GENERALIZED PROCESSOR SHARING (GPS) TECHNIQUES

By

Onah F. I, Mom J. M. and Ani C. I.
Centre for Communication Networking
(ETF Centre of Excellence)
Department of Electronic Engineering
University of Nigeria, Nsukka, Nigeria.

ABSTRACT

In ***this era of service integration and proliferation of value-added services, ICT networks have traffic volume that tends to overwhelm network resources. Satisfaction of quality of services (QoS) is usually difficult especially with static network resources. A flexible and popular technique, Generalized Processor Sharing (GPS), provided an effective and efficient utilization of the available resources at the face of stringent and varied QoS requirements. This paper, therefore, presents the comparison of two GPS techniques – PGPS and CDGPS – based on performance with limited resources in an isolated node. The better performing PGPS technique was further investigated under varying QoS, resources and traffic load. PGPS resource utilization behaviors for given QoS parameters (delay, and reliability) were presented and applied in the realization of a network performance management system at the ETF centre of excellence in Communication Networking, University of Nigeria, Nsukka.***

Introduction

It has severally been stated that no matter how accurate a network may have been designed and implemented, the operation may not be effective and efficient, especially with the ever rising traffic volume, without appropriate network performance management technique [1]. In this era of service integration and proliferation of value-added services, network traffic volume continuously increases and overwhelms network resources with consequent QoS degradation. Generalized Processor Sharing (GPS) traffic routing techniques provide flexible and efficient utilization of the available resources at the face of stringent and varied QoS requirements [2,3,4,5]. There are variations in the GPS technique such as Packet Generalized Processor Sharing (PGPS), Code Division Generalized Processor Sharing (CDGPS) and Dynamic Weight Generalized Processor Sharing (DWGPS).

The problem of increase in the volume of traffic as a result of introduction of new and varied services was successfully tackled using GPS technique through the optimization of the use of network resources while maintaining standard QoS. Definition of optimum resource utilization for any given QoS requirements was a tricky issue since QoS parameters place conflicting demands on the network resources, specifically buffer capacity [1]. PGPS packet-by-packet service discipline was implemented with leaky bucket admission control that allows the network to make a wide range of performance guarantees on throughput and delay [2,3]. The PGPS, some times referred to as Weighted Fair Queuing (WFQ), technique works by assigning a session a weight at each link on the session path. The weight of a session at a link decides the share of the resources to be made available and consequently the resources define the QoS. The weight reflects the amount of resources needed to guarantee the QoS specified for a particular traffic class [2,3,6,7]. Sessions were grouped into classes and sessions having the same parameters are grouped into same traffic class.

CDGPS is a dynamic bandwidth allocation technique that allows channel rates to be dynamically and fairly allocated in response to the variation of traffic rates in WCDMA systems [8]. It adjusts service rates of each traffic flow by varying the spreading factor and/or using multiple code channels, rather than allocating service time to each packet. The technique provides a trade-off between the Generalized Processor Sharing (GPS) fairness and efficient resource utilization. Simulation results and the analysis show that resource utilization for a given guaranteed QoS is enhanced.

In DWGPS, a session is defined as an active batch in the transmission queue. An important weight is selected and assigned based on the criterion that a batch from a higher-priority class will be assigned a relatively larger weight corresponding to a higher transmission rate. Other factors considered in the DWGPS technique in addition to those considered are the batch size and timer value.

This paper compares the performance of the PGPS technique with the CDGPS technique. Analytical results presented for the PGPS technique show its performance with respect to the allocation of resources bearing in mind QoS parameters of delay and packet-loss. Analysis of the model results showed that PGPS has better resource utilization with respect to delay and data loss. Furthermore, the performance of PGPS technique was investigated under varying traffic load (packet length).

Architecture

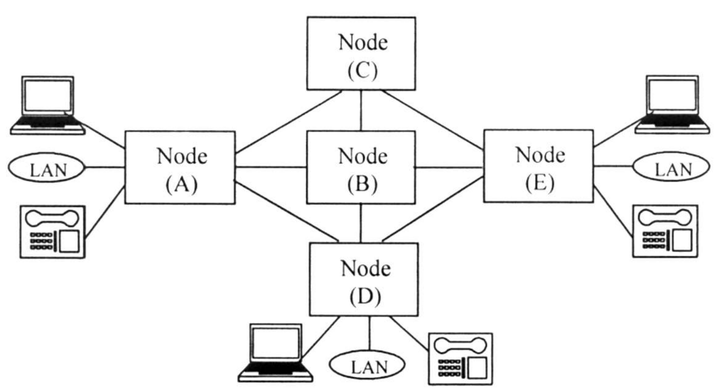

Figure 1: Network architecture

In this work, simple network architecture illustrated in figure 1 was employed. The network supported voice, video and data traffic. The network was based on virtual circuit switching technique. Each node realized the usual functions of traffic switching, processing and transmission. The function of traffic processing involved traffic routing and scheduling. GPS technique which is one of the several scheduling techniques employed in the integrated services (voice, data and video services) network to achieve optimum resource utilization and standard QoS provisioning was applied at a node in the network. The input traffic is from varied sources comprising of voice, video, and data traffic, which are bundled into flow classes [8].

The resources considered were buffer spaces and link transmission bandwidth. QoS parameters such as delay and data loss (reliability) were considered against the resources along a given traffic determined path. A definite path was defined for a given traffic at a given moment using a shortest path algorithm.

Traffic was scheduled into the transmission buffer attached to the path with the shortest distance to the destination and needed resources to guarantee the required QoS. Where the specified buffer space and the resources were not available then the next alternative path would be specified. The GPS scheduling technique was applied as a service discipline in order to realize QoS (delay, packet-loss) improvement for specified transmission buffer space and link bandwidth.

Model

The model was developed on the principle of an isolated node bearing in mind that all the network nodes are identical [1]. Figure 2 presents the network graphical model. Traffic flow path through the network was chosen arbitraly from Node (A) to Node (B) via Node (B), see figure 1. The network is modeled as a directed graph, $G=(V,E)$, with set of nodes V_i and a set of links E_i. Transmission rate R_i, and buffer space, K_i, defined as transmission link capacity, C, in a definite path from source to destination.

The set of nodes V_i may be represented as a function of transmission buffer space, K_i, and transmission bandwidth, R_i. Service discipline defines the output interface. For individual network service, the transmission buffer capacity varies as, K_i, while the bandwidth varies as, R_i.

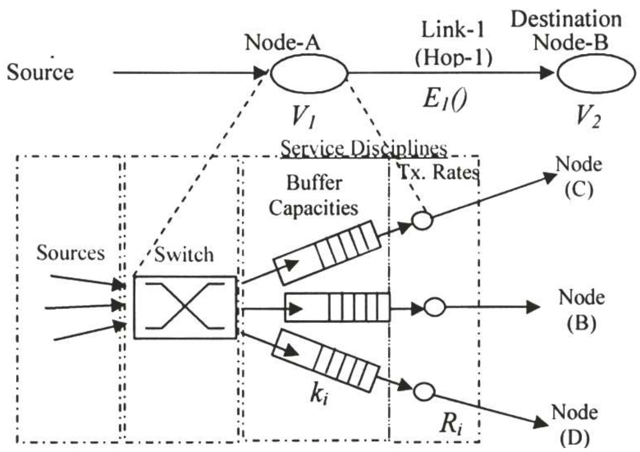

Figure 2: .Network traffic flow path model

Therefore, a transmission link (V_i,V_{i+1}) connecting two nodes, for instance, V_1 and V_2 may be characterized by QoS and the consequent resources – transmission capacity (transmission buffer space and bandwidth), *C*, transmission delay, *D*, and reliability(packet-loss), *Q*.

For a chosen path $p=f(V_1,V_2)$, the reliability, *Q(p)*, and delay, *D(p)*, were all expressed as functions with arguments corresponding to the required amount of resources, *C(p)*. The buffer space, *K(p)*, allocated for a given data stream in the service disciplines along the path and the bandwidth, *R(p)*, allocated for the same data stream along the same path were referred to as network resources. A path *p* is said to satisfy the QoS requirement of a given data stream if,

$Q(p)$ ³ Q_{req}; $D(p) \leq D_{req}$; $R(p) \leq R_{req}$; $k(p) \leq k_{req}$ (1)

Where, Q_{req} - minimum reliability,
D_{req} - maximum end-to-end delay,
R_{req} - maximum bandwidth, and
k_{req} - maximum buffer space.

Therefore, delay and data losses (reliability) in a network depended on three factors - traffic, interaction between service disciplines connected in tandem, and the allocated resources [2]. In order to guarantee the QoS for all users, each traffic source is shaped by a leaky bucket regulator. The regulator was defined by parameters s and r that represented the token buffer size and token generating rate respectively [3].

Server service discipline differentiates the two scheduling techniques considered in this work. In the case of PGPS technique, in addition with the parameters represented by *D*, *Q*, *R* and *K* every session is characterized by *s*, *ρ* and maximum length of packet, *L*, parameters. The weighting of the *D*, *Q*, *R* and K parameters were bounded by equation (1) and defined by equations (2) – (5) for each session and each link in the path [2,3,6,7]. Sessions are grouped into classes and sessions having the same parameters are grouped into same traffic class.

$$Q = \left[\frac{k}{\sigma + L}, 1 \right]_{\min} \quad ($$

$$D = \left[\frac{\sigma + L}{R}, \frac{k}{R} \right]_{\min} \quad ($$

$$K = [\sigma + L]_{\min} \quad (4)$$

$$R = g \quad (5)$$

where, g = minimum guaranteed rate for a user.

In the case of CDGPS, the server allocates *R* over all. The delay bound for the traffic sessions flow can be derived if the flow is regulated by leaky bucket with token bucket buffer size *s* and token generating rate, *r*. Maximum flow backlog per session, $Q_{backlog,max}$, was defined in equation (6).

$$Q_{backlog,\,max} \leq \sigma + \rho T \qquad (6)$$

where, $g \geq \rho$

Sessions delay bound was defined by equation (7).

$$D \leq \left[\frac{\sigma + \rho T}{g}\right] \leq \frac{\sigma}{g} + T$$

where, T = scheduling time ≈ mean packet inter arrival time.

Simulation

The model was simulated for two QoS parameters, delay and reliability (packet loss), for given network resource capacity - buffer sizes and bandwidths, packet lengths and leaky bucket parameters. The objectives of the simulation were to compare the two main GPS techniques (CDGPS and PGPS) and carryout further detailed analysis of the better technique. The simulations were run using block oriented network simulator – Network II.5.

In order to compare the two GPS techniques, in one hand, the simulation was run with bandwidth (rate of transmission) varied against constant packet sizes and buffer storage capacities to determine the corresponding delay. On the other hand, the simulation was also run with buffer size varied against constant packet lengths in order to determine the corresponding reliability.

The better technique was simulated using the same model to determine, in a greater detail, delay-bandwidth and reliability-buffer size relationships for varied transmission bandwidths and buffer sizes.

Results

The results illustrated in graph-form in figures 3-10 were obtained from the above simulations. The results were obtained in two parts using a block oriented network simulator. The results were obtained for traffic with varying packet lengths, L=200 bits, 500 bits, 750 bits, 1000 bits, and varying token bucket sizes, *s*=0.2 kbits, 0.8 kbits, 2 kbits.

PGPS technique [9] was compared with CDGPS technique [8] and results presented for resources (transmission buffer and bandwidth) utilization against QoS parameters (delay and reliability) in figures 3-6. Figures 7-10 presented further results on PGPS technique alone.

In figure 3, the bandwidth was varied between 0.1Mbps and 2Mbps and the corresponding delay determined for 0.2 and 0.8 Kbits leaky bucket sizes while mean packet size was kept at 1000 bits. The results show that with increase in transmission bandwidth the associated delay decreases exponentially for both CDGPS and PGPS techniques.

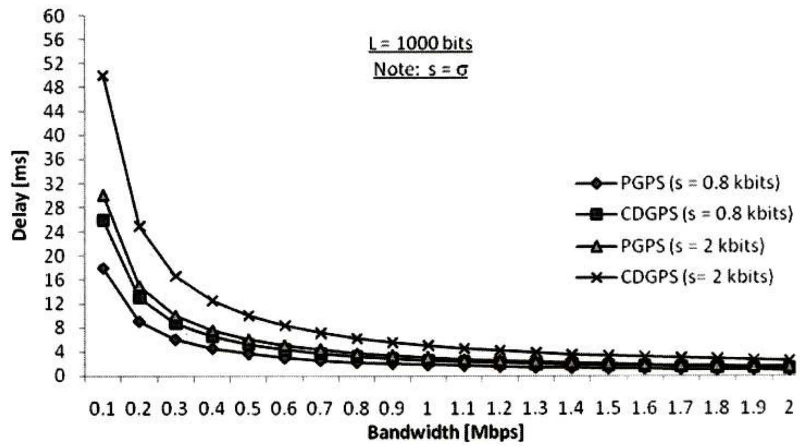

Figure 3: Delay [ms] versus bandwidth [Mbps] for both CDPGS and PGPS

In the case of figure 4 the bandwidth was also varied between 0.1Mbps and 2Mbps and the corresponding delay determined for 0.2 and 0.8 Kbits leaky bucket sizes with mean packet size kept at 500 bits. The results show the same trend with increase in transmission bandwidth as in figure 3 for both CDGPS and PGPS techniques.

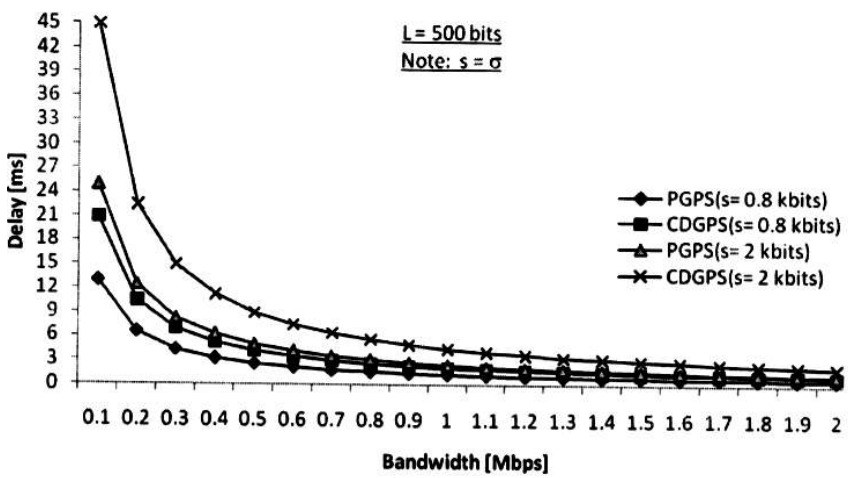

Figure 4: Delay versus bandwidth [ms] [Mbps] for both CDPGS and PGPS

Both figures 3 and 4 show that with increase in transmission bandwidth the exponential decay in the delay is more in PGPS technique than in CDGPS. Therefore, PGPS exhibits better delay performance than CDGPS technique. In order words, results show that for a given packet size, leaky-bucket parameter value and transmission rate, CDGPS technique exhibits greater delay than PGPS technique.

Variation in the leaky-bucket token buffer size parameter values influences the exponential parameter and therefore the rate of delay variation. The result shows that the delay with CDGPS is close to that with PGPS for small token buffer size, but disparity increases with increasing token buffer sizes.

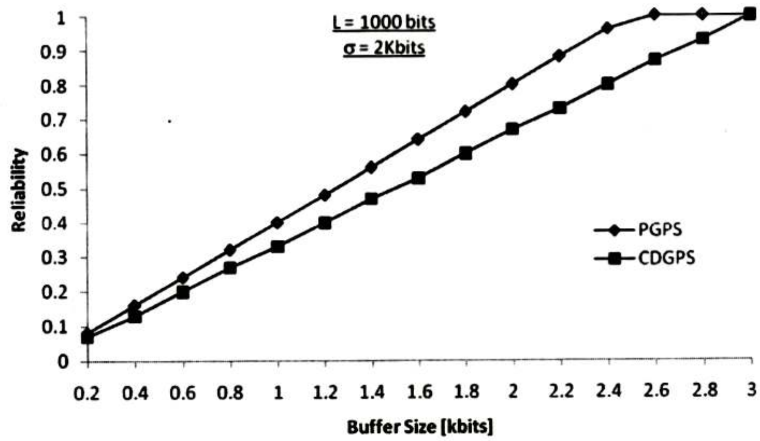

Figure 5: Reliability versus buffer size for both CDPGS and PGPS

In figure 5 the buffer size was varied between 0.2Kbits and 3Kbits and the corresponding reliability values determined for 2Kbits leaky bucket size while mean packet size was kept at 1000 bits. The results show that with increase in transmission buffer size the associated reliability also increases though tends to unity for both CDGPS and PGPS techniques.

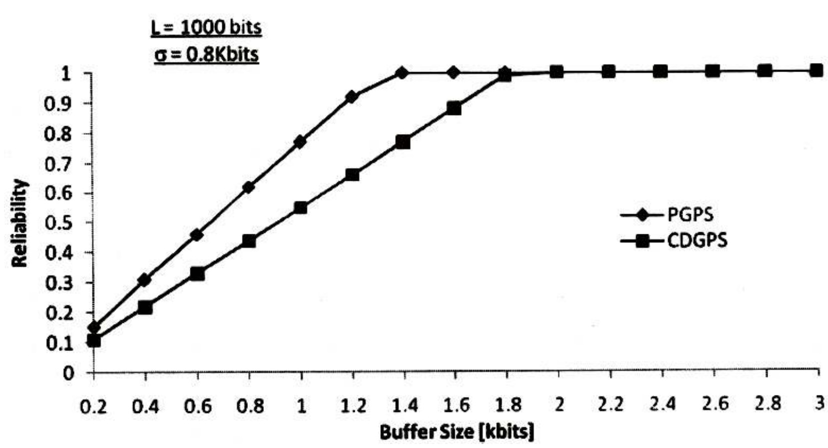

Figure 6: Reliability versus buffer size for both CDPGS and PGPS techniques

Also, in figure 6 the buffer size was varied between 0.2Kbits and 3Kbits and the corresponding reliability values determined for 0.8Kbits leaky bucket size while mean packet size was kept at 1000 bits. The results show that with increase in transmission buffer size the associated reliability increases and also tends to unity for both CDGPS and PGPS techniques, as in figure 5.

Figures 5 and 6 show that for a given packet size, leaky-bucket parameter value and transmission buffer sizes, PGPS technique exhibits greater reliability than that of CDGPS technique.

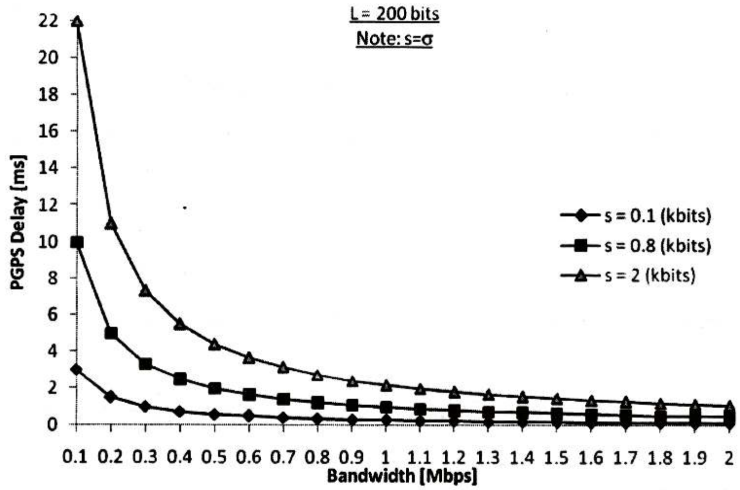

Figure 7: Delay [ms] versus bandwidth [Mbps] for only PGPS technique

As in figures 3 and 4, figures 7-10 have the bandwidth varied between 0.1Mbps and 2Mbps and the corresponding delay determined for 0.1, 0.8 and 2Kbits leaky bucket sizes with mean packet size kept at 200, 500 and 1000 bits. The results show the same trend with increase in transmission bandwidth as in figure 3 and 4 for both CDGPS and PGPS techniques.

Figure 7 has the bandwidth varied between 0.1Mbps and 2Mbps and the corresponding delay determined for 0.1, 0.8 and 2Kbits leaky bucket sizes with mean packet size kept at 200 bits.

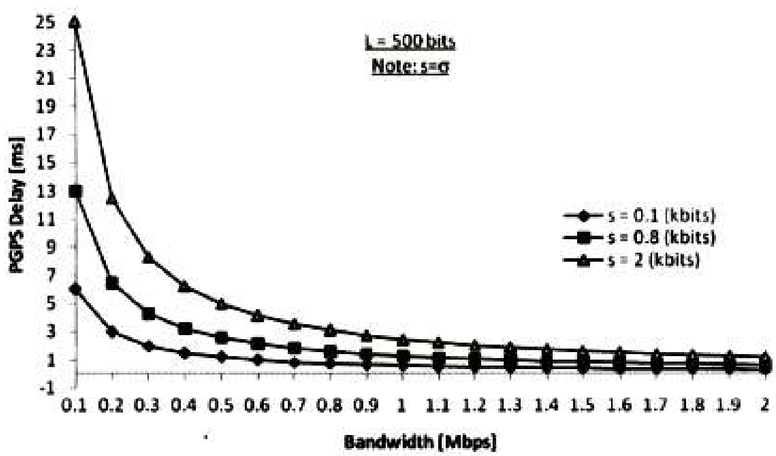

Figure 8: Delay [ms] versus bandwidth [Mbps] for only PGPS technique

Also, figure 8 has the bandwidth varied between 0.1Mbps and 2Mbps and the corresponding delay determined for 0.1, 0.8 and 2Kbits leaky bucket sizes with mean packet size kept at 500 bits.

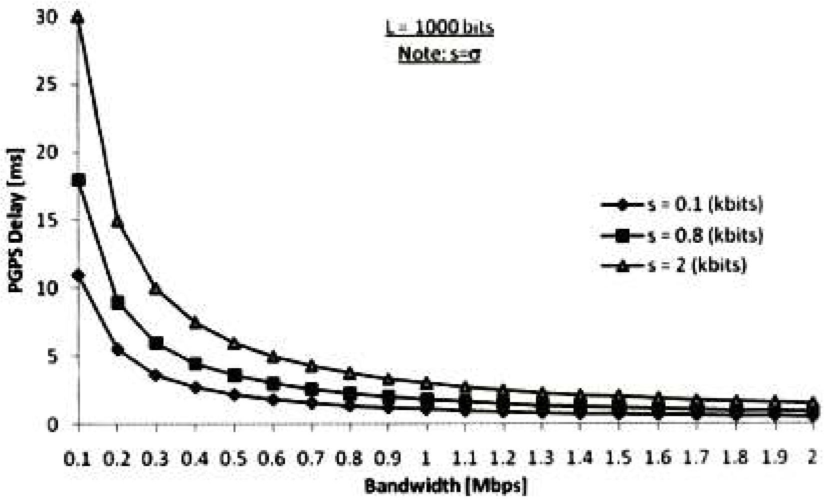

Figure 9: Delay [ms] versus bandwidth [Mbps] for only PGPS technique

Figures 9 has the bandwidth varied between 0.1Mbps and 2Mbps and the corresponding delay determined for 0.1, 0.8 and 2Kbits leaky bucket sizes with mean packet size kept at 1000 bits.

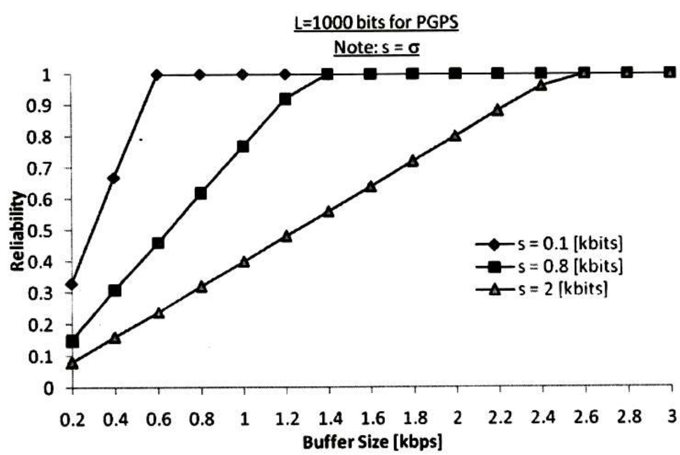

Figure 10: Reliability versus buffer size for only PGPS technique.

As in figure 6, figure 10 has the buffer size varied between 0.2Kbits and 3Kbits and the corresponding reliability determined for 0.1, 0.8 and 2Kbits leaky bucket sizes with mean packet size kept at 1000 bits. The results show the same trend with increase in transmission buffer size as in figure 6 for only PGPS technique. Increase in the buffer size brings about relative increase in reliability for a given leaky bucket sizes parameter value. Reliability decreases with increase in the value of leaky bucket parameter at a constant transmission buffer capacity.

Conclusion

This paper presented the comparison of two GPS techniques - CDGPS and PGPS. The performance of the winning technique, PGPS technique, was investigated under varying traffic load (packet length).

The results of the analysis show that PGPS has better delay and reliability (packet-loss) performance than the CDGPS. Both techniques have the average of 4.0 average deviation in the results for both delay and reliability parameters. Specific observation made on the PGPS technique indicated that it is capable of allocating resources and can use idle network resources to improve network performance. Delay and data losses in a network depend on three factors - traffic, interaction between service disciplines connected in tandem, and the allocated resources.

In this work, the interaction between service disciplines connected in tandem was not considered while comparing the two GPS techniques since the model was based on an isolated node. The implication is that this work considered one-hop routing. It may be equally, important for this aspect to be considered while making a conclusive choice on the appropriate GPS technique.

Acknowledgement

The authors are always grateful to other Networks Research Team members for their ever constructive comments. Authors are also grateful to the University of Nigeria, Nsukka for providing the enabling environment and the Federal Government of Nigeria for endowing the University with Centre of Excellence in Networking.

References

[1] Ani, C.I., Ahmad, R., and Halsall, F., Methodology for the Derivation of Network Resources to Support Video-Related Services in ATM Based Private Wide Area Networks, *IEE Proc. Communication*, Vol. 142, Number 4, 1995.

[2] Parekh, A.K. and Gallager, R.G., A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case, *IEEE/ACM Trans. Networking*, Vol. 1, Number 3, 1993, pp. 344-357.

[3] Parekh, A.K. and Gallager, R.G., A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple-Node Case, *IEEE/ACM Trans. Networking*, Vol. 2, Number 2, 1994, pp. 137-150.

[4] Xu, L., Shen, X., and Mark, J.W., Dynamic Fair Scheduling with QoS Constraints in Multimedia Wideband CDMA Cellular Networks, *IEEE Trans. Wireless Commun.*, Vol. 3, Number 1, 2004, pp. 60-73.

[5] Jiang, H., Zhuang, W., and Shen, X., Cross-Layer Design for Resource Allocation in 3G Wireless Networks and Beyond, *IEEE Commun. Magazine*, Vol.43, Number 12, 2005, pp. 120-126.

[6] Demers, A., Keshav, S., and

Shenker, S., Analysis and Simulation of a Fair Queueing Algorithm," *Internetworking, Res. And Experience.*, Vol. 1, Number 1, 1990, pp. 3-26.

[7] Shafiq, B., *el al*, Wireless Network Resource Management for Web-Based Multimedia Document Services," *IEEE Commun. Magazine,* Vol. 41, Number 3, 2003, pp.138-145.

[8] Xu, L., Shen, X., and Mark, J.W., Dynamic Bandwidth Allocation with Fair Scheduling for WCDMA Systems, *IEEE Wireless Commun., Magazine*, Vol. 9, Number 2, 2002, pp. 26-32.

[9] Bashandy, A.R., Chong, E.K.P. and Ghafoor, A., Generalized Quality-of-Service Routing with Resource Allocation, *IEEE J. Sel. Areas Commun.*, Vol. 23, Number 2, 2005, pp. 450-463.