THE ECONOMICS OF SEGREGATED AND INTEGRATED SYSTEMS IN DATA
COMMUNICATION WITH GEOMETRICALLY DISTRIBUTED MESSAGE LENGTHS

P. K. Verma and A. M. Rybczynski

Bell Canada Computer Communications

Ottawa, Canada

ABSTRACT

Integrating two different message switching systems into one with the resultant integrated system having the combined channel capacity, need not necessarily result in improved performance. This segregated system vs. integrated system question has been adequately answered in a recent communication for the specific case when the message arrivals are Poisson processes and the message lengths for the two systems are (separately) exponentially distributed.

While the Poisson assumption is probably adequate for message arrivals, recent investigations indicate that in computer communications traffic, a better approximation of the message lengths is provided by the geometric distribution. This paper presents a new algorithm which establishes the relative superiority of the segregated or the integrated system for the geometrically distributed message lengths.

In order to make an equitable comparison, the average delay per message for the segregated systems is computed after making an optimum reallocation of the total available channel capacity. This is done by numerically solving a sixth order equation. The first and the second moments of the combined traffic, which need not necessarily have a geometric distribution, is then computed. Using this statistics and the combined channel capacity, the delay for the integrated system is obtained. A comparison between the performances of the segregated and the integrated systems is then effected.

INTRODUCTION

A computer communication system is characterized by a set of nodes connected to each other by communication links or channels in order to effect the transmission of messages from one node to another. Both the lengths of the messages as well as the time between successive arrivals are random quantities. The resulting stochastic flow of the message traffic gives rise to queues at the nodes and an analysis of this process is an essential part of the successful design of any computer communication system.

The average delay encountered by a message in passing through a communication net is the criterion commonly employed as a measure of its performance.^{1,2} Since the transmission costs can be considered to be most heavily dependent upon the channel capacity required,³ system optimization consists essentially in minimizing the average delay under the constraint of a fixed channel capacity.

Generally speaking, an increase in the size of a message switching system with the utilization factor ρ * held constant, results in improved performance in the sense of obtaining a lower average message delay. However, integrating two different message switching systems into one with the resultant integrated system having the combined channel capacity need not necessarily result in improved performance. This segregated system vs. integrated system question has been the subject of recent investigations^{2,4} and must be carefully analyzed for every computer communication system catering to more than one class of traffic.

Rudin² has obtained appropriate results for the specific case when the message arrivals are Poisson processes and the message lengths for the two systems are (separately) exponentially distributed. da Silva⁴ has introduced priorities in the two classes of users and has established a critical value of message length ratio such that a lesser disparity in message lengths will lead to the integrated system being superior in performance. da Silva's analysis also assumes exponentially distributed message lengths.

The Poisson assumption is probably adequate for message arrivals but recent investigation on computer communication traffic models indicate that a better approximation to message lengths is provided by the geometric distribution.^{5,6} While the mathematical complexity of the geometric distribution defies a closed form solution, it is nevertheless essential to obtain an

*The utilization factor ρ of a link can be defined as the ratio of the average number of bits transmitted per second to the channel capacity of the link in bits per second.

algorithmic solution since a significant difference may exist between the performance criteria derived under the assumption of exponential or geometric distribution. As will be seen later, this difference is particularly accentuated as the message lengths ratio for the two classes of users differs more and more from unity. This investigation therefore, presents an algorithmic approach to answer the segregated vs. integrated question for data communication systems with geometrically distributed message lengths. In the case of segregated systems a method of optimally allocating the available channel capacity so as to minimize the combined average message delay is also furnished. This is in fact essential since an equitable comparison between segregated and integrated systems can be made only if the channel capacity allocation in the former is such that the average delay per message is minimized.

STOCHASTIC DELAYS THROUGH A COMMUNICATIONS NETWORK WITH POISSON ARRIVALS AND EXPONENTIALLY (M/M) OR GEOMETRICALLY DISTRIBUTED (M/G) MESSAGE LENGTHS

Kendall⁷ has shown that for a single channel system with Poisson arrivals and an arbitrary distribution of message lengths, the expected time that a message spends in the system is:

$$T = \frac{1}{\mu C} + \frac{\rho^2 + \lambda^2 \sigma^2}{2\lambda(1-\rho)} \quad (1)$$

where $1/\mu$ = mean message length (No. of Characters)

C = channel capacity (Characters /Sec)

σ^2 = variance of service time (Sec²)

λ = Arrival rate of messages (Sec⁻¹)

and ρ = utilization factor = $\lambda/\mu C$

Assuming an exponential distribution of message lengths, the variance of the service time can be shown to be:

$$\sigma^2 \text{ exp} = \frac{1}{\mu^2 C^2} \quad (2)$$

Using (2) in (1) above, we obtain,

$$T \text{ exp} = \frac{1}{\mu C - \lambda} \quad (3)$$

With geometrically distributed message lengths of minimum length 1 we obtain, after some manipulations:

$$\sigma^2 \text{ geom} = \frac{1}{C^2} \frac{1-\mu}{\mu^2} \quad (4)$$

and

$$T_{\text{geom}} = \frac{1}{2C} \frac{2C-\lambda}{\mu C-\lambda} \quad (5)$$

A comparative study of T_{exp} and T_{geom} is made in Figure 1, using the average message length $\bar{l}$ ($\approx 1/\mu$) as a parameter. The time is computed in terms of character service times. In all cases, the expected time a message spends in the system is greater for exponentially distributed message lengths. In addition, the relative difference between T_{exp} and T_{geom} rapidly increases as the average message length is reduced. This is better illustrated in Figure 2 in which the relative difference between T_{exp} and T_{geom} is plotted as a function of ρ ; once again $\bar{l}$ is the parameter. We now proceed to compare the delay performances of two segregated systems each catering to a different class of users against an integrated system carrying the combined traffic. The constraint in each is a fixed channel capacity C . The message arrivals are assumed to be Poisson processes while the message lengths are geometrically distributed with the minimum message length being equal to unity.

It should be noted in passing that the geometric distribution is a discrete analog of the exponential distribution; both are in fact unique in that they are the only possible "memoryless" distributions in the discrete and the continuous domain respectively. As the mean increases a geometric distribution can be approximated by an exponential distribution of equal mean with relatively less error. This also follows from a comparison of (2) and (3) with (4) and (5) respectively.

FIGURE 1: AVERAGE DELAY PER MESSAGE VERSUS UTILIZATION FACTOR FOR EXPONENTIALLY AND GEOMETRICALLY DISTRIBUTED MESSAGE LENGTHS, PARAMETER $\bar{l}$ ($=1/\mu$)

FIGURE 2: NORMALIZED RELATIVE DIFFERENCE IN AVERAGE DELAYS PER MESSAGE FOR EXPONENTIALLY AND GEOMETRICALLY DISTRIBUTED MESSAGE LENGTHS VERSUS UTILIZATION FACTOR; PARAMETER: $T (= 1/\mu)$

AVERAGE DELAY PER MESSAGE IN TWO SEGREGATED SYSTEMS EACH CATERING TO A DIFFERENT CLASS OF USERS

Let S_1 and S_2 be two segregated systems with the following specifications:

TABLE 1

SYSTEM	AVE. NO. OF MESSAGES PER SEC.	MEAN MESSAGE LENGTH IN CH.	AVE. MESSAGE DELAY WITH AN OPTIMUM CHL. CAP. ALLOCATION
S_1	λ_1	$1/\mu_1$	T_1
S_2	λ_2	$1/\mu_2$	T_2

Before proceeding to examine whether an integrated system, S , with the given channel capacity C (which is the sum of the channel capacities of S_1 and S_2) will have an improved performance, the allocation of C between S_1 and S_2 must be optimized. The combined average message delay, T_{seg} for the segregated systems S_1 and S_2 will have to be computed on this basis, in order to obtain a fair comparison between the performance of segregated and integrated systems.

We have from (5),

$$T_1 = \frac{1}{2C_1} \frac{2C_1 - \lambda_1}{\mu_1 C_1 - \lambda_1} \quad (6)$$

$$T_2 = \frac{1}{2C_2} \frac{2C_2 - \mu_2}{\mu_2 C_2 - \lambda_2} \quad (7)$$

Further, T_{seg} is given by

$$T_{seg} = \frac{\lambda_1}{\lambda_1 + \lambda_2} T_1 + \frac{\lambda_2}{\lambda_2 + \lambda_2} T_2 \quad (8)$$

and substituting (6) and (7) into (8),

$$T_{seg} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \frac{1}{2C_1} \frac{2C_1 - \lambda_1}{\mu_1 C_1 - \lambda_1} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \frac{1}{2C_2} \frac{2C_2 - \lambda_2}{\mu_2 C_2 - \lambda_2} \quad (9)$$

If the total channel capacity C is fixed, the minimum value of T_{seg} as C_1 (or C_2) is varied can be obtained by setting:

$$\frac{dT_{seg}}{dC_1} = 0 \quad (10)$$

(10) can be shown to yield the following equation:

$$\lambda_1 \frac{2\mu_1 C_1^2 - 2\lambda_1 \mu_1 C_1 + \lambda_1^2}{(\mu_1 C_1^2 - C_1 \lambda_1)^2} + \lambda_2 \frac{2\mu_2 (C - C_1)^2 - 2\lambda_2 \mu_2 (C - C_1) + \lambda_2^2}{[\mu_2 (C - C_1)^2 - (C - C_1) \lambda_2]^2} = 0 \quad (11)$$

(11) results in a sixth order equation in C_1 , making a closed form solution impossible. This is unlike the well treated case ¹ of exponential message length distributions where a closed form solution is feasible.

From (11), we have,

$$A_1 C_1^6 + A_2 C_1^5 + A_3 C_1^4 + A_4 C_1^3 + A_5 C_1^2 + A_6 C_1 + A_7 = 0 \quad (12)$$

where

$$\begin{aligned} A_1 &= 2\mu_2/\lambda_2 - 2\mu_1/\lambda_1 \\ A_2 &= -8\mu_2 C/\lambda_2 + 8 - 2\mu_2 \lambda_1/\lambda_2 + 4\mu_1 C/\lambda_1 - 2\mu_1 \lambda_2/\lambda_1 \\ A_3 &= 12\mu_2 C^2/\lambda_2 - 20C + 2\lambda_2/\mu_2 + 8\mu_2 \lambda_1 C/\lambda_2 - 4\lambda_1 + \mu_2 \lambda_1^2/(\mu_1 \lambda_2) - 2\mu_1 C^2/\lambda_1 - 2\lambda_1/\mu_1 + 2\mu_1 \lambda_2 C/\lambda_1 + 4\lambda_2 - \mu_1 \lambda_2^2/(\mu_2 \lambda_1) \\ A_4 &= -8\mu_2 C^3/\lambda_2 + 16C^2 - 4\lambda_2 C/\mu_2 - 12\mu_2 \lambda_1 C^2/\lambda_2 + 12\lambda_1 C - 2\lambda_1 \lambda_2/\mu_2 - 4\mu_2 \lambda_1^2 C/(\mu_1 \lambda_2) + \end{aligned}$$

$$+ 2\lambda_1^2/\mu_1 + 4\lambda_1 C/\mu_1 - 4\lambda_2 C - 2\lambda_1\lambda_2/\mu_1$$

$$+ 2\lambda_2^2/\mu_2$$

$$\begin{aligned} A_5 = & 2\mu_2 C^4/\lambda_2 - 4C^3 + 2\lambda_2 C^2/\mu_2 + 8\mu_2\lambda_1 C^3/\lambda_2 \\ & - 12\lambda_1 C^2 + 4\lambda_1\lambda_2 C/\mu_2 + 6\mu_2\lambda_1^2 C^2/(\mu_1\lambda_2) \\ & - 6\lambda_1^2 C/\mu_1 + \lambda_1^2\lambda_2/(\mu_1\mu_2) - 2\lambda_1 C^2/\mu_1 \\ & + 2\lambda_1\lambda_2 C/\mu_1 - \lambda_1\lambda_2^2/(\mu_1\mu_2) \end{aligned}$$

$$\begin{aligned} A_6 = & 4\lambda_1 C^3 - 2\lambda_1\lambda_2 C^2/\mu_2 - 4\mu_2\lambda_1^2 C^3/(\mu_1\lambda_2) \\ & + 6\lambda_1^2 C^2/\mu_1 - 2\lambda_1^2\lambda_2 C/(\mu_1\mu_2) - \\ & 2\mu_2\lambda_1 C^4/\lambda_2 \end{aligned}$$

$$A_7 = \mu_2\lambda_1^2 C^4/(\mu_1\lambda_2) - 2\lambda_1^2 C^3/\mu_1 + \lambda_1^2\lambda_2 C^2/(\mu_1\mu_2)$$

(12) can be solved numerically giving the optimum value of C_1 and therefore of C_2 . Using these value T_{seg} can be computed from (6), (7) and (8). A simplification of equn. (12) results when $\mu_2/\lambda_2 = \mu_1/\lambda_1$ making it a fifth order equation.

AVERAGE DELAY PER MESSAGE FOR AN INTEGRATED SYSTEM CATERING TO TWO CLASSES OF USERS

Let the integrated system S with channel capacity C transmit both classes of messages as given in Table 1 from one communication centre (node) to another. System S will have the following traffic parameters:

- (i) The arrival process will be a Poisson distribution with the parameter $\lambda = \lambda_1 + \lambda_2$.

- (ii) The mean service time will be given by:

$$\frac{1}{\mu C} = \frac{\lambda_1}{\lambda} \frac{1}{\mu_1 C} + \frac{\lambda_2}{\lambda} \frac{1}{\mu_2 C} \quad (13)$$

- (iii) The variance of the service time can be shown to be given as:

$$\begin{aligned} \sigma^2_{int} = & \frac{1}{C^2} \left[\frac{\lambda_1}{\lambda} \frac{2-\mu_1}{\mu_1^2} + \frac{\lambda_2}{\lambda} \frac{2-\mu_2}{\mu_2^2} - \right. \\ & \left. \left(\frac{\lambda_1}{\lambda\mu_1} + \frac{\lambda_2}{\lambda\mu_2} \right)^2 \right] \quad (14) \end{aligned}$$

The mean delay per message for the integrated system can now be written using equation (1) as:

$$T_{int} = \frac{1}{\mu C} \frac{\lambda/\mu C + \sigma^2_{int} \lambda \mu C}{2(\mu C - \lambda)} \quad (15)$$

Figure 3 illustrates the variation of mean message delays for the segregated and integrated system with $\lambda_1 = 25$, $1/\mu_1 = 20$ and $C = 1200$. The channel utilization is held constant at 0.92 while the ratio μ_1/μ_2 is varied to illustrate the effect of relative message lengths in determining the suitability of a segregated or an integrated system. It can be seen from the curve that for this particular example, the preferred region for the integrated system lies when $.21 < \lambda_1/\lambda_2 < 6.6$. For all other ratios of lengths the segregated system is preferable.

It can be remarked that in general as the disparity between the average lengths increases, the relative performance of the segregated system improves. An interesting special case arises when the two classes of users generate traffic with exactly the same characteristics. While it is obvious that the integrated system will have a better performance in this case, the following theorem establishes a quantitative relationship between the average message delays.

FIGURE 3: COMPARISON OF AVERAGE MESSAGE DELAYS FOR SEGREGATED AND INTEGRATED SYSTEMS AS A FUNCTION OF RATIO OF MESSAGE LENGTHS. ($\lambda_1 = 25$, $\lambda_2 = 20$, $C = 1200$, $\lambda_1/\lambda_2 = 600$)

Theorem: If the traffic generated by the two classes of users has the same mean and variance, then using the same total channel capacity, the average delay per message for the segregated system is twice the average delay for the integrated system.

Proof: We have $\mu_1 = \mu_2 = \mu$ and $\lambda_1 = \lambda_2 = \lambda/2$. For the segregated system, $C_1 = C_2 = C/2$ from (11) and $T_{seg} = T_1$ since $T_1 = T_2$. Let σ_{λ}^2 be the variance of each of the message lengths, then $\sigma_{seg}^2 = 4\sigma_{\lambda}^2/C^2$. T_{seg} can now be obtained from (1) as:

$$T_{seg} = \frac{2}{\mu C} + \frac{\lambda/\mu C + \lambda \mu \sigma_l^2/C}{\mu C - \lambda} \quad (16)$$

Similarly, for the integrated system,

$$\sigma_{int}^2 = \sigma_l^2/C^2$$

and

$$T_{int} = \frac{1}{\mu C} \frac{\lambda/\mu C + \lambda \mu \sigma_l^2/C}{2(\mu C - \lambda)} \quad (17)$$

From (16) and (17) it can be readily seen that,

$$\frac{T_{seg}}{T_{int}} = 2$$

Corollary: As an important conclusion of the above theorem, it can be noted that the average delay per message for the segregated systems can never be more than twice the average delay of the integrated system. On the other hand, depending upon the disparity in the average message lengths, the ratio (average message delay for the integrated system/average delay per message for the segregated system) could be indefinitely large.

The algorithm for computing the average delay per message for the segregated and the integrated system can now be presented in the form of a flow-chart as follows:

CONCLUSIONS

It has been shown that the mean delay per message of geometrically distributed message lengths may differ substantially from that of the exponentially distributed message lengths with the same parameter. Using geometrically distributed messages from two classes of users, an algorithmic solution to determine whether an integrated or a segregated system would be more desirable has been obtained. The method is general enough to accommodate message lengths with arbitrary distribution as long as the first and second moments are known. A flow chart based on the algorithm developed has also been furnished.

The segregated vs. integrated question obviously involves several more considerations than the average delay per message, even though the latter is the most important one. Segregation may be desirable from the point of view of achieving simplicity in hardware or software or in attaining message security between different classes of users with the least effort. Further, commissioning a separate system is essential when a modular system reaches the physical or the economic limit of its modularity. In general it can be said that segregated systems are most cost effective whenever wide variations exist between the requirements of the different classes of users as regards message statistics, response time, data security, error performance, availability, reliability and cost considerations.

ACKNOWLEDGMENT

The authors would like to thank Dr. M. Karnaugh for his comments.

REFERENCES

1. L. Kleinrock, Communication Nets - Stochastic Message Flow and Delay, McGraw Hill Book Co., New York, 1964, Chs. I and II.
2. H. R. Rudin, Jr., On Economics of Scale and Integration of Services in Certain Queued Information Transmission Systems, International Communications Conference, Philadelphia, June 1972, pp. 32-27 to 32-31.
3. Same as Ref. 1, pp. 18.
4. J. S. da Silva, On the Desirability of Integrating a Communication System for Two User Classes, Compcon 73, San Francisco, pp. 113-117.
5. E. Fuchs and P. E. Jackson, Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models, Proc. ACM Symp. Probl. in the Optimization of Data Commun. Syst., (Pine Mountain, Ga., October 13-16, 1969), pp. 201-207.

6. W. W. Chu, Buffer Behavior for Batch Poisson Arrivals and Single Constant Output, IEEE Trans, on Comm. Tech., Vol. COM-18, No. 5, October 1970, pp. 613-618.
7. D. G. Kendall, Some Problems in the Theory of Queues, J. Roy Statist. Soc., Series B, Vol 13, pp. 151-185 (1951).