Buffer Allocation for Frame Reassembly and Queueing in ATM Networks

Pierre A. Humblet
Massachusetts Institute of Technology, Cambridge, MA, and Institut Eurecom, France

Whay C. Lee & Michael G. Hluchyj
Motorola Codex, Mansfield, MA

Abstract

Buffer allocation for frame reassembly and queueing in the destination node of an ATM subnetwork supporting multiple frame relay connections is examined. The finite peak rate of a frame relay connection is explicitly modeled and shown to be an important factor in buffer sizing. Specifically, the average number of buffers required is shown to be large when many low-peak-rate frame relay connections are multiplexed for delivery to a single high-speed frame relay interface port, but the tail of the distribution is determined only by aggregate traffic statistics.

1: Introduction

In the network under consideration, frames are generated at multiple sources, with each source having a frame relay interface to a cell-based Asynchronous Transfer Mode (ATM) subnetwork. As they enter the ATM subnetwork, frames for a frame relay connection are segmented into cells for transmission at high speed through the subnetwork via an ATM virtual channel. At a destination edge, the cells from the various frame relay connections are reassembled into frames, then queued at the output port of the frame relay interface, and eventually transmitted to the destination. In keeping with the frame relay service definition, any error recovery is between the end users, and is outside the subnetwork.

This paper is concerned with sizing the buffer pool at the destination port frame relay interface. Because the speed of the output port can be much larger (say 1.5 Mb/s) than the peak rates of the individual frame relay connections (say 9.6 or 56 kb/s), the number of buffers needed for frame reassembly and queueing can be large.

Previous work on buffer allocation for frame reassembly and queueing appears to be scarce. In [1], the authors analyze buffer sizing at a host in an ATM network, using an M/G/1 processor sharing model for the link that feeds cells to the host and a queueing model with finite buffers for the broadband terminal adaptor that collects ATM cells for reassembly. The result in [1] does not apply here, as it does not capture the finiteness of the peak rates of the frame relay connections, nor the fact that the frames are transmitted sequentially from the output queue. Reference [2] gives an elaborate analysis of the output queue with finite buffers, but does not consider reassembly.

In Section 2, we develop and analyze a queueing model that characterizes the reassembly and transmission queues. In Section 3, we evaluate the number of buffers that are required to avoid overflow (or the activation of congestion control) with high probability, for the case where each output port has its own buffer pool.

2: Models for Frame Reassembly and Transmission Queues

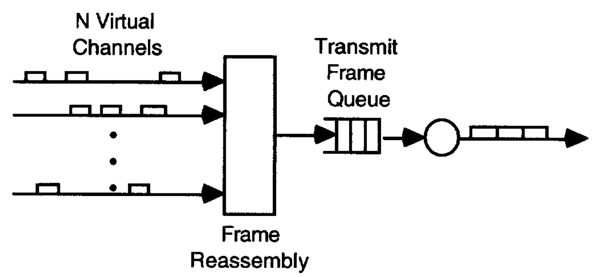

Figure 1: Frame Reassembly and Queueing

2.1: System Description

We characterize the traffic in each virtual channel i to be generated by an ON-OFF source. The idle time of source i is denoted by the random variable I_i, while the frame length is denoted by the random variable G_i (measured in cells), with probability mass function $g_i(\bullet)$. Throughout the paper we assume that all cells are filled. This is an excellent approximation if the cells are much shorter than the frames. The peak rate of source i is C_i cells/s. The source of virtual channel i is thus busy with probability

$$b_i = \frac{E(G_i)}{E(I_i)\,C_i + E(G_i)} \qquad (1)$$

where $E(\bullet)$ denotes expected value.

0743-166X/94 $3.00 © 1994 IEEE

At the destination edge (Figure 1), buffers are drawn from a buffer pool shared by the reassembly and transmit queues, and allocated in the reassembly area on a per cell arrival basis[1]. When an entire frame is buffered, the frame is placed in the transmit queue. The buffers are returned to the buffer pool at once when the frame is eventually transmitted from the transmit queue. This is a common policy.

The following variables are defined for the frame buffer pool sizing:

N: Total number of frame relay connections at the destination port that originating at different source edge ports.

X_i: Number of buffers used for reassembly for virtual channel i.

Y: $\sum_{i=1}^{N} X_i$; Total number of buffers used for frame reassembly.

M: Number of buffers used by the transmit frame queue.

W: Y + M; Total buffers drawn from the buffer pool.

2.2: Reassembly Queue

Assume that the sources are independent and ergodic, and that delay through the ATM network fluctuates little compared to the intercell time $1/C_i$ (thus the network does not "bunch"cells). Then a frame of length k>0 will be in progress on virtual channel i at a random time with probability

$$P_i(k) = \frac{k\, g_i(k)}{E(I_i)C_i + E(G_i)} = b_i \frac{k\, g_i(k)}{E(G_i)} \quad \text{for } k>0. \qquad (2)$$

The probability that no frame is in progress is $P_i(0)$, where

$$P_i(0) = 1 - \sum_{k=1}^{\infty} P_i(k) = 1 - b_i.$$

If no frame is in progress, $X_i = 0$, while if a frame of length k is in progress, X_i will be uniformly distributed between 1 and k. The z-transform of the probability mass function of X_i is thus given by

$$X_i(z) = P_i(0)\, z^0 + \sum_{k=1}^{\infty} P_i(k)\, U_k(z)$$

[1]The theory actually holds if buffers contain multiple cells, as long as the unused space at the end of the last buffer is negligible.

where $U_k(z)$ denotes the z-transform for a uniformly distributed random integer between 1 and k. It follows that

$$X_i(z) = (1-b_i)\, z^0 + b_i \sum_{k=1}^{\infty} \frac{g_i(k)}{E(G_i)} \sum_{j=1}^{k} z^j$$

$$= (1-b_i) + b_i\, z \left\{ \frac{G_i(z) - 1}{(z-1)\, E(G_i)} \right\} \qquad (3)$$

By taking derivatives, one finds that

$$E(X_i) = b_i \left\{ \frac{E(G_i^2) + E(G_i)}{2\, E(G_i)} \right\} \qquad (4)$$

$$E(X_i^2) = b_i \left\{ \frac{2\, E(G_i^3) + 3\, E(G_i^2) + E(G_i)}{6\, E(G_i)} \right\} \qquad (5)$$

Because the reassembly queues are independent (this is a reasonable assumption if the internal network speed is large compared to C_i), the z-transform of the probability mass function of Y is given by

$$Y(z) = \prod_{i=1}^{N} X_i(z) \qquad (6)$$

2.3: Transmission Queue

Following reassembly, frames are queued waiting for transmission, and buffers are freed when a frame has been completely transmitted. Modeling that queue is very difficult, as it is fed from a finite number of sources that have an ON/OFF behavior with general distributions. In addition, the transmission queue is negatively correlated with the frame reassembly queue. To make progress, we model the arrivals of frames in the transmission queue by a Poisson process, independent of the reassembly queue. We expect that model to be conservative, specially when $\sum_{i=1}^{N} C_i < C$, i.e., when the sum of the virtual circuit peak rates cannot keep the output line busy. However the approximation should become quite accurate when the number of sources is large, or when the sources have a very large aggregate peak rate.

The total frame arrival rate is given by

$$\lambda = \sum_{i=1}^{N} \frac{b_i\, C_i}{C\, E(G_i)}, \qquad (7)$$

where C is the capacity of the output link in cells/s and λ is given in frames per cell transmission time. We denote the frame length of the output traffic by the unsubscripted G and its probability mass function by

$$g(k) = \sum_{i=1}^{N} \frac{b_i\, C_i}{\lambda\, C\, E(G_i)}\, g_i(k). \qquad (8)$$

The load on the output link is given by

$$\rho = \lambda\, E(G) = \sum_{i=1}^{N} \frac{b_i\, C_i}{C} < 1 \tag{9}$$

Denote by M the number of cells in the transmit queue at a random time, by k the number of cells in the frame whose transmission is in progress (if any), and by x the time elapsed in the current frame transmission. x is distributed uniformly between 0 and k (in the normalized time). M is related to the waiting time in an M/G/1 queue, but it is not a standard quantity, so a short derivation is in order. We use many standard results from queueing theory. They can be found, for example, in [3]. We have

$$M(z) = E(z^M)$$

$$= (1-\rho) + \sum_{k=1}^{\infty} \int_{x=0}^{k} \rho \frac{k\, g(k)}{E(G)} z^k \frac{1}{k} \exp(\lambda x (G(z)-1))\, H(z)\, dx \tag{10}$$

where

$$H(z) = W(\lambda(1-G(z))).$$

The first term in (10) is the probability that the output line is idle. The second term comprises the following factors: the probability that the output line is busy, the distribution of the number of cells in the frame in progress, the z-transform of the number of cells in that frame, the density of the time x since the frame started, the z-transform of the number of cell arrivals since the frame started, and the z-transform of the number of cell arrivals while the frame was waiting. W(s) denotes the Laplace transform of the density of the waiting time of a frame in the transmit queue, which is an M/G/1 system. It is known [3, p200] that

$$W(s) = \frac{s(1-\rho)}{s-\lambda+\lambda G(\exp(-s))}, \tag{11}$$

as the density of the service time of a frame has Laplace transform G(exp(-s)).

We can perform the average over x, then over k, obtaining

$$M(z) = (1-\rho) + \sum_{k=1}^{\infty} \rho \frac{k\, g(k)}{E(G)} z^k \frac{1}{k} \frac{\exp(\lambda\, k\, (G(z)-1))-1}{\lambda\, (G(z)-1)} H(z)$$

$$= (1-\rho) + \rho \frac{1}{E(G)} \frac{G(z \exp(\lambda(G(z)-1)))-G(z)}{\lambda\, (G(z)-1)} H(z)$$

Using the expression given in (11), with $s = \lambda(1-G(z))$, substituting it for H(z), and then simplifying the resulting expression, we obtain

$$M(z) = (1-\rho) \left\{ \frac{G(z \exp(\lambda(G(z)-1)))-G(\exp(\lambda(G(z)-1)))}{G(z)-G(\exp(\lambda(G(z)-1)))} \right\} \tag{12}$$

This transform has a pole at the largest real solution of

$$z = \exp(\lambda(G(z)-1)) \tag{13}$$

and this pole will determine the tail of the complementary distribution function.

The mean number of cells can be obtained from the first derivative of M(z), evaluated at z=1.

$$E(M) = \frac{\rho\, (2-\rho)}{2(1-\rho)} \left\{ \frac{E(G^2)}{E(G)} \right\} \tag{14}$$

with ρ given in (9).

When all frames have γ cells, $G(z) = z^\gamma$ and the previous formula simplifies to

$$M(z) = (1-\rho) \left\{ \frac{(z^\gamma-1) \exp(\rho(z^\gamma-1))}{z^\gamma - \exp(\rho(z^\gamma-1))} \right\} \tag{15}$$

which is the value at z^γ of the z transform of the probability mass function of the number of customers in an M/D/1 system [3, p. 194], as expected.

3: Buffer Pool Sizing

The total amount of buffers used is given by W = Y + M. It is interesting to first compare the average numbers of buffers Y needed for reassembly and M for transmit queueing. Later we will consider the distributions.

3.1: Average Buffer Requirements

On the average, the total number of buffers used for frame reassembly is

$$E(Y) = \sum_{i=0}^{N} E(X_i) = \sum_{i=1}^{N} b_i \frac{E(G_i^2) + E(G_i)}{2\, E(G_i)} \tag{16}$$

while the average number of buffers used by the transmit frame queue is E(M) as given in (14).

Suppose that all sources are identical (i.e., the subscripts in (16) may be dropped, and C_0 will later denote the common peak rate of each source). Both E(Y) and E(M) are proportional to the ratio of the second moment to the first moment of the frame length. It is striking that E(Y) depends strongly on the probabilities that the virtual channels are active, and thus on the peak rates (see (1)). E(M) on the other hand depends only on the load. To pursue the comparison, let us assume that E(G) is small compared to $E(G^2)$.

We then have

$$E(Y) = N\, b \left\{ \frac{E(G^2)}{2\, E(G)} + \frac{1}{2} \right\} \sim N\, b \frac{E(G^2)}{2\, E(G)} \tag{17}$$

$$E(M) = \rho \frac{2-\rho}{1-\rho} \frac{E(G^2)}{2\, E(G)} \tag{18}$$

where ρ, derived from (9), is simplified as follows.

$$\rho = \lambda\, E(G) = \sum_{i=1}^{N} \frac{b_i\, C_i}{C} = Nb\, \frac{C_0}{C}. \tag{19}$$

One sees that the average number of buffers needed for reassembly exceeds the average amount required for queueing when Nb exceeds $\rho\left(\frac{2-\rho}{1-\rho}\right)$. As indicated by (19), Nb is also given by $\rho\,\frac{C}{C_0}$. Hence, the above condition is equivalent to $\frac{C}{C_0} > \frac{2-\rho}{1-\rho}$. This occurs in many situations of interest. Even for a high load ($\rho = 0.8$), it only takes Nb ~ 5 or $\frac{C}{C_0} = 6$ for the reassembly area to require as much storage as the transmit queue, on average. On the other hand, $Nb < \frac{C}{C_0}$ for stability, thus E(Y) is upperbounded as follows,

$$E(Y) \le \frac{C}{C_0}\left\{\frac{E(G^2) + E(G)}{2\, E(G)}\right\}, \tag{20}$$

independent of the number of virtual circuits.

On the average, the total number of buffers required is upper-bounded as follows.

$$E(W) \le \frac{C}{C_0}\left\{\frac{E(G^2)}{2\, E(G)} + \frac{1}{2}\right\} + \rho\,\frac{2-\rho}{1-\rho}\,\frac{E(G^2)}{2\, E(G)}$$

$$\sim \left\{\frac{C}{C_0} + \rho\,\frac{2-\rho}{1-\rho}\right\}\frac{E(G^2)}{2\, E(G)} \tag{21}$$

It is interesting to note that if the frames were buffered at the source nodes, and if the network had a speed much greater than C_0, then the need for reassembly buffers at the destination would be greatly reduced. Essentially C_0 would be increased to the network speed, and b would be decreased in inverse proportion. The delay suffered by a frame would be little affected by this modified strategy. The traffic internal to the network would become more bursty, increasing the buffering requirement at the intermediate nodes. More buffers would also be required at the source, but they can be dimensioned precisely, as only a single frame would be buffered there.

Now that we have compared the means, let us consider how fluctuations in the frame reassembly process are affected by the source characteristics, without assuming that all sources are identical. We have obtained previously $E(X_i)$ and $E(X_i^2)$ (see (4) and (5) respectively). Thus, $Var(X_i)$ has the form

$$Var(X_i) = b_i\, \alpha - b_i^2\, \beta. \tag{22}$$

(When all sources are identical, the subscripts in (22) may be dropped.) When b_i is small, the variance increases linearly with b_i, starting from 0. It reaches a maximum at

$$b_i = \frac{E(G_i)\,(2\,E(G_i^3) + 3\,E(G_i^2) + E(G_i))}{3\,(E(G_i^2) + E(G_i))^2} \tag{23}$$

(if this quantity is less than 1) and it then decreases somewhat as b_i approaches 1.

If the lower moments of G_i can be neglected compared to the higher ones, one finds that the variance is largest, about $\left\{\frac{E(G_i^3)}{3\,E(G_i^2)}\right\}^2$, at $b_i = \frac{2\,E(G_i^3)\,E(G_i)}{3\,(E(G_i^2))^2}$, and that it decreases to $\frac{E(G_i^3)}{3\,E(G_i)} - \left\{\frac{E(G_i^2)}{2\,E(G_i)}\right\}^2$ as b_i approaches 1.

The precise values depend strongly on the distribution of G_i. If G_i is constant, the slope of the variance at $b_i = 0$ is $\frac{(E(G_i))^2}{3}$, the maximum is $\frac{(E(G_i))^2}{9}$ at $b_i = 2/3$, and the variance decreases to $\frac{(E(G_i))^2}{12}$ when $b_i = 1$. If G_i is geometric, recall that $E(G_i^2) \sim 2\,(E(G_i))^2$ and $E(G_i^3) \sim 6(E(G_i))^3$, so that that the slope of the variance at $b_i = 0$ is $2(E(G_i))^2$, and that the variance increases with b_i to a maximum $(E(G_i))^2$ at $b_i = 1$.

In summary, in the common situation where b_i is not close to 1, the variance of X_i is almost linear in b_i, and it saturates as b_i increases. In either case the ratio of the variance to the square of the mean decreases as b_i increases.

3.2: Distribution of the Buffer Utilization.

By numerical techniques we have inverted the distribution of the buffer usage, first for the reassembly queue and the transmit queue separately, and then for the two together, plotting the complementary distribution functions in Figures 2 and 3.

The results can be used to bound the number of buffers necessary to avoid overflows with high probability, or to evaluate the probability that a congestion control mechanism will need to be invoked, if the network provides one.

For the reassembly of cells from source i, the critical moment is when a frame from the source is complete, and not at a random time. Thus, we have actually inverted the transform

$$G_i(z) \prod_{j=1, j\ne i}^{N} X_j(z)\, M(z)$$

rather than Y(z) M(z), i.e., we have replaced $X_i(z)$ by $G_i(z)$ to take into account that an entire frame from source i is present in the reassembly area.

In all cases considered, there are 30 identical but statistically independent sources, with a frame distribution that can be either deterministic or geometric. The average number of cells per frame is 10. As the results are plotted in terms of average frame size, this number is of marginal importance. The ratio of the speeds C/C_0 is 24. In all figures, we change the load ρ from 0.1 to 0.9 in steps of 0.1 by adjusting the probability b_i that a source is busy.

Turning to Figure 2 and the results for the deterministic frame size, we see that the space needed for reassembly increases with the load, but that its density is only significant in a narrow range. Indeed, at most 30 frames can be in reassembly at any time. To the contrary, the amount of buffers needed for queueing is characterized by an exponential tail (defined in (13))[2], that becomes more and more significant as the load increases[3], although the mean number required for the transmit queue is well below that required for reassembly (5% at $\rho = 0.1$, 25 % at $\rho = 0.8$, 50% at $\rho = 0.9$). The distribution of the total number of buffers has the tail of the transmit queue distribution, but with an offset due to the reassembly queue. For $\rho > 0.6$, the offset appears to be about equal to 12, independent of the load. We note that

$$12\ E(G) \sim \frac{C}{C_0}\left\{\frac{E(G^2) + E(G)}{2\ E(G)}\right\}, \qquad (24)$$

i.e. the value of E(Y) when $\rho = 1$. The fact that the offset appears to be insensitive to the load can be explained by the saturation of Var(Y) discussed earlier.

The situation is quite different for the geometric frame lengths, as shown in Figure 3. Now the reassembly queue distribution also has a tail, but it is given by the tail of the frame length distribution, and it does not depend on the load. Similarly at light loads the tail of the transmit queue is almost independent of the load. Again the total number of buffers in use has the tail of the number in the transmit queue, but shifted. At high load, the shift is about 22, which again is closely related to

$$\frac{C}{C_0}\left\{\frac{E(G^2) + E(G)}{2\ E(G)}\right\}$$

(for geometric distribution, $E(G^2) = 2\ (E(G))^2 - E(G)$).

Results for 100 sources were also obtained, but not displayed here. They are not significantly different from those for 30 sources. However, the model introduced here is probably more accurate in that case.

4: Conclusion

This paper provides a buffer pool model and its sizing analysis. It appears that the frame reassembly function can require a large number of buffers on average. It shifts the distribution of the total number of buffers in use, but the tail of that distribution is determined by the transmit queueing process.

---

[2] It is likely that the tail is overly pessimistic for the number of sources used in this example.

[3] The staircase nature of the curve is due to the deterministic frame length.

Acknowledgement

The authors would like to thank Javier Morales for many valuable suggestions and Ray Wang for his help on earlier versions of the work.

References

[1] D. E. Smith and H. J. Chao, "Buffer Sizing at a Host in an ATM Network," *Proceedings IEEE INFOCOM '92*, pp. 536-543, May 1992.

[2] J.F. Chang and R.F. Chang, "The Behavior of a Finite Queue with batch Poisson Inputs Resulting from Message Packetization and a Synchronous Server," *IEEE Transactions on Communications*, Vol.COM-32, No.12, December 1984.

[3] L. Kleinrock, *Queueing Systems, Volume 1: Theory*, John Wiley & Sons, 1975.

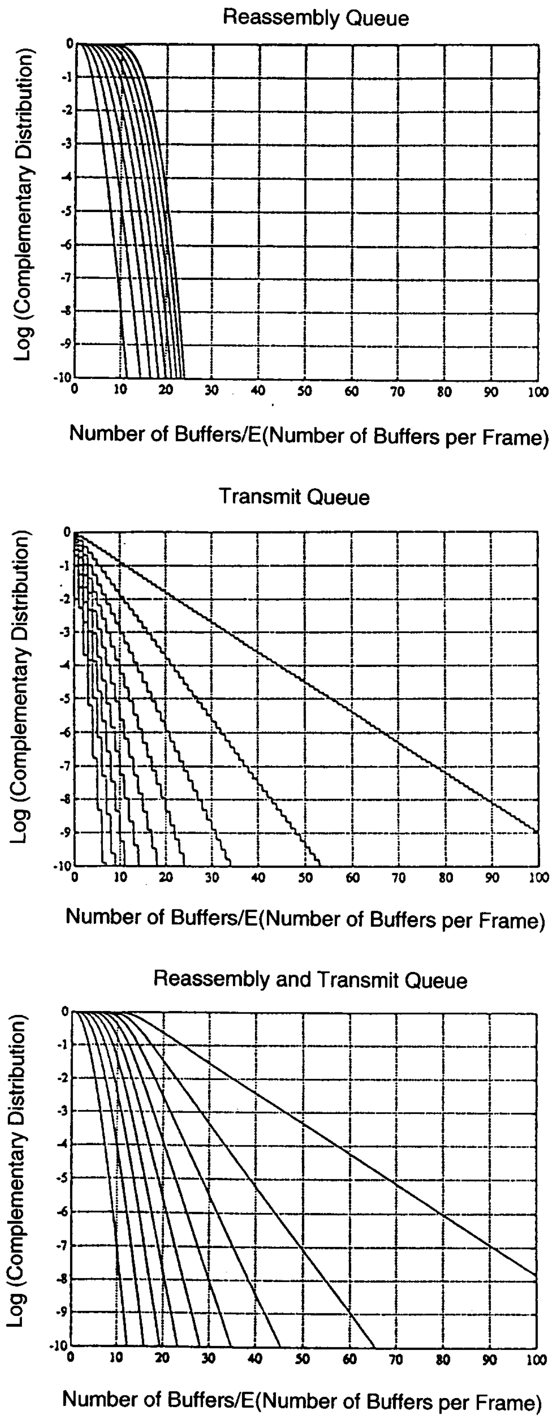

Figure 2: Deterministic Frame Size, ρ varies from 0.1 to 0.9

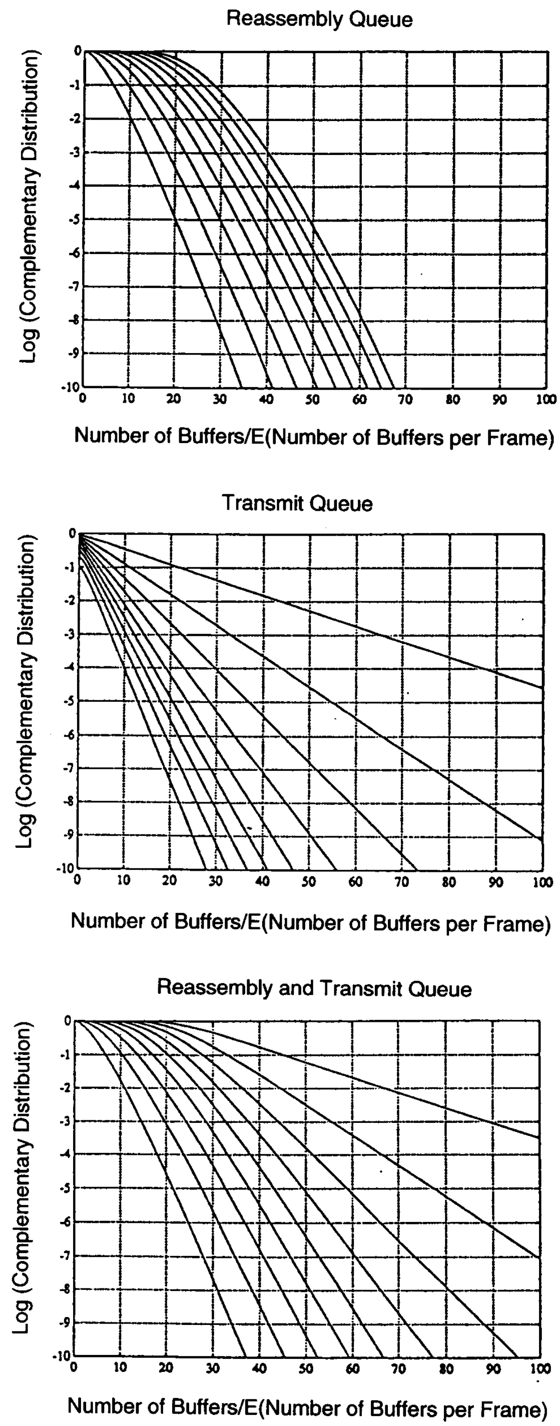

Figure 3: Geometrically Distributed Frame Size, ρ varies from 0.1 to 0.9