ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1991

Seria: INFORMATYKA z. 15 Nr kol. 1044

Zdzisław SZCZERBIŃSKI

Zakład Systemów Automatyki Kompleksowej PAN
ul. Bałtycka 5, 44-100 GLIWICE

SELECTED PROBLEMS IN PERFORMANCE EVALUATION OF REGISTER INSERTION, RING SHAPED LOCAL AREA NETWORKS

Summary. This paper discusses some performance issues characterizing the register insertion, ring structured local computer networks, noted for their high throughput and relative ease of implementation. The questions relate to a new modeling approach and two methods of solving the relevant model: the "partial" one, viewing the network as a collection of individual queueing servers, and the "global" one, based on the image of the network as a large single server. The two approaches are shown to give identical results for symmetric rings. The latter technique is later applied to asymmetric rings and a comparison with simulation results is presented. Furthermore, the proposed technique is used to analyze performance of a selected practical implementation of the ring network, where a file server provides file storage and retrieval services to intelligent workstations. Though performance analyst addressed, the paper also gives insight into ring local computer networks in general.

1. INTRODUCTION

Local area networks (LANs) offer high speed communication between distributed system components, such as intelligent workstations and shared resource controllers (e.g. file and printer servers). LAN links support data rates in the Mbps range, or about three orders of magnitude above the Kbps offered by long distance networks. Application areas for local networks range from campus computing facilities (data processing, interchange and storage) through office automation (word processing, electronic mail and copying), factory automation (CAD, CAM, inventory control), and library systems (book inventory and checkout, document retrieval and electronic copying) to legal systems (case retrieval, billing and word processing).

There are a wide variety of local area networks available. They can be categorized in respect of their topology, transmission media, access techniques and interfaces. As regards topology, most of the LANs feature one of the following configurations: the star, the ring and the bus (a fourth, the tree, is in fact a modified bus), the star being the least common.

Our interest concerns ring local area networks, also called loop networks (the words "loop" and "ring" are used interchangeably in the literature). A ring network (Fig. 1) can be characterized as a sequence of processing elements (stations) interconnected via a communications path, closed on itself. The station is attached to the ring by an interface unit. Messages travel from station to station around the loop, passing through network interfaces, which act as relays. Traffic is usually unidirectional, although double loop topologies have been proposed to enhance network performance and reliability. Messages may be of fixed or variable length and, depending on the particular network, only one or more than one message may be present on the loop at a time.

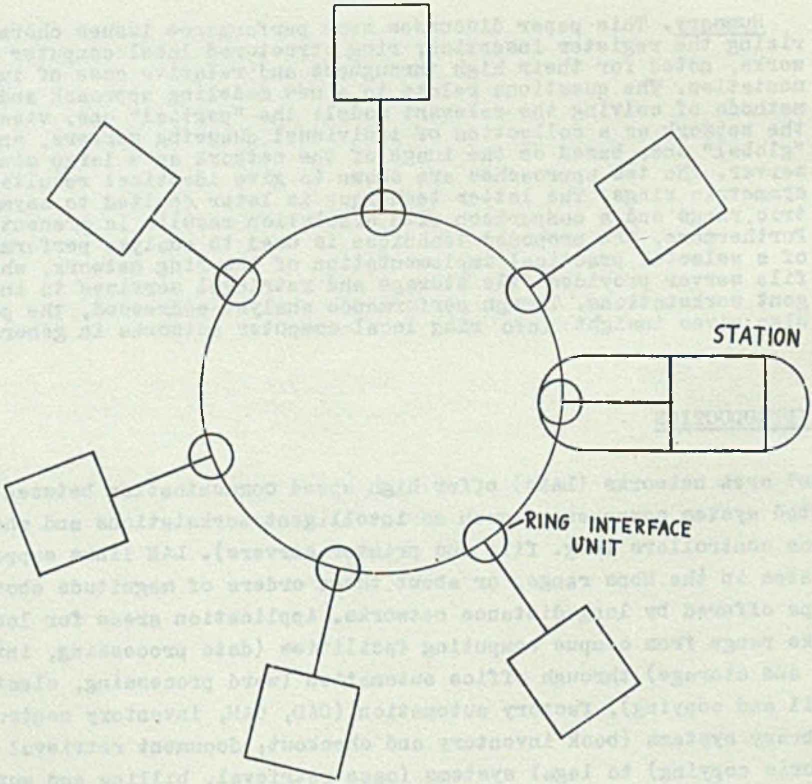

Fig. 1. The ring network
Rys. 1. Sieć typu pierścieniowego

A number of different types of ring networks have been designed and implemented. One reason for this is the relative ease of implementation. Besides, rings can cover large distances since messages can be regenerated at each station. Also, both simulation and live operation have shown the efficiency to be fairly stable with load. Obviously, there are also drawbacks, including reliability, which depends on the reliability of each component (i.e. channel or processor). Moreover, in order to add or delete processing elements, one has to break the ring, which is not the case with bus LANs.

Most of the differences in ring systems are due to loop access protocols. Three basic protocols have been developed: slotted rings, token rings and register insertion rings.

The slotted ring (also called the Pierce loop) is a succession of fixed-length slots circulating around the ring. Each slot incorporates a lead field indicating whether it is empty or full. If the slot is empty, a message may be placed in the slot and the slot marked as full. If the slot is full, it may not be filled. As the slots circulate, each slot passing a station is examined to determine whether it is empty (messages can be transmitted) or full (check to see if message is for this station). Although several messages may be on the network simultaneously, the method's drawback is the slot's fixed length. Obviously, length of messages is seldom constant, which results in a waste of space in the slot (short message) and the necessary software for message assembly/disassembly (long message, disassembled since it would not fit in one slot). Nevertheless, the slotted ring protocol is popular, especially in Europe (e.g. the Cambridge ring [9]).

The token ring (also called the Newhall loop) supports variable - length messages. Control of the loop is passed from station to station through use of a "token" (dedicated bit or two-state bit structure). When a station has control of the loop (i.e. its interface unit has received an "idle" token) and wishes to transmit a message, it transmits and passes control downstream, toggling the token to the "busy" state. Except for the destination station, all stations passed by the message note the "busy"token and refrain from transmitting their own messages (if any), passing control on. The destination station detects the message for itself, removes it from the ring and changes the token back to the "idle" state. Because the message circulating on the ring is of variable length and the control token is passed from station to station, there can only be a single message on the loop at a given time. The token passing protocol has been selected as ANSI/IEEE 802.5 standard.

The register insertion ring (buffer insertion ring, delay register loop) allows multiple variable-length messages to be handled at the same time, thus supporting higher throughput. At each station there are two shift registers, operating as switches to control traffic into and out of the ring, and as extensions to the ring, enabling transmission of long messages.

In this paper, attention is limited to register insertion rings, which have a great potential for applications requiring heavy traffic, such as those dealing with graphics, facsimile, video and voice, and networks which share data heavily across machines.
Section 2 explains the register insertion technique in more detail, presenting a relevant queueing model and the author's proposal for its solution, which is claimed to be more convenient and less time-consuming, especially when the asymmetric flow is assumed. In section 3, an exemplary asymmetric network is analyzed for its performance with the adoption of the proposed method, and a comparison with simulation results is presented. Section 4 describes a selected practical implementation of the ring LAN in database retrieval systems and discusses cases of different asymmetric flows. Concluding remarks comprise section 5.

2. The Register Insertion Ring

A register insertion ring shares the general structure of Fig. 1. Its distinctive features become apparent from the station interface, shown in Fig. 2 [3].

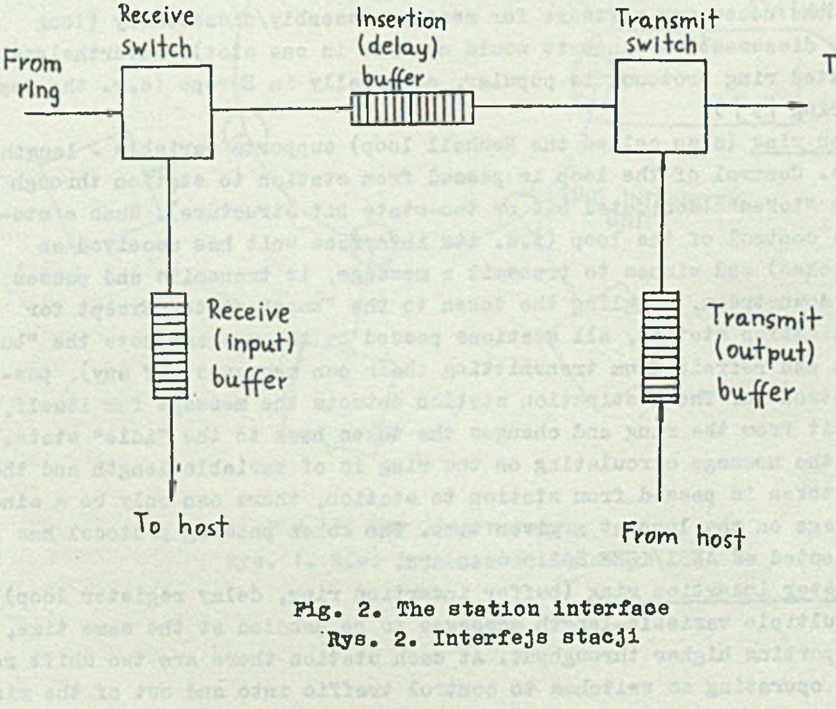

Fig. 2. The station interface
Rys. 2. Interfejs stacji

Three shift registers are present:

- the receive (input) buffer, which accepts packets that are addresses to the station,
- the insertion (delay) buffer, which stores packets not addressed to the station,
- the transmit (output) buffer, which contains packets generated by the station and destined to other stations.

Thus, the station interface receives two streams of packets:

- incoming packets from the ring,
- packets entering the ring from the station.

It is capable of merging them without mutual interference. Conflicts, occurring when a simultaneous arrival of packets from both streams takes place, are resolved by adopting one of the following techniques:

- station priority, meaning that the transmit buffer is allowed to transmit first,
- ring priority; in this case, priority is given to the insertion buffer.

Control is provided by the switch software and is a fixed feature of the network.

Our analysis concerns the Distributed Loop Computer Network (DLCN) [7], where locally generated packets have priority as long as there is enough space in the delay buffer for holding incoming packet(s) until the current packet is transmitted. Otherwise, priority is given to the insertion buffer. Mathematically, this condition may be expressed [14] as

$$D_i - \sum_{j=1}^{k} m_j \geqslant S_i,$$

where S_i is the length the first packet in the output buffer, k the number of packets in the insertion buffer, m_j the length of the jth packet in the insertion buffer, and D_i the length of the insertion buffer. This feature prevents the ring from being overloaded by external streams with large throughputs.

In terms of queueing theory, all buffers are FIFO queues, and packets are customers with defined mean arrival rates and service times. To simplify our further analysis, it is assumed that the delay and output queues can grow without bounds, Also, it has been shown [6,4,2] that, as far as the mean delay time is concerned, if the priority rules are independent of the packet service time (directly or indirectly), then this mean delay is the same for ring priority and station priority, as discussed above. Since storage at the nodes is usually not a problem for ring LANs, performance is reflected primarily in packet delay. Thus, performance analysis of ring networks consists in determining packet delay and how it is affected by various parameters of the network. Hence, the station interface model is simplified and the two packet streams are merged into one queue and served on a first-come-first-served basic.

The queueing model of the register insertion ring is a cyclic collection of models for single station interfaces. It is an open queueing network, resembling the well-known Jacksonian network. Differences are due to the following facts:

1. Kleinrock's "independence assumption" [6] is not invoked. This assumption demands that the interarrival times of packets and their lengths be independent.

2. Service is of the check-and-forward type. As in other types of ring local area networks, each packet contains a header, informing about the packet's destination. The station interface checks the header to see if the packet is addressed to this station. If this is not the case, and providing the channel is not busy, the packet is transmitted immediately without waiting for the rest of it to arrive. In this way, a packet may be simultaneously served by more than one channel. Queueing theory unfortunately assumes that a customer may be served by only one server at a time.

Several analyses have been published to estimate the mean delay time of the register insertion ring [8,12,1]. They, in turn, have been discussed and compared in surveys, including [14,3], and in an interesting study by Hilal et al [5], which also presents approximate formulas for calculating the average waiting time at a station. Since we apply these formulas in our work, some attention will be devoted to the problem of the mean waiting time and how it has been coped with in [5]. The results are then applied to the whole ring. Next, our alternative approach is presented.

2.1. The Average Waiting Time in Tandem Queues

Under assumptions made above, an arbitrary station interface may be represented as a queueing model, depicted in Fig. 3. The model includes the preceding station interface, significant in our analysis.

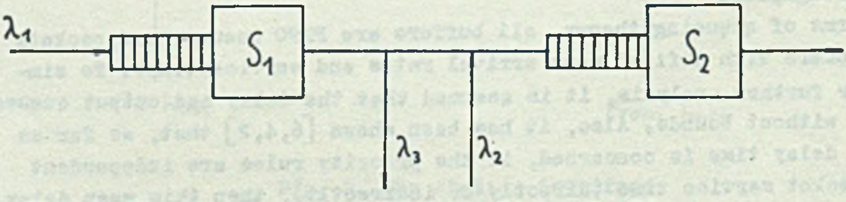

Fig. 3. Two-tandem queues with interfering traffic
Rys. 3. Dwuetapowy system obsługi z zewnętrznym strumieniem zgłoszeń

We are interested in finding the average waiting time for packets passing through station interface s_2. We will use the term "stream λ_i" meaning "stream of packets with arrival rate λ_i packets/second". The following assumptions are made:

1. λ_1 is a Poisson stream. Packets have an exponentially distributed length and an average service time of $\bar{x}$ seconds.

2. λ_2 is a Poisson stream. Packets have an exponentially distributed length with the same mean length and service time as those of λ_1.

3. $\varrho_1 = \lambda_1\bar{x}, \qquad \varrho_2 = \lambda_2\bar{x}$.

4. $\bar{w}_1$ is the mean waiting time for packets of λ_1 at station s_2.
$\bar{w}_2$ is the mean waiting time for packets of λ_2 at station s_2.
$\bar{w} = \bar{w}_1 + \bar{w}_2$.

Rubin [10,11] studied the problem of mean waiting time in N-tandem queues with and without traffic interference (he refers to λ_3 as the interfering traffic). As regards the latter case, he obtained formulas for the mean waiting time by using the superposition technique, consisting in handling only one of the streams λ_1, λ_2 at a time. He concluded that effect of stream λ_2 on stream λ_1 can be accounted for deterministically at s_2 by increasing the traffic intensity from λ_1 to $\lambda_1(1 - \varrho_2)^{-1}$. One of his main assumptions was that the number of arrivals of stream λ_2 packets between the n^{th} and $(n + 1)^{th}$ arrivals of stream λ_1 packets can be approximated by $\lambda_2 T_{n+1}$, where T_{n+1} is the interarrival time between the two arrivals. Unfortunately, this approximation results in poor accuracy at low loads. Hence, Hilal et al [5] propose to estimate the number by $\varrho_2/(1 - \varrho_2)$ at low loads ($\varrho_1 + \varrho_2 < 0.5$). Accordingly, they compute $\bar{w}_1$ by increasing the mean service time from $\bar{x}$ to $\bar{x}(1 - \varrho_2)^{-1}$. Consideration of the effect of stream λ_1 on stream λ_2 is analogous. Thus, $\bar{w}$ is the sum of the mean waiting time in two systems:

- M/M/1 queue, entered by stream λ_2, with the effect of stream λ_1 on the waiting time taken into account,
- two-tandem queue, entered by stream λ_1, without traffic interference but with the effect of stream λ_2 on the waiting time at station s_2 accounted for.

Furthermore, the effect of stream λ_3 on the average waiting time is estimated. It is concluded that:

- the effect of streams λ_1 and λ_3 on stream λ_2 is best expressed by increasing the mean service time at low loads ($\varrho_1 + \varrho_2 - \varrho_3 < 0.5$) from $\bar{x}$ to $\bar{x}(1 - \varrho_1 + \varrho_3)^{-1}$ and by increasing the average traffic intensity at high loads ($\varrho_1 + \varrho_2 - \varrho_3 > 0.5$) from λ_2 to $\lambda_2(1 - \varrho_1 + \varrho_3)^{-1}$,

- the effect of streams λ_2 and λ_3 on stream λ_1 is taken into account by increasing the intensity to $\lambda_1(1 - \rho_3)^{-1}$ and by increasing the mean service time to $\bar{x}(1 - \rho_2)^{-1}$ at all loads.

Accordingly, the authors introduce two symbols for $\bar{w}$: $\bar{w}_{ll}$ at low loads and $\bar{w}_{hh}$ at high loads, whereas the average of $\bar{w}_u$ and $\bar{w}_{hh}$ is accepted when $\rho_1 + \rho_2 - \rho_3 = 0.5$.

$$\bar{w}_{ll} = \frac{\rho_2\bar{x}}{(1 - \rho_1 - \rho_3)(1 - \rho_1 - \rho_2 + \rho_3)} + \frac{\rho_1(1 - \rho_3)\bar{x}}{(1 - \rho_1 + \rho_1\rho_3 - \rho_2)(1 - \rho_2)} - \frac{\rho_1\bar{x}}{1 - \rho_1} \tag{1}$$

$$\bar{w}_{hh} = \frac{\rho_2\bar{x}}{1 - \rho_1 - \rho_2 + \rho_3} + \frac{\rho_1(1 - \rho_3)\bar{x}}{(1 - \rho_1 + \rho_1\rho_3 - \rho_2)(1 - \rho_2)} - \frac{\rho_1\bar{x}}{1 - \rho_1} \tag{2}$$

2.2. The Average Transfer Delay in a Register Insertion Ring

As in other networks, the performance criterion of interest for register insertion ring networks is the average transfer delay on the network (network response time), measured from the time the packet is completely stored in the transmit buffer of the source station to the time the packet is completely received by the receive buffer of the destination station. Hital et al [5] consider a symmetric traffic pattern on the ring (each station generates the same ammount of traffic as it receives, i.e. $\lambda_2 = \lambda_3$. Denoting the number of stations by M and the total traffic arrival rate entering the ring by λ, the total rate of packets transmitted from station i to station i+1 equals

$$\sum_{j=2}^{M-1} \frac{(M - j)\lambda}{M(M - 1)} + \frac{\lambda}{M} = \frac{\lambda}{2}$$

which is independent of both the number of stations and the station's position on the ring. Hence, in order to compute the average waiting time on the ring, it suffices to multiply the waiting time at one station by the average number of stations that a packet passes on its way from source to destination, and which equals M/2 in this case. Since a packet needs an average transmission time $\bar{x}$ at the source station and the header inspection time equals H/C, where H is the header length and C is the channel transmission speed, the formula for the average transfer delay is as follows

$$\bar{R} = h\bar{w} + \bar{x} + h\frac{H}{C} \tag{3}$$

The above approach is hardly applicable in an asymmetric traffic environment. In this case, one cannot compute the mean waiting time by simply adding "partial" waiting times at stations, since the number of such components h is not explicitly known. Besides, $\bar{w}$ now varies with the station index, as do station throughputs. Our proposal is to adopt a combination of a conventional Jacksonian network solution and the above one. The proposed formula for the average transfer delay is

$$\bar{R} = \bar{W} + \bar{x} + h\frac{H}{C} \tag{4}$$

where $\bar{W}$ is the "global" mean waiting time on the ring, i.e. the average time that a packet spends waiting in queues. Below we present the algorithm that enables to calculate unknown parameters $\bar{W}$ and h.

First, one has to compute station throughputs λ_i, $i=1,2,\dots,M$ (from now on, subscripts indicate station numbers), by solving a set of M linear equations

$$\lambda_m = \sum_{i=1}^{M} \lambda_i P_{im} + \lambda_{Om}, \qquad m = 1,2,\dots,M$$

where λ_{Om} is the arrival rate of packets entering the ring from station (host) m. Next, h may be computed as follows:

Since h indicates the mean number of stations visited by a packet, it may be expressed as

$$h = \sum_{i=1}^{M} e_i = \sum_{i=1}^{M} \frac{\lambda_i}{\lambda} = \frac{1}{\lambda} \sum_{i=1}^{M} \lambda_i$$

where λ is the "global" throughput of the whole network

$$\lambda = \sum_{i=1}^{M} \lambda_{oi}$$

In order to compute the "global" mean waiting time $\bar{W}$ on the ring, we apply Little's formula

$$\bar{W} = \frac{Q}{\lambda}$$

where Q is the "global" mean queue length, i.e. the sum of mean numbers of packets q_i in all queues on the ring:

$$Q = \sum_{i=1}^{M} q_i$$

q_i is obtained by invoking Little's law again:

$$q_i = \lambda_i \bar{w}_i \,, \qquad i = 1,2,\dots,M$$

$\bar{w}_i$ is computed according to formulas (1) and (2).

Clearly, formula (4) holds true for a symmetric traffic. As we remember, in that case

$$\bar{w}_1 = \bar{w}_2 = \dots = \bar{w}_M = \bar{w}$$

$$\lambda_1 = \lambda_2 = \dots = \lambda_M = \frac{\lambda}{2}$$

Hence,

$$\bar{R} = \frac{1}{\lambda} \sum_{i=1}^{M} \frac{\lambda}{2} \bar{w} + \bar{x} + \frac{M}{2}\frac{H}{C} = \frac{M}{2}\bar{w} + \bar{x} + \frac{M}{2}\frac{H}{C}$$

which is equivalent to (3). The equivalence was borne out by comparison of analytical results obtained by both methods for 2 symmetric rings, analyzed by Hilal et al [5] . The results were, unsurprisingly, identical. They are presented in Figs. 4 and 5. Also, plotted are relevant simulation results (discussion of agreement between analysis and simulation can be found in [5]). The load is defined as the total number of input bits that enter the ring per second, divided by the channel capacity. The average transfer delay is normalized by dividing its value by the mean packet transmission time $\bar{x}$.

There is, however, a slight difference between our curve in Fig. 4 and the one published in [5] . It turned out that, for a small number of 6 nodes, a better approximation is obtained if one takes the average of $\bar{w}_{11}$ and $\bar{w}_{hh}$ for $\rho_1 + \rho_2 - \rho_3 = 0.6$ instead of 0.5. Accordingly, the two curves differ at loads 1.0 and 1.2.

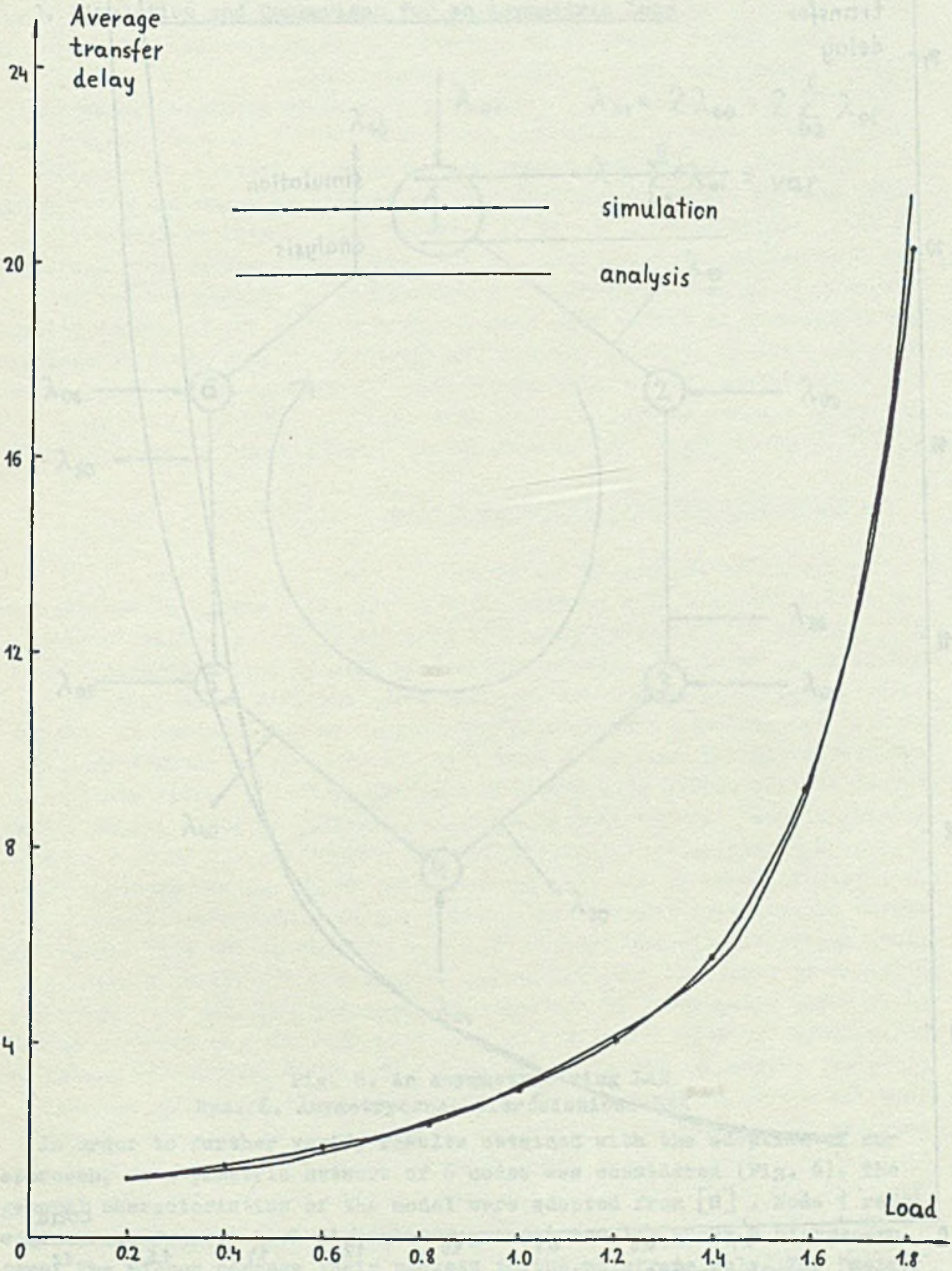

Fig. 4. Average transfer delay for a 6-station symmetric register insertion ring

Rys. 4. Średni czas zwłoki dla 6-stanowiskowej, symetrycznej sieci pierścieniowej z wstawianiem bufora

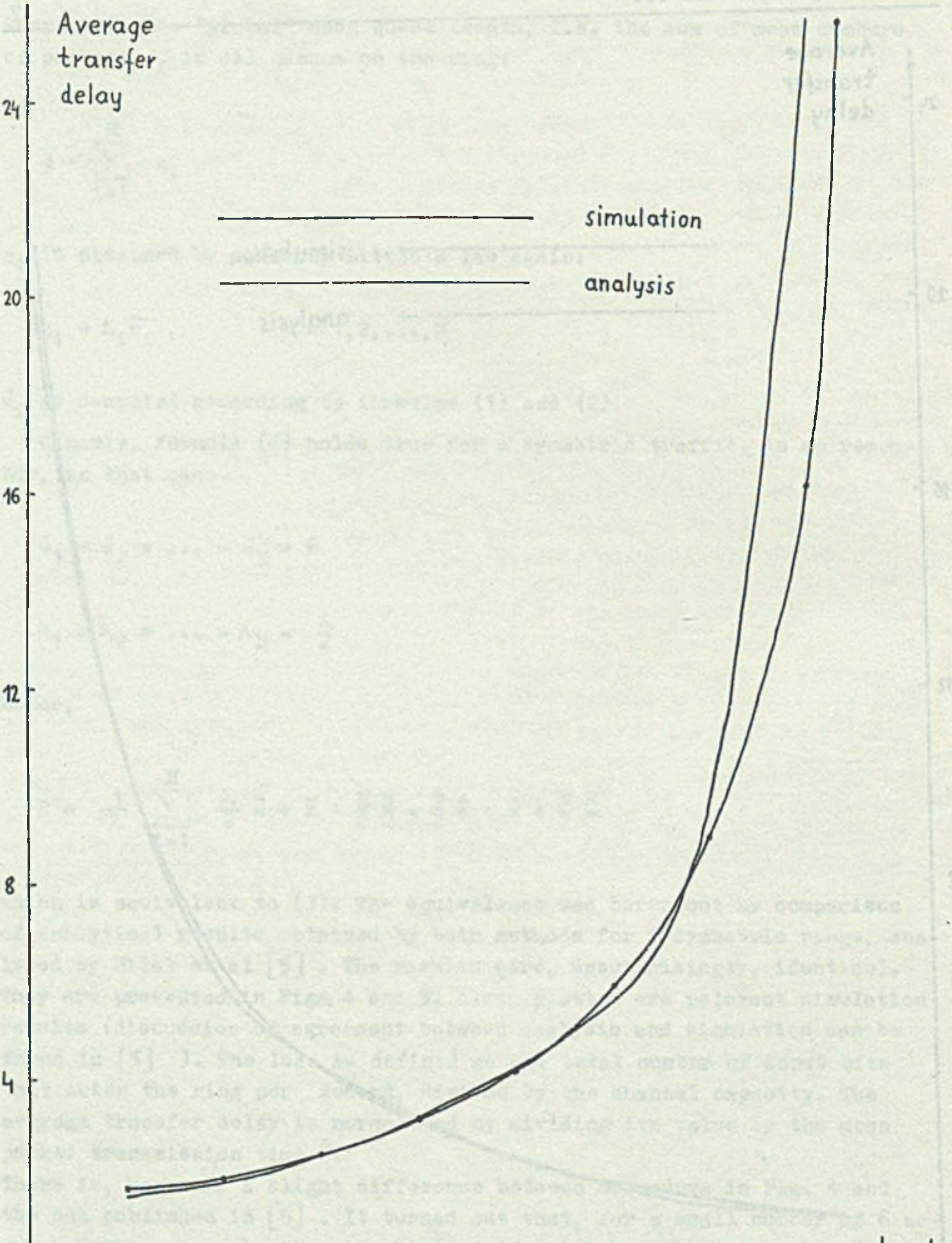

Fig. 5. Average transfer delay for a 40-station symmetric register insertion ring

Rys. 5. Średni czas zwłoki dla 40-stanowiskowej, symetrycznej sieci pierścieniowej z wstawianiem bufora

3. Simulation and Comparison for an Asymmetric Loop

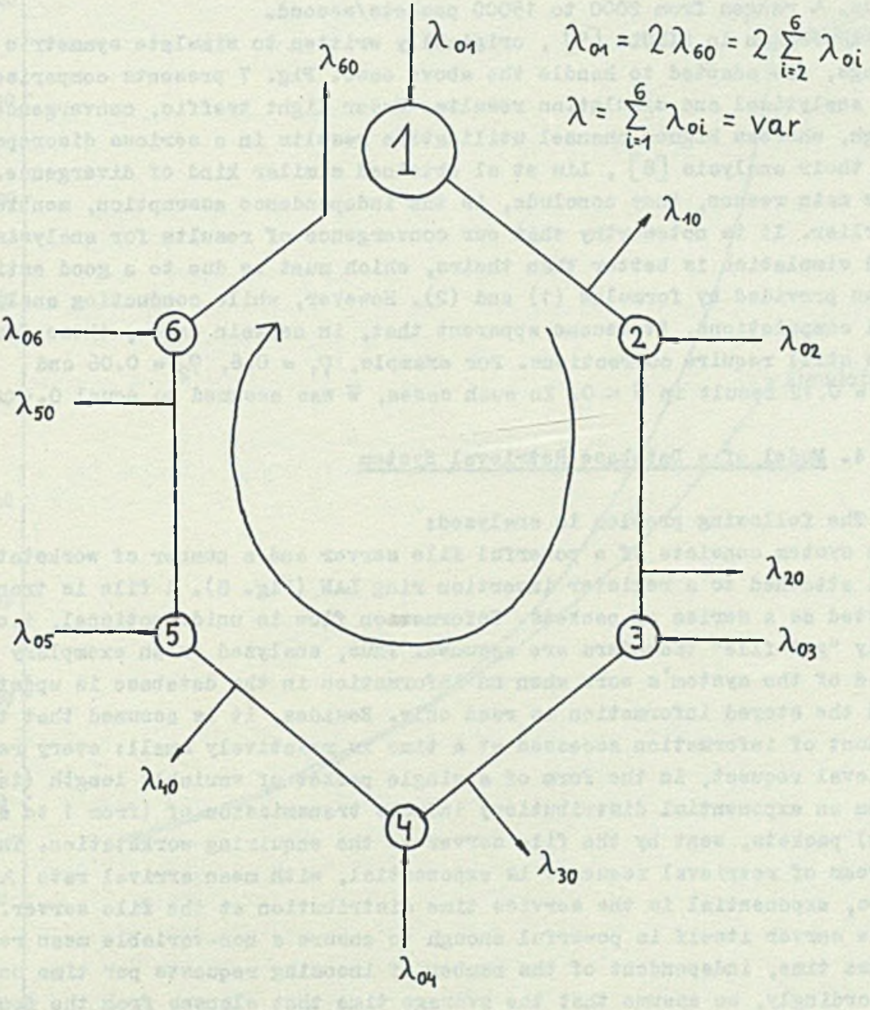

Fig. 6. An asymmetric ring LAN
Rys. 6. Asymetryczna, pierścieniowa LSK

In order to further verify results obtained with the adoption of our approach, an asymmetric network of 6 nodes was considered (Fig. 6). The general characteristics of the model were adopted from [8] . Node 1 represents a mainframe, which distributes packets equally among 5 microcomputers. The micros address their packets to the mainframe only. The "mainframe-to-micros" traffic is assumed to be twice as much as the "micros-to-mainframe" traffic. Thus, packets generated by the mainframe constitute two thirds of the global input traffic, whereas the micros produce one third. Channel capacity is measured in characters/second and equals C=1M. The average packet length is 60 characters. The header inspection time is 2 μs. λ ranges from 2000 to 15000 packets/second.

A program in SIMULA [5] , originally written to simulate symmetric rings, was adapted to handle the above case. Fig. 7 presents comparison of analytical and simulation results. Under light traffic, convergence is high, whereas higher channel utilization results in a serious discrepancy. In their analysis [8] , Liu et al obtained similar kind of divergence. The main reason, they conclude, is the independence assumption, mentioned earlier. It is noteworthy that our convergence of results for analysis and simulation is better than theirs, which must be due to a good estimation provided by formulas (1) and (2). However, while conducting analytical computations, it became apparent that, in certain cases, these formulas still require corrections. For example, $\rho_1 = 0.6$, $\rho_2 = 0.06$ and $\rho_3 = 0.12$ result in $\bar{w} < 0$. In such cases, $\bar{w}$ was assumed to equal 0.

4. Model of a Database Retrieval System

The following problem is analyzed:
The system consists of a powerful file server and a number of workstations, all attached to a register insertion ring LAN (Fig. 8). A file is transmitted as a series of packets. Information flow is unidirectional, i.e. only "get-file" transfers are assumed. Thus, analyzed is an exemplary period of the system's work when no information in the database is updated and the stored information is read only. Besides, it is assumed that the amount of information accessed at a time is relatively small: every retrieval request, in the form of a single packet of variable length (taken from an exponential distribution) induces transmission of (from 1 to a few) packets, sent by the file server to the enquiring workstation. The stream of retrieval requests is exponential, with mean arrival rate λ_r: also, exponential is the service time distribution at the file server. The file server itself is powerful enough to ensure a non-variable mean response time, independent of the number of incoming requests per time unit. Accordingly, we assume that the average time that elapses from the moment a request enters the server to the moment the first information packet enters the server's output buffer is identical at all loads. In terms of queueing theory, the file server can be modeled as an infinite, exponential server, preceded by a "fork" operator (Fig. 9). Our aim is to analyze the above system in respect of its mean response time's dependency on both the request frequency and the amount of information accessed by a workstation at a time.

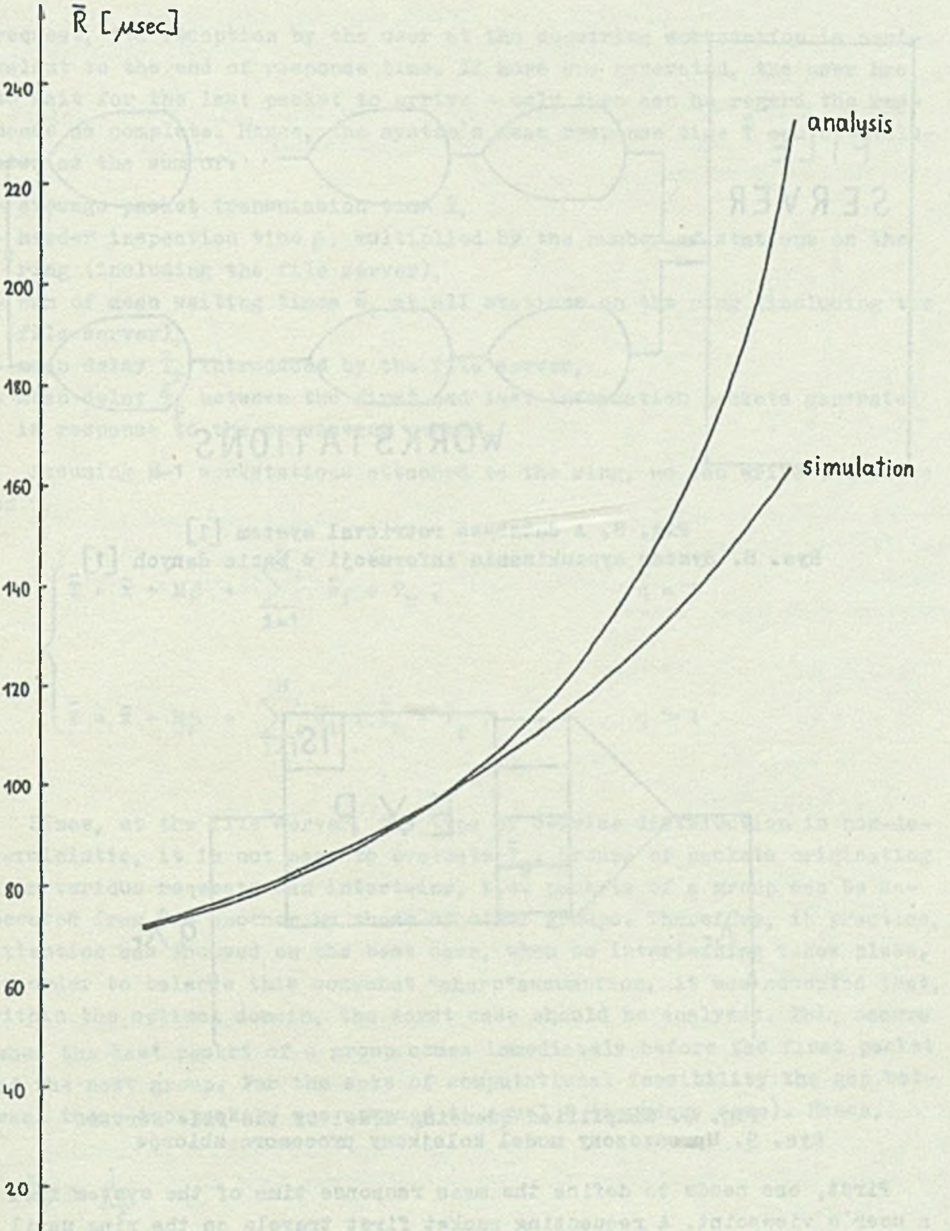

Fig. 7. Average transfer delay for an asymmetric loop
Rys. 7. Średni czas zwłoki dla sieci asymetrycznej

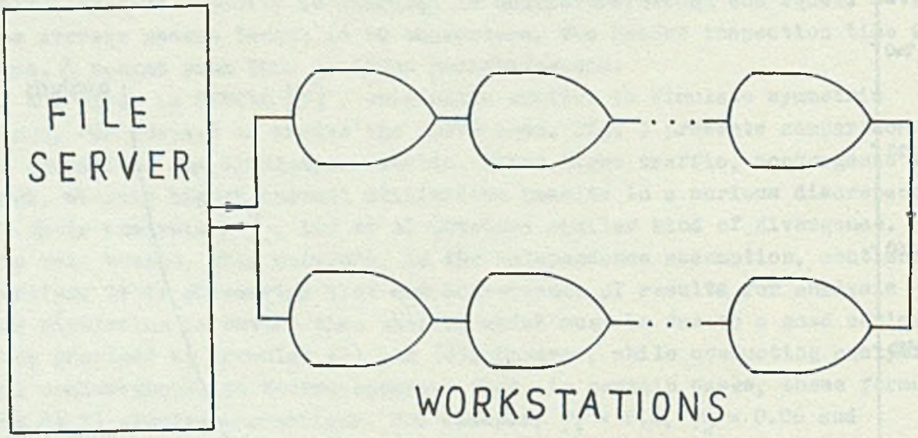

Fig. 8. A database retrieval system [1]
Rys. 8. System wyszukiwania informacji w bazie danych [1]

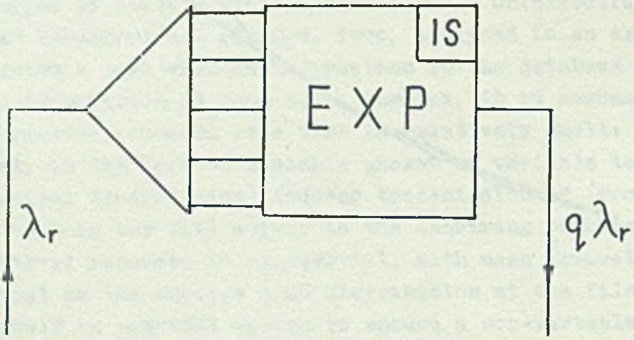

Fig. 9. Simplified queueing model of the file server
Rys. 9. Uproszczony model kolejkowy procesora zbiorów

First, one needs to define the mean response time of the system from a user's viewpoint. A requesting packet first travels on the ring until it reaches the file server. Next, demanded information is retrieved, which takes a certain amount of time, taken from an exponential distribution. Finally, the first packet of a group of q "responding" information packets is ready to be transmitted. It enters the ring and travels to the enquiring workstation. If only one information packet is generated per request, its reception by the user at the enquiring workstation is equivalent to the end of response time. If more are generated, the user has to wait for the last packet to arrive - only then can he regard the response as complete. Hence, the system's mean response time $\bar{T}$ can be evaluated as the sum of:

- average packet transmission time $\bar{x}$,
- header inspection time β, multiplied by the number of stations on the ring (including the file server),
- sum of mean waiting times $\bar{w}_i$ at all stations on the ring (including the file server),
- mean delay $\bar{T}_r$ introduced by the file server,
- mean delay $\bar{T}_p$ between the first and last information packets generated in response to the requesting packet.

Assuming M-1 workstations attached to the ring, we can write the above as

$$\begin{cases} \bar{T} = \bar{x} + M\beta + \sum_{i=1}^{M} \bar{w}_i + \bar{T}_r , & q = 1 \\ \bar{T} = \bar{x} + M\beta + \sum_{i=1}^{M} \bar{w}_i + \bar{T}_r + \bar{T}_p , & q > 1 \end{cases}$$

Since, at the file server, the type of service distribution is non-deterministic, it is not easy to evaluate $\bar{T}_p$. Groups of packets originating from various requests can intertwine, i.e. packets of a group can be seperated from one another by those of other groups. Therefore, in practice, attention was focused on the best case, when no intertwining takes place, In order to balance this somewhat "sharp"assumption, it was accepted that, within the optimal domain, the worst case should be analyzed. This occurs when the last packet of a group comes immediately before the first packet of the next group. For the sake of computational feasibility the gap between these two packets was assumed to equal 0 (boundary case). Hence,

$$\bar{T}_p = \frac{1}{\lambda_r}$$

Let us note that the interfering traffic (retrieval requests) at the nodes on the way from the file server to the analyzed workstation does not affect $\bar{T}_p$, although the arrival rate of all packets changes.

As regards $\bar{T}_r$, since it is unaffected by q and λ_r, we can ignore it in the analysis, realizing that it increases the mean response time of the system independently of parameters. Hence, any resulting graphs will only be shifted upwards and the shape of curves, which is of most interest to us, will remain unchanged.

Denoting the resulting "global response time minus file server delay" (referred to as simply "response time" in the sequel) by $\bar{T}\_$ and following assumptions regarding $\bar{T}_p$:

$$\begin{cases} \bar{T}\_ = \bar{x} + M\beta + \sum_{i=1}^{M} \bar{w}_i \,, & q = 1 \\ \bar{T}\_ = \bar{x} + M\beta + \sum_{i=1}^{M} \bar{w}_i + \frac{1}{\lambda_r} \,, & q > 1 \end{cases}$$

The model was solved analytically with the adoption of formulas (1) and (2) for evaluating $\bar{w}_i$, $i=1,2,\dots,M$. Values for $\bar{T}\_$ were obtained for the cases M = 6 (5 workstations) and M = 21 (20 workstations). Other parameters were as follows:

$\bar{x} = 60\ \mu s$

$\beta = 2\ \mu s$

$q = 1, 2, 4, 8$

$\lambda \in [2000; 16000]$, $\lambda = \sum_{i=1}^{M} \lambda_{oi}$ (see Fig. 6)

(The upper limit for λ for this model's parameters is approximately $16500(q+1)/q$; however, 16000 was adopted as the limit since $\lambda > 16000$ results in a serious divergence of simulation and analysis, as shown in section 3).

Figures 10 and 11 present relevant graphs. For $q = 1$ the mean response time of the system is equivalent to the average transfer delay on the ring (not global, though, but from a user's viewpoint).

Results for $q > 1$ show that the mean response time has its local minimum, which is shifted upwards and to the right when q increases. Obviously, the whole curves are shifted upwards as well. They tend to become sharper with increasing q.

The parabolic shape of curves may be explained from observations of results for $\bar{w}_i$ (not quoted here). For small values of λ, gradual increase of all waiting times on the ring was observed. At the same time the mean delay between packets decreased significantly, which resulted in $\bar{T}\_$ decreasing. Next, there came the moment when the file server's interface unit (station interface 1) became seriously loaded.

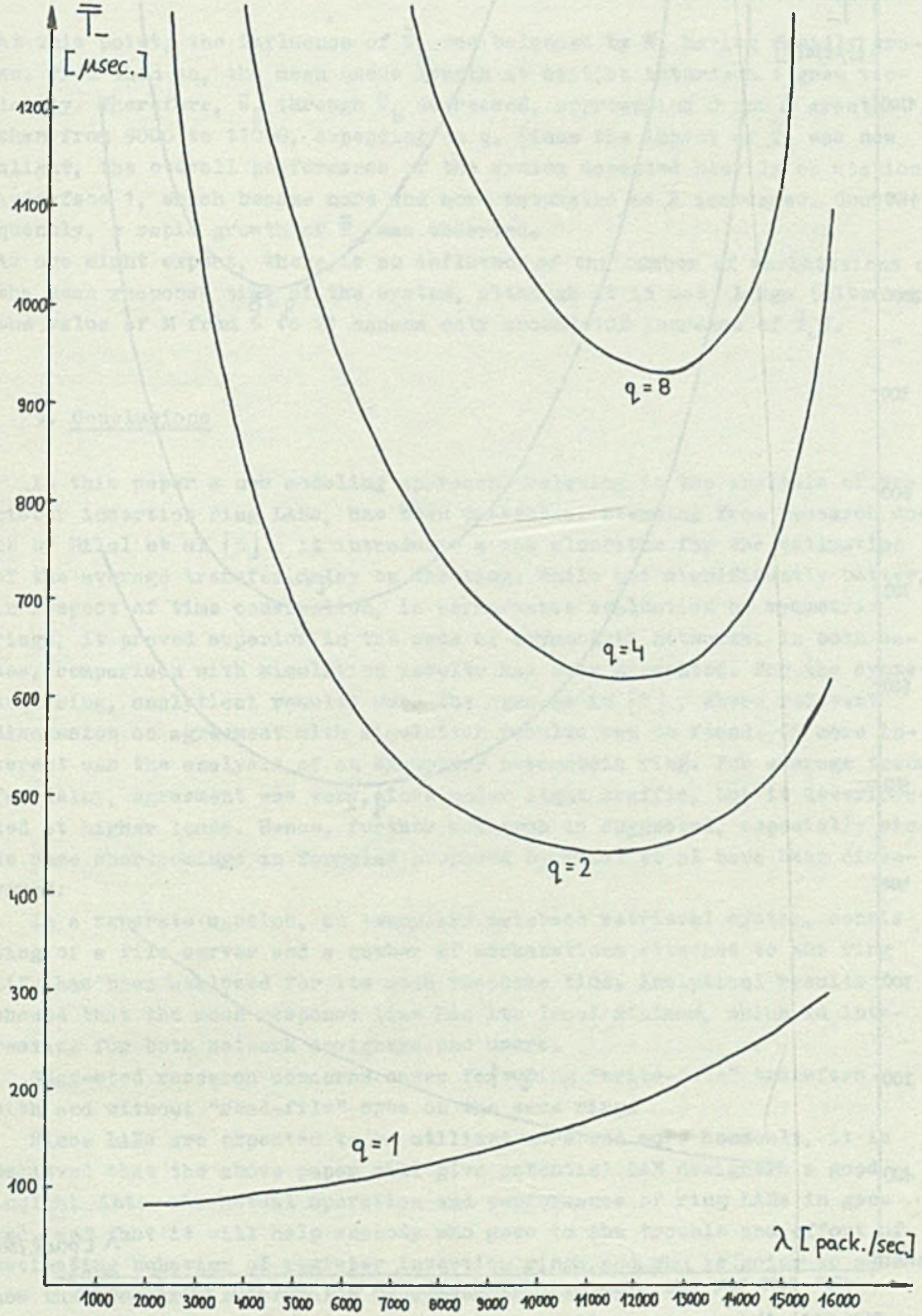

Fig. 10. Mean response time of the database retrieval system with 5 users accessing the file server

Rys. 10. Średni czas reakcji systemu wyszukiwania informacji w bazie danych z 5 użytkownikami korzystającymi z procesora zbiorów

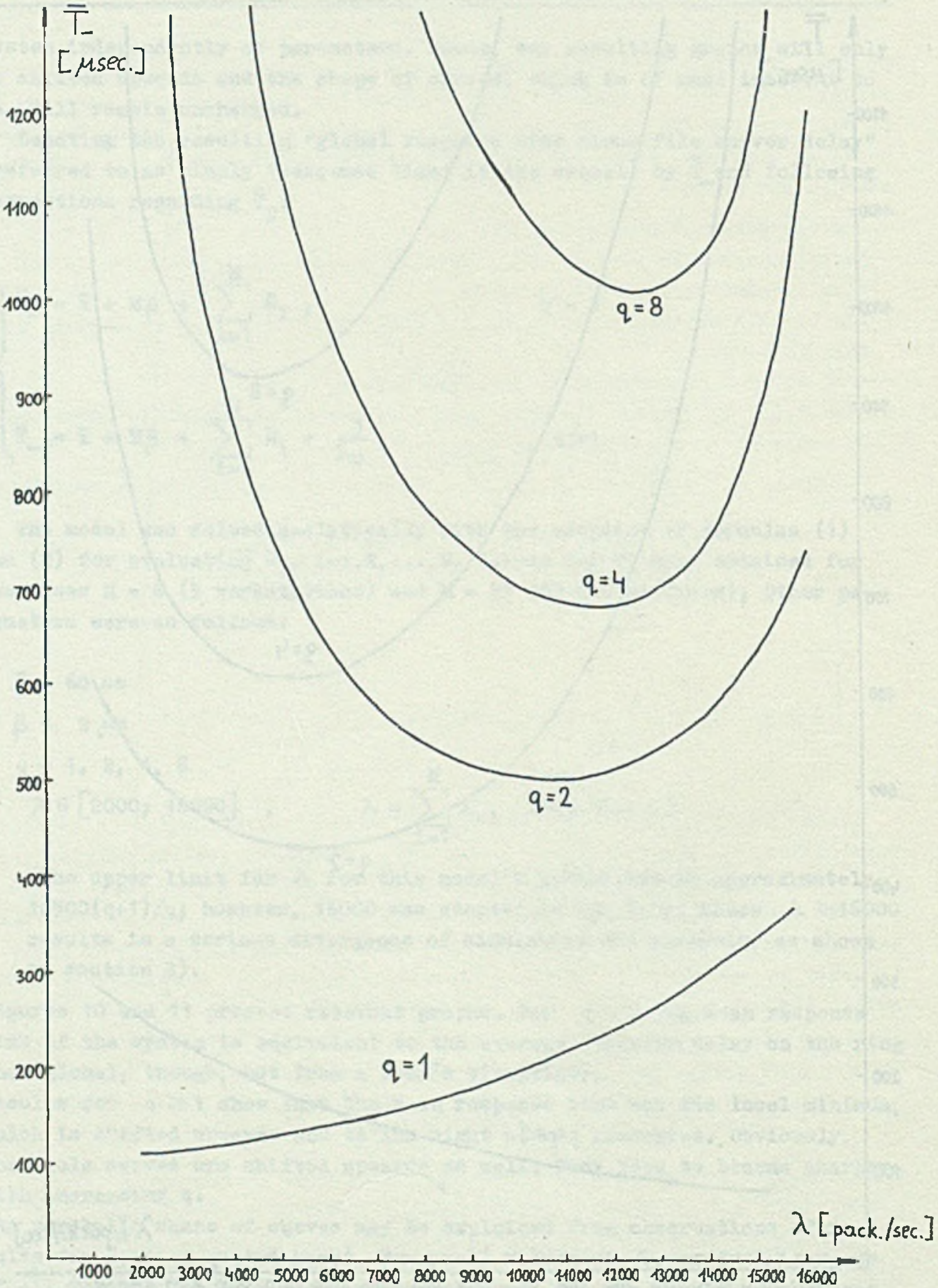

Fig. 11. Mean response time of the database retrieval system with 20 users accessing the file server

Rys. 11. Średni czas reakcji systemu wyszukiwania informacji w bazie danych z 20 użytkownikami korzystającymi z procesora zbiorów

At this point, the influence of $\bar{T}_p$ was balanced by $\bar{w}_1$ having rapidly grown. From then on, the mean queue length at station interface 1 grew violently. Therefore, $\bar{w}_2$ through $\bar{w}_M$ decreased, approaching 0 for λ greater than from 9000 to 11000, depending on q. Since the impact of $\bar{T}_p$ was now slight, the cverall performance of the system depended heavily on station interface 1, which became more and more saturated as λ increased. Consequently, a rapid growth of $\bar{T}_-$ was observed.

As one might expect, there is an influence of the number of workstations on the mean response time of the system, although it is not large (altering the value of M from 6 to 21 causes only around 10% increase of $\bar{T}_-$).

5. Conclusions

In this paper a new modeling approach, relating to the analysis of register insertion ring LANs, has been presented. Stemming from research done by Hilal et al [5], it introduces a new algorithm for the estimation of the average transfer delay on the ring. While not significantly better, in respect of time consumption, in performance evaluation of symmetric rings, it proved superior in the case of asymmetric networks. In both cases, comparison with simulation results has been presented. For the symmetric ring, analytical results were the same as in [5], where relevant discussion on agreement with simulation results can be found. Of more interest was the analysis of an exemplary asymmetric ring. For average transfer delay, agreement was very close under light traffic, but it deteriorated at higher loads. Hence, further research is suggested, especially since some shortcomings in formulas proposed by Hilal et al have been discovered.

In a separate section, an exemplary database retrieval system, consisting of a file server and a number of workstations attached to the ring LAN, has been analyzed for its mean response time. Analytical results showed that the mean response time has its local minimum, which is interesting for both network designers and users.

Suggested research concerns cases featuring "write-file" transfers with and without "read-file" ones on the same ring.

Since LANs are expected to be utilized more and more commonly, it is believed that the above paper will give potential LAN designers a good insight into the actual operation and performance of ring LANs in general, and that it will help anybody who goes to the trouble and effort of estimating behavior of register insertion rings and who is going to assess how this behavior is affected by system parameters.

Acknowledgement

The author wishes to thank Dr Wael Hilal from the Department of Computer Science at the University of Houston, USA, for many helpful discussions and for providing simulation results.

REFERENCES

[1] W. Bux, "Performance issues in local-area networks", IBM Systems Journal, vol. 23, no. 4, 1984, pp.351-374.

[2] W. Bux, M. Schlatter, "An approximate method for the performance analysis of buffer insertion rings", IEEE Transactions on Communications, vol. COM-31, no. 1, Jan. 1983, pp.50-55.

[3] J.L. Hammond, P.J.P. O'Reilly, "Performance analysis of local computer networks", Addison-Wesley, 1986.

[4] W. Hilal, M.T. Liu, "Analysis and simulation of the register-insertion protocol", Proc. Computer Networking Symposium, Dec. 1982, pp.91-102.

[5] W. Hilal, J.F. Chiu, M.T. Liu, "A study of register insertion rings", to be published.

[6] L. Kleinrock, "Qeuueing systems", vol. II, Wiley, 1976.

[7] M.T. Liu, "Distributed loop computer networks", in "Advances in computers", ed. M.C. Yovits, vol. 17, pp.163-221, Academic Press 1978.

[8] M.T. Liu, G. Babic, R. Pardo, "Traffic analysis of the Distributed Loop Computer Network (DLCN)", Proc. 1977 National Telecommunication Conference, Dec. 1977, pp.31:5-1 - 32:5-7.

[9] R.M. Needham, A.J. Herbert, "The Cambridge Distributed Computing System", Addison-Wesley, 1982.

[10] I. Rubin, "Message path delays in packet-switching communication networks", IEEE Transactions on Communications, vol. COM-23, no. 2 Feb. 1975, pp.186-192.

[11] I. Rubin, "An approximate time-delay analysis for packet-switching communication networks", IEEE Transactions on Communications, vol. COM-24, no. 2, Feb. 1976, pp.210-222.

[12] A. Thomasian, H. Kanakia, "Performance study of loop networks using buffer insertion", Computer Networks, vol. 3, 1979, pp.419-425.

[13] K.J. Thurber, "Local computer networks", in "Advances in computers", ed. M.C. Yovits, vol. 20, pp.85-114, Academic Press, 1981.

[14] C. Tropper, "Local computer network technologies", Academic Press, 1981.

Recenzent: Doc. dr hab. inż. Krzysztof Zieliński

Wpłynęło do Redakcji 29.04.1988 r.

WYBRANE ZAGADNIENIA Z DZIEDZINY OCENY EFEKTYWNOŚCI DZIAŁANIA PIERŚCIENIOWYCH, LOKALNYCH SIECI KOMPUTEROWYCH Z WSTAWIANIEM BUFORA

Streszczenie

W pracy przedstawiono podstawowe problemy, charakteryzujące ocenę efektywności działania lokalnych sieci komputerowych typu pierścieniowego, z tzw. wstawianiem bufora jako metodą sterowania dostępem do kanału komunikacyjnego. Prezentacja zagadnień oceny poprzedzona jest krótkim wstępem na temat ogólnego pojęcia lokalnych sieci komputerowych oraz pewnej ich klasy: sieci o konfiguracji pierścieniowej. Następnie przedstawiona została zasada działania oraz model kolejkowy sieci pierścieniowej z wstawianiem bufora, na przykładzie sieci DLCN. Po zaprezentowaniu znanej z literatury metody rozwiązania w/w modelu w ujęciu "cząstkowym", przedstawiony został sposób korzystający z ujęcia "globalnego". Ilustracją są wyniki programu, implementującego tą metodę, uzyskane dla sieci o symetrycznym i niesymetrycznym rozpływach pakietów i porównanie ich z wynikami symulacji w funkcji intensywności generacji pakietów. W dalszej części opisano przykład praktycznego wykorzystania metod analizy omawianego typu sieci: przedstawiony został system gromadzenia danych, z którego korzystają użytkownicy dysponujący inteligentnymi końcówkami interaktywnymi; dla systemu tego określony został wpływ niektórych jego parametrów na średni czas reakcji, z punktu widzenia użytkownika przy końcówce interaktywnej.

ВЫБРАННЫЕ ПРОБЛЕМЫ ИЗ ОБЛАСТИ ОЦЕНКИ ЭФФЕКТИВНОСТИ РАБОТЫ КОЛЬЦЕВЫХ ЛОКАЛЬНЫХ СЕТЕЙ ЭВМ С ВСТАВКОЙ РЕГИСТРОВ

Резюме

В работе представлены основные проблемы, характеризующие оценку эффективности работы локальных вычислительных сетей кольцевого типа, с так называемой вставкой регистров как принятым способом управления доступом к каналу связи. После краткого обзора классов локальных сетей ЭВМ, с особенным учётом сетей кольцевого типа, внимание сосредоточено на кольцевой сети с вставкой регистров. Представлен принцип работы такой сети и её модель как системы массового обслуживания. Примером является здесь разработания в США сеть DLCN . В статье описываются два способа решения проблемы оценки средного времени отклика пакета информации: частный, представлен другими исследователями в их работах, и валовой - предлагаемый автором статьи. Иллюстрацией являются итоги работы программы, внедряющей второй способ и применённой для анализа кольцов с симметрической и несимметрической передачей пакетов. После того следует пример практичкского применения методов анализа обсуждаемого типа кольцевой сети: система накапливания информации, используемой обладателями терминалов. Для представленной системы определено влияние некоторых её параметров на среднее время отклика, с точки зрения пользователя терминала.