Response Time Calculation for VSAT Networks

Lung-Sing Liang[1] and Jin-Fu Chang[2]

Electrical Engineering Department, National Taiwan University

Taipei, Taiwan 10764

Abstract

This paper calculates the response time of a VSAT network. The response time consists of network access time from VSAT terminals to the hub station, query waiting and processing time, plus response queueing and transmission delay. In addition to mathematical derivations, numerical examples are given to demonstrate several interesting phenomena. Computer simulations are also provided to verify the validity of the derived results.

I. Introduction

Very Small Aperture Terminal(VSAT) systems have received great attention recently due to their small size and market potential.

There have been extensive discussions on VSAT systems in the open literature. [1]–[12] VSAT system has many applications. It can be used to carry data, voice and video signals. It can also be used for the purpose of information retrieval. In this paper we concentrate on the application of retrieving information from a database installed in the hub station. Sketched in Fig. 1 is the system configuration of a VSAT network. The system provides two pairs of channels: one to be used by the VSAT terminals to access the central hub, the other to be used by the central hub to return the response to the requesting terminal. For conveniences as sketched in Fig. 1 these four channels are named uplink, downlink, forward link, and return link.

Any VSAT terminal wishing to access the database in the central hub should first transmit a request or query on the uplink channel following an appropriate channel access protocol. The access protocol could be a contention–type protocol such as random accessing. It could also be a conflict–free protocol such as TDMA.

The satellite in Fig. 1 can play a role as simple as a repeater. In other words the satellite needs only to repeat whatever is received on the uplink onto the forward link in broadcast mode. The broadcast on the forward link serves an additional purpose of acknowledging the query sent from any VSAT terminal. To be more precise when protocol such as slotted ALOHA is used, a terminal must also monitor transmissions on the downlink to tell whether its previous request has been destroyed by collision.

It is also feasible to assume that the hub station is able to determine whether a query received on the downlink can be processed or not through the use of proper error detection codes. A collided query definitely can not be processed. If processible the query is first placed into the inbound queue in the hub station waiting to be processed. After the query is processed, the generated response is first placed in the outbound queue of the hub station waiting to be transmitted on the return link and rebroadcast by the satellite on the downlink to all the VSAT terminals so that the response can be received by its destined user.

Evaluation of various channel access protocols was previously done in [2]. These protocols are evaluated on the basis of their capacity, delay and stability. However a more meaningful measure of the performance is the so called response time which is measured from the instant that a query is transmitted from a VSAT terminal for the first time to the time that the response returns. More specifically, the response time consists of the channel access delay for VSAT terminals to access the central hub, query waiting time in the inbound queue, query processing time, response waiting time in the outbound queue, response transmission time and response propagation time from central hub to VSAT terminals via the satellite. The purpose of this paper is to derive the response time of the VSAT network depicted in Fig. 1.

II. Mathematical Derivation

In this paper we assume that the channel times of the four links in Fig. 1 are all slotted so that each slot equals the transmission time of a packet. For convenience we shall momentarily assume that all the four links have the same bandwidth. The effect of allocating different bandwidths to different links will be later discussed in Section III when we present numerical examples.

Let TR denote the response time. We have mentioned previously that TR contains the following components: TA— the channel access time for a VSAT terminal to access the central hub; TI— the waiting time of a query in the inbound queue; TP— the query processing time; TO— the waiting time of the response in the outbound queue of the hub, and TB the transmission delay for the response to travel from the hub back to the VSAT terminal via the satellite. In other words,

(1) $TR=TA+TI+TP+TO+TB$

TA is measured from the instant that a query is transmitted from a VSAT terminal for the first time till it successfully arrives at the input of the central hub. Clearly TA is protocol dependent. Many channel access protocols have been proposed and the corresponding TA can be found in the open literature, e.g.[2]. TB is measured from the time that a response commences its transmission at the output of the hub till it is accepted by the destined VSAT terminal. Let B be the random variable representing the length of a response message expressed in number of packets or slots. Clearly $TB=B+R$ with R denoting the round–trip propagation time (expressed also in number of slots) between the

---

[1] Also with Telecommunication Laboratories, Directorate General of Telecommunications, Ministry of Communications, Chung–Li, Taiwan.

[2] Also with the Institute of Information Sciences,Academia Sinica, Taipei, Taiwan.

CH2831-6/90/0000-1184 $1.00 © 1990 IEEE

satellite and the earth. Clearly our main job in this paper is to derive the expected value of TI and TO.

Sketched in Fig. 2 is the queueing model used to derive the expectation TI and TO of (1). As shown in Fig. 2 the inputs to the inbound queue of the hub are the queries arriving from the forward link. These query arrivals are represented by upward arrows with their heights proportional to their required processing time. Each of these arrivals is first placed in the inbound queue, should the processor be not available at the time of arrival. Each query consumes TP seconds from the processor to generate a response of B packets. The responses are also represented by upward arrows with their heights proportional to their lengths. They are again buffered in the outbound queue and sent by the output transmitter to the satellite via the return link in the manner of first come first served. Since the queries are sent from the satellite in synchronous mode, they appear only at slot boundaries on the input axis to the inbound queue. For simplicity in this paper we also assume that each query consumes an integer number of slots of service time from the processor. Therefore the response messages in Fig. 2 also emerge at slot boundaries. The input process to the inbound queue in Fig. 2 as a matter of fact can be characterized by a Bernoulli process with parameter σ due to our use of only a single uplink channel. The value of σ of course depends on the protocol that the VSAT terminals use to access the channel. For example if slotted ALOHA is adopted and if the number of VSAT terminals is so large that the query generation process can be modeled as Poisson then $\sigma = Ge^{-G}$ where G is the mean rate of the Poisson process.

Let us first consider the case in which TP is a fixed value. Without loss of generality we may assume that each slot is of length one second. Since we have assumed previously that TP equals an integral number of slots, the possible values that TP may take are 0, 1, 2,...

In deriving $\overline{\mathrm{TI}}$, the expectation of TI, first denote N_j to be the number of queries left behind by the completion of the processing of the jth query and A_j be the number of queries which arrive within the service time of the jth query. Clearly,

$$(2)\quad N_{j+1}=(N_j-1)^{+}+A_j$$

where $(a)^{+}$ is defined to be a if $a>0$ and 0 if $a\leq 0$. Base on (2) we obtain

$$(3)\quad N(z)=\frac{N_0(z-1)A(z)}{z-A(z)}$$

where N(z) and A(z) are the PGF of N_j and A_j, respectively and $N_0=N(0)$. Setting $N(1)=1$ we obtain $N_0=1-A'(1)$. Since queries arrive only at the slot boundaries, A(z) of (3) can be expressed as follows

$$(4)\quad A(z)=[(1-\sigma)+\sigma z]^{TP}$$

Finally $\overline{\mathrm{TI}}$ can be obtained as follows via the application of Little's formula

$$(5)\quad \overline{TI}=N'(1)/\sigma =\frac{TP(2-\sigma-TP\sigma)}{2(1-TP\sigma)}$$

In order to obtain $\overline{\mathrm{TO}}$ we need to first characterize the output process of the inbound queue, or equivalently, the input process to the outbound queue. Let I be the random variable representing the length of idle period (expressed in number of slots) on the time axis at the input of the inbound queue. Clearly

$$(6)\quad \Pr[I=i]=(1-\sigma)^{i-1}\sigma,\ i=1,2,\ldots$$

and

$$(7)\quad I(z)=\frac{\sigma z}{1-(1-\sigma)z}$$

Let X be the random variable denoting the interdeparture time of the output process of the inbound queue or the interarrival time of the input process of the outbound queue. Then

$$(8)\quad X(z)=(1-N_0)S(z)+N_0S(z)I(z)$$

where S(z) denotes the PGF of the query processing time, i.e., $S(z)=z^{TP}$. In (8), $(1-N_0)S(z)$ is due to the fact that upon the completion of a query if the inbound queue remains nonempty the interdeparture time is of course TP slots. However if the completion of the processing of the query leaves the inbound queue empty then the interdeparture time is I+TP slots.

With the input process to the outbound queue characterized in (8) we may model the outbound queue as an G/G/1. Discrete time queueing systems are discussed in [13]–[15]. In particular, use the technique developed in [14]. We now proceed to obtain $\overline{\mathrm{TO}}$. First let us summarize Koinheim's method as follows. Let X(z) and B(z) be the PGF of interarrival time and service time, respectively. Define

$$(9)\quad Y(z)=\frac{1-X(1/z)B(z)}{1-z}$$

Write $Y(z)=Y^{+}(z)Y^{-}(z)$ where $Y^{+}(z)$ has roots outside the unit disc $|z|=1$ and $Y^{-}(z)$ has roots within the unit disc. Furthermore $Y^{+}(z)$ and $Y^{-}(z)$ are given as follows

$$(10)\quad Y^{+}(z)=c_1\prod_i(z-\eta_i^{+})^{r_i^{+}}\ \text{with}\ |\eta_i^{+}|>1$$

and

$$(11)\quad Y^{-}(z)=c_2z^{-c_3}\prod_i(z-\eta_i^{-})^{r_i^{-}}\ \text{with}\ |\eta_i^{-}|<1$$

in which c_1 can be determined from $Y^{+}(1)=1$. Finally W(z), PGF of the waiting time can be obtained as follows

$$(12)\quad W(z)=\frac{1}{Y^{+}(z)}$$

III. Numerical Examples and Discussions

In the following figures, all the numerical calculations are verified by computer simulations. The curves are traced from numerical calculations using our derived results. The simulated results are marked by blackened dots.

Fig. 3 shows the average waiting time in the inbound queue. In Fig. 3 $\overline{TI}$ is plotted against σ. Four values of TP are considered. In Fig. 3 we observe that under fixed σ, larger TP yields higher waiting time. This result is of course reasonable. Furthermore we observe that when TP=1, the waiting time is in fact zero which is also intuitively correct. Finally, for the average waiting time to be finite we require $(\sigma)(TP) < 1$. This is supported by the way our curves in Fig. 3 approach infinity.

Fig. 4 shows the average waiting time in the outbound queue. In Fig. 4 $\overline{TO}$ is plotted versus p. Five values of σ are considered in Fig. 4. For each value of σ, $\overline{TO}$ decreases as p increases. This trend is reasonable since larger p implies that response messages have shorter average length. Furthermore larger σ gives higher waiting time which is also reasonable since larger σ means heavier load. For the outbound queue to have steady–state performance we require $\sigma/p < 1$ or equivalently $\sigma < p$. This is supported by the curves in Fig. 4.

Fig. 5 presents another view of the waiting time in the outbound queue. In Fig. 5 $\overline{TO}$ is now plotted as a function of σ using TP as parameter. The values of p in Fig. 5(a) and (b) are 0.4 and 0.5, respectively. Quite different from Fig. 4, here we observe that larger TP now creates shorter waiting time. This is still explainable since under fixed p larger TP somehow yields longer interarrival time which in turn reduces the amount of time that a response has to await to be transmitted.

Fig. 6 shows the average response time. In this figure we assume that each of the uplink, downlink, forward, and return link has the same bandwidth BW. In Fig. 6(a) we assume BW=100kbps while in Fig. 6(b) BW=10kbps. If each packet contains 1024 bits and the round–trip signal propagation delay R=270ms then in Fig. 6(a) each slot is of size τ=10.24ms and R= 27slots while in Fig. 6(b) each τ=102.4ms and R=3 slots. For the network considered in Fig. 6(a) we further assume that the hub processor needs two slots to process a query, i.e. TP=2(slots)=20.48(msec). Since the networks considered in Figs 6(a) and (b) differ only in BW, the hub processor of the network using BW=10kbps needs the same 20.48ms, which is 0.2 slots, to process a query, i.e. TP=0.2. For simplicity we approximate the TP of Fig. 6(b) by TP=0. Finally we assume in both figures that each response message is geometric in length with parameter p=0.4, i.e. 2.5 packets on the average.

Although in Fig. 6 our main interest is the response time which is the end–to–end delay from and back to a VSAT terminal, we also plot several other curves in Fig. 6. In Fig. 6 we also plot the average inbound waiting time, the average outbound waiting time, and the average hub delay. These three curves are labeled by $\overline{TI}$, $\overline{TO}$ and $\overline{TH}$, respectively.The average hub delay TH now is defined to be $\overline{TI}+\overline{TO}+1/p$, where $1/p$ is the average transmission time of the response message which has geometric distribution in length.

In order to calculate the response time we need to further specify the protocol used by the VSAT terminals to access the hub via the uplink channel. Two popular channel access protocols are considered in Fig. 6, S–ALOHA and TDMA. When using S–ALOHA the average delay for a VSAT terminal to successfully access the hub can be approximated by R+1+E(R+1+(K–1)/2) for a network which has sufficiently large number of users [16]. In the above E is the average number of times that a packet has to be retransmitted which is $e^{G}-1$ and (K–1)/2 is the average rescheduling delay of a retransmitted packet. The values of K used in Figs 6(a) and (b) are 9 and 4 respectively. When using TDMA the average delay can be obtained from [17] with modifications since our arrivals are restricted to occur only at slot boundaries. The average access delay using TDMA has been derived to be $\{(1-N\sigma)[N+1+(N^2-1)\sigma/3-N(N-1)\sigma] + N(N-1)\sigma[1-N\sigma+(N+1)\sigma/2]\}/\{2[1-(N-1)\sigma/2](1-N\sigma)\}$ where N denotes the number of VSAT terminals so that each TDMA frame contains N slots. In Fig. 6 we use N=30 so that the network can be viewed as a network which contains sufficiently large number of users and the results can be compared with S–ALOHA on a somehow fair base.

In Fig. 6 we also plot the average of the two access delays described above. These two curves are respectively labeled by 1 and 2 for TDMA and S–ALOHA and both verified by computer simulations. Of course in Fig. 6 we finally plot the curve of the average response time when either TDMA or S–ALOHA is used in accessing the hub. These two curves are labeled by 3 and 4 for TDMA and S–ALOHA, respectively.

From Fig. 6 we observe the followings.

1. As reported in earlier literature S–ALOHA performs better than TDMA in access delay at low input rate while TDMA does better from medium to high input rate, this phenomenon is again observed in Fig. 6.

2. Compare Figs. 6(a) and (b) we observe that for a network with larger BW such as that of Fig. 6(a) the access delay seems to play a dominant role in the overall response time. This can be explained. In a network with larger bandwidth, each response message gets shorter transmission time on the return link from hub to satellite. Therefore the bottleneck is in accessing the hub from the VSAT terminals. In other words, if a network with reasonable large bandwidth can be given, then the selection of access protocol becomes crucial.

3. The reason that $\overline{TI}$=0 in Fig. 6(b) is due to our approximation that TP=0. Furthermore in Fig. 6(b) we observe that in the range of σ which gives finite hub delay, i.e. $\sigma(1/p)<1$ or equivalently $\sigma<0.4$ for p=0.4, S–ALOHA seems to perform better than TDMA in overall response time. This is again reasonable since $\sigma<0.4$ is considered low.

Suppose now we can work with a system which can have different bandwidths on different links. Nevertheless we still assume that uplink and forward link have the same bandwidth since both of them carry the same traffics. Similarly we assume that both the return and downlink have the same bandwidth. Let α be the ratio of the bandwidth of the return link to that of the forward link. In Fig. 7 we demonstrate the effect of α on the average outbound waiting time $\overline{TO}$ and observe that as α gets bigger $\overline{TO}$ becomes smaller. This is again reasonable since larger bandwidth on the return link means shorter message transmission time.

ACKNOWLEDGEMENT

The study reported in this paper is supported by the National Science Council of the Republic of China under Grant NSC 79–0404–E–002–49.

References

[1] D. Chakraborty, "VSAT Communications Networks —An Overview, " IEEE Communications Magazine, vol. 26, pp.10–24, May 1988.

[2] D. Raychaudhuri and K. Joseph, "Channel Access Protocol for Ku–band VSAT Networks: A Comparative Evaluation, " IEEE Communications Magazine, vol. 26, pp. 34–44, May 1988.

[3] D. M. Chitre and J. S. McCoskey, "VSAT Networks: Architectures, Protocols, and Management, " IEEE Communications Magazine, vol. 26, pp. 28–38, July, 1988.

[4] F. M. Naderi and W. W. Wu, "Advanced Satellite Concepts for Future Generation VSAT Networks, " IEEE Communications Magazine, vol. 26, pp. 13–22, September 1988.

[5] K.M. Sundara Murthy and K. G. Gordon, "VSAT Networking Concepts and New Applications Development, " IEEE Communications Magazine, vol. 27, pp. 43–49, May 1989.

[6] K.M. Sundara Murthy et al., "VSAT User Network Examples, " IEEE Communications Magazine, vol. 27, pp. 50–57, May 1989.

[7] L. S. Golding, "VSATs: Expert Views on Future Trends, " IEEE Communications Magazine, vol. 27 pp. 58–64, May 1989.

[8] D. Raychaudhuri, "Ku–band Satellite Data Networks Using Very Small Aperture Terminals —Part 1: Multi–access Protocols, " Int. Journal on Satellite Communications, pp. 195–212, June 1987.

[9] D. Raychaudhuri, "Ku–band Satellite Data Networks Using Very Small Aperture Terminals —Part 2: System Design, " Int. Journal on Satellite Communications, pp. 265–278, Sep. 1987.

[10] T. T. Ha, "Personal Computer Communication via VSAT Networks," IEEE J. Selected Areas in Communications, vol. SAC–7, pp. 235–245, February 1989.

[11] T. T. Ha, "ALOHA VSAT Networks with Data Link Controls," Int. Conf. Communications, pp. 798–801, Philadelphia, PA, June 1988.

[12] A. L. McBride, "Application of Renewal Theory to Unslotted ALOHA in VSAT Networks, "IEEE Trans. on Communications, vol. COM–36, pp. 617–621, May 1988.

[13] H. Kobayashi and A. G. Konheim, "Queueing Models for Computer Communications System Analysis," IEEE Transactions on Communications, vol. COM–25, pp. 2–28, Jan. 1979.

[14] A. G. Koiheim, "An Elementary Solution of the Queueing System G/G/1," SIAM J. Computing, vol. 4, no. 4 , pp. 540–545, Dec. 1975.

[15] W . V. Grassman, "Numerical Solutions of the Waiting Time Distribution and Idle Time Distribution of the Arithmetic GI/G/1 Queue, " Operations Research, vol. 37, pp. 141–150 , Jan–Feb 1989.

[16] M. Schwartz, Computer Communication Network :Design and Analysis. Englewood Cliffs, N. J. : Prentice Hall, 1977.

[17] S. S. Lam, "Delay Analysis of a Time Division Multiple Access (TDMA) Channel, " IEEE Transactions on Communications, vol. COM–25, pp. 1489–1494, December, 1977.

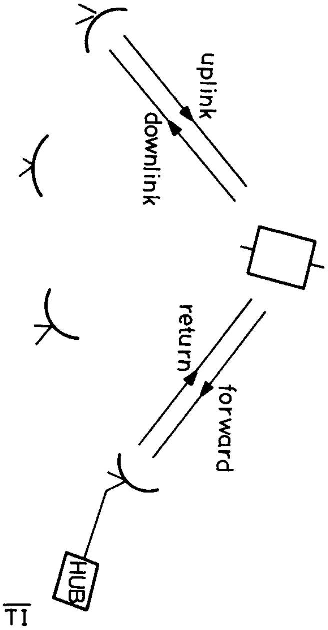

Fig. 1

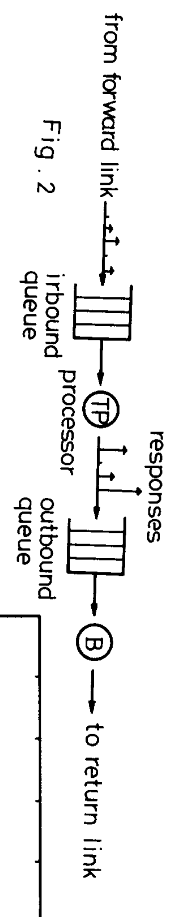

Fig. 2

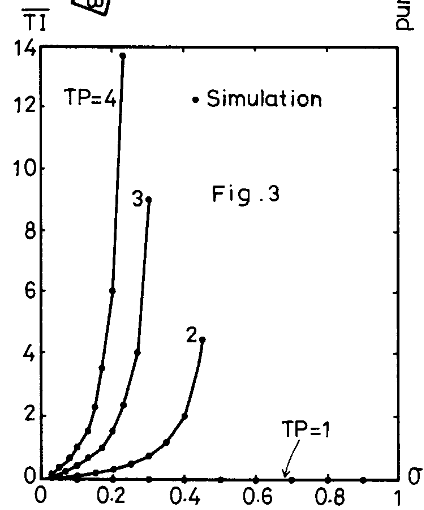

Fig .3

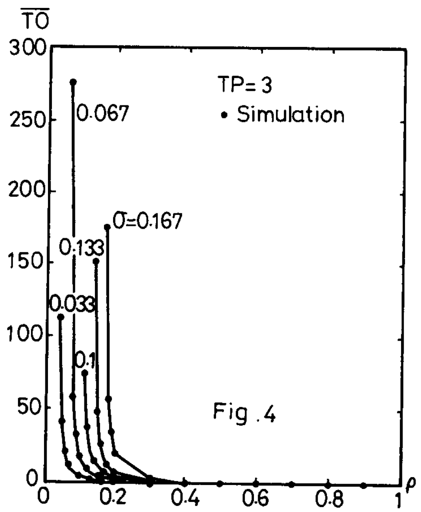
Fig. 4

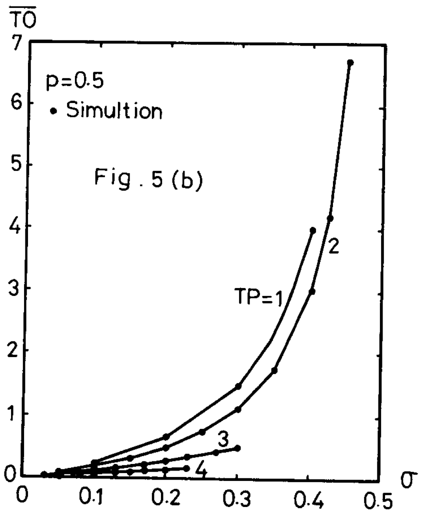
Fig. 5 (b)

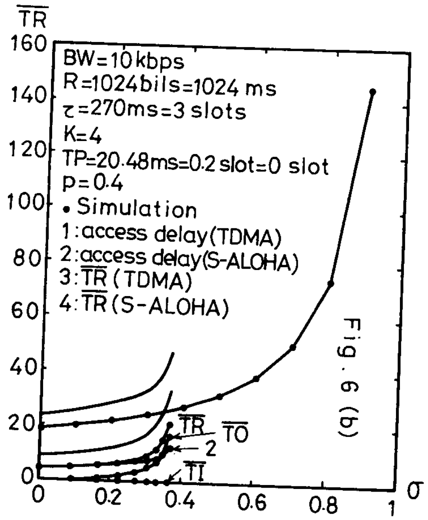
Fig. 6 (b)

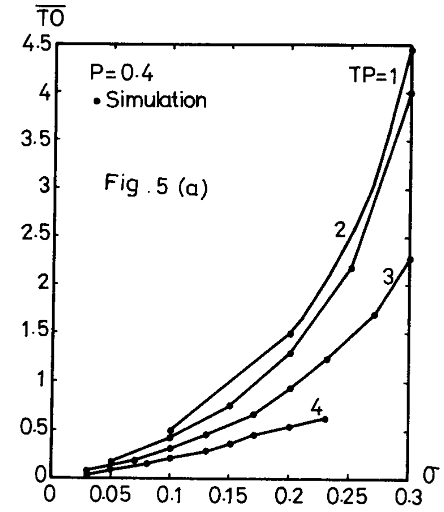
Fig. 5 (a)

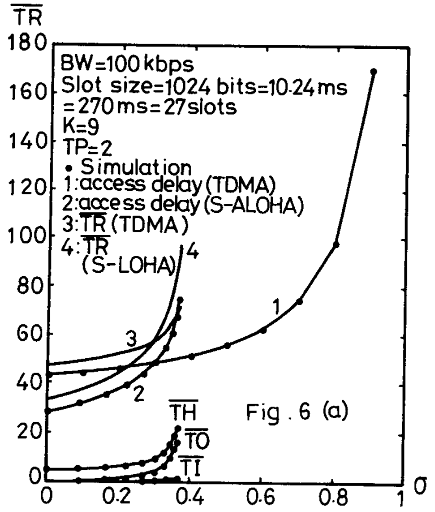
Fig. 6 (a)

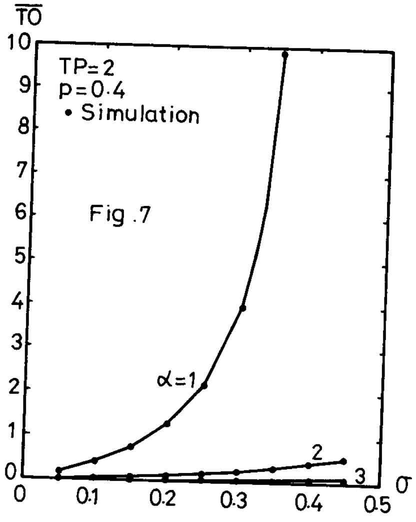
Fig. 7