degree of aggregation. However, in the case of one-way delay and IPDV it was not differentiated whether the effects are caused simply by the increase in the load with the number of micro-flows. To overcome this, in our simulation, we keep the total aggregated traffic constant irrespective of the number of micro flows.

One-way delay of a packet transmitted from a given source to a given destination is defined as dT if the source sends the first bit of the packet to the destination at wire-time T and the destination receives the last bit of the packet at wire-time T+dT [5]. The major portion of one-way delay is composed of the delays at the intermediate nodes from the source to the destinations. That is, for a packet passing through n intermediate nodes the one-way delay *D* is given by

$$D = \sum_{i=1}^{n} D_i$$

$D_i$ is the delay at $i^{th}$ node.

The IPDV of a pair of packets of a given stream, transmitted from a given source to a given destination, is the difference between the one-way delays of the selected packets [6]. That is, the IPDV of a pair of packets is defined as ddT if the source sends the first bit of the first packet to the destination at wire-time T1 and the first bit of the second packet at wire-time T2 and the destination receives the last bit of the first packet at T1+dT1 and the last bit of the second packet at T2+dT2 where ddT = | dT2 – dT1|. This is illustrated in Figure 1.

The IPDV between packet *i* and packet *k* is

$$\mathrm{IPDV}_{ik} = \mathrm{dd}T_{ik} = |\, \mathrm{d}T_k - \mathrm{d}T_i \,|$$

In real time packet streaming the packets that arrive relatively early at the destination are delayed in a play-out buffer so that the data is relayed uniformly to the application. This makes it more meaningful to select the 2 consecutive packets of the same stream for IPDV calculation and hence we followed this criterion for all our simulations.

## 3. Simulation Model

The variations of one-way delay and IPDV against the number of micro flows of real time traffic were obtained. As a measure of total aggregated real time traffic we introduce the term *real time fraction* (RTF); it is defined as the ratio of amount of real time traffic serviced to the total amount of traffic serviced by the queue for a given period of time. That is,

$$RTF = \frac{n_R^{\tau}}{n_T^{\tau}}$$

$n_R^{\tau}$ is the number of real time packets serviced during the period $\tau$ and $n_T^{\tau}$ is the number of total packets serviced by the queue during the period $\tau$.

A second term *micro flow granularity index* (MFGI) is defined as a measure of fraction of packets of a single micro flow transmitted over given period of time. That is, if $n_m^{\tau}$ is the number of packets of real time micro flow m serviced during the period $\tau$ the micro flow fraction (MFGI) with respect to the flow m is defined as,

$$MFGI = \frac{n_m^{\tau}}{n_R^{\tau}}\ .$$

The results were obtained for packet inter-arrival times of exponential distribution. The combined packet arrival rate of all traffic for all simulations was 1000 packets per second. The queue service rate was equal to the combined arrival rate always, irrespective of the value of RTF. That is the fraction of non-real time traffic serviced by the queue was (1 - *RTF*). The packet size for all traffic was kept constant at 512 bits for all simulations. The packet service time was proportional to the size of the packet, thus giving constant service time to each packet. OPNET Modeller 7.0 was used for the simulation and for simplicity only a single node was considered for measurements [Figure 3]. The results were obtained for aggregation of 2, 4, 8 and 12 micro flows, apart from the single micro flow case. In each of the above cases 2 separate sources generated the non real time packets. Priority queue scheduling scheme was used with 3 levels of priority with real time packets being serviced with the highest priority.

## 4. Observations and results

### 4.1 Effects on average queue delay

The average queue delay for real time traffic remained constant except for statistical variations and irrespective of the number of micro flows. (Table 1 and Figure 2). In this simulation, despite the varying number of micro flows the total aggregate remained constant.

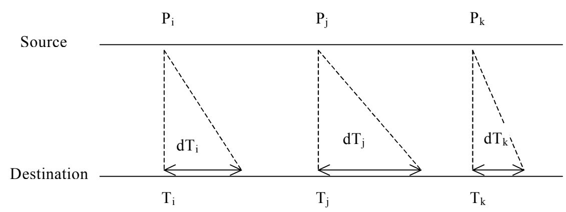


**Figure 1. Illustration of IPDV definition**

**Table 1. Average queue delay variation with number of micro flows**

| Number of micro flows | Average queue delay in milli seconds | | | |
|---|---|---|---|---|
| | RTF 0.2 | RTF 0.4 | RTF 0.6 | RTF 0.8 |
| 1 | 1.62 | 1.86 | 2.22 | 3.61 |
| 2 | 1.61 | 1.87 | 2.26 | 3.48 |
| 4 | 1.62 | 1.83 | 2.24 | 3.63 |
| 8 | 1.61 | 1.83 | 2.26 | 3.54 |
| 12 | 1.62 | 1.83 | 2.24 | 3.36 |

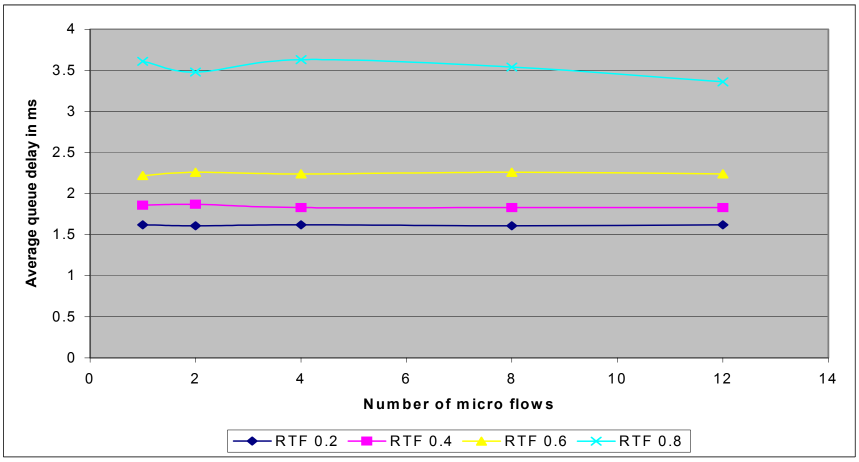


**Figure 2. Average queue delay variation with number of micro flows**

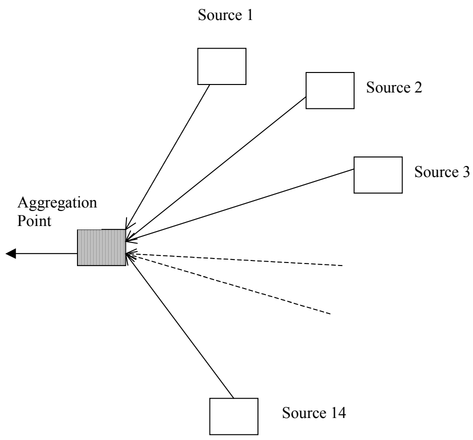


**Figure 3. Aggregation at a single node**

### 4.2 Effect on IPDV

The IPDV of a given micro flow remained unchanged, except for the statistical variations, with the increase of the number of micro flows when MFGI of the given flow is constant. (Table 2 and Fig 4). Further, for a number of aggregated micro flows the IPDV of the given micro flow increased when the MFGI of the flow decreased. This is explained in detail later.

## 5. Analytical comparisons

### 5.1 Effect on Average Queue Delay

The average total time in the system by a packet of priority group p ($T_p$) is calculated as the sum of the average waiting time in the queue ($W_p$) and the average service time ($\overline{x}_p$) [7,8]. That is,

$$T_p = W_p + \overline{x}_p \tag{1}$$

$W_p$ is given by [7, 8]

$$W_p = \frac{W_0}{(1-\sigma_p)(1-\sigma_{p+1})} \quad p = 1, 2, \dots P \tag{2}$$

where,

$$W_0 = \sum_{i=1}^{P} \rho_i \frac{\overline{x_i^2}}{2\overline{x}_i} \text{ and} \tag{3}$$

**Table 2. IPDV variation with MFF and number of micro flows**

| Number of micro flows | IPDV in milli seconds | | | |
|---|---|---|---|---|
| | MFGI 0.5 | MFGI 0.25 | MFGI 0.125 | MFGI 0.083 |
| 2 | 0.86 | 1.08 | 1.39 | 1.53 |
| 4 | 0.86 | 1.11 | 1.31 | 1.48 |
| 8 | 0.86 | 1.09 | 1.35 | 1.49 |
| 12 | 0.86 | 1.10 | 1.31 | 1.45 |

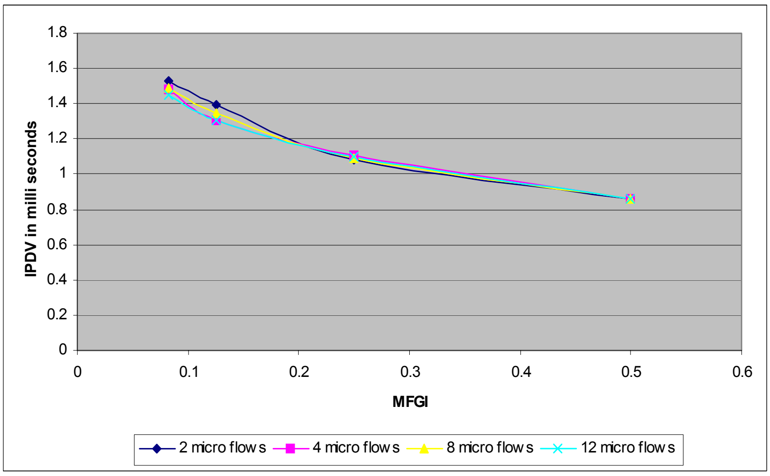


**Figure 4. Variation of IPDV with MFGI**

$$\sigma_p = \sum_{i=p}^{P} \rho_i \qquad (4)$$

Note: $P$ is the highest priority; that is the priority increases with the number.

$\overline{x_i^2}$ and $\overline{x}_i$ are the first and the second moments of the service time of i[th] group, the ratio $\overline{x_i^2} / 2\overline{x}_i$ is the residual life if an i[th] group packet is in service and $P$ is the number of priority groups or the highest priority.

The variance $\sigma_x^2$ of service time $x$ is given by

$$\sigma_x^2 = \overline{x^2} - (\overline{x})^2 \qquad (5)$$

Since the service time is constant for all packets

$$\sigma_x^2 = 0 \qquad (6)$$

Hence (5) gives

$$\overline{x^2} = (\overline{x})^2 \qquad (7)$$

Further, for mean arrival rate $\lambda$

$$\rho = \lambda\, \overline{x} \qquad (8)$$

From (3), (7) and (8) we get

$$W_0 = \frac{(\overline{x})^2}{2} \sum_{i=1}^{P} \lambda_i \qquad (9)$$

Using (1), (2), (4) and (9) we calculate the queue delay values for different RTF values and different number of micro flows.

We see that the analytical results in Table 03 are in concurrence with the direct results measured from the simulation in Table 01.

### 5.2 Effect on IPDV

How the traffic aggregation affects the IPDV of each micro flow is elaborated in the following diagrams and the paragraphs.

**Table 3. Queue delay (analytical) variation with number of micro flows**

| Number of micro flows | RTF 0.2 | | RTF 0.4 | | RTF 0.6 | | RTF 0.8 | |
|---|---|---|---|---|---|---|---|---|
| | Inter arrival time for real time packets in ms | Average queue delay in ms | Inter arrival time for real time packets in ms | Average queue delay in ms | Inter arrival time for real time packets in ms | Average queue delay in ms | Inter arrival time for real time packets in ms | Average queue delay in ms |
| 1 | 5.02 | 1.62 | 2.43 | 1.92 | 1.65 | 2.27 | 1.23 | 3.73 |
| 2 | 5.06 | 1.63 | 2.49 | 1.84 | 1.66 | 2.26 | 1.26 | 3.41 |
| 4 | 5.02 | 1.61 | 2.48 | 1.87 | 1.65 | 2.29 | 1.23 | 3.66 |
| 8 | 5.07 | 1.62 | 2.52 | 1.83 | 1.66 | 2.26 | 1.25 | 3.55 |
| 12 | 4.91 | 1.63 | 2.5 | 1.84 | 1.67 | 2.24 | 1.26 | 3.38 |

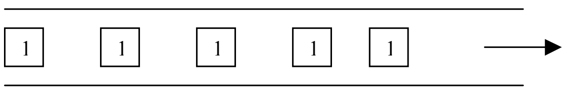


**Figure 5. Single micro flow**

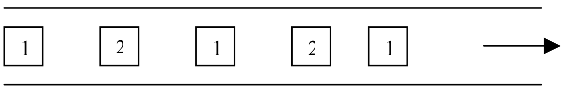


**Figure 6. Two micro flows**

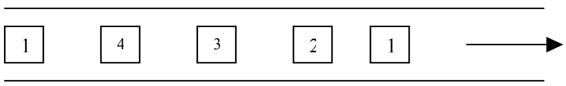


**Figure 7. Four micro flows**

Consider a queue serving aggregated traffic of equal packet size. In Figure 5 only the packets of a single micro flow (flow 1) are waiting to be served. In Figure 6 there are packets of 2 micro flows (1 and 2). In Figure 7 there are packets of 4 micro flows (1, 2, 3 and 4) to be served. In each of the 3 figures the aggregate packet arrival rates or packet inter arrival times are equal. Also consider a generic order of packets with respect to each micro flow. IPDV is defined as the difference of one- way delay between 2 consecutive packets of the same stream. Assume that the first packet of micro flow 1 of each figure starts servicing at time = 0. The waiting times for the 2nd packet of micro flow 1 in figures 5, 6 and 7 are 1 packet service time, 2 packets service time and 4 packets service time respectively. Since the servicing of 1st packet of the micro flow 1 was started at time = 0 its waiting time is zero. Therefore, the delay differences or IPDVs between the first 2 packets of micro flow 1 of each figure are 1 packet service time, two packets service time and 4 packets service time respectively. However, since the aggregate packet inter arrival times are the same, the average aggregate queue delays are the same for each case.

The MFGI in the above scenarios decreased clearly with the increase of the figure number. For example, the micro flow 1 of Figure 7 had more packets of other micro flows to be served in a given period of time compared to the other two figures.

In a non-preemptive queuing system, the waiting time of a packet of the highest priority group at a single node is composed of 2 parts:

- Delay due to the residual service time of the packet in service at the time of arrival of the packet.
- Delay due to the packets of highest priority group already in the queue at the time of arrival of the packet.

Note: the packet in service is not considered here.

We find an analytical expression for IPDV between i^th and (i-1)^th packet of a stream of highest priority below.

Let $W_i$ be the waiting time (delay due to highest priority group packets already in the queue) of packet i, its residual time (due to the packet currently in service) be $R_i$ and its delay at the node is $D_i$. Then

$$D_i = W_i + R_i \ \forall \ i > 0$$

IPDV between packets i and (i-1) is

$$|D_i - D_{i-1}| = |(W_i + R_i) - (W_{i-1} + R_{i-1})|$$
$$= |(W_i - W_{i-1}) + (R_i - R_{i-1})|$$
$$\forall \ i > 1 \qquad (10)$$

$W_i$ is equal to the summation of service times of all the packets in the queue at the time of arrival of i^th packet. That is,

$$W_i = \sum_{j=1}^{N_i} S_j^i \qquad (11)$$

where $S_j^i$ is the service time of the j[th] packet in the queue at the time of arrival of the i[th] packet of the stream and $N_i$ is the number of packets waiting to be served at arrival of the i[th] packet.

Similarly,

$$W_{i-1} = \sum_{j=1}^{N_{i-1}} S_j^{i-1} \tag{12}$$

$R_i$ is equal to the difference between the service time of the packet in service and the length of time it has been in service, at the time of arrival of the i[th] packet. That is,

$$R_i = \begin{cases} S_a^i - (t_a^i - t_s^i) & (t_a^i - t_s^i) \le S_a^i \\ 0 & (t_a^i - t_s^i) > S_a^i \end{cases} \tag{13}$$

where $S_a^i$ is the service time of the packet in service at the time of arrival of the ith packet, $t_a^i$ is the arrival time of the ith packet and $t_s^i$ is the service start time of the packet in service at the time of the arrival of ith packet.

Similarly,

$$R_{i-1} = S_a^{i-1} - (t_a^{i-1} - t_s^{i-1}) \tag{14}$$

From (10), (11), (12), (13) and (14) we get,

$$|D_i - D_{i-1}| = |\sum_{j=1}^{N_i} S_j^i - \sum_{j=1}^{N_{i-1}} S_j^{i-1} + [\{S_a^i - (t_a^i - t_s^i)\} - \{S_a^{i-1} - (t_a^{i-1} - t_s^{i-1})\}]| \tag{15}$$

For simplicity assume that all the packets are of the same size. Then,

$$S_j^i = S_j^{i-1} = S_a^i = S_a^{i-1} = S$$

Now (15) simplifies to,

$$|D_i - D_{i-1}| = S(N_i - N_{i-1}) - (t_a^i - t_a^{i-1}) + (t_s^i - t_s^{i-1})| \tag{16}$$

The constituents of the IPDV measured at the single intermediate node verify the end-to-end IPDV measured from the simulation. The equation (16) relates the constituents of IPDV and it is seen that the residual time of the packet in service is negligible compared to the number of real time packets in the queue at the time of arrival of the packets of concern (Tables 4, 5, 6, and 7).

## 6. Conclusion

We have observed through simulations and analytical results that the IPDV of a given flow increases when MFGI of the flow decreases for a given number of aggregated micro flows. Further, IPDV of a given flow remained unchanged, except for the statistical variations, with the increase of number of micro flows for a given MFGI of the flow. On the other hand the average queue-delay is not affected purely by traffic aggregation. As seen from Table 3, other than for statistical reasons, there is no variation in inter arrival times of real time packets due to traffic aggregation. A variation in the inter arrival times of real time packets is the only cause that could have made an impact on average queue delay.

## 7. Future Direction

We plan to extend this work to a more realistic multi-node network. There are two inherent problems associated in priority queuing disciplines: the burstiness nature that accumulates over the nodes causing negative effects on quality of service parameters and the starvation of lower priority queues. We observed that IPDV had significant effects whereas delay had no effects purely due to aggregation. By introducing some delay on the packets that arrive relatively early at a node the level of burstiness can be reduced. This would also reduce the average IPDV at the node.

**Table 4. IPDV variation of a flow with 0.5 MFGI and for 80% RTF**

| | 2 micro flows | 4 micro flows | 8 micro flows | 12 micro flows |
|---|---|---|---|---|
| $\overline{\lvert D_i - D_{i-1} \rvert}$ | 0.87 | 0.88 | 0.87 | 0.88 |
| $\overline{S \lvert (N_i - N_{i-1}) \rvert}$ | 0.81 | 0.83 | 0.82 | 0.83 |
| Effects of residual time of the packet in service on IPDV | 0.06 | 0.05 | 0.05 | 0.05 |
| IPDV (direct simulation) | 0.86 | 0.86 | 0.86 | 0.86 |

**Table 5. IPDV variation of a flow with 0.25 MFGI and for 80% RTF**

| | 2 micro flows | 4 micro flows | 8 micro flows | 12 micro flows |
|---|---|---|---|---|
| $\overline{\lvert D_i - D_{i-1} \rvert}$ | 1.09 | 1.14 | 1.12 | 1.12 |
| $\overline{S \lvert (N_i - N_{i-1}) \rvert}$ | 1.05 | 1.09 | 1.06 | 1.07 |
| Effects of residual time of the packet in service on IPDV | 0.04 | 0.05 | 0.06 | 0.05 |
| IPDV (direct simulation) | 1.08 | 1.11 | 1.09 | 1.10 |

**Table 6. IPDV variation of a flow with 0.125 MFGI and for 80% RTF**

| | 2 micro flows | 4 micro flows | 8 micro flows | 12 micro flows |
|---|---|---|---|---|
| $\overline{\lvert D_i - D_{i-1} \rvert}$ | 1.31 | 1.34 | 1.38 | 1.32 |
| $\overline{S \lvert (N_i - N_{i-1}) \rvert}$ | 1.25 | 1.30 | 1.31 | 1.26 |
| Effects of residual time of the packet in service on IPDV | 0.06 | 0.04 | 0.07 | 0.06 |
| IPDV (direct simulation) | 1.39 | 1.31 | 1.35 | 1.31 |

**Table 7. IPDV variation of a flow with 0.083 MFGI and for 80% RTF**

| | 2 micro flows | 4 micro flows | 8 micro flows | 12 micro flows |
|---|---|---|---|---|
| $\overline{\lvert D_i - D_{i-1} \rvert}$ | 1.56 | 1.51 | 1.52 | 1.49 |
| $\overline{S \lvert (N_i - N_{i-1}) \rvert}$ | 1.50 | 1.45 | 1.46 | 1.44 |
| Effects of residual time of the packet in service on IPDV | 0.06 | 0.06 | 0.06 | 0.05 |
| IPDV (direct simulation) | 1.53 | 1.48 | 1.49 | 1.45 |

# References


[1] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services": RFC 2475, 1998.

[2] D. D. Clark, Scott Shenker, and Lixia Zhang, "Supporting Real-Time Applications in an Integrated Services packet Network: Architecture and Mechanism", Proc. ACM SIGCOMM 92, Aug 1992.

[3] Abhay K Parekh, and Tobert G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case", IEEE Transactions on Networking, Vol. 1, No 3, June 1993.

[4] V. Jacobson, K. Nichols, and K. Poduri, "An expedited Forwarding PHB", RFC 2598.

[5] G. Almes, S. Kalidindi, and M. Zekauskas, "A One-way Delay Metric for IPPM": RFC 2679, 1999.

[6] C. Demichelis and P. Chimento, "IP Packet Delay Variation Metric for IPPM", Internet Draft (draft-ietf-ippm-ipdv-09.txt)

[7] T. Ferrari and P. F. Chimento, "A Measurement-based Analysis of Expedited Forwarding PHB Mechanisms", Proceedings of IWQoS'2000, Pittsburgh, June 2000.

[8] Leonard Kleinrock, "Queueing Systems Volume II: Computer Applications", John Wiley and Sons, 1976, pp. 106-122.

[9] Leonard Kleinrock, "Queueing Systems Volume I: Theory", John Wiley and Sons, 1975.

[10] J. Mao, W. M. Moh, and B. Wei, "PQWRR Scheduling Algorithm in Supporting of DiffServ", Communications, 2001 ICC 2001 IEEE International Conference, Volume: 3, 2001 Page(s): 679 -684.

[11] C. Dovrolis, D. Stiliadis, and P. Ramanathan, "Proportional differentiated services: Delay differentiation and packet scheduling", ACM SIGCOMM-99, September 1999.

[12] T. N. Quynh, H. Karl, A. Wolisz, and K. Rebensburg, "Using only Proportional Jitter Scheduling at the boundary of a Differentiated Service Network: simple and efficient", Universal Multiservice Networks, 2002. EDUMN 2002. 2nd European Conference, 2002, Page(s): 116 -123.

[13] H. Wang, C. Shen, and K. G. Shin, "Adaptive-Weighted Packet Scheduling for Premium Service", Communications, 2001 ICC 2001. IEEE International Conference, Volume: 6, 2001, Page(s): 1846 -1850, vol.6.

[14] C. LI, S. Tsao, M. C. Chen, Y. Sun, and Y. Huang, "Proportional Delay Differentiation Service based on Weighted Fair Queuing", Computer Communications and Networks, 2000. Proceedings. Ninth International Conference, 2000, Page(s): 418 -423.