RESOURCE ALLOCATION METHODOLOGY FOR INTERNET HETEROGENEOUS TRAFFIC

Cosmas I Ani
Department of Electronic Engineering
University of Nigeria Nsukka, Nigeria

ABSTRACT
The mode of operation of internet can lead to congestion which, in turn, leads to degradation in the quality of service (QoS). Congestion can be seen as overflow in the input and/or output buffers of switches at a node. Research issue relating to internet services is determining the optimum network resources - in terms of transmission link bandwidth and buffer capacity in switches - that are required for heterogeneous internet traffic which guarantees a given QoS, even under high network loading conditions.
This paper, therefore, presents a method for determining the optimum internet resources required for heterogeneous (data and voice only) traffic services to guarantee given QoS requirements

INTRODUCTION

Internet is a global network with nodes that transfer frames based on TCP/IP protocol. The mode of operation of internet can lead to congestion which, in turn, leads to degradation in the quality of service (QoS). Congestion can be seen as overflow in the input and/or output buffers of switches at a node. Congestion control, III this case, involves determining the optimum network resources - in terms of link bandwidth (leased lines) and buffer capacity at each node. The most important QoS parameters associated with the buffer queuing process at each node are traffic frame loss rate, traffic frame delay and frame delay variation (jitter). Therefore, the method in which resources are allocated to internet traffic is very vital in providing standard QoS to internet users. If the resources in a network are optimally allocated to support network services, the associated QoS will be optimal and satisfactory.

Statistical multiplexing method has been adopted for the investigation of data traffic (variable bit rate (VBR) traffic) in private networks (corporate networks) since it gives improved levels of utilization for data services II I. The effect of statistical multiplexing, however, is that congestion can occur within the backbone switches if a set of arriving frames relating to different internet accesses all require forwarding on the same output link. To allow for such eventualities, therefore, buffering is provided within the switch. For voice traffic (constant bit rate (CBR) traffic), deterministic multiplexing has been adopted and in this case, congestion can occur if bandwidth is allocated at less than its peak rate [I].

Internet traffic is basically comprised of data and voice frames. Data services are known to be sensitive to traffic frame loss and hence it is necessary to consider traffic frame loss rate parameter when allocating resources for data services; while voice services are known to be sensitive to both frame delay and jitter. In principle, for a particular number of accesses that are loss sensitive, this can be resolved simply by providing a large butler. In practice, however, increasing the buffer capacity in order to resolve the problem of frame loss increases the delay frames experience within the buffer switches. Also. the level of jitter that is introduced is increased. As can be concluded from this, the requirements for frame loss and frame delay/jitter are interrelated and conflicting in terms of buffer capacity provisioning [2]. Research reported to date, however, has treated each parameter and each traffic service separately 13, 4, 5, 6, 7, 8]. The reason for this is that the level of complexity of the analytical processes involved when the parameters are jointly considered is too high and very difficult to track 191. The situation is worse when heterogeneous traffic is considered.

In this paper, therefore, a new method is presented for deriving the amount of buffer storage and transmission link bit rate required to support heterogeneous data and voice services. The derived resources are required to ensure that the maximum frame loss, frame delay and jitter per access - due to buffer overflow - are within defined limits. The new methodology therefore, involves the following two steps:

1. Allocation of resources to separate homogenous traffic to determine its individual resource requirements. This is necessary for the validation of next step, and
2. Allocation of resources heterogeneous internet traffic. The two steps were realized using computer simulation approach based on Designer BONeS (Block Oriented Network Simulator') [9]. The new approach is much less complex.

SIMULATION MODEL

The QoS parameters relating to specific traffic load and transmission rates are obtained by evaluating the performance of the queuing process at a node for a given buffer size. It has been shown that the operation of a network can be satisfactorily modelled and analyzed by considering an isolated node [2, 10, 11]. Computer simulation modelling method is used in this case [10, 11]

Figure I shows the simulation modcl for an internet node. The model. as can be seen, is comprised of a traffic Sources, a first-in-first-out *(FIFO)* transmission buffer queue and associated transmission link.

Each internet data traffic (VBR) source may be represented by a markov two state system- a sequence of frame burst and silence periods [12,13,14] the burst and silence durations are independently random and exponentially distributed. The traffic sources generate frames at the rate of π frames per unit time during each burst period [12,13,14]. The parameter π is express in equation (1)

$$\pi = \frac{1}{\delta} \qquad (1)$$

Where δ is the average frame generation period defined by average frame length. The period ($T_s+\tau$) unit time, shown in Figure 2, is used to define the average rate of frame generation,λ, and is expressed as in equation (2).

$$\lambda = \frac{average\ number\ of\ frames\ ina\ burst}{T_s+\tau} \qquad (2)$$

Hence a Source model that generates frames at peak rate π and mean rate λ can be represented by equation (3).

$$T_s = \tau^*(1/\rho - 1) \qquad (3)$$

Where *p* is the ratio of mean-to-peak rate and is known as the traffic burstiness. This call also be expressed as a fraction of the on time by equation (4).

$$\rho = \frac{\tau}{T_s+\tau} \qquad (4)$$

Voice traffic (CBR) Sources generates frames at the rate of π frames per unit time 112.13,41. The parameter π is expressed as in equation (I). Voice traffic graph is illustrated in figure 3.

Performance parameters - frame loss rate, frame delay and frame jitter - relating to specific traffic loads and transmission bit rates, arc obtained by evaluating the performance of the queuing process at the node for a given buffer size. The statistics collected using the choice computer simulation package include:

number of frames entered into the buffer queue; number: of frames rejected entry into the queue;

i. maximum, minimum and average time frames spend in the queue,
ii. time each frame spends in the queue:
iii. buffer occupancy;
iv. interval between frame arrivals;
v. interval between frame departures.

The frame loss rate, delay and jitter parameters are then calculated on the basis of these statistics. A frame is rejected if the buffer is full while it is delayed if it has to be quelled in the buffer frame delay is calculated using equation (5). Jitter (frame delay variation) is simply the variation in the time frame spent waiting in the queue. Hence the maximum value of jitter is computed employing equation (6).

$$Frame\ Delay_{mean} = [T_j]_{mean} \qquad (5)$$

where T_j is the time spent in the queue by the jth frame.

$$frame\ jitter_{maximum} = [T_{j+1} - T_j]_{maximum} \qquad (6)$$

Frame loss rate is calculated for buffer capacities from zero to the maximum buffer occupancy, K, using equation (7).

$$frame\ loss\ rate = \frac{number\ of\ frame\ rejected}{number\ of\ rames\ through\ queue + number\ of\ frames\ rejected} \qquad (7)$$

The first-in-first-out *(*FIFO) transmission buffer queue is represented as a server While transmission link is the rate at which queue is served.

MODEL SIMULATION RESULTS

The simulation results are shown in figures 4-10 Figures 4 and 5 present the results for homogeneous data traffic; figures 6 - 8 present the results for homogeneous voice traffic; figures 9 and 10 present results for the internet heterogeneous data and voice traffic.

For homogeneous data traffic sources

In this case, however, only frame loss rate is considered since it has been established that data service is only sensitive to this. The relationship between frame loss rate and transmission rate was considered within the transmission bandwidth range defined between the aggregate average and maximum/peak bit rate of the sources (2 Mbps - 10 Mbps, 10 Mbps - 50 Mbps, 20 - 100 Mbps. and 40 Mbps - 200 Mbps for simulations with one, five, ten. and twenty sources respectively).

The results are presented without unnecessary repetitions by presenting only those relating to five sources with 10 Mbps average bit rate and 50 Mbps peak bit rate for demonstration.

For data services, the quantification of frame loss rate involves the determination of frame loss rate as a function of buffer capacity for a given transmission link bit rate and traffic load (number of sources). The set of results that show the relationship between frame loss and buffer capacity at given transmission link bit rate are obtained by simulating the multiplexing process at a switching node.

As can be seen from these results, for a given transmission bandwidth value, increasing the size of the transmission buffer queue, decreases the frame loss rate. The results indicate that transmission bandwidth has significant influence on the frame loss rate. Frame transmission close to the source mean bit rare results in a considerable frame loss even with large buffer storage capacity. The behaviour of the relationship in figure 4 agrees with the findings in 12].

Figure 5 shows the relationship between buffer storage capacity and transmission bandwidth. Figure 5 shows a shaded region (Area A) with the curve indicating its boundary. The shaded area represents the region of operation in which no frame loss is recorded during the simulation. Hence, operation within Area A should satisfy the specification for data traffic frame loss while operation within Area B involves frame loss. The curve was derived from figure 4 and for frame loss rate that is equal to zero. In the same way multiple curves can be defined for varied frame loss rates. Therefore, for any desired frame loss rate and a corresponding buffer size intended for five data traffic sources the precise value of the optimum transmission link bandwidth can be read easily from the graph. The larger the available buffer size, the smaller the transmission bandwidth that would be required for specific QoS value.

For homogeneous voice traffic sources

The nature of voice traffic makes it necessary for the non-statistical method of resource allocation to be adopted for multiplexing. This multiplexing approach requires that peak bit rate be allocated for each source in order to guarantee the required QoS. Clearly it is vital that, irrespective of the method of multiplexing, the specified QoS are guaranteed.

The objective of this allocation methodology is to ensure that the resources allocated for voice traffic using the peak bit rate method is the optimum (quantity of resources that ensures maximum link utilization and guarantees the specified QoS). This is done by determining and analysing the relationship between transmission link bandwidth and QoS parameters. The simulation model used in this case is the same as that used for data traffic sources except that the traffic sources are for voice (see Figure I). Standard CBR voice peak frame rate that is equal to 2.048 Mbps was assumed lor the simulation.

Figure 6 shows the pattern of buffer storage occupancy that was investigated with varying levels of bandwidth and fixed buffer storage size.

. Therefore, Figure 6 shows the relationship between buffer storage occupancy and transmission link bit rate. At a transmission link bit rate less than the peak frame arrival rate (2.048 Mbps), the buffer occupancy increases from zero rapidly. The graph shows that a 2.34%, decrease in the multiplexer queue service rate results in an approximately 1.13E3 [frames] increase in the maximum buffer occupancy. Such a sharp rise in the frame occupancy for a slight decrease in bandwidth indicates that operation below the allocated bandwidth (non- statistically allocated peak rate) results in a long frame queue build-up. Hence, frame loss rate can be excessive especially in the case where the buffer storage capacity is limited. This implies that the bandwidth allocated must not be less than the allocated peak rate.

Figure 7 shows the relationship between frame delay and transmission bit rate. The graph illustrates that delay is directly proportional to queue length and therefore its pattern is quite similar to that in Figure 6. A 2.34% decrease in the queue service rate (transmission link capacity) results in the rise of maximum frame delay of 234 ms that is close to the specified delay <250ms 121. Buffer occupancy increases very rapidly with decrease in transmission link bit rate below voice peak frame bit rate (2.048 Mbps), Frame delay is zero at a link bit rate equal to or more than 2.048 Mbps. The implication is that transmission link bit rate assignment is effective only at rates equal to or more than the frame arrival rate.

Figure 8 shows the relationship between frame jitter and transmission link bit rate. The graph illustrates that frame jitter is within the specified range for the link bit rate range investigated. In this case, the implication is that using only the frame jitter parameter for quantifying voice service performance would have necessitated the use of statistical multiplexing approach that could have resulted in a significant multiplexing gain.

Basically, it has been established that the adoption of the non-statistical method of traffic multiplexing for voice traffic is appropriate since any violation of the principle will result in a substantial degradation in the QoS. Therefore, the resources allocated using the peak rate allocation method are the optimum. For heterogeneous (data and voice) internet traffic sources

Figure 9 shows the curves that may be used for the allocation of transmission bandwidth required for a given number of heterogeneous internet sources and specific buffer storage capacity. Figure 9 is derived using Figures 4-8. Similar curves can be derived for other numbers of traffic sources, frame loss rates and buffer sizes analogously.

It is clear that the curves in figure 9 perfectly agree with resource allocation graphs in figures 4-8. Figure 6 shows that frames may experience delay up to 234ms as against ISO maximum standard delay (250ms) with 1.13E3 maximum buffer occupancy. In figure 9, the maximum buffer size is 30 (frames). Therefore for such size of buffer, frame loss rate and frame delay/jitter are maintained within specified limit.

Figure 10 shows that the allocated bandwidth per source decreases with an increasing number of sources and tends towards the mean bit rate associated with a single source (-2M bps). Basically the graphs for bandwidth per source demonstrate the statistical multiplexing gain obtained II hen multiplexing heterogeneous internet traffic sources together. Figure 10 can be used to determine the number of traffic sources that a particular leased circuit can support at a given buffer storage size without compromising the specified QoS.

CONCLUSION

A novel methodology for allocating resources to heterogeneous internet traffic using all three QoS parameters jointly has been described. It allows the optimum resources (least quantity of resources) that are required for a given QoS and traffic load to be determined. In addition, an allocation factor that can be used to determine the optimum number of internet traffic sources that a particular leased circuit can support without compromising the specified QoS and buffer storage capacity has also been defined.

REFERENCES

1. Decina M., Toniatti T., Vaccari P. and Verri L., "Bandwidth Assignment and Virtual Call Blocking in ATM Networks", *IEEE Proc. INFOCOM* 90, pp 881 - 888.

2. C.1. Ani, Ahmad R. and Fred Halsall, "Methodology for the derivation of network resources to support video-related services in ATM based private wide area networks", *IEE Proc. Communication* vol 142, No.4, August. 1995. [Agnew84]

3. Roch Guerin, Hamid Ahmadi and Mahmoud Naghshineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High Speed Networks" *IEEE Journal on Selected Areas in Communications.* Vol. 9 .No. 7, September 1991 pp.968-981.

4. Sykas D, et. al., "Congestion Control - Effective Bandwidth Allocation in ATM Networks", *High Performance Networking. IV (C-14), IFIP*, Belgium, 1993, pp.65-80.

5. Ibrahim Wahby Habib and Tarek N. Saadawi, "Dynamic Bandwidth Allocation and Access Control of virtual Paths in ATM Broadband Networks", *High Performance Networking*, IV (C-14), IFIP, Belgium, 1993, pp.49-63.

6. Gordon B. Agnew and Jon W. Mark, "Performance Modeling for Communications Networks at a Switching Node", *IEEE Transactions on Communications* vol. Com-32, No.8, August 1984.

7. Bijan Jabbari and Ferit Ycgcnoglu. "An efficient Method for Computing Cell Loss Probability for Heterogeneous Bursty Traffic in ATM Networks", *International Journal of Digital and Analog Communication Systems*, vol.5, 1992. 39-48

8. George Kesidis, Jean Walrand and Cheng-Shang Chang, "Effective Bandwidth for Multiclass Markov Fluids and Other ATM .Sources*". IEEE ACM Transactions on Networking*, Vol. 1, No.4, August 1993.

9. COM DISCO SYSTEMS, "BONcS (Block Oriented Network Simulator) Designer" a business Unit of Cadence Design Systems, Inc., USA, 199

10. C. l. Ani & Fred Halsali, "Simulation Technique for evaluating Cell-Loss Rate In ATM Networks"; *The Society for Computer simulation. SIMULATION Journal* California; Vol. 64, No.5: May 1995; pp. 320-329.

11. Gordon B. Agnew and Jon W. Mark, "Performance Modeling for Communications Networks at a Switching Node", *IEEE Transactions on Communications* vol. Com-32, No.8, August 1984.

12. Roch Guerin, Hamid Ahmadi and Mahmoud Naghshineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High Speed Networks" *IEEE Journal on Selected Areas in Communications*. vol. 9, No.7. September 1991.

13. Yamanaka Naoaki, Youichi Sato, and Ken-ichi Sato, "Usage Parameter Control and Bandwidth Allocation Methods Considering Frame Delay Variation in ATM Networks", *IEICE TRANS. COMMUN.*, Vol. E76-B, No.3 March 1993, pp, 270-279.

14. Anick D., Mitra D and Sondhi M. M., "Stochastic Theory of a Data-Handling System with multiple Sources", *The Bell System Technical Journal*, October 1982, USA, pp 1871 - 1894.

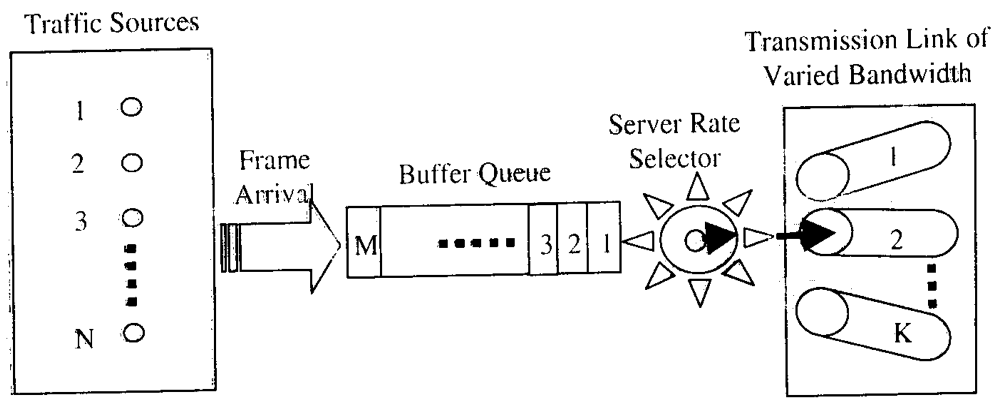

Fig. 1: Model of Internet Node

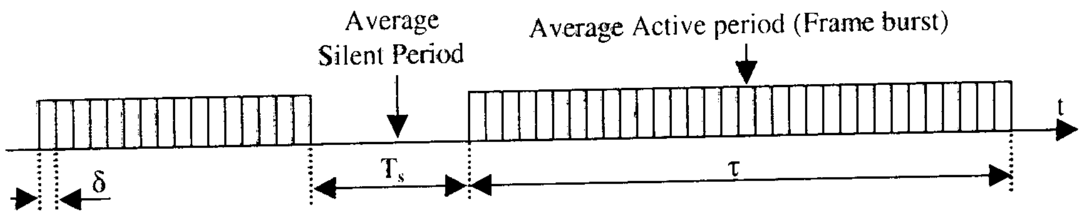

Fig. 2: Data Traffic Time Graph

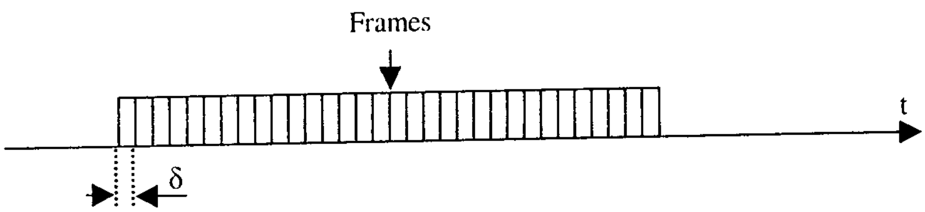

Fig. 3: Voice Traffic Time Graph

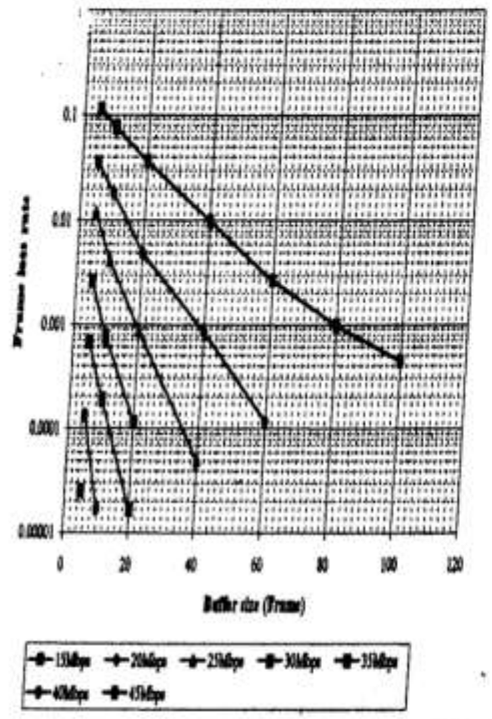

Figure 4: Data frame loss rate as a function of buffer size at various link bit rates for 5-sources

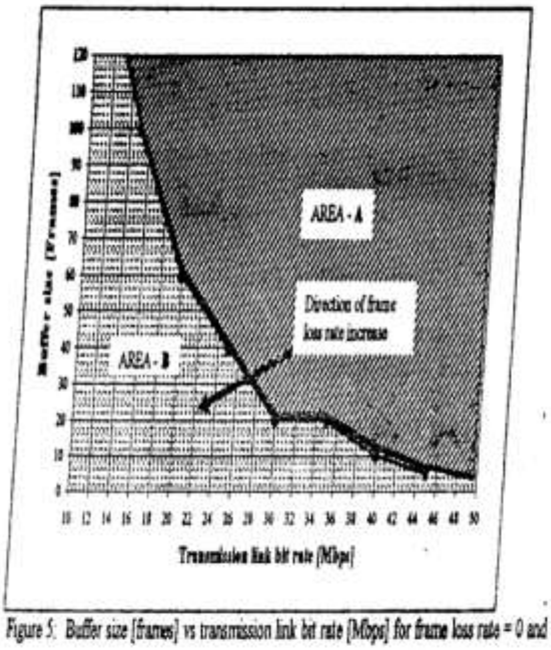

Figure 5: Buffer size [frames] vs transmission link bit rate [Mbps] for frame loss rate = 0 and 5 sources

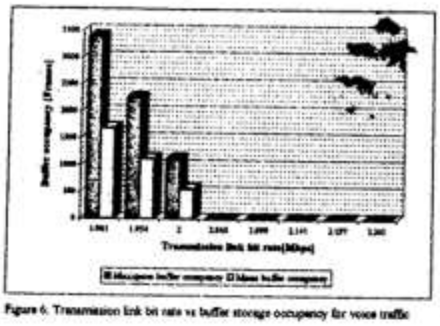

Figure 6: Transmission link bit rate vs buffer storage occupancy for voice traffic

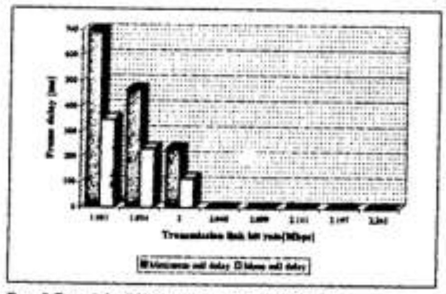

Figure 7: Transmission link bit rate vs frame delay for voice traffic

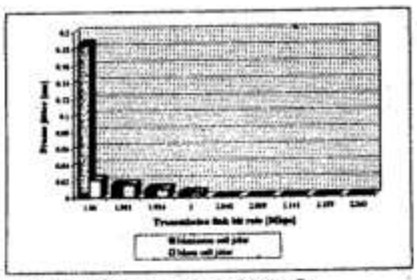

Figure 8: Transmission link bit rate vs frame jitter for voice traffic

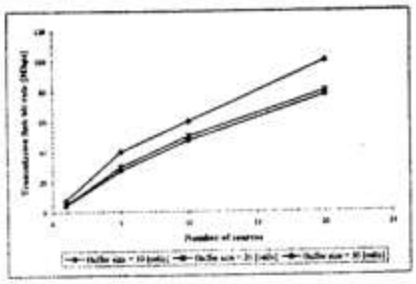

Figure 9: Optimum allocation curves

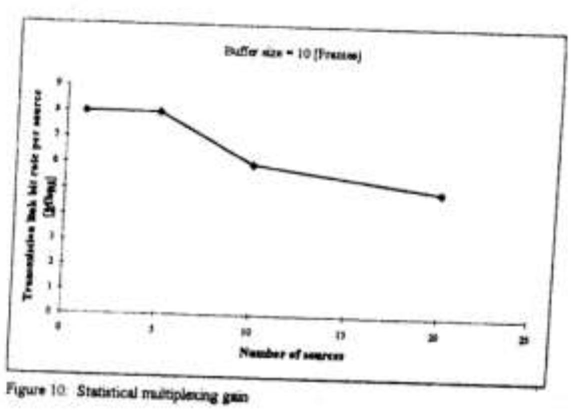

Figure 10: Statistical multiplexing gain