目次

1．デジタル交換方式

○ Time slot interchange in three-stage time-division switching networks

○ An optimum design theory for PCM toll switching networks

○ Early development history of digital time division switching network using time slot interchange

○ VTDM: A variable bit rate TDM switch architecture for video stream

○ 可変速度 TDM 交換方式を用いた最低帯域保証機能をもつ動画像トラヒック制御方式

○ 時分割スイッチ回路網における対関係の構成

○ Ｔ－Ｓ－Ｔ形スイッチ回路網における経路探索処理量の評価

○ 時分割交換方式

2．通信網同期方式

○ 相互同期系の理論

○ 相互同期系の動作モード

○ 相互同期系の一般化された構成概念

3．大容量パケット交換方式

○ High capacity packet switching system by means of Multi-Microprocessors

○ Evaluation of alternatives for intra-network inplementation of X.25 compatible packet switching networks

○ A proposal of high capacity switching fabric for Tbps-level IP Router

○ 多段接続網を用いた大容量ＩＰデータグラムスイッチング方式

4．回線割当制御方式

○ 衛星データ通信における順次予約方式

○ 超高速ＬＡＮ／ＭＡＮのための高能率アクセス方式

5．プロトコル工学

○　オートマンモデルによるＨＤＬＣ　プロトコル製品検証の一方式

○　多層構造プロトコルをもつネットワークの伝送効率の評価

○　計算機網におけるウィンドウ方式　フロー制御の効果の一評価

○　トランスピュータによる多層プロトコルの性能評価用シミュレータ

6．ハブ型ローカルエリアネットワーク

○　A star type local area network -MUITISTAR- having simultaneous multiple transmission capability

○　同時多重伝送機能をもつスター形ＬＡＮ

7．分散処理・コンピュータネットワーク

○　ネットワーク・オペレーティング・システムにおけるプロセス管理方式

○　分散形データフロー計算機の構成と実験的評価

○　ＦＩＦＯ付きブロードキャストメモリを用いた共有メモリの実現とその評価

8．標準化

○　A proposal for dejure standardization by implementation

○　Regional activities for telecommunication standardization

１．　デジタル交換方式

○ Time slot interchange in three-stage time-division switching networks
○ An optimum design theory for PCM toll switching networks
○ Early development history of digital time division switching network using time slot interchange
○ VTDM: A variable bit rate TDM switch architecture for video stream
○ 可変速度ＴＤＭ交換方式を用いた最低帯域保証機能をもつ動画像トラヒック制御方式
○ 時分割スイッチ回路網における対関係の構成
○ Ｔ－Ｓ－Ｔ形スイッチ回路網における経路探索処理量の評価
○ 時分割交換方式

TIME SLOT INTERCHANGE IN THREE-STAGE TIME-DIVISION SWITCHING NETWORKS

by

H. INOSE, T. SAITO and M. KATO

Department of Electronic Engineering
Faculty of Engineering
University of Tokyo, Tokyo, Japan

1. Introduction

In a conventional four-wire three-stage time-division switching network [1] shown in Figure 1, a connection is set up by means of a time slot which is commonly idle in the three time-division stages, namely the send highway, one of the junctor pairs and the receive highway. In case that a commonly idle time slot is unavailable, an internal blocking results which may be called the time slot mismatch loss. For a network with n time slots, t junctor pairs and with offered traffic of a_1, a_2 and a_3 erlangs to each of the send highways, the junctors and the receive highways respectively, the following blocking probability formula has been derived [2] assuming that the distribution of calls on the highways and the junctors is independent and random.

$$B = \frac{1}{\left(\sum_{i=0}^{n} \frac{a_1^i}{i!}\right)\left(\sum_{j=0}^{n} \frac{a_2^j}{j!}\right)^t\left(\sum_{k=0}^{n} \frac{a_3^k}{k!}\right)} \sum_{m=0}^{n} \frac{(n-m)!}{n!}\left(\frac{m!}{n!} a_2^{n-m}\right)^t$$

$$\times \sum_{i=0}^{m} \sum_{k=m-i}^{m} \frac{a_1^i a_3^k}{(i+k-m)!} \binom{n-i}{m-i} \binom{n-k}{m-k} \left(\sum_{j=0}^{m} \frac{a_2^j}{j!}\right)^t \tag{1}$$

In case that T out of t junctor pairs are to be matched to the send and receive highways t in the above equation should be replaced by T.

On the other hand, if different time slots can be used in each of the time division stages and if sufficiently large number of junctor pairs are provided, the network becomes

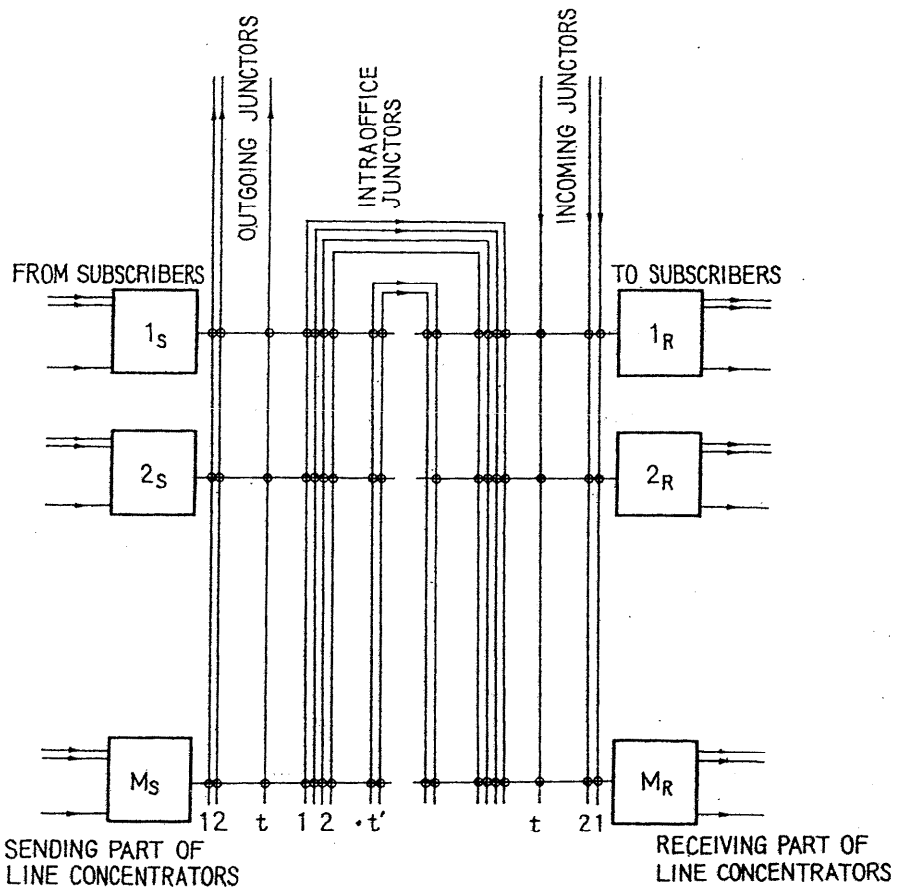


FIG. 1.

free from the time slot mismatch and the blocking probability reduces to

$$B_0 = 1 - \left(1 - \frac{\dfrac{a^n}{n!}}{\displaystyle\sum_{\mu=0}^{n} \frac{a^\mu}{\mu!}}\right)^2 \tag{2}$$

The difference between B and B_0 represents the time slot mismatch loss. The objective of synthesizing the three-stage time-division switching networks with the time slot interchange capability is the elimination, or from the economical point of view the minimization of the time slot mismatch loss.

2. Three-Stage Network Free from Time Slot Mismatch

The knowledge of space-division non blocking network [3] may be conveniently adopted for the realization of a time slot mismatch free three-stage time-division switching network. The time-division switching element in the conventional network is the junctor gates which connect M highways and t junctors by a same time slot. Therefore, in case that the number of time slots in one frame is n, the junctor gates is analogical to a group

of n independent $M \times t$ crossbar switches, with the restriction that a terminal of a crossbar switch can only be connected to a terminal of another crossbar switch representing the same time slot.

To avoid the restriction, an appropriate delay or memory device which delays or shifts a time slot to any other time slot should be introduced. In a time-division exchange employing digital modulation techniques, such a device is economically realizable by a pulse shifter, composed of a shift register, or a tapped delay line with one frame delay and with output taps having one time slot delay in each other [4]. The space-division analogy of the pulse shifter is a $n \times n$ crossbar switch. Thus the time-division network analogous to the space-division non-blocking network of Figure 2*a* is realized by using $2M-1$ junctors with the pulse shifters as shown in Figure 2*b*. The junctors with pulse shifters may be called hereafter as the pulse-shifting-junctors.

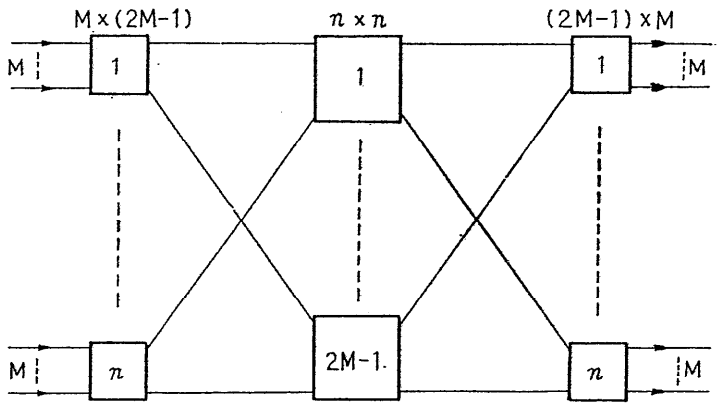


FIG. 2*a*.

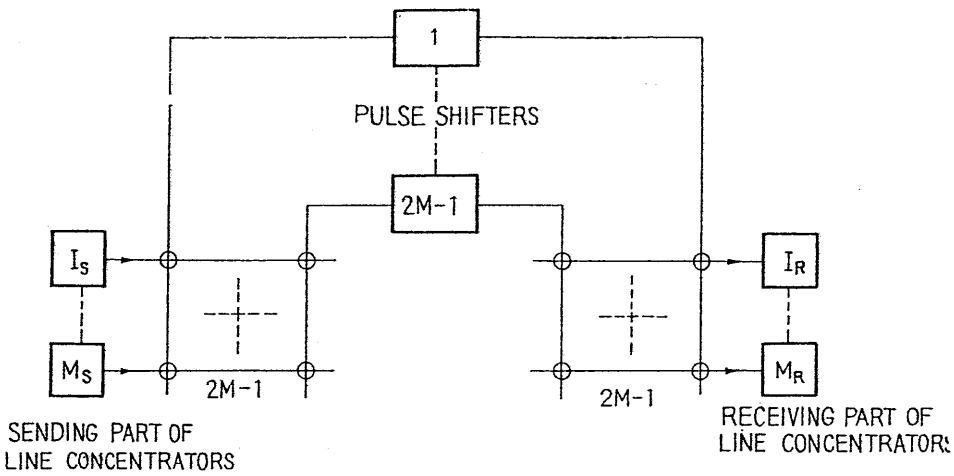


FIG. 2*b*.

3. Three-Stage Network with Pulse-Shifting-Junctors

The network of Figure 2*b* though strictly free from the time slot mismatch, is not quite economical since the traffic efficiency of the junctors are reduced by $M/M-1$ as

is compared with that of the highways. For the economical realization of the network with time slot interchange capability, it is essential to reduce the number of junctors to such extent that the time slot mismatch loss is essentially negligible if not strictly eliminated. Let us assume that the number of the junctors in Figure 2*b* is reduced to t.

The junctor is divided by the pulse shifter into two parts, namely the parts at the input and at the output of the pulse shifter which may be called the primary and the secondary parts of the junctor respectively. In setting up a connection, a pair of junctors are involved in such a way that the same time slot is used in the primary part of a junctor and the secondary part of the junctor in pair. The send and the receive highway of a line concentrator is also used in pair and thus the same time slot is used in transmitting speech to and from a subscriber. The originating and terminating subscribers, however, can use different time slots, which are interchanged by the pulse shifters.

To establish a connection, a commonly idle time slot must be found, in the first place, between the send highway of the originating line concentrator and the primary part of a junctor. The procedure may be repeated on several or all of the junctors connecting the originating and the terminating line concentrators. If the number of the repetition is limited to T, this may be called as taking primary junctor matching T times. On the other hand, if a commonly idle time slot is not obtained between the receive highway of the terminating remote line concentrator and the secondary part of the junctor, the primary part of which is matched to the send highway of the originating remote line concentrator, the junctor must be abandoned and the procedure of matching by another junctor should be performed again including primary junctor matching. The number of repetition caused by the mismatch between the secondary part of a junctor and the receive highway is called the number of secondary junctor matching, the maximum number of which may be limited to S ($S \leqq T$).

The blocking probability formula is derived as follows assuming that the distribution of calls on the highways and the each part of the junctors are independent and random and that time slot assignment is also random.

$$B' = \sum_{U=0}^{S} \binom{T}{U} \sum_{i=0}^{n} \frac{\dfrac{a_1^i}{i!}}{\sum\limits_{\nu=0}^{n} \dfrac{a_1^\nu}{\nu!}} \left\{ 1 - \sum_{j=n-i}^{n} \frac{\dbinom{i}{n-j}}{\dbinom{n}{j}} \frac{\dfrac{a_2^j}{j!}}{\sum\limits_{\nu=0}^{n} \dfrac{a_2^\nu}{\nu!}} \right\}^U$$

$$\times \left\{ \sum_{j=n-i}^{n} \frac{\dbinom{i}{n-j}}{\dbinom{n}{j}} \frac{\dfrac{a_2^j}{j!}}{\sum\limits_{\nu=0}^{n} \dfrac{a_2^\nu}{\nu!}} \right\}^{T-U}$$

$$\times \sum_{k=0}^{n} \frac{\dfrac{a_3^k}{k!}}{\sum\limits_{\nu=0}^{n} \dfrac{a_3^\nu}{\nu!}} \left\{ \sum_{j=n-k}^{n} \frac{\dbinom{k}{n-j}}{\dbinom{n}{j}} \frac{\dfrac{a_2^j}{j!}}{\sum\limits_{\nu=0}^{n} \dfrac{a_2^\nu}{\nu!}} \right\}^U$$

$$+ \sum_{U=S+1}^{T} \binom{T}{U} \sum_{i=0}^{n} \frac{\frac{a_1^i}{i!}}{\sum_{\nu=0}^{n} \frac{a_1^\nu}{\nu!}} \left\{ 1 - \sum_{j=n-i}^{n} \frac{\binom{i}{n-j}}{\binom{n}{j}} \frac{\frac{a_2^j}{j!}}{\sum_{\nu=0}^{n} \frac{a_2^\nu}{\nu!}} \right\}^U$$

$$\times \left\{ \sum_{j=n-i}^{n} \frac{\binom{i}{n-j}}{\binom{n}{j}} \frac{\frac{a_2^j}{j!}}{\sum_{\nu=0}^{n} \frac{a_2^\nu}{\nu!}} \right\}^{T-U}$$

$$\times \sum_{k=0}^{n} \frac{\frac{a_3^k}{k!}}{\sum_{\nu=0}^{n} \frac{a_3^\nu}{\nu!}} \left\{ \sum_{j=n-k}^{n} \frac{\binom{k}{n-j}}{\binom{n}{j}} \frac{\frac{a_2^j}{j!}}{\sum_{\nu=0}^{n} \frac{a_2^\nu}{\nu!}} \right\}^S \tag{3}$$

where n, a_1, a_2, a_3 are the same to that in Eq. (1).

Numerical calculation of the blocking probabilities is performed by means of an electronic computer under various conditions. Figure 3 is an example of the results in case of $a_1 = a_2 = a_3 = a$ and $n = 20$. The blocking probabilities of the conventional network, the network with the pulse-shifting-junctors and the network free from time slot mismatch are shown by the chained lines, the solid lines and the broken line respectively. It is known from the figure and other results of calculations that the internal blocking is essentially eliminated by performing the junctor matching four times [5]. Figure 4 shows one of the comparisons made between the results of calculations and the results of computer simulation in the case of $n = 20$ and $T = 4$. It is known that the assumptions made in deriving the blocking probability formula generally gives rather safety side results especially in the case of the conventional network. The method of simulation will be described in the succeeding section.

4. Three-Stage Network with Pulse-Shifting-Junctors as Alternate Route

Considering the fact that the majority of cases which require the time slot interchange are such that a commonly idle time slot is unavailable between both highways involved, the network in Section 3 may be modified more economically by providing the pulse shifters to only a limited number of junctors and utilize the pulse-shifting-junctors as the alternate route.

In setting up a connection, a pair of the direct junctors without the pulse shifters are matched in the first place to the highways of the originating and the terminating line concentrators and if the first trial fails, the procedure is repeated on other direct junctor pairs. The repetition may be limited to T_1. If the direct junctor matching is unsuccessful, the matching by means of the alternate pulse-shifting-junctors is performed, the repetition of which may be limited to T_2. If a commonly idle time slot between the highways of the originating and the terminating line concentrators is unavailable, the alternate junctor matching is performed immediately without performing the direct junctor matching. The number of repeated matching may be limited to T_3. In the case of intra-concentrator

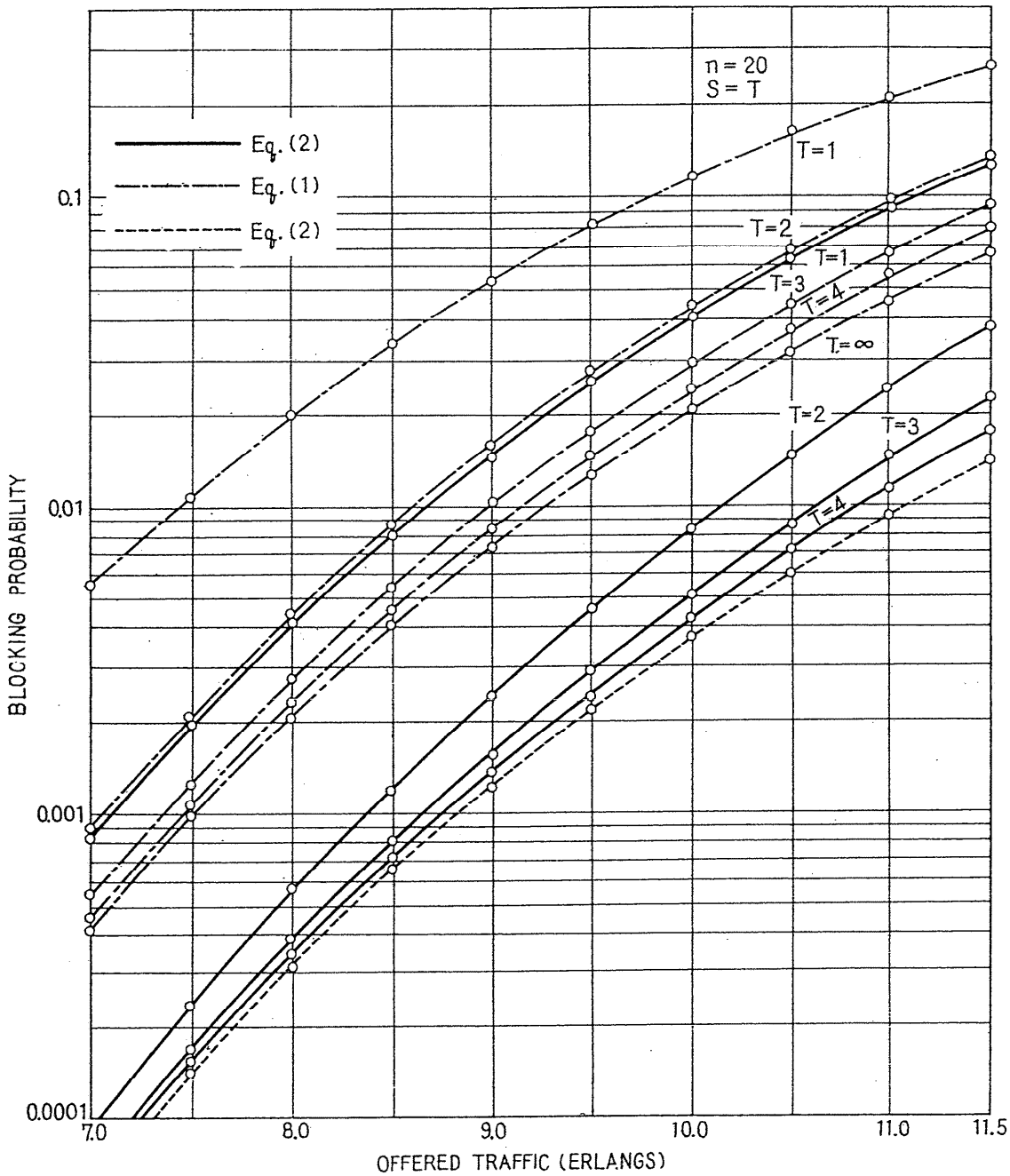


FIG. 3.

connection, two time slots must be assigned and interchanged by an alternate junctor pair. The number of repeated matching may be limited to T_4

The system described above is evidently an alternate routing system and therefore, the theoretical consideration is quite difficult as the statistical properties of the overflow traffic to the alternate junctors should be taken into account. For this reason simulation studies by means of IBM 7090 computer have been performed. Another purpose of the simulation has been the evaluation of the Eqs. (1) and (2), an example of which has already been shown in Figure 4.

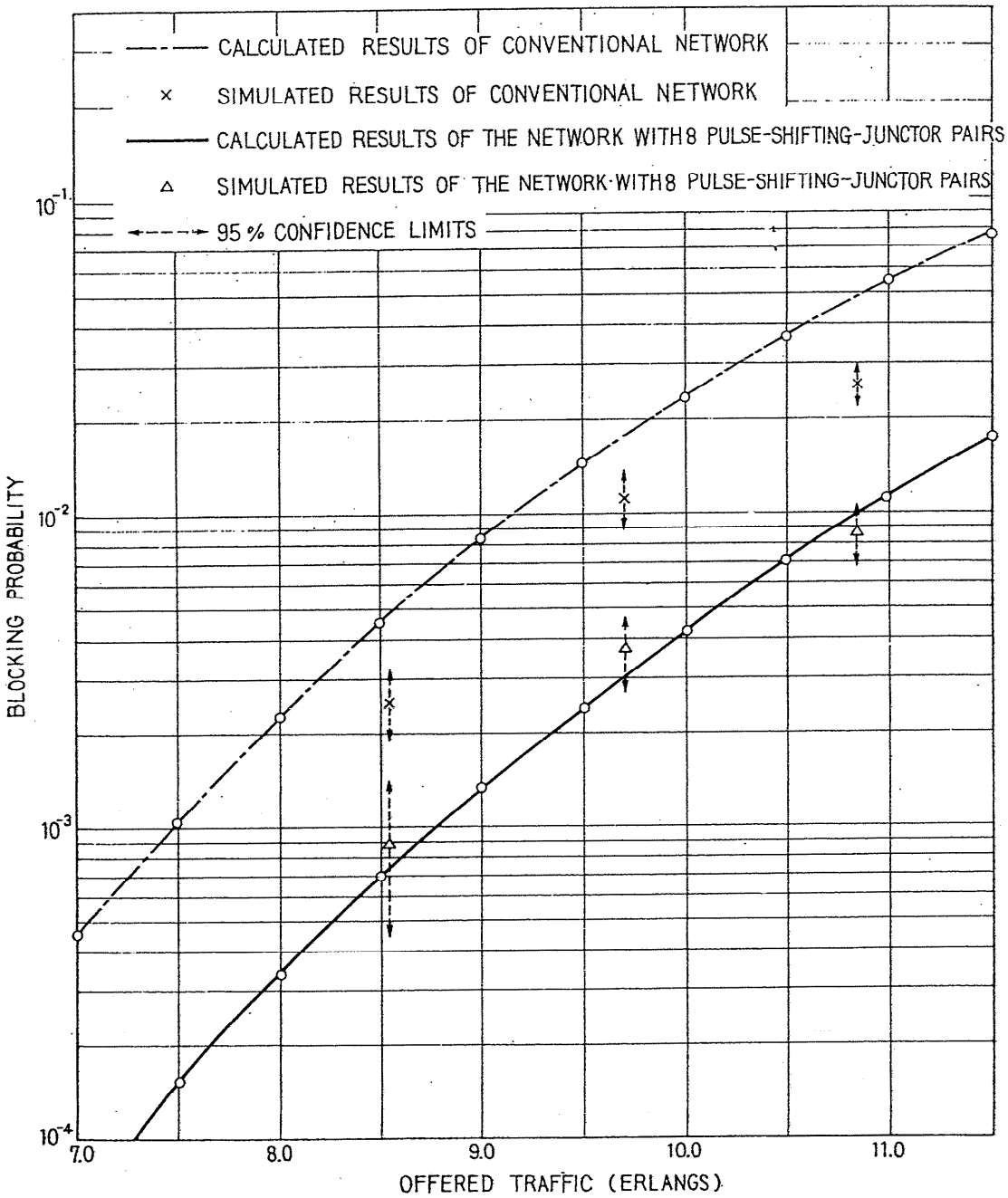


FIG. 4.

Assuming the Poisson arrival and the negative exponential holding time distribution, the simulation is performed by the Markov chain method [6]. Uniformly distributed pseud random numbers Q generated by the multiplicative congruential method are so arranged as to specify the origination of a call to a particular remote line concentrator, the termination of a call occupying a particular time slot of a particular line concentrator. In addition to this in the case of the origination of a call, another independent pseud random number R is generated to specify a particular remote line concentrator to which the originated call is directed.

As an illustrative example, Figure 5 shows the flow chart of the simulation on the network with the alternate pulse-shifting-junctors [7]. If the originating call is an intra-concentrator call up to T_4 junctor matching on the pulse-shifting-junctors is performed. And if it is an inter-concentrator call, testing is made in the first place whether a commonly idle time slot between the both highways exists. If exists, up to T_1 junctor matching on the direct junctors is performed, and if this fails, up to T_2 junctor matching on the pulse-shifting-junctors is performed. If not, up to T_3 junctor matching on the pulse-shifting-junctors is performed. In the case of the termination of a call, the numbers of the concentrator and the time slot are computed and if the time slot in actually occupied, the members of the concentrator, the junctor and the time slot involved in setting up the connection are read out, the time slots are released, and the contents of the memory are up dated accordingly.

The simulation is performed on a system with 16 concentrators, 8 junctors, 127 subscribers per each concentrator and 20 time slots per frame, and on 50,000 calls per each test. The calling rate of each idle inlet is assumed as 0.03605, 0.04130 and 0.04675 which corresponds to the offered traffic of 8.540, 9.690 and 10.560 erlangs per each concentrator respectively. An example of the results of simulation on the conventional network and

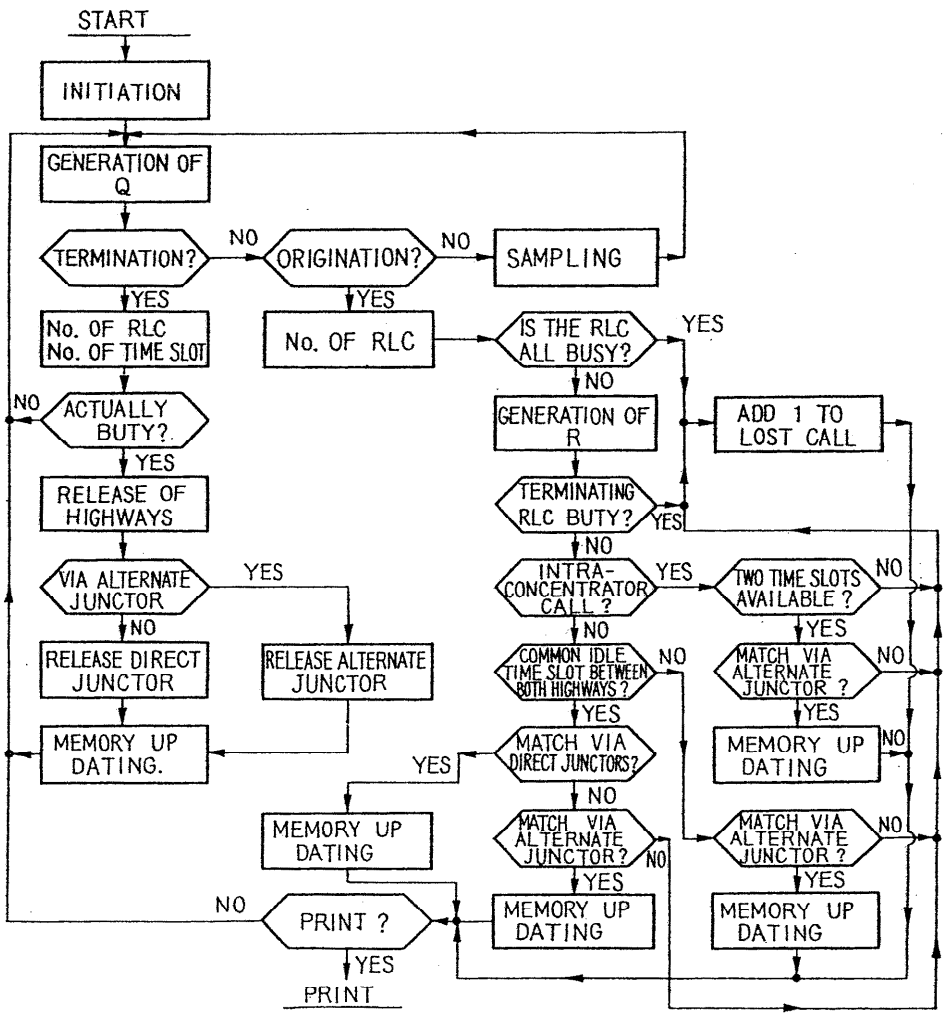


FIG. 5. — Flow chart of simulation on network with pulse-shifting-junctors as alternate route.

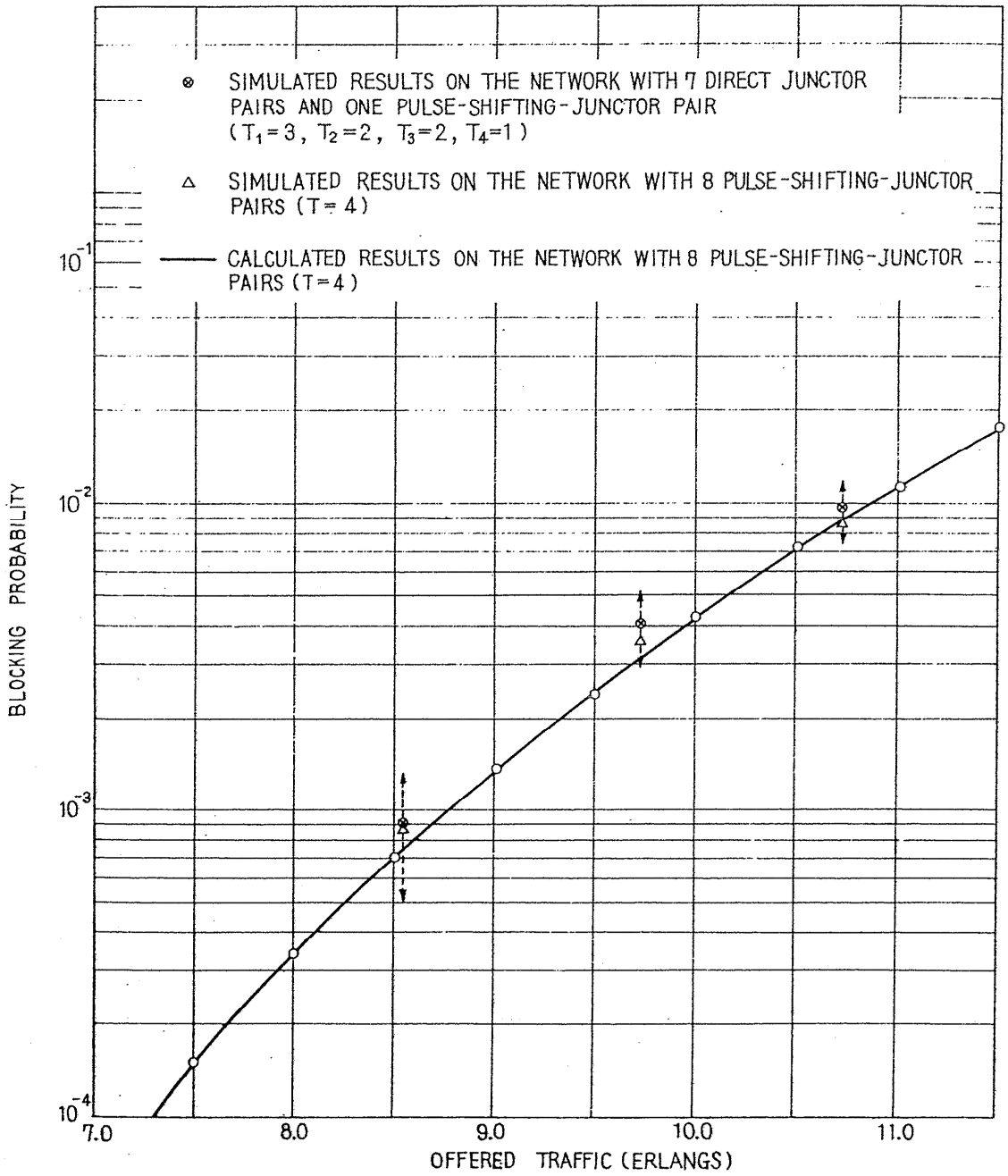


FIG. 6.

on the network with pulse shifters on all of the junctors has been shown in Figure 4. In the case of the network with the pulse-shifting-junctors as the alternate route, various configurations are possible as to the number of the pulse-shifting-junctor pairs and the number of the junctor matching. A preliminary simulation has been performed to find out a preferable configuration and as the result, it has been known that the traffic efficiency of the network with a pair of pulse-shifting-junctors is sufficiently high if the junctor matching is performed as $T_1 = 3$, $T_2 = 2$, $T_3 = 2$ and $T_4 = 1$. A result of simulation on this case is shown in Figure 6. It is known from the figure that the network essentially attains more economically the traffic efficiency of the network with the pulse shifters

on all the junctors and performing four junctor matching. It is also known from the experiment that the traffic carried by the pulse shifting-junctor is about 70 % of that carried by the direct junctor and about 70 % of the traffic carried by the pulse-shifting-junctor is the intra-concentrator traffic. This reveals that a pair of the pulse-shifting-junctors provided as the alternate route is generally suffice in essentially eliminating the time slot mismatch loss with such additional features as the simplification of intra-concentrator connection. Additional pulse-shifting-junctors may be provided especially in such cases that the traffic to be carried by the junctors is higher than that of the highways.

Acknowledgments

The authors wish to acknowledge with thanks the discussions of the members of the High Frequency Laboratories, the University of Tokyo and the sponsorship of the Bell Telephone Laboratories, Inc.

REFERENCES

[1] VAUGHAN H.E., « Research Model for Time Separation Integrated Communication », *B.S.T.J.*, Vol. 138, pp. 909-932 ; July 1959.

[2] INOSE H., « Blocking Probability and Junctor Efficiency of a Three Stage Time-Division Switching Network », *Journal of the Institute of Electrical Communication Engineers of Japan*, Vol. 44, pp. 935-941 ; June 1961.

[3] CLOS C., « A Study of Non-Blocking Switching Network », *B.S.T.J.*, Vol. 32, pp. 406-424 ; March 1953.

[4] INOSE H., YOSHIDA Y., YASUDA Y. and KOONO Z., « A Time Slot Interchange System in Time-Division Electronic Exchanges », *Trans IEEE*, Vol. CS-11, pp. 336-345 ; September 1963.

[5] INOSE H. and SAITO T., « A Three-Stage Time-Division Switching Network by Time-Slot Interchange Principle, » *Journal of the Institute of Electrical Communication Engineers of Japan*, Vol. 49, n° 3, pp. 409-417 ; March 1966.

[6] WALLSTROM B., « Artificial Traffic Trials on a Two-State Link System Using a Digital Computer », *Ericsson Techniques*, Vol. 14, n° 2, pp. 259-289 ; 1958.

[7] INOSE H., SAITO T. and KATO M., « Three-Stage Time-Division Switching Network including Alternate Junctors having Time Slot Interchange Capability », to be published in the *Journal of the Institute of Electrical Communication Engineers of Japan.*

Imprimerie Marcel Bon, Vesoul — D. L. 1296.22-3-1967

AN OPTIMUM DESIGN THEORY FOR PCM TOLL SWITCHING NETWORKS

By Tadao Saito
Department of Electrical Engineering, University of Tokyo
7–3–1, Hongo, Bunkyo–ku, Tokyo, 113 Japan

ABSTRACT

This paper presents a design theory to obtain the optimal channel graph characterizing a time division switching network using practical switch sizes and satisfying specified blocking probability conditions.

A method for composing a channel graph which gives lower blocking probability is described. For networks using space switches smaller than 16 x 16, the superiority of T–S–T type structures over S–T–S type structures is shown. Various pair relationship schemes for 4-wire time division connections are described in relation to the constraints on channel graphs.

In this design procedure, the network is specified by the number of PCM lines to be switched and by the blocking probability at a specified occupancy. This theory generates a limited number of feasible switching networks from the number of PCM lines to be switched and predicts the relative value of blocking probability. By calculating the blocking probability for a limited number of candidates, a network with minimum cost satisfying the given blocking probability constraint can be selected.

1. INTRODUCTION

PCM time division switching system has started to have economical advantage over conventional space division switching system due to the development of recent electronic circuit technology.(1) Especially in the application for toll switching to handle toll pulse code modulation (PCM) transmission circuits, the advantage of PCM time division switching system is apparent. In toll PCM switching system, the time division multiplexity of the PCM system to be handled is thought to be rather high, e.g., 128 channels or 256 channels.

PCM switching networks are composed of time slot interchange (TSI: T-stage) and time multiplexed switches (TMS: S-stage). Networks can be classified like time-space-time (T–S–T) types of networks or space-time-space (S–T–S) types of networks based on the combination of switches. Many networks have been studied until recently mainly for the purpose of local switching networks using relatively low speed PCM systems.(2,3)

Denoting the time division multiplexity of PCM by t, one time slot interchange corresponds to a t by t crossbar switch. This means that in highly multiplexed PCM switching network, large crossbar switches are easily realizable using T-stage. In the switching system for toll network using high speed PCM systems, networks should be designed to take advantage of this fact.

The optimum design theory described in this paper tries to give the optimum network using highly multiplexed PCM system. The actual size of each switch is practical one, i.e., a power of two as is usually used in electronic switching systems.

The relationship between the channel graph and the actual switching network is discussed. Then rules to select channel graphs to give lower blocking probability, keeping the number of crosspoints constant, are discussed. T–S–T types of networks and S–T–S types of networks are compared from this viewpoint, and it is concluded that T–S–T types of networks are advantageous in time division switching network using highly multiplexed PCM system.

Pair relationship (the method of connecting originating and terminating parties) is important in time division switching networks. Various pair schemes are discussed and requirements on the channel graph to satisfy the specified pair relationship are described.

If a network specification is given in terms of the total number of trunks, the number of time slots and the blocking probability at a specified occupancy, several channel graphs of networks which can accommodate the specified number of trunks can be constructed. Each channel graph gives different cost, but minimum blocking probability for the cost. It is also possible to predict the relative value of blocking probability for some of these networks. Thus, by evaluating the blocking probability for some of the networks, it is possible to reject candidates which do not satisfy blocking probability specifications. The network with the lowest cost which satisfies the blocking probability requirement is the optimum network.

2. EXPRESSION OF CHANNEL GRAPH

Denoting the number of time slots by t, a time slot interchange corresponds to a t by t crossbar switch. A time multiplexed switch with n inputs and m outputs corresponds to t, n by m crossbar switches. Therefore, time division switching networks can be considered to be a kind of crossbar bar type network.

In typical crossbar type switching networks, the function of each of the stages is expressed in terms of a channel graph. Figure 1 (b) is an example of a channel graph for a 6-stage crossbar network shown in Figure 1 (a). The function of stages in a crossbar type switching network can be classified into the following classes.(4)

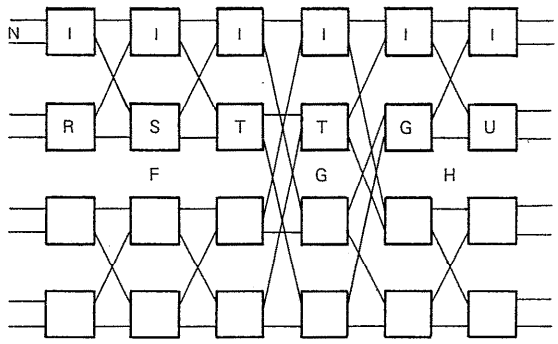
Distribution stage: The stage having the function of increasing the number of available paths to a given destination. Corresponding to this stage, a collection stage should be provided in the network.

Routing stage: The stage having the function of increasing paths without a corresponding collection stage. The paths separated by this stage are directed to different output lines. In terms of a channel graph, this stage is expressed by a node connecting an inlet link to only one outlet link.

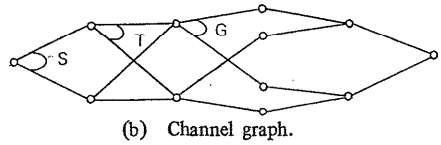
Collection stage: The stage having the function of collecting distributed paths.

Collection-distribution stage: The stage having the function of collection and distribution.

In some networks, two or more collection stages may be used for one distribution stage, or two or more distribution stages may be provided for one collection stage. In the simple example of a network shown in Figure 2, two of the paths distributed by stage 1 are collected by stage 3 and the other paths are collected by stage 4. The paths collected by stage 4 are distributed by stages 1 and 2. In such a case, the channel graph may be expressed by specifying the stage of distribution and collection and the number of paths distributed and collected.



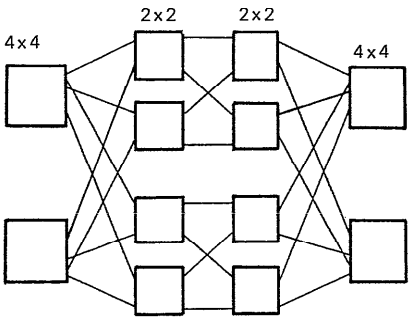
(a) Network configuration.



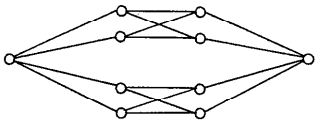
(b) Channel graph.

Fig. 1. An example of six stage network.

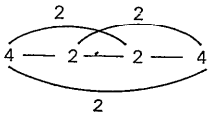
Denothing the number of paths distributed at stage i and collected by stage j by $l_{i,j}$, the network shown in Figure 2 is expressed by $l_{1,3} = 2$, $l_{1,4} = 2$, $l_{2,4} = 2$. If the switch in each stage is square, each stage of the network can be specified by the size of the switch like 4–2–2–4 for the example shown in Figure 2. The distribution and collection between stages can also be given in simplified manner as is shown in Figure 2. (c).



(a) Network configuration.



(b) Channel graph.



$l_{1,3} = 2,\ l_{2,4} = 2,\ l_{1,4} = 2$

(c) Symbolic notation. 2 2 4–2–2–4. 2

Fig. 2. Network expressed by

3.1 Number of Distribution

Denoting the total number of stages by σ, total number of distribution at stage i is:

$$l_{i\cdot} = \prod_{j=i+1}^{\sigma} l_{i,j} \quad \text{.....} \quad (1)$$

and the total number of collection at stage j is:

$$l_{\cdot j} = \prod_{i=1}^{j-1} l_{i,j} \quad \text{.....} \quad (2)$$

Denoting the size of a stage i switch as n_i by m_i:

$$l_{i\cdot} \leq m_i, \qquad l_{\cdot i} \leq n_i \quad \text{.....} \quad (3)$$

The total number of distribution in the network is denoted by L and defined as:

$$L = \prod_{i=1}^{\sigma-1} l_{i\cdot} = \prod_{j=2}^{\sigma} l_{\cdot j} = \prod_{i=1}^{\sigma-1} \prod_{j=i+1}^{\sigma} l_{i,j} \quad \text{.....} \quad (4)$$

The number of inlets and outlets for a network which may be denoted by N_1 and M_σ respectively is given by:[5]

$$N_1 = \prod_{i=1}^{\sigma} n_i/L \quad \text{.....} \quad (5)$$

$$M_\sigma = \prod_{i=1}^{\sigma} m_i/L \quad \text{.....} \quad (6)$$

3.2 Number of Time Division Crosspoints

A time multiplexed switch is a rectangular array of time shared space-division switches. Denoting the number of input and output buses of the stage i switch by n_{is} and m_{is} respectively, and the number of time slots at this stage by t_i, a time multiplexed switch is equivalent to t_i crossbar switches having the dimension of n_{is} by m_{is}. Let us denote the expansion ratio at stage i in terms of number of time division busses by C_{is},

$$C_{is} = m_{is}/n_{is}$$

Let us denote the total number of time division buses at the output of stage i by M_{is}, and the total number of time division buses at the input of the switching network by N_{1s}. The total number of time division buses is:

$$M_{is} = N_{1s} \prod_{j=1}^{i} C_{js} \quad \text{.....} \quad (7)$$

Total number of time division crosspoints for time multiplexed switch is:

$$X_p = \sum_{i=1}^{\sigma} n_{is}\, m_{is} \ (i\text{: for time multiplexed switches}) \quad \text{.....} \quad (8)$$

From Equations (5), (6) it is apparent that:

$$N_{1s} = \prod_{i=1}^{\sigma} n_{is}/L_s \quad \text{.....} \quad (9)$$

$$M_{\sigma s} = \prod_{i=1}^{\sigma} m_{is}/L_s \quad \text{.....} \quad (10)$$

where L_s is total number of distribution by time multiplexed switches. Therefore, if N_{1s}, $M_{\sigma s}$, and L_s are given, the products of each switch's inlets and outlets are given:

$$N_{1s} L_s = \prod_{i=1}^{\sigma} n_{is} \quad \text{.....} \quad (11)$$

$$M_{\sigma s} L_s = \prod_{i=1}^{\sigma} m_{is} \quad \text{.....} \quad (12)$$

3.3 Subnetwork Structure

An arbitrarily selected continously arranged sequence of switches can form a subnetwork. Usual telephone switching networks are composed of subnetworks. Denoting the number of distribution originated and terminated within the subnetwork by L_g, the total input (N_g) and the total output (M_g) of a subnetwork composed of stage h through stage k is given by:

$$N_g = \frac{n_h\, n_{h+1} \cdots\cdots n_k}{L_g} \quad \text{.....} \quad (13)$$

$$M_g = \frac{m_h\, m_{h+1} \cdots\cdots m_k}{L_g} \quad \text{.....} \quad (14)$$

The total number of subnetworks is:

$$\text{total number of subnetworks} = g = \frac{N_h}{N_g} = \frac{M_k}{M_g}$$

From Equations (5) and (13):

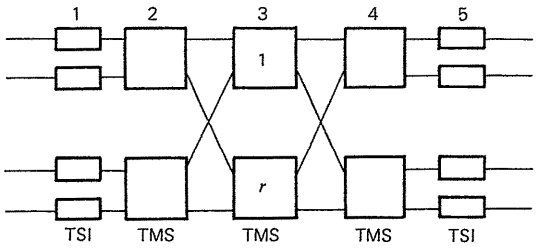
$$\text{total number of subnetworks} = (n_{k+1} \cdots n_\sigma)(m_1 \cdots m_{h-1}) \frac{L_g}{L} \quad \text{.....} \quad (15)$$

A subnetwork is a complete network in a sense that any input can be connected to any output. In a given network, two subnetwork stages may be formed. If d links exist between two subnetwork stages, the number of distribution starting at one subnetwork and terminating at the other subnetwork is d.

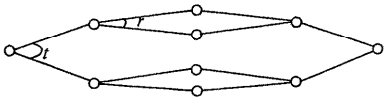
4. CHANNEL GRAPHS OF TYPICAL TIME DIVISION NETWORK

Since the invention of PCM switching system, various time division networks have been discussed. Typical among these are T–S–T and S–T–S types of networks. Figure 3 shows a T–S–T type network including three S-stages. The channel graph of the network is shown in Figure 3 (b). In this network, it is apparent

that the first T-stage switch distributes a path into t paths and the last T-stage switch collect t paths. The second switch is S-stage and the r paths distributed at the stage 2 are collected by the stage 4. The stage 3 is a routing stage.



(a) Network configuration.



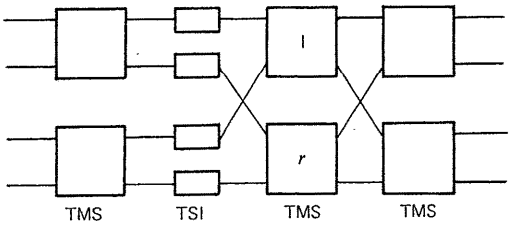
(b) Channel graph.

Fig. 3. A T–S–T type network having total distribution of tr.

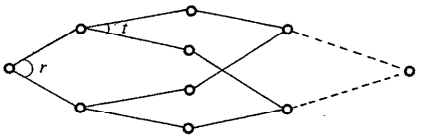
Considering the fact that the time slot interchange at the stage 1 provides the access to any time slot within the network, and the time slot interchange at stage 5 connects the arbitrarily selected time slot to a specified PCM channel at the output of the network, the collection stage corresponding to a distribution obtained by an S-stage must be collected by an S-stage.

Figure 4 shows an S–T–S type network, having the total distribution of tr. In this network, the stages 1, 3, and are S-stages and stage 2 is a T-stage. The channel graph is shown in Figure 4 (b). In this network, there is no collection stage corresponding to the T-stage, however, so long as the signal in all the time slots in one output PCM bus is directed to the same location, there should be a time slot interchange or an equivalent device at the destination location, so that we can conceive an imaginary collection stage at the end of the network.

Various kinds of S–T–S types of networks can be designed, however, common characteristics of S–T–S types of networks are that the first distribution is given by a space stage, and the distribution is collected before the collection of distribution by the T-stage. On the other hand, in T–S–T types of networks, the first distribution and the last collection are given by T-stages. In terms of channel graph, T–S–T networks are characterized by the fact that the distributions by T-stages and S-stages are separated, S–T–S networks are characterized by the fact that distributions by T-stages and S-stages are intermixed.



(a) Network configuration.



(b) Channel graph.

Fig. 4. An S–T–S type network having total distribution of tr.

5. CHANNEL GRAPH TO GIVE LOWER BLOCKING PROBABILITY

1. In the channel graph shown in Figure 5 (a) the stage 1 gives a distribution of l_1 which is collected by stage 5. The stage 2 gives a distribution of l_2 which is collected by stage 4. In the channel graph shown in Figure 5 (b), l_1 and l_2 are interchanged. In such a graph, if $l_1 > l_2$ the graph shown in Figure 5 (a) gives lower blocking probability if the occupancy of each link in two networks is identical. In other words greater distribution should start eariler and should end later. This is apparent because total number of available links at stage 3 is l_1 l_2 in both networks, but in the case of (a) the number of available links at the output of stage 1 and at the input of stage 5 is greater than the case of (b).

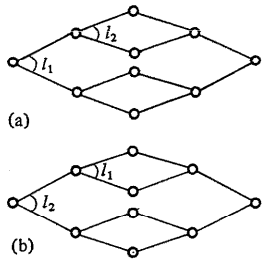


Fig. 5. Comparison of channel graphs–1.

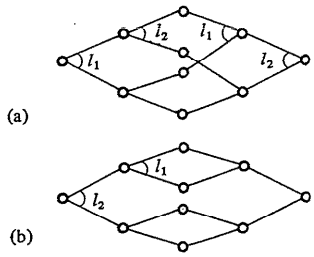


Fig. 6. Comparison of channel graphs–2.

2. In the case of the channel graph shown in Figure 6 (a), if $l_1 > l_2$ and the occupancy of each link in two networks is identical, the graph shown in Figure 6 (a) gives lower blocking probability. This is apparent because if a link between stages 1 and 2 is busy, all the links connected to the link cannot be used in the case of (b), whereas in the case of (a), even in such a case at least the link between stages 4 and 5 can be used to get an alternate route.

3. However, if $l_2 \gg l_1$, the network shown in Figure 6 (b) may give lower blocking probability.

Figure 7 shows an examples of blocking probability comparison between two types of networks. The blocking probability is obtained assuming binomial distribution of calls at each stage.

Figure 7 shows the blocking probability of types (a) and (b) networks of Figure 6 when l_2 = 128. The occupancy is assumed to be 0.9 multiplied by 105/128.[1] The figure show that if l_1 is

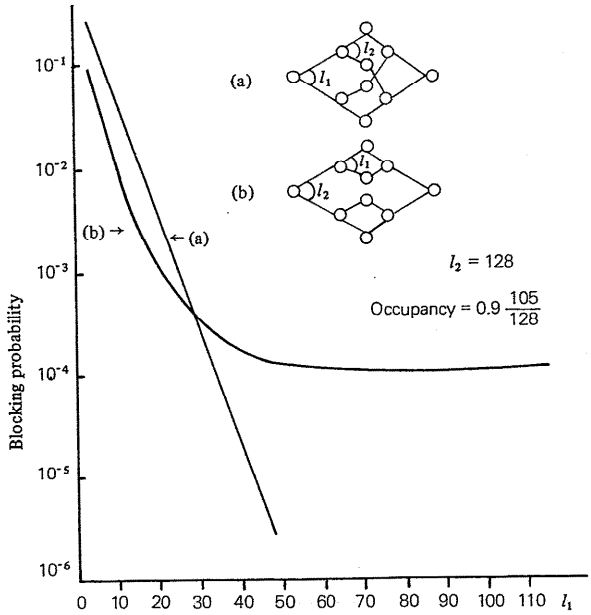


Fig. 7. Comparison of two types of networks when l_2 = 128.

ess than 30, type (b) network will give lower blocking probability.

The value of l_1 at which the blocking probability of the two types ›f network crosses over depends upon the value of l_2 and the ›ccupancy. If the value of l_2 is the greater, the value of l_1 at the ›rossover point increases.

Similar comparisons were made for the cases l_2 = 64, 128 and 256 s well, for heavy link occupancy. Generally speaking, if the value of ı is 16, the network of type (b) is better for the value of l_2 = 64, 28 and 256. If the value of l_1 is 32, the preference of the network epends upon the value of l_2 and the occupancy. As has been escribed in Section 4, corresponding l_2 to t and l_1 to r the channel raph shown in Figure 6 (b) typically corresponds to T–S–T type of ıetworks and the channel graph shown in Figure 6 (a) correponds to S–T–S type of networks.

In time-division switching network with highly multiplexed PCM ıses, it is relatively easy to obtain a large switch such as 128 by 28 or 256 by 256 in the case of time slot interchange. On the other ıand, because of the limit of fan-out of each electronic gate circuit nd to reduce the cost of the switch array, too big a space switch is ıot desirable. Usually the maximum space switch size available is imited to 16 by 16 or 32 by 32.

This means that the channel graph of time division switching etwork using highly multiplexed PCM bus should be as is shown ı Figure 5 (a), i.e., the network should be T–S–T type.

From this reason, the scope of consideration in the following iscussion is limited to T–S–T type networks. Since the design of he T-stage is fixed, the remaining problem for network design is he design of the S-stage. The first stage of the space network is tage 2 and the last stage of the space network is the stage σ-1. 'he total number of distribution in a network, and the number of ıbnetworks are discussed in terms of the space network.

6. PAIR RELATIONSHIP IN TIME DIVISION SWITCHING NETWORK

In a 4-wire time division network, to set up a connection, two aths, i.e., the path from the originating party to the terminating arty and the path from the terminating party to the originating arty, are needed. The former connection may be called a forward onnection and the latter connection may be called a backward onnection.

So called mirror image or pair relationship is a way to set up the ackward connection automatically with the forward connection y giving a particular relation between connection links in the etwork. By this relation, path hunting is needed only once to set p a pair of connections and the blocking probability is reduced. ince a pair of gates can be controlled by a single control signal, ontrol signal memory and the decoder gates to control transmis- on path can be reduced in many cases.

.1 Combined Network and Separated Network

Pair relationship can be classified into several classes. The first lassification is combined network and separated network shown ı Figure 8 (a) and (b) respectively.

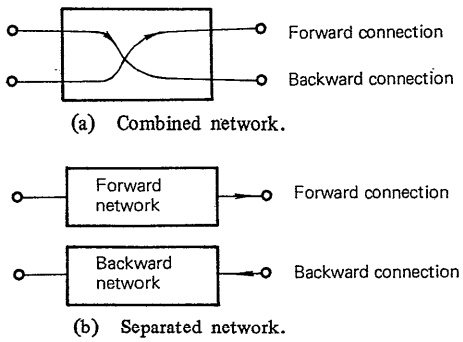


Fig. 8. Combined network and separated network.

In a combined network a pair of connection is established using ne network. On the other hand in the separated network two etworks having identical structure but directed reversely are used ı pair. One may be called the forward network; the other may be called the backward network. In this network a connection between an input to the forward network and an output from the backward network is impossible. Therefore, if there are 2-way trunks to be connected, the separated network will cause inefficiency.

6.2 Concurrent Pair and Noncurrent Pair

The second classification is given by the difference that pair should be established between concurrently operating elements or noncurrently operating elements in the network. Figure 9 shows an example of a network consisted of two TSI stages and three TMS stages. Figure 9 (a) shows a concurrent pair network. When the TSI 0 and TSI 3 are to be connected, time slot i is used in all the links in both of forward and backward connections. In the nonconcurrent pair network, if the time slot i is used in the forward connection, in the backward connection the time slot $(i + \beta)$ is used, where β is a fixed integer.

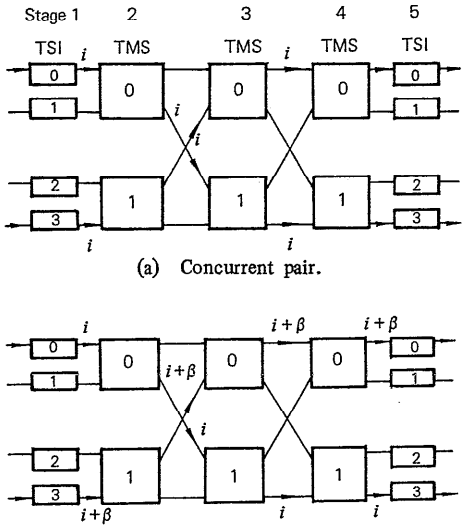


Fig. 9. Concurrent pair and non-concurrent pair.

A problem of concurrent pair is the intra-TSI connection, to connect identically numbered TSIs. To accomodate this, two independent connections should be established between the identically numbered input TSI and the output TSI.

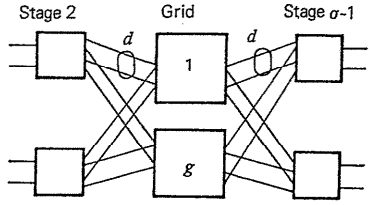
In this sense, nonconcurrent pair is the more flexible and powerful solution; however, this scheme requires that signals to control the TMS and TSI should be read out in two different time slots in each frame. This means that the operating speed of every control memory must be doubled. To cope with a small percentage of calls, this will be a large sacrifice, therefore non-concurrent pair is an expensive solution for a large scale network.

6.3 Intragrid Pair and Intergrid Pair

Discarding the possibility of separated networks and non-concurrent pair networks, we can continue our discussion to combined concurrent pair networks. If the space stage consists of more than three stages of switches, we can classify the network by intragrid pair and intergrid pair.

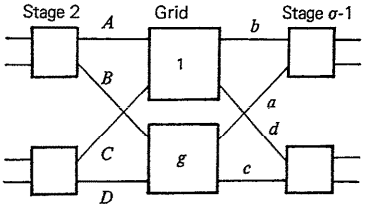
We can set up a subnetwork in any part of a time division network. When a subnetwork is formed in the center of TMS network, it may be called a grid. For the convenience of discussion let us define the grid as a subnetwork which is formed in the space network by removing the first and the last stage of the given space network. Each of first and last stage switches must be connected to each of grids.

As is shown in Figure 10 (a), if there are at least two links between each first stage switch and each grid and each grid and each last stage switch, and the connection pattern between first stage switches and grids and grids and the last stage switches is symmetric, each pair of connections can be completed within a grid if a grid has a sufficient structure to establish a pair of connections. Such a pair scheme is called intragrid pair.



(a) Intragrid pair.

$l_{2,g} = l_{g,\sigma\text{-}1} = d \quad l_{2,\sigma\text{-}1} = g$



(b) Intergrid pair.

$l_{2,\sigma\text{-}1} = g$

Fig. 10. Intragraid pair and intergrid pair.

Denoting the number of paths between a first stage switch and a rid by d, since a grid is a complete network, the number of distri- ution and concentration between the first stage and the grid which ıay be denoted by $l_{2,g} = d$. Since the network is symmetric $_{,\sigma\text{-}1} = d$. Each of first stage switches and last stage switches are onnected to each grid, therefore denoting the number of grid by , $m_2 = n_{\sigma\text{-}1} = gd$. This further means that $l_{2,\sigma\text{-}1} = g$.

In the intergrid pair system, a pair of grids is used to set up a pair f connections. As is shown in Figure 10 (b), link A is paired with nk a and link B is paired with link b and so on. In this case the umber of connection between a first stage switch and a grid and ıat between a grid and a last stage switch may be one. Thus, the equirement for the channel graph is $l_{2,\sigma\text{-}1} = g$.

Table 1 summarizes the classification. To design networks having ss than four space stages, concurrent pair system of intragrid pair r intergrid pair may be considered.

Table 1. Classification of pair system.

Combined network	Concurrent pair	Intragrid pair
		Intergrid pair
	Non-concurrent pair	
Separated network		

7. DESIGN OF TOLL STAGE TIME DIVISION SWITCHING NETWORK

As has been discussed in Section 4, for a time division switching network with highly multiplexed PCM system, T–S–T type configuration is advantageous. Deciding to adopt T–S–T configuration, the design problem is reduced to the design of S-stages.

In electronic switching system, switch size is usally selected to be a power of two. In this discussion also, the switch size is assumed to be a power of two. In this assumption, when concentration or expansion is to be performed within TMS, even if the value of C_{is} is as small as one-half or two, concentration or expansion will result in too high blocking probability or too low efficiency in toll stage systems. Thus, the value of C_{is} is assumed to be one. In other words all the switches are assumed to be square switches. If some concentration or expansion is needed, it can be accomplished at time stages at the input and the output of the network. Defining

$$L_s = L/t_1 \quad \text{.....} \quad (16)$$

L_s is the total number of distribution obtained at TMS. Denoting the number of time-division buses at the input and output of TMS by N_2 and $M_{\sigma\text{-}1}$ respectively. Since $C_{is} = 1$, $N_2 = M_{\sigma\text{-}1}$, and $m_i = n_i$, the total number of crosspoint in the TMS is:

$$X_p = \sum_{i=2}^{\sigma\text{-}1} n_i N_2 \quad \text{.....} (17)$$

If the value of N_2 is given and the value of L_s is assumed, the value:

$$\prod_{i=2}^{\sigma\text{-}1} n_i \quad \text{.....} \quad (18)$$

that is the product of all the switch sizes in TMS is obtained. From the value $L_s N_2$, the possible combination of n_i can be obtained.

Once a size of switch at each stage and the interconnection between them is obtained, a network is defined. The per trunk cost of crosspoints for the TMS is proportional to the sum of n_i and the per trunk cost of the control memory of the TMS is proportional to the sum of $\log_2 (n_i + 1)$. Therefore, it is possible to define the crosspoint cost measure (CM) and memory cost measure (MM) by:

$$\text{CM} = \sum_{i=2}^{\sigma\text{-}1} n_i \quad \text{.....} \quad (19)$$

$$\text{MM} = \sum_{i=2}^{\sigma\text{-}1} \log_2 (n_i + 1) \quad \text{.....} \quad (20)$$

If each of n_i is a power of two, MM is reduced to $\log_2 L_s N_2 + \sigma\text{-}2$. It means that the memory cost does not depend upon the selection of the value of n_i. To reduce the value of CM, the value of n_i should be selected as uniform as possible.

If the size of switch is thus obtained, the optimum connection pattern between the switches can be obtained by the discussion of Section 5. In these optimum networks, if the value of L_s is identical a network with a greater number of stages has the higher blocking probability. Thus, computing the blocking probability for a limited number of networks, it is possible to determine if a particular network satisfies the given constraint for blocking probability. Then a network with the least value of CM and MM satisfying the blocking probability requirement is selected. In the case of intragrid pair network some modification in the link pattern may be needed to cope with some types of connection.

8. DESIGN EXAMPLES

In this section let us design networks in which blocking probability is less than 12 percent for trunk occupancy of 0.9, in the ase that the number of time slots is 128. Expansion of 105/128 assumed at the input of the network.[1]

Example 1

Let us design a network having 1024 time division buses with the multiplexity of 128. Pair relationship is assumed to be intergrid pair.

Assuming L_S = 32, 16 and 8, the factorization of the value $N_2\ L_S = 2^{15}$, 2^{14}, and 2^{13} will give the network shown in Table 2. The network which gives the least crosspoint cost measure and the least memory cost measure satisfying the blocking probability of 12 percent for the occupancy of 0.9 is the optimum network. If L_S = 16, a 5-stage netowrk (D) cannot satisfy the blocking probability requirement. This means that if the value of L_S is less than 16 and the number of stages is more than five, the network cannot satisfy the brobability requirement. Therefore it is certain that any network which is not listed in Table 2 is more expensive or does not satisfy the blocking probability requirement. This means that the least expensive network in Table 1 which satisfies the blocking probability requirement is the optimum network. Thus, the network (C) is obtained as the optimum network.

Example 2

Let us design a network having 1024 time-division buses with the multiplexity of 128. Pair relationship is assumed to be intragrid pair.

Since the condition for intragrid pair is more strict, the number of possible combinations of switch size decreases. Table 3 shows the possible network structures. Since the network (B) satisfies the conditon, the network (A) will give over quality. The optimum network (B) is the network for No. 4 ESS.

Table 2. Feasible networks for 1024 lines (intergrid pair).

	L_S	Network	CM	MM	g	Blocking Prob.
A	32	32 32–32–32	96	18	32	0.299×10^{-3}
B	32	4 4 8–8–8–8–8 2	40	20	2	0.295
C	16	16 16–8–8–16	48	18	16	0.108
D	16	4 8–8–4–8–8 2 2	36	19	4	0.533
E	8	8 8–4–8–4–8	32	18	8	0.723

Blocking probability is obtained assuming the occupancy of 0.9 and expansion of $\frac{105}{128}$.

Table 3. Feasible networks for 104 PCM lines (intragrid pair).

	L_S	Network	CM	MM	g	Blocking Prob.
A	64	4 16–16–16–16 4 4	64	20	4	0.332×10^{-2}
B	16	4 8–16–16–8 2 2	48	18	4	0.119

Blocking probability is obtained assuming the occupancy of 0.9 and expansion of $\frac{105}{128}$.

9. CONCLUSION

In this paper a design theory of time division switching network using highly multiplexed PCM time division buses is descussed. A general theory of the channel graph to give lower blocking probability is described and T–S–T types of networks are selected. Then the design theory of space stage networks is described in relation to the requirement for pair relationship. As practical pair relationship schemes, intragrid pair system and intergrid pair system are discussed.

Using the theory, networks with the multiplexity of 128 are designed. The optimum network obtained for intragrid pair system is identical to that of No. 4 ESS.

ACKNOWLEDGEMENT

Thanks are due to Mr. H. E. Vaughan, former director of Bell Telephone Laboratories and Mr. G. D. Johnson and Dr. J. N. Denenberg for their help in preparing the manuscript.

REFERENCES

(1) H. E. Vaughan: "An Introduction to No.4 ESS" ISS, 1972, p. 19, June, 1972.

(2) H. Inose, T. Saito and M. Kato: "Time Slot Interchange in Three-Stage Time Division Switching Networks", Colleque International de Communication Electronique, p. 555, March, 1966.

(3) H. Inose and T. Saito: "Four-Stage and Five-Stage Switching Networks in Integrated Communication Systems", Conference on Switching Techniques for Telecommunication Networks, p. 165, April, 1969.

(4) H. Inose, T. Saito, S. Asano and M. Ito, "A General Purpose Simulator for Crossbar Type Switching Networks", ICC, 1973 39–1, June, 1973.

(5) K. Takagi, "Optimum Channel Graph of Links System", Electronics and Communication of Japan, Vol. 54–A, No. 7, July, 1971.

TELETRAFFIC AND DATATRAFFIC
Socio-Economic Aspects, ITC-13
A. Jensen and V.B. Iversen (Editors)
Elsevier Science Publishers B.V. (North-Holland)
© 1992 IAC. All rights reserved.

Early Development History of Digital Time Division Switching Network Using Time Slot Interchange

Tadao SAITO*

*The University of Tokyo, Department of Electronic Engineering, Faculty of Engineering

Abstract

Present day digital switching system uses time division switching networks composed of space switches and time switches. In contrast to traditional space division switching networks based on space switches only, design of time division switching network includes complex problems.

Concept of time switch is invented in early 60's by Prof. H.Inose of the University of Tokyo and named time slot interchange. Since then, variety of ideas on switching network design were proposed. In this paper, variety of design of time deivision networks were surveyed histrically. Most of traditional ideas on network to reduce blocking probability became not very important at present due to rapid progress in high speed transmission technology. Nevertheless, if number of time division multiplexity is reduced due to increase in high bandwidth demand, these ideas will recover importances again in future.

I Introduction

When electronic switching systems were conceived in early days, one of major concern is how to realize economical switching fabrics using expensive electronic components. Time division multiplexed (TDM) switching principle was developed to share an expensive electronic component by a number of time division channels. Before 1960 many experimental PAM (pulse amplitude modulation) based TDM switching system was developed and a few systems were experimentally installed in the field. TDM switching system based on PAM principle was only applicable to very simple switching system because of difficulty to maintain transmission quality.

The most complex large scale TDM-PAM switching system which was tried at those period was British Highgatewood office system. The system has a number of PAM buses (highways) which are interconnected by a TDM switch array. PAM modulators are connected to a time division send buses and PAM demodulators are connected to a time division receive buses. Send buses and receive buses are connected by a square switch array. This network was called two stage time division network because two time division buses are used for interconnection[1].

In TDM system, large scale switching system can be realized by much simpler switching fabrics compared with crossbar switching fabrics. In evaluation of blocking probability, if the switch fabrics are simpler, traffic distribution in each of highway are requested to be modeled more precisely.

In time division switch to be used as a local switch, since a highway is directly connected to subscriber lines, the probability that i channels out of n channels are used when applied traffic is a Erlangs is assumed to be

$$A(i,a,n) = \frac{\dfrac{a^n}{i!}}{\displaystyle\sum_{\nu=0}^{n} \frac{a^\nu}{\nu!}} \tag{1}$$

This distribution is called Erlangian distribution. Under the assumption the blocking probability of two stage switching network is given by[1]

$$B = \frac{\dfrac{a^n}{n!}}{\left(\displaystyle\sum_{\nu=0}^{n} \frac{a^\nu}{n!}\right)^2} \sum_{\gamma=0}^{n} (n-\gamma+1)\frac{a^\gamma}{\gamma!} \tag{2}$$

II Application of PCM to TDM switches

The transmission quality problem TDM switch leads the study of PCM digital switch. Using PCM, much more complex TDM switch fabrics can be realized without difficulty in maintaining transmission quality in field.

In AT&T Bell Laboratories, the group of H.E. Vaughan started digital switching fabric studies[2] which continued nearly 20 years until the realization of the first large scale TDM switching system No.4 ESS in 1975. In Japan a study group of Professors Usatake and Akiyama of the University of Tokyo proposed a PCM switching network by parallel transmission, which was one of major outcome from their study on PAM switching network realization[3].

The experimental system of the group of H.E. Vaughan was called ESSEX. The switch fabric employed in the ESSEX was a three stage time division network as shown in Fig.1. Subscriber lines are connected to line

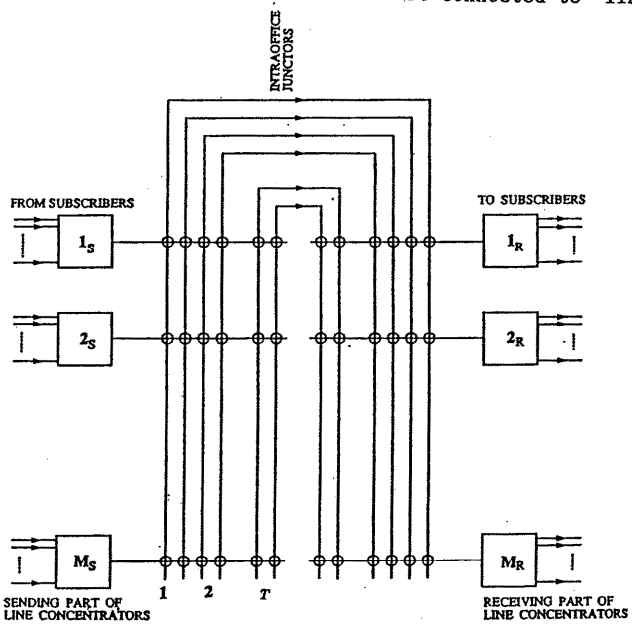


FIG.1 BASIC DIAGRAM OF THREE STAGE TIME DIVISION SWITCH FABRIC

concentrators having AD and DA converters. Send and receive buses from line concentrators are connected by the third time division buses called junctors. In setting up connection, a commonly available time slots must be found among the three time division buses.

Professor Hiroshi Inose was with the Moore School of Electrical Engineering, the University of Pennsylvania as an associate from Sept. 1956 to Feb. 1958. During the summer of 1957, Professor Inose was with Bell Laboratories as a consultant and join the PCM switching group. He worked to evaluate the traffic carrying capability of ESSEX. By his analysis for a network with n time slots, T junctor pairs with offered traffic of a_1, a_2 and a_3 erlangs to each of the send buses, the junctors and the receive buses respectively, the blocking probability is given by the following formula assuming that the distributions of calls on the buses and the junctors are independent and random[4].

$$B = \frac{1}{\left(\displaystyle\sum_{i=0}^{n}\frac{a_1{}^i}{i!}\right)\left(\displaystyle\sum_{j=0}^{n}\frac{a_2{}^j}{j!}\right)^t\left(\displaystyle\sum_{k=0}^{n}\frac{a_3{}^k}{k!}\right)} \sum_{m=0}^{n}\frac{(n-m)!}{n!}\left(\frac{m!}{n!}a_2{}^{n-m}\right)^T$$
$$\times \sum_{i=0}^{m}\sum_{k=m-i}^{m}\frac{a_1{}^i a_3{}^k}{(i+k-m)!}\binom{n-i}{m-i}\binom{n-k}{m-k}\left(\sum_{j=0}^{m}\frac{a_2{}^j}{j!}\right)^T \tag{3}$$

In those days, since very high speed transmission s difficult and number of PCM channels used in a bus only 20 or 24. The blocking caused by the difficulty find commonly available time slot between three buses lled time slot mismatch loss was serious.

If different time slots can be connected between nd and receive buses, time slot mismatch loss can be iminated. For this purpose, Prof. Inose invented a vice called time slot interchange[5]. Using time slot terchange, if time slot mismatch loss can be iminated completely, the blocking probability is to be ven by

$$B = 1 - \left(1 - \frac{\frac{a^n}{n!}}{\sum_{\mu=0}^{n} \frac{a^\mu}{\mu!}} \right)^2 \qquad (4)$$

To construct time slot interchange, a eech memory is needed which have random cess capability. In those days, magnetic re memory was still slow in speed and pensive. Shift register was expensive cause it was constructed by discrete ansistor flip-flops. The first perimental time slot interchange was nstructed in the University of Tokyo in

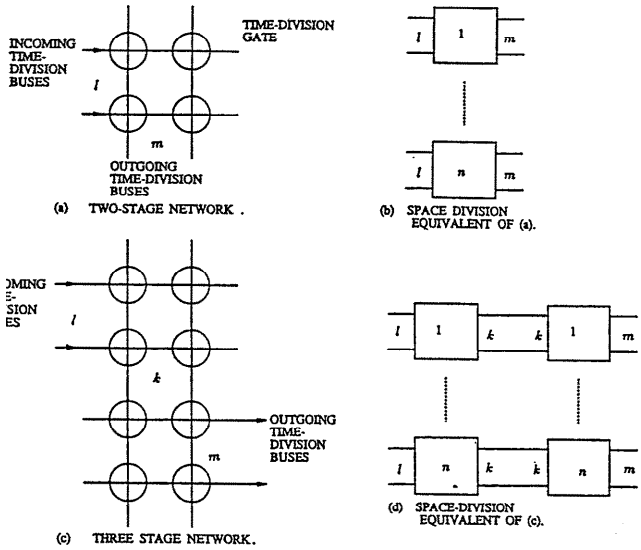


FIG.2 SPACE-DIVISION ANALOGY OF TIME-DIVISION NETWORKS OF n TIME SLOTS WITH TIME-DIVISION GATES.

61 utilizing magnotostrictive delay ne having a number of output taps to rm a long delay line[6]

Prof. Inose also constructed a full perimental digital switching system lled CAMPUS having automatic control pability by means of delta modulation ing the time slot interchange. The itching system has only a pair of send d receive buses which are connected by time slot interchange.

I Various Switch Fabric Design using Time Slot Interchange

-1 Time Division Space Division Equivallence

After confirming the realizability time slot interchange, various types time division switch fabrics were signed in the University of Tokyo. ıce space division switches were ıdied intensively before time division chnology, traffic characteristics of ne-division network was understood by ıce-division analogy[7].

Fig.2 shows space division analogy a simple rectangular array of $l \times m$ ne division gate which connects incoming and outgoing time division ses each having n time slots. this configuration a time slot in any one of l incoming buses has access to the same time slot in any one of m outgoing buses, so that the network is equivalent to n separate $l \times m$ space-division rectangular arrays as shown in Fig. 2(b). It should be noted that no connection between time slots of different phase is possible by time-division gates. This constraint imposes considerable restriction on trunking aspects of the network. For example, if the time slot is preassigned in the first time-division stage of Fig.2(a), the network simply works as a $l \times m$ rectangular array and no advantage in availability owing to the time-division multiplicity is taken. The restriction may also be observed in the space-division

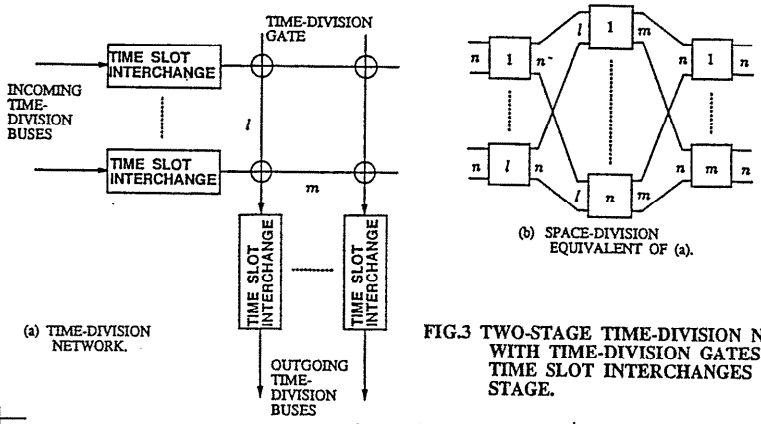


FIG.3 TWO-STAGE TIME-DIVISION NETWORK WITH TIME-DIVISION GATES AND WITH TIME SLOT INTERCHANGES IN EACH STAGE.

analogue shown in Fig.2(d) which is equivalent to the time-division switching network connecting three time-division stages shown in Fig.2(c). Here, no distribution between the equivalent space-division rectangular arrays is available.

The inflexibility is remedied by the ultilization of appropriate delay devices. A time slot interchange of a bus having n channels equivalent to a $n \times n$ space-division square array in which each inlet or outlet corresponds to the respective time slot. Employing the time slot interchange in the two time-division buses of the network of Fig.2(a), the preassigned time slot in its first time division bus may be shifted to any one of n time slots, thus the network becomes analogous to the three-stage space division network as shown in Fig.3. Similarly, the network in Fig. 2 (c) is equivalent to the three-stage space division network as shown in Fig.4 if the time slot interchanges are employed in its second time-division bus. This network may be suffice for a local switch in which a convenient time slot may be assigned to a subscriber. In a toll switch in which the incoming time slot is preassigned, the time slot interchange may be employed in the first and the third time-division stages similar to that in Fig 3.

3-2 Non-Blocking Time-Division Switch Fabrics[7]

The network in Fig.4(a) is apparently nonblocking if $k = 2l - 1$ since its space division equivalent becomes Clos type three-stage nonblocking network.

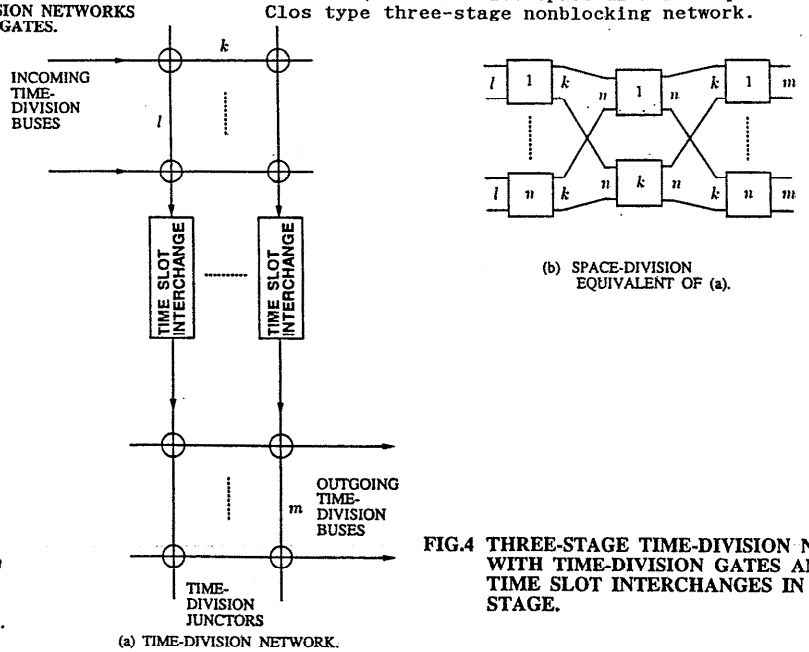


FIG.4 THREE-STAGE TIME-DIVISION NETWORK WITH TIME-DIVISION GATES AND WITH TIME SLOT INTERCHANGES IN SECOND STAGE.

The switch fabric of Fig.3 is a time-space-time (T-·T) network and the switch fabric of Fig.4 is a space-.me-space (S-T-S) network. In the network of Fig.3 if .me division multiplexity of incoming bus and outgoing ıses is n and the time division multiplexity of space /itch is 2n, the network is also equivalent to Clos 'pe non-blocking network. Inose and Saito filed the ıtent for those two types of non-blocking time division /itching networks in 1964 and granted patent in)67[8].

If cost of time switch is higher than that of space /itch, S-T-S networks are more economical than T-S-T ;tworks. In late 1960's random access memory cost is ;pensive so that S-T-S types of networks are preferred › reduce network cost.

3-3 Three Stage Network with TSI

In S-T-S types of network shown in Fig.4, if the umber of time division buses having TSI is smaller, hen the structure of switch fabrics is an improvement f ESSEX network having TSI in junctors. This type of witch fabric is shown in Fig.5[9].

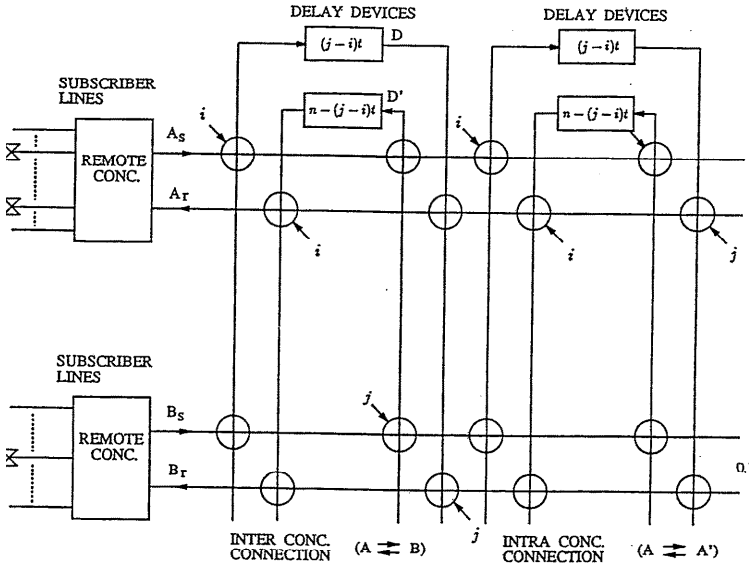


FIG.5 S-T-S TYPES OF NETWORK APPLIED TO ESSEX TYPE TDM SWITCH

To establish a connection, a commonly idle time slot must be found, in the first place, between the send ɔus of the originating line concentrator and the primary ɔart of a junctor. The procedure may be repeated on several or all of the junctors connecting the ɔriginating and the terminating line concentrators. If the number of the repetition is limited to T, this may ɔe called as taking primary junctor matching T times.)n the other hand, if a commonly idle time slot is not ɔbtained between the receive bus of the terminating remote line concentrator and the secondary part of the Junctor, the primary part of which is matched to the send bus of the originating remote line concentrator, the junctor must be abandoned and the the procedure of matching by another junctor should be performed again including primary junctor matching. The number of repetition caused by the mismatch between the secondary part of a junctor and the receive highway is called the number of secondary junctor matching, the maximum number of which may be limited to $S(S \leq T)$.

The blocking probability formula is derived as follows assuming that the distribution of calls on the highways and the each part of the junctors are independent and random and that time slot assignment is also random[9].

$$
{}_{,}B = \sum_{U=0}^{S} \binom{T}{U} \sum_{i=0}^{n} \frac{\frac{a_1{}^i}{i!}}{\sum_{\nu=0}^{n} \frac{a_1{}^\nu}{\nu!}} \left\{ 1 - \sum_{j=n-i}^{n} \frac{\binom{i}{n-j} \frac{a_2{}^j}{j!}}{\binom{n}{j} \sum_{\nu=0}^{n} \frac{a_2{}^\nu}{\nu!}} \right\}^{U}
$$

$$
\times \left\{ \sum_{j=n-j}^{n} \frac{\binom{i}{n-j} \frac{a_2{}^j}{j!}}{\binom{n}{j} \sum_{\nu=0}^{n} \frac{a_2{}^\nu}{\nu!}} \right\}^{T-U}
$$

$$
\times \sum_{k=0}^{n} \frac{\frac{a_3{}^k}{k!}}{\sum_{\nu=0}^{n} \frac{a_3{}^\nu}{\nu!}} \left\{ \sum_{j=n-k}^{n} \frac{\binom{k}{n-j} \frac{a_2{}^j}{j!}}{\binom{n}{j} \sum_{\nu=0}^{n} \frac{a_2{}^\nu}{\nu!}} \right\}^{U}
$$

$$
+ \sum_{U=S+1}^{T} \binom{T}{U} \sum_{i=0}^{n} \frac{\frac{a_1{}^i}{i!}}{\sum_{\nu=0}^{n} \frac{a_1{}^\nu}{\nu!}} \left\{ 1 - \sum_{j=n-i}^{n} \frac{\binom{i}{n-j} \frac{a_2{}^j}{j!}}{\binom{n}{j} \sum_{\nu=0}^{n} \frac{a_2{}^\nu}{\nu!}} \right\}^{U}
$$

$$
\times \left\{ \sum_{j=n-i}^{n} \frac{\binom{i}{n-j} \frac{a_2{}^j}{j!}}{\binom{n}{j} \sum_{\nu=0}^{n} \frac{a_2{}^\nu}{\nu!}} \right\}^{T-U}
$$

$$
\times \sum_{k=0}^{n} \frac{\frac{a_3{}^k}{k!}}{\sum_{\nu=0}^{n} \frac{a_3{}^\nu}{\nu!}} \left\{ \sum_{j=n-k}^{n} \frac{\binom{k}{n-j} \frac{a_2{}^j}{j!}}{\binom{n}{j} \sum_{\nu=0}^{n} \frac{a_2{}^\nu}{\nu!}} \right\}^{S} \tag{5}
$$

where n, a_1, a_2, a_3 are the same to that in Eq.(1).

Fig.6 shows the numerical values of blocking probability of networks of the types shown in Figs. 1 and 5 in the cases $a_1 = a_2 = a_3 = a$ and $n = 20$. The blocking probabilities of the conventional network, the network with the junctors with TSI and the network free from time slot mismatch are shown by the chained lines, the solid lines and the broken line respectively. It is known from the figure and other results of calculations that the internal blocking is essentially eliminated by performing the junctor matching four times.

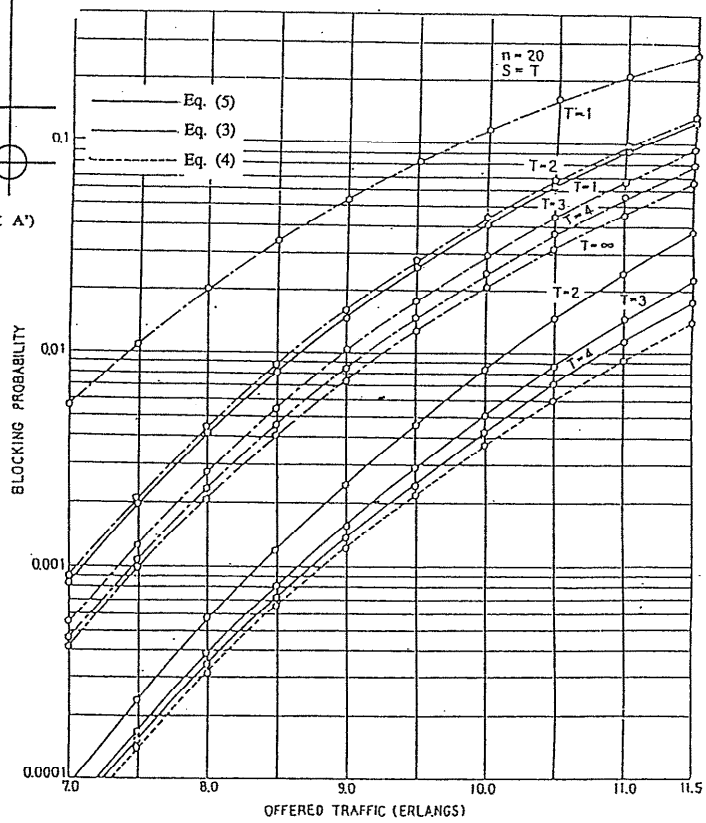


FIG.6 BLOCKING PROBABILITY OF VARIOUS SWITCH FABRICS

3-4 Three Stage Network Having TSI on Limitted Number of Junctors

Considering the fact that the majority of cases which require the time slot interchange are such that a commonly idle time slot is unavailable between both highways involved, the network of Fig.5 may be modified by providing the time slot interchanges only in a

limited number of junctors and utilize the junctors with TSI as the alternate route.

In setting up a connection, a pair of the direct junctors without the time slot interchanges are matched in the first place to the buses of the originating and the terminating line concentrators and if the first trial fails, the procedure is repeated on other direct junctor pairs. The repetition may be limited to T_1. If the direct junctor matching is unsuccessful, the matching by means of the alternate junctor with time slot interchanges is performed, the repetition of which may be limited to T_2. If a commonly idle time slot between the highways of the originating and the terminating line concentrators is unavailable, the alternate junctor matching is performed immediately without performing the direct junctor matching. The number of repeated matching may be limited to T_3. In the case of intra-concentrator connection, two time slots must be assigned and interchanged by an alternate junctor pair. The number of repeated matching may be limited to T_4. The performance of the network was evaluated by simulation. The results of simulation indicated that smaller number of pulse shifting junctors are enough to essentially eliminate time slot mismatch loss.

The simulation was performed by means of IBM 7090 computer. Computer time of needed to obtain the data of one point was approximately 3 minutes. Simulation study at those times was so expensive.

3-5 Networks for Inter-office connection

The three-stage network of Fig.5 is suffice so far as the interaoffice connection is concerned. For interoffice connection in which interoffice junctor loading as high as 70 % is generally required, two time-division stages which work as a concentration and an expansion stages respectively may be added before and after the interoffice junctor stage as shows in Fig.7.

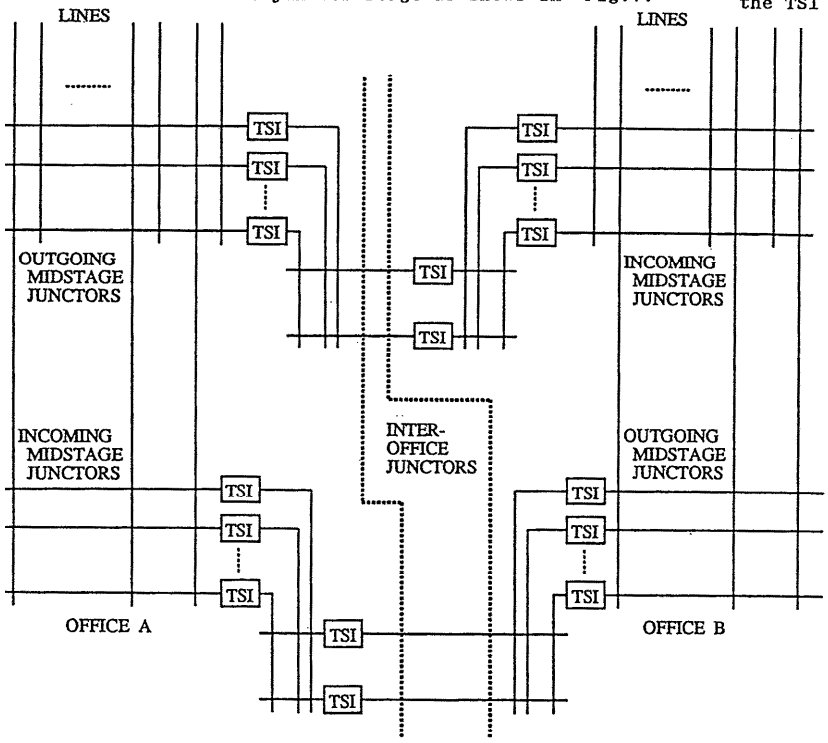


FIG.7 FIVE STAGE INTEROFFICE SWITCHING NETWORK.

These additional stages which are provided with the pulse shifters, concentrate the traffic on the interoffice junctors and as the result the permissible interoffice trunk loading reaches 85%[10].

Combinatorial calculations similar to that on the preceeding cases give the blocking probabilities of the five-stage network shown in Fig.7 in which two additional stages that are called the outgoing midstage junctors and the incoming midstage junctors are provided before and after the interoffice junctor stage respectively. The blocking probability for the interoffice incoming connection which involves an interoffice junctor, one of the incoming midstage junctors and a link, and the blocking probability for the interoffice outgoing connection which involves a bus, one of the outgoing midstage junctors and one of the interoffice junctors, are given respectively as follows.

$$B_{in} = \sum_{\gamma=0}^{n} \frac{\gamma/n}{1-\frac{a_3}{n}} A(n-\gamma, a_3, n) \sum_{\tau=0}^{T} \binom{T}{\tau} \{K(\gamma, a_2, n)\}^{T-\tau} \{1-K(\gamma, a_2, n)\}^{\tau} \times \sum_{\lambda=0}^{n} \binom{n}{\lambda} \{K(\tau, a_2, T)\}^{n-\lambda} \{1-K(\tau, a_2, T)\}^{\lambda} K(\lambda, a_1, n) \quad (6)$$

$$B_{out} = \sum_{\gamma=0}^{n} A(n-\gamma, a_1, n) \sum_{\tau=0}^{T} \binom{T}{\tau} \{K(\gamma, a_2, n)\}^{T-\tau} \{1-K(\gamma, a_2, n)\}^{\tau} \times \sum_{\lambda=0}^{n} \binom{n}{\lambda} \{K(\tau, a_2, T)\}^{n-\lambda} \{1-K(\tau, a_2, T)\}^{\lambda} \times \{K(\lambda, a_3, n)\}^{R} \quad (7)$$

where

$$K(i, a, n) = \sum_{\gamma=i}^{n} \frac{\binom{n-i}{n-\gamma}}{\binom{n}{\gamma}} A(\gamma, a, n) \quad (8)$$

In the above equations, T and R are the numbers of incoming (or outgoing) midstage junctors and the interoffice junctors available for connection, a_1, a_2 and a_3 are the offered traffic to a bus, a midstage junctor and an interoffice junctor respectively[10].

3-6 Networks assuming Rearrangement

Time slot mismatch loss can also be avoided by the rearrangement. The possibility may readily be understood in the two-stage timing division network with the TSI shown in Fig.3 the space division equivalent of which is apparently a rearrangeable space-division network[11]. The rearrangement is performed in such a way that, if a commonly idle time slot is unavailable for a new call, one of the calls in connection is shifted to another time slot so as to accomodate the new call. If this rearrangement is repeated unlimittedly, the network becomes completely non-blocking. Rearrangement of only once or at most twice however, is generally suffice to reduce the time slot mismatch loss essentially negligible. The blocking probability when the rearrangement is performed once is given as

$$B_{rl} = \sum_{i=0}^{n} \sum_{j=n-i}^{n} \frac{A(i,a,n)A(j,a,n)\binom{i}{n-j}}{\binom{n}{j}} \times \left\{\frac{K(n-i+1,a,n)}{K(1,a,n)}\right\}^{n-j} \left\{\frac{K(n-j+1,a,n)}{K(1,a,n)}\right\}^{n-i} \quad (9)$$

The blocking probability when the rearrangement is performed twice is similarly obtained. The results of numerical calculation indicates that the time slot mismatch loss may be essentially eliminated by means of the rearrangement of only once. Similar conclusion has been reached as the result of analysis on the three-stage time division switching networks[12][13].

3-7 Early Implementations

In late 1960's, after commercialization of the first electronic switching system by AT&T Bell Laboratories, time division switching systems are also considered as candidates for commercial electronic switching system. In Japan DEX-T1 experimental digital switching system was constructed by Electrical Communication Laboratories using a switch fabric of the type shown in Fig.5[14].

However, it is only after 1975 that cost of digital system is reduced enough to justify digital switching system.

IV Networks having higher multiplexity

One of the reason why digital system gains economical justification is the improvement of digital technology to implement higher speed circuit. In the time division switch fabrics of after 1970's, number of time division channels increased. Using transmission speed of 8 Mb/s, number of channels can be 128. If 8 bit parallel transmission is assumed 1024 channel system can be easily realized. When digital switching

technology was started to be widely used, in mid 1970's, switching speed higher than 8 Mb/s was easily realizable. In addition, memory cost to realize time slot interchange was also reduced drastically. Therefore TSI having 128 channels or more became economical enough.

When TSI was expensive due to high memory cost, S-T-S type of fabrics was more economical. However, if TSI is economical enough, T-S-T type of switch fabrics becomes more economical. In addition, as is apparent from the space division equivalent circuit shown in Fig. 3(b), channel selectivity at the first stage switch is equal to time division multiplexity in T-S-T types of switch fabrics. This means that channel selectivity of 128 or more can be easily attained in T-S-T types of switch fabrics. In this types of networks, very low blocking probability is achieved at high traffic. As for the space stage in T-S-T types of switch fabrics, one through 4 stage space switch can be used depending on the number of time division buses handled by the switch fabric.

For example, in No.4 ESS which was first implemented in 1975, a 4 stage space switch composed of 8x8, 16x16, 16x16 and 8x8 matrices is used as shown in Fig.8[15]. In the switch fabric of No.4 ESS, the TSI

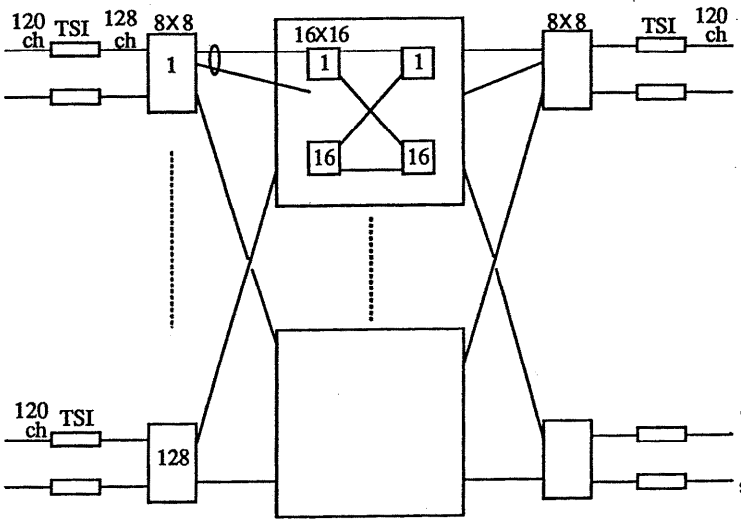


FIG.8 T-S-T TYPE SWITCH FABRIC FOR NO.4 ESS

have 120 channels for external transmission and 128 channels for internal transmission. If the number of channels for internal transmission is twice that of the external transmission and if the S stage is single stage switch, the fabric could be complete non-blocking switch. However since the number of time division channels is so large that the network have very low blocking probability even if the number of channels for internal transmission is less than the value to give non-blocking condition. Because of this reason, the switch fabrics for No.4 ESS is said to be essentially non-blocking[15][16].

In No.4 ESS switch, the blocking probability is very low. If the usage of transmission line which gives the blocking probability of 1% is about 84%. In No.5 ESS, one stage S switch having time division multiplexity of 256 is used[17]. This switch gives further lower blocking probability. In D-70 switching system developed by NTT, Japan one stage S switch having multiplexity of 1024 channels is used[18].

In these switch fabrics, blocking probability of switches is low enough even if the network is used by the efficiency of 90% or more. In such an environment, traffic estimation error of 10% is enough to bring the network into catastrophe. This means that subject of traffic study related with switch design has shifted from switch fabric design to traffic load estimation.

V Concluding remarks

This paper described a short history of traffic engineering related with time-slot interchange (TSI) invented by Prof. Hiroshi Inose. Variety of ingenuities were important to make the TSI based switch fabric practical. All the switch fabrics utilized in present day digital switching system are based on TSI principle.

Traffic engineering aspect to be applied to TSI based switch fabric changed drastically in this history. Initially switch fabric structure for economical realization of switching function is important. However, because of advancement of electronics, design of higher speed switch fabrics became easily enough. The importance of traffic engineering has shifted switch design to traffic load estimation and overall network management independent from specific design of switch fabrics.

References

1) L.R.F.Harris : "Time sharing as a Basis for Electronic Telephone Switching", Proc. IEE, Vol. 103B, pp.722-742 (Nov., 1956)

2) H.E.Vaughan : "Research Model for Time Separation Integrated Communication", Bell Syst. Tech. Jour., No.38, No.4, pp.909-932 (July, 1959)

3) T.Osataka and M.Akiyama et al: "An Experimental Electronic Switching System on a Basis of Coded Pulse Parallel Transmission", Journal Inst. of Elec. Commun. Engs. of Japan, Vol.46, No.10, pp.1402-1410 (Oct., 1963)

4) H.Inose: "Blocking Probability and Junctor Efficiency of a Three Stage Time-Division Switching Network" Jour of Inst. of Elec. Commun. Engs of Japan, Vol.44, No.6, pp.935-941 (June, 1961)

5) H.Inose and J.P.Runyon : "Time Division Switching System for Telephone System Utilizing Time Slot Interchange" (filed on April 27, 1960) U.S. Patent 3,172,956 (March 9, 1965)

6) H.Inose, Y. Yoshida, Y.Yasuda and Z.Koono : "A Time Slot Interchange System in Time-Division Electronic Exchanges" Trans. IRE., Vol. CS-10, pp.397-467 (Dec., 1967)

7) H.Inose, T.Saito and M.Kato: "Time Slot Interchange in Three-Stage Time-Division Switching Networks", Coloque International de Commutation Electronique, Paris, pp.409-418,(1966)

8) H.Inose and T.Saito : "Time Division Switching System" (filed on Dec. 29, 1964) U.S. Patent 3,446,917 (May 27, 1969)

9) H.Inose and T.Saito : "A Three-Stage Time-Division Switching Network by Time Slot Interchange Principle" Jour. of Inst. of Elec. Commun. Engs. of Japan, Vol.49, No.3, pp.409-417 (March, 1966)

10) H.Inose and T.Saito : "Four-Stage and Five-Stage Switching in Integrated Communication Systems" Conference on Switching Technique for Telecommunication Networks, London, pp.165-168 (April, 1969)

11) V.E.Benes : "On Rearrangeable Three-Stage Connecting Networks", Bell. Syst. Tech. Jour., Vol.41, No.5, pp.1481-1492 (Sep., 1962)

12) M.Akiyama and H.Okada : "Rearrangements in Time Division Switching System", Trans. Inst. of Elec. Commun. Engs. of Japan, Vol.51-A, No.4, pp.145-152 (April, 1968)

13) M.Akiyama and S.Suzuki : "Rearrangements in Two Stage Time Division Switching Network", Trans. Inst. of Elec. Commun. Engs. of Japan, Vol.52-A, No.3, pp.137-42 (March, 1969)

14) J.Yamato and K.Habara : "Speech Path Network of PCM Switching System DEX-T1", Elec. Commun. Lab. Tech. Jour., Vol.16, No.7, pp.2551-2580 (Nov., 1967)

15) H.E. Vaughan: "An Introduction to No.4 ESS", ISS 1972, p.19 (June, 1977)

16) T.Saito: "An Optimum Design Theory for PCM Toll Switching Networks", National Telecommunication Conference, pp.46.6-1 - 46.6-6 (Nov. 1977)

17) D.L.Carney et al: "The 5 ESS Switching System : Architectural Overview", ATT Tech. Jour., Vol.64, No.6, p.1339 (July-Aug., 1985)

18) T. Egawa, T.Ueno et al. : "Speech Path Equipment for D60 and D70 Digital Switching System", Elec. Commun. Lab. Tech. Jour., Vol.33, No.7, pp.1737-1752 (July, 1984)

VTDM: A VARIABLE BIT RATE TDM SWITCH ARCHITECTURE FOR VIDEO STREAM

Udomkiat Bunworasate, Takayuki Muranaka, Terumasa Aoki, Hitoshi Aida and Tadao Saito
Department of Information and Communication Engineering, University of Tokyo
udomkiat,taka,aoki,aida,saito@sail.t.u-tokyo.ac.jp

Abstract: This paper proposes the Variable-bit-rate TDM (VTDM) switch architecture, which not only has the advantage of TDM but also can support VBR traffic. The VTDM frame is designed for accommodating video data by enlarging the size of time slot and synchronizing the VTDM frame length with the video frame interval. The double-buffer scheme allows a guarantee of maximum switching delay and a slight delay jitter. It will be shown that adequate performance can be achieved by simply applying traffic shaping. By utilizing the header information of VTDM frame and the scheduler operation, an intelligent priority control, leading to picture quality attainment, can be implemented.

1 Introduction

Digital motion video is an increasingly substantial type of network traffic and most of this traffic is expected to comply with the MPEG video compression standard. MPEG stream, however, may require an output bit rate to vary from frame to frame which increases network resource allocation complexity. Frame-to-frame rate variation with variable bit rate (VBR) MPEG video may be severe, up to an order of magnitude, due to video content and frame type in a MPEG sequence.

It is broadly acknowledged that video traffic is generated at synchronous timing and requests a guarantee of maximum transmission delay together with a slight delay jitter. These requirements can be easily fulfilled by TDM (Time Division Multiplexing) technology. A shortcoming of conventional TDM, however, due to its static nature makes it difficult to efficiently support continuous and bursty video traffic. Although TDM-based circuit switching can be modified to support multirate services by allocating multiple slots for high-bandwidth applications [1], the approach still does not eliminate the problem of handling VBR stream.

This has motivated a search for alternative solutions. Asynchronous Transfer Mode (ATM), a predominant technology, was principally designed to accommodate various types of network traffic. For VBR video over ATM networks, cell losses may be incurred due to limited buffer capacity at the switches and result in severe picture quality degradation. Difficulty in satisfying end-to-end delay jitter and high transmission efficiency simultaneously has been reported [2].

Fast packet switching [3], burst switching [4], and DTDM [5] were proposed with the goals of providing bit-rate flexibility. Their approaches, however, appeared to be based on the original concept of ATM. DTM [6] was also developed to support bursty traffic. Unfortunately, its characteristic of establishing a circuit whenever a burst of data occurs, results in a complicated connection setup problem. Moreover, all of the above mentioned methods were not particularly set for handling VBR video traffic.

In this paper, we propose a new switching architecture, referred to here as Variable bit rate TDM (VTDM), which is capable of handling video stream efficiently. The VTDM not only has the advantage of TDM but also can support VBR traffic. One distinctive feature is that intelligent video stream controls can be performed at the switch level by utilizing time slot content information in VTDM frame header.

2 VTDM- Variable bit rate TDM

VTDM allows flexible transport for bursty traffic. Note that a time slot in TDM or multirate TDM is allocated to an input channel even during the absence of data on that channel. VTDM removes this inefficiency by allocating time slots in VTDM frame only for the active channels and is, therefore, more efficient in resource allocation. In addition, the time slot allocation, which is varied in every VTDM frame, also eliminates the fixed channelization constraints of the TDM approach. A subscriber uses sufficient slots to match the bit amount of each generated video frame instead of dealing with integer multiples of the basic channel rate. This makes the support of VBR traffic possible.

0-7803-5796-5/99/$10.00 © 1999 IEEE

Fig. 1 shows VTDM frame configuration. The size of each time slot in VTDM frame is preliminarily set to 16 kbit. Note that the selection of an optimal slot size is usually a difficult engineering problem involving tradeoffs between service quality, transmission efficiency, and equipment simplicity. The larger slot size, compared with 1 octet in TDM, 390 octets in DTDM and 48 octets in ATM payload, is chosen to simplify the processing of video frames and to reduce the overhead of header. Furthermore, this would result in simple time slot synchronization and the possibility of high capacity switch. The fixed length time slot also permits us to simplify buffering in multiplexer and to ease the bandwidth allocation among multiple users.

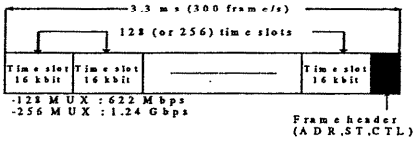


Fig. 1 VTDM frame configuration

The frame header carrying all time-slot contents is attached to VTDM frame to cope with VBR stream. This header consists of address (ADR) field, slot type (ST) field and other control information (CTL). The address field can be a connection identification number, routing information or physical address. The ST field identifies which priority class each slot belongs to. During network congestion, VTDM switch carries out priority control in slot level by utilizing slot content information through the ST field. The CT field is reserved for future use. (e.g. to carry time slot sequence numbers or time stamps)

The duration of VTDM frame is determined as 3.3 ms. (1/300 sec), which is a common divisor of two frame time intervals of NTSC (1/30 sec) and PAL (1/25 sec) video signals. One reason to choose VTDM frame duration being a divisor of picture frame time interval is to simplify time slot allocation procedure. For a transmission rate of 622 Mbps (1.24Gbps), VTDM frame will contain 128(256) multiplexed MPEG2 video channels. By assigning one slot in VTDM frame to each connection, the average bit rate per connection is 16 (Kbit) x 300 (frame/s) = 4.8 Mbps.

3 Switch Architecture and Operation

The VTDM switch consists of four main parts: a variable-bit-rate time (VT) switch per input port, a variable-bit-rate space (VS) switch, a centralized scheduler, and frame generators (FG) as shown in fig. 2. The VT switch adopts the structure of double sequential-write / random-read buffer. One buffer is used for accumulating time slot arriving at the input port of VT switch, while the other buffer is used for storing time slot to be transmitted to VS switch. The size of the both buffers is exactly VTDM frame length. The function of each buffer interchanges on frame-to-frame basis.

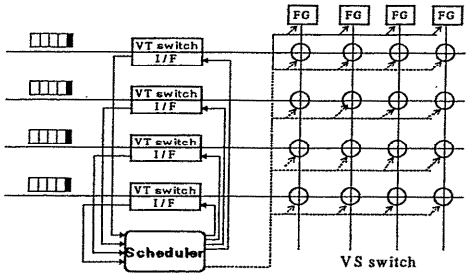


Fig. 2 Architecture of VTDM switch

The main difference between conventional time switch and the VT switch is that a scheduler interface module is attached to each port. According to scheduling result from the scheduler, if a time slot has not been transmitted within a VTDM frame period, it will be compulsively discarded. This policy has two advantages of being practical to implement and making the guarantee of maximum slot delay possible (fig. 3).

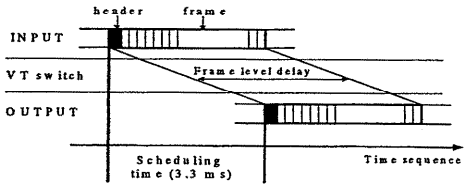


Fig. 3 Movement of VTDM frame in time sequence

VS switch fabric is fundamentally identical to crossbar space switch in TDM. As there is no buffer at every VS switch crosspoint, the centralized scheduler must conduct an arbitration to avoid collision at the VS switch plane. Based on the scheduling result, the time-slot reallocation operation is performed at the random-write buffer in VT switch. The scheduler also has to control the on-off operation of all crosspoint gates in VS switch for transferring each time slot correctly to its destination. The scheduler performs time slot assignment by referring to the information in VTDM frame header at every cycle of frame time. This is apparently different from the time slot assignment of TDM and multirate TDM, which is

conducted only when a connection is first established. Moreover, the scheduler implements priority control by exploiting VTDM header information. However, as shown in fig. 3, the scheduling process must be completed within one VTDM frame period (3.3 ms).

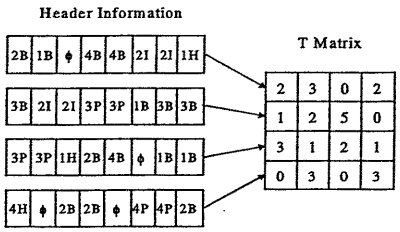


Fig. 4 Header information and T matrix

The scheduling policy in VTDM switch exploits a concept of group traffic matrix (T matrix). Fig. 4 illustrates an example of T matrix formation from header information (switch size 4x4, the amount of multiplexed channels 8). I, P, B symbol represent the picture type of MPEG2 stream where H represents the time slot that accommodates the header of each picture. (Data amount of a picture frame is generally bigger than a time slot size.) Each figure in the header information depicts the output port, to which each time slot will be destined. T can be represented by $n \times n$ matrix, where n is the number of input-output port of VTDM switch. The element t_{ij} represents the amount of time slots destined to output port j from input port i (or time slots that desire to pass through any crosspoint (i,j) of VS switch). By generating T matrix for every VTDM frame, the macro information of time slots is visualized. For instance, the sum of elements in a column indicates the number of time slots vying for the same output port. If this amount exceeds the number of multiplexed channels, some time slots will be inevitably discarded.

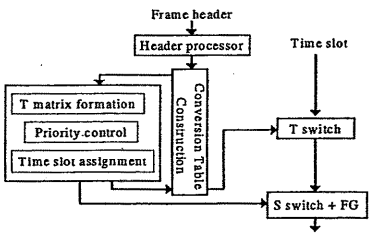


Fig. 5 Functional module of VTDM switch

Fig. 5 demonstrates functional module of VTDM switch. All functions performed by the scheduler are indicated in the left region. As every T matrix component matches every crosspoint gate of VS switch, the scheduler is able to configure these gates directly. However, the matrix components do not correspond to memory addresses of time slots in VT switch buffer. Hence, the modules in the middle part of fig. 5, implemented in scheduler interface module at each VT switch, are required to support time-slot reallocation operation. The header processor unit acquires header information and then conveys it to the conversion table construction unit. Table 1 illustrates an example of the table, built from the 1st VTDM header shown in fig. 4. The table groups the memory addresses of time slots by output number. By searching time slot address from this table on first-in-first-out basis, the proper address that corresponds to the selected component of T can be retrieved and time slot sequence integrity can be achieved. Furthermore, constructing the table with separated priority level categories would enable the implementation of priority control.

Input port no	Destined output	Pic. header	I pic.	P pic.	B pic.
1	1	1			7
	2		2,3		8
	3				
	4				4,5

Table 1 Example of priority-based conversion table

Time slot assignment problem in VTDM can be viewed as the decomposition of T into nonconflicting matrices: $T = U_1 + U_2 + \ldots + U_h$. A nonconflicting matrix, U_i, is a unit matrix with at most one nonzero entry in any row or any column and h is the number of multiplexed channel in a VTDM frame. A feasible time slot assignment algorithm in VTDM switch must have a execution time of less than 3.3 ms. We have proposed VCA+ (Vector Comparison Algorithm +) [7], a systematic iterative algorithm that provides excellent efficiency and is also practical for hardware implementation. With a combination of hardware (FPGA) and software (C language) implementation, we have designed a scheduler and confirmed that the scheduler can make its decision within 0.64 ms. The total number of gates are calculated to be 260,924, feasible for implementation on a single LSI chip at current technology. (16x16 switch, 128 channel, 100 MHz sampling clock) [8]

The scheduler forwards the renewed header information to all frame generators. Then, each frame generator creates the output VTDM frame that will contain the time slots, consecutively sent from VT switch.

4 Evaluation of VTDM switch

VTDM network allocates transmission capacity for each connection depending on its average bandwidth. Since MPEG2 stream has a bit rate variation, there is a definite possibility that a large number of time slots hold the same destination address in the same VTDM frame time. It is highly likely that some time slots may be discarded due to a lack of output channel despite of exploiting the ideal time slot assignment algorithm. In this section, we will investigate the performance of VTDM switch in terms of time slot loss ratio and delay characteristic by means of extensive simulations. A multiplexer is attached to each input port of the switch model illustrated in fig. 2. MPEG2 stream with average bit rate of 4.8 Mbps is generated at each source. The number of multiplexed sources per one port (128 and 256) as well as per port capacity (622 Mbps and 1.24 Gbps) are set as parameters.

4.1 Traffic shaping

It is widely recognized that networks generally tend to behave inefficiently when a traffic concentration has occurred at some nodes. As VTDM network is aimed to accommodate bursty video stream, we expect a worthy effect of traffic shaping on the switch performance.

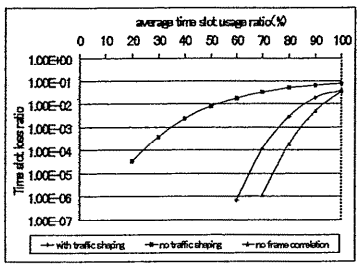


Fig. 6 Effect of traffic shaping

The VTDM frame duration (3.3 ms.) is equal to one-tenth of the time interval of ordinary video frame. (33 ms.) We adopt a straightforward traffic shaping method, which divides MPEG2 frame into time slots and uniformly accommodates them among 10 VTDM frames. To investigate the effect of this technique, another 2 cases were also studied. The former case is without traffic shaping, i.e., all time slots, segmented from the same picture frame, are successively accommodated into a VTDM frame. In the latter case, each traffic source generates time slots with random distribution (no frame correlation). The simulation results are shown in fig. 6 (16 ports, 128 channels). It can be inferred that the effect of the traffic shaping is non-trivial in VTDM switch even though time slot loss ratio of MPEG2 stream with traffic shaping was slightly above the one with no frame correlation. This finding lends support to the conclusion that traffic shaping is ideal for suppressing the bursty feature of MPEG2 frame and deconcentrating load distribution.

4.2 Priority Control

Our research work on subjective evaluation of MPEG2 stream in VTDM [9] has indicated that, the slot that accommodates picture frame header has the highest effect on picture quality degradation when it is discarded, followed by I, P, and B slot. Two-level priority control is established inside the VTDM switch. At the first level performed by priority control unit in the centralized scheduler, slots with low priority are selectively discarded whenever a lack of output channel occurs. At the second level, slots with high priority are selected from the priority-based conversion table. Fig. 7 shows the time slot loss ratios for all slot types when VCA+ assignment algorithm and 2-level priority control were applied. It is obvious that 2-level priority control shows satisfying result, as the majority of lost time slots was B slots. Our subjective evaluation has also shown that, by employing slice optimization together with selective B-slot discard policy, the quality of motion picture decoded by a standard MPEG2 decoder will be compatible to that of TV if time slot loss ratio is below 10^{-3}. Consequently, from fig. 7, we can conclude that VTDM switch is able to maintain, at least TV quality if the average time slot usage ratio is below 75 percent. Note that the P slot loss is negligible.

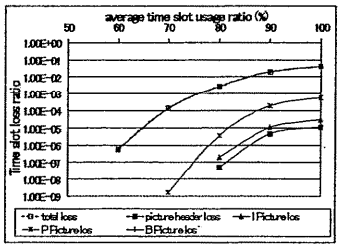


Fig.7 time slot loss ratio (applying priority control)

4.3 Switch capacity

VTDM switch capacity is given by nxh, where n is the total input-output ports, and h denotes the number of multiplexed channels in VTDM frame. The effect of both parameters was studied. Fig. 8 shows time slot loss ratio characteristics where n varied from 16 to 128 while h was set to either 128 or 256. One can say that the parameter h had an essential dependence on time slot loss ratio in

contrast to n, which did not show any significant effect. This study would lead us to an important finding that the approach of increasing the number of multiplexed channels in VTDM frame and decreasing the number of ports is the more feasible in realizing a specified switch capacity.

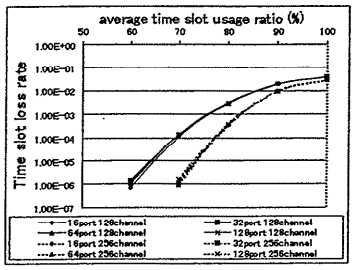


Fig. 8 Effect of switch capacity

In spite of this, it is impractical to increase the number of multiplexed channel endlessly. By maintaining VTDM frame duration at 3.3 ms and slot size at 16 kbit, the only way to increase the number of multiplexed channels is to employ a higher link capacity. Nevertheless, there is a limitation on the link capacity. This is caused by the effect of signal reflection while augmenting the capacity of TDM bus (several tens of Gbps, in general) and the restriction on RAM access cycle of time switch (3.2Gbps for SRAM, hundreds of Mbps to several Gbps for Direct Rambus DRAM). One can conclude that it is essential to adequately consider switch performance together with these bottlenecks when designing a switch capacity.

4.4 Delay characteristic

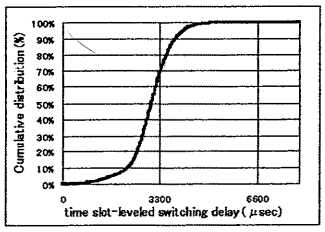


Fig. 9 Slot-level delay characteristic

By forbidding the transfer of unscheduled time slots to the subsequent VTDM frame and the employment of double buffer architecture, the ability to guarantee maximum switching delay is attained. Fig. 9 represents slot-level switching delay distribution obtained from a simulation (n = 16, h = 128,average time slot usage ratio = 60%). From the result, the maximum switching delay was bounded at 6.6 ms (two times as long as VTDM frame duration) and 99% of time slots encountered the switching delay lower than 4.63 ms. It should be noted that the picture frame-level delay equals picture frame interval due to traffic shaping. This delay characteristic is indeed suited for real-time motion picture transmission.

5. Summary

VTDM, a simple switching architecture was proposed to handle video stream effectively. Intelligent stream controls, such as priority control, can be implemented at the switch level by utilizing time slot content information in VTDM frame header. By enlarging the slot size, overheads and slot synchronization problem can be reduced. It was shown that adequate performance could be achieved by applying only traffic shaping without flow, congestion or connection admission controls. Moreover, double buffer architecture resulted in superior delay characteristic, a crucial QoS parameter for real-time video transmission. Extensions of the present work would include designs of large-scale video network and support for multicast connection.

References

[1] Yiu-Wing Leung, et al., "A Modular Multirate Video Switch – Design and Dimensioning", ICC'93

[2] Wenjun Luo and Magda El Zarki, "Quality Control for VBR Video over ATM Networks", IEEE JSAC, VOL. 15, No. 6, Aug. 1997.

[3] R.W. Muise, et al., "Experiments in wideband packet technology", Proc.1986 Int. Zurich Seminar on Digital Communications, pp.135-138

[4] S.R. Amstutz, "Burst Switching-An Introduction", IEEE Commun. Mag., Nov. 1983.

[5] L.T. Wu, et al., "Dynamic TDM – A Packet Approach to Broadband Networking", ICC'87

[6] C. Bohm, et al., "Fast Circuit Switching for the Next Generation of High Performance Networks", IEEE JSAC, VOL.14, NO.2, Feb.1996

[7] T. Muranaka, et al., "Scheduling Algorithms in TS typed Variable Bit Rate TDM Switch", Proc. of the 1999 Communications Society Conference of IEICE

[8] T. Muranaka, "A study on scheduling algorithms in TS typed VTDM video switch", Master thesis (Japanese), 1999, University of Tokyo

[9] T. Onishi, "Video compression method for VTDM switching network", Master thesis (Japanese), 1999, University of Tokyo

論　文

可変速度TDM交換方式を用いた最低帯域保証機能をもつ動画像トラヒック制御方式

森野　博章[†]　ウドムキャット ブンウォラセト[††]　相田　仁[†††]
齊藤　忠夫[†]

A Control for Video Traffic with Minimum Bandwidth Guarantee Function Using Variable Bit Rate TDM Switch Architecture

Hiroaki MORINO[†], Udomkiat BUNWORASATE[††], Hitoshi AIDA[†††], and Tadao SAITO[†]

あらまし　統計多重化効果を利用する可変速度動画像トラヒックの交換の制御は，高い伝送効率を得られるが，大容量かつ大規模な数のコネクションを収容するネットワークへ適用する場合には，スイッチにおけるトラヒックの監視等の品質保証制御をなるべく簡素化することが重要である．本論文では，既に提案されている可変速度TDM交換方式を用い，固定時間長のTDMフレームの構造を利用したトラヒックシェーピングの方式と，それに適した動画像情報への優先度付与の方法を提案している．提案するシェーピング方式では，1枚の画像フレームの情報を含む複数のタイムスロットは，画像フレーム時間間隔内で送られる各TDMフレームに均等な数だけ存在するように伝送される．TDMフレームの時間長は画像フレーム時間間隔の約数に設定されていることから，これによりトラヒックパラメータの監視制御はタイムスロット単位でなくTDMフレームの単位で行われ，処理量が削減される．本論文では，提案方式により標準的な可変速度動画像トラヒックを交換する場合の伝送効率をシミュレーションによって評価し，実用上十分な性能が得られることを明らかにしている．

キーワード　可変速度動画像交換，TDM，トラヒックシェーピング，監視制御，無損失伝送

1. まえがき

通信ネットワークは，長距離伝送及び加入者線伝送技術の進展によって急速に大容量化している．一方，トラヒック需要としては，動画像伝送を代表とする広帯域の需要が増加しつつある．こうした状況のもとで，スループット，転送遅延時間，損失率等の品質保証を目的とするトラヒック制御技術は重要な役割を果たすと考えられるが，今後の大容量化と大規模化に対応していくためには，回線利用の高能率化という目的を追求する一方，制御機構をなるべく簡素化する方向で検討を行うべきである．

† 東京大学大学院工学系研究科，東京都
School of Engineering, University of Tokyo, Tokyo, 113–8656 Japan
†† プライスウォーターハウスクーパースコンサルタント株式会社
PricewaterhouseCoopers, Tokyo, 150–6014 Japan
††† 東京大学大学院新領域創成科学研究科，東京都
Graduate School of Frontier Sciences, University of Tokyo, Tokyo, 113–8656 Japan

マルチメディアトラヒックを扱うネットワークのトラヒック制御方式の従来の研究は，広帯域ISDNを実現するという目標のもとにATM方式を対象として集中的に行われ，音声，固定速度・可変速度動画像，データ通信の個々のトラヒックの特性と要求品質に応じた最適な制御方式の検討が行われてきた[1]．なかでも，今後の需要の増加が期待される可変速度動画像伝送に対して規定されるSBR（Statistical Bit Rate）サービスクラスについては，実現機構の各要素として，コネクション受付制御，トラヒックパラメータの監視制御，そしてセルスケジューリングのそれぞれの技術について研究が行われ，制御パラメータと得られる統計多重化効果について多くの知見が得られている．

SBRサービスクラスの実現のかぎとなる技術は，トラヒックパラメータの監視制御であり，トラヒックシェーピング機能や違反トラヒックへのタグ付与の機能など様々な処理が検討されている．しかし，これらをすべて正確に行うためには，制御機能にコネクショ

ン単位のバッファ管理が必要であり，大規模数のコネクションを扱う将来の通信ネットワークで実現するには困難が多い．また，品質保証においてはスイッチにおけるセル損失の優先制御も重要であり，出力バッファ量があるしきい値を超えると低優先度セルの入力を受け付けない選択的セル廃棄により高優先度セルを保護する手法が一般的に知られているが，端末でのシェーピング及び UPC（Usage Parameter Control）の状況により，一時的に過負荷状態が生じる場合に，本方式で保証できる損失率については予測しにくい．

これらの問題点を解決するため，本論文では，基幹網において筆者らが新たに提案する可変速度 TDM 交換方式 [2] により，動画像伝送網を構築することを前提とし，可変速度 TDM 交換方式において各トラヒックソースに対する最低帯域の保証が可能なトラヒック制御方式を提案する．可変速度 TDM 交換方式は，TDM 方式を動画像伝送用に最適化した方式であり，各 TDM フレームが先頭にヘッダを有し，ここに各タイムスロットのあて先・優先度等の情報をもつ．TDM フレームの時間長は，標準的な画像伝送方式の画像フレーム周期の約数に設定されている．このため，送信端末がコネクション受付制御で決定するピークレート・平均レートと，端末が各 TDM フレーム周期で送出するタイムスロットの数とは対応づけすることができる．

提案方式においては，各送信端末は，画像フレームの情報の送出にあたって，各 TDM フレームに割り当てるタイムスロットに優先度の付与を行う．この優先度の設定は，決定したピークレートと平均レートに基づいて行われる．各端末が，各 TDM フレームごとに最も高い優先度を与えるタイムスロットの数を平均レートに相当する数以下とすれば，これらのタイムスロットについてエンド・ツー・エンドで無損失になることが保証できる．送信端末の側では，符号情報のなかで復号に必要な制御情報あるいは知覚品質にとって重要な画像情報を含むタイムスロットに，あらかじめ無損失を保証する最高優先度を与えることで，受信側での画像品質を保証できる．これは ATM 方式では実現できない優先制御である．

提案方式について，種々の実画像を MPEG2 方式 [3] を利用し，I，P，B の三つのピクチャにより符号化した画像情報を入力トラヒックとして評価を行い，その効果を示す．

2. 可変速度 TDM 交換方式

従来の通信ネットワークの研究はネットワークの提供者の立場からすべてのサービスを統合的に提供するという考え方で行われてきたが，実際にはサービスごとにネットワークが形成される場合が多い．特に交換技術においては比較的短い期間に多様なサービスごとの技術が現れて競争的に導入されてサービスが進展する．可変速度 TDM 交換方式は，動画像伝送がサービスの主要な部分を占めると予想される将来の通信ネットワークにおいて有効となる交換技術である．

2.1 スイッチの動作原理

図 1 に本方式での TDM フレームフォーマットを示す．TDM フレーム長の 3.3 ms は，画像伝送フォーマットの 2 大標準である NTSC ビデオのフレーム間隔の 33 ms と，PAL のビデオのフレーム間隔の 40 ms の約数となっている．またフレームの先頭のヘッダには，各タイムスロットのあて先，優先度等が記載される．この交換方式では，スイッチの各部の名称に TDM 交換方式に対応した用語を用い，1 TDM フレームに収容されるタイムスロット数の総和を多重度と呼ぶ．伝送速度が 622 Mbps の場合，このフレームの多重度は 128 であり，フレームの一つのタイムスロットの割当てが，平均 4.8 Mbps の帯域割当てに相当する．次に，図 2 には，可変速度 TDM 交換方式のスイッチの一つであるクロスポイントバッファ型スイッチを示す．スイッチの各クロスポイントバッファは，図 3 のように二つの FIFO バッファからなっている．入力 TDM フレームの各タイムスロットは，Traffic Monitor で監視制御を受けた後，アドレスフィルタを通り，あて先の出力ポートに対応するバッファにいったんすべて蓄積された後，優先制御を経て出力ポートへと転送される．図 2〜図 4 にかけては，例として，サイズが 4×4，多

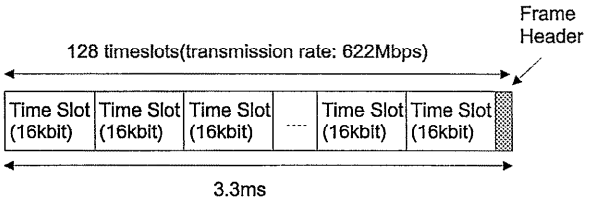


図 1 可変速度 TDM 交換方式の TDM フレームフォーマット
Fig. 1 TDM frame format of variable bit rate TDM switch architecture.

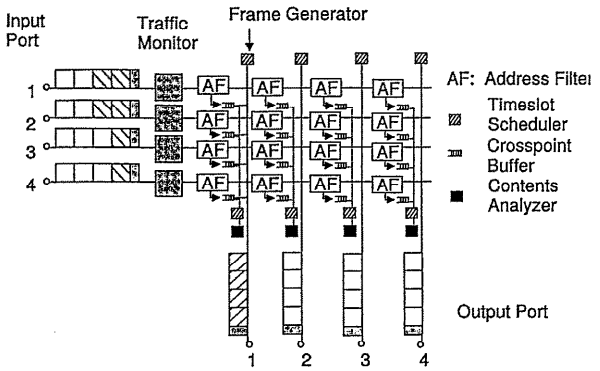


図 2　クロスポイントバッファ型スイッチ
Fig. 2　Cross-point buffer type variable bit rate TDM switch.

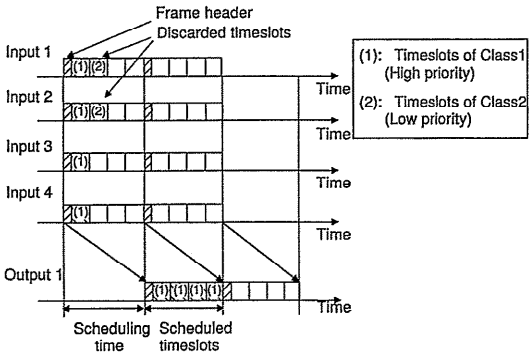


図 4　スイッチ動作のタイムチャート
Fig. 4　Timechart of switch operation.

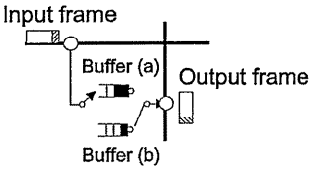


図 3　クロスポイントバッファの構成
Fig. 3　Configuretion of cross-point buffer.

重度 4 のスイッチにおいてある TDM フレーム周期に合計で 6 タイムスロットが入力され，それらがすべて出力ポート 1 に向かう場合の動作を示している．本方式では，各出力ポートに向かうタイムスロットの合計数が多重度より多い場合，多重度数分のタイムスロットのみが転送されて，残りは廃棄される．動作の時間順序は図 4 に示すとおりであり，入力されるタイムスロットに対してあらかじめ優先度の高い順に Class 1, Class 2 の二つの優先度が設定されていれば，優先度の低い Class 2 のタイムスロット 2 個が廃棄され，四つのタイムスロットが次の交換機に向かう．本方式の通信品質は，TDM 交換方式の性質を受けついでおり，エンド・ツー・エンドのタイムスロットの遅延分散は 2 TDM フレーム時間以内におさえられる．

2.2　可変速度トラヒックの交換に関する性質

可変速度動画像トラヒックの交換を行ううえで重要な監視制御を，既存のセル交換機及びパケット交換機の上で行う場合，セルまたはパケットが到着するごとに，到着間隔の監視を行う必要があった．これに対して本方式では，TDM フレームヘッダの情報を解析することで，ヘッダ読込みの直後にフレーム内のすべてのタイムスロットに対する監視制御を開始できる．つまり，監視制御を並列化することで処理速度の高速化が可能である．

また，コネクション受付制御で申告するピークレート平均レートに従って適切なトラヒックシェーピング及び優先度設定を行うことにより，各 TDM フレームで，平均レートに相当する数のタイムスロットに関して，無損失であることが保証される．この性質は以下のようにして説明される．あるコネクション i について，コネクション受付制御によりピークレート及び平均レートを 4.8 Mbps の整数倍に設定したとして，それぞれを $P_i = M_i \times 4.8$[Mbps]，$R_i = N_i \times 4.8$[Mbps] とする．画像フレーム周期を F[s] とし，符号量制御により，各画像フレームの情報量のピーク値，平均値がそれぞれ $P_i \times F$[bit]，$R_i \times F$[bit] でおさえられる場合，その情報を 1 画像フレーム周期の中でトラヒックシェーピングを行って送出すれば，各 TDM フレームに格納されるタイムスロットの数は最大値は M_i 以下，平均値は N_i 以下となる．コネクション受付制御を行えば，エンド・ツー・エンドの各伝送路において，トラヒックの収容効率が 100% を超えることはないから，その伝送路における TDM フレームの多重度を L とすると，$\left(\sum_i N_i < L\right)$ が成り立つため，各コネクションが各 TDM フレームに対して，N_i 個以下の数のタイムスロットに最高の優先度を付与すれば，それらはすべて無損失であることが保証できる．

なお，符号器の生成する 1 枚の画像フレームの符号量が $P_i \times F$ を超える場合には，超える情報量をもつタイムスロットは，次の画像フレーム周期に持ち越して送出することでこの性質は保たれる．

3. 最低帯域保証を実現するトラヒック制御方式

提案するトラヒック制御方式は，次の二つの要素で構成される．

(1) コネクション受付制御

(2) トラヒックシェーピング

本章では，図 5 に示される，ピークレートが 11.5 Mbps，平均レートが 4.8 Mbps のトラヒックを伝送する場合を例にとって説明する．この例では，画像フレーム間隔は 33 ms であるものとする．

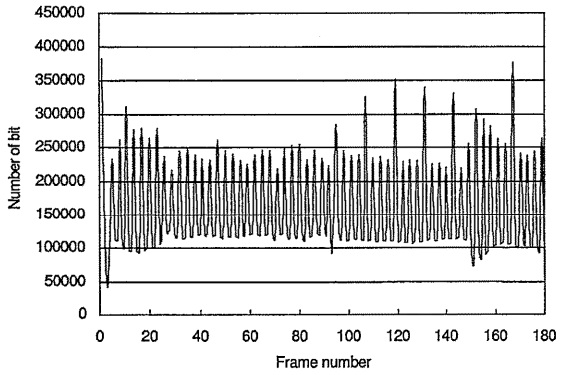


図 5 3. で説明の対象となる入力トラヒックの情報量変動（4. においては条件 1 の入力トラヒックとして利用）
Fig. 5 Example of bit rate fluctuation of applied traffic.

3.1 コネクション受付制御

コネクションのセットアップにあたり，送信端末とネットワークとの間でトラヒックのピークレートと平均レートが決定される．これらのパラメータからタイムスロット損失率を算出する方法については，4. で述べる．

3.2 トラヒックシェーピング

情報転送の段階に入ると，送信端末は，1 枚の画像フレームの情報が格納されたタイムスロットを，画像フレーム時間間隔の中で伝送される各 TDM フレームに均等に割り当てて送る．

図 6 は，タイムスロットの優先度クラスとして Class 1 と Class 2 の 2 段階の優先度を設け，図 5 のトラヒックを伝送する場合のトラヒックシェーピングと優先度付与方法を示している．

トラヒックの平均レートは 4.8 Mbps であるが，これは各 TDM フレームごとに一つのタイムスロットを送出することに相当する．そこで送信端末は，各画像フレーム情報の送出の際，10 のタイムスロットに Class 1 を，残りの数のタイムスロットに Class 2 の優先度をそれぞれ与える．

符号器では，符号化における制御情報の含まれるピクチャヘッダや画像情報のなかでも重要な低周波成分を含む情報を Class 1 タイムスロットに格納することになる．

なお，MPEG2 方式の動画像ストリームの場合にお

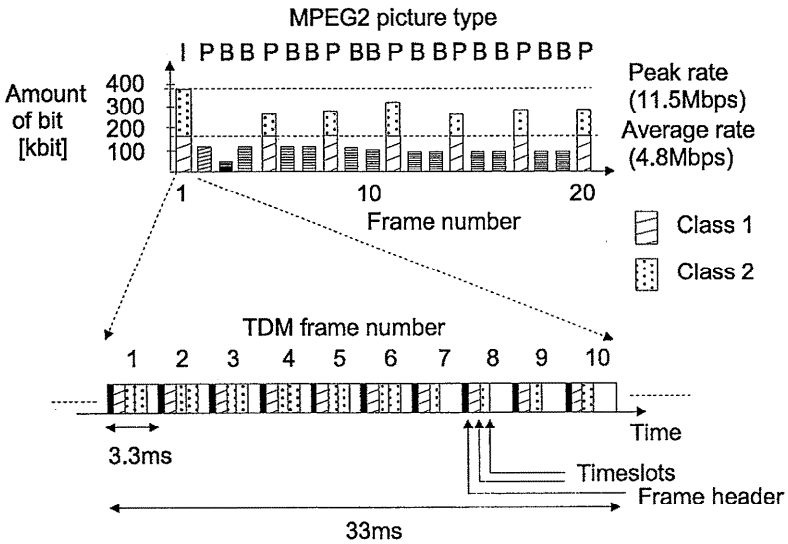


図 6 一般の動画像情報への優先度付与とトラヒックシェーピング
Fig. 6 Assignment of priority level and traffic shaping for normal video streams.

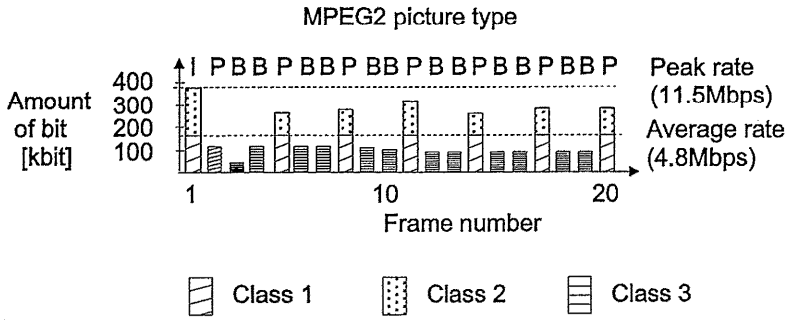


図 7　I，P，B の三つのピクチャタイプをもつ MPEG2 動画像情報への優先度付与
Fig. 7　Assignment of priority level for video streams encoded with MPEG2 where I,P and B pictures are included.

けるBピクチャのように，損失による画像劣化の影響が比較的小さい画像フレームを伝送する場合には，他のフレームよりも，より多くのタイムスロットに低い優先度を与える手法が考えられる．例として，1枚のBピクチャが160 kbit の情報量をもつ場合，提案方式の標準的な動作ではこれらのタイムスロットにすべてClass 1が与えられることになるが，制御情報を含む先頭のタイムスロットにのみ Class 1 を与え，他のタイムスロットにより低い，例えば Class 3 の優先度を与えることにより，タイムスロットの損失が生じるときには Class 3 のタイムスロットが選択的に廃棄され，全体として画像品質の維持が図られる（図 7）．

なお，図 6，図 7 にはトラヒックシェーピングの均等化区間を1画像フレームとする例を示している．均等化区間を長くすれば，ピークレートが減少するため収容できるコネクション数が増加するが，同じ条件でピークレートのみでコネクション受付制御を行う場合の収容可能コネクション数も増加するため，統計多重化効果が改善されることはないと考えられる．本論文では，可変速度交換の長所である，低転送遅延時間の特徴を生かした条件における統計多重化効果を評価するため，次章の性能評価では特に断りがない限り均等化期間を1画像フレーム周期として検討を行っている．

4. 性能評価

提案するトラヒック制御方式の優先制御の効果を明らかにするため，種々の条件のトラヒックを入力し，損失が生ずるタイムスロットの種類とその損失率をシミュレーションにより評価する．また，可変速度交換では，収容するコネクション数とタイムスロット損失率との関連をなるべく簡単な処理で予測できるコネクション受付制御が重要であり，制御に適用できる数式を導出する目的で損失率の数学解析を行う．

4.1 シミュレーションによる性能評価

性能評価は，実画像を MPEG2 方式により符号化し，Test Model 5 方式によりレート制御を行って得た情報を入力トラヒックとして行った．

4.1.1 シミュレーション条件

シミュレーション条件は以下のとおりである．

（a）スイッチのパラメータ

伝送速度	622 Mbps
TDM フレーム長	3.3 ms
スイッチサイズ	16 × 16

（b）入力トラヒックの特性

入力トラヒックとして，表 1〜表 3 に示す三つの条件を設定した．条件 1 と条件 2 は，標準画像 “Table Tennis” を符号化して得られたデータをスイッチへの入力トラヒックとして用いた．画像サイズは 352×240，フレーム数は 180 である．符号化された動画像情報を伝送する際には，ネットワーク上で生じる情報損失による画質劣化を防止するためと，早送り，巻戻しなどの編集を容易にするという主な二つの理由から，ある時間周期で画面全体をイントラ符号化する手法がとられる．このイントラ符号化の方式はトラヒックの変動に大きな影響を与えるため，表 1，表 2 に示す方式で符号化したデータを用意した．

条件 1 のトラヒックの情報変動は図 5 に既に示されている．また条件 2 のトラヒックの情報量変動は図 8 に示している．条件 1 は，MPEG2 で一般的に用いられる符号化方式であり，12 フレームごとに I ピクチャを挿入しているため，周期的に大きなバーストが生じている．一方，条件 2 は，放送・通信用途を考慮して

表1 入力トラヒックの符号化パラメータ（条件1）
Table 1 Encoding parameters of applied traffic under the first condition.

画像の種類	標準画像 "Table Tennis"
GOP サイズ	12
ピクチャタイプ順序	IBBPBBP...
ピークレート	11.5 Mbps
平均レート	4.8 Mbps
ピーク対平均比	2.4
画像サイズ	352 × 240
フレーム数	180

表2 入力のトラヒックの符号化パラメータ（条件2）
Table 2 Encoding parameters of applied traffic under the second condition.

画像の種類	標準画像 "Table Tennis"
GOP サイズ	180
ピクチャタイプ順序	IBBPBBP...
ピークレート	9.9 Mbps
平均レート	4.8 Mbps
ピーク対平均比	2.1
画像サイズ	352 × 240
フレーム数	180

表3 入力トラヒックの符号化パラメータ（条件3）
Table 3 Encoding parameters of applied traffic under the third condition.

画像の種類	テレビドラマ番組
GOP サイズ	12
ピクチャタイプ順序	IBBPBBP...
ピークレート	11.2 Mbps
平均レート	4.8 Mbps
ピーク対平均比	2.1
画像サイズ	720 × 480
フレーム数	1,800

スライスイントラ符号化方式を取り入れている．1枚のPピクチャでイントラ符号化される領域は352×16であり，15枚のPピクチャの伝送を通して画面全体のイントラ符号化が終了する．その結果，周期的なバーストの発生が除かれており，伝送効率の向上が期待できる．更に条件3は，GOPサイズを条件1と同じに設定しているが，画像のサイズが720×486と現在のテレビ放送サービスに近い画像を符号化することにより得られた情報量を用い，フレーム数を1,800としている．

（c） 伝送方法

タイムスロットに与える優先度は優先度の高い順にClass 1，Class 2，Class 3の3種類として，以下のように設定を行った．

[Class 1]

- 各Iピクチャ，Pピクチャについて，10個のタイムスロット

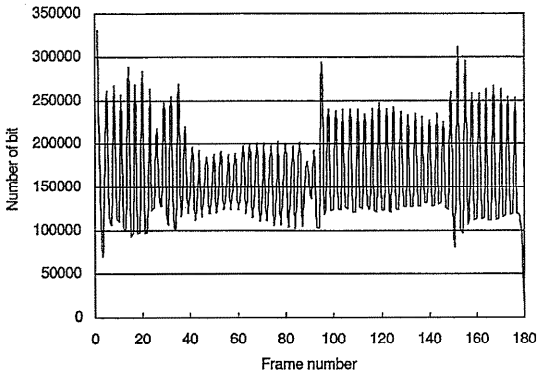


図8 入力トラヒックの情報量変動（条件2）
Fig. 8 Bit rate fluctuation of input traffic under the second condition.

- 各Bピクチャについて，1個のタイムスロット

[Class 2]

各Iピクチャ，Pピクチャの各フレームについて（ピクチャを構成する総タイムスロット数−10）個のタイムスロット

[Class 3]

各Bピクチャの各フレームについて，(ピクチャを構成する総タイムスロット数−1）個のタイムスロット

なお，IピクチャPピクチャにおけるClass 1のタイムスロット数は，各入力トラヒックの平均レートである4.8 Mbpsに対応し，BピクチャにおけるClass 1のタイムスロット数は，ピクチャヘッダの格納に必要な情報量に対応している．なお，シミュレーションの開始にあたり，スイッチに入力される各トラヒックソースのあて先は，スイッチの16本の出力ポートに均等に分布するように選択される．この際，トラヒックの情報量変動は，図5〜図9に示すように周期的であるため，各ソースのIピクチャあるいはPピクチャの情報がスイッチに同じ時刻に入力されないように伝送方法を調節することでタイムスロット損失率を低くすることができる．シミュレーションでは，スイッチに入力する画像情報源の開始フレーム番号をランダムに選択することでこの条件に近い状況を実現している．また，実際の放送サービスでは，多様な画像情報が多重化・伝送されるのに対し，本論文では，同種のトラヒックを多重化し評価を行っていることから，条件を実サービス提供条件に近づけるため，条件1〜3ではスイッチに入力される画像フレーム数が180に達すると，新たに開始フレーム番号をランダムに選び直して

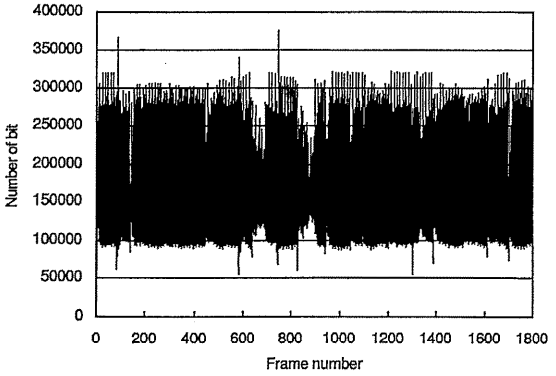


図 9　入力トラヒックの情報量変動（条件 3）
Fig. 9　Bit rate fluctuation of input traffic under the third condition.

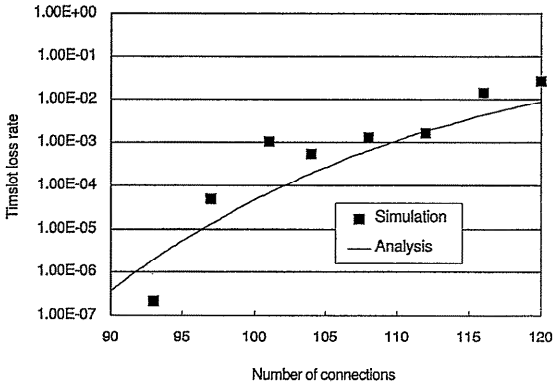


図 11　条件 2 におけるタイムスロット損失率
Fig. 11　Timeslot loss rate measured under the second condition.

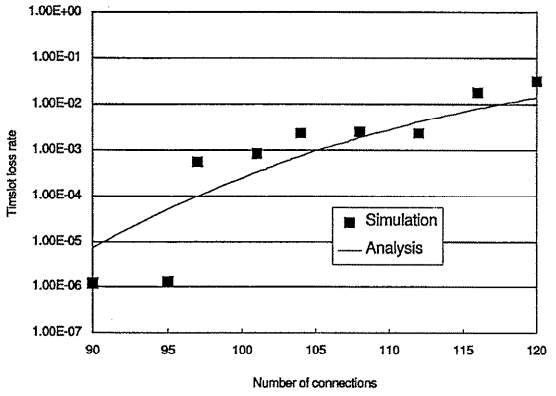


図 10　条件 1 におけるタイムスロット損失率
Fig. 10　Timeslot loss rate measured under the first condition.

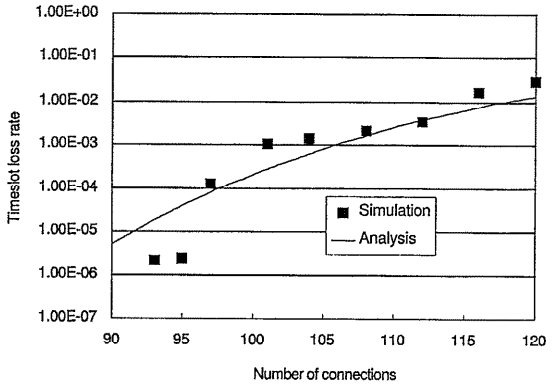


図 12　条件 3 におけるタイムスロット損失率
Fig. 12　Timeslot loss rate measured under the third condition.

入力を再開する．

以上の条件のもとで，各々のトラヒックを入力して，合計で 10^8 個のタイムスロットを送り終わるまでシミュレーションを行い，タイムスロットの損失率の評価を行った．

4.1.2　評価結果

条件 1〜3 の各条件のもとで，与えられる負荷とタイムスロット損失率の関係を図 10〜図 12 に示す．シミュレーションに得られた値は図中で "Simulation" で表された部分である．"Analysis" は数学解析により得られた値であり，**4.2**で述べる．なお，いずれの場合でも，Class 1 のタイムスロットは損失を生じておらず，失われたのは B ピクチャの低優先度情報を有する Class 3 のタイムスロットのみであった．これにより，提案方式の実効性が確認された．

次に，提案方式により得られる統計多重化効果を以下のように定義する．すなわち，ピークレートのみでコネクション受付制御を行う場合に収容できるコネクションの数を N_p，可変レートでコネクション受付制御を行う場合に収容できるコネクションの数を N_a とするとき，統計多重化効果を N_a/N_p と定義する．

まず，ピークレートのみによってコネクション受付制御を行う場合，ピークレートの値は，(a) 任意の実数値を受け入れる場合と，(b) 4.8 Mbps の整数倍のみを受け入れる場合の 2 通りが考えられる．ピークレートでのコネクション受付制御では，伝送される情報の損失率はほぼ 0 に保証する必要があるが，ATM その他の一般の非同期式の交換方式では前者の方法でそれを比較的容易に実現できる．一方，可変速度 TDM 交換方式においても本論文の **2.** で述べたスイッチの動作から，画像情報の損失率が 0 であることを厳密に保証するためには後者を採用すべきである．そこで条件 1〜

3の各トラヒックに対してピークレートを9.6 Mbpsとして伝送することを想定して，収容できるコネクション数を N_p とする．なお，画像フレームのなかには，毎秒当りのビットレートに換算して9.6 Mbpsを超える情報量をもつフレームがあるが，このフレームは，次の画像フレーム周期まで持ち越してトラヒックシェーピングを行って伝送を行う．

次に，可変レートによるコネクション受付制御では，ピークレートにおける整数倍の制限はなく任意の値をとることができる．各送信端末が画像フレームを必ず1画像フレーム周期内でシェーピングして伝送するものとし，表1〜表3に示されたピークレートの値をそのまま用いる．

これらの条件のもとで，ピークレートによるコネクション受付制御では，条件1〜3の各トラヒックで収容できるコネクション数は622/9.6=64と計算される．

次に，可変レートによるコネクション受付制御に関して，損失率と画像品質の関連に関する評価は，既に文献[5]において行われている．ここでは，条件1のトラヒックについて画像の劣化に関する許容度を現行テレビ放送程度に設定して，劣化の度合を5段階で評価している．その結果，損失がBピクチャに限られ，かつその割合が 10^{-3} 以下であれば，評価の平均は，「基準画像に比べ劣化がわかるが気にならない」レベルであることが示されている．条件2についても許容損失率はほぼ同等であると考えられる．

また，条件3で用いられている画像については，被験者6人に対して文献[5]と同じ条件で新たに主観評価実験を行った．損失を受けるピクチャをBピクチャに限定し，損失率を 10^{-2} から 10^{-4} まで変化させて評価を行った結果，損失率が 10^{-3} 未満であれば「基準画像に比べ劣化がわかるが気にならない」レベルであるという結果を被験者全員から得た．

以上のタイムスロット損失率のもとで，収容コネクション数を図10〜図12から算出すると，条件1〜3のいずれの場合も97本となる．

3条件とも，ピークレートと平均レートの比が2倍強であることを考慮すれば，統計多重化効果は十分な値であるということができる．

4.2 タイムスロット損失率の数学解析

可変速度TDM交換方式では，トラヒック制御がTDMフレームごとに独立しているため，ON-OFFモデルで，かつON区間の継続時間がTDMフレーム長に等しい固定長であるモデルで損失率を正確に近似できることが期待できる．本論文では，以下では，ONの発生確率がベルヌーイ到着分布に従うとして損失率を求める．

表4 各条件のトラヒックに対して，可変レートでコネクション受付制御を行う場合に得られる統計多重化効果

Table 4 Statistical multiplexing gain under each condition.

条件	許容損失率	N_p	N_a	統計多重化効果
条件1	10^{-3} 以下	64	97	1.5
条件2	10^{-3} 以下	64	97	1.5
条件3	10^{-3} 未満	64	97	1.5

可変速度TDM交換方式のパラメータのうち，伝送速度を Cl，TDMフレームの多重度を L と定義し，一つのコネクション i に割り当てられたピークレートを RPi，平均レートを RAi とするとき，一つのTDMフレームに存在するそのコネクションのタイムスロット数を l として，l の確率分布 $Pi(l)$ を次の式で表す．

$$Pi(l) = \begin{cases} p & (l = RPi/Cl \times L) \\ 1-p & (l = 0) \\ 0 & (\text{その他の } l \text{の値}) \end{cases}$$

$$(\text{ただし}, p = RAi/RPi)$$

このとき，平均のタイムスロット損失率が次の式で与えられる．

$$\sum_{k=1}^{N} P_1 \times P_2 \times \ldots P_N(k) \times \frac{\max\left(\sum_{i=1}^{k} RPi/Cl \times L - L, 0\right)}{\sum_{i=1}^{N} RAi/Cl \times L}$$

（ただし，$P_1 \times P_2 \times \ldots P_N(k)$ は，N 本のコネクションのなかの合計で k 本がONになる確率を表す）

特にすべてのコネクションのトラヒックパラメータが等しい場合は，ピークレートを RP，平均レートを RA と置くと次のとおりになる．

$$\sum_{k=1}^{N} {}_NC_k \times p^k \times (1-p)^{N-k} \times \max(k \times RP - Cl, 0)/(N \times RA)$$

図10〜図12には，それぞれ解析により得られる値を合わせて示しているが，いずれもシミュレーション

結果をよく近似できており，コネクション受付制御に有効であるといえる．

5. む す び

本論文では，可変速度 TDM 交換機による動画像伝送専用網の上で動作するトラヒック制御方式として，可変速度トラヒックに対し，無損失保証及び遅延分散保証の機能をもつトラヒック制御方式の提案を行った．また，無損失が保証されない情報量の損失率について，MPEG2 Test Model 5 方式により符号化して得られる画像情報を加入者交換機に入力するという条件のもとで，シミュレーションによる評価を行い，得られる統計多重化効果の議論を行った．この結果，送信端末が各画像フレームを 1 画像フレーム周期でトラヒックシェーピングして伝送し，かつ端末の間でトラヒックの送出タイミングを適切に調節するという制御を行えば，ピークレート対平均レートが 2 倍程度のトラヒックに対して，実用上十分な統計多重化効果を得ることができることを明らかにした．更に，タイムスロット損失率を近似する解析式として，TDM フレームを周期としてトラヒック量を ON-OFF モデルで表すことにより得る式が有効であり，コネクション受付制御への適用可能性があることを示した．

なお，本論文で示した，規模拡張性をもつトラヒック制御は，動画像に限らず一般のデータ通信にも有効である．今後は，TCP/IP のデータ通信トラヒックに適用する場合の性能評価を行っていくことが課題である．

謝辞　本研究は日本学術振興会「未来開拓学術研究振興事業」に基づく研究プロジェクト「ポスト ATM 高度情報通信ネットワークアーキテクチャ」(JSPS-RFTF-96P00601) の一部として行われたものであり，ここに謝意を表する．

文　　献

[1] "Traffic Control and Congestion Control in B-ISDN," ITU-T Recommendation I.371, July 1995.

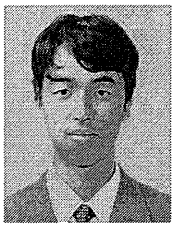
[2] ウドムキャット ブンウォラセト，青木輝勝，相田　仁，齊藤忠夫，"連続メディアに適したクロスポイントバッファ型可変速度時分割スイッチアーキテクチャ，"1998 信学ソ大，B-6-14, Sept. 1998.

[3] "Information Technology - Generic coding of moving pictures and associated audio," ISO-IEC 13818-2, Draft International Standard, March 1994.

[4] "Test Model 5," ISO/IEC JTC1/SC29/WG11 MPEG 93/457, April 1993.

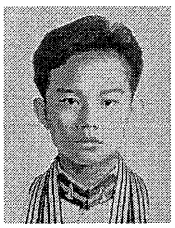
[5] 大西隆之，青木輝勝，相田　仁，齊藤忠夫，"可変速時分割スイッチを用いた MPEG2 ストリーム伝送の画質主観評価，"1999 信学総大，B-6-1, March 1999.

(平成 11 年 9 月 3 日受付，12 年 5 月 24 日再受付)



森野　博章　(正員)

平 6 東大・工・電子卒．平 11 同大大学院博士課程了．現在，同工学系研究科リサーチアソシエイト．リング型 LAN のアクセス制御方式，ATM の可変速度動画像トラヒック制御方式，大容量動画像伝送用交換方式の研究に従事．博士 (工学)．



ウドムキャット ブンウォラセト (学生員)

平 6 タイ・チュラロンコーン大・工・電気卒．平 9 東大大学院工学系研究科電子工学専攻修士課程了．平 12 同電子情報工学専攻博士課程了．在学中，高速交換ネットワーク，マルチメディア通信プロトコル，ATM トラヒック制御に関する研究に従事．博士 (工学)．

相田　　仁　(正員)

昭 55 東大・電気卒．昭 60 同大大学院博士課程了．同大工・電気・助手，昭 60 同大講師，SRI インターナショナル (アメリカ) 客員研究員を経て，現在，東大・教授．工博．高品質ネットワーク，並列・分散処理などの研究に従事．著書「第 5 世代コンピュータ入門」(共著)．情報処理学会，人工知能学会，ソフトウェア科学会，IEEE，ACM 各会員．

齊藤　忠夫　(正員)

昭 38 東大・工・電子卒．昭 43 同大大学院博士課程了．東大工学部講師，同助教授，カリフォルニア工科大研究員などを経て，現在，東大教授．ディジタル通信方式，コンピュータネットワーク，分散処理などの研究に従事．工博．著書「電子回路入門」(昭晃堂，昭 52)，「ディジタル通信網入門」(オーム社，平 2)．昭 42 年度，昭 61 年度本会論文賞，昭 48 年度同業績賞受賞．電気学会，情報処理学会，IEEE 各会員．

UDC [621.395.345:621.394.42]:621.395.658.062

時分割スイッチ回路網における対関係の構成

正 員 斉藤 忠夫† 正 員 荻野 長生†
非会員 後藤 文広†* 正 員 猪瀬 博†

On Pair Relationship in Time Division Switching Networks

Tadao SAITO† *and* Nagao OGINO†, *Regular Members*, Fumihiro GOTO†*, *Nonmember*, Hiroshi INOSE†, *Regular Member*

あらまし ＰＣＭ伝送路では音声信号を一方向にしか伝送することができないため，ＰＣＭ電話交換方式で用いられる時分割スイッチ回路網は本質的に４線式となる．そのため一つの接続を設定するのに，１対の経路が必要になる．これを解決するためスイッチ回路網中のスイッチやリンクの間に対関係を設け，片方の経路から自動的にもう一方の経路を設定することが一般に行われている．対関係を設ければ，スイッチ回路網の呼損率，経路探索処理量，制御用ハードウェア量などを小さくすることができる．このように，対関係は能率的な時分割スイッチ回路網を構成する上で重要であるが，対関係についての一般的な理論はまだ存在しない．本論文では，時分割スイッチ回路網の対関係について統一的な検討を行う．まず対関係の定義を述べ，対関係を幾つかの形に分類し，対関係の具体的な実現例を示す．更に重要な対関係の構成法を述べ，対関係を設けることによって生じる接続形式とリンク接続パターンの変更についても簡単に触れる．最後にこれらの検討結果から，望ましい対関係の構成法を明らかにしている．

1. まえがき

最近のディジタル技術の発展によって，時分割交換方式が従来の空間分割交換方式に取って代わろうとしている．ＰＣＭ信号は１方向にしか伝送することができない．従って，電話などの両方向性サービスを実現するＰＣＭ時分割スイッチ回路網は本質的に４線式となる．

４線式時分割スイッチ回路網では，一つの接続を設定するのに，発信側から着信側への経路と着信側から発信側への経路が必要である．前者の接続を順方向接続，後者の接続を逆方向接続と呼ぶ．順方向接続と逆方向接続は独立に設定することも可能であるが，この方式ではスイッチ回路網の制御処理量が２線式のときと比べて２倍になってしまう．

対関係（pair relationship）は順方向接続から自動的に逆方向接続を設定するために，スイッチ回路網中のリンクの間に特定の関係を与える方法である．この関係を与えておけば，１対の接続を設定するのに１回の経路探索を行えば良く，又，呼損率も低減することができる．１対のスイッチを単一の制御信号で制御できるので，多くの場合制御信号用メモリと復号用ゲートの数を節約することができる．

このように，対関係は能率的な時分割スイッチ回路網を構成する上で重要である．小規模な時分割ハイウェイスイッチ回路網については従来からも対関係の研究があり[1]，又，Ｔ－Ｓ－Ｔ形スイッチ回路網を対象として対関係の分類も行われている[2]．しかし，対関係についての一般的な理論はまだ存在しない．

本論文では，時分割スイッチ回路網における対関係を統一的に検討する．まず対関係の定義を述べ，対関係を幾つかの形に分類する．次に各時分割スイッチ回路網上で実現可能なさまざまな形の対関係を示し，幾つかの重要な対関係の構成法を明らかにする．最後に対関係を設けることによって生じる接続形式と，特定の接続形式において対関係を設けるために必要となるリンク接続パターンの修正方法について述べる．

† 東京大学工学部電子工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan
* 現在，通産省
論文番号：昭 56-414[B-133]

2. 対関係の定義

本論文では，対象とするスイッチ回路網を均質なスイッチ回路網に限っている．均質なスイッチ回路網とは，以下の条件を満足する回路網のことである．

〔性質１〕 回路網の各入線は，すべての出線に接続することができる．

〔性質２〕 第 i スイッチ段のスイッチのすべての出線は，第 $i+1$ スイッチ段のスイッチに接続されている．第 i スイッチ段のスイッチのすべての入線は，第 $i-1$ スイッチ段のスイッチに接続されている．

〔性質３〕 同一段のスイッチはすべて大きさが等しく，同じ機能を持つ．すなわち，拡大，あて先決定，収束，拡大収束のうちの一つの機能を持つ．同一スイッチ段においては，経路の拡大数，収束数は一定であり，それぞれ対応する収束段，拡大段も等しい．

性質１～３より，均質なスイッチ回路網では，すべての入線出線対に対するチャネルグラフが等しくなり，議論が簡単になる．現実のほとんどのスイッチ回路網は性質１～３を満たす．均質なスイッチ回路網においては，対関係は数学的に以下のように定義される．

〔定義〕 対関係 σ は，次のような条件を満足するスイッチ回路網 $\boldsymbol{N}$ からスイッチ回路網 $\boldsymbol{N}$ への写像である．

(1) $\sigma^2 = \mathrm{id}$．（恒等写像）

(2) スイッチ回路網 $\boldsymbol{N}$ は第０スイッチ段から第 n スイッチ段までのスイッチ段で構成されているものとする．第 i スイッチ段のスイッチの集合を $\boldsymbol{V}_i$ で表せば，

$$\sigma(\boldsymbol{V}_i) \subset \boldsymbol{V}_{n-i}$$

(3) 写像 σ は連続である．すなわち，リンク又はスイッチが連結している状態を $x \cdot y$ で表せば，

$$\sigma(x \cdot y) = \sigma(x) \cdot \sigma(y)$$

(4) 経路すなわち幾つかのリンクが連結したものを $\boldsymbol{P}$ で表し，すべてのリンクの集合を $\boldsymbol{L}$ とすれば，

$$\sigma(\boldsymbol{P}) \cap \boldsymbol{P} \cap \boldsymbol{L} = \phi$$

写像 σ が上に述べた四つの条件を満足するとき，x と $\sigma(x)$ は対関係にあると定義する．

条件(1)は１本のリンクが１本のリンクと対になり，一つのスイッチが一つのスイッチと対になることを示す．条件(2)は第 i 段における一つのスイッチが第 $n-i$ 段における一つのスイッチと対になることを示す．条件(3)はリンクやスイッチが互いに接続されていれば，それらと対関係にあるリンクやスイッチも互いに接続できることを示す．条件(4)は互いに対関係にある２本の経路は共通なリンクを含んではいけないことを示す．

3. 対関係の分類

対関係はさまざまな種類が考えられる．小規模時分割ハイウェイスイッチ回路網における対関係は，その構成方法によって時間４線式，空間４線式，混合４線式に分類される[1]．本章では一般の均質なスイッチ回路網を対象として，対関係を４とおりの方法で分類する．初めの２とおりの方法は，対関係の構成の仕方に着目して分類する方法であり，後の２とおりの方法は，対関係の性質に着目して分類する方法である．

3.1 分離網と一体網[2]

分離網（separated network）では，１対の接続を設定するのに同様な構造を持つが逆方向を向いた二つの回路網が使用される．一方を順方向スイッチ回路網，他方を逆方向スイッチ回路網と呼ぶ．これに対して一体網（combined network）では，１対の接続は一つの回路網を使って設定される．すなわち，互いに対関係にある２本のリンク又は二つのスイッチは，一つの回路網に含まれる．

一体網においてはスイッチ回路網は対称でなければならない．一方，分離網においてはスイッチ回路網は対称である必要はないが，一方の回路網の入線と他方の回路網の出線を接続することはできない．従って，加入者線や両方向中継線を接続するときには，分離網は能率が低い．分離網は入中継線を出中継線に接続する場合にのみ有利である．

分離網を両方向回線の交換に適用する手段としては，折り返し経路を付加する方法と回路網の両側に回線を収容する方法がある．しかし前者の方法では，呼が回路網を２回通るので，能率の低下を招く．又，接続品質が不均一になり，チャネルグラフが一様でなくなるために接続制御が複雑化する．後者の方法では，収容する入線と出線の数が増加するので，回路網の規模が大きくなってしまう[3]．

3.2 グリッド内対とグリッド間対[2]

時分割スイッチ回路網は，パルスシフタ段またはメモリスイッチ段である時間段と，時分割多重スイッチ段である空間段とから構成される．ここでパルスシフタとは空間的に１×１の時間スイッチを指し，メモリスイッチとはそれ以外の時間スイッチを指す．時分割スイッチ回路網の空間的構成を表すために，時分割スイッチ回路網からパルスシフタのみを取り除いた回路網を，空間回路網（space network）と呼ぶことにする．

空間回路網の中では種々の部分回路網が構成できる．本論文では，特に空間回路網の初段と終段を除くことによって構成される部分回路網をグリッドと呼ぶことにする．もちろん，グリッドが1個しか存在しない回路網もあり得る．スイッチ段数が2段の場合には，それぞれの中間リンクを一つのグリッドと考える．

各グリッド内に1対の接続を設定するのに十分な構造があれば，1対の接続を一つのグリッドによって実現することができる．このような対方式をグリッド内対（intragrid pair）と呼ぶ．一方，1対の接続を設定するのに1対のグリッドを利用するとき，これをグリッド間対（intergrid pair）と呼ぶ．

3.3 同時対と非同時対[2]

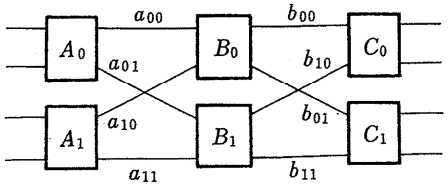
この分類は，同時に動作する素子の間で対を作るか，同時に動作しない素子の間で対を作るかという分類である．同時対（concurrent pair）においては，後で述べるように，1対の接続を設定することができないような接続形式が存在し得る．しかし，一つの制御系のみで対関係にある二つのスイッチを制御することができる．

一方，非同時対（nonconcurrent pair）においては，対関係にある二つのスイッチを制御するために，二つの制御系が必要になる．一つの制御メモリのみで対関係にある一つのスイッチを制御する場合には，特別なハードウェアを付加しなければならない[4]．あるいは各フレーム内で2回制御信号を読み出すことができるような2倍の動作速度を持つ制御メモリが必要になる．

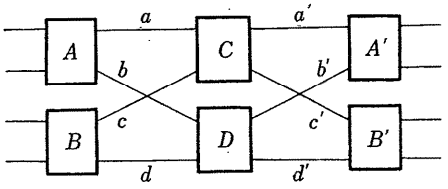
3.4 弱対，準弱対，準強対，強対

2.で述べた条件のうち最初の三つだけしか満足しない対関係を弱対（weak pair）と呼ぶ．特定の入側スイッチと出側スイッチを接続するために特定のタイムスロットが選択される場合を除き，**2.**の条件をすべて満足する対関係を準弱対（quasi weak pair）と呼ぶ．常に**2.**の条件をすべて満足するが，接続すべき入側スイッチと出側スイッチの相対的位置関係によって，利用可能な経路数が変化するような対関係を準強対（quasi strong pair）と呼ぶ．**2.**の条件をすべて満足し，しかも利用可能な経路数が常に一定であるような対関係を強対（strong pair）と呼ぶ．強対は分離網においてのみ構成することが可能であり，逆に分離網上で構成される対関係はすべて強対になる．

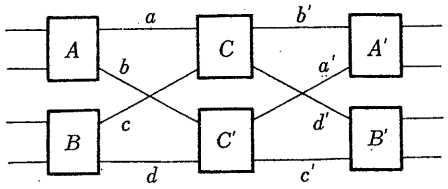
図1は2×2スイッチ3段から成るスイッチ回路網を示している．図1(a)に示したように，スイッチ回路網を構成する各スイッチと各リンクには記号が割り当てられている．図1(b)はリンク $a_{i,j}$ とリンク $b_{j,i}$ が対になる場合を示している．図1(b)では，互いに対関係にある2本のリンクと二つのスイッチを同一の記号で表している．この場合，リンク $a_{0,0}$ とリンク $b_{0,1}$ を使用して入側スイッチ A_0 と出側スイッチ C_1 を接続する経路は，リンク $a_{1,0}$ とリンク b_{00} を使用して入側スイッチ A_1 と出側スイッチ C_0 を接続する経路と対になる．一方，リンク a_{00} とリンク $b_{0,0}$ を使用して入側スイッチ A_0 と出側スイッチ C_0 を接続する経路は，互いに対となる経路を持たない．すなわちこの場合，一部の経路はそれ自身と対になる．このため，ある入側スイッチと出側スイッチ間では，1対の経路が設定



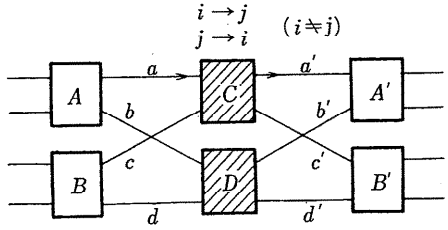
(a) Switching network composed of three stages.



(b) An example of weak pair.



(c) An example of quasi strong pair.



(d) An example of quasi weak pair.

図1 弱対，準強対，準弱対
Fig.1 – Weak pair, quasi strong pair and quasi weak pair.

できないか，経路対の数が減少する場合が起り得る．このような対関係を弱対と呼ぶ．

図1(c)はリンク a_{ij} とリンク b_{ki} ($k=j+1 \bmod 2$) が対になる場合を示している．この場合，自分自身と対になる経路は存在しない．しかし，互いに対関係にない入側スイッチと出側スイッチ間には2組の経路対が設定できる反面，互いに対関係にある入側スイッチと出側スイッチ間には1組の経路対しか設定できない．これはある入側スイッチと出側スイッチを接続する際に利用可能な経路群と，その経路群と互いに対関係にある経路群が重なるためである．従って，この対関係は準強対になる．

図1(d)は中間段の空間スイッチをメモリスイッチで置き換えたスイッチ回路網を示している．リンク $a_{i,j}$ とリンク b_{ji} が対関係にあると仮定する．経路を設定するのにメモリスイッチの入側と出側でそれぞれタイムスロット i と j を使用すれば，その経路と対関係にある経路はメモリスイッチの入側と出側でそれぞれタイムスロット j と i を使用する．このスイッチ回路網は図1(b)のスイッチ回路網と似ているが，すべての入側スイッチと出側スイッチ間で，ほとんど同数の経路対を設定することが可能である．互いに対関係にある入側スイッチと出側スイッチを接続する場合でも，メモリスイッチの両側で同一のタイムスロットを使用しない限り，自分自身と対になることはない．すなわちこの対関係は準弱対になる．

4. 対関係の具体例

本章では時分割スイッチ回路網を対象として，実現可能なさまざまな対関係を示す．時間スイッチとしては，フルアクセスのパルスシフタとフルアクセスのメモリスイッチを仮定する．すなわち，入側の各タイムスロットは，出側のすべてのタイムスロットに接続できるものとする．又，時間スイッチの入側と出側のフレーム位相は，互いに同期しているものと仮定する．分離網の場合には，順方向回路網と逆方向回路網のフレーム位相は，互いに同期しているものとする．これより同時対とは，同一番号のタイムスロットの間で対を作ることに等しくなる．

表1　T-S-T形スイッチ回路網における対関係

	同時対		非同時対	
	グリッド間対	グリッド内対	グリッド間対	グリッド内対
一体網	準強対	弱対	準強対	準強対
分離網	強対	強対	強対	強対

4.1　T－S－T形スイッチ回路網

初段と終段に時間スイッチを使用し，それ以外の段では時間スイッチを使用していない回路網を，ここではT－S－T形スイッチ回路網と定義する．表1は，T－S－T形回路網上で実現可能な対関係を示す．

図2は一体網同時グリッド間準強対方式の例を示す．図2では例えばスイッチ A と A' が互いに対をなす．又，1対の接続を設定するのに同一のタイムスロットが使用される．時間スイッチとしてパルスシフタを使用する場合には，この対方式は厳密には常に同時対になるとはいえない．すなわち対関係にある二つのパルスシフタ間に1対の経路を設定するには，二つの異なるタイムスロットが必要になる．時間スイッチとしてメモリスイッチを使用する場合には，常に同時対を構成することができる．

図3は一体網同時グリッド内弱対方式の例を示す．この場合には自分自身と対になる経路が存在し，グリッド間対の場合と比較して，利用可能な経路の数が半分に減少してしまう．これは後で述べるように，リンクの接続パターンを変更することによって解決することができる．すなわち同時グリッド内対の場合でも，一部のリンク接続パターンを変更することによって，準強対が構成可能である．しかし，リンク接続パターンを変更して得られた回路網は，もはや均質なスイッ

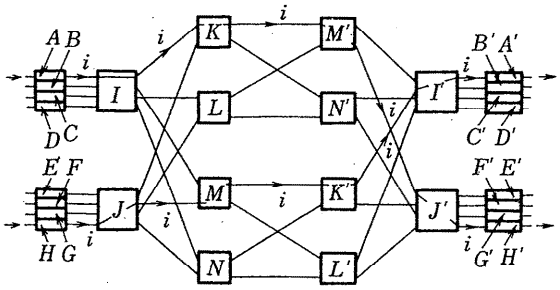


図2　一体網同時グリッド間準強対方式
Fig.2 – Combined concurrent intergrid quasi strong pair scheme.

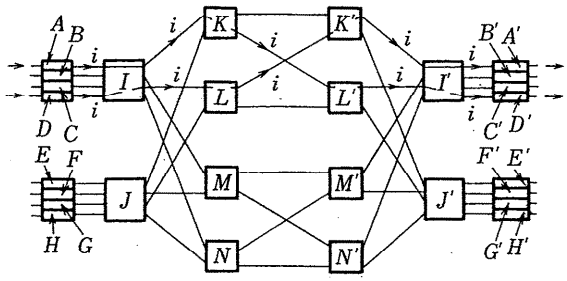


図3　一体網同時グリッド内弱対方式
Fig.3 – Combined concurrent intragrid weak pair scheme.

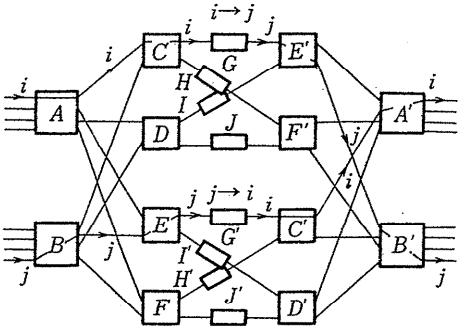


図4　一体網同時グリッド間準強対方式
Fig.4 – Combined concurrent intergrid quasi strong pair scheme.

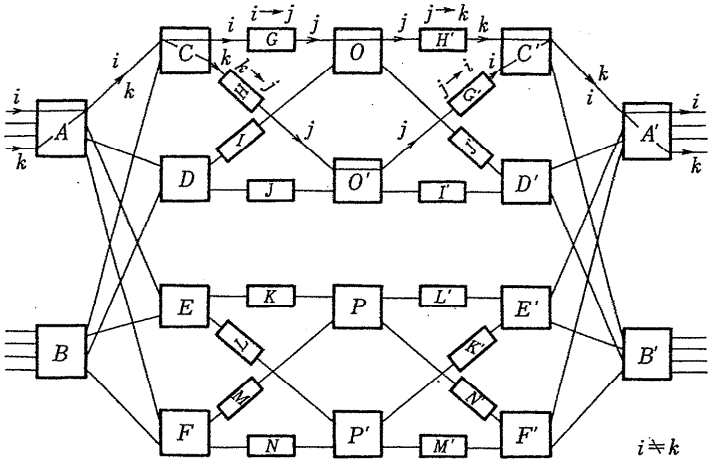


図5　一体網同時グリッド内準弱対方式
Fig.5 – Combined concurrent intragrid quasi weak pair scheme.

表2　S－T－S形スイッチ回路網における対関係

	同時対		非同時対	
	グリッド間対	グリッド内対	グリッド間対	グリッド内対
一体網	準強対 準弱対	準弱対	準強対 準弱対	準弱対
分離網	強対	強対	強対	強対

チ回路網ではなくなる可能性がある[5]．グリッド内対の場合も，時間スイッチとしてパルスシフタが用いられていれば，互いに対となるパルスシフタを接続する際に二つのタイムスロットを用いなければならない．一対網非同時対の場合には，グリッド内対であっても準強対が成立する[4]．時分割ハイウェイスイッチ回路網における時間4線式と混合4線式は，この非同時対の一つの例である[1]．

4.2　S－T－S形スイッチ回路網

初段と終段を除く少なくとも一つのスイッチ段で時間スイッチが使用されている回路網を，ここではS－T－S形スイッチ回路網と定義する．表2はS－T－S形回路網上で実現可能な対関係を示す．

図4は一体網同時グリッド間準強対方式の例である．この場合，初段と終段にメモリスイッチ又は時分割多重スイッチが使用されている限り，準強対が成立する[6]．しかし，初段と終段にパルスシフタが使用されているときは，互いに対関係にあるパルスシフタ間を接続する場合，入側パルスシフタでタイムシロット i が選択されたならば，出側パルスシフタでタイムスロット i を選択することができなくなる．すなわち，このときは準弱対になる．

図5は一体網同時グリッド内準弱対方式の例を示す．S－T－S形回路網では，グリッド内対であっても準弱対が成立する．但し，時間段の数が偶数であり，同時対でなければならないときは，中央の二つの時間段の間では同一のタイムスロットを用いて1対の経路を設定しなければならなくなる．このため図5で示した回路網のように，中央の二つの時間段の間では，利用可能な空間分割経路が少なくとも2本存在しなければならず，自分自身と対になる空間分割経路があってはならない．

一体網非同時グリッド間対は，同時対の場合と同様に，初段と終段にパルスシフタが用いられているか否かによって，それぞれ準弱対，準強対になる．又，一体網非同時グリッド内対では，二つの異なったタイムスロットが対になるので，時間段の数が偶数であっても，同時対の場合と違って中央の二つの時間段の間で2本の空間分割経路が存在する必要はなく，常に準弱対が成立する．

5. 対関係の構成法

本章では一体網の構成法について述べる．分離網の構成は容易なので，ここでは触れない．

まず，弱対 σ_w が構成可能な空間回路網を σ_w 回路網（σ_w－network）と呼べば，σ_w 回路網について以下の定理が成り立つ．

〔定理1〕

(1)　σ_w 回路網は対称な回路網でなければならない．

(2)　すべての対称な回路網は σ_w 回路網になる．

（証明略）

ここで対称な回路網とは，スイッチサイズとチャネルグラフが線対称な回路網のことを指す．

図6は弱対関係の構成例を示す．図6(a)には五つの

スイッチ段から成る空間回路網が示されている．図6(b)に示したように，まず，この回路網をそれぞれ段1から段3，段3から段5までで構成される二つの回路網に分割する．この回路網は対称なので，右側か左側の段3におけるスイッチの順序を変えることによって，互いに線対称な二つの回路網が得られる．この状態で，鏡像関係（mirror relationship）によって二つのリンクや二つのスイッチの間に対応を付ける．再び元の回路網を構成すれば，この対応が弱対関係になっている．図6(a)では，このようにして得られた弱対関係にあるスイッチを同一の記号で表している．

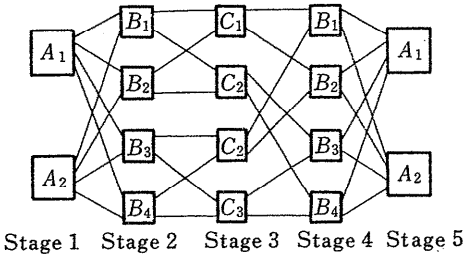
σ_w 回路網の入線と出線にパルスシフタを接続するか，又は σ_w 回路網の初段と終段の空間スイッチを同一の大きさを持つメモリスイッチで置き換えれば，T－S－T形回路網におけるグリッド内対が構成される．

準強対 σ_s が構成可能な空間回路網を σ_s 回路網（σ_s－network）と呼べば，σ_s 回路網について以下の定理が成り立つ．

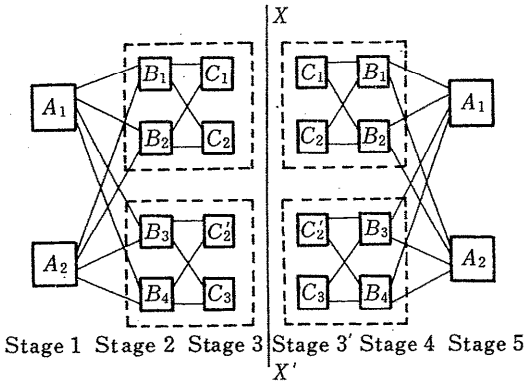
〔定理2〕

(1) σ_s 回路網においては，グリッドの数は偶数でなければならない．

(2) グリッドの数が偶数である σ_w 回路網は，σ_s 回路網になる．（証明略）



(a) Weak pair relationship.



(b) Mirror relationship.

図6　σ_w 回路網の構成
Fig.6 – Construction of σ_w – network.

図7は準強対の構成法を示している．回路網1はグリッド G_1 と，グリッド G_1 を初段スイッチと終段スイッチに接続するリンク群とから構成されている．回路網2はグリッド G_2 と，グリッド G_2 を初段スイッチと終段スイッチに接続するリンク群とから構成されている．回路網1と回路網2は同一の構造を持つ．回路網1と回路網2の構成要素をそれぞれ対応する回路網2と回路網1の構成要素に移す写像を τ とすれば，写像 σ_s（$=\sigma_w \cdot \tau = \tau \cdot \sigma_w$）が準強対関係になる．

σ_s 回路網の入線と出線にパルスシフタを接続するか，又は σ_s 回路網の初段と終段の空間スイッチを同一の大きさを持つメモリスイッチで置き換えれば，T－S－T形回路網におけるグリッド間対が構成される．

σ_s 回路網のリンク段に対称にパルスシフタをそう入するか，σ_s 回路網の空間スイッチ段を対称にメモリスイッチ段で置き換えれば，S－T－S形回路網におけるグリッド間対が構成される．この場合，前章で述べたように，対関係は準強対か準弱対になる．

σ_w 回路網のリンク段に対称にパルスシフタをそう入するか，σ_w 回路網の空間スイッチ段を対称にメモリスイッチ段で置き換え，更に二つの異なったタイムスロット間で対関係を設定すれば，S－T－S形回路網における非同時グリッド内準弱対が構成される．又，偶数個のスイッチ段を持つ σ_w 回路網の対称な奇数個のリンク段にパルスシフタをそう入するか，奇数個のスイッチ段を持つ σ_w 回路網の対称な奇数個のスイッチ段をメモリスイッチ段で置き換え，同一のタイムスロット間で対を設ければ，奇数時間段S－T－S形回路網における同時グリッド内準弱対が構成される．

偶数時間段S－T－S形回路網における同時グリッド内準弱対を構成する際には，次の定理が成り立つ．

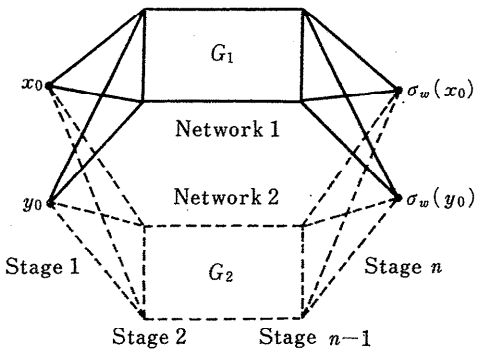


図7　σ_s 回路網の構成
Fig.7 – Construction of σ_s – network.

〔定理3〕 スイッチ段1からスイッチ段nまでで構成されるσ_w回路網において，次の(1)，(2)，(3)は同値である．

(1) 段iで拡大して段jで収束する経路数を$l_{i,j}$で表せば，$l_{k,n-k+1}$（$2 \leqq k \leqq [n/2]$）の少なくとも一つは偶数である．

(2) 弱対σ_wを変形することにより，スイッチ段kとスイッチ段$n-k+1$の間では，二つの空間的に異なった経路を互いに対にすることができる．

(3) 偶数個のリンク段にパルスシフタをそう入するか，偶数個のスイッチ段をメモリスイッチ段で置き換えなければならない場合でも，同時グリッド内準弱対を構成することができる． （証明略）

奇数個のスイッチ段を持つσ_w回路網のリンク段に偶数個のパルスシフタをそう入して同時グリッド内準弱対を構成するには，スイッチ段kとスイッチ段$n-k+1$の間に少なくとも1対のパルスシフタをそう入しなければならない．又，偶数個のスイッチ段を持つσ_w回路網の偶数個のスイッチ段をメモリスイッチ段で置き換えることによって同時グリッド内準弱対を構成するには，スイッチ段kからスイッチ段$n-k+1$までの少なくとも1対のスイッチ段をメモリスイッチで置き換えなければならない．

6. 接続形式とリンク接続パターン

6.1 接続形式の数

一体網では強対関係を構成することはできない．すなわち，ある1対の入側スイッチと出側スイッチを接続する場合，利用可能な空間分割経路群と，それらと対関係にある空間分割経路群とが重なり合ってしまうのである．そのため利用可能な空間分割経路の数が，実質的に減少してしまう．

利用可能な空間分割リンクの数は，接続すべき入側スイッチと出側スイッチの相対的位置関係によって変化する．利用可能なリンク数の違いによって，幾つかの接続形式（connection case）が生じる．接続形式の数は，時分割スイッチ回路網の空間的な構成のみに依存する．時分割スイッチ回路網からパルスシフタを取り除いた回路網を前に空間回路網と定義したが，接続形式の数は空間回路網の段数が増える程多くなる．

〔定理4〕 n個のスイッチ段から成る空間回路網における接続形式の数は，2以上$[n/2]+1$以下である． （証明略）

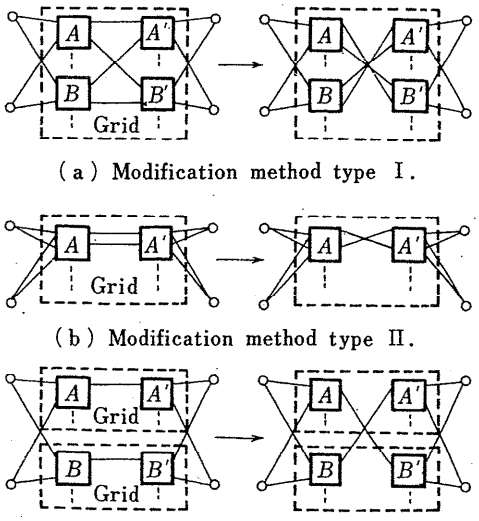
パルスシフタが空間回路網の入線と出線に接続される場合には，互いに対関係にあるパルスシフタ間の接続か否かによって，接続形式の数は更に一つ増加する．

6.2 リンク接続パターンの変更

T－S－T形スイッチ回路網の一体網同時グリッド内対方式の場合は，弱対しか成立しない．しかし，空間回路網の初段における経路の拡大数と，終段における経路の収束数が2以上のときは，リンク接続パターン（link pattern）を変更することにより，準強対を構成することができる．4段の空間回路網を持つT－S－T形回路網を考えると，中央のリンク段の接続パターンを変えることによって，準強対を構成することができる．この場合，リンク接続パターンの変更方法は三つのタイプが考えられる．

図8(a)は1対の初段スイッチと終段スイッチを接続するのに，1グリッドの中で複数の2段目スイッチと3段目スイッチが利用可能な場合を示す．図8(b)は1グリッドの中で1個の2段目スイッチと3段目スイッチしか利用できないが，各段のリンクは複数本使える場合を示す．図8(c)はあるリンク段で1本のリンクしか使用できない場合を示す．それぞれ図に示したように，互いに対関係にある2段目スイッチと3段目スイッチの間のリンクの接続パターンを変えることによって，準強対を構成することができる．タイプⅢでは二つの異なったグリッド間にリンクを接続し直す．

1グリッド中に含まれるスイッチ数が各段それぞれ2個しかないときは，タイプⅠは1グリッドを2グリッドに分割して，その2グリッドを使用してグリッド



(a) Modification method type Ⅰ.

(b) Modification method type Ⅱ.

(c) Modification method type Ⅲ.

図8 リンク接続パターンの変更
Fig.8 – Modification of link pattern.

間対を設けるのと等価になる．又，このとき，タイプⅢは二つのグリッドから別の二つのグリッドを再構成し，それらを使用してグリッド間対を設けるのと等価になる．1グリッド中に含まれるスイッチ数が各段それぞれ2個より多く，1対の入線と出線間の経路を説定するのにそれらのスイッチの一部しか利用できないときは，タイプⅠの変形を施した回路網はチャネルグラフの一様性が失われ，均質なスイッチ回路網ではなくなる[5]．

7. むすび

本論文では，時分割スイッチ回路網における対関係について述べてきた．対関係の構成という観点から時分割スイッチ回路網を評価した結果をまとめると次のようになる．一般に一体網同時対が有利である．スイッチ回路網のグリッドが偶数個あるときは，T－S－T形回路網の場合も，S－T－S形回路網の場合も，グリッド間対を用いることによって望ましい対関係である準強対を一般に構成することができる．S－T－S形回路網の場合は，ほとんどの回路網に対して，グリッド内対であっても準強対とほぼ同程度に望ましい準弱対を構成することができる．T－S－T形回路網の場合は，リンク接続パターンを変更することによって，グリッド内対であっても準強対を構成することができるが，このようにして得られた回路網は均質なスイッチ回路網ではなくなる可能性がある．

スイッチ回路網の呼損率や制御量は，用いる対関係にも依存すると思われるが，これらの関係を明らかにすることが今後の課題である．

文　献

(1) 五嶋一彦：“時分割4線式交換回路網の構成についての考察”，研実報，15,7,pp.1171-1194（昭41-07）．

(2) Saito, T.：“An optimum design theory for PCM toll switching networks”, NTC'77, 46:6 (Dec.1977).

(3) Bieszczad, E. S., Krylow, K. E., Kowalik, R. F. and Macrander, M. S.：“No.3 EAX network and master clock”, ICC'77, 27, 2, pp.199-203 (June 1977).

(4) 俵，浜里，井上，高橋：“時間スイッチによる通話路構成”，研実報，28,7,pp.1277-1291（昭54-07）．

(5) Huttenhoff, J.H., Janik Jr., J., Johnson, G. D., Schleicher, W.R., Slana, M.F. and Tendick Jr., F.H.：“No.4 ESS peripheral system”, Bell Syst. Tech. J., 56, 7, pp. 1029-1055 (Sept.1977).

(6) 回陽，坂井，矢谷，吉田：“DDX-2通話路系装置”，研実報，25,12,pp.1853-1868（昭51-12）．

（昭和55年12月23日受付，56年3月6日再受付）

UDC [621.395.4 : 621.376.56] : 621.395.345 : [621.395.658 : 681.323]

T-S-T形スイッチ回路網における経路探索処理量の評価

正員 斉藤 忠夫† 正員 猪瀬 博† 正員 荻野 長生†

Evaluation on Dynamic Steps of Path Hunting in T-S-T Types of Switching Networks

Tadao SAITO†, Hiroshi INOSE† *and* Nagao OGINO†, *Regular Members*

あらまし 高速PCM方式を用いたPCM交換においては，スイッチ回路網の与えられた入出力点間を接続する経路の数が増加するため，高能率・低呼損率のスイッチ回路網を実現できる．しかし，接続を行うために探索される経路数も増加するので，経路探索のための中央制御装置の負荷は大きくなる．一般にスイッチ回路網を高能率とするにつれて，経路探索のための負荷は大きくなる傾向にある．従来スイッチ回路網の最適設計法には種々の提案があるが，スイッチ回路網自体の能率に重点が置かれ，その制御上の問題には考慮が払われていない．スイッチ回路網のタイプによって従来能率の点で優れていると考えられるものでも，探索処理量の点から好ましくない場合も生じうる．本論文はこうした目的から，スイッチ回路網における経路探索処理量を評価しようとするものである．対象としてはT−S−T形の時分割スイッチ回路網を取り上げている．まず，スイッチ回路網の制御方法と探索処理量の理論的ならびにシミュレーションによる評価方法を述べ，最後に典型的なスイッチ回路網について探索処理量の評価を行い，呼損率の小さなスイッチ回路網程，経路探索処理量が増加することを明らかにしている．

1. まえがき

交換システムの設計において，スイッチ回路網をいかに能率良く設計するかは基本的な問題である．スイッチ回路網の形態を仮定し，その呼損率とコストの観点から，各スイッチのサイズを最適なものとする最適化の研究は古くから行われている[1]．これに対してどのようなタイプのスイッチ回路網が能率が良いかを，チャネルグラフの形で比較する研究も行われている[2]．これによれば，スイッチ回路網の規模と段数を与えるだけで，最も能率の良いスイッチ回路網のチャネルグラフを求めることができる．この方法は，複雑な対関係が要求される時分割スイッチ回路網にも適用することができる[3]．

しかしながらこの方法で求められる能率の良いスイッチ回路網では，より複雑な経路探索が要求され，経路探索に必要なダイナミックステップ，すなわち探索処理量は，能率の良いタイプのスイッチ回路網ほど増加する傾向がある．

特に高速PCM方式による時分割交換方式のスイッチ回路網では，与えられた一対の入線と出線を結ぶのに利用できる時分割経路の数が増加するので，非常に小さな呼損率を持つスイッチ回路網の設計が可能となる．このため非常に高能率なスイッチ回路網が設計されている．高能率を保ちながら低い呼損率を与えるスイッチ回路網を実現するためには，交換機の中央制御装置はより強力な経路探索を行わなければならない．このことを考慮すれば，時分割PCM交換方式のスイッチ回路網では特にその設計に当って，その探索処理量の評価が重要となる．

本論文では，高速PCM交換方式における制御処理について，典型的なスイッチ回路網を取り上げて，経路探索に必要なダイナミックステップ数を，理論的解析とシミュレーションによって評価している．中央制御装置の命令体系としては，電子交換機で使用されている典型的な命令を仮定している．又，語長は32bitであると仮定している．

本論文では，まずT−S−T形スイッチ回路網の制御

†東京大学工学部電子工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan
論文番号：昭 55-181[B-56]

方法を述べる．次に経路探索処理に必要なダイナミックステップ数の理論的評価方法とシミュレーションによる評価方法を述べる．最後に典型的なスイッチ回路網についてダイナミックステップ数を評価し，経路探索処理量の観点からスイッチ回路網を比較検討する．

2. スイッチ回路網の制御方法

2.1 対制御方式

PCM時分割スイッチ回路網は本質的に4線式であるため，一つの接続を行うためには，発信側から着信側への経路と着信側から発信側への経路の2本の経路が必要となる．対関係とは，回路網内のリンク間に特定の関係を与えることによって，2本の経路を同時に設定する方法である．対関係を設けることによって，1回の経路探索処理で一対の接続経路を設定することが可能となる．

対関係はさまざまな方法で実現することができるが[3]，本論文で用いた対方式は，一体網同時対グリッド間対方式である．一体網においては，一対の接続は一つの回路網を使って実現される．同時対では，回路網内の同時に動作する素子の間で対が設けられる．すなわち，T-S-T形回路網では，同一のタイムスロットを使って一対の接続が設けられる．グリッド間対では，一対のグリッドを使用して一対の接続が実現される．

図1は，簡単なスイッチ回路網に一体網同時対グリッド間対方式を適用した例である．例えば，リンクAとリンクa，あるいはリンクBとリンクbの同一タイムスロットが互いに対となり，一対の接続経路を設けるために利用される．

2.2 経路探索方式

本論文では，中央制御装置への経路探索負荷の大きさを，回路網を通して一つの接続を設定するのに必要となる平均ダイナミックステップ数により評価している．

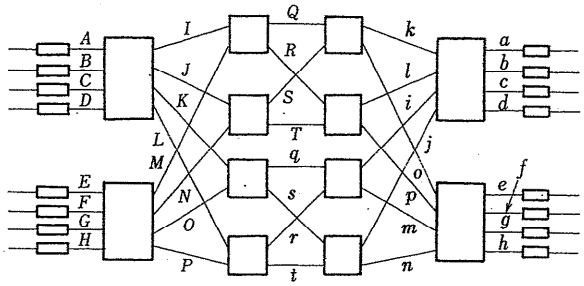


図1　一体網同時対グリッド間対方式の例
Fig.1－An example of combined concurrent intergrid pair scheme.

る．回路網の時分割多重度は128チャネルであると仮定している．本論文で述べる経路探索方式は，次のような条件を満足するものとする．

1. 空そく表においては，1本のリンク内の一つのタイムスロットの空そく状態を表すのに，1 bitのみ使用する．
2. 経路探索の際に利用できる作業用レジスタの数は無制限とする．すべての作業用レジスタは，経路探索処理の終了後，中央制御装置によってクリアされる．
3. あるタイムスロット入換装置を通して回路網に加えられた呼は，それと対になっているタイムスロット入換装置には接続されないものと仮定する．これは回路網が中継交換用であることを意味する．対関係において例外が認められるならば，回路網は加入者交換にも適用できる．
4. 探索されるリンクの選択方法として，本論文では定位式選択と不定位式選択について評価が行われている．経路探索に必要なダイナミックステップ数は，これら二つの選択方法の間で著しく異なってくる．

次章で述べる経路探索プログラムにおいては，実際には約40個のレジスタが，経路探索情報を一時的に蓄積するために使用される．現実の処理装置では，このような多数のレジスタは利用できないので，一時的データは主メモリに記憶されることになる．このためには，ストア命令とロード命令が使用される．これらの命令に要するプログラムステップ数は，利用可能なレジスタの数に依存する．これを避けて，経路探索プログラムの特徴についてのより一般的な結論を得るために仮定2を設けている．

制御装置の語長が32 bitであれば，整合操作において一度に32本の経路が探索できる．一度に探索される32本の経路が，互いに独立なものであれば，経路探索の効率は良くなるものと期待される．1本の空間的経路に含まれるタイムスロットの異なる経路は，互いに独立である．このため空そく表の1語中には，1本の時分割バスに含まれる32個のタイムスロットの空そく情報が記憶されているものとし，経路探索処理はこのような空そく表を使って行われるものとする．

時分割多重度が128チャネルであれば，1本の空間分割経路の探索は4(＝128／32)回の整合操作で完了する．一方空間的経路の探索は，互いに従属性の強い経路を選択して行えば，探索効率は低くなるかもしれ

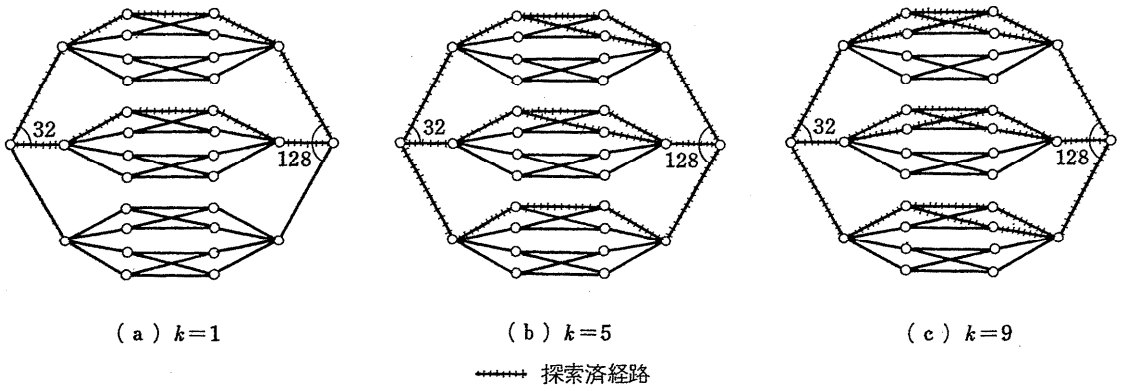


図2 経路探索方式
Fig.2—Path hunting strategy.

ないが，前の整合結果を後の整合操作の際に利用できるので，整合操作に必要な論理演算数を減少できる．本論文でも示すが，経路探索に要する平均ダイナミックステップ数は，このような探索方式を用いることによって小さくなる．つまりシミュレーションを行った結果によると，従属な経路から探索することによって論理演算数が減少するので，経路探索の成功確率が減少するにもかかわらず，全体のダイナミックステップ数は小さくなる．

図2は，図1に示したスイッチ回路網に上述の経路探索方式を適用する例を示す．整合操作の回数を k としたとき，それぞれ $k=1,5,9$ である場合の被探索経路が図2(a), (b), (c)に示されている．まず $k=1$ のときには，1本の空間的経路に含まれるタイムスロットの異なる32本の経路が探索される．$k=5$ のときには，1本の空間的経路に含まれる128本の経路と，その空間的経路と最も従属性の強い空間的経路に含まれる32本の経路が探索される．$k=9$ のときには，2本の空間的経路と，それらと最も従属性の強い空間的経路に含まれる32本の経路が探索される．

3. 経路探索処理量の評価方法

3.1 理論的方法

一対の入線と出線を接続する経路の数を n とし，k 本の経路を探索するのに必要なダイナミックステップ数を S_k とし，k 本の探索で探索を終了したときの呼損率を B_k で表せば，経路探索プログラムの平均ダイナミックステップ数ADSは次式で与えられる．

$$\mathrm{ADS}=\sum_{k=0}^{n-1} B_k(S_{k+1}-S_k),\quad B_0=1,\quad S_0=0 \qquad (1)$$

B_k はトラヒック理論から導かれた通常の呼損式によって得られる．S_k の値は次のような仮定を用いて概算することができる．

1. 経路探索処理の機能は，指定された入線と出線を結ぶ経路を探索し，空き経路が存在するか否か，存在するときは空き経路の番号を示すことである．
2. 本論文で仮定されている機械語命令は次のようなものである．ロード命令（メモリ→レジスタ，レジスタ→レジスタ），ストア命令，シフト命令，アンド命令（マスク命令），加算命令，比較命令，条件付ジャンプ命令，FRM命令[5]．
3. 経路探索プログラムは，スタティックステップ数は無視してダイナミックステップ数が最小となるように書かれている．

経路探索処理は二つの部分に分けられる．第1の部分は，必要な空そくデータのメモリ中におけるアドレスを指定する部分である．第2の部分は，それらの空そくデータを用いて整合操作を行う部分である．空間的パスの空そくデータのアドレスは，パラメータと呼ばれるそのパスのライン番号やそのパスが収容されているスイッチ番号によって指定される．経路探索に利用可能な空間的パスを指定するのに必要となるパラメータの総数を P で表し，第 i リンク段の空間的パスを指定するのに必要となるパラメータの数を P_i で表す．作業用レジスタの数は無制限なので，入線番号と出線番号から得られた P 個のパラメータの値は，レジスタに保存して置くことができる．第 i リンク段の空そくデータのアドレスは，レジスタから P_i 個のパラメータの値を読み出し，それらを結合することによって得られる．

ロード命令とマスク命令の2ステップが，パラメータの値を得るのに必要となる（図3(b)）．パラメータ

を結合するためには，最初はロード命令とシフト命令の2ステップ，2回目からは加算命令とシフト命令の2ステップが必要となる（図3(b)）．以上より空そくデータのアドレスを得るのに必要なステップ数S_{00}は次式で与えられる．

$$S_{00}=2P+\sum_{i=1}^{p}2P_i \quad (p=\text{リンク段数}) \tag{2}$$

不定位式選択の場合には，回路網における経路の拡大収束対数をP'で表せば，式(2)に$2P'$が加算される．これは一群の拡大された経路の中から1本の経路を指定する情報をロードし，探索処理終了後再びストアするためのステップ数である．従ってS'_{00}は次のようになる．

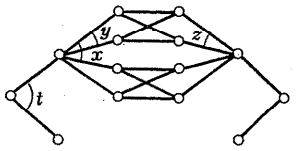
$$S'_{00}=2P+\sum_{i=1}^{p}2P_i+2P' \tag{3}$$

式(2)と式(3)はそれぞれ定位式と不定位式の場合の前処理に必要な最少のダイナミックステップ数を与える．この値は，以下の議論で$k=0$のときのダイナミックステップ数に相当する．

PやP'，P_iの値は，図3(a)に示されるように，スイッチ回路網の2項表現[4]から容易に知ることができる．又，PとP_iの値は，次式を用いてチャネルグラフからも得られる．

$$P=2(p-1)-(D_1+C_1) \tag{4}$$

$$P_i=P_{i-1}+\begin{cases}1: \text{条件1}\\ -1: \text{条件2}\end{cases};\ P_0=\frac{P}{2}+1 \tag{5}$$

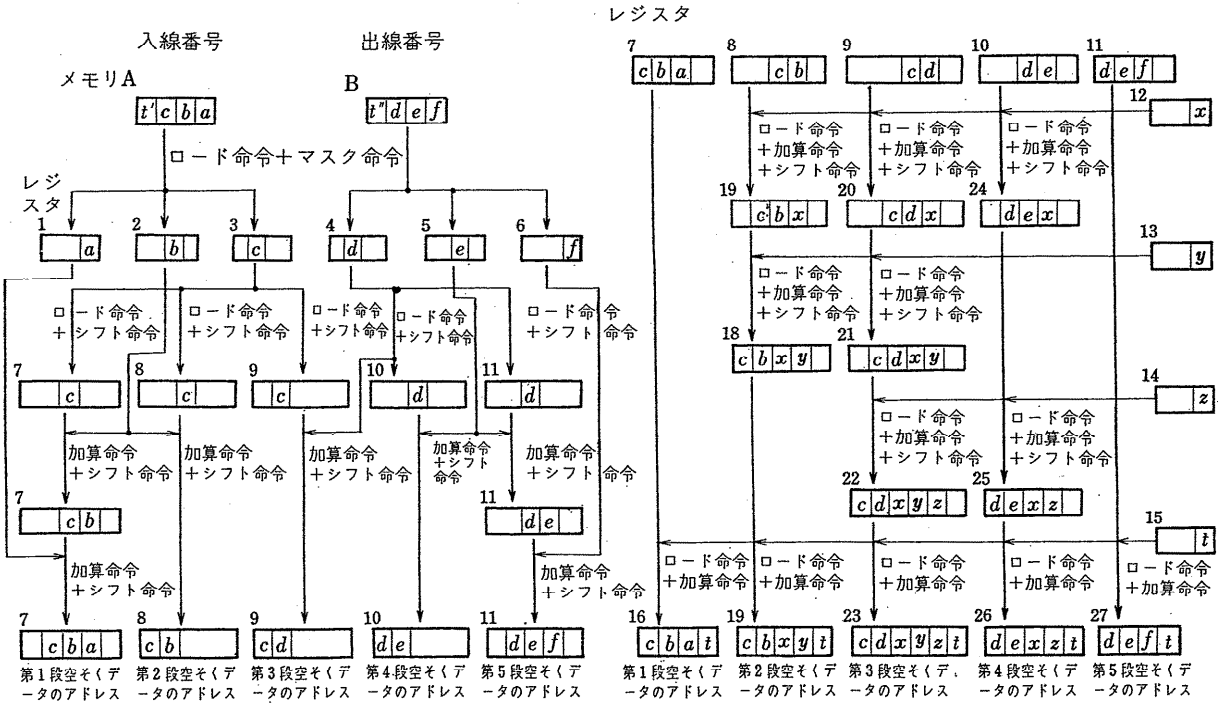


t', abc
a, bct
b, $cyxt$
cy, $dzxt$
zx, edt
t, fed
t'', fed

⇓

$P=6\,(a,b,c,d,e,f)$
$P'=4\,(t,x,y,z)$
$P_1=3\,(a,b,c);\ P'_1=1\,(t)$
$P_2=2\,(b,c);\ P'_2=3\,(x,y,t)$
$P_3=2\,(c,d);\ P'_3=4\,(x,y,z,t)$
$P_4=2\,(d,e);\ P'_4=3\,(x,z,t)$
$P_5=3\,(e,f,d);\ P'_5=1\,(t)$

（a）2項表現とパラメータ数の関係



（b）定位式選択　　（c）不定位式選択

図3　空そくデータのアドレスの導出
Fig.3—Address derivation for busy-idle data.

0：その他

D_1は拡大経路数がスイッチの出線数と一致するような拡大段と拡大収束段の総数である．C_1は収束経路数がスイッチの入線数と一致するような収束段と拡大収束段の総数である．条件1は第i段が，収束経路数がスイッチの入線数と一致するような収束段，又は収束経路数がスイッチの入線数と一致し，拡大経路数がスイッチの出線数と一致しない拡大収束段である場合に適用される．条件2は第i段が，拡大経路数がスイッチの出線数と一致するような拡大段，または拡大経路数がスイッチの出線数と一致し，収束経路数がスイッチの入線数と一致しない拡大収束段である場合に適用される．

各リンクの空そくデータが得られた後，整合操作が行われる．整合操作においては，特定のアドレスの空そくデータがロードされ，論理積を取られ，FRM命令が実行される．1本の空間的パスに含まれる1群の32本のリンクの整合を行うのに必要となるダイナミックステップ数は，次のように評価することができる．

1. 定位式選択の場合
   i) 空そくデータの整合を取るためのアンド命令1ステップ．
   ii) 整合の取られる空そくデータを保持して置きたい場合は，空そくデータを他のレジスタへ移すためのロード命令1ステップ．
   iii) 入線から出線までの整合が完了したとき，空き経路が存在するか否かを調べるFRM命令1ステップ．
   iv) 経路が拡大している場合は，これから整合を行う経路の番号を示す情報を残して置くためのロード命令1ステップ．
2. 不定位式選択の場合
   i), ii), iii) は定位式選択の場合と同様．
   iv) 経路が拡大している場合，整合を行うべき32本の経路を示す情報を得るための加算命令とマスク命令2ステップ．
   v) iv) で述べたように経路が拡大している場合は，整合を行うべき32本の経路を示すパラメータを求めなければならない．第iリンク段の経路を指定するのに必要なパラメータの総数をP'_iとする．空そくデータのアドレスを得るために，パラメータを結合しなければならないが，このためにロード命令と加算命令とシフト命令が必要となる．パラメータに対応している拡大収束間のリンク段数をdとすれば，最初のP'_i-1個のパラメータに対しては$3d$ステップが必要となる．最後のパラメータに対してはロード命令と加算命令のみで済むため，$2d$ステップが必要となる（図3(c)）．

上述の概算方法は，空そくデータのアドレスを求める手続きにおけるステップ数の評価に誤差を生じる．つまりアドレスは独立に求められた各パラメータの値を使って計算されると仮定しているので，S_{00}やS'_{00}の評価値は実際よりも大きく見積もられる．又，アドレスを求める手続きにおいて対制御が無視されているので，kに対するS_kの増加率は実際よりも小さく見積もられる．

後で述べるように，前者の誤差は定位式選択の場合に支配的であり，後者の誤差は不定位式選択の場合に支配的である．しかし，これらの誤差はすべての回路網において一様に加わるため，異なる回路網の性質を比較する目的に対しては無視することができる．

3.2 シミュレーションによる方法

スイッチ回路網の経路探索処理を，マルコフ鎖法を用いたモンテカルロ法によって模擬した．32本ずつ経路を探索するのに実際に必要となるダイナミックステップ数を，経路探索プログラムを作って評価した．この方法によって，32本の経路を探索するための各手続きに必要な実際のダイナミックステップ数を正確に知ることができる．

経路探索に要する平均ダイナミックステップ数は，各処理段階において必要となるステップ数の加算命令をシミュレーションシステム中の経路探索プログラムにそう入することによって測定される．経路探索プログラムは**3.1**で述べられた仮定を満足するように書かれている．回路網構成に対応して空そく表と経路表を指定し，回路網制御方式に対応して経路復旧手続きと経路探索手続きを指定するだけで，さまざまな回路網のシミュレーションが可能である．

4. 経路探索処理量の評価結果

4.1 対象としたスイッチ回路網

時分割多重度が128である3種類のT−S−T形スイッチ回路網について評価を行った．回路網のS段構成を，シンボル表示[3]を用いて図4に示している．三つのスイッチ回路網共，S段は4×4スイッチ4段で構成されている．入側と出側のPCMラインの数は，それぞれ32本である．

タイプⅠとⅡの結果を比較すれば，経路の拡大収束

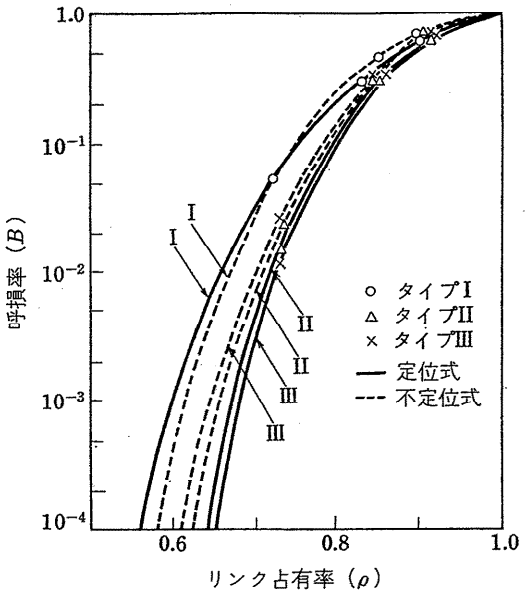


図6 リンク占有率と呼損率の関係
Fig.6—Blocking probability versus link occupancy.

なる．

図7にリンク占有率と経路探索に要する平均ダイナミックステップ数の関係を示す．図7における理論値とは，S_kとして図5における理論値を使用し，確率リニアグラフ[1]からB_kを求め，最後に両者から式(1)を使って求めた値である．入線数有限で，各リンク段における呼分布は2項分布に従うと予想されるので，確率リニアグラフを使ってB_kを求めた．シミュレーション値とは，S_kとして図5における実測値を使用し，**3.2**で述べたシミュレーションシステムを使って測定した値である．

図7(a)は定位式選択を行った場合のADSの値を比較している．スイッチ段数と経路多重数が等しい場合は，図6に示されているようにB_kの差が小さいので，S_kの差がそのままADSに反映している．タイプIとIIを比較すれば，初期値はIの方が大きく，傾きはIIの方が大きい．タイプIIとIIIを比較すれば，傾きはIIIの方が大きい．図7(b)より不定位式選択を行った場合も，傾向はほぼ同じである．

理論値とシミュレーション値を比較すると，定位式の場合はS_kの評価誤差を除いて，リンク占有率が小さいとき，理論値の方が小さくなっている．これは確率リニアグラフの方法が，各リンク段の呼の分布を一様であると仮定しているためである．リンク占有率が0.6～0.7を越えると呼分布は実際にも一様に近付く．この場合は確率リニアグラフの方法が，各段における

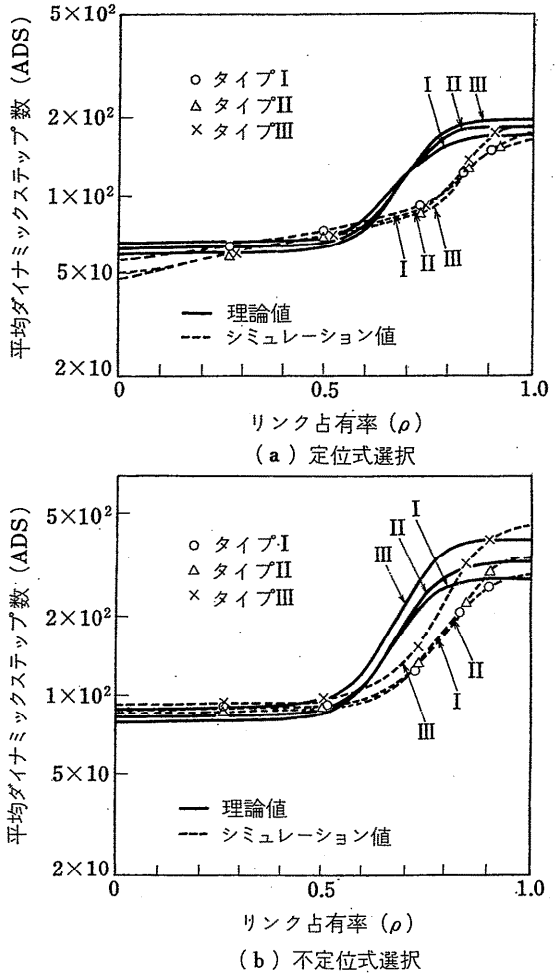


図7 リンク占有率と平均ダイナミックステップ数の関係
Fig.7—Average number of dynamic steps versus link occupancy.

呼の占有を独立と仮定し，又，時間ふくそう率を評価するので，理論値の方が大きくなる．不定位式の場合は呼分布はリンク占有率に関係なく一様であるため，理論値の方が大きくなる．

不定位式を定位式と比較すると，呼の分布が一様なので空き経路を発見するのに探索しなければならない経路数は減少する．しかし，S_kの値は増加するので，結局ADSの値もかなり増加してしまう．

4.4 スイッチ回路網形式と経路探索処理量の関係

経路多重数とスイッチ段数が一定のとき，回路網形式とADSの関係を要約すると次のようになる．

1. ある段において経路拡大数がスイッチの出線数と等しいとき，又は経路収束数がスイッチの入線数と等しいとき，ADSの初期値は減少する．
2. 経路拡大数がスイッチの出線数と一致しない拡

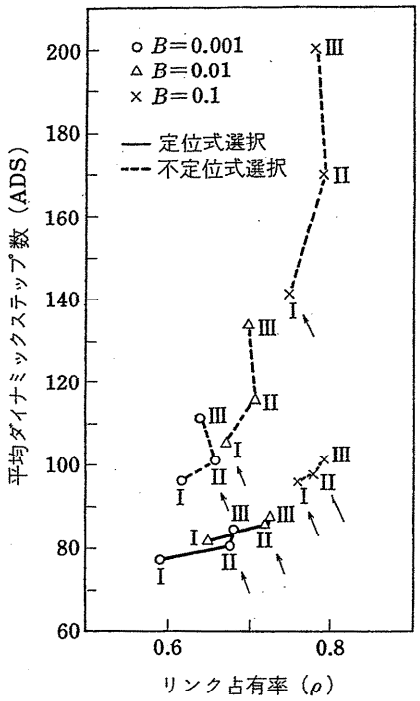


図8　呼損率一定のときのリンク占有率と平均ダイナミックステップ数の関係
Fig.8—Average number of dynamic steps versus link occupancy when the blocking probability is given.

大段はより後段に，経路収束数がスイッチの入線数と一致しない収束段はより前段に置いた方が，ADSの初期値は減少する（タイプⅠとⅡ，タイプⅠとⅢ）．

3. 拡大収束対数が減少すると，ADSの傾きは減少する（タイプⅠとⅢ，タイプⅡとⅢ）．
4. 拡大収束間の段数が減少すると，ADSの傾きは減少する（タイプⅠとⅡ）．
5. 拡大収束を交ささせると，ADSの傾きは増加する（タイプⅡとⅢ）．

上述の2，4，5は，呼損率の低いスイッチ回路網が大きなADSの値を持つことを示している．従って，中央制御装置を含めた全体的な経済性を満足する回路網を実現するためには，呼損率や回路網コストに加えて経路探索処理量を設計の際の評価基準として考慮すべきである．図8は呼損率を0.001，0.01，0.1に固定したときのリンク占有率とADSのシミュレーション値を示している．ADS/ρの値が最小である回路網を矢印で示している．リンク占有率すなわち運ばれた呼量だけから判断すると，タイプⅢが最適である．しかし経路探索処理量を考慮に入れると，タイプⅡの方が好ましくなる．

5. むすび

本論文では，T-S-T形スイッチ回路網における経路探索処理量の評価について述べた．T-S-T形回路網の制御方法，経路探索処理量の評価方法について述べた．経路探索処理量の評価結果から，呼損率の低いスイッチ回路網程，経路探索処理量の増加することが示された．これよりスイッチ回路網の最適設計の際の設計基準として，経路探索処理量を加える必要性のあることが明らかとなった．どのような形で経路探索処理量をスイッチ回路網設計に反映させて行くかが，今後の課題である．

謝辞　本研究を進めるに際し，御討論を頂いた猪瀬・斉藤研究室の諸氏に深謝します．

文　献

(1) Lee, C.Y.:"Analysis of switching networks", Bell Syst. Tech. J.,34,6,p.1287(Nov.1955).
(2) 高木謙三："任意格子形リンク方式の最適経路形式", 信学論(A),54-A,7,p.379(昭46-07).
(3) Saito, T.:"An optimum design theory for PCM toll switching networks", NTC '77, 46:6.
(4) 浅野，新開，斉藤，猪瀬："交換網記述の一手法", 信学会交換研資，SE72-22(1972-08).
(5) 日本電信電話公社編："D-10形自動交換機 第6部中央処理装置"，47，話技シリーズ，126(昭47-06).

（昭和54年11月12日受付）

96 F 8 ①,②

特　許　庁

特　許　公　報

特許出願公告
昭42-1895
公告　昭42. 1.28
（全12頁）

時分割交換方式

特　　願　　昭　39－74259
出 願 日　　昭　39．12．29
発 明 者　　出願人に同じ
出 願 人　　猪瀬博
　　東京都渋谷区青葉町13
同　　斎藤忠夫
　　東京都品川区大崎2の10
代 理 人　　弁理士　岡部正夫

図面の簡単な説明

第1図は中央局と該中央局に接続された複数個の遠隔集線装置よりなる交換局、該遠隔集線装置に接続された複数個の加入者及び他の交換局へのハイウエイを示すブロックダイヤグラム、第2図は第1図に示した遠隔集線装置のスイッチング及び制御部のブロックダイヤグラム、第3図は中央局における制御部と共に本発明による時分割接続回路網の一実施例を示すブロックダイヤグラム、第4A図は第3図に示した本発明による時分割接続回路網に使用するのに適したパルスシフタの構成部分を制御部と関連して示す結線図、第4B図は第4A図のパルスシフタに用いられる結合回路を記憶装置との関連に於て示した結線図、第5図は制御部の一部を含んだ本発明による時分割接続回路網の他の実施例を示すブロックダイヤグラム第6図と第7図は夫々第5図の回路網に有用な送りハイウエイにおかれたパルスシフタと受けハイウエイにおかれたパルスシフタの構成を制御部との関連に於て示した結線図、第8図は本発明による時分割接続回路網の他の実施例を示すブロックダイヤグラム、第9図は第8図に示した方式に使用するのに適したパルスシフタの構成を制御部と関連して示した結線図、第10図、第11図及び第12図は夫々第3図、第5図及び第8図に示した本発明による回路網の等価空間分割接続回路網である。

発明の詳細な説明

本発明は通信方式特に内部呼損のない時分割交換方式に関する。

現行の空間分割交換方式に代る時分割交換方式においては時分割多重化されたハイウエイを経て発呼回線と被呼回線間に情報を能率よく伝送できることは既に周知であり、従来よりこの種交換方式について幾多の提案がなされている。これらの提案の多くは発呼回線と被呼回線に一定周期で繰返される同一のタイムスロットを割当てることによつて情報の授受が行われるものである。しかし乍ら、ハイウエイスイッチング方式を用いた時分割接続回路網において時分割多重化された2個のハイウエイを接続する場合リング接続の場合の！ク呼損と類似のタイムスロットのミスマッチ（不整合）による損失が生ずる場合がある。

従つて、本発明の目的は改良された時分割交換方式を提供することにある。本発明の他の目的はタイムスロット入替機能を有するパルスシフタを交換網に導入し、空間分割接続回路網における非閉塞形の3段リンク回路網と等価な時分割交換回路網を構成することによつて内部呼損のない時分割交換方式を提供することにある。

本発明の他の目的は非閉塞形の時分割接続回路網を構成することによつて発信側加入者に割当てられたタイムスロットに関する情報を必要とせずに着信側加入者に独自にタイムスロットの割当を行い制御を簡単にすることにある。更に本発明の目的はハイウエイスイッチング方式を用いた時分割接続回路網の送りハイウエイと受けハイウ間のタイムスロットのミスマッチによる損失を除去する新規な時分割電話交換方式を提供することにある。

伝送と交換が一体化された時分割交換方式において、タイムスロットの入替えを行うためには何らかの遅延又は記憶機能が要求されるがこれを実現するためには波形歪みが問題とならない符号化変調方式を採用する必要がある。符号化変調方式を採用すると伝送路および交換機通話回路の歪み漏話、雑音の影響を受けにくゝすることができ、又この方式によると音声信号に復調することなくチヤンネルの入替えが出来るのでチヤンネルの割当てを簡単化することができる。

本発明の特徴によると、符号化変調方式を本発明に係る時分割接続回路網に有利に適用し得るも

のであつて上記の利点を悉く具備するものである

本発明の他の特徴によると、時分割接続回路網をして内部呼損のない空間分割接続回路網と等価な網を構成することができる。空間分割接続回路網を非閉塞形とした3段配置の回路網については既に公知であり、例えば「ベルシステム　テクニカル　ジヤーナル」第32巻第2号第406頁以下記載のチヤールスクロス著「非閉塞形スイツチ回路網の研究」と題する論文に発表されており、そこで著者は例えば入出線容量を等しくとつた対称構成のスイツチ回路網においては閉塞を避けるために2次即ち中間段スイツチの数は1次又は3次スイツチ当りの入り線容量をn個とした場合にn－1個以上としなければならないことについて述べている。以下上記のような3段配置における完全構成の回路網をクロス形の空間分割接続回路網として引合いに出すことが、本発明はこのクロス形の回路網と等価な時分割接続回路網を構成するものである。本発明の彼此の目的と特徴は図面と共に詳細に説明する以下の記載から容易に理解し得るものである。

先づ第1図を参照すると、交換局は遠隔集線装置101，102，103及び局間中継装置100を含み、これらの装置は夫々対応するハイウエイ104a，104b，104c及び104dを経て中央局105に接続されている。各ハイウエイは遠隔集線装置から中央局105に情報を伝送するためのS線中央局105から各遠隔集線装置を制御するためのR線及び遠隔集線装置を制御するためのC線とから成つている。遠隔集線装置1J1，102及び103には加入者線を通して複数個の加入者が接続され、集線装置101には加入者110及び112が接続されていることが例示されている。同様に、集線装置102及び102及び103には夫々複数個の加入者111及び113が接続されている。而して各集線装置間或いは1つの集線装置内の接続は時分割方式の適当なハイウエイ104a～104dを経て中央局105を介して完結される。中央局105の制御装置は1つの発信加入者の起呼を検出し、この加入者に対しタイムスロットを割当てる。発信加入者からのダイヤル情報はそのタイムスロットによつて中央局に到り、中央局は着信加入者を識別しこれにタイムスロットを割当てる。従来の方式においては発信加入者と着信加入者に共通なタイムスロットが空でなければ接続できなかつた。例えば集線装置101に接続された加入者110が集線装置102に接続された加入者111に対して呼を発するものと仮定すると、中央局の制御装置は加入者110のオフフツクに基づく接続要求を検出して加入者110に空いている1つのタイムスロットを割当てる。次に中央局105の制御装置は集線装置102のタイムスロットの状態を決定しもし同一タイムスロットが空であればそのタイムスロットで加入者110と111間に接続が完了する。しかし発信加入者110を収容する集線装置101のハイウエイと着信加入者を収容する集線装置102のハイウエイとの間に共通に空きであるタイムスロットが存在しない場合には同一タイムスロットで加入者110と111を接続することは不可能である。このタイムスロットの入替えができるならば同一タイムスロットで接続されるべき場合に予想されるタイムスロットのミスマツチによるブロツキングの問題は解消されるはづである。本発明においては発信加入者に割当てられるタイムスロット着信加入者に割当てられるタイムスロットは異つたものでよい。例えば加入者110と111に異つたタイムスロットが割当てられ、情報は加入者110から加入者線、集線装置101及びハイウエイ104aを通して中央局に送られ、該情報がハイウエイ104b、集線装置102加入者線を通して加入者に送られるように中央局105の制御装置で加入者111に割当てたタイムスロットに入替えられる。同じように、加入者111からの情報はその回線に割当てられたタイムスロットから加入者110に割当てられたタイムスロットに入替えられる。

第2図は第1図の集線装置の具体的な構成を示すものであるが、これは公知技術に属するものであつてその他いかなる形式の符号化変調方式を用いた時分割集線装置も本発明に使用し得ること勿論である。第2図に於て、集線装置に接続された複数個の加入者はS線とR線で中央局に相互接続されている。加入者110は常套的なライン回路207を経てラインゲート209に接続され、ラインゲート209はハイウエイ104aの送信ゲート214と受信ゲート215に接続されているオフフツク状態の加入者に接続されたラインゲートは一定周期で繰返されるタイムスロットで動作され、送信ゲート214と受信ゲート215は各タイムスロットの間連続的に動作される。これらのゲートの操作は遠隔集線装置制御200により

制御される。この遠隔集線装置はR線とC線を経て中央局の共通制御から指令信号を受信する。

加入者線は一定周期で繰返す監視用タイムスロットで周期的に監視されていて加入者からの接続要求があるとライン走査器204の走査によつて検出される。加入者のオフフツク状態はライン走査器204、アンドゲート231、オアゲート226、ハイウエイ104aのS線、可変遅延線315を経て走査制御314に送られる。走査制御314の機能に関連する諸装置の動作については第3図の説明にゆづることゝし、中央局の共通制御で発生された加入者番号がハイウエイ104aのC線を通して集線装置制御200に送られ、該制御200はこの符号化された番号を翻訳して割当てられたタイムスロツト中に発信加入者の関連している特定のラインゲートを動作する。ラインゲートの動作と共に送信ゲート214が動作する従つて発信加入者からの情報は符号化回路網216は情報を符号化した後オアゲート226を経て符号化した情報をハイウエイ104aのS線に送出する。而して送信ゲート214が動作を止めると同じタイムスロツトで受信ゲート215が動作し、そのためR線から受信された情報を復号化回路網217で復号して該情報を発信加入者線に送る。呼を発した加入者線とハイウエイ間の斯様な情報伝送の切換操作は集線装置制御200の制御のもとに発信加入者に割当てられた特定タイムスロツトで行われる。

第2図に関連して集線装置の各ゲートの制御について述べたが本方式の種々のゲートは正確な時間の信号パルスで制御され、そのため予め割当てられたタイムスロツトで発信加入者と着信加入者間の情報の伝送が出来るものである。フレーム周期は適当に選択されたものであるが1フレームは複数個のタイムスロツトから成り、フレーム内のタイミングはクロツクパルスによつて行われ、このクロツクパルスは第3図に示したクロツク源330によつて発生される。このクロツク源は必要により全ての集線装置と制御装置に使用される。

本発明においては互いに接続される2つの加入者が同一集線装置に属すると否とにかゝわらず全く同じ制御方式によつて接続が行われるものである。

第3図についてみると、ハイウエイ104aのS線、可変遅延線315を経て送られてきた情報によつて走査制御314は集線装置第2図に収容された加入者の接続要求を検出して主制御器300に伝える。加入者が走査の結果オフフツクであると判ると、ラインゲートナンバーシフトレジスタ312に遅延線路311に書き込まれた情報を読出しその加入者番号がこれまで遅延線路311に書込まれていなければ新しい呼びであるとみなされる。而して、この呼は主制御器300に伝えられる。これに対し主制御器300は複数個のレジスタ317のうちの空レジスタをさがすと共に接続段階記憶用の遅延線路318から空チヤンネルをさがしてその空チヤンネルを接続段階シフトレジスタ319読出す。この空チヤンネルに対応する加入者番号記憶用の遅延線路311に加入者番号を書込むと同時にこのとき補促されたレジスタ317と共働してトーン源320からトーンゲート321を経て加入者にダイヤル音を送る。

上記のような方法で発信加入者とレジスタが接続される加入者がダイヤルを開始するとそのダイヤル情報は割当てられた時分割チヤンネルにおいてハイウエイ140aのS線を通して送られ補促されたレジスタ317に記憶される。加入者がダイヤルを終るとレジスタ317はダイヤル情報の内容を主制御器に伝える。主制御器300は以下述べるように相手加入者の話中状態を調べる。

本方式のスイツチング回路網の特定の実施例は第3図に示す如く時分割多重度nの複数個の送り時分割ハイウエイを同数の受け時分割ハイウエイに接続する。こゝで図面を簡略化するため1つの系のみについて示し他は点線で示しこれを一線鎖線で囲んでいるが両者全く同一の構成であることは容易に理解されるであろう。各送りハイウエイには伝送遅れを補償するための可変遅延線路315が設けられ、該可変遅延線路315は遅延制御によつて常套的な方法で制御される。又各送りハウエイには可変遅延線路315と直列にパルスシフタ301が接続され、該パルスシフタは第4図に示すように構成されている。その出力は時分割多重度2nである。従つてこの種パルスシフタを以下n×2nパルスシフタと称することにする。パルスシフタ301からの出線は受け時分割ハイウエイに接続されたパルスシフタ304の入り線と格子状に配列され、その交さ点には時分割多重度2nのジヤンクダゲート307を有する。パルスシフタ304の入力は時分割多重度が2nであるからこの種のパルスシフタを以下2n×nパルスシフタと称する。パルスシフタ301，304

は夫々パルスシフタ制御302，305により制御され、これらのパルスシフタ制御に必要な情報は夫々フレーム周期に等しい遅延線路303，306によつて記憶されている。第3図にはこれらの遅延線路が夫々1つのブロックで示されているが遅延線路は採用された変調方式や遅延線路の形式によつて任意数のものが使用されるものであるハイウエイの交さ点に接続されたジヤンクタゲート307は遅延線路309とゲート制御308の制御のもとに動作する。

前に述べたようにレジスタ317はダイヤル情報を主制御器に伝え相手加入者の話中状態を調べるが、これは相手加入者を収容する集線装置に関連したラインゲートナンバー記憶用の遅延線路（図示せず）の内容を読み出すことによつて行われる。もしも相手加入者が空であれば相手加入者を収容する集線装置のハイウエイの任意のタイムスロットが相手加入者に割当てられる。そこでジヤンクタゲート307を制御する遅延線路309の対応するタイムスロットに相手加入者を収容するハイウエイの番号が書き込まれ、接続のために共通の空き時分割チヤンネルが用いられる。この共通の空き時分割チヤンネルが必ず存在することに関しては本方式のスイッチング回路網の等価回路を示す第10図のところで詳述するであろう。

着信接続も同様な手順で行われる。即ち相手加入者の集線装置に関連するハイウエイと発信加入者のハイウエイは共通に空いた時分割チヤンネルで接続され、相手加入者を収容する集線装置のハイウエイに設けられたゲート制御用の遅延線路（図示せず）の対応するタイムスロットに発信加入者を収容する集線装置の番号が書き込まれる。一方パルスシフタの1次側と2次側で割当てられたタイムスロットはそれぞれ主制御器300で比較されパルスシフタを制御するための情報がパルスシフタ制御用の遅延線路303，306に書き込まれる。この方法については後に述べるがこの様にして発信加入者と着信加入者間の接続が完成される。発信加入者と着信加入者間の通話が終ると、走査制御314はこれを検出し、主制御器300によつてその通話に関与した情報がラインゲートナンバー記憶用の遅延線路311、接続段階記憶用の遅延線路318、ゲート制御用の遅延線路309及びパルスシフタ制御用の遅延線路303，306から消される。

次に、第4A図には第3図に1実施例として示した方式に有効に用いられ得るn×2nパルスシフタの構成が制御装置と関連して示されている。パルスシフタは1つの入力と1つの出力を有し、複数個のゲートと記憶装置とから構成される。送りハイウエイにおかれたパルスシフタでは1フレーム内のタイムスロット数は入力側においてはn出力側においては2nである。又受けハイウエイにおかれたパルスシフタでは入力側においては2n、出力側においてはnである。

パルスシフタの記憶装置としてはシフトレジスタ又は磁歪遅延線路等が用いられる。例えば、シフトレジスタを用いた場合にはシフトレジスタの1段は1タイムスロットを構成するビット数に等しいフリツプフロツプ及びその周辺回路より構成されており、パルスシフタは斯様なシフトレジスタ2n段より成る。出力はこのシフトレジスタの1段ごとにとられる。磁歪遅延線路を用いた場合には各タイムスロットごとにおかれたタツプから出力がとり出される。図示実施例に於ては0乃至2n－1の2n段のシフトレジスタ403が用いられ、各シフトレジスタは相互に1タイムスロットの遅延をとるように配置される。第4A図に示したこのパルスシフタは第1と第2の入力ゲート401，402を有し、第1の入力ゲート401の出線はシフトレジスタの第1段目（0で示す）に接続され、第2の入力ゲート402は結合回路404を介してシフトレジスタの2段目（1で示す）に接続されている。第4A図における結合回路404の詳細を関連したシフトレジスタと共に第4B図に示す。この結合回路はオアゲート404aとインバータ404bから構成されている。今入力ゲート402から入力1が到来する場合は第1段目のシフトレジスタ403－0の出力如何にかゝわらず第2段目のシフトレジスタ403－1の初段のフリツプフロツプ403aをセツトする必要があるが、オアゲート404aはこのセツト入力を提供する。インバータ404bは第1段目のシフトレジスタ403－0の出力が0である場合もフリツプフロツプ403aがリセットされるのを防止する。尚第4B図のシフトレジスタの構成について附言すると、オアーゲート404aからの出力を受けるゲート403cとインバータ－404bの出力を受けるゲート403bがフリツプフロツプ403aの入力側にありこれらのゲートにクロツクパルスが加えられており、フリツプフロツプ403aの出力は抵抗R_1、コンデン

サC_1、抵抗R_2、コンデンサC_2によつて構成される積分回路を通つてそれぞれ次のフリツプフロツプを制御する入力ゲートに供給される。第4A図にかえつて、パルスシフタを制御するパルスシフタ制御記憶装置は入力ゲート用と出力ゲート用に分かれ入力ゲート用はシフトレジスタ411とダイオードマトリツクス414よりなるパルスシフタ制御と遅延線路412より構成され、後者即ち出力ゲート用はダイオードマトリツクス413とシフトレジスタ415よりなるパルスシフタ制御と遅延線路416より構成されている。入力ゲート用のものはf（f＝n×時分割システムの標本化周波数）のクロツク周波数で読出される（n×1）ビツトの記憶である。これはfのクロツク周波数の遅延線路を1個用いればよく、各チヤンネルに対応して読出され入力ゲート401と402を制御する。一方、出力ゲート用のものは本実施例では2fの周波数で動作することが必要であり、2n×10g_2（n＋1）ビツトの記憶を必要とする（この際10g_2（n＋1）が整数でない場合は端数が繰上げられる。）。第4A図ではこれが1個の遅延線路に記憶される場合が例示されている。遅延線路416に記憶された情報は10g_2（n＋1）ビツトの2fのクロツク周波数で動作するシフトレジスタ415に読出され翻訳回路を構成するダイオードマトリツクス413によつて翻訳されて各出力ゲート405，406及び407に分配される。第4A図に示すパルスシフタを制御するには夫々の情報を主制御器からこれらの遅延線路412と416に書込むことによつてなされる。

第4A図のパルスシフタのタイムスロツト入替え機能について述べると、例えば、入力の第h時分割チヤンネルを出力の第k時分割チヤンネルに移す場合もしkが偶数ならば第2の入力ゲート402が動作され、シフトレジスタ403ごとに1タイムスロツトずつ遅延を受けて第$\left\{\frac{k}{2}-h\right\}$（mod n）番の出力ゲートが開かれる。もしもkが奇数ならば第1の入力ゲート401と第$\left\{\frac{k-1}{2}-h\right\}$（mod n）番の出力ゲートが開かれる。この出力はオアゲート408によつて集成されてパルスシフタの出線に至る。この出力は入力周波数の2倍の周波数であり時分割多重度は2nである。第4A図に示したパルスシフタは送りハイウエイに接続されるものを示すものであるが、受けハイウエイに設けられるパルスシフタもこれと同様な構成にすることが出来、この場合タイムスロツトの入替えは次のようになる。即ち入力の第h′時分割チヤンネルを出力の第k′チヤンネルに移す場合もしもk′が偶数ならば第2の入力ゲートと第$\left\{h'-\frac{k'}{2}\right\}$（mod n）番の出力ゲートが開かれる。一方、k′が奇数ならば第1の入力ゲートと第$\left\{h'-\frac{k'-1}{2}\right\}$（mod n）番の出力ゲートが開かれる。パルスシフタを受けハイウエイに設ける場合には入力周波数2fがfの出力周波数となり、時分割多重度も2nからnとなることは容易に理解できるものと思われる。又入力制御用の装置は2fの周波数で出力制御用の装置はfの周波数で動作することに留意すべきである。

次に本発明に係るスイツチング回路網の他の実施例を第5図に示す。こゝでも第3図と同様に1つの系についてのみ実線で示すが、制御装置は第3図に示した実施例とは、同一のものが適用し得るので説明に必要な最少限の制御装置をブロツクダイヤグラムで示すに留め他の構成部分は第3図を参照され度い。たゞ主制御器も第3図のものと同様なものが使用され得るが主制御器の共通の空時分割チヤンネルをさがす機能は第5図の実施例では以下述べられるような理由で時分割多重度nの時分割ジヤンクタに対して実行される。

第5図の回路網において、送りハイウエイSに置かれたパルスシフタ501には第1の出力と第2の出力がある。他方、受けハイウエイに接続されたパルスシフタ504には第1と第2の入力があり、格子状に配列された送りハイウエイにおかれたパルスシフタ501の第1と第2の出線と受けハイウエイにおかれたパルスシフタ504の第1と第2の入線の交さ点は夫々ジヤンクタゲート507，508の群で接続される。第5図の各ジヤンクタゲートは常套的な表示法で示されているが実質は第3図のものと同じであることを了解され度い。この場合送りハイウエイにおかれたパルスシフタ第1の出力は各受けハイウエイにおかれたパルスシフタの第1の入力にジヤンクタゲート群によつて接続され、同様に各送りハイウエイにおかれたパルスシフタに第2の出力は各受けハイウエイにおかれたパルスシフタに第2の入力は接続される。これらの構成から容易に判るように第5図に示した方式のパルスシフタは第3図の方式に使用されるものとは異なり、従つてパルスシフタ制御用の遅延線路も第3図の実施例で使用され

たものとは異つた構成の遅延線路が使用される。又パルスシフタの出力ハイウエイは2本あるから第3図の如きゲート制御用の遅延線路511の外に第2のゲート制御514及び遅延線路513が設けられる必要がある。

第5図の実施例において、主制御器（図示せず）による制御は着信加入者に空タイムスロットを割当てる段階までは第3図の回路網について述べた手順と同じである。ところが第5図の方式では発信加入者のハイウエイにあるパルスシフタの出力側と着信加入者のハイウエイにあるパルスシフタの入力側で共通の空タイムスロットをさがす操作は夫々のパルスシフタを結ぶ2組の時分割ハイウエイについて行われる。着信加入者から発信加入者への接続についても同様な操作が行われて接続が完了する。その際各操作が使用されるタイムスロットにおいて受けハイウエイに設けられたゲート制御用の遅延線路506にはジヤンクタゲートで接続される送りハイウエイの番号が書き込まれパルスシフタの1次側と2次側のタイムスロットの番号の差の情報及び2本の送りハイウエイ若しくは受けハイウエイの内の1本を選択する情報がパルスシフタ制御用の遅延線路に書込まれる。次に記憶方式を説明するに当り、制御回路と関連して送りハイウエイに有利に用い得るパルスシフタの構成を第6図に示す。

第5図に示した時分割接続回路網の送りハイウエイに用いられるパルスシフタはn個の記憶装置601と2n個のゲート602，603から第6図に示す如く接続して構成される。この場合記憶装置601はいかなる形式のものでもよいが本実施例ではシフトレジスタが用いられている。このパルスシフタはハイウエイからのクロツク周波数と同じ周波数でタイミングがとられるように各シフトレジスタ601がクロツク源に接続されている。このパルスシフタの制御は第1出力と第2出力用に分配された遅延線路612，616とパルスシフタ制御502によてなされる。遅延線路612，616には$\{10g_2(n+1)\}$ビットを各n個のタイムスロットに記憶する$n\times\{10g_2(n+1)\}$ビットの記憶容量を有するものが用いられる。（この際$10g_2(n+1)$が整数でない場合は端数が繰上げられる）。情報は$\{10g_2(n+1)\}$本の遅延線路に並列に記憶されてもよいし、1つのタイムスロットに$\{10g_2(n+1)\}$ビット直列に記憶されてもよい

前記パルスシフタ制御502は遅延線路612，616に記憶された情報を読出すシフトレジスタ611，615とダイオードマトリツクス610 613とから成り、遅延線路612，616からシフトレジスタ611，615に読出された情報はダイオードマトリツクス610，613で翻訳され制御信号として各出力ゲート602，603に分配され出力ゲート602，603を制御するパルスシフタ601に記憶された通話パルスはパルスシフタ制御502からの制御信号によつて出力ゲート602，603から取出され、第1出力オアゲート604と第2出力オアゲート605により集成される。而して出力は指定された時分割チヤンネルにおいて指定された番号の出力ハイウエイに読み出される。

第7図は第5図の方式の受けハイウエイに有利に用いられるパルスシフタの構成を示すものである。このパルスシフタには2つの入力と1つの出力があり第1入力と第2入力に夫々接続された入力ゲート702と703を含み、該入力ゲートからの出力がオアゲート704で各タイムスロットごとに集成されてシフトレジスタ701に加えられるようになつている。又このパルスシフタの出力の時分割チヤンネルはたゞ1つであるからシフトレジスタ701に同時に存在する通話パルスは1つに限られる。各段シフトレジスタは第4B図の結合回路404と同様の結合回路705によつて接続されている。各シフトレジスタ701にはクロツク源（図示せず）からクロツクパルスが供給されタイミングがとられる。図示実施例のパルスシフタを制御する方法は第6図に示したパルスシフタの制御方法と同様に行われるが、この場合入力ゲート702，703で制御するのが好ましい。制御装置は第6図に示した装置と同様であつて$n\times\{10g_2(n+1)\}$ビットの記憶容量を持つた遅延線路712，716とパルスシフタ制御504から成り、パルスシフタ制御は情報を読出すシフトレジスタ711，715と制御信号を各入力ゲート702，703に分配するダイオードマトリツクス710，713から構成され、これから分配された制御信号のもとに出力はシフトレジスタ701から読み出され1つのハイウエイに至るものである。

第8図は本発明による交換方式の他の実施例を示すものであつて接続段階記憶用遅延線路、ラインゲートナンバー記憶用遅延線路、走査器、集線

装置制御器等各集線装置あたりに必要な装置及び走査番号発生器、トーン源クロック源等全体に共通な装置は第3図の実施例で示したものと同様であるから第8図からは省略されている。

第8図に示した時分割接続回路網はr個の時分割送りハイウエイをr個の時分割受けハイウエイに接続するもので両ハイウエイは（2r－1）本のジヤンクタで相互接続される。各ジヤンクタには1個あてパルスシフタ801，811が設けられ合計（2r－1）個のパルスシフタがある。これらのパルスシフタは入出力とも時分割多重度がnであるから以下n×nパルスシフタと称することにする。ジヤンクタと送りハイウエイの交さ点は夫々ジヤンクタゲート804で接続されておりこのジヤンクタゲート804を制御するための循環形記憶部があり、この記憶部はゲート制御807と遅延線路806から成つている。同様に、ジヤンクタと受けハイウエイの交さ点はジヤンクタゲート805で接続され、これらのゲートはゲート制御809と遅延線路808よりなる循環形記憶部によつて動作される。尚第8図中送りハイウエイに設けられた可変遅延820は遠隔集線装置から中央装置に信号を送るときに生ずる時間的遅延を補償する装置であつて遅延制御821によつて夫々制御される。

第8図の方式に於て主制御器（図示せず）による制御手順は着信加入者に空タイムスロツトを割当てるまでは第3図の方式と同じであるが、その論理機能は第3図について述べたものとは若干の相違がある。即ちこの実施例では主制御器によつて1次側と2次側で夫々発信加入者と着信加入者に割当てられたタイムスロツトの空いている1つのパルスシフタが（2r－1）個のパルスシフタ中より選び出される。斯様なパルスシフタが存在することは第12図と関連して後に説明することから容易に判るものと思われる。1つのパルスシフタが選択されると、送りハイウエイに設けられたゲート制御用の遅延線路806と受けハイウエイのゲート制御用の遅延線路808にその割当てられたタイムスロツトにおいて選出されたパルスシフタの番号が夫々書き込まれる。他方、パルスシフタ制御用遅延線路803又813には使用されるタイムスロツトの差の情報が書き込まれる。この様にして送りハイウエイと受けハイウエイの接続が完了する。

第9図は第8図の方式に適用するのに好ましいパルスシフタの構成を制御装置に関連して示したものである。このパルスシフタはn個のゲートとn個の記憶装置を含み第9図に示す如く接続される。記憶装置としてはいかなる形式のものでもよいが代表的なものとしてシフトレジスタ901が用いられている。入力側から入つた情報はシフトレジスタ901に記憶され、シフトレジスタからの出力は出力ゲート902に入る。一方、遅延線路803又は813に書き込まれた情報はシフトレジスタ912に読み出され、この読み出された情報がダイオードマトリツクス911で翻訳された制御信号として出力ゲート902に加えられている。この出力ゲートからの出力はオアゲート903で集成されてパルスシフタの出力側に取出される。

従つて、入力の任意のタイムスロツトを任意のタイムスロツト分例えば2タイムスロツト分遅延せしめるためにはシフトレジスタ901の数字2で示した段に対応するゲート群902のうちの1つのゲートをダイオードマトリツクス911及びシヤフトレジスタ912よりなるパルスシフタ制御802によつて開けばよい。同様の操作を行えば入力の任意のタイムスロツトを任意のタイムスロツト分だけ遅延することができるから出力の任意のタイムスロツトに入替えることができる。

通話が終ると走査器がそれを検出しその通話に関与した全ての遅延線路に書き込まれた情報が主制御器によつて消されることは第3図に関連して述べたことゝ同じである。

以上本発明による時分割接続回路網の構成と機能をその制御装置と関連して述べたが、各実施に示した回路網の等価回路は次のようになる。

先づ、第3図の時分割接続回路網に於て、送りハイウエイにおかれたパルスシフタ301の出力と受けハイウエイに設けられたパルスシフタ304の入力は夫々2nの時分割チヤンネルを有するが、夫々2nの時分割チヤンネルを有するr個のハイウエイをr個のハイウエイに接続するジヤンクタゲート307の群はr×rのクロスバースイツチが2n個並んだものと等価と考えられる。従つて第3図の等価空間分割接続回路網は第10図の如くなる。第10図に示された箱はクロスバースイツチの群を示す。第10図に示した空間分割接続回路網は所謂クロス型の非閉塞回路である。従つて、第3図に示した時分割接続回路網では送りハイウエイの任意の時分割チヤンネルに与えら

れた呼は受けハイウエイの任意の空時分割チヤンネルに接続することができる。

次に第5図に示した時分割接続回路網では前述の如く各送りハイウエイにおかれたパルスシフタ501の第1の出力は各受けハイウエイにおかれたパルスシフタ504の第1の入力にジヤンクタゲート507の群によつて接続され、各送りハイウエイに設けられたパルスシフタ501の第2の出力は各受けハイウエイにおかれたパルスシフタ504の第2の入力にジヤンクタゲート508の群によつて接続されている。従つて第5図の時分割接続回路網の等価空間分割接続回路網は第11図の如くなる。第11図に於て、奇数番の中間段スイツチ（一点鎖線内のブロック1－－－－2n－1）は第5図における送りハイウエイにおかれたパルスシフタ501の第1の出力と受けハイウエイにおかれたパルスシフタ504の第1の入力を接続するジヤンクタゲート507群に対応し、他方偶数番の中間段スイツチ（一点鎖線内のブロック2－－－2n）は送りハイウエイに設けられたパルスシフタ501の第2の出力と受けハイウエイに設けられたパルスシフタ504の第2の入力を接続するジヤンクタゲート508群に対応する。従つて、この等価回路から判るように第5図の時分割接続回路網もクロス形の非閉回路と等価である。

第8図に示した時分割接続回路網はr個の時分割送りハイウエイをr個の時分割受けハイウエイに（2r－1）個のn×nパルスシフタを設けジヤンクタで接続するものであるから、第8図の時分割接続回路網の等価空間分割接続回路網は第12図に示す如くなる。第12図から明らかなように、これもクロス型の非閉塞回路の完全群である従つて第8図の時分割接続回路網では呼に対して発信加入者と着信加入者に割当てられたタイムスロツトがそれぞれの1次側及び2次側で空いているパルスシフタが必ず存在し、呼の損失を生ずることは決してない。

以上述べた如く本発明はタイムスロツト入替機能を有するパルスシフタを用いて3段配置のリンク接続非閉塞接続回路と等価な回路網を構成できるので内部呼損のない時分割変換方式が提供されブロツキングの問題が解決されるものである。

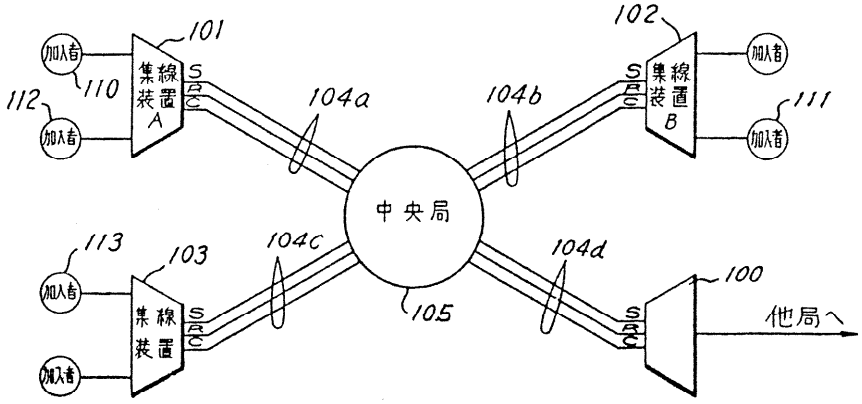
本発明の原理を理解するために図面に実施例を示してこれを説明したが本発明はこれらの実施例に限られるものではなく本発明の精神を逸脱しない限りいかなる変更や改良もなされ得ること勿論である。例えば、本実施例に使用したパルスシフタ制御用若しくはゲート制御用の遅延線路、ラインゲートナンバーシフトレジスタの内容の記憶及び接続段階符号内容の記憶に用いられる遅延線路はこれらと同一機能を有する記憶装置又は遅延装置によつて代替されることができる。

特許請求の範囲

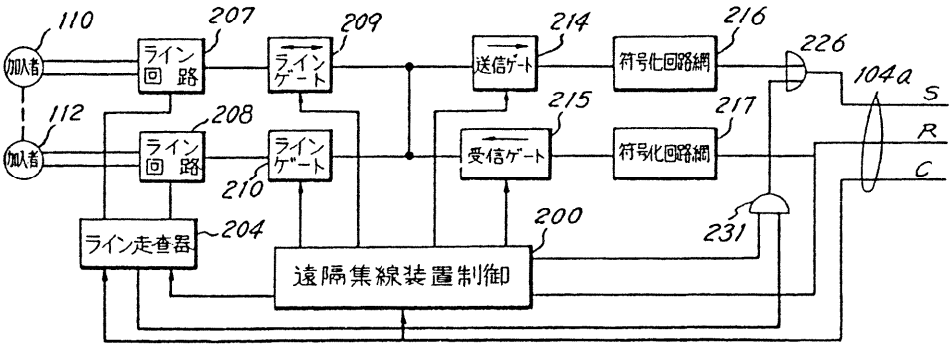
1　発信加入者の情報を送信するための送り時分割ハイウエイと、該発信加入者の情報を受信するための受け時分割ハイウエイと、情報を遅延又は記憶するための中央局における手段と、該手段によつて発信加入者と着信加入者に別個に割当てられたタイムスロツトをその位相のいかんにかゝわらず相互に入替える機能を有するパルスシフタを複数個含む方式において、該パルスシフタを両時分割ハイウエイに設けることによつて非閉塞形空間分割交換網と等価な時分割交換網を構成したことを特徴とする時分割交換方式。

2　発信加入者の情報を送信するための送り時分割ハイウエイと、該発信加入者の情報を受信するための受け時分割ハイウエイと、情報を遅延又は記憶するための中央局における手段と、該手段によつて発信加入者と着信加入者に別個に割当てられたタイムスロツトをその位相のいかんにかゝわらず相互に入替える機能を有するパルスシフタを複数個含む方式において、該方式は更に送り及び受けハイウエイを相互接続するための複数個のジヤンクタを含み、前記パルスシフタを該ジヤンクタに夫々設け該ジヤンクタの各々にその入力側の任意のタイムスロツトをその出力側の任意のタイムスロツトに入替える機能を付与して非閉塞形空間分割交換網と等価な時分割交換網を構成したことを特徴とする時分割交換方式。

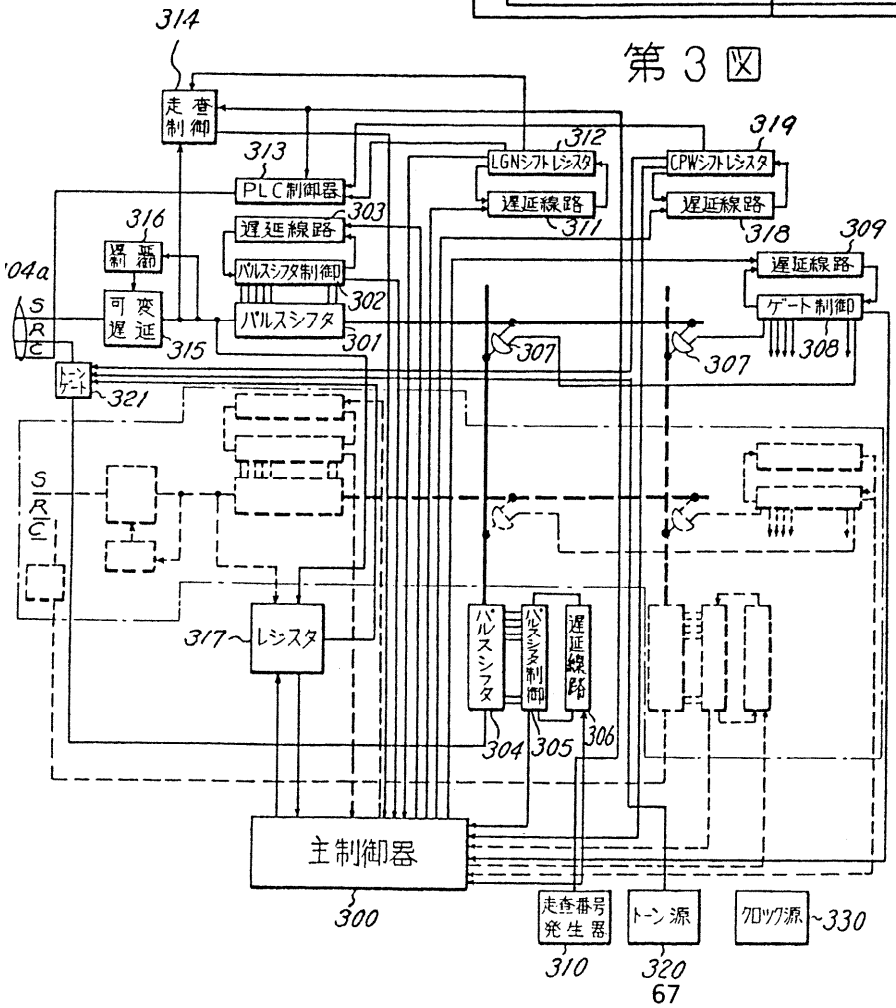
第1図


加入者
集線装置A
集線装置B
集線装置
中央局
他局へ
S
R
C

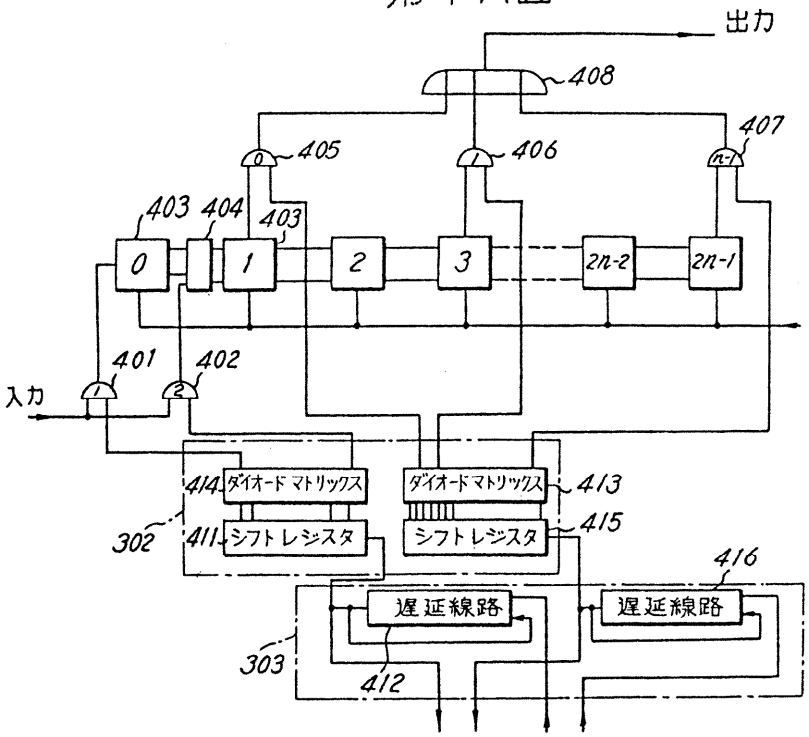
第2図


加入者
ライン回路
ラインゲート
送信ゲート
受信ゲート
符号化回路網
ライン走査器
遠隔集線装置制御
S
R
C

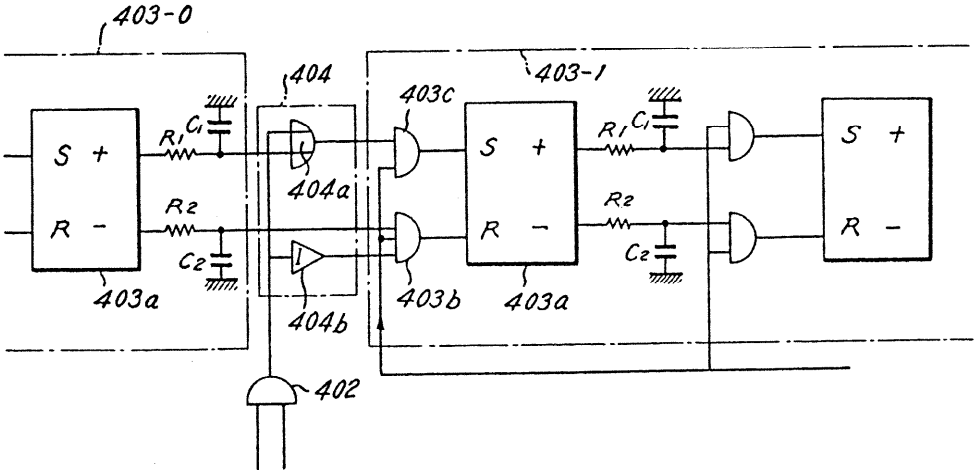
第3図


走査制御
PLC制御器
LGNシフトレジスタ
CPWシフトレジスタ
遅延線路
パルスシフタ制御
パルスシフタ
可変遅延
ゲート制御
レジスタ
主制御器
走査番号発生器
トーン源
クロック源
S
R
C

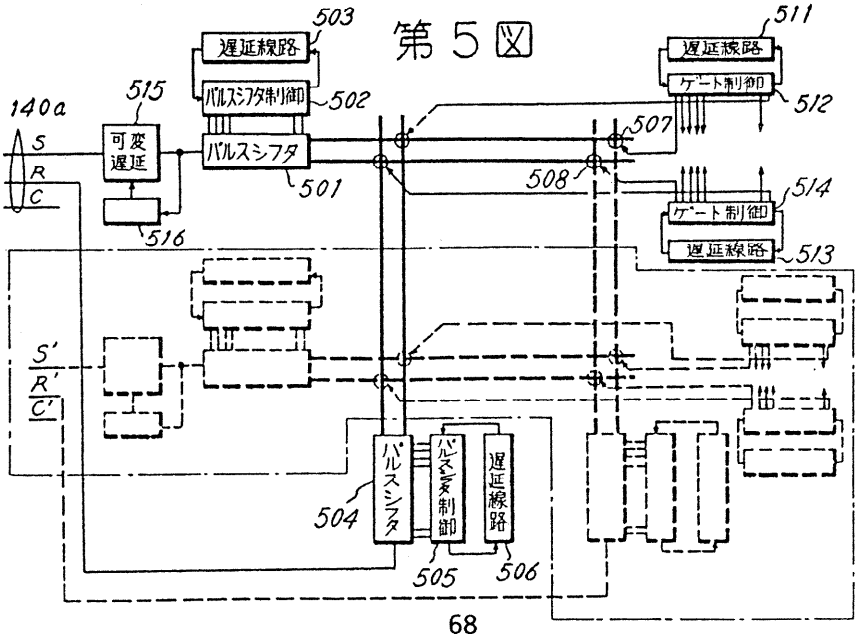
第4A図



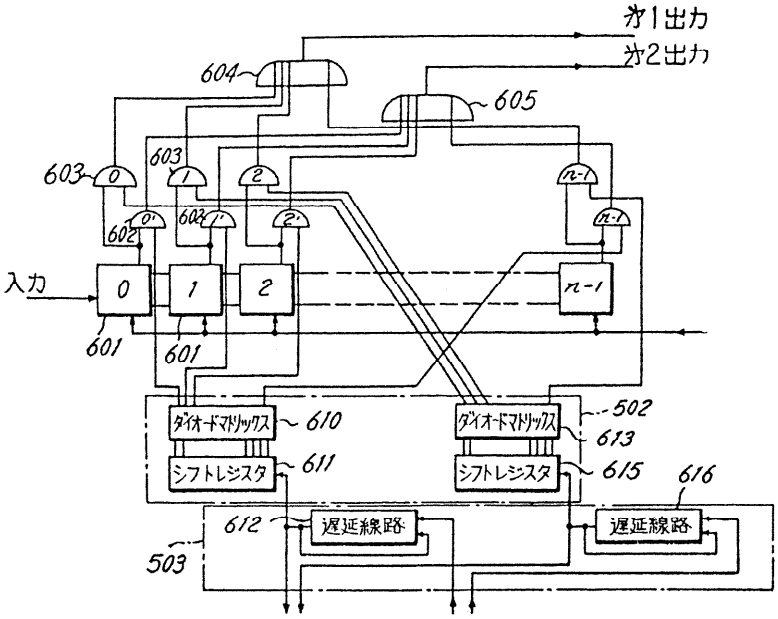
第4B図



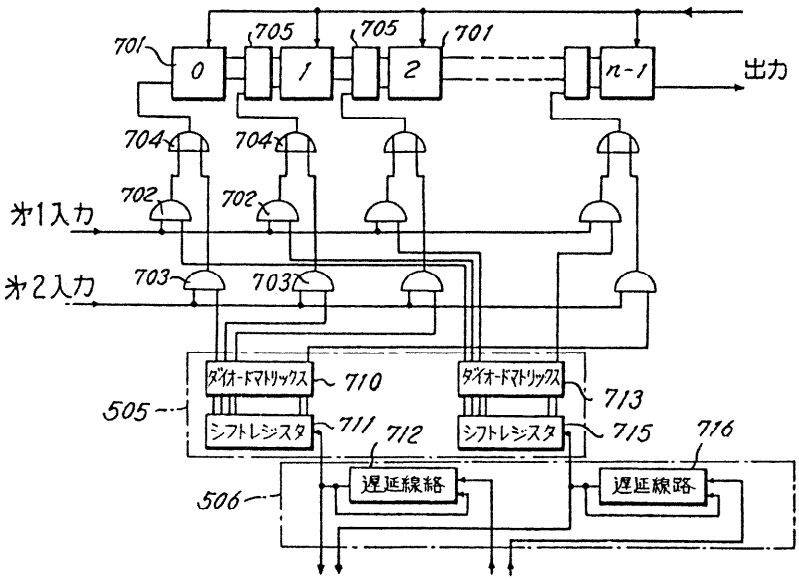
第5図



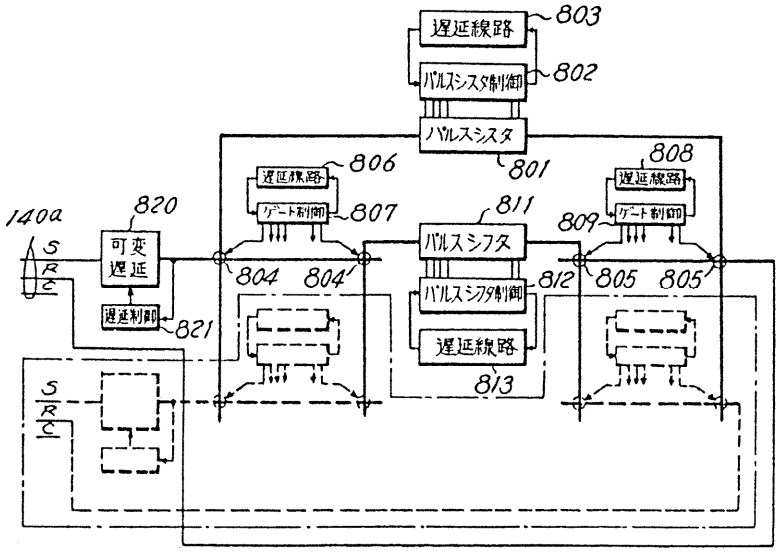
第6図


第1出力
第2出力
604
605
603
602
入力
0
1
2
n-1
601
ダイオードマトリックス
610
613
502
シフトレジスタ
611
615
616
612
遅延線路
503

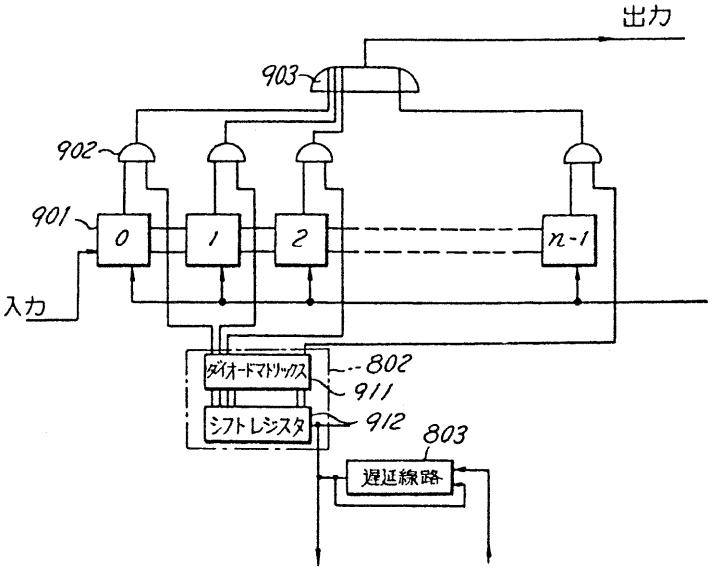
第7図


701
705
0
1
2
n-1
出力
704
702
第1入力
703
第2入力
ダイオードマトリックス
710
713
505
シフトレジスタ
711
712
715
716
遅延線路
506

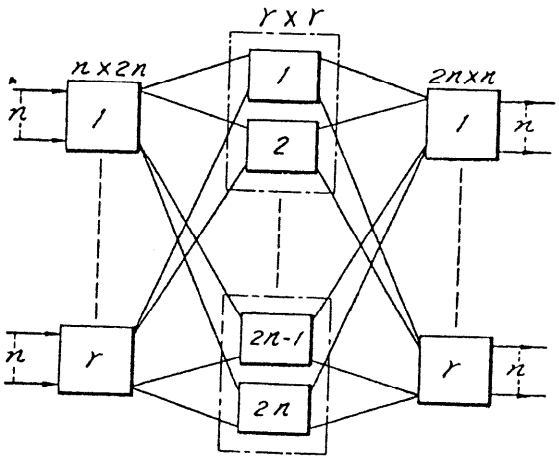
第8図


遅延線路
803
パルスシスタ制御
802
パルスシスタ
801
806
808
ゲート制御
807
811
809
140a
820
可変遅延
804
805
812
遅延制御
821
813

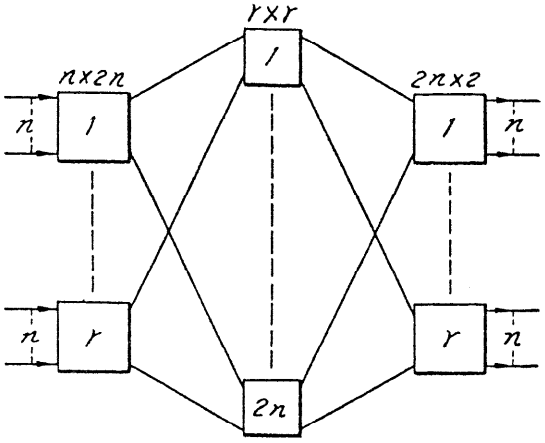
第9図



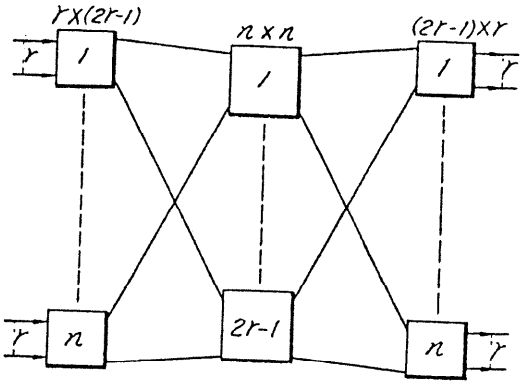
第11図



第10図



第12図



2.　通信網同期方式

○　相互同期系の理論
○　相互同期系の動作モード
○　相互同期系の一般化された構成概念

UDC 621.373.018.51

相互同期系の理論*

正員 猪瀬 博† 正員 藤崎博也† 正員 斎藤忠夫†

あらまし 多数の局より成る統合通信系統において，各局のクロック源として位相制御発振器を設け，相互に接続されたすべての局の位相制御発振器が互いに他を制御することにより同期して動作するよう構成した相互同期系は，発振器の故障または局間の接続断を生じても系の動作への影響は局限されるなどの利点をもつ従来にない新しい同期方式である．本論文はこのような相互同期系の一般的構成を明らかにするとともに周波数同期に関し理論的検討を行なったものであって，系統周波数の一般式を導きその具体例を示し，また系統内の障害あるいはパラメータの変化に対する系統周波数の変動を論じ，さらに位相の微少変動に対する系の動特性につき解析を行なって，二，三の実用的構成例に関し安定な系とするための設計法を明らかにした．

1. はじめに

大規模な統合通信系統 (integrated communication system)[1] においては，地域的に分散して配置された多数の局において周波数と位相の同期をとらなければならない．そのためには独立同期[2]，従属同期[1]および相互同期[3]の3つの方式が考えられる．独立同期方式においては，各局に独立に高安定クロック源を設け，他局からの入来信号を記憶装置に一時蓄積して自局のクロックにより読出し処理を行なうが，各局のクロック周波数間の相違のために情報の一部が失なわれるおそれがあり，これを防ぐためには装置が複雑化する．従属同期方式は系の中心となる1局に主クロック源を設け他をこれに直接・間接に従属させるものであって，主クロック源の故障は全系統の障害をひきおこすから，局の間に上下の階級を設け中心となっている上位の局のクロック源の障害の際は次位の局が順次中心となるよう配列する必要があり系の構成が著しく複雑となることは避けられない．

相互同期方式は系統内の各局のクロック源を，互いに他の局のクロック源により制御し相互に接続されたすべての局のクロック源が同期して動作するよう構成したもので，局の間に上下の階級を設ける必要がなく，クロック源の故障または局間の接続断を生じても系の動作への影響は局限され，また個々のクロック源発振器の安定度に対する要求も前述の二方式に比し緩和されるなど従来に見られない利点をもつ新しい方式である．本報告においてはこのような相互同期系の構成ならびに各局のクロック源として位相制御発振器 (phase locked oscillator) を用いた相互同期系の周波数同期に関し静的および動的特性を理論的に検討した結果の概要を述べてご参考に供したい．

2. 相互同期系の構成

2.1 相互同期の成立に必要な局間接続

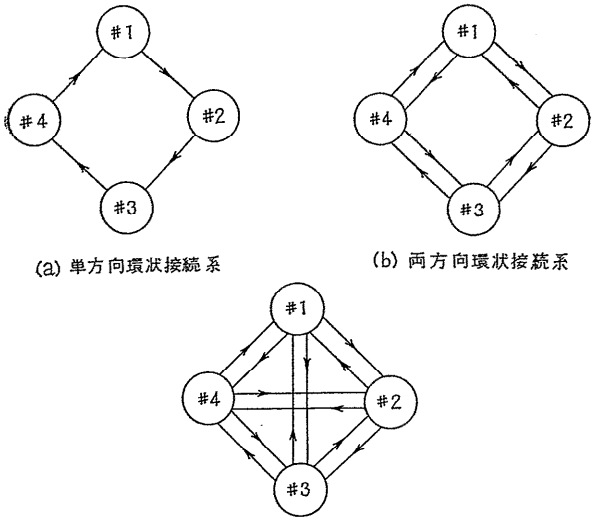
一般に通信網は相互に連繋された多数の局から構成され，局間の伝送路は両方向の通信に用いられるが，ここでは同期信号の授受のみに着目するものとし，局 i から発せられた同期信号が直接局 j へ伝達され，後述する方法によって局 j のクロック源を制御しうる場合にのみ，局 i から局 j への直接の接続が存在するものとみなす．さらに局 j から局 k へもこのような直接の接続があるとき，局 i は局 j を介して間接に局 k へ接続しているとみなす．このような系に属するすべての局の間に，序論に述べたごとき相互同期が成り立つためには，系統内のどの1局をとっても，直接または間接に他のすべての局への接続が存在することが必要である．

このような条件をみたす系統の典型的な例として，特に各局のクロック源が m 個の他局のクロック源から直接の接続によって制御をうけ，逆に同数の局のクロック源を直接に制御する系統を対称接続系と呼ぶこととする．この種の系には単方向環状接続系，両方向環状接続系，あるいは各局が他のすべての局と両方向の直接接続を有する完全接続系などがあり，それぞれ $m=1, 2, n-1$ に対応する．ただし n は系に含まれる局の総数である．図1には，$n=4$ の場合のこれらの例を示す．$n>4$ の場合には $n-1>m>2$ に対応する中間的な対称接続系も存在する．また，さらに一般には対称接続系から一部の接続を除いた多くの場合にも相互同期は可能である．

相互同期に必要な接続条件が満たされているか否か

* Theory on Mutually Synchronized Systems. By HIROSHI INOSE, HIROYA FUJISAKI and TADAO SAITO, Members (Faculty of Engineering, University of Tokyo, Tokyo). [論文番号 4026]

† 東京大学工学部



(a) 単方向環状接続系　(b) 両方向環状接続系　(c) 完全接続系

図 1　局数 4 の場合の対称接続系
Fig. 1—Symmetrically connected systems of four offices.

を視察により判定できるのは比較的簡単な系統の場合のみであり，複雑な系統でこれを代数的に判定するには，接続行列の性質[4]を利用することができる．

3.2　位相制御発振器を用いた相互同期系の構成

相互同期に必要な接続条件[4]を満足する系統内で，各局を相互に同期せしめるには，クロック源発振器として図 2 の構成による多入力の位相制御発振器を各局に配置すればよい．図 2 に示したのは一般に m 個の局

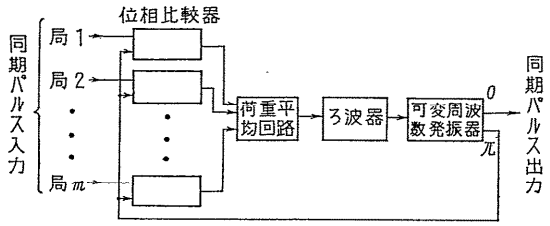


図 2　各局のクロック源となる多入力位相制御発振器
Fig. 2—Multiple-input phase locked oscillator as clock source of each office.

からの同期信号を入力とする場合であるが，それぞれの局からの同期信号と自局の発振器の出力との位相差を検出するための m 個の位相比較器，それらの出力に適当な荷重を付して加え合わせる荷重平均回路，位相制御ループの周波数特性を調整するフィルタ，およびその出力電圧によって制御される可変周波数発振器から成り，発振器の出力は各位相比較器に帰還される．特に位相比較器としてフリップ・フロップを用いる場合には，フリップ・フロップのセット入力に他局からの同期パルスを加え，カウンタ入力に自局の同期パルスを半周期移相して加えると，両局の位相差に対してフリップ・フロップの平均出力は周知のきょ歯状特性を示す[3]．もし何らかの障害によって他局からの同期パルスがなくなると，フリップ・フロップはカウンタ入力のみによって駆動され，その平均出力は 0 となり，障害の影響を受けずに動作を続ける．

3．相互同期系の系統周波数

相互同期系の特性を解析するに当たって，系統の諸要素の特性を定式化することが必要であるが，本節ではまず，それらの特性が時間的に変化せず，系が定常状態にあるものとし，この条件のもとで各局の発振器の間に位相の差はあるが全体として単一の周波数で動作し得るための条件を明確にし，かつその周波数（以下系統周波数と呼ぶ）を求め，その特性を検討する．

3.1　特性の定式化と系統周波数

まず，位相比較器が理想的なきょ歯状特性を有するものとすれば，局 j において局 i からの同期パルスを受ける位相比較器の出力の直流分は

$$e_{ij}=\lambda_j\{\theta_{ij}-k_{ij}\} \qquad (1)$$

ただし　$\theta_{ij}=f_j(\tau_{ij}+d_{ij})$,

$$k_{ij}=\left[\theta_{ij}+\frac{1}{2}\right],\quad [\ \ ]\ \text{はガウス記号}$$

と表わされる．ここに λ_j は位相差に対する感度，θ_{ij} は位相比較器の入力における局 i の同期パルスと自局（局 j）の同期パルスの位相差を 2π で割ったもの，τ_{ij} は両局の発振器の位相差を時間で表わしたもの，また d_{ij} は局 i から局 j に到る伝送遅延を表わす．θ_{ij}, τ_{ij} は自局が進み位相の場合に正とする．また f_j は自局の発振周波数であるが，定常状態にあっては系統周波数と一致する．

つぎにこれらの位相比較器の出力は荷重平均回路に加えられる．その出力は一般に次式で表わされる．

$$v_j=\sum_{i=1}^{n}\mu_{ij}e_{ij} \qquad (2)$$

ただし　$\mu_{ij}\geqq 0$,　$\sum_{i=1}^{n}\mu_{ij}=1$　とし，特に

$\mu_{jj}=0$　である．

ここで荷重係数　μ_{ij}　として零をとることも許容すれば，局 i から局 j への接続のない場合も含めることができる．さらに荷重平均回路の出力はフィルタに加えられるが，その直流分に対する伝達比は 1 に等しいとしても一般性を害わず，定常状態における系統周波数の決定においてはフィルタの影響を無視できる．

フィルタの出力はつぎに可変周波数発振器を制御するが，制御特性が直線的であるものとすればその周波数は

$$f_j=f_{0j}-\nu_j v_j \qquad (3)$$

と表わされる．ただし f_{0j} は中心周波数であり，制御入力が零の場合の発振周波数に等しい．

これらの特性を総合すれば，局 j の発振周波数に関して，

$$f_j=f_{0j}-\sum_{i=1}^{n}\alpha_{ij}\{f_j(\tau_{ij}+d_{ij})-k_{ij}\} \qquad (4)$$

が得られる．ただし $\alpha_{ij}=\lambda_j\mu_{ij}\nu_j$ は局 i からの同期信号に対する局 j での総合荷重係数であるが，系統の安定のためには，位相差の変化に対して負帰還を与えることが必要であり，一般に $\alpha_{ij}\geq 0$ となるよう λ_j, ν_j の符号を定める必要がある．また，局 i から局 j へ直接の接続がない場合，$\alpha_{ij}=0$ とし，特に $\alpha_{jj}\equiv 0$ である．

ここで全系統が共通の系統周波数 f_s で動作するためには，各局の f_j がすべて f_s に等しく，したがって

$$\left\{f_{0j}+\sum_{i=1}^{n}\alpha_{ij}k_{ij}\right\}f_s^{-1}-\sum_{i=1}^{n}\alpha_{ij}\tau_{ij}=1+\sum_{i=1}^{n}\alpha_{ij}d_{ij},\quad j=1,2,\cdots,n \qquad (5)$$

の n 個の方程式が満足されねばならない．未知数 τ_{ij} については $\tau_{ij}=\tau_{ik}+\tau_{kj}$, $\tau_{ji}=-\tau_{ij}$ なる関係があるため，その中で独立なものの数は $(n-1)$ 個であり，結局式 (5) は系統周波数 f_s を含めて n 個の未知数の連立方程式として一般に解くことができるが，系が相互同期系として成立するに必要な局間の接続条件[4]が満たされない系では方程式は縮退するか，または解が存在せず，それぞれ従属同期および同期不可能の場合に相当する．さらにまた，きょ歯状特性の位相比較器の出力には上・下限が存在するため，各局の発振周波数には，

$$f_{0j}+\frac{1}{2}\sum_{i=1}^{n}\alpha_{ij}>f_j>f_{0j}-\frac{1}{2}\sum_{i=1}^{n}\alpha_{ij} \qquad (6)$$

で与えられる同期可能の範囲がある．したがって相互同期が成立するためには，すべての f_j の可変範囲が重なり合っていること，換言すれば

$$\min_j\left\{f_{0j}+\frac{1}{2}\sum_{i=1}^{n}\alpha_{ij}\right\}>\max_j\left\{f_{0j}-\frac{1}{2}\sum_{i=1}^{n}\alpha_{ij}\right\} \qquad (7)$$

が成り立っていることが必要であり，この式の左辺・右辺をそれぞれ f_u, f_l と書けば，それらは f_s の上限および下限を表わす．

これらの条件のもとに，式 (5) の解はつねに存在して，系統周波数はつぎの一般式で与えられる．

$$f_s=\frac{\begin{vmatrix} f_{01}+\sum\alpha_{i1}k_{i1}, & \alpha_{21}, & \cdots, & \alpha_{n1} \\ f_{02}+\sum\alpha_{i2}k_{i2}, & -\sum\alpha_{i2}, & \cdots, & \alpha_{n2} \\ \vdots & \vdots & & \vdots \\ f_{0n}+\sum\alpha_{in}k_{in}, & \alpha_{2n}, & \cdots, & -\sum\alpha_{in} \end{vmatrix}}{\begin{vmatrix} 1+\sum\alpha_{i1}d_{i1}, & \alpha_{21}, & \cdots, & \alpha_{n1} \\ 1+\sum\alpha_{i2}d_{i2}, & -\sum\alpha_{i2}, & \cdots, & \alpha_{n2} \\ \vdots & & & \\ 1+\sum\alpha_{in}d_{in}, & \alpha_{2n}, & \cdots, & -\sum\alpha_{in} \end{vmatrix}} \qquad (8)$$

ただしここで Σ は i に関する総和を表わす．また k_{ij} は前記式 (1) におけると同様の意味を持つが，

$$f_u>f_s>f_l \qquad (9)$$

の制限のなかで可能な多くの組合せがある．

3.2 各種の系におけるパラメータと系統周波数

式 (5) は相互同期の可能なすべての系に関して成り立つ最も一般的な表現であるが，ここでは比較的簡単な二，三の例を通じて相互同期系の特性の把握に資する．

（i） 荷重係数が均等な対称接続系　この場合には局 j における総合荷重係数 α_{ij} をあらためて α_j/m と書くことができ，系統周波数は

$$f_s=\frac{\sum_{j=1}^{n}f_{0j}/\alpha_j+K/m}{\sum_{j=1}^{n}1/\alpha_j+D/m} \qquad (10)$$

で表わされる．特に $m=1$, 2, $n-1$ の場合はそれぞれ，単方向環状接続系，両方向環状接続系，完全接続系に対応する．また

$$D=\sum_{i,j}d_{ij},\quad K=\sum_{i,j}k_{ij} \qquad (11)$$

である．ここで $\sum_{i,j}$ は，直接接続の存在するすべての局の対に関する和を示し，したがって D は系統内の伝送遅延時間の総和を表わす．また K は一般に零を含む正・負の整数であるが，f_{0j}, α_j, D が与えられている場合には，式 (9) の制約から K の値の範囲は

$$f_uD+m\sum_{j=1}^{n}\frac{f_u-f_{0j}}{\alpha_j}>K>f_lD-m\sum_{j=1}^{n}\frac{f_{0j}-f_l}{\alpha_j} \qquad (12)$$

で与えられる．

式 (10) で表わされる f_s と D との関係は K をパラメータとする双曲線群の一部をなす．したがって K は系統周波数のモードを表わす指標とみなされ，同一のモードに着目すれば，式 (12) から逆にそのモードに対応する D の範囲が定まり，次式で与えられる．

$$\frac{1}{f_l}\left\{K+m\sum_{j=1}^{n}\frac{f_{0j}-f_l}{\alpha_j}\right\}>D>\frac{1}{f_u}\left\{K-m\sum_{j=1}^{n}\frac{f_u-f_{0j}}{\alpha_j}\right\} \qquad (13)$$

一定のモードに対応する D の範囲は，互いに隣接のモ

ードと部分的に重複しており，f_s は D に対して多値関数となる．D が準静的に変化する場合には，f_s は一定のモードに従って変化するが，個々の位相比較器の出力の k_{ij} が階段的に変化する際に隣接のモードへ転移し，同時に系統周波数は $\left(D+m\,\Sigma\frac{1}{\alpha_j}\right)^{-1}$ ずつ跳躍する．各局のパラメータがすべて等しく，d_{ij} もすべて相等しい系では，この跳躍は系統周波数が f_l または f_u に等しくなった場合にのみ起こるが，一般の系ではそれ以外の周波数においても起こる．現実には系統の位相同期の必要性から，局間の位相差が 2π のほぼ整数倍に等しくなるよう，何らかの手段によって補正を行なう上，各局のパラメータにもほとんど等しいものを用いることが望ましいので，モードの跳躍点の近傍を避けて系統の動作点を設定することが可能である．この場合の系統周波数 f_s と系統内総遅延時間 D との関係を図3に示す．ただしここでは簡単のため，さらに各局のパラメータが等しいものとし，$f_{0j}=f_0$, $\alpha_j=\alpha$ とおいてある．

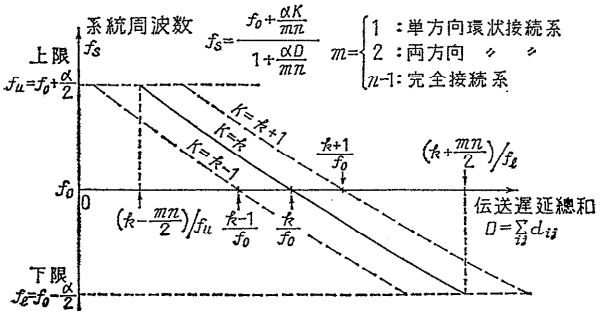


図 3　パラメータの等しい n 局から成る対称接続系の系統周波数

Fig. 3—System frequency f_s versus total transmission delay D in a symmetrically connected system of n offices.

(ii)　2 局間の接続が常に両方向で荷重が対等な系

相互に接続された1対の局 i，局 j において，常に $\alpha_{ij}=\alpha_{ji}$ の関係が成立すれば，系統周波数は一般に次式で表わされる．

$$f_s=\frac{\sum_{j=1}^{n} f_{0j}+\sum_{i,j}\alpha_{ij}k_{ij}}{n+\sum_{i,j}\alpha_{ij}d_{ij}} \tag{14}$$

ただし $\sum_{i,j}$ は直接接続の存在するすべての局の対に関する和を表わす．この場合にも，式(9)によって k_{ij} のとりうる整数の範囲が局限されることは(i)の場合と同様であるが，一般に α_{ij} の値が互いに異なる場合には，個々の k_{ij} に対応するモードが生ずるため，系統の特性はより複雑である．

3.3　系統の部分的障害の影響

相互同期系においては，一部の伝送路あるいは局設備に障害が生じた場合にも，障害部分を除く全系統の同期が維持される．一般の障害に対して相互同期が成立するか否かは，前述の方法[4]により判定が可能であり，またその条件が満たされていれば式(8)から，障害発生後，系が再び定常状態に達した場合の系統周波数を求めることができるが，ここでは前節に述べた対称接続系を例にとり，特に単方向伝送路および両方向伝送路の不通，および障害による系統内の1局の離脱の影響を考察する．

(i)　単方向伝送路が不通となる場合　　単方向環状接続系は，この種の障害によって単方向の鎖状接続系となり，鎖の一端の局を主局とする従属同期系に変化する．これに対して両方向環状接続系は障害発生後も相互同期系として動作し，その系統周波数は式(10)の $m=2$ の場合に対応して次式で与えられる．

$$f_s=\frac{\sum_{j=1}^{n}\left(\frac{n-j+1}{n}\right)\left(\frac{f_{0j}}{\alpha_j}+\frac{1}{2}k_j\right)}{\sum_{j=1}^{n}\left(\frac{n-j+1}{n}\right)\left(\frac{1}{\alpha_j}+\frac{1}{2}d_j\right)} \tag{15}$$

ただし　$j=1: k_j=k_{21},\qquad d_j=d_{21}$

$1<j<n: k_j=k_{j-1\,j}+k_{j+1\,j},\quad d_j=d_{j-1\,j}+d_{j+1\,j}$

$j=n: k_j=k_{n-1\,n}+k_{1n},\qquad d_j=d_{n-1\,n}+d_{1n}$

とする．ここで局の番号は環に沿う順序に従うものとし，また障害によって局 n から局1への接続が断たれるものとしたが，一般性を害わない．

完全接続系においても，障害によって相互同期が失なわれることなく，その系統周波数は

$$f_s=\frac{\sum_{j=1}^{n}\frac{f_{0j}}{\alpha_j}+\frac{1}{n-1}\left(\frac{f_{01}}{\alpha_1}-\frac{f_{0n}}{\alpha_n}\right)}{\sum_{j=1}^{n}\frac{1}{\alpha_j}+\frac{1}{n-1}\left(\frac{1}{\alpha_1}-\frac{1}{\alpha_n}\right)}*\frac{+\frac{1}{n-1}\left\{K^*+\frac{1}{n-1}(k_1-k_n)\right\}}{+\frac{1}{n-1}\left\{D^*+\frac{1}{n-1}(d_1-d_n)\right\}} \tag{16}$$

となる．ここで

$$K^*=\sum_{i,j}k_{ij},\qquad D^*=\sum_{i,j}d_{ij}$$

$$k_1=\sum_{i=2}^{n}k_{i1},\qquad d_1=\sum_{i=2}^{n}d_{i1}$$

$$k_n=\sum_{i=1}^{n-1}k_{in},\qquad d_n=\sum_{i=1}^{n-1}d_{in}$$

であり，$\sum_{i,j}$ は障害をうけた伝送路を除く和を示す．ただしこの場合にも局 n から局1への伝送路の不通を仮定している．

（ii） 両方向伝送路が不通となる場合　ここで対象となるのは $m \geq 2$ の系統であるが，この場合にも相互同期は害われず，系統周波数は依然式（10）で表わされる．ただし式中の $\sum_{i,j}$ は障害を受けた局間を除外するものとする．

（iii） 障害により1局が離脱する場合　単方向環状接続系は，この種の障害によって（i）の場合と同じく従属同期系と化する．これに対して，$m \geqq 2$ の系統では，障害発生後も相互同期は維持され，系統周波数は，離脱した局に関する項のみを除いて式（10）で表わされる．

これらの例から明らかな如く，相互同期系は，最も簡単な単方向環状接続系の場合でも上記の種々の部分的障害によって同期を害われることがなく，また，障害が系統全体に及ぼす影響は，式（15），式（16）に示されるように系統周波数の変化となって現われるが，系統内の局数が大きく，それらの間の接続が完全接続に近いほど，軽微なものとなる．

3.4 系統パラメータの統計的変動の影響

現実の系統では，各パラメータの統計的変動を避けることはできない．それらの速い変動に対しては次節に述べる系統の動特性が問題となるが，ここでは準静的な変動を対象とし，かつそれらが局相互間，あるいは伝送路相互間において無相関に生起すると考えられる一般的な場合について，荷重係数の等しい対称接続系を例にとり，系統周波数の変動を考察する．

（i） 各局の発振器の中心周波数のランダムな変動

この場合，f_{0j} および f_s の変動の標準偏差を σ_{f0j} および σ_{fs} とすると，式（10）から

$$\sigma_{fs}=\frac{\sqrt{\sum_{j=1}^{n}(\sigma_{f0j}/\alpha_j)^2}}{\sum_{j=1}^{n}\frac{1}{\alpha_j}+\frac{D}{m}} \tag{17}$$

が得られる．したがって系統周波数の安定度は一般に系統内の局数の平方根に比例して向上する．

（ii） 局間遅延時間のランダムな変動　d_{ij} がそれぞれ標準偏差 σ_{dij} をもって変動する場合，系統周波数の標準偏差は近似的に

$$\sigma_{fs}\cong f_s\frac{\frac{1}{m}\sqrt{\sum_{i,j}\sigma_{dij}^2}}{\sum_{j=1}^{n}\frac{1}{\alpha_j}+\frac{D}{m}} \tag{18}$$

で与えられ，一般に局数の大なるほど，また完全接続系に近いほど安定度が向上することを示している．

4. 相互同期系の動特性および安定性

本節では相互同期系において，ある局に入来する同期信号の位相変動またはその局の発振位相の変動によって，位相差に微少な変動を生じた場合の，系統の各局における発振位相の動特性を解析し，これらの局よりなる相互同期系を安定に構成するための設計条件を明らかにする．

4.1 位相の微少変動に対する局の動特性

ここでは動特性の解析は，系統が 3．に示された動作状態にあり，微少位相変動によって各位相比較器が直線的に動作する範囲で行なう．静特性を表わす式（4）において

$$f_j\tau_{ij}=\phi_j(t)-\phi_i(t) \tag{19}$$

であることを考慮し，さらに局 j の位相制御発振器のフィルタの効果を表わすため，そのインパルス応答を $h_j(t)$ で示せば，局 j の位相制御発振器の動特性は近似的に次式で与えられる．

$$f_j(t)+\sum_{i=1}^{n}\alpha_{ij}\,h_j(t)*\{\phi_j(t)-\phi_i(t)+d_{ij}f_j(t)\}$$
$$=f_{0j}+\sum_{i=1}^{n}\alpha_{ij}k_{ij} \tag{20}$$

ただし＊は，たたみ込み積分を示す．式（20）にラプラス変換を施せば，

$$\mathfrak{L}[f_j(t)]=F_j(s),\quad \mathfrak{L}[\phi_j(t)]=\Phi_j(s)$$
$$\mathfrak{L}[h_j(t)]=H_j(s)$$

とおいて，また $F_j(s)=s\Phi_j(s)$ であるから，

$$s\Phi_j(s)+\sum_{i=1}^{n}\alpha_{ij}H_j(s)\{\Phi_j(s)-\Phi_i(s)$$
$$+s\Phi_j(s)d_{ij}\}=X_j(s) \tag{21}$$

となる．ここで $X_j(s)$ は定常状態における固定項および外乱に関する項より成る．これを変形すれば，

$$\left[1+\frac{H_j(s)}{s}\sum_{i=1}^{n}\alpha_{ij}(1+sd_{ij})\right]\Phi_j(s)$$
$$-\frac{H_j(s)}{s}\sum_{i=1}^{n}\alpha_{ij}\Phi_i(s)=\frac{X_j(s)}{s} \tag{22}$$

を得る．

相互同期系では，位相の変動が直接各局の発振周波数に影響を与えることは望ましいことではない．したがって系統の周波数特性を決定する位相制御発振器のフィルタの伝達関数の時定数は，局間伝送路の遅延時間 d_{ij} より十分大きな値に選ぶ必要がある．したがって問題となる周波数領域において，式（22）の sd_{ij} は

$$sd_{ij}\ll 1 \tag{23}$$

と考えることができる．以上の考察において，d_{ij} は

一定であるとしたが，これも温度変化，経年変化等によって変動する．しかしこの変化の速さは十分遅いものであると考えられ，別途補償することができるから，d_{ij} の時間的変動の系の動特性への影響は無視してさしつかえない．

4.2 系の動特性

式 (23) の成立する範囲では，式 (22) は

$$\left(1+\frac{H_j(s)}{s}\sum_{i=1}^{n}\alpha_{ij}\right)\Phi_j(s)-\frac{H_j(s)}{s}\sum_{i=1}^{n}\alpha_{ij}\Phi_i(s)=X_j(s)/s \qquad (24)$$

となる．　　$j=1,2,\cdots,n$

ここで

$$\boldsymbol{\Phi}=\begin{bmatrix}\Phi_1(s)\\ \Phi_2(s)\\ \vdots\\ \Phi_n(s)\end{bmatrix}$$

$$\boldsymbol{B}=\begin{bmatrix}\Sigma\alpha_{i1}\frac{H_1(s)}{s}, & -\alpha_{21}\frac{H_1(s)}{s}, & \cdots & -\alpha_{n1}\frac{H_1(s)}{s}\\ -\alpha_{12}\frac{H_2(s)}{s}, & \Sigma\alpha_{i2}\frac{H_2(s)}{s}, & \cdots & -\alpha_{n2}\frac{H_2(s)}{s}\\ \cdot & \cdot & \cdots & \cdot\\ -\alpha_{1n}\frac{H_n(s)}{s}, & -\alpha_{2n}\frac{H_n(s)}{s}, & \cdots & \Sigma\alpha_{in}\frac{H_n(s)}{s}\end{bmatrix} \qquad (25)$$

（ただし，Σ は i に関する総和）

とすれば，系の固有動作を表わす特性方程式は，式 (24) の右辺をすべて 0 とおいたものとなり，

$$\boldsymbol{B}\boldsymbol{\Phi}+\boldsymbol{E}\boldsymbol{\Phi}=0 \qquad (26)$$

となる．ただし $\boldsymbol{E}$ は単位行列である．ここで方程式

$$|\boldsymbol{B}+\boldsymbol{E}|=0 \qquad (27)$$

を満足する s の根が，系の極となり，これが右半平面にあれば，系は不安定となる．

実用上各局の位相制御発振器のフィルタの伝達関数は等しいことが望ましいので，$H_j(s)=H(s)$ ($j=1, 2, \cdots, n$) とおくことができ，このような場合式 (27) は

$$y=s/H(s) \qquad (28)$$

とおいて

$$y^{-n}\begin{vmatrix}y+\Sigma\alpha_{j1}, & -\alpha_{21}, & \cdots & -\alpha_{n1}\\ -\alpha_{12}, & y+\Sigma\alpha_{j2}, & \cdots & -\alpha_{n2}\\ \cdot & \cdot & \cdots & \cdot\\ -\alpha_{1n}, & -\alpha_{2n}, & \cdots & y+\Sigma\alpha_{jn}\end{vmatrix}=0 \qquad (29)$$

となる．明らかに $H(s)/s=0$ は式 (27) の根ではない．したがって

$$\boldsymbol{A}=\begin{bmatrix}\alpha_{11}, & \alpha_{21}, & \cdots & \alpha_{n1}\\ \alpha_{12}, & \alpha_{22}, & \cdots & \alpha_{n2}\\ \cdot & \cdot & \cdots & \cdot\\ \alpha_{1n}, & \alpha_{2n}, & \cdots & \alpha_{nn}\end{bmatrix} \qquad (30)$$

ただし　$\alpha_{jj}=-\sum_{\substack{i=1\\ i\neq j}}^{n}\alpha_{ij}$　　(31)*

であるとおけば，

$$g(y)=|y\boldsymbol{E}-\boldsymbol{A}|=0 \qquad (32)$$

の y に関する根を求め，これを式 (28) に代入することによって系の特性周波数を得ることができる．式 (32) は行列 $\boldsymbol{A}$ の固有方程式であり，y はその固有値であるが，その実数部は，すべて負となる[6]．

固有方程式 (32) はよく知られているように

$$g(y)=y^n+a_1y^{n-1}+\cdots\cdots+a_{n-1}y+a_n \qquad (33)$$

$$a_k=\sum_{q_1<q_2\cdots<q_k}\begin{vmatrix}\alpha_{q_1q_1}, & \alpha_{q_2q_1} & \cdots & \alpha_{q_kq_1}\\ \alpha_{q_1q_2}, & \alpha_{q_2q_2} & \cdots & \alpha_{q_kq_2}\\ \cdot & \cdot & \cdots & \cdot\\ \alpha_{q_1q_k}, & \alpha_{q_2q_k} & & \alpha_{q_kq_k}\end{vmatrix}$$

のように展開できる．行列 $\boldsymbol{A}$ の固有値を一般的にもとめることは困難であるが，式 (30)，(31) から，$a_n=0$ であり，したがって固有方程式 $g(y)$ は，$y=0$ にすくなくとも 1 つの根を有し，これを y_0 とする．0 以外の固有値は高々 ($n-1$) 個であるから，それらを $y_1, y_2, \cdots y_{n-1}$ で示せば固有方程式は

$$g(y)=\left(\frac{s}{H(s)}-y_0\right)\left(\frac{s}{H(s)}-y_1\right)\left(\frac{s}{H(s)}-y_2\right)\cdots\left(\frac{s}{H(s)}-y_{n-1}\right) \qquad (34)$$

となる．

系が安定であるためには，式 (34) が s の右半平面に零点を持たないことが必要である．すなわち第 1 項については $Z=\frac{H(s)}{s}$ は s の右半平面で極をもってはならず，また残りの項については $\frac{1}{y_i}=z_i$ とおいて，$\left(\frac{H(s)}{s}-z_i\right)$ は s の右半平面で零点をもってはならない．

s の右半平面を図 4 (a) に示すように，半径 ε の半円で原点を避けるような閉ループにそって一周した場合の $\left(\frac{H(s)}{s}-z_i\right)$ の偏角の増分は，Cauchy の定理により，$2\pi\xi$ となる．ただし ξ は s の右半平面における $\left(\frac{H(s)}{s}-z_i\right)$ の零点の数である．位相制御発振器のフィルタとしては，低域通過特性が要求されるから，その伝達関数 $H(s)$ は一般に

$$\lim_{|s|\to\infty}\frac{H(s)}{s}=0 \qquad (35)$$

を満足する．図 4 (a) に示した閉ループにそって s を

* ここで行列 $\boldsymbol{A}$ の対角要素 α_{jj} は 0 ではなく式(31)で表わされすべて負の実数である．これ以外の行列 $\boldsymbol{A}$ の要素はすべて正の実数である．

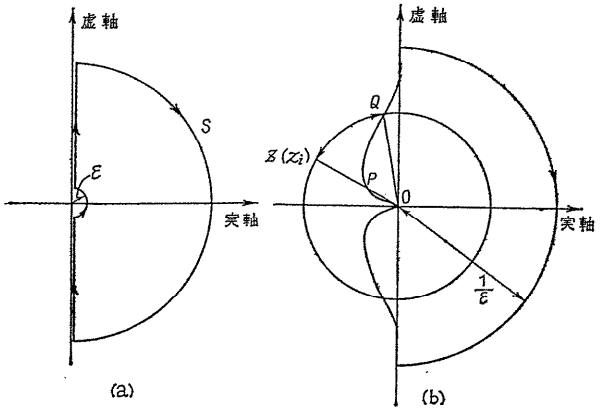


図4 系の安定判別法
(a) 極の存在してはならない s の右半平面
(b) z 平面における特性基準値 z_i と $H(s)/s$ のベクトル軌跡
Fig. 4—Stability criterion of system.
(a) Right half of s-plane where no pole should exist.
(b) z_i and Vecter locus of $H(s)/s$ in z-plane.

変化させた場合，$\frac{H(s)}{s}$ は図4(b)で示すような閉ループのベクトル軌跡をえがく．図4(b)の閉ループは半径 $1/\varepsilon$ の半円で，z の右半平面全体をおおう．

z_i を $H(s)/s$ に対する特性基準値と呼ぶことにすれば，ベクトル軌跡に囲まれる面の中に，その特性基準値があれば系は不安定となる．したがって，特性基準値が z の右半平面にあれば，系は必ず不安定となる．すでに示したように相互同期系では，行列 A の固有値 y_i の実数部は常に正ではないから，その逆数 z_i，すなわち特性基準値は常に z の左半平面にある．したがって伝達関数を適当に選定することによって，必ず安定な系を設計することが可能であり，この目的のためには Bode 線図による設計法を利用すればよい．

z 平面上で原点 O と，特性基準値に相当する点 Z をむすぶ直線がベクトル軌跡と交る点を P とすれば，PZ/PO が系のゲイン余裕に対応する．また原点を中心として，OZ を半径とする円がベクトル軌跡とまじわる点を Q とすれば，角 ZOQ が位相余裕に対応する．

以上の解析では，主として系の動作点の安定性を対象として微小じょう乱に対する動特性に着目したが，さらに大振幅のじょう乱に対する系の特性の解明には，位相比較器の非直線特性までも考慮した取扱いが必要であろう．

4.3 各種の系における動特性および安定性

(i) 2局間の接続が両方向で荷重も同等の場合

この場合 A は実対称行列であるから，これに対する固有値は常に実数で，さらに式(31)の関係より負となる．したがって $H(s)/s$ に対する特性基準値もすべて負の実数となる．

(ii) 荷重係数のすべて等しい完全接続系の場合

各局における伝達関数がすべて等しく $\alpha H(s)/s$ であれば，固有方程式は

$$\begin{vmatrix} \frac{y}{\alpha}+(n-1), & -1, & -1 & \cdots, & -1 \\ -1, & \frac{y}{\alpha}+(n-1), & -1 & \cdots, & -1 \\ \cdot, & \cdot, & \cdot & \cdots & \cdot \\ -1, & -1, & -1 & \cdots, & \frac{y}{\alpha}+(n-1) \end{vmatrix} = \left(\frac{y}{\alpha}+n\right)^{n-1}\frac{y}{\alpha}=0 \qquad (36)$$

となる．すなわち荷重係数のすべて等しい完全接続系の場合は特性方程式は

$y/\alpha=-n$ に $(n-1)$ 重根

$y/\alpha=0$ に1根

を有する．いいかえるならば，各局の正規化された一巡伝達関数 $H(s)/s$ に対する特性基準値として，$-1/\alpha n$ を用いて安定設計を行なえばよいことを示している．このことは，n 局からなる完全接続系において，$H(s)$ の等しい1個の位相制御発振器と同じ位相余裕を持たせるには，α を $1/n$ 倍する必要のあることを示している．図2に示した多入力位相制御発振器において，位相比較器の出力を平均化抵抗回路網で加算すれば，α は $1/(n-1)$ 倍となり，上記の要求は n が十分大きい場合近似的に満足される．

5. おわりに

以上相互同期系の構成およびその特性に関し若干の理論的検討を行なった．すなわちまず網の構成の見地から相互同期が可能な系を構成するための条件を考察し，具体的な例として単方向環状接続系，両方向環状接続系，完全接続系などを示した．つぎにクロック源として用いられる位相制御発振器の特性を示し，これらを用いて構成された相互同期系の系統周波数をあらわす一般式を導き，若干の実用的構成に関してその具体例を示すとともに局間の接続断，クロック源の障害，クロック周波数の変化，局間遅延時間の変化に対する系統周波数の変動を論じ，一般に多数の局から成る大規模な系ほど系統周波数の変動は局限されることを示した．さらに相互同期系の動特性について検討を行なって，位相差の微少変動に対する位相制御発振器の動特性をあらわす式を導き，これらの位相制御発振

器より成る系の動特性を解析することにより，二，三の実用的構成例に関し安定な系る得るための設計法を明らかにした．この設計法に従えば，自動制御理論において周知の手法を援用して系の安全余裕度を算定することができる．

以上本報告においては理論的な面から相互同期系の構成とその特性に関し論じたが，位相補償など相互同期系における局の回路構成上の諸問題については機会をあらためてご報告することとしたい．

謝辞　おわりにご討論いただいた本学阪本工学部長はじめ高周波談話会諸氏ならびにご援助いただいたベル電話研究所に対し深甚の謝意を表する次第である．

文　　献

(1) H.E. Vaughan : "Research model for time-separation integrated communication", Bell Syst. tech. J., **38**, p. 909 (1959-07).
(2) 尾佐竹徇，秋山 稔，高羽禎雄，西沢正雄："独立同期 PCM 中継方式"，信学誌，**47**, p. 210 (1964-02).
(3) 藤崎博也，斎藤忠夫，猪瀬 博："相互同期系の特性について"，昭 39 信学全大，S 7-6.
(4) 斎藤忠夫，藤崎博也，猪瀬博："統合通信系における相互同期について"，昭 40 信学全大，S 11-12.
(5) C.J. Byrne : "Properties and design of the phase controlled oscillator with a sawtooth comparator"; Bell Syst. tech. J., **41**, p. 559 (1962).
(6) 例えば浅野啓三，永尾汎："行列と行列式"，共立出版数学演習講座，p. 130 (1956).

(昭和 40 年 7 月 7 日受付，11 月 4 日再受付)

[UDC 621.391 : 621.376.56 : 621.3.018.51]

相互同期系の動作モード

斎藤　忠夫

あらまし　相互同期系は，時分割統合通信系の同期制御系として，信頼性および安定性の見地からすぐれた特性をもっているが，位相比較器の入出力特性が周期性を有するために多数の動作モードが発生し，系統周波数および局間位相関係の一意性が失われるおそれがある．本報告はこの動作モードについて一般的に解析を加えてその特性を定式化するとともに，局間伝送遅延の適切な調整および位相モード選択装置の使用により単一の動作モードが選択できることを示し，この動作モードは系統のじょう乱に対して安定に保存されることを示した．

1. まえがき

時分割統合通信系統においては，系統を構成する各局におけるクロック周波数の同期が必要であるばかりでなく，各局に到来する符号パルスのフレーム位相がその局のフレーム位相と同期していることが必要である．筆者らは先にクロック周波数の同期を実現するための相互同期方式を提案し[1]，さらに各局のフレーム位相を制御し，各局の入力におけるフレーム位相の同期を取る装置を有する相互同期装置を試作した結果を御報告した[2]．

この方式では系統の周波数および各局の位相は局間の伝送遅延時間 d_{ij}，各局の中心周波数 f_{0j} のみならず，各局の位相制御発振器に用いられている位相比較器ののこぎり波状特性の周期性を表わすパラメータ k_{ij} の影響を受けるため，系統が与えられても系統の周波数および各局の位相は必ずしも一義的には定まらず，種々の動作モードが存在する．相互同期系を実現するためにはこのような動作モードの性質を知り，適切なモードで動作する系統を構成することが必要である．本論文では相互同期系の動作モードの発生する条件につき一般的に論じ，各局の入力における位相差を小さくするための局間遅延に要求される条件を示した．局間遅延はそれぞれの動作モードに応じて調整されるが，局間遅延を調整する際に定められた動作モードは十分安定に保持される．さらに具体的な例として n 局の荷重が均等な完全接続系について定量的に論じている．

斎藤忠夫：正員　東京大学工学部電子工学科

Operation Modes in Mutually Synchronized Systems. By TADAO SAITO, Member (Faculty of Engineering, University of Tokyo, Tokyo).
論文番号：昭 43-61 [A-21]

2. 相互同期系における発振位相

のこぎり波状特性の位相比較器を用いた相互同期系では，局 j および局 i の 2π で正規化された発振位相を ϕ_j および ϕ_i，系統周波数を f_s，局 i の局 j に対する制御の総合荷重係数を α_{ij}，局 i から局 j に至る伝送路の遅延時間を d_{ij} で表わすと，n 局より成る系ではすでに御報告したように[3]局 j に関して式（1）で表わされる関係が成立する．

$$f_s=f_{0j}-\sum_{\substack{i=1\\ i\neq j}}^{n}\alpha_{ij}(\phi_j-\phi_i+f_s d_{ij}-k_{ij}) \qquad (1\mathrm{a})$$

$$k_{ij}=\left[\phi_j-\phi_i+f_s d_{ij}+\frac{1}{2}\right] \qquad (1\mathrm{b})$$

[ ] はガウス記号

$(j=1,2,\cdots,n)$

式（1a）は各局の発振位相の基準として任意の一局たとえば局1を選びその位相 ϕ_1 を $\phi_1=0$ とおくことによって解くことができる．すなわちこの連立方程式は，$\alpha_{jj}=-\sum_{\substack{i=1\\ i\neq j}}^{n}\alpha_{ij}$ として

$$\begin{bmatrix} f_{01}-f_s\sum_i\alpha_{i1}d_{i1}+\sum_i\alpha_{i1}k_{i1}\\ f_{02}-f_s\sum_i\alpha_{i2}d_{i2}+\sum_i\alpha_{i2}k_{i2}\\ \vdots\\ f_{0n}-f_s\sum_i\alpha_{in}d_{in}+\sum_i\alpha_{in}k_{in}\end{bmatrix}=\begin{bmatrix}1,-\alpha_{21},\cdots,-\alpha_{n1}\\ 1,-\alpha_{22},\cdots,-\alpha_{n2}\\ \vdots\\ 1,-\alpha_{2n},\cdots,-\alpha_{nn}\end{bmatrix}\begin{bmatrix}f_s\\ \phi_2\\ \vdots\\ \phi_n\end{bmatrix} \qquad (2)$$

で表わされる．ここで右辺の係数となっている行列を $\boldsymbol{A}'$ とし，$\boldsymbol{A}'$ の逆行列を $\boldsymbol{P}'$ とすれば，式（2）の解は $\boldsymbol{P}'$ の要素を p_{ij}' で表わして

$$\begin{bmatrix}f_s\\ \phi_2\\ \vdots\\ \phi_n\end{bmatrix}=\begin{bmatrix}p_{11}',p_{21}',\cdots,p_{n1}'\\ p_{12}',p_{22}',\cdots,p_{n2}'\\ \vdots\\ p_{1n}',p_{2n}',\cdots,p_{nn}'\end{bmatrix}$$

$$\cdot\begin{bmatrix} f_{01}-f_s\sum_i \alpha_{i1}d_{i1}+\sum_i \alpha_{i1}k_{i1} \\ f_{02}-f_s\sum_i \alpha_{i2}d_{i2}+\sum_i \alpha_{i2}k_{i2} \\ \vdots \\ f_{0n}-f_s\sum_i \alpha_{in}d_{in}+\sum_i \alpha_{in}k_{in} \end{bmatrix} \quad (3)$$

となる．式（3）は式（2）を形式的に解くことによって得られたものであるが，式中の k_{ij} は式(1b)によって与えられるものである．式(1)は f_s および ϕ_j の項を含んでいるから，式(3)が連立方程式(2)の解であるためには，式(3)に適当な k_{ij} のグループを与え f_s および ϕ_j を求め，これを式(1b)に代入してこれによりさきに与えた k_{ij} のグループが得られるかどうかの吟味がなされなければならない．後述のように k_{ij} のグループは必ずしも一義的に定まらないが，このことによって系統の動作モードの問題が生ずる．

各位相比較器の入力における二つの入力の見掛けの時間差を T_{ij} とおけば，

$$f_sT_{ij}=\phi_j-\phi_i+f_sd_{ij}-k_{ij} \quad (4)$$

である．式(1b)の k_{ij} の定義は

$$\frac{1}{2}>f_sT_{ij}>-\frac{1}{2} \quad (5)$$

であるよう k_{ij} が定まると言い換えてもよい．

$\phi_1=0$ と定義したことに対応して，$\boldsymbol{P}'$ の第1行をすべて0とおいた行列を $\boldsymbol{P}$ としその要素を p_{ij} で示そう．行列 $\boldsymbol{P}$ を用いると，局 $\#j$ の位相は

$$\phi_j=\sum_{q=1}^{n}p_{qj}\left(f_{0q}-f_s\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}d_{rq}+\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}k_{rq}\right) \quad (6)$$

で与えられるから，これを式（4）に代入して

$$\begin{aligned} f_sT_{ij}=&\sum_{q=1}^{n}(p_{qj}-p_{qi})f_{0q} \\ &-\sum_{q=1}^{n}(p_{qj}-p_{qi})f_s\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}d_{rq} \\ &+\sum_{q=1}^{n}(p_{qj}-p_{qi})\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}k_{rq}+f_sd_{ij}-k_{ij} \end{aligned} \quad (7)$$

となる．位相制御発振器の入力の位相差は d_{ij} に f_s を乗じた形で d_{ij} の影響を受けている．f_s もまた d_{ij} によって変化するので，f_sT_{ij} は d_{ij} の非線形の関数となる．局間の遅延を時間 d_{ij} で測定するのではなく，その時の系統周波数における位相で測定することによって，式（7）は線形化される．すなわち g_{ij} を 2π を1で正規化した位相で表わした d_{ij} とすれば

$$g_{ij}=f_sd_{ij} \quad (8)$$

とおくことができ

$$\begin{aligned} f_sT_{ij}=&\sum_{q=1}^{n}(p_{qj}-p_{qi})f_{0q} \\ &-\sum_{q=1}^{n}(p_{qj}-p_{qi})\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}g_{rq} \\ &+\sum_{q=1}^{n}(p_{qj}-p_{qi})\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}k_{rq}+g_{ij}-k_{ij} \end{aligned} \quad (9)$$

となる．これを式（5）に代入すれば，

$$\begin{aligned} &\frac{1}{2}+\sum_{q=1}^{n}(p_{qi}-p_{qj})f_{0q} \\ &>\sum_{q=1}^{n}(p_{qj}-p_{qi})\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}k_{rq} \\ &-\sum_{q=1}^{n}(p_{qj}-p_{qi})\sum_{\substack{r=1\\r\neq q}}^{n}\alpha_{rq}g_{rq}+g_{ij}-k_{ij} \\ &>\frac{-1}{2}+\sum_{q=1}^{n}(p_{qi}-p_{qj})f_{0q} \end{aligned} \quad (10)$$

となる．式(10)は局 j における局 i からの信号の到来位相を受信する位相比較器ののこぎり波状特性に関する式である．系の接続の数を Ω で表わすと式(10)はこれらの接続に関係した Ω 個の位相比較器のすべてについて成立することになり式(10)は Ω 元連立不等式となる．k_{ij} および g_{ij} もそれぞれ Ω 個ずつ存在するから，Ω 個の k_{ij} および g_{ij} の作るベクトルを，

$$\boldsymbol{K}=\begin{bmatrix} k_{12} \\ k_{21} \\ \vdots \\ k_{n,n-1} \end{bmatrix} \qquad \boldsymbol{G}=\begin{bmatrix} g_{12} \\ g_{21} \\ \vdots \\ g_{n,n-1} \end{bmatrix} \quad (11)$$

で表わすことにする．式（9）の k_{ij} および g_{ij} に関する項は符号のみが異なる絶対値が等しい係数を持っている．ここで g_{rq} が f_sT_{ij} に与える影響を $f_{(i,j)(r,q)}g_{rq}$ で与えることにする．これは $(i,j)\neq(r,q)$ であれば

$$f_{(i,j)(r,q)}=(p_{qi}-p_{qj})\alpha_{rq} \quad (12)$$

であり，$(i,j)=(r,q)$ であれば，これに1を加えて，

$$f_{(i,j)(r,q)}=(p_{qi}-p_{qj})\alpha_{rq}+1 \quad (13)$$

で与えられる．つぎに

$$\sum_{q=1}^{n}(p_{qj}-p_{qi})f_{0q}$$

をすべての接続についてならべた Ω 元のベクトルを $\boldsymbol{\Phi}(f_0)$ で示そう．これは各位相比較器の出力に対する各局の中心周波数の影響を表わすものであり式(14)によって与えられる．

$$\boldsymbol{\Phi}(f_0)=\begin{bmatrix} \sum_q (p_{q2}-p_{q1})f_{0q} \\ \sum_q (p_{q1}-p_{q2})f_{0q} \\ \vdots \\ \sum_q (p_{qn-1}-p_{qn})f_{0q} \end{bmatrix} \quad (14)$$

また $\boldsymbol{I}$ としてつぎのような Ω 元のベクトルを考える．

$$\boldsymbol{I}=\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} \tag{15}$$

このようなベクトルおよび行列を用いて式 (10) の Ω 元連立不等式を式 (16) のようにベクトル表示する．

$$\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0)+\boldsymbol{FG}>\boldsymbol{FK}>-\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0)+\boldsymbol{FG} \tag{16}$$

相互同期系においてその接続，g_{ij} および f_{0j} が指定されたとき不等式 (16) を満足する整数ベクトル $\boldsymbol{K}$ に対応する動作モードが存在する．このように相互同期系の動作モードの問題は与えられた系統において式 (16) を満足する整数ベクトル $\boldsymbol{K}$ を求める問題に帰着する．またすべての f_{0j} および g_{ij} が与えられたときに式 (16) を満足する $\boldsymbol{K}$ が存在するということは，このような $\boldsymbol{K}$ を仮定して f_s および ϕ_j を求め，これを式 (1 b) に代入すれば，ここで仮定した $\boldsymbol{K}$ が再び得られることを示している．したがって式 (16) を満足する $\boldsymbol{K}$ が存在することが，系統周波数が存在する必要十分条件となる．

一例として行列 $\boldsymbol{F}$ を図 1 に示すような完全接続系で各 $\alpha_{ij}(i\neq j)$ が等しい 3 局の場合につき求めて見ると接続 (i,j) を (12, 21, 23, 32, 31, 13) の順に並べるとして

$$\boldsymbol{F}=\begin{bmatrix}2/3, & 1/3, & 0, & -1/3, & 1/3, & 0\\1/3, & 2/3, & 0, & 1/3, & -1/3, & 0\\1/3, & 0, & 2/3, & 1/3, & 0, & -1/3\\-1/3, & 0, & 1/3, & 2/3, & 0, & 1/3\\0, & -1/3, & 1/3, & 0, & 2/3, & 1/3\\0, & 1/3, & -1/3, & 0, & 1/3, & 2/3\end{bmatrix} \tag{17}$$

となる．

本論文では以下本節の定式化を基礎にして，動作モードの問題を明らかにする．局間遅延を g_{ij} でなく

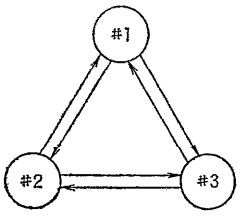


図 1　3 局の完全接続系
Fig. 1—Complete connection of three offices.

d_{ij} で与える必要のある場合には，その時与えられる f_s を用い，式 (8) によってこれらを換算すればよい．

3. 局間伝送遅延の調整と遅延の調整された系の動作モード

PCM 統合通信系においては，各局の位相と到来符号パルスの位相が同様となるように位相同期が行なわれる．このため既に御報告した相互同期系では[2]多入力位相制御発振器を含む周波数同期装置の他に位相同期装置を持っている．位相の同期を容易にするために，各局の位相比較器の出力は f_{0j} の偏差によるもの以外はすべて 0 となるように局間遅延の調整を行なうことが望ましく，既に御報告した方式ではすべての局間遅延をフレーム周期の整数倍に選ぶようになっている．各局の位相比較器の出力を f_{0j} の偏差によるもの以外すべて 0 とするためには，すべての局間遅延をフレーム周期の整数倍にする必要は必ずしもなく，前節の表現に従って

$$\boldsymbol{FG}=\boldsymbol{FK} \tag{18}$$

となるよう局間遅延を調整すればよい．

n 個の局の内の 1 局を基準として他の $(n-1)$ 局の位相を任意に指定したとき，その位相差に等しい局間伝送遅延を持つ伝送路で各局間を接続することが可能であることを考えると，式 (18) の解 $(\boldsymbol{G}-\boldsymbol{K})$ は一般に $(n-1)$ 個の自由度を持っている．したがって行列 $\boldsymbol{F}$ の rank は一般に $\Omega-(n-1)$ となっている．

局間伝送遅延が式 (18) を満足していれば，明らかに系統周波数は d_{ij}, k_{ij} の影響を受けることはなく，

$$f_s=\begin{vmatrix}f_{01}, & \alpha_{21}, & \cdots, & \alpha_{n1}\\f_{02}, & \alpha_{22}, & \cdots, & \alpha_{n2}\\\vdots & \vdots & \cdots\cdots & \vdots\\f_{0n}, & \alpha_{2n}, & \cdots, & \alpha_{nn}\end{vmatrix}\Bigg/\begin{vmatrix}1, & \alpha_{21}, & \cdots, & \alpha_{n1}\\1, & \alpha_{22}, & \cdots, & \alpha_{n2}\\\vdots & \vdots & & \vdots\\1, & \alpha_{2n}, & \cdots, & \alpha_{nn}\end{vmatrix} \tag{19}$$

で与えられる．換言すれば，式(18)が満足されているならば，系統周波数は式(19)によって一義的に定まる．このように局間遅延を調整するためにはまず適当な $\boldsymbol{K}$ を仮定し，式 (18) に従って $\boldsymbol{G}$ を定めればよく，このような $\boldsymbol{K}$ が保たれている限りにおいては，各局の位相比較器の出力は f_{0j} の偏差によるもの以外すべて 0 であり，各局の位相は見かけ上一致している．このような $\boldsymbol{K}$ が保たれている系は位相一致モードにあると呼ばれる．系はこのような状態で動作させることが望ましく，望ましい動作モードは位相一致モードである．

しかしながら，このように調整された $\boldsymbol{G}$ を持つ系において，式 (16) を満足する $\boldsymbol{FG}\neq\boldsymbol{FK}'$ なる他の $\boldsymbol{K}'$ が存在したとしよう．このような $\boldsymbol{K}'$ によって定まる動作モードもまた系統中に安定に存在できる．こ

の動作モードでは各局の位相比較器の出力には0でないものがあり，各局の位相は見かけ上一致していないので位相不一致モードと呼ばれる．位相不一致モードにある系の位相同期は困難であり，このような動作モードは望ましくない．位相不一致モードには位相一致モードの場合と局間位相関係のみが異なり，系統周波数が同一である場合と，局間位相関係および系統周波数が共に異なる場合とがあり，前者は位相モードのみが異なり後者は周波数モードもまた異なっていると呼んでよいであろう．

遅延が調整された系においてどのような動作モードが存在しうるかを考えるためには，このような $\boldsymbol{K}'$ を考え，さらに $\boldsymbol{FG}=\boldsymbol{FK}$ であるから

$$\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0)+\boldsymbol{FK} > \boldsymbol{FK}' > -\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0)+\boldsymbol{FK} \qquad (20)$$

したがって

$$\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0) > \boldsymbol{F}_1(\boldsymbol{K}'-\boldsymbol{K}) > -\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0) \qquad (21)$$

ここで $\boldsymbol{K}'-\boldsymbol{K}$ を改めて $\boldsymbol{K}$ とおけば，

$$\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0) > \boldsymbol{FK} > -\frac{\boldsymbol{I}}{2}+\boldsymbol{\Phi}(f_0) \qquad (22)$$

となる．したがって，式 (18) を満足するように局間遅延が調整された状態における系統の動作モードは，局間遅延がすべて0である場合の動作モードの問題に帰着して考えて良い．

このような $\boldsymbol{K}'$ で表わされる動作モードでは各位相比較器の出力が，$\boldsymbol{K}$ で表わされる動作モードの場合と大幅に異なることは式 (9) から明らかであり，先に御報告した位相モード選択装置[2]の位相不一致モード検出器はこれを検出することができるので，このような動作モードは位相モード選択装置を用いて抑圧することができる．

4. 動作モードの存在範囲および動作モードの変化

相互同期系において，g_{ij}, f_{0j} および α_{ij} がすべて与えられれば系統は与えられたことになるが，式(16)を満足する $\boldsymbol{FK}$ は一義的には定まらず，$\boldsymbol{FK}$ の値によって種々の動作モードが生ずる．行列 $\boldsymbol{F}$ の rank は $\Omega-(n-1)$ であるから，

$$\boldsymbol{FK}=\boldsymbol{L} \qquad (23)$$

によってベクトル $\boldsymbol{L}$ を定義すれば，一つの $\boldsymbol{L}$ に対して $(n-1)$ 次元部分空間の格子点を占めるベクトル $\boldsymbol{K}$ が対応する．

系統が与えられその動作モードが定まった後に f_{0j} または g_{ij} の内のあるものが変動し，式 (16) を満足しなくなると，$\boldsymbol{K}$ は式 (16) を満足させるように変化する．これを動作モードの変動と呼ぶが，式 (16) の Ω 個の不等式の内局 i から局 j に至る接続に対応する不等式の下限または上限が満足しなくなると，k_{ij} が1だけそれぞれ減少または増加する．系統が動的に十分安定であれば[5]，この時の動作モードの変動によるじょう乱でこれ以外の動作モードは生じない．こうしてできた新しい $\boldsymbol{K}$ が，その時の g_{ij} および f_{0j} に対し式 (16) を満たすとは限らず，式 (16) が満たされなければ，さらに満たさない要素に対応する k_{ij} は1だけ減少または増加する．これを繰返して適切な $\boldsymbol{K}$ において式 (16) が満足されれば，この時相互同期系は相互同期状態を保つことができる．$\boldsymbol{F}$ の rank は $(\Omega-n+1)$ であるので，$\boldsymbol{G}$ が任意に変化したときに不等式の内，1個ないし $(\Omega-n+1)$ 個および Ω 個を同時に満足しなくなることは有り得るが，これ以外の個数を同時に満足しなくなることはない．したがって動作モードの変動は，一時には1ないし $(\Omega-n+1)$ または Ω の k_{ij} についてのみ生じ，これ以外の数の k_{ij} の同時変化は生じ得ない．

ある $\boldsymbol{K}$ が与えられたとき

$$\frac{\boldsymbol{I}}{2}-\boldsymbol{\Phi}(f_0)+\boldsymbol{FK} > \boldsymbol{FG} > -\frac{\boldsymbol{I}}{2}-\boldsymbol{\Phi}(f_0)+\boldsymbol{FK} \qquad (24)$$

を満足する $\boldsymbol{G}$ の開空間の中を $\boldsymbol{G}$ が変動しても動作モードすなわち $\boldsymbol{K}$ の変化はない．これを満足する二つの $\boldsymbol{G}$ を $\boldsymbol{G}_1$, $\boldsymbol{G}_2$ とすれば，$1\geq x\geq 0$ として

$$\boldsymbol{G}=x\boldsymbol{G}_1+(1-x)\boldsymbol{G}_2 \qquad (25)$$

は式 (24) を満足する．すなわち空間 $\boldsymbol{G}$ は凸であり，もし系統のあるモードを満足する g_{ij} の二つのグループ $\boldsymbol{G}_1$ および $\boldsymbol{G}_2$ を知ることができれば，その一方の状態 $\boldsymbol{G}_1$ から他方の状態 $\boldsymbol{G}_2$ にはモードの変動をともなわずに移ることができる．

以上動作モードの性質を一般的に示したが，実用上の相互同期系においては，正常の動作状態では局間伝送遅延 g_{ij} は **3.** に従って調整されており，各局の中心周波数 f_{0j} の偏差は α に比して十分小さく作られていることを考えれば，g_{ij} もしくは f_{0j}/α が系統周期もしくはその 1/2 程度変動しなければ，このような動作モードの変化が生ずることはない．既に御報告したようなフレーム周波数を系統周波数とする相互同期

系[2]ではこれらのパラメータの変動は系統周期に比べて十分小さい．したがって相互同期系には多数の動作モードが存在するにもかかわらず系統は初期に設定された動作モードを十分安定に保つことができる．さらにもし事故によって g_{ij} または f_{0j} が大幅に変動して動作モードの変化が生じても，変化する k_{ij} は事故の生じた部分に関係するものが主として変化するので，じょう乱の波及は局限される．

5. 均等完全接続系の動作モード

以上の所論によって系統の動作モードの主な性質を定式化した．以下相互同期系の一例として n 局がすべて均等な荷重で接続された，n 局の均等完全接続系について具体例を示す．

5.1 系統周波数と局間位相関係

n 局の完全接続系において各 α_{ij} を

$$\alpha_{ij}=\alpha/(n-1) \quad (i \neq j) \tag{26}$$

と定めこれを式（3）の方法で解けば，

$$f_s=\sum_{j=1}^{n}\frac{f_{0j}}{n}+\frac{\alpha(\sum_{i,j}k_{ij}-\sum_{i,j}g_{ij})}{n(n-1)} \tag{27}$$

$$\phi_j=\frac{n-1}{n\alpha}(f_{0j}-f_{01})+\frac{1}{n}(\sum_{i=2}^{n}g_{i1}-\sum_{\substack{i=1\\i\neq j}}^{n}g_{ij})$$

$$-\frac{1}{n}(\sum_{i=2}^{n}k_{i1}-\sum_{\substack{i=1\\i\neq j}}^{n}k_{ij}) \tag{28}$$

となる．したがって

$$f_sT_{ij}=\frac{n-1}{n\alpha}(f_{0j}-f_{0i})+\frac{1}{n}(\sum_{\substack{q=1\\q\neq i}}^{n}g_{qi}-\sum_{\substack{q=1\\q\neq j}}^{n}g_{qj})$$

$$+g_{ij}-\frac{1}{n}(\sum_{\substack{q=1\\q\neq i}}^{n}k_{qi}-\sum_{\substack{q=1\\q\neq j}}^{n}k_{qj})-k_{ij} \tag{29}$$

である．

5.2 周波数モードの一意性と安定性

系の同期の必要性は $f_{0i}\neq f_{0j}$ であるところから生じてくるものであるが，現実の相互同期系では $(f_{0j}-f_{0i})(n-1)/n\alpha$ を無視できる程度に小さくすることは容易であり，この項は動作モードには大きな影響を与えない．この節ではしたがって $\boldsymbol{\Phi}(f_0)$ の項を無視するが，f_{0j} に大幅な変動がある場合にはさらにこの項を修正項として入れればよい．

前述のように局間伝送遅延が式（18）を満足するように調整されていれば，動作モードの問題は局間遅延がすべて0である場合の問題に帰着する．すなわち

$$\frac{\boldsymbol{I}}{2}>\boldsymbol{FK}>-\frac{\boldsymbol{I}}{2} \tag{30}$$

なる $\boldsymbol{K}$ を求める問題となる．k_j を

$$k_j=\sum_{\substack{q=1\\q\neq j}}^{n}k_{qj} \tag{31}$$

と定義すれば，式（28）から明らかに，

$$k_j=\sum_{\substack{q=1\\q\neq j}}^{n}\left[\frac{k_q-k_j}{n}+\frac{1}{2}\right] \tag{32}$$

である．ガウス記号の性質から

$$\left[x+\frac{1}{2}\right]=-\left[-x+\frac{1}{2}\right] \tag{33}$$

であるから，式（31）の k_j をすべての j について加えれば，絶対値が等しく符号が逆の要素が相殺して，

$$\sum_j k_j=\sum_{i,j}k_{ij}=0 \tag{34}$$

である．したがって明らかに伝送遅延を式（18）を満足するように調整した n 局の均等完全接続系では，系がいかなる動作モードをとっても周波数モードは一つに定まり，

$$f_s=\frac{\sum_{j=1}^{n}f_{0j}}{n} \tag{35}$$

で与えられる系統周波数が定まる．

g_{ij} に変化が生じたために局 j における局 i からの入力位相と自局の位相を比較する位相比較器の k_{ij} が変化するのは式（29）から明らかなように

$$\frac{1}{n}(\sum_{\substack{q=1\\q\neq i}}^{n}g_{qi}-\sum_{\substack{q=1\\q\neq j}}^{n}g_{qj})+g_{ij}$$

$$=\frac{1}{n}(\sum_{\substack{q=1\\q\neq j}}^{n}k_{qi}-\sum_{\substack{q=1\\q\neq j}}^{n}k_{qj})+k_{ij}\pm\frac{1}{2} \tag{36}$$

が成立する場合である．$\boldsymbol{G}$ のすべての要素 g_{rq} が最悪の方向に g だけ変動し，$\boldsymbol{G}$ が $\boldsymbol{G}'$ になった場合，すなわち，

$$\left.\begin{aligned}g_{ij}'&=g_{ij}+g\\g_{qi}'&=g_{qi}+g\\g_{qj}'&=g_{qj}-g \quad (q\neq i)\end{aligned}\right\} \tag{37}$$

のように変動した場合，g_{qj} には g_{ij} が含まれていることを考慮して，式（37）を式（36）に代入すれば，$\boldsymbol{FG}=\boldsymbol{FK}$ が成立するよう元の $\boldsymbol{G}$ が定められていた場合，

$$|g|(3n-4)=\frac{n}{2} \tag{38}$$

の時式（36）が満足する．換言すればすべての g が調整された値から最悪の方向に $n/2(3n-4)$ フレームだけ変動すると動作モードの変動が生ずることになる．これはいかなる n に対しても 1/6 より大きい．すなわちすべての局間遅延がその時のフレーム周期の 1/6 にわたって大幅にしかも最悪の方向に変動しない限り，動作モードの変化は生じない．

5.3 系統周波数存在の必要十分条件

前述のように系統が与えられたとき式 (16) を満足する $\boldsymbol{K}$ が存在することが系統周波数が存在する必要十分条件である．これは $\boldsymbol{G}$ および各 f_{0j} の許容範囲を定める．一例として，本節の系統で n 局中の一局たとえば局1の中心周波数 f_{01} が変動し，他の局の中心周波数が f_{00} に保たれている場合につき，たとえば g_{12} 以外の g_{ij} を0に調整し，g_{12} のみを変化させたとき g_{12} の変化が $(f_{00}-f_{01})$ の許容値にいかなる影響を与えるかを示そう．中心周波数の許容偏差は g_{12} の変化に応じて k_{12} のみが変化し，他の k_{ij} が0に保たれる動作モードで最大になるからこれについて考える．$\sum_{q=1}^{n}(p_{qj}-p_{qi})f_{0q}=\frac{n-1}{n\alpha}(f_{0j}-f_{0i})$

であり，g_{12} の変化は f_sT_{12} に $\frac{n-1}{n}g_{12}$ の影響を，$f_sT_{p2}(p\neq 1,2)$ に $-\frac{1}{n}g_{12}$ の影響を，$f_sT_{p1}(p\neq 1)$ に $\frac{1}{n}g_{12}$ の影響を与え，他の f_sT_{ij} には影響を与えない．したがって式 (16) の連立不等式のうち典型的な二つをすべての k_{ij} が0である動作モードについて示せば

$$\left.\begin{aligned}&\frac{n-1}{n}g_{12}-\frac{1}{2}+\frac{(n-1)(f_{00}-f_{01})}{n\alpha}\\&\quad<0<\frac{n-1}{n}g_{12}+\frac{1}{2}+\frac{(n-1)(f_{00}-f_{01})}{n\alpha}\\&\frac{1}{n}g_{12}-\frac{1}{2}+\frac{(n-1)(f_{01}-f_{00})}{n\alpha}\\&\quad<0<\frac{1}{n}g_{12}+\frac{1}{2}+\frac{(n-1)(f_{01}-f_{00})}{n\alpha}\end{aligned}\right\}\tag{39}$$

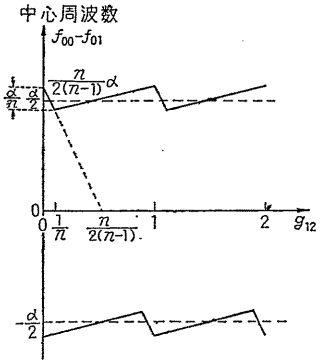


図 2　n 局均等完全接続系における系統周波数存在の必要十分条件を示す許容しうる中心周波数の偏差と正規化位相で表わした伝送遅延 g_{12} の関係．g_{12} 以外の伝送遅延はすべて0としている．

Fig. 2—Relation of permissible deviation of center frequency versus g_{12}, transmission delay in terms of normalized phase, showing necessary and sufficient condition for existence of system frequency, in completely connected n offices where all transmission delays except g_{12} are assumed to be zero.

式 (39) は

$$\frac{n}{2(n-1)}\alpha+\frac{g_{12}}{n-1}\alpha<f_{00}-f_{01}<\frac{n}{2(n-1)}\alpha-g_{12}\alpha\tag{40}$$

と変形できる．さらに $k_{12}=1,2$ と変化させて同様に求めた結果をあわせて図2に示す．これによれば一局の中心周波数のみがずれる場合の許容される範囲は，ほぼ $\pm\frac{\alpha}{2}$ である．これは系統内のパラメータがこのように変化した場合の系統周波数存在の必要十分条件を示す．

6. むすび

以上相互同期系の動作には複数個の動作モードが存在しうること，これらの動作モードが位相比較器の入出力特性の周期性を表わす整数のパラメータ k_{ij} を要素とするベクトル $\boldsymbol{K}$ によって一般的に表現できることを示した．系統が与えられたときその局間遅延の内の $(\Omega-n+1)$ 個を適当に選択すれば系統の動作モードは適切に定められ，このようにして定められた動作モードは系統のパラメータの変化に対して変動を生ずることはほとんどなく，系統の動作モードは初期に選択された値に十分安定に保つことができる．

一般の相互同期系についての定量的な解析は，系統が与えられれば，本報告の定式化に従って行ないうるが，本報告ではその典型的な一例として n 局から成る荷重の等しい完全接続系について具体例を示し，上記安定性の最悪値を定量的に示した．

謝辞　おわりに本研究に対し御指導いただいた猪瀬博教授，藤崎博也助教授に深甚の謝意を表する．また日頃御討論いただく本学高周波談話会の諸氏ならびに御援助いただいたベル電話研究所に対し謝意を表する次第である．

文　献

(1) 藤崎博也，斎藤忠夫，猪瀬 博："相互同期系の特性について"，昭 39 信学全大，S7-6.
(2) 猪瀬 博，藤崎博也，斎藤忠夫："相互同期系の構成"，信学誌，**49**，11，p. 2095 (昭 41-11).
(3) 猪瀬 博，藤崎博也，斎藤忠夫："相互同期系の理論"，信学誌，**49**，4，p. 755 (昭 41-04).
(4) 斎藤忠夫，片岡雅憲，藤崎博也，猪瀬 博："相互同期系の局間位相関係"，昭 40 連大，1492.
(5) 斎藤忠夫，片岡雅憲，藤崎博也，猪瀬 博："相互同期系の動特性"，昭 41 信学全大，928.

付録　3局の均等完全接続系の動作モード

本論文に示した行列 $\boldsymbol{F}$ による動作モードの解析は，比較的小規模な系統でも Ω の数はかなり大きくなるため，筆算は困難となる．一般に系統が具体的に与えられれば，電子計算機を用いて動作モードの特性を求めることができるが，ここでは以上の議論の理解に資するために，3局の均等完全接続系について動作モードの解析を行ない相互同期系の典型的な動作モードの変動の例を示す．

A-1　局間伝送遅延の変化による周波数モードの変動

各局の中心周波数の偏差を無視すれば，**4.** に示したように系統があるベクトル $\boldsymbol{L}(=\boldsymbol{FK})$ で与えられる動作モードにあるとき，局間伝送遅延 $\boldsymbol{G}$ が

$$\frac{\boldsymbol{I}}{2}+\boldsymbol{L}>\boldsymbol{FG}>\boldsymbol{L}-\frac{\boldsymbol{I}}{2} \qquad (41)$$

を満足しなくなると，満足しなくなった不等式に対応する k_{ij} がその不等式を満足させるように変化する．

$$\boldsymbol{L}^T=(0,\ 0,\ 0,\ 0,\ 0,\ 0)$$

とすれば $\boldsymbol{K}$ の代表的解は

$$\boldsymbol{K}^T=(0,\ 0,\ 0,\ 0,\ 0,\ 0)$$

である．この時式 (41) は

$$\left.\begin{array}{llll}
\frac{1}{2}>\frac{2}{3}g_{12}+\frac{1}{3}g_{21} & -\frac{1}{3}g_{32}+\frac{1}{3}g_{31} & >-\frac{1}{2} \\
\frac{1}{2}>\frac{1}{3}g_{12}+\frac{2}{3}g_{21} & +\frac{1}{3}g_{32}-\frac{1}{3}g_{31} & >-\frac{1}{2} \\
\frac{1}{2}>\frac{1}{3}g_{12} & +\frac{2}{3}g_{23}+\frac{1}{3}g_{32} & -\frac{1}{3}g_{13}>-\frac{1}{2} \\
\frac{1}{2}>-\frac{1}{3}g_{12} & +\frac{1}{3}g_{23}+\frac{2}{3}g_{32} & +\frac{1}{3}g_{13}>-\frac{1}{2} \\
\frac{1}{2}> & -\frac{1}{3}g_{21}+\frac{1}{3}g_{23} & +\frac{2}{3}g_{31}+\frac{1}{3}g_{13}>-\frac{1}{2} \\
\frac{1}{2}> & +\frac{1}{3}g_{21}-\frac{1}{3}g_{23} & +\frac{1}{3}g_{31}+\frac{2}{3}g_{13}>-\frac{1}{2}
\end{array}\right\} \qquad (42)$$

となる．

上式の $\boldsymbol{G}$ は式 (18) を満足するよう調整された $\boldsymbol{G}$ からの変化分のベクトルであると考えても良いし，$\boldsymbol{G}$ そのものであると考えても良い．g_{12} 以外の g_{ij} をすべて0に保ち，g_{12} だけを変化させたとすれば，$g_{12}=\frac{3}{4}$ で k_{12} が1増加し

$$\boldsymbol{K}^T=(1,\ 0,\ 0,\ 0,\ 0,\ 0)$$

となる．この動作モード変動前の f_s は $f_0-\frac{\alpha}{8}$，変動後の f_s は $f_0+\frac{\alpha}{24}$ である．g_{12} を d_{12} に換算すれば，$d_{12}=6/(8f_0-\alpha)$ でモードの変動が生じていることになる．ここで g_{12} を元に戻せば同様に $g_{12}=\frac{1}{4}$ でモードが変化し，元のモードに戻る．この時の f_s は $f_0+\frac{\alpha}{8}$ で，$d_{12}=6/(8f_0+\alpha)$ である．このようにして計算した d_{12} と f_s の関係を図 A-1 に示す．この場合たとえば $\frac{6}{8f_0+\alpha}<d_{12}<\frac{6}{8f_0-\alpha}$ の区間では f_s は一義的に定まらず，系統の履歴を考慮することによってはじめて求められる．他の g_{ij} またはいくつかの g_{ij} が同時に変化する場合についても同様に求められるが，たとえばすべての g_{ij} が同時に等しい大きさで変化すれば，f_s が $f_s-\frac{\alpha}{2}$ となるまで，すなわち f_s の存在範囲の下限まで動作モードは変化しない．

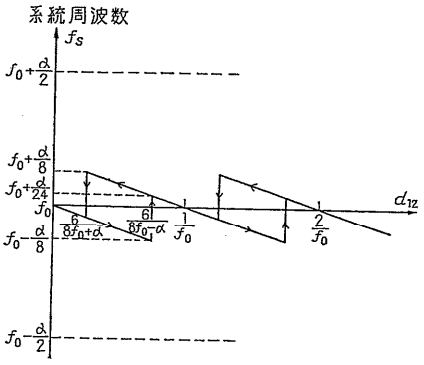


図 A-1　3局完全接続系における伝送遅延 d_{12} と系統周波数の関係．d_{12} 以外の伝送遅延はすべて0としている．

Fig. A-1—Relation of system frequency versus transmission delay d_{12} of completely connected three offices, when all d_{ij}'s besides d_{12} are kept to zero.

A-2　局間伝送遅延の変化による位相モードの変動

ある与えられた $\boldsymbol{G}$ を持ち適当な $\boldsymbol{K}$ で動作している状態から，$\boldsymbol{G}$ を変化させて後再び $\boldsymbol{G}$ を元に戻す過程によって位相モードを変化させることができる．

(例 1)　$\boldsymbol{G}^T=(0,\ 0,\ 0,\ 0,\ 0,\ 0)$

$\boldsymbol{K}^T=(-1,\ 1,\ 0,\ 0,\ 0,\ 0)$

で表わされる動作モードは，$\phi_1=0$，$\phi_2=-\frac{1}{3}$，$\phi_3=\frac{1}{3}$ の動作を行なう位相不一致モードである．この状態から g_{21} を増加させると

$$\boldsymbol{G}^T=\left(0,\ \frac{1}{2},\ 0,\ 0,\ 0,\ 0\right)$$

で k_{12} が1だけ変化し

$$\boldsymbol{K}^T=(0,\ 1,\ 0,\ 0,\ 0,\ 0)$$

となる．g_{12} をさらに増加させると

$$\boldsymbol{G}^T=(0,\ 1,\ 0,\ 0,\ 0,\ 0)$$

でも上記 $\boldsymbol{K}$ は保存され，これはこの $\boldsymbol{G}$ における位相一致モードで，また $\boldsymbol{G}$ を元に戻しても位相一致モードは保たれる．

(例 2)　$\boldsymbol{G}^T=(1,\ 1,\ 1,\ 1,\ 1,\ 1)$

$\boldsymbol{K}^T=(1,\ 1,\ 1,\ 1,\ 1,\ 1)$

は明らかに式 (19) を満足する位相一致モードである．ここで，

$$\boldsymbol{G}^T=(0,\ 2,\ 1,\ 1,\ 1,\ 1)$$

としても $\boldsymbol{K}^T$ は保存される．さらに g_{21} を 1/4 増加させると

$$\boldsymbol{G}^T=\left(0,\ \frac{9}{4},\ 1,\ 1,\ 1,\ 1\right)$$

$$\boldsymbol{K}^T=(0,\ 2,\ 1,\ 1,\ 1,\ 1)$$

に変化し，ここで $\boldsymbol{G}$ を元に戻すと $\boldsymbol{K}'$ は不変で

$$\boldsymbol{G}^T=(1,\ 1,\ 1,\ 1,\ 1,\ 1)$$

$$\boldsymbol{K}^T=(0,\ 2,\ 1,\ 1,\ 1,\ 1)$$

となる．これは例1の最初の状態と等価な位相不一致モードである．

A-3 動作モードの存在範囲

ある $\boldsymbol{L}$ で与えられる動作モードの限界は

$$\boldsymbol{L}+\frac{\boldsymbol{I}}{2}>\boldsymbol{FG}>\boldsymbol{L}-\frac{\boldsymbol{I}}{2} \qquad (43)$$

を与える $\boldsymbol{G}$ によって定まるが，図A-1のようなグラフが，$G=\sum_{i,j} g_{ij}$ と f_s の間に成立する動作モードの限界は

$$\boldsymbol{FG}>\boldsymbol{L}-\frac{\boldsymbol{I}}{2},\ \boldsymbol{G}\geq 0 \qquad (44)$$

なる条件下で $\min G=\sum_{i,j} g_{ij}$ を可能なすべての $\boldsymbol{L}$ について求めて見れば決定できる．これは通常の linear programming の問題であり，図A-2に示す実線の範囲内でのみ，G と f_s の関係が成立し得ることが結論

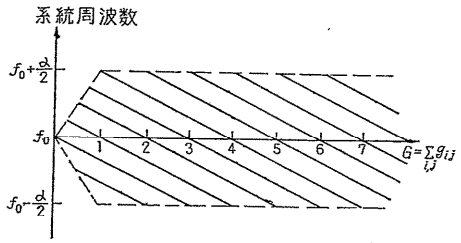


図 A-2　3局完全接続系の動作モードの存在範囲
Fig. A-2—Range of operation modes possible in completely connected three offices.

できる．

A-4 ベクトル $\boldsymbol{K}$ の占める部分空間

例に示す系では $\boldsymbol{FK}=\boldsymbol{L}$ は

$$\left.\begin{array}{llll}
\frac{2}{3}k_{12}+\frac{1}{3}k_{21} & -\frac{1}{3}k_{32}+\frac{1}{3}k_{31} & & =l_{12} \\
\frac{1}{3}k_{12}+\frac{2}{3}k_{21} & +\frac{1}{3}k_{32}-\frac{1}{3}k_{31} & & =l_{21} \\
\frac{1}{3}k_{12} & +\frac{2}{3}k_{23}+\frac{1}{3}k_{32} & -\frac{1}{3}k_{13} & =l_{23} \\
-\frac{1}{3}k_{12} & +\frac{1}{3}k_{23}+\frac{2}{3}k_{32} & +\frac{1}{3}k_{13} & =l_{32} \\
-\frac{1}{3}k_{21}+\frac{1}{3}k_{23} & & +\frac{2}{3}k_{31}+\frac{1}{3}k_{13} & =k_{31} \\
+\frac{1}{3}k_{21}-\frac{1}{3}k_{23} & & +\frac{1}{3}k_{31}+\frac{2}{3}k_{13} & =l_{13}
\end{array}\right\} \qquad (45)$$

で示される．この式の $\boldsymbol{K}$ に関する一般解は，C_1, C_2 を任意の整数として

$$\boldsymbol{K}_0=C_1\begin{bmatrix}1\\-1\\-1\\1\\0\\0\end{bmatrix}+C_2\begin{bmatrix}1\\-1\\0\\0\\-1\\1\end{bmatrix}$$

であり，可能な特解は

$$\boldsymbol{K}_1=\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}\quad \boldsymbol{K}_2=\begin{bmatrix}1\\-1\\0\\0\\0\\0\end{bmatrix}\quad \boldsymbol{K}_3=\begin{bmatrix}-1\\1\\0\\0\\0\\0\end{bmatrix}$$

であり，これらに対応する $\boldsymbol{L}$ はそれぞれ

$$\boldsymbol{L}_1=\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}\quad \boldsymbol{L}_2=\begin{bmatrix}1/3\\-1/3\\1/3\\-1/3\\1/3\\-1/3\end{bmatrix}\quad \boldsymbol{L}_3=\begin{bmatrix}-1/3\\1/3\\-1/3\\1/3\\-1/3\\1/3\end{bmatrix}$$

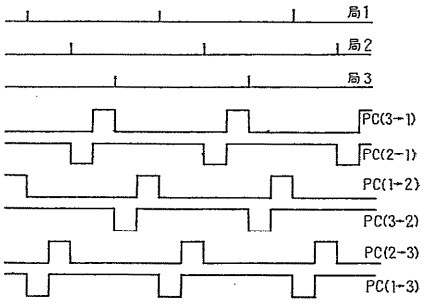


図 A-3　3局完全接続系の位相不一致モードの一例
Fig. A-3—Timing of completely connected three offices in an out-of-phase mode.
(PC : Phase Comparator)

である．$\boldsymbol{K}_1$, $\boldsymbol{K}_2$, $\boldsymbol{K}_3$ の各々に $\boldsymbol{K}_0$ を加えた $\boldsymbol{K}$ の部分空間がこの系の三つの位相モードに対応する．一例として $\boldsymbol{K}_3$ で示される動作モードにおける各局の発振位相と各位相比較器出力のタイミング図を図 A-3 に示す．

(昭和42年4月5日受付)

UDC 621.394.423

相互同期系の一般化された構成概念

正員　斎藤　忠夫†　正員　丸山　勝巳†　正員　猪瀬　博†

Generalized Arrangement of Mutually Synchronized System

Tadao Saito†, *Member*, Katsumi Maruyama†, *Member* *and* Hiroshi Inose†, *Member*

あらまし　相互同期方式によれば，大規模な通信系で安定な同期を確保できるが，各局の中心周波数差，局間遅延変動によって局入力における位相差が生ずること，あるいは局間遅延変動によって系統周波数の変動が生ずることが問題となっている。これらの問題を解決するためには，各局の位相比較器の出力を相互に伝送して，各局の周波数を制御する方法，伝送路に設けた可変遅延線路を制御する方法，位相比較器の出力を積分して各局の周波数を制御する方法があり，また系を安定化するにもいくつかの方法がある．本論文では，これらの各制御方式を利用した 10 種類の相互同期方式を示し，その特性を解析し，シミュレーションによってその実現性を確認するとともに各方式の特徴を論じている．

Summary　In mutually synchronized system although stable frequency synchronization is established in the network, phase difference at the input of each office and fluctuation of system frequency are caused due to fluctuation of the parameters of offices and transmission delay between offices. Techniques to improve the characteristics of the system by integration and transmission of phase difference information between offices are proposed. Combining these techniques 4 schemes without integrater and 6 schemes with integrater are proposed and demonstrated to be feasible by computer simulation.

1. まえがき

相互同期の概念が発表されてから[1]，各方面で行なわれた相互同期方式により，このような概念を時分割統合通信系の同期方式に適用することが可能であることが，理論的，実験的に明らかにされてきた[2]~[7]．相互同期方式の概念とは，地理的に離隔した多数の局に設けられた発振器の相互作用によって，各局に共通な系統周波数を得ようとするものである．従来研究の対象となっている相互同期方式はこのような概念を実現するものとして，各局で受信された同期信号と各局の同期信号の位相差によって局の周波数を制御する基本相互同期方式[2]と，相互に接続された局における位相比較器出力の差によって各局の周波数を制御する平衡相互同期方式が知られている[3]．

相互同期方式によれば，系全体において安定な周波数同期を確保することができるが，各局の中心周波数差，あるいは局間遅延の変動に起因する各局の入力における位相差と，局間遅延の変動に起因する系統周波数の変動が問題となっている．各局の入力における局間遅延変動に起因する位相差は各局間の伝送路に可変遅延線を設けて，これを制御することによって軽減できる．一方，中心周波数差に起因する位相差は位相比較器の出力を積分して各局の周波数を制御すれば，積分制御の原理に従って除去できる．このようにして局入力における位相差が軽減できれば，位相調整のための位相同期装置[8]は除去できるか，あるいはその容量を大幅に減ずることができ，系を経済的に実現できる．局間遅延の変動に起因する系統周波数の変動は，上述のようにして遅延変動を補正することによって除去することもできるが，また相互に接続された2局の位相比較器の出力の差を得て，これで各局の周波数を制御することによっても除去することができる．

本論文ではこのような相互同期系の各制御方式を示し，これらを用いた各種の相互同期方式について，そ

† 東京大学工学部，東京都
Faculty of Engineering, University of Tokyo, Tokyo, Japan 113
論文番号：昭 45-815 [A-212]

の構成，特性，特徴を論じている．

2. 相互同期方式の原理

ここでは各方式の原理を明確にするために基本的な2局の相互同期系について系を制御するための種々の情報を用いる方式について述べる．ここで2局について示した結果は一般的に任意の局を含む系に拡張できる．

2.1 基本相互同期方式

図1は基本相互同期方式の各局に用いる多入力位相制御発振器の構成図である．システムの各局からのクロック信号は相互に伝送され，他局から送られたクロック信号の位相はその局の位相と位相比較器（PC）によって比較される．本論文では位相比較器としてはのこぎり波状特性を持つものを仮定する．加算回路が位相比較器の出力の和を形成し，可変周波数発振器（VFO）の出力を制御する．局 $\#i$ からのクロック信号を受信する局 $\#j$ の位相比較器の出力を θ_{ij} で表わし，局 $\#j$ から局 $\#j$ にいたる伝送遅延を d_{ij} で，局 $\#i$ の位相を ϕ_i で，系統周波数を f_s で表わせば，θ_{ij} の値は，

$$\theta_{ij}=\phi_j-\phi_i+f_s d_{ij}-k_{ij} \qquad (1)$$

$$k_{ij}=\left[\phi_j-\phi_i+f_s d_{ij}+\frac{1}{2}\right] \qquad (2)$$

［　］はガウス記号

となる．k_{ij} は位相比較器ののこぎり波状特性を表わすものである．位相制御ループのループ利得を α とすれば，局 $\#j$ の中心周波数を f_{0j} として，2局の場合の f_s と θ_{21} はつぎのようになる．

$$f_s=\frac{f_{01}+f_{02}+\alpha(k_{21}+k_{12})}{2+\alpha(d_{12}+d_{21})} \qquad (3)$$

$$\theta_{21}=\frac{f_{01}-f_{02}}{2\alpha}+\frac{d_{21}+d_{12}}{2}f_s-\frac{k_{21}+k_{12}}{2} \qquad (4)$$

2.2 遅延制御情報[9]

相互に接続された2局において，両局の位相比較器の出力の和をとることができるとしよう．式(1)から，

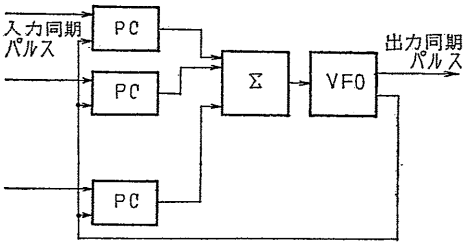


図1　基本相互同期方式（分類1）の各局のクロック源
Fig. 1—Structure of clock source of each office in a basic mutually synchronized system (class 1).

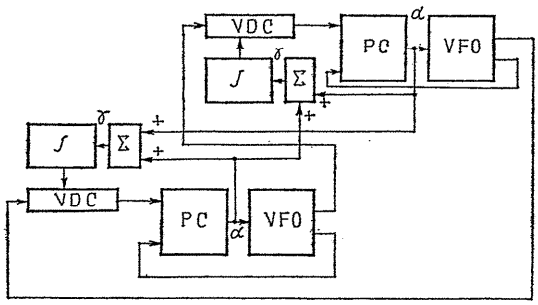


図2　可変遅延回路を持つ相互同期方式（分類2）
Fig. 2—Mutually synchronized system having delay compensation unit (class 2).

$$\theta_{12}+\theta_{21}=f_s(d_{21}+d_{12})-(k_{21}+k_{12}) \qquad (5)$$

局間の遅延は系統周期（$=1/f_s$）の整数倍に調整されるが，式(5)は局間の遅延の調整値からの偏差を示すものである．したがって図2に示すように局間に可変遅延回線（VDC）を設け，式(5)の値が零となるよう可変遅延回路を制御すれば，遅延は常に系統周期の整数倍に保たれるから，

$$f_s=\frac{f_{01}+f_{02}}{2} \qquad (6)$$

$$\theta_{21}=\frac{f_{01}-f_{02}}{2\alpha} \qquad (7)$$

となる．遅延制御方式を用いれば系統周波数も局の入力における位相差も局間遅延の影響を受けない．

2.3 平衡相互同期方式[3]

図3に示すように各局の可変周波数発振器を自局の位相比較器の出力と相手局の位相比較器の出力の差で制御し，

$$f_s=f_{0j}-\alpha(\theta_{ij}-\theta_{ji}) \qquad (8)$$

とすれば，

$$f_s=\frac{f_{01}+f_{02}}{2} \qquad (9)$$

$$\theta_{21}=\frac{f_{01}-f_{02}}{4\alpha}+\frac{d_{21}+d_{12}}{2}f_s-\frac{k_{21}+k_{12}}{2} \qquad (10)$$

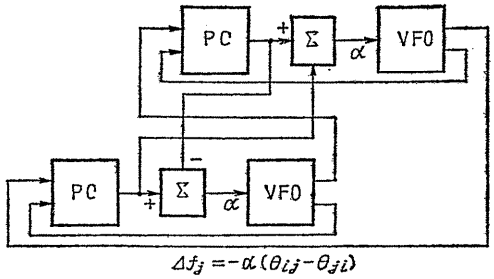
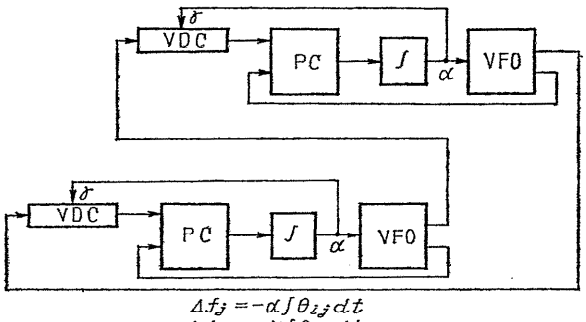


図3　平衡相互同期方式（分類3）
Fig. 3—Balanced control of mutually synchronized system (class 3).

となる．すなわちこの方式では，系統周波数は遅延を制御した方式と同様に局間遅延の影響を受けずに決まるが，各局の入力における位相差は局間遅延変動の影響を受ける．

2.4 位相比較器と可変周波数発振器の間に積分要素を含む方式

上述した三つの方式ではいずれも局間の位相差に各局の中心周波数差に比例した項が生ずる．この項は位相比較器の出力の積分値で各局の可変周波数発振器を制御することによって除くことができる．このような積分を行なうためには位相比較器の出力の定常値は零であることが必要で，この目的で局間遅延の補正を行なうことが必要であるから，このような積分要素を含む方式の基本的構成は図4に示すようになる．この方式では位相比較器の出力には中心周波数差に起因する定常値がないので，一つの位相比較器の出力だけで可変遅延回路を制御することができる．



$\Delta f_j = -\alpha \int \theta_{ij} dt$
$\Delta d_{ij} = -\gamma \int \theta_{ij} dt$

図 4 FSU に積分器を持つ相互同期方式（分類5）
Fig. 4—Mutually synchronized system with integrators in FSU (class 5).

この方式で，ϕ_i, θ_{ij} のラプラス変換を $\Phi_i(s)$, $\Theta_{ij}(s)$ で表わせば可変遅延回路の制御の定数を γ_{ij} として，局間の遅延を考えなければ，

$$\Theta_{ij}(s) = \Phi_j(s) - \Phi_i(s) - \frac{\gamma_{ij}}{s}\Theta_{ij}(s) \qquad (11)$$

となる．局 #j の初期位相を $\phi_j(0)$ とすれば，

$$s\Phi_j(s) - \phi_j(0) = \frac{f_{0j}}{s} - \frac{\alpha_{ij}}{s}\Theta_{ij}(s) \qquad (12)$$

となる．したがってこの方式は自然周波数 ω_n，制御系数 ξ として，

$$\left.\begin{aligned} \omega_n &= \sqrt{2\alpha} \\ \xi &= \gamma/2\sqrt{2\alpha} \end{aligned}\right\} \qquad (13)$$

で表わされる二次の過渡応答を持つ．可変遅延回路はこのように局間伝送遅延の偏差を補正する他に系に制動を与える役割を持つものであることがわかる．

2.5 積分＋比例制御による制動

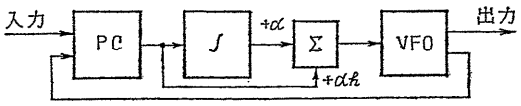


図 5 比例分制動方式
Fig. 5—Proportional component damping scheme.

積分要素を含む位相制御発振器に制動を与えるにはその特性方程式に一次の項を与えればよく，図5に示すように可変周波数発振器を位相比較器の出力に比例した信号と積分器の出力の荷重和によって制御することによっても系に制動を与えることができる．図5の位相制動発振器では，入出力信号の位相のラプラス変換を $\Phi_i(s)$, $\Phi_0(s)$ で示せば，比例の項の荷重を αh，積分の項の荷重を α，中心周波数を f_0 とし，積分および位相の初期値を無視して，

$$s\Phi_0(s) = \frac{f_0}{s} - \frac{\alpha}{s}(\Phi_0(s) - \Phi_i(s)) - \alpha h(\Phi_0(s) - \Phi_i(s)) \qquad (14)$$

となるから系は，

$$\left.\begin{aligned} \omega_n &= \sqrt{\alpha} \\ \xi &= \alpha h/2\sqrt{\alpha} \end{aligned}\right\} \qquad (15)$$

なるパラメータを持つ二次の過渡応答を示すことになる．このように積分＋比例制御方式で制動力を与える方式を比例分制動方式と呼ぶ．

3. 一般化された相互同期方式の分類

相互同期系の各局におけるクロック源は 2. で述べた各要素，すなわち可変遅延回路，位相比較器，積分器，可変周波数発振器からなる．可変周波数発振器を含む部分を周波数同期装置（FSU）と呼び，可変遅延回路を含む部分を前置位相同期装置（PPSU）と呼ぶ．もし FSU が積分器を含めば，その方式を FSU 積分制御（FSU-I）方式と呼び，もし可変周波数発振器を位相比較器の出力に比例して制御すれば，その方式を FSU 比例制御（FSU-P）方式と呼ぶ．システムの中で，異なる2局の位相比較器の出力の差あるいは和を用いて可変遅延回路あるいは可変周波数発振器を制御するには，このための位相比較器出力情報を局間で伝送しなければならない．このような制御方式を伝送制御方式（T）と呼び，このような各装置の制御を自局の位相比較器の出力のみで行なう方式を無伝送制御方式（NT）と呼ぶ．

FSU 積分制御方式を用いたシステムでは，その系の制動力を与えるのに PPSU を用いてもよいし，比例分制動を行なってもよい．PPSU 無伝送制御方式を

用いたシステムではシステムの制動力を PPSU の動作で与えることができる．これに対して PPSU 伝送制御方式を用いたシステムでは制動力は PPSU の動作で与えてもよいし，比例分制動方式を用いてもよい．FSU を比例制御する方式では，制動力に関して特に考慮を払う必要はない．

したがって相互同期系は表 1 に示すように，FSU 比例制動方式による四つの方式と，FSU 積分制御方式による六つの方式に分類できる．もちろんこれらの分類の中間に，たとえば PPSU 制動と比例分制動を合わせ持つ方式などの種々の方式を考えることができるが，これらの構成を持つ方式の特徴は以上 10 に分類された方式の特性から類推できる．

表 1　相互同期方式の分類

分類	FSU P/I	FSU-NT/T	PPSU	制動方式	図面番号
1	P	NT	PPSUなし	——	1
2			PPSUあり	——	2
3		T	PPSUなし	——	3
4			PPSUあり	——	
5	I	NT	NT	——	4
6			T	PPSU制動	6
7				比例分制動	7
8		T	NT	——	
9			T	PPSU制動	
10				比例分制動	

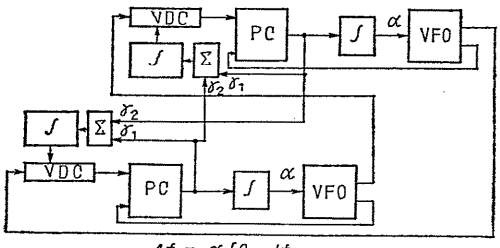


図 6　FSU-I-NT, PPSU-T, PPSU 制動方式（分類 6）
Fig. 6—FSU-I-NT, PPSU-T, PPSU damping scheme (class 6).

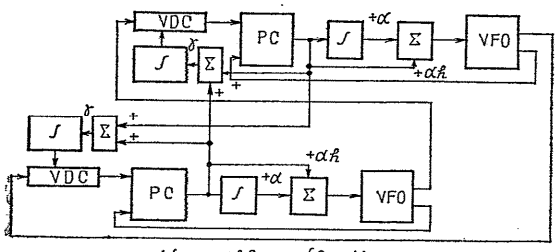


図 7　FSU-I-NT, PPSU-T 比例分制動方式（分類 7）
Fig. 7—FSU-I-NT, PPSU-T proportional damping scheme (class 7).

分類 1 の方式は図 1 に，分類 2, 3, 5 の方式はそれぞれ図 2, 3, 4 に示してあるが，分類 6, 7 の方式を図 6, 7 に示した．他の 4 方式についてはこれらの組み合わせから明らかである．

4. FSU 積分制御系の特性解析

表 1 に示した 10 の方式のうち FSU 比例制御の基本方式については，すでにその特性の解析結果を報告した[2]．その方式に伝送制御あるいは PPSU を付加した方式の特性は **2.** に示した各制御方式の特性から明らかであると思われるので，FSU-I 以下の方式についてその特性の解析を示すことにする．

4.1 各方式の基本方程式

（i）FSU-I-NT, PPSU-NT 方式の解析

この方式では n 局からなる系において，

$f_s d_{ij}-k_{ij}=\mathfrak{G}_{ij}$ とすれば，

$$\theta_{ij}(t)=\varphi_j(t)-\varphi_i(t)-\gamma_{ij}\int\theta_{ij}(t)\,dt+\mathfrak{G}_{ij} \tag{16}$$

$$\frac{d}{dt}\varphi_j(t)=f_{0j}-\sum_{i=1}^{n}\alpha_{ij}\int\theta_{ij}(t)\,dt \tag{17}$$

ただし $\alpha_{jj}=0,\ \gamma_{jj}=0,\ \mathfrak{G}_{jj}=0$

両式をラプラス変換すれば，積分器出力の初期値を無視して，

$$\left\{1+\sum_{i=1}^{n}\frac{\alpha_{ij}}{s(s+\gamma_{ij})}\right\}\Phi_j(s)-\sum_{i=1}^{n}\frac{\alpha_{ij}}{s(s+\gamma_{ij})}\Phi_i(s)$$
$$=\frac{\phi_j(0)}{s}+\frac{f_{0j}}{s^2}-\sum_{i=1}^{n}\frac{\alpha_{ij}\mathfrak{G}_{ij}}{s^2(s+\gamma_{ij})} \tag{18}$$
$$(j=1,2,\cdots,n)$$

したがって，

$$\sum_{i=1}^{n}\frac{\alpha_{ij}}{s(s+\gamma_{ij})}=\beta_{jj}(s)$$
$$-\frac{\alpha_{ij}}{s(s+\gamma_{ij})}=\beta_{ij}(s)\quad(i\neq j) \tag{19}$$

と置けば，連立方程式 (18) は，

$$\begin{bmatrix}1+\beta_{11}(s) & \beta_{21}(s)\cdots\cdots\beta_{n1}(s)\\ \beta_{12}(s) & 1+\beta_{22}(s)\cdots\beta_{n2}(s)\\ \vdots & \vdots\\ \beta_{1n}(s) & \beta_{2n}(s)\cdots 1+\beta_{nn}(s)\end{bmatrix}\begin{bmatrix}\Phi_1(s)\\ \Phi_2(s)\\ \vdots\\ \Phi_n(s)\end{bmatrix}=\begin{bmatrix}X_1(s)\\ X_2(s)\\ \vdots\\ X_n(s)\end{bmatrix} \tag{20}$$

ただし，

$$X_j(s)=\frac{\phi_j(0)}{s}+\frac{f_{0j}}{s^2}-\sum_{i=1}^{n}\frac{\alpha_{ij}\mathfrak{G}_{ij}}{s^2(s+\gamma_{ij})} \tag{21}$$

となる．

これ以外の 5 種の FSU 積分制御方式についても同

様に解析を行なうことができる．

(ii) FSU-I-NT, PPSU-T, PPSU 制動方式

この方式では，局 $\#j$ の入力における局 $\#i$ からの入力に設けられた前置位相同期装置を制御する自局の係数，他局の係数が各局においていずれも，それぞれ γ_1, γ_2 に等しいとすれば，式 (20) において，

$$\beta_{ij}(s)=\frac{-\alpha_{ij}}{s(s+\gamma_1-\gamma_2)} \quad (i\neq j)$$

$$X_j(s)=\frac{\phi_j(0)}{s}+\frac{f_{0z}}{s^2}-\sum_{i=1}^{n}\alpha_{ij}\frac{\frac{\mathfrak{G}_{ij}}{s}\left(1+\frac{\gamma_1}{s}\right)-\frac{\mathfrak{G}_{ij}}{s}\frac{\gamma_2}{s}}{s^2+2s\gamma_1+\gamma_1^2-\gamma_2^2} \tag{22}$$

と置けば，同様にして各種の特性を求めることができる．

以下同様に各方式の $\beta_{ij}(s)$, $X_j(s)$ を示せば，つぎのようになる．

(iii) FSU-I-NT, PPSU-T 比例分制動方式

$$\beta_{ij}(s)=-\alpha_{ij}\frac{1+sh}{s^2} \quad (i\neq j)$$

$$X_j(s)=\frac{f_{0j}}{s^2}+\frac{\phi_j(0)}{s}+\sum_{i=1}^{n}\alpha_{ij}\frac{\frac{\mathfrak{G}_{ij}}{s}\{s+\gamma\}-\frac{\mathfrak{G}_{ji}}{s}\gamma}{s^2(s+2\gamma)}(1+sh) \tag{23}$$

(iv) FSU-I-T, PPSU-NT 方式

$$\beta_{ij}(s)=\frac{-2\alpha_{ij}}{s(s+\gamma)} \quad (i\neq j)$$

$$X_j(s)=\frac{f_{0j}}{s^2}+\frac{\phi_j(0)}{s}-\sum_{i=1}^{n}\alpha_{ij}\frac{\mathfrak{G}_{ij}-\mathfrak{G}_{ji}}{s^2(s+\gamma)} \tag{24}$$

(v) FSU-I-T, PPSU-T, PPSU 制動方式

$$\beta_{ij}(s)=\frac{-2\alpha_{ij}}{s+\gamma_1-\gamma_2} \quad (i\neq j)$$

$$X_j(s)=\frac{f_{0j}}{s^2}+\frac{\phi_j(0)}{s}-\sum_{i=1}^{n}\alpha_{ij}\frac{\mathfrak{G}_{ij}-\mathfrak{G}_{ji}}{s^2(s+\gamma_1+\gamma_2)} \tag{25}$$

(vi) FSU-I-T, PPSU-T 比例分制動方式

$$\beta_{ij}(s)=\frac{-2\alpha_{ij}(1+sh)}{s^2} \quad (i\neq j)$$

$$X_j(s)=\frac{f_{0j}}{s^2}+\frac{\phi_j(0)}{s}-\sum_{i=1}^{n}\alpha_{ij}\frac{(\mathfrak{G}_{ij}-\mathfrak{G}_{ji})(1+sh)}{s^2(s+2\gamma)} \tag{26}$$

これらの式を用いた式 (20) の解をいくつかの例について示せば，つぎのようになる．

4.2 FSU-I-NT, PPSU-NT 方式の解

各局の接続が任意であるときに式 (20) を一般的に解くことは困難であるが，すべての局間接続における α_{ij}, γ_{ij} が等しければ，式 (20) を解くことができ，各局間の位相差 $\Phi_i(s)-\Phi_j(s)$ は，

$$\Phi_i(s)-\Phi_j(s)=\frac{X_i(s)-X_j(s)}{1-n\beta(s)} \tag{27}$$

となる．これを用いて，局入力における位相差を求めれば，

$$\theta_{pq}(t)=\{\phi_q(0)-\phi_p(0)\}A(t)+(f_{0q}-f_{0p})B(t)-\frac{1}{n}\sum_{i=1}^{n}(\mathfrak{G}_{iq}-\mathfrak{G}_{ip})(e^{-\gamma t}-A(t))+\mathfrak{G}_{pq}e^{-\gamma t} \tag{28}$$

となる．ここで，

$$A(t)=\varepsilon^{-\xi\omega_n t}\left\{\cosh\omega_n\sqrt{\xi^2-1}\,t-\frac{\xi}{\sqrt{\xi_2-1}}\sinh\omega_n\sqrt{\xi^2-1}\,t\right\}$$

$$B(t)=\frac{\varepsilon^{-\xi\omega_n t}}{\omega_n\sqrt{\xi^2-1}}\sinh\omega_n\sqrt{\xi^2-1}\,t$$

$$C(t)=1-A(t)-2\xi\omega_n B(t) \tag{29}$$

$$\omega_n=\sqrt{n\alpha} \tag{30}$$

$$\xi=\gamma/2\sqrt{n\alpha} \tag{30}$$

である．このように FSU-I 制御系では式 (30) に示す自然周波数 ω_n と制御係数を持つ二次の過渡応答が生ずる．

また，正規化された位相で表わされた，前置位相同期装置の遅延時間 $\delta_{pq}(t)$ を求めると，

$$\delta_{pq}(t)=\{\phi_q(0)-\phi_p(0)\}\gamma B(t)+\frac{\gamma}{n\alpha}(f_{0q}-f_{0p})\cdot\{1-A(t)-\gamma B(t)\}-\frac{1}{n}\sum_{i=1}^{n}(\mathfrak{G}_{iq}-\mathfrak{G}_{ip})\cdot[1-e^{-\gamma t}-\gamma B(t)]+\mathfrak{G}_{pq}[1-e^{-\gamma t}] \tag{31}$$

$$f_q(t)=\{\sum_{i=1}^{n}\phi_i(0)-n\phi_q(0)\}B(t)+f_{0q}\{1-C(t)\}+\frac{1}{n}\sum_{j=1}^{n}f_{0j}C(t)-\sum_{i=1}^{n}\mathfrak{G}_{iq}\alpha B(t)-\sum_{i=1}^{n}\sum_{j=1}^{n}\mathfrak{G}_{ij}\cdot\frac{\alpha}{n\gamma}\cdot C(t) \tag{32}$$

となる．

したがって，この方式では定常状態では，位相比較器入力の位相差および前置位相同期装置の遅延時間は

$$\theta_{pq}=0 \tag{33}$$

$$\delta_{pq}=\frac{\gamma}{n\alpha}(f_{0q}-f_{0p})-\frac{1}{n}\{(\sum_{i=1}^{n}\mathfrak{G}_{iq}-\sum_{i=1}^{n}\mathfrak{G}_{ip})+\mathfrak{G}_{pq}\} \tag{34}$$

となり，系統周波数は，

$$f_s=\frac{1}{n}\sum_{j=1}^{n}f_{0j}-\sum_{i=1}^{n}\sum_{j=1}^{n}\mathfrak{G}_{ij}\cdot\frac{\alpha}{n\gamma} \tag{35}$$

となる．したがってこの方式では，おのおのの前置位相同期装置が，局間遅延の変動を補正するから，各位相比較器の入力位相差は定常的に零となっているが，局間遅延の変動に応じて系統周波数の変動がみられる．

4.3 FSU-I-T, PPSU-T 比例分制動方式の解

比例分制動方式の例として 10 方式の中で最も複雑なものである．分類 10 の方式の例を示す．この方式については式(27) が成立するから，これを用いて解くことができ，

$$\omega_n=\sqrt{2n\alpha},\ \xi=2n\alpha h/\sqrt{2n\alpha} \tag{36}$$

として，

$$\theta_{pq}=\{\phi_q(0)-\phi_p(0)\}A(t)+\frac{1}{2}\mathfrak{G}_{pq}(1+\varepsilon^{-2\gamma t})$$
$$-\frac{1}{2}\mathfrak{G}_{pq}(1-\varepsilon^{-2\gamma t})-\frac{1}{2n}\{\sum_{i=1}^{n}(\mathfrak{G}_{iq}-\mathfrak{G}_{qi})$$
$$-\sum_{i=1}^{n}(\mathfrak{G}_{ip}-\mathfrak{G}_{pi})\}\{1-A(t)\} \tag{37}$$

$$\delta_{pq}(t)=\frac{1}{2}(\mathfrak{G}_{pq}+\mathfrak{G}_{qp})\{1-\varepsilon^{-2\gamma t}\} \tag{38}$$

$$f_q(t)=\left\{\frac{\alpha}{n}\sum_{i=1}^{n}\phi_i(0)-2n\alpha\phi_q(0)\right\}\{hA(t)+B(t)\}$$
$$+f_{0q}A(t)+\frac{1}{n}\sum_{i=1}^{n}f_{0i}\{1-A(t)\}$$
$$-\alpha\sum_{i=1}^{n}\{\mathfrak{G}_{iq}-\mathfrak{G}_{qi}\}\{hA(t)+B(t)\} \tag{39}$$

したがって，定常状態では，

$$\theta_{pq}=\frac{1}{2}(\mathfrak{G}_{pq}-\mathfrak{G}_{qp})$$
$$-\frac{1}{2n}\{\sum_{i=1}^{n}(\mathfrak{G}_{iq}-\mathfrak{G}_{qi})-\sum_{i=1}^{n}(\mathfrak{G}_{ip}-\mathfrak{G}_{pi})\} \tag{40}$$

$$\delta_{pq}=\frac{1}{2}(\mathfrak{G}_{pq}-\mathfrak{G}_{qp}) \tag{41}$$

$$f_q=\frac{1}{n}\sum_{i=1}^{n}f_{0i} \tag{42}$$

となる．したがって位相比較器の出力には定常的な値が残るが，相互に接続された2局の位相比較器の出力の和 $(\theta_{pq}+\theta_{qp})$ は零となり，また $\sum_{i=1}^{n}\theta_{ik}$ もまた零となる．したがって積分器の入力には，各位相比較器の出力の和を接続すればよい．

5. 各方式の特性とその比較

各方式に対するシミュレーションを周波数 10^4 Hz の2局の系について行なった[10]．FSU-P 制御方式の場合には各方式とも時定数1ミリ秒の一次の応答を持つパラメータを選択し，FSU-I 制御方式の場合には $\omega_n=10^3$ Hz, $\xi=1$ の二次の応答を持つパラメータを選択した．

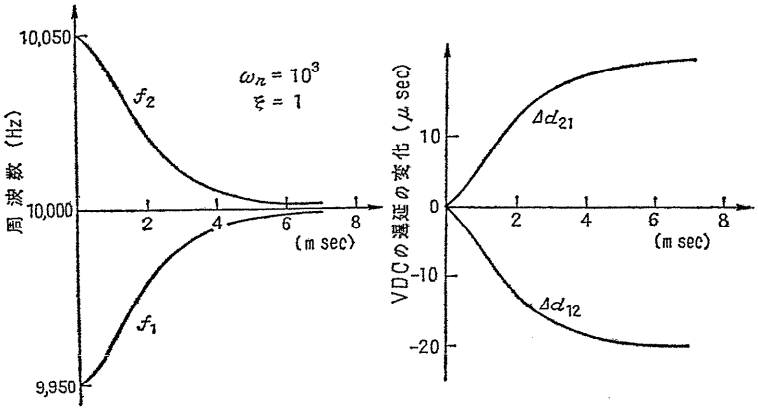


図 8 FSU-I, PPSU 制動の4方式（分類 5, 6, 8, 9）の周波数変動に対する応答

Fig. 8—Response of 4 schemes employing FSU-I, PPSU damping scheme (5), (6), (8) and (9) to step change of frequency.

図8は FSU-I 制御の PPSU 制動の4方式において，系を構成する2局の初期周波数が 100 Hz 異なった場合に両局の周波数が変化して同期にいたる過渡状態を示している．図には同時に PPSU の可変遅延回路の応答が示されており，それぞれ初期の状態から増減した遅延時間をとるようになることを示している．本論文のいずれの方式でも各局の周波数の，周波数変動，位相変動に対する過渡特性はかなり自由に設計することができ，FSU-P 制御方式でも，すでに報告したように[11]系に適切なフィルタを設ければ，ほぼ同様の応答を得ることができる．しかし FSU-P 制御方式，あるいは FSU-I 制御方式でも比例分制動方式の系では，局の周波数変動や位相変動で可変遅延回路の遅延時間が変化することはないのに対して，FSU-I 制御方式の PPSU 制動方式の場合には，このような

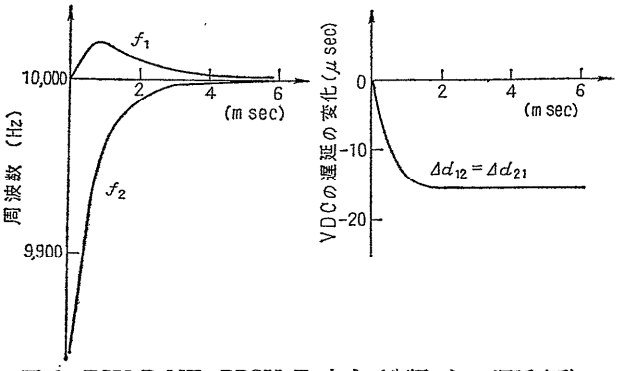


図 9 FSU-P-NT, PPSU-T 方式（分類2）の遅延変動に対する応答遅延は d_{12} が 30 μ/sec 増加

Fig. 9—Response of FSU-P-NT, PPSU-T scheme (class 2) to step change of delay. d_{12} changes 30 μ/sec.

場合にも可遅変延回路が動作するため，これらの回路にかなり高速の応答が要求され，またその長さも遅延変動を補正するに要するものより長くなければならないことになる．しかし FSU-I 制御方式では周波数，位相が変動しても，位相比較器に定常的な出力が生ずることはなく，この点では FSU-P 方式より有利である．

表 2 各方式の比較

分類	各局の入力での定常的位相差の原因	位相同期装置の容量	遅延変動によって定常的周波数偏差が生ずるか	PPSU の容量	位相比較器出力伝送の必要の有無
1	周波数変動 遅延変動	大	有	不要	無
2	周波数変動	小	無	小	有
3	周波数変動 遅延変動	大	無	不要	有
4	周波数変動	小	無	小	有
5	なし	不要	有	大	無
6	なし	不要	有	大	有
7	遅延変動 (小)	小	有	小	有
8	なし	不要	無	大	有
9	なし	不要	無	大	有
10	遅延変動 (小)	小	無	小	有

一方遅延変動に対する応答はこれら 10 方式のおのおのですべて異なっている．FSU-P 方式では，分類 1 の方式を除き，定常的な系統周波数は遅延変動の影響を受けることはない．たとえば図 9 は分類 2 の方式の遅延変動に対する応答を示している．遅延変動は局 #1 から局 #2 にいたる遅延が 30 μ/sec だけステップ的に変動したものとして与えてあるが，これに対して両局の r が等しければ時間で示した両局の PPSU の遅延時間 Δd_{ij} は，指数関数的に減少してこれを平等に分担する．これに対して FSU-I 制御方式の PPSU 制動方式の各系では，PPSU の遅延は指数関数的な単純な変化ではなく，その過渡状態において終極値より大きな遅延の変動を生ずる．この場合でも両局の r が等しければ，遅延変動は両局の PPSU で平等に分担される．図 10 は分類の方式について示したもので，このときには系統周波数も変化している．この系統周波数の変動は FSU を伝送制御することによって除かれる．

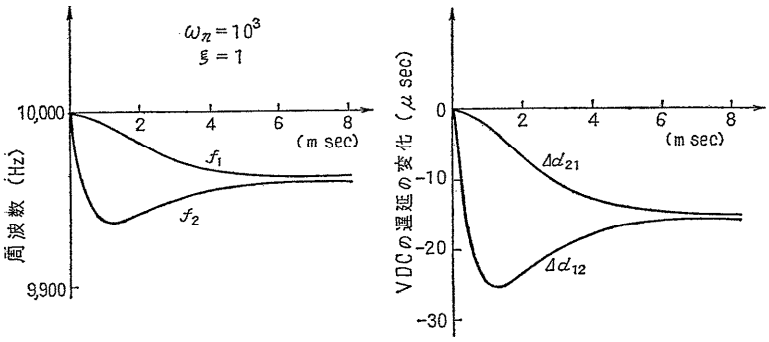


図 10 FSU-I-NT, PPSU-NT 方式（分類 5）遅延変動に対する応答 d_{12} が 30 μ/sec 増加
Fig. 10—Response of FSU-I-NT, PPSU-NT scheme (class 5) to step change of delay. d_{12} changes 30 μ/sec.

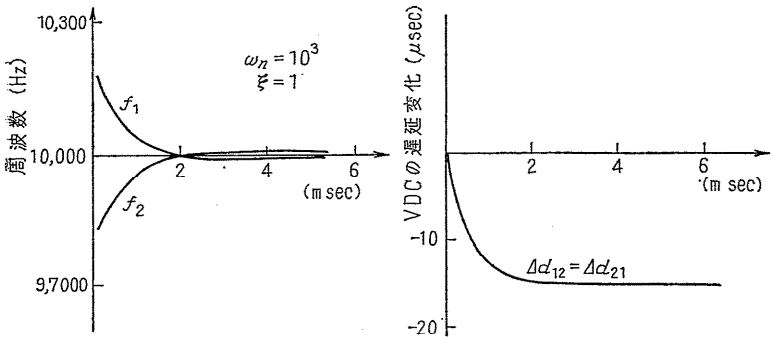


図 11 FSU-I-T, PPSU-NT, 比例分制動方式（分類 10）の遅延変動に対する応答 d_{12} が 30 μ/sec 増加
Fig. 11—Response of FSU-I-T, PPSU-T proportional damping scheme (class 10) to step change of delay. d_{12} changes 30 μ/sec.

図 11 は FSU-T, PPSU-T 比例分制動方式（分類 10）の場合の遅延変動に対する応答を示している．この方式では系統周波数に遅延変動の影響はなく，PPSU の遅延も指数関数的に応答するが，**4.3** に述べたように遅延変動の結果各位相比較器の出力には相互に接続された 2 局間の両方向の遅延の差に比例した値が残留する．

表 2 はこれらの各方式の性質をまとめたものである．各局の入力での定常的位相差の原因には周波数変動と遅延変動があり得るが，各方式でその両方あるいは一方を除くことができる．その一方が除かれれば位相同期装置の遅延補償メモリの容量は小さくなり，両者を除くことができれば，原理的には位相同期装置は不要となる．つぎに系統周波数が遅延変動の影響を受けるかどうかを示してある．PPSU の容量は PPSU が遅延変動のときしか動作しないものでは小さくてよいことになる．

これらを総合的に考えれば，構成が簡単であるという点で，分類 1 の基本相互同期方式がすぐれていると思われる．こ

れに対して，局間の遅延変動が大きく，その変動がゆるやかな場合には分類2の方式がすぐれている．局間遅延の変動より，各局の中心周波数の変化による位相差が支配的であれば，FSU-I 方式が適している．この方式としては分類5の方式が最も簡単であるが，周波数の変動に際しても PPSU が動作するため，PPSU の遅延補償メモリの容量は遅延変動を処理するためだけのものより容量の大きいものとなる．これに対して，分類7あるいは 10 の方式では PPSU は分類2の場合と同様の容量のものでよく，特に分類 10 の方式がすぐれている．

7. むすび

相互同期方式としては従来本論文の分類 1, 2, 3 に示す各方式の研究が行なわれていたが，本論文では相互同期方式に積分制御方式を採り入れる場合を含めて方式の制御に用い得る情報を考察し，これを用いたさらに7種類の相互同期方式を考察し，その特性を解析するとともに，シミュレーションによってその実現性を明らかにした．これらのうちで従来広く研究されている基本相互同期方式は簡単で特性もすぐれているが，局間の遅延変動が大きく，その変動がゆるやかな場合には分類2の方式が，また各局の中心周波数の変化によって生ずる位相差が大きい場合には分類 10 の方式にみられる改良を施すことが有利であると考えられる．なお本論文のシミュレーションは東京大学大型計算機センタの HITAC 5020 E を用いて行なった．

文　献

(1) 藤崎，斉藤，猪瀬："相互同期方式の特性について"，昭 39 信学全大，S. 7-6.

(2) 猪瀬，藤崎，斉藤："相互同期系の理論"，信学誌，**49**，4，p. 755 (昭 41-04).

(3) 大和，小野："PCM 交換における平衡相互同期"，通研実報，**16**，12，p. 2853 (昭 43-12).

(4) A. Gersho and B.J. Karafin : "Mutual synchronization of geographically separated oscillators", Bell Syst. tech. J., **45**, 10, p. 1689 (Dec. 1966).

(5) M. Karnaugh : "A model for the organic synchronization of communication systems", Bell Syst. tech. J., **43**, 10, p. 1075 (Dec. 1966).

(6) N. West : "Synchronous digit switching in highly interconnected communication network", Proc. Instn Elect. Engrs, **114**, 10, p. 1378 (Oct. 1967).

(7) M.R. Miller : "Feasibility studies of synchronized-oscillator system for p.c.m. telephone network", Proc. Instn Elect. Engrs, **116**, 7, p. 1135 (July 1969).

(8) 猪瀬，藤崎，斉藤："相互同期系の構成"，信学誌，**49**，11 p. 2095 (昭 41-11).

(9) 斉藤，杉原，藤崎，猪瀬："相互同期系における局間遅延偏差の補正方式"，昭 42 信学全大，1154.

(10) 斉藤，猪瀬："一般化された相互同期系の概念"，信学会交換研資，SE 68-8 (昭 43-05).

(11) 斉藤，安田："相互同期系の動特性"，信学論 (A)，**51-A**，4，p. 137 (昭 43-04).

(昭和 44 年 11 月 22 日受付)

３．　大容量パケット交換方式

○　High capacity packet switching system by means of Multi-Microprocessors
○　Evaluation of alternatives for intra-network inplementation of X.25 compatible packet switching networks
○　A proposal of high capacity switching fabric for Tbps-level IP Router
○　多段接続網を用いた大容量ＩＰデータグラムスイッチング方式

HIGH CAPACITY PACKET SWITCHING SYSTEM BY MEANS OF MULTI-MICROPROCESSORS

T. SAITO, H. INOSE, M. WADA and H. SHIBAGAKI
Department of Electrical Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113 Japan

Versatility of packet switching principle gave rise to a new possibility to realize a integrated communication service for various information including data, voice and image. However, since the complexity of packet switching made it difficult to realize a high capacity switch, the throughput of the switch ever realized is as small as 0.5 Mbit/sec, in contrast to usual PCM switch having the throughput of 100 Mbit/sec. To use as the switch for more general information, the capacity of a packet switching system should be expanded. This paper proposes a principle of a packet switching system based on multi-micro processor. A number of processors are interconnected through a common memory. Each of processors has its own local memory and to facilitate interprocessor communication, address space of the common memory is designed to be continueing from each local memory. An experimental packet switch with skeltonized scale has been constructed and the results of evaluation for the experimental system indicate the realizability of high capacity packet switch based on the principle.

1. INTRODUCTION

Variety of ideas has been presented to realize combined voice and data communication servece in future. Proposal of integrated service digital network is based on circuit switched system. Combined voice and data communication systems based on packet switching principle has also been proposed.

Comparing the combined voice and data service based on these two principles, the packet switched system generally provides more versatile services.

In conventional data packet switching system for data, however, the highest data speed is 48 kb/sec and total amount of data traffic is much lower than that of telephone traffic. Therefore, traffic handling capacity of a conventional data packet system is too low to handle combined voice and data traffic. To realize a combined voice and data packet switching system therefore a new switching concept is needed to enhance the traffic handling capacity of packet switching systems.

Although packet switching systems are characterized by the use of store and forward concept, the same concept has been basically employed by PCM switching systems in their time slot interchanges. The major differences of the latter to the former are such that the store and forward interval is restricted within the Niquist interval and that the amount of information to be stored is much less. Another major difference to be pointed out exists in the control mechanism of the switching system. In the PCM switching system, the control information to determine the destination of buffered information is stored in the switching system prior to the connection set up. This may be called "preset up control". On the other hand in the packet switching system, the destination of the packet is determined by the information carried in the header of the packet itself. This control mechanism may be called "header driven control".

In preset up control system, processing for switching is performed by a central control which is used only at set up and clear down phase a communication. In the remainder of the communication period, information can be handled by the speech path which is operated independent from the central control. This is the reason why high capacity can be achieved by preset up controlled system.

In packet switching systems also, if control subsystems and information path are able to be separated by providing a number of speech path control circuits, a high-capacity packet switching system may be realized.

The packet switching system described in this paper is essentially X.25 compatible. Types of packets are classified into two categories one is to be handled by the main processor and the other is handled by input and output processors. The packets of the former type includes the packets involved in the setup and release of virtual circuits. Other packets like data packet receive ready packet etc. are handled by input and output processors without intervention of the main processors.

2. PRINCIPLE FOR CAPACITY ENHANCEMENT

Figure 1 shows the outline of the high-capacity packet switching system. An input output (IO) processor module are provided for each of lines which are assumed to be PCM repeatered lines having the transmission rate of 1.544 MHz. The total number of IO processors depends on the percentage of packet to be handled by the main processor module. In typical case, the expected number of possible IO processors is about 50. The main processor module is used as the central control of the system.

An IO processor module is composed of a microprocessor, a local memory and a communication control LSI which is used for HDLC processing.

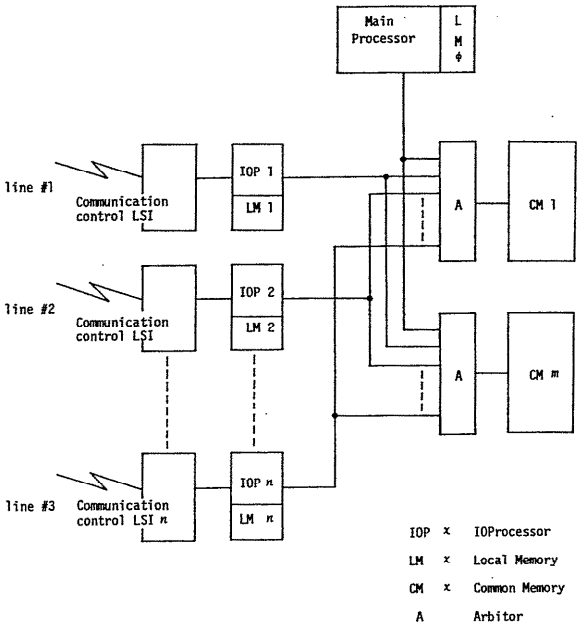


Fig. 1 Overall configuration of high-capacity packet switching system

Table - 1 Assignment of function to each of module

Information path		Control path
IO processor module		Main processor module
Communication control LSI	Processor	
HDLC procedure * Flag detection * Insertion and deletion of "0" * Check and compute FCS * Detection of underrun * Compute number of byte	HDLC procedure * Generate address and command * Check command part * Transmission of ACK and NAK * Monitoring of timers Data packet handling procedure * Determine packet types * Flow control * Check packet sequence	Virtual circuit handling procedure * Analysis of call request * Analysis of call termination * Establish and release of virtual circuit * Establish routing table

The main processor module is composed of a micro-processor and a local memory. The processors share common memory modules each of which has a number of ports. The access to the common memory is controlled by an arbiter which schedules the access on first-come-first-served basis. A local memory is used to store programs, parameter tables as well as temporary data for each of the processor. The common memory is used for packet buffering and inter-processor communication.

Any processor can access to all the address space of the common memory to read the information. A processor, however, is permitted to write only in the assigned address space for the processor. For this purpose the address space of the common memory is divided into portions each of which is assigned to an IO processors for writing. By this mechanism, the control processor for the common memory and the overhead by the control processor is eliminated. Since the address space of the common memory is divided, the efficiency of the common memory usage is degraded, however, in view of the drastic cost reduction of the LSI memory, the problem is not essential. The assignment of switching functions to the IO processors and the main processors is summerized in Table 1. Since the IO processors are provided for each of lines, functions of data path are assigned to IO processors. The functions which needs the status information of the whole switching system or other lines are assigned to the main processor. The computation of the frame check sequence and the framing pattern processing for the HDLC procedure is to be performed by the communication control LSI.

The processing procedure for a packet is as follows. The packet received by a communication control LSI is transferred to an input buffer located in a common memory. During the reception, the frame check sequence is verified by the communication control LSI. If an error is found, the IO processor discards the information received at the common memory and start the usual recovery procedure. If the HDLC frame is found to be correct, the address and the command of the HDLC frame are checked. After this procedure, the packet type is identified. If the packet is a data packet, the outgoing route is determined by the link table established at the call establishment phase and the output request is fed to the IO processor which handled the line of the outgoing route. When the IO processor of the outgoing route finds the request, the processor generates the HDLC frame which includes the packet stored in the common memory and transmits the frame through the communication control LSI. At the transmission phase, the content of the packet is directly read out from the address space at which the IO processor of the input side has written the packet.

On the other hand, if the packet is related to setup or release of a virtual circuit, the IO processor informs the arrival of the packet to the main processor to establish the link table.

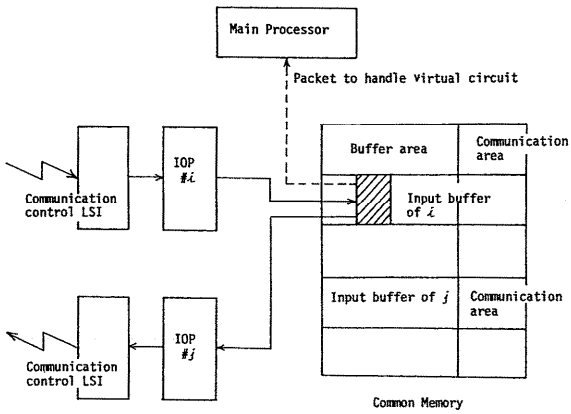


Fig. 2 Flow of packet from IOP #i to IOP #j

The outline of the flow of packet handling is summerized in Fig. 2.

3. ESTIMATION OF TRAFFIC HANDLING CAPABILITY

The traffic handling capacity of the system is expected to be limitted by the following factors.

1. processing capability of an IO prossor.
2. processing capability of the main processor.
3. access contention of the common memories.

The average processing step for a packet is reported to be about 700 steps in typical packet switching system. Viewing from the results, in the present system using micro-processors, required processing step needed for a packet is less than 1000 steps in average. To handle 1.5 Mbit/sec PCM line, if average packet size is 1000 bits, an IO processor module having the processing capability of 1.5 M step/sec is needed.

The traffic applied to the main processor depends on mixture ratio of packet which require attention by the main processor. If the majority of traffic is to carry voice or facsimile, the mixture ratio of packet to be handled by the main processor is small enough. However, if the traffic of interactive data or traffic for X.25-type fast select packets is dominant, the total throughput of the system will be limited by the throughput of the main processor.

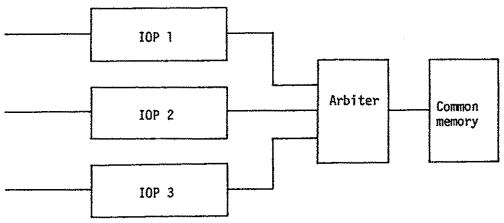


Fig. 3 Model for the analysis of memory congestion

As has been described previously, the common memories are accessed on first-come-first-served basis. The access mechanism can be evaluated by

1. average waiting time before served by the memory.
2. probability of reception overrun and transmission underrun.

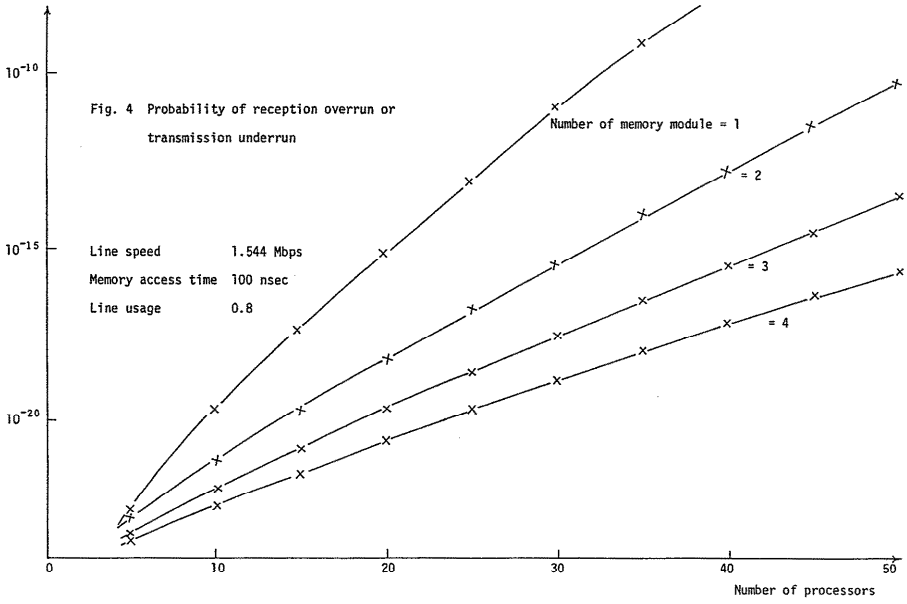
The reception overrun occurrs in case that, before completion of writing of a data byte, next data byte arrives to be stored. If this case happened, since the access mechanism is still involved in the processing of prior byte, the next byte may be lost. The transmission overrun occurs in case that readout of byte from the memory is delayed to cause interruption of transmission because of the congestion of the memory. If the average waiting time is too long, processing capability of the processors are degraded.

The memory access mechanism can be evaluated using the simplified model as is shown in Fig. 3. In this model, processors are represented by IO processors. Each of n IO processors requests access to the common memory independently. Each of processors waits until the request is permitted.

Assuming the Poisson arrival of request and the negative exponential distribution of the memory service time, the model of the waiting system can be expressed by M/M/1/n.

The average service time corresponds to the memory access time. Since the memory access time is constant, the evaluation by M/M/1/n model will give safety side results. In this evaluation, the memory access time is assumed to be 100 nsec which is not too difficult to attain in near future.

Access to the common memory is requested for packet reception and transmission as well as



inter-processor communication. Access for inter-processor communication is assumed to be 10 times for a packet.

If a plurality of common memory units are provided, access rate to a memory unit will be reduced. Denoting the number of memory unit by m, access rate to one memory will be simply λ/m.

Figure 4 shows the reception overrun and transmission underrun probabilities respectively when the number of common memory modules is 1 through 4. The transmission speed is assumed to be 1.544 MHz and reception overrun probabilitiy is defined as the probability that waiting time is longer than one byte duration, i.e., 5.18 μsec.

If the number of memory module is 3, the average waiting time is as small as 100 nsec and the reception overrun probability is less than 10^{-15}. This means that the structure of the packet switching system using the common memory for inter-processor communication is feasible from the view point of memory access control.

4. HARDWARE OF THE HIGH-CAPACITY PACKET SWITCHING SYSTEM

For the feasibility demonstration of the high-capacity packet switching system, and experimental system has been designed and constructed based on the principle described in 3.

Figure 5 shows the experimental system. This system includes a skeletonized high-capacity

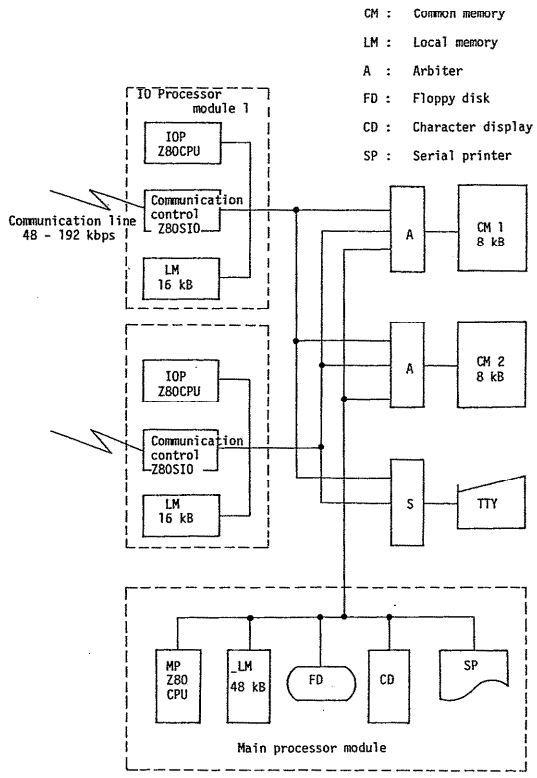


Fig. 5 Hardware configuration of experimental system

packet switching system and some development aids including floppy disk drives, a character display and a serial printer.

The high-capacity packet switching system is composed of two IO processor modules, one main processor and two common memories. in this configuration, three processors may request access to the common memories, so that the feasibility study of the access control to the common memory was made possible which is considered to be the most critical problem in the system design. Z80s are used as the micro-processors as an experiment.

The transmission and the reception is controlled by the Z80SIO and Z80DMA as follows. When a data pattern which is not identical to the HDLC flag pattern (01111110) is received at the receive line of the SIO, the reception start interrupt is generated, which enables the reception operation by the DMA. The received data is transferred from the buffer register in the SIO to the input buffering area in the common memory through the DMA control on byte-by-byte basis. After the completion of data reception, which is identified by the detection of a flag pattern, the SIO will check the frame check sequence which is found just before the framing pattern and SIO will interrupt the IO processor. The interrupted processor disables the DMA and confirms that the result of FCS check is correct and other errors was not found.

When a packet is to be transmitted from the common memory, the IO processor establishes the starting address and the byte count to be transmitted in the DMA. Then the processor enables the DMA. The data is transmitted to the line through the DMA and the SIO byte-by-byte. When the specified number of bytes has been transmitted, the frame check sequence is automatically generated by the SIO and transmitted to the outgoing line. The termination of transmission is noticed to the IO processor by means of an interruption. Then the IO processor disables the DMA.

The common memory module include the arbiter to control the access to the common memory, the request generators which generate the request to the arbiter and to the switches. The request generator and the switch are provided for each of processors which have access to the common memory.

The RAM used for the common memory of the experimental system is static RAM having the access time of 200 nsec. The clock of the common memory module is 4 MHz, which is independent from the processors.

Figure 6 shows the principle of the arbiter used for the common memory. If two requests are applied to the input of the S-R flip-flop, the output of the S-R flip-flop indicates which of the input is given earlier. The output of the S-R flip-flop is inverted and applied to J-K flip-flops.

If n inputs are applied, the required number of flip-flops is ${}_nC_2$. In the experimental

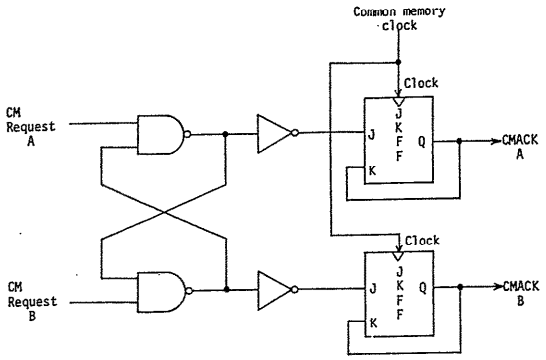


Fig. 6 Principle of the arbiter

system, since two IO processors and a main processor is provided, 3 S-R flip-flops are used for the arbiter.

Extending the principle to such cases as the number of IO processors is as many as 50, the arbiter circuit may become too complicated. In such a situation, tree type arbiter can be designed to reduce the number of required components.

When one output from the S-R flip-flop becomes logical "1", the corresponding J-K flip-flop is enabled. The output of the J-K flip-flop is applied to the processor which generated the request for the memory access permission. The signal is called CMACK. The CMACK resets the memeory request signal from the processor and the J-K flip-flop is also resets at the next clock. Thus the selection of the processor is performed in every 250 nsec cycle time of the common memory.

The main-processor module is composed of a CROMEMCO micro-computer system. The system is is also Z80 based system having 48 kbyte local memory and two floppy disk drives.

Total scale of hardware is 700 integrated circuit chips including LSI, which are mounted on 26 circuit boards.

5. SOFTWARE OF THE HIGH-CAPACITY PACKET SWITCHING SYSTEM

The software of the high-capacity packet switching system is composed of the HDLC procedure handling module, the data packet handling module, the virtual circuit handling module and the basic control module. The HDLC procedure handling module and the data packet handling module are assigned to the IO processors. The virtual circuit handling module is assigned to the main processor. All the program has been written by Z80 assembler.

The data packet handling module of the IO processor has the structure as is shown in Fig. 7. The request for data packet processing is applied from the HDLC procedure of the IO processor for packet reception of from other processors for packet transmission. The former request is queued at the processing queue to which request from timer is also applied. The packet is first analyzed by the packet analyzing program. If the packet is to be

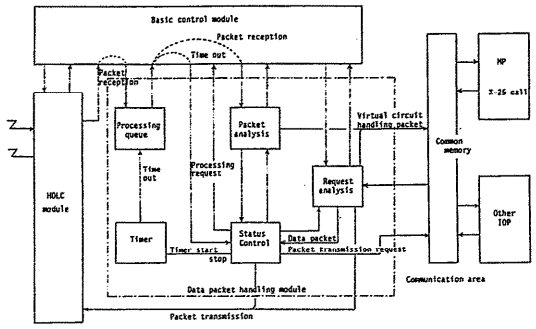


Fig. 7 Program of the IO processor

processed by the main processor, it is noticed to the main processor. The request from other processor is analyzed by the request analysis program. If it is the request for transmission, the request is passed to the HDLC module. If a packet routed through a different IO processor is acknowledged, the IO processor receives a buffer release request from the IO processor used for packet transmission. The buffer release request is passed to the buffer management program in the HDLC module to release the buffer.

Figure 8 shows the software structure of the main processor used for X.25 virtual circuit handling. When the IO processor #J finds a packet which is related to X.25 virtual circuit handling, the IO processor will give the request to main processor by writing into the communication area. The main-processor maintains the monitoring of the communication area in the common memory. If the request is found, the request analysis module of the main processor program will analyze the content of the request, the logical channel number and the originating IO processor identity. The result of the analysis is reported to the basic control module. The virtual circuit handling module is started by this request and the task to be excuted is determined using the IO processor number, logical channel number, content of the request and the present status of the virtual circuit.

For the control of a virtual circuit, status of all the logical channels of all the IO processors are needed. This table is called the link table which is stored in the common memory and shared by data packet handling module of IO processors.

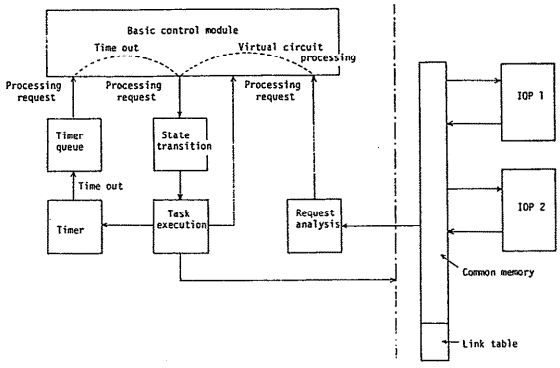


Fig. 8 Program of the main processor module

If some packets are to be transmitted to handle a virtual circuit, the order to transmit the packet is requested from the main processor to the IO processor.

6. EXPERIMENT FOR PERFORMANCE EVALUATION

To obtain the design parameters for an operational high-capacity packet switching system, performance of the experimental system has been evaluated.

6.1 Dynamic Steps for Packet Handling

The throughput of the packet switching system depends upon the kinds of packets transmitted through the network. If the number of data packets transmitted for each of virtual call is large, the loading to the main-processor which handles virtual circuits are reduced. If the acknowledgment is piggybacked on the data packet, the number of recive ready packet used for acknowledgment is reduced.

The experiment was performed in the case that data packets are continuously transmitted after the establishment of virtual circuit and all the HDLC frames are acknowledged by piggyback operation. The average number of dynamic steps for a packet reception was about 430 steps, and that of packet transmission was about 400 steps. In total, dynamic steps to process a packet was about 830 steps.

The average instruction exec ution time of Z80 is about 4.2μ sec when the clock cycle is 400 nsec. Since the CPU is not operational during DMA transfer, assuming that DMA transfer is performed once in every instruction, average execution time is about 4.5μ sec. Therefore the processing capability of one IO processor is about 270 packet/sec. If the average packet length is 1000 bits, the IO processor can handle a line of 270 kbit/sec. The required number of dynamic steps may increase by further development. Using faster microprocessors available in the market in near future, however, it is expected that one IO processor can easily handle a PCM line of 1.544 Mbit/sec.

6.2 Requirement for Memory Capacity

The memory capacity for each IO processor used in the experimental system was 10 kbytes for program memory and 4 kbytes for input and output buffer. If the number of IO processors is 50, the total memory capacity is 700 kbytes. The amount can be attained economically.

6.3 Requirement for Main Processor

Dynamic steps to set up and release a virtual circuit was 150 steps in the experimental system. To handle 50 IO processors by main processor, if identical processors are used for both purposes, capacity ballances when the number of data packets is 10 times larger than the number of virtual circuit controlling packets. If the data packets transmitted by a virtual call is smaller in number, more powerfull processors may be needed for the main processor.

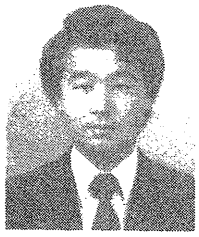
7 CONCLUSION

This paper described the experimental model and the results of the first phase of experiment on the high-capacity packet switching system. The objective of the experiment is to obtain the detailed design parameter to design a prototype model which can handle 50 ∿ 100 T-1 class PCM lines. If the capacity can be realized, the packet switching system can be applicable to commercial quality voice signal as well as handwidth compressed voice and data. The experimental result indicates that, with reasonable effort to enhance the capacity, above described scale of packet switching system is realizable.

Tadao SAITO is an associate professor of electrical engineering of the University of Tokyo. He received his B.S., M.S. and Ph.D. degrees in electronic engineering from the University of Tokyo in 1963, 1965 and 1968 respectively. Since 1969 he is an associate professor of the University of Tokyo. He was a research associate of Caltech in 1974-1975. His study subject includes digital communication systems, computer networks and protocol standardization.

Hiroshi INOSE has been Professor of Electronic Engineering since 1961, and the Director of Computer Center since 1977, at the University of Tokyo. His work has been concerned with electronic switching, digital modulation, data communication and road traffic control systems. He has received a number of awards for his work including the Second Marconi International Fellowship and Japan Academy Prize. He served as the Program Committee Chairman of the Fourth ICCC. Dr. Inose is a foreign associate of the National Academy of Sciences (U.S.A.), a foreign member of the American Philosophical Society and a Fellow of IEEE.

Masahiro WADA is an engineer of Kokusai Denshin Denwa (KDD) of Japan. He received his B.S. and M.S. degrees from the University of Tokyo in 1979 and 1981 respectively. His study for his M.S. thesis is included in this paper.



Hitoshi SHIBAGAKI is an engineer of NTT Yokosuka Electrical Communication Laboratory. He received his B.S. and M.S. degrees from the University of Tokyo in 1978 and 1981 respectively. His current interest is software engineering to produce application software of on-line systems.

EVALUATION OF ALTERNATIVES FOR INTRA-NETWORK IMPLEMENTATION OF X.25 COMPATIBLE PACKET SWITCHING NETWORKS

by Tadao Saito, Shinichi Kobayashi, Yoshio Takeuchi and Hirochi Inose

Department of Electronic Engineering, Faculty of Engineering, University of Tokyo

7-3-1, Hongo, Bunkyo-ku, Tokyo, 113 Japan; phone 03-812-2111 ext 6662, telex 2722111

SUMMARY X.25 protocol specifies the way of communication for DTE to realize error free communication by intensive error control. In usual implementation, the link level error control is applied link-by-link. However, conforming to X.25 protocol, the error control section can be devised up in other ways. For example, the error control section may be set between DTE's, in this case error control can be selectively applied for each communication. If link-by-link error control is not needed at every nodes, store and forward switching principle may be applied in a different way. This paper classifies various possibility for realization of error control and switching, and proposes several combinations and evaluate the performance from the viewpoint of delay characteristics.

1. INTRODUCTION

Recently communication protocol is specified in a layered model. In layered model, variety of functions like error control and flow control are specified in each layer and executed repeatedly. The layered architecture may simplify the protocol implementation but has a tendency to cause excessive overhead. The overhead will cause excessive delay in lower levels of a communication architecture.

This paper is intended to propose several alternatives to reduce the packet transfer delay caused by error control in X.25 compatible packet switching network. The principal cause of the delay is buffering at each node for store and forward switching. The buffering delay may be reduced if link level error control is not executed at each node. If link level error control is performed only between nodes which are directly connected to transmitting and receiving DTE's, intermediate switching nodes do not need buffering of a packet. Each node may only store a part of a packet which is necessary for header analysis for routing. This switching principle was proposed by Kermani [1], and is referred to in this paper as semi-store and forward (SSF). The nodes between which link level error control is executed must be operated on usual full store and forward (FSF) principle.

The error recovery is possible either by means of REJ (or SREJ) frame in the link level for the case of FSF or by means of timeout in the network level. Routes for packet transfer can be selected for each of packets or for each of logical channel. Flow controls for the link level and the packet level may be combined in some of the implementations.

Performance of the network depends on combination of these alternatives. This paper analyzes end-to-end packet delay to evaluate the network performance, considering various combination of these alternatives and various values of error rate, network size and other parameters. In chapter 2., function of X.25 is classified in terms of several principles, and several alternatives of these principles for the purpose of data communication or voice communication are proposed. In chapter 3., three models are specified. End-to-end packet delay is evaluated for these models by theoretical analysis and simulation, respectively, in chapters 4. and 5..

2. VARIOUS IMPLEMENTATION OF X.25

2.1 General description of X.25 protocol

X.25 protocol which is adopted in majority of public packet switching networks in the world specifies the way of communication between the user and the network. According to the terminology of X.25, the user and the network are called DTE (Data Terminating Equipment) and DCE (Data Circuit-terminating Equipment), respectively. X.25 consists of the following three distinct levels:

1) the physical level
2) the link level
3) the packet level

The functions described in the X.25 protocol in each level are independent each other. And they can be implemented in various ways in the network.

The link level specifies the use of data link control procedures which are compatible with HDLC(High-Level Data Link Control). HDLC specifies three types of frames: Information frame (I-frame), Supervisory frame (S-frame) and Unnumbered frame (U-frame). I-frames are used to transfer user data, which is numbered sequentially. This number is called "frame sequence number", and denoted by $N(S)$. The sending node transmits I-frames to the receiving node. When a I-frame is lost or some bits in it are disturbed, the receiving node returns a S-frame for error recovery by specifying the frame sequence number. The error recovery is possible by means of either REJ frame or SREJ frame. REJ frames are used to request retransmission of all I-frames following a disturbed one. SREJ frames are used to request retransmission of a single I-frame previously transmitted. If REJ is used, some I-frames may be retransmitted due to the error of other frame. Obviously, end-to-end packet delay is smaller in the case using SREJ than the case using REJ. Due to its complexity, SREJ is adopted as an optional function of HDLC and only rarely implemented in current systems.

The packet level is the highest level of the X.25 protocol and specifies the manner in which control information and user data are structured into packets.

In case of virtual circuit service, a logical channel is established between the sending DTE and the receiving DTE, before user data are transmitted. Then user data are decomposed into packets, and each packet is numbered sequentially; this is called "packet sequence number". Functions of flow control and confirmation of packet are implemented by using the packet sequence number in each logical channel.

2.2 Classification of implementation of the function of X.25

Various implementation of X.25 protocol may be devised keeping the communication procedure between users and the network unchanged. The intra-network implementation of the protocol can be classified in terms of the following four principles.

1) the switching principle;
2) the error control principle;
3) the routing principle;
4) the flow control principle.

2.2.1 Switching principle

In FSF packet switching, the DTE and the DCE store a complete packet before further transmission to the next node. In SSF packet switching, the switching node can store just the header of the packet. As soon as the route is determined by the analysis of the header, it begins packet transmission to the next node, unless the transmission channel is busy. The SSF principle can be applied in all the switching node or combined with usual full-store and forward (FSF) switching, It is more reasonable to adopt SSF switching in the following three cases: (1) when the packet switching network is used for voice communication in which small end-to-end delay is essential, (2) when transmission error rate is so small that the effect of error is negligible, or (3) when network size is small. In a network using optical fiber transmission having very low error rate, SSF switching is more advantageous due to the second reason.

The way to use the switching principles can be classified as follows.

[1] The FSF principle is applied to all the switching node. This method is adopted in conventional packet switching networks.

[2] The FSF principle is applied to the DCE's directly accommodating DTE's. For the other intermediate switching nodes, the SSF principle is applied.

[3] The SSF principle is applied at all the switching nodes. In this system, error control and flow control are only possible between end nodes. This is preferable for voice communication especially, where small end-to-end delay is more essential than the correctness of transmitted packet, transmission speed of each voice channel is standardized to be 64 kbps. As for data communication, the optimal principle depends on the parameters.

In the switching system using SSF principle, it is difficult to separate error control and flow control in the link level and the packet level in the switching node where SSF principle is applied. Therefore, the switching principles should be considered in relation with the following alternatives.

2.2.2 Error control principle

Error control in the packet switching network can be performed in link level or packet level. The error can be controlled by REJ or SREJ frames or by timeout of link level or packet level. The packet level error control can only be applied between DTE and DTE. Data link level error control can be applied in various sections shown in Fig.1.

(1) Data link level error is recovered by each physical link.

(2) Data link level error is recovered by two subscriber links and one B-link, where subscriber link is defined to be the physical link between the DTE and its adjacent DCE, and B-link is defined to be a series of physical links between the sending DCE and the receiving DCE as shown in Fig.1..

(3) Data link level error is recovered only by C-link. Here, C-link is defined to be a series of the physical links between the sending DTE and the receiving DTE.

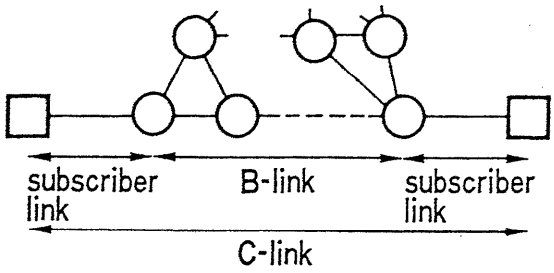


Fig.1 — Logical Channel and Link in Packet Switching Network

As for the method for error recovery, there can be following alternatives.

(a) Frame error is checked by FCS (Frame Check Sequence) in HDLC, and error is recovered by REJ .

(b) Frame error is checked by FCS, and error is recovered by SREJ.

(c) Error is recovered by timeout in the link level;

(d) No error recovery is executed in link level; (if error control is necessary, it can be realized by the timeout of the packet level).

2.2.3 Routing principle

If error control is performed in the B-link, and if link level error control is to be performed, a series of packets belonging to a logical channel must be routed through a chain of data link. In other words, route should be selected for each logical channel rather than for each packet. Therefore two alternatives of routing are:

(i) Route is selected for each packet.

(ii) Route is selected for each logical channel.

Generally speaking, the method (i) is more powerful and more flexible.

2.2.4 Flow control principle

Variety of flow control is executed in actual network. As for the flow control method adopted in X.25, they are classified as follows.

(A) packet level flow control

(B) link level flow control

Some of flow control principles are difficult to be applied depending on the relation of the error control section.

2.3 Combinations of principles

Principles described in 2.2 can be combined in various ways, however, some combinations are meaningless or impossible. Therefore in Table 1, the combinations of switching principle and error control principle are shown. On the error control principle, it is itemized whether error is recovered in link level or packet level. The alternatives of the error control section are also shown.

On routing and flow control principle, possible alternatives depend on the switching principle and the error control principle. Possible alternatives are listed in the table.

<table>
<tr><th rowspan="2">switching principle</th><th rowspan="2">error recovery pinciple</th><th colspan="3">error control section</th></tr>
<tr><th>(1) in all physical link</th><th>(2) in B-link and subscriber links</th><th>(3) in C-link</th></tr>
<tr><td rowspan="4">(1) FST in all nodes</td><td rowspan="2">(a) frame error is recovered in the frame level</td><td>routing : O(i)(ii)
FC* : O(1)(2)</td><td colspan="2">routing : O(ii), X(i)
FC : O(1), X(2)</td></tr>
<tr><td colspan="2">MA** : O</td><td>MA : X</td></tr>
<tr><td rowspan="2">(b) frame error is not recovered in the frame level</td><td>routing : O(i)(ii)
FC : O(1)(2)</td><td colspan="2">routing : O(i)(ii)
FC : O(1), X(2)</td></tr>
<tr><td colspan="3">MA : O</td></tr>
<tr><td rowspan="3">(2) SSF in switching nodes in network</td><td rowspan="2">(a) frame error is recovered in the frame level</td><td rowspan="6">not adaptable</td><td colspan="2">routing : O(i), X(ii)
FC : O(1), X(2)</td></tr>
<tr><td>MA : O</td><td>MA : X</td></tr>
<tr><td>(b) frame error is not recovered in the frame level</td><td colspan="2">routing : O(i)(ii)
FC : O(1), X(2)
MA : O</td></tr>
<tr><td rowspan="3">(3) SSF in all switching nodes</td><td rowspan="2">(a) frame error is recovered in the frame level</td><td colspan="2"></td><td>routing : O(ii), X(2),
FC : O(1), X(2),
MA : X</td></tr>
<tr><td colspan="2"></td><td>routing : O(i)(ii)
FC : O(1), X(2)
MA : O</td></tr>
<tr><td>(b) frame error is not recovered in the frame level</td><td colspan="3"></td></tr>
</table>

* FC : Flow control
** MA : Multiple Access

Table.1 — Combination of Principles

In addition to them, two characteristics of the network, *i.e.* line speed adaptation and multiplexing of a data link by logical channels depends on the combinations of the principles. Combination of [2]-(1), [3]-(1) and [3]-(2) is not possible, because error control is possible only when the switching nodes at the ends of the error control section has FSF function.

2.3.1 Data communication oriented implementation

In data communication, error control function is essential. Then error control principle (a) or (b) is more suitable than (c) or (d). As for switching principle [1], error principle (1) is most suitable, because its error control section is shorter. As for switching principle [2], error control principle(2) is the best by the same reason. In the combination [3]-(3), SSF principle used for all the switching node and link level error control is applied only in C-link. In this combination, since a data link cannot be multiplexed by a plurality of logical channels, and only one data link is possible in one physical link, the function of X.25 will be partially lost. However, X.25 compatible DTE can establish a single logical channel between end nodes and can establish logical channel which is compatible with X.25.

Since error control is usually needed for networks for data communication, link level error control is preferable to packet level error control. Therefore three choices are advantageous:

[1]-(1)-(a) This combination is FSF system having link level error control in every data link. This combination is commonly used in usual X.25 implementation and adequate to realize reliable communication using communication link with high error rate.

[2]-(2)-(a) In this combination, link level error control is performed in B-links. In this combination DTE can use full X.25 capability, thus DTE can establish a plurality of data links with various line speed.

[3]-(3)-(a) In this combination, data link error control is performed in C-link. This combination is preferable if error rate of communication links is low and if delay characteristics is more critical than error characteristics.

2.3.2 Voice communication oriented implementation

In voice communication, delay should be kept as short as possible. For this purpose switching principle [3] is the best. In the sense, [3]-(3) is adequate. If the communication link is also used for data, error control principle (c) or (b) may be used. In voice communication, error control can be applied selectively,

In the following chapters, three network models, *i.e.*, [1]-(1)-(a), [2]-(2)-(a) and [3]-(3)-(a) will be analyzed in detail.

3. THE NETWORK MODEL FOR EVALUATION

This chapter describes the network model in order to analyze three types of intra-network implementation preferable for data communication.

3.1 General assumption for network model

To simplify the analysis, the operation of the packet switching network is assumed to be as follows:

(1) A logical channel is modeled to be a tandem connection of nodes and physical links as shown in Fig.2.

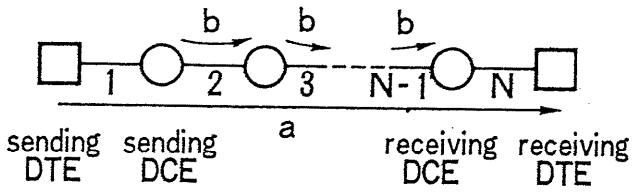


Fig.2 — Model and Traffic of Logical Channel

(2) A subscriber link is not multiplexed by logical channels. On the other hand, physical link between switching nodes may be multiplexed by a number of logical channels.

(3) Just the data transmission phase is analyzed, both in the link level and in the packet level.

(4) Traffic is classifyed into two kind; one is internal traffic between sending DTE and receiving DTE, denoted by a, the other is transit traffic between a pair of adjacent switching nodes, denoted by b.

(5) Packet to a data link arrives according to Poisson distribution. The nodes connected in tandem is analyzed independently.

(6) Transmission error is recovered by REJ frame or SREJ frame. The difference of the characteristics is analyzed theoretically in chapter 4..

(7) The delivery of I-frame is confirmed by RR frame.

(8) The traffic of S-frame, such as REJ, SREJ and RR frame, is negligible and neglected in the analysis.

(9) The header and the trailer of a frame is assumed to be error-free. Thus I-frame is assumed to be delivered to the destination, even if the frame has error or not. This is for simplicity of analysis.

(10) The length of user data in a DATA packet is constant.

(11) Since transmission time is usually longer than processing time in node, the effect of processing time in switching nodes is neglected.

(12) Size of transmission queue of each node is assumed to be long enough.

(13) To simplify the analysis, the effect of flow control is neglected.

3.2 Three network models

The following three network models are analyzed in chapter 4.

MODEL I

The FSF switching principle is applied to all the switching node. Error control is executed by each physical link. This protocol model is [1]-(1)-(a) type described in 2.3, which is identical to that of existing typical packet switching networks.

MODEL II

The SSF principle is applied to the switching nodes between the sending DCE and the receiving DCE, and the FSF principle is applied in the sending DCE and receiving DCE. Error control is executed by each subscriber link and by B-link. This protocol model is [2]-(2)-(a) type.

MODEL III

The SSF principle is applied to all switching nodes. Error control is executed only by C-link. This model is [3]-(3)-(a) type.

4. THEORETICAL ANALYSIS OF PACKET DELAY

Delay characteristics of the three models are described in this chapter. Two error control schemes, *i.e.*, by REJ frame and SREJ frame are analyzed.

4.1 Packet delay caused by error control using SREJ frame

Packet transfer delay between DTE's using SREJ frame for error recovery can be analyzed in the following 3 steps.

(Step 1) Assuming n physical links in a particular error control section, I-frame transfer time (t) between the sending node and the receiving node is evaluated. This does not include the time needed for acknowledgement and retransmission.

(Step 2) I-frame transfer time including acknowledgement time and retransmission time between the error control section, is evaluated.

(Step 3) For Model I, Model II and Model III, packet transfer delay between DTE's (D_{1SREJ}, D_{2SREJ} and D_{3SREJ}, respectively), is evaluated.

4.1.1 Step 1

The following symbols are used in the analysis:

n = number of physical links in the error control section (if $n>2$ then the SSF principle is adopted at intermediate switching nodes)

e = frame error rate in one physical link

α_0 = internal traffic transferred between the sending node and the receiving node [Erlang]

α_i = transit traffic at physical link i [Erlang]

L = length of user data [octet]

H = length of link level and packet level header [octet]

C = line speed of physical link [bit/sec]

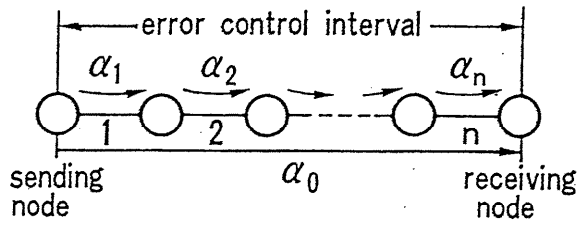


Fig.3 — Model of Nodes of the Error Control Section

Assuming that the identical error control scheme is applied for the transit traffic, the actual traffic at physical link i including the retransmitted frame, is given by Eq.(1).

$$\rho_i = \frac{\alpha_0 + \alpha_i}{1 - P_e} \qquad (1 \leqslant i \leqslant n) \tag{1}$$

where, P_e is the frame error rate in the error control section given by

$$P_e = 1-(1-e)^n \tag{2}$$

Assuming each X.25 packet has control field of H octets as a header, I-frame transmission time is evaluated by Eq.(3).

$$\frac{1}{\mu} = \frac{8(L+H)}{C} \tag{3}$$

Therefore, the sum of I-frame transmission time and the time needed to store control fields at $(n-1)$ switching nodes is given by Eq.(4).

$$u = \frac{8(L+H)}{C} + \frac{8\times H(n-1)}{C} = \frac{8(L+Hn)}{C} \tag{4}$$

Mean waiting time for transmission at link i, is evaluated by applying the following M/D/1 formula.

$$w_i = \frac{\rho_i}{2(1-\rho_i)\mu} \qquad (1 \leqslant i \leqslant n) \tag{5}$$

Then, I-frame transfer time from the sending node to the receiving node is given by

$$t = u + \sum_{i=1}^{n} w_i = \frac{8(L+Hn)}{C} + \sum_{i=1}^{n} \frac{\rho_i}{2(1-\rho_i)\mu} \tag{6}$$

Attention must be focused on the fact that t does not contain the acknowledgement time and retransmission time.

The transmission time for S-frames is similarly evaluated. Variables related to S-frame is denoted by a prime attached to each notation. Assuming that the length of packet containing S-frame is again H, the sum of SREJ frame transmission time and time needed to store control field at switching nodes, is given by Eq.(7).

$$u' = 8\times\frac{H}{C} + \frac{8\times H(n-1)}{C}$$
$$= \frac{8Hn}{C} \qquad (7)$$

Then, SREJ frame transfer time from the receiving node to the sending node is given by Eq.(8).

$$t' = u' + \sum_{i=1}^{n} w_i$$
$$= \frac{8Hn}{C} + \sum_{i=1}^{n} \frac{\rho_i}{2(1-\rho_i)\mu} \qquad (8)$$

4.1.2 Step 2

Using the values of t and t', I-frame transfer time including acknowledgement time and retransmission time in the error control section is evaluated by Eq.(9).

$$d_{SREJ} = t + P_e(t+t') + P_e^2(t+t') + \cdots$$
$$= \frac{t + P_e t'}{1 - P_e} \qquad (9)$$

Since d_{SREJ} is a function of e, n and α_i under several constants such as L, H or C, let it be denoted as

$$d_{SREJ}(e;n;\alpha_0,\alpha_1,\cdots,\alpha_n)$$

4.1.3 Step 3

Considering a logical channel to be a series of the error control sections, packet transfer delay between DTE's can be calculated as a sum of I-frame transfer time. As in Fig.2, following symbols are used.

N= total number of physical links between DTE's

a= internal traffic transferred between DTE's

b= transit traffic at each physical link.

(1) MODEL I

For MODEL I, error control is executed at all physical links; therefore n=1.
As for subscriber link, internal traffic and transit traffic are respectively given by

$$\alpha_0 = a \quad , \quad \alpha_1 = 0.$$

As for the other link, two kinds of traffic are given by

$$\alpha_0 = a+b \ , \quad \alpha_1 = 0.$$

Therefore, packet transfer delay between DTE's is evaluated by Eq.(10).

$$D_{1SREJ}(e;N;a,b)$$
$$= 2\, d_{SREJ}(e;1;a,0) + (N-2)\, d_{SREJ}(e;1;a+b,0) \qquad (10)$$

(2) MODEL II

For MODEL II, error control is executed at each subscriber link and at B-link.
As for subscriber link, n=1 and internal traffic and transit traffic are respectively,

$$\alpha_0 = a \quad , \quad \alpha_1 = 0.$$

As for B-link, n=N−2 and the traffic of each link is give by

$$\alpha_0 = a \quad , \quad \alpha_1 = \alpha_2 = \cdots = \alpha_{N-2} = b.$$

Therefore, packet transfer delay between DTE's, D_{2SREJ}, is given by Eq.(11):

$$D_{2SREJ}(e;N;a,b)$$
$$= 2d_{SREJ}(e;1;a,o) + d_{SREJ}(e;N-2;a,\underbrace{b,b,\ldots,b}_{N-2}) \qquad (11)$$

(3) MODEL III

For MODEL III, error control is executed only by C-link, hence n=N. Then traffic is given as below:

$$\alpha_0 = a \quad , \quad \alpha_1 = \alpha_N = 0 \quad ,$$
$$\alpha_2 = \alpha_3 = \cdots = \alpha_{N-1} = b$$

Therefore, packet transfer delay, D_{3SREJ}, is given by Eq.(12).

$$D_{3SREJ}(e;N;a,b) = d_{SREJ}(e;N;a,0,\underbrace{b,b,\ldots,b}_{N-2},0) \qquad (12)$$

4.2 Packet delay cause by error control using REJ frame

Packet transfer delay between DTE's when REJ frame is used for error recovery, can be analyzed in the following three steps.

(Step 1) I-frame transfer time (t) in an error control section is evaluated by estimating traffic applied to each link. The traffic is evaluated considering the number of frames retransmitted due to a frame error. The I-frame transfer time (t) does not include the time needed for acknowledgement and retransmission.

(Step 2) I-frame transfer time including acknowledgment time and retransmission time between the error control section is evaluated.

(Step 3) For Model I, Model II and Model III, packet transfer delay between DTE's (D_{1REJ}, D_{2REJ} and D_{3REJ}, respectively), is evaluated.

4.2.1 Step 1

Assume that the transit traffic is transferred under the same error control as the internal traffic. Then probability of I-frame error in the error control section is given by Eq.(13).

$$P_e = 1-(1-e)^n \qquad (13)$$

Therefore, one frame error occurs for every $\frac{1}{P_e}$ frames transmitted.
Denoting

m = number of frames which is involved in retransmission due to frame error of one frame

ξ = probability that REJ is returned for each I-frame

The value of ξ is a little smaller than that of P_e, because error may have already been detected in the m-proceeding frames. Then, ξ is given by Eq.(14).

$$\xi = \frac{1}{\frac{1}{P_e}+m} = \frac{P_e}{P_e m+1} \tag{14}$$

If the sending node receives a REJ frame, it retransmit $(m+1)$ I-frames consecutively. Then, actual traffic by physical link i including retransmission frames is given by

$$\rho_i = \frac{\alpha_0+\alpha_i}{1-\xi(m+1)} \qquad (1 \leqslant i \leqslant n) \tag{15}$$

Similarly to the case for SREJ, I-frame transfer time from the sending node to the receiving node and REJ frame transfer time from the receiving node to the sending node are evaluated as follows.

$$t = \frac{8(L+Hn)}{C} + \sum_{i=1}^{n} \frac{\rho_i}{2(1-\rho_i)\mu} \tag{16}$$

$$t' = \frac{8Hn}{C} + \sum_{i=1}^{n} \frac{\rho_i}{2(1-\rho_i)\mu} \tag{17}$$

The value of m is evaluated as the product of frame arrival rate and round trip time by

$$m = \rho_0 \mu (t+t') \tag{18}$$

where μ is give by Eq.(3). By Eqs.(14)–(18), m is evaluated, then t, t' and ρ_i are evaluated.

4.2.2 Step 2

Then, frame transfer delay including acknowledgement time and retransmission time (d_{REJ}) is evaluated. Usually, I-frames arrives at the receiving node separatedly according to the arrival at the sending node. But retransmitted I-frames are consecutively transmitted from the sending node and they tend to arrive at the destination node in a cluster. And the time needed for retransmission depends on the difference of sequence number between the frame and the frame which includes the error. To simplify the notation, let p_l be the probability that a certain I-frame is retransmitted due to the error of frame proceeding $m-l$. In other words,

p_l = probability that I-frame with $N(S)=m$ is retransmitted due to the error of I-frame with $N(S)=l$.

Denoting

d_l = time duration from arrival of I-frame with $N(S)=m$ to arrival of the retransmitted one with $N(S)=m$, due to the error of I-frame with $N(S)=l$.

P_{RET} = retransmission probability for each I-frame

the value of P_{RET} is given by

$$P_{RET} = \sum_{l=0}^{m} p_l \tag{19}$$

Then

$$d_{REJ} = t + \sum_{l=0}^{m} p_l d_l + P_{RET} \sum_{l=0}^{m} p_l d_l + P_{RET}^2 \sum_{l=0}^{m} p_l d_l + \cdots$$

$$= t + \frac{1}{1-P_{RET}} \sum_{l=0}^{m} p_l d_l \tag{20}$$

Therefore, d_{REJ} will be evaluated, if p_l and d_l are known.

The value of p_l is evaluated as follows. Denoting

Q = probability that the receiving node is in abnormal state after transmission of REJ frame

Where the abnormal state is defined as the state of the receiving node, after the transmission of REJ frame until following m frames are received.

Therefore,

$$Q = (m+1)\xi \tag{21}$$

The value of p_0, *i.e.*, the probability that frame with $N(S)=m$ is retransmitted due to the error of frame with $N(S)=0$ is evaluated as the joint probability that

(1) The receiving node is not in abnormal state when the I-frame with $N(S)=0$ arrives at the receiving node, and

(2) the I-frame includes error.

Therefore, p_0 is given by

$$p_0 = (1-Q)P_e \tag{22}$$

As for $l \geqslant 1$, p_l is the joint probability of two cases. The first case is when

(1) the receiving node is not in abnormal state when the I-frame with $N(S)=0$ arrives at the receiving node, and

(2) I-frames with $0 \leqslant N(S) \leqslant l-1$ does not include error, and

(3) I-frame with $N(S)=l$ includes error.

The second case is when

(1) the receiving node is in abnormal state when the I-frame with $N(S)=0$ arrives at the receiving node, and

(2) the receiving node recovers from abnormal state at some I-frame (for example, with $N(S)=j$), and

(3) I-frames with $j \leqslant N(S) \leqslant l-1$ does not include error, and

(4) I-frame with $N(S)=l$ includes error.

The probability that the receiving node recovers from abnormal state when an I-frame is received, is given by $\frac{1}{m+1}$. Therefore,

$$p_l = (1-Q)(1-P_e)^l P_e + Q\sum_{j=1}^{l} \frac{1}{m+1}(1-P_e)^{l-j} P_e$$

$$= (1-Q)(1-P_e)^l P_e + \frac{Q}{m+1}(1-(1-P_e)^l) P_e$$

$$(1 \leqslant l \leqslant m) \tag{23}$$

Eq.(23) also holds for $l=0$. Thus p_l can be evaluated

by Eq.(23) for $0 \leqslant l \leqslant m$.

Then, d_l is evaluated as follows. Assume that I-frame with $N(S)=l$ arrives at the receiving node at time 0. After time 0 the following events will happen:

(1) at time 0 : I-frame with $N(S)=l$ arrives in the receiving node

(2) at time $\frac{m-l}{\alpha_0 \mu}$: I-frame with $N(S)=m$ arrives in the receiving node

(3) at time t' : REJ arrives in the sending node

(4) at time $\frac{t'+t}{\mu}$: retransmitted I-frame with $N(S)=l$ arrives in the receiving node

(5) at time $t'+t+\frac{m-l}{\mu}$: retransmitted I-frame with $N(S)=m$ arrives in the receiving node

Therefore,

$$d_l = t'+t+\frac{m-l}{\mu}-\frac{m-l}{\alpha_0 \mu}$$

$$= t+t'+\frac{m-l}{\mu}\left(1-\frac{1}{\alpha_0}\right) \tag{24}$$

By substituting Eq.(23) and Eq.(24) to Eq.(20), d_{REJ} is evaluated.

4.2.3 Step 3

Using the results obtained by step 2, packet transfer delay between DTE's can be calculated for MODEL I, MODEL II and MODEL III.

$$D_{1REJ}(e;N;a,b)$$

$$= 2\, d_{REJ}(e;1;a,0) + (N-2)\, d_{REJ}(e;1;a+b,0) \tag{25}$$

$$D_{2REJ}(e;N;a,b)$$

$$= 2 d_{REJ}(e;1;a,o) + d_{REJ}(e;N-2;a,\underbrace{b,b,\ldots,b}_{N-2}) \tag{26}$$

$$D_{3REJ}(e;N;a,b)$$

$$= d_{REJ}(e;N;a,0,\underbrace{b,b,\ldots,b}_{N-2},0) \tag{27}$$

4.3 Packet delay without error recovery : Voice communication case

If error recovery can be applied selectively, error control should be omitted for voice communications. The transmission delay for each model can be obtained from Eqn.(10)-(12) and Eqs.(25)-(27) by substituting $e=0$. Therefore, for Model I, Model II and Model III, packet delay is respectively given by

$$D_{1VOICE}(N;a,b)$$

$$= D_{1SREJ}(0;N;a,b) = D_{1REJ}(0;N;a,b) \tag{28}$$

$$D_{2VOICE}(N;a,b)$$

$$= D_{2SREJ}(0;N;a,b) = D_{2REJ}(0;N;a,b) \tag{29}$$

$$D_{3VOICE}(N;a,b)$$

$$= D_{3SREJ}(0;N;a,b) = D_{3REJ}(0;N;a,b) \tag{30}$$

4.4 Some example of theoretical analysis

Fig.4 shows the relation of frame error rate (e) versus packet delay (D) under the parameters in Table 2.

internal traffic	a	0.2
transit traffic	b	0.2
length of user data	L	256 [octet]
length of header	H	10 [octet]
line speed	C	48 [kbps]
number of physical links	N	6

Table.2 — Parameters for Fig.4

As expected in **2.2.1**, Model II and MODEL III have shorter delay when $e<0.01$. When error rate is high, for example, if $e \geqslant 0.1$ for SREJ case or $e \geqslant 0.05$ for REJ case, packet delay in MODEL II and MODEL III increases rapidly. Obviously, packet delay for SREJ case is shorter than that for REJ case.

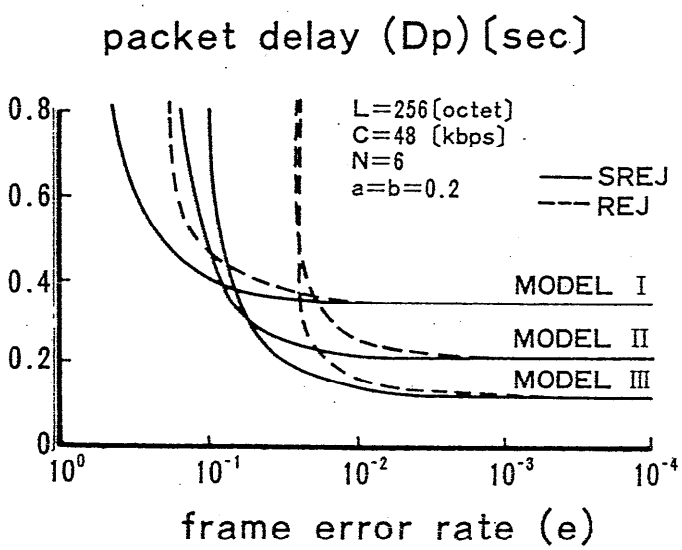


Fig.4 — Packet Delay versus Frame Error Rate

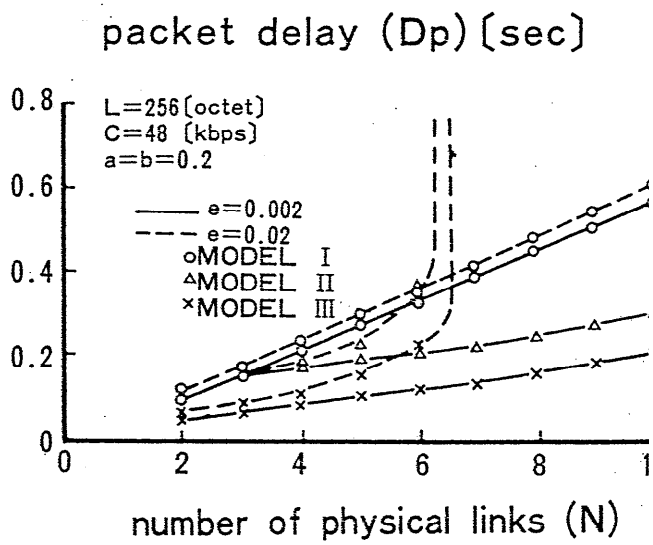


Fig.5 — Packet Delay versus Number of Physical Links

Next, Fig.5 shows the relation of packet delay (D) versus number of physical links (N), taking the error rate as a parameter. The results indicates that D and N is linearly related for MODEL I, and D increases rapidly when e=0.02 and N>7 for MODEL II and MODEL II. However, if error rate is low enough even if the number of physical links is large, packet delay in MODEL II and MODEL II is sufficiently small. Therefore, the SSF principle gives acceptable characteristics for the packet switching networks.

5. EVALUATION OF DELAY BY SIMULATION

For the purpose of validating the theoretical evaluation described in chapter **4.**, computer simulation has been made.

Using the simulator based on Fortran and Gasp-II[2], statistics are collected for thousands of packet in stationary states. As for acknowledgement, S-frames such as RR and REJ (or SREJ) are assumed to reach the sending node in a fixed time after the receiving node transmits it.

Figs.6 and 7 show the relation packet delay (D) versus frame error rate (e), for the case when REJ and SREJ are use for error control respectively. Theoretical analysis in chapter **4.** is confirmed by the results.

Fig.8 shows the relation packet delay (D) and total traffic ($a+b$), when N=6, e=0.01 and a=b. The results indicates that the error recovery principle using REJ is not suitable in heavy loaded condition, when error rate is very high.

6. CONCLUSION

This paper evaluates end-to-end packet transfer delay for various implementation of X.25 compatible packet switching network.

The functions of the implementation are analyzed theoretically in the following three points:

(1) switching principle (SSF and FSF)
(2) error control principle
(SREJ, REJ and No error control)
(3) error control section (link-by-link)

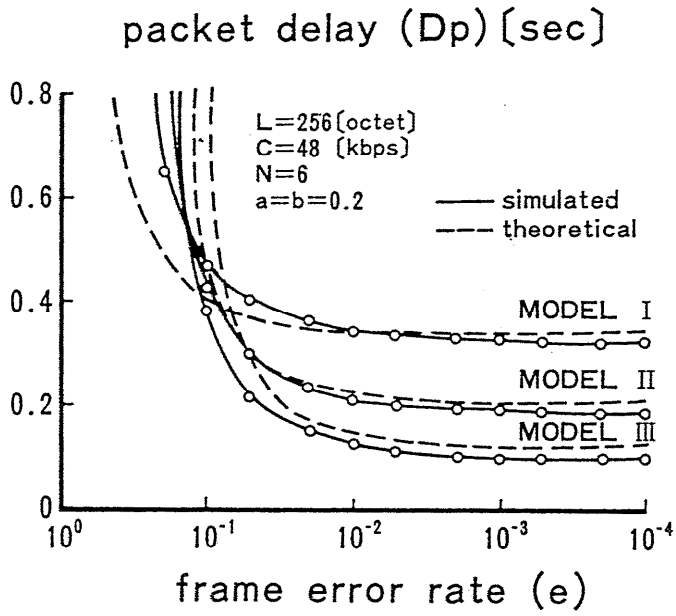


Fig.7 — Packet Delay versus Frame Error Rate (SREJ)

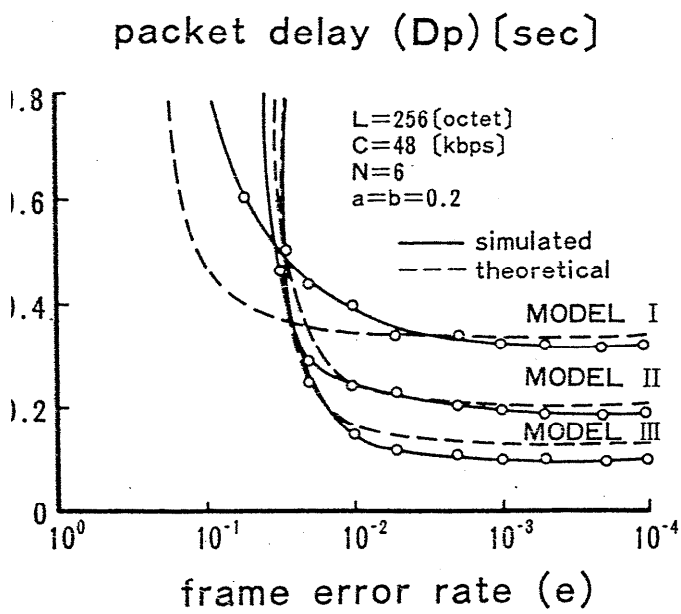


Fig.6 — Packet Delay versus Frame Error Rate (REJ)

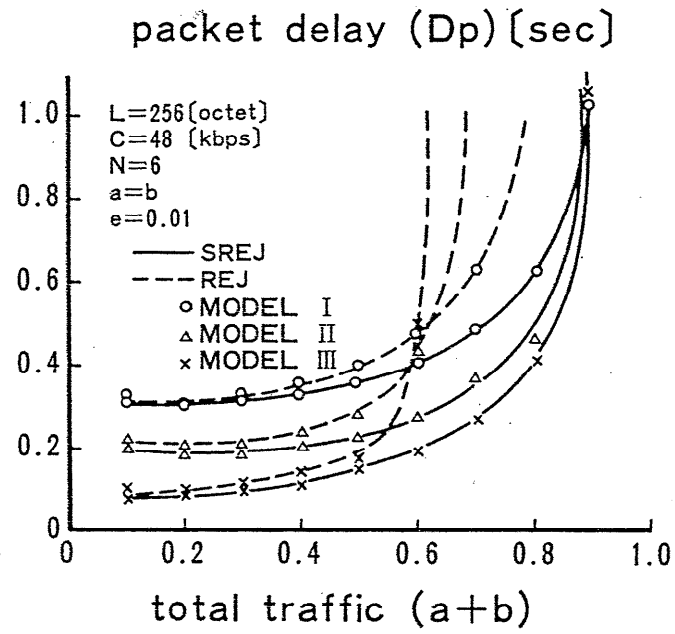


Fig.8 — Relation : Packet Delay versus Total Traffic

The error control scheme using DTE to DTE section as the error control section is found to have acceptable characteristics for both data and voice communication, and especially for voice and data integrated service in packet switching networks.

Although this paper evaluated packet delay, the evaluation of buffer usage for each model and the evaluation of the mutual interference of the flow control and the error control are left for further studies.

REFERENCE

[1] P.Kermani and L.Kleinrock:
"Virtual cut-through:A new computer communication switching technique,"
Computer Networks, 3, pp.267-286 (Sept. 1979)

[2] Kiviat,P.K., : *Simulation with Gasp-II*, Prentice Hall Inc. (1969)

[3] Kobayashi,S., Takeuchi,Y., Saito,T. and Inose,H.
"One relation of packet delay to the switching principle in X.25 compatible network",
National Conference Record of IECE of Japan, No.1634 (April, 1983)

[4] Kobayashi,S:
"On end-to-end delay and throughput of packet switching networks,"
Ph.D. dissertation, Univ. of Tokyo, Japan

GLOSSARY

n = number of physical links in the error control section (if $n>2$ then the SSF principle is adopted at intermediate switching nodes)

e = frame error rate in one physical link

α_0 = internal traffic transferred between the sending node and the receiving node [Erlang]

α_i = transit traffic at physical link i [Erlang]

L = length of user data [octet]

H = length of link level and packet level header [octet]

C = line speed of physical link [bit/sec]

μ = service rate of physical link [bit/sec]

N = total number of physical links between DTE's

a = internal traffic transferred between DTE's

b = transit traffic at each physical link

m = number of frames which is involved in retransmission due to frame error of one frame

ξ = probability that REJ is returned for each I-frame

A Proposal of High Capacity Switching Fabric for Tbps-level IP Router

Tredej Toranawigrai[+], Naoki Ueda[+], Terumasa Aoki[++], Hitoshi Aida[+], Tadao Saito[+]
[+] Department of Information and Communication Engineering, University of Tokyo
7-3-1 Hongo Bunkyo-ku Tokyo Japan 113-8656
tredej,nueda,aida,saito@sail.t.u-tokyo.ac.jp
[++] Research Center for Advanced Science and Technology, University of Tokyo
4-6-1 Komaba Meguro-ku Tokyo Japan 153-8904
aoki@mpeg.rcast.t.u-tokyo.ac.jp

ABSTRACT

We propose a new architecture for high capacity switching fabric in a router. The current technologies for fast packet switching achieve Gigabit-level speed, however, the cost of overhead due to fix-length data switching is unavoidable. Instead of fragmenting variable-length packets to fixed-length cells, we switch them without any modification in order to reduce the overhead by using the architecture called Ring Banyan switch. Ring Banyan switch architecture is based on Multi-stage Interconnection Network (MIN) with non-buffer switch-element. Applying Ring Banyan switch to IP packet achieves low packet-loss rate, scalability and high capacity. As a result, a scalable switching fabric at Tbps-level can be created by Ring Banyan structure. However, performance of Ring Banyan is mainly limited by non-buffer switch-element. We improve the performance of Ring Banyan switch by replacing non-buffer switch-element with a novel switch-element called Misroute Reduction Shared Buffer (MRSB) switch-element. We compare the performance of previously proposed Ring Banyan switch and Ring Banyan with our MRSB switch-element by doing a simulation. The simulation results obviously show the improvement of Ring Banyan performance due to using MRSB switch-element.

1. Introduction

Technologies of data transmission have been developed rapidly. Crossbar switch and ATM switching technologies have been proposed recently to enable packet switching at the level of tens and hundreds Gbps. However, the present Gbps IP routers are insufficient to handle the explosive growth of Internet nowadays. As a result, we aim to create Tbps-level IP router with thousands ports and hundreds of Mbps speed per port. We propose a design for such a very high capacity router based on Ring Banyan architecture [7]. In order to enhance the performance of Ring Banyan, originally proposed non-buffer switch-element is replaced with our new switch called Misroute Reduction Shared Buffer (MRSB) switch-element which use a small amount of buffer memory but can improve the packet-loss rate in Ring Banyan dramatically.

The next section presents existing technologies of IP router, including the architecture of Ring Banyan, briefly. Then, section 3 mentions how to enhance the switch-element in Ring Banyan, and proposes our MRSB switch-element architecture. In section 4, MRSB performance evaluation by simulation is presented to show the improvement of Ring Banyan comparing to the previous one. Finally, section 5 concludes this research and mentions about our future work.

2. Existing Technologies

Transmission IP packet across the network can be separated into 2 ways, i.e. by ATM switching and by IP datagram. ATM technology is a connection-oriented fast packet-

switching protocol that has been standardized for B-ISDN. All types of user information are carried in 53-byte cells consisting of 5-byte header and 48-byte information field. ATM enables high throughput, low-delay cell switching by leaving a number of common field from ATM cell header, as a result, cells can be relayed through the network with minimal processing delay. Various researches have been proposed methods to improve ATM technology, such as IP switching [1] and Tag switching [2], by using cut-through technique. Although ATM gains a lot of advantages, including high speed transmission and scalability due to usage of multi-stage interconnection network (MIN) in ATM switch, applying ATM with variable-length IP packet faces a lot of problems.

Fragmentation and reassembling

Fragmenting IP packet into ATM cells before transmission and reassembling them at the next router increases the complexity of network system, consuming processing power, and also reducing efficiency of transmission packets.

Header fragmentation

The header of IP and TPC with any option may be longer than ATM cell, and fragmented into many cells. As a result, the protocol error due to header fragmentation may occur.

Loss due to dead cell

Fragmentation a packet into ATM cells results in the loss of a whole packet for each cell loss, potentially leading to a form of congestion due to retransmission.

The other method of data transmission currently used is packet switching. The largest-scale router products, such as Tiny Tera [3], Cisco 12000 series [4], and MultiGigabit Router [5], all apply crossbar switch with input-queued architecture. In order to avoid head of line (HOL) blocking, Virtual Output Queue (VOQ) [6] method is used in these routers. In this solution, the number of buffers required in the switch is the order of N^2, where N is the number of ports. As a result, this factor becomes the main limitation of implementation high capacity switching-fabric on real hardware. So far, only tens Gbps router, with a few tens ports can be achieved by these architectures.

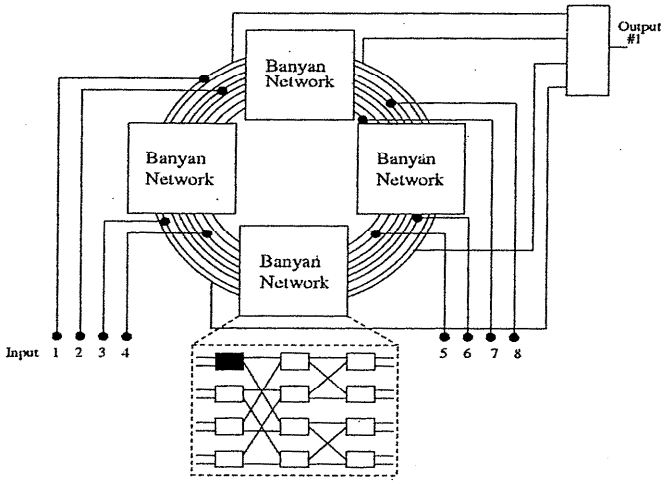


Figure 1: Ring Banyan Switch Architecture

A new architecture called Ring Banyan [7] is proposed to enable both high speed and scalability by using multi-stage interconnection network (MIN) with variable-length packet. Thanks to MIN characteristics, scalability can be achieved by connecting small-size switches together. However, internal blocking problem in MIN must be solved to enable full operation of the switch. To deal with this problem, Ring Banyan concatenates multiple Banyan networks together in ring topology, as illustrated in figure 1. If there is an internal blocking at a certain Banyan network, blocked packets are added misroute-tag, then transmitted to the next Banyan plane to be re-routed. To reduce the probability of blocking in Banyan planes, the input ports are distributed to various planes. As a result, an additional buffer is necessary in each input port to prevent collision of new-coming packet and misrouted packet from previous Banyan plane. At the output side,

output ports from each Banyan plane are connected to their correspondent output-port modules, consisted of an output buffer per port, in order to manage non-misrouted packets before sending them to the network. Since buffer are provided at input and output ports, switching-element in Banyan plane is proposed to be a non-buffer switch. A packet that cannot be switched to its destination port is automatically added by misroute-tag immediately, then transmitted out of the current Banyan plane.

Comparing with Tandem Banyan switch [8], Ring Banyan achieves lower packet-loss rate since internal blocking is reduced by input port distribution and misroute concept.

Ring Banyan achieves the scalability issue by connecting small-scale switches together, for example, 262144x262144 Ring Banyan switch can be created from 9 stages of 4x4 switch-element, or 6 stages of 8x8 switch-element, or 3 stages of 64x64 switch-element. The larger scale is, the lower stages are required, that means smaller delay, and less hardware. If the link speed is 622 Mbps/port, the bandwidth of Ring Banyan switch becomes 163.05 Tbps.

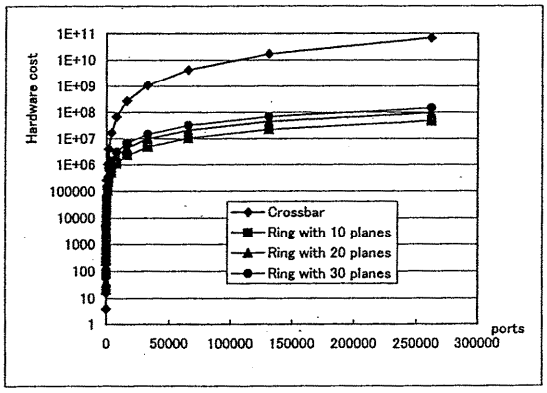


Figure 2: Comparison of hardware between Ring Banyan and Crossbar switch

Since the amount of hardware in NxN Ring Banyan is the proportion of kNlog(N), where Nlog(N) represents the number of switch-element in Banyan planes, and k is the number of planes. The amount of hardware of Ring Banyan and that of Crossbar switch are compared in figure 2. It is obvious that Ring Banyan costs much less hardware than Crossbar switch does.

Although Ring Banyan has a lot of advantages mentioned above, the performance of Ring Banyan is still mainly limited by packet-misroute rate caused by non-buffer switch structure. The next section describes a new architecture of switch-element to enhance the performance of Ring Banyan.

3. Proposal of Switch Element

Enhancement of switch element in Ring Banyan Network

In Ring Banyan switch, providing that buffer at both input port and output port are large enough, a packet is lost when the packet is added by misroute-tag and passes all Banyan planes. As a misrouted packet occurs, it may block other packets and results more misrouted packets, which is the source of packet loss in the switch. In order to relief this situation, the packet-misroute rate should be as low as possible. We add some buffer into switch-element to lower the number of misrouted packets. However, the simplicity of buffer management and switch are preferable. We decide to propose a switch-element that can reduce the packet-misroute rate and not too complicated comparing to non-buffer switch-element, and we call it as Misroute Reduction Shared Buffer (MRSB) switch.

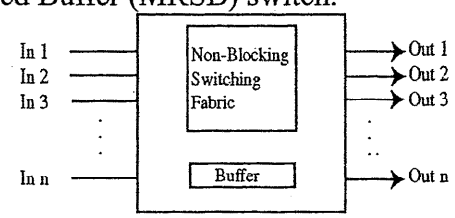


Figure 3: Switch-element concepts

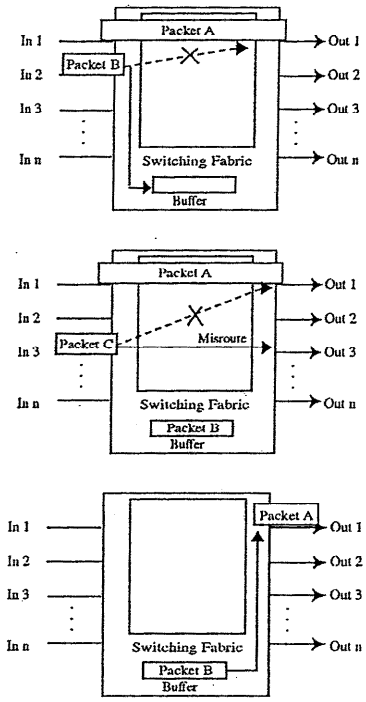


Figure 4: Switch-element operations

Switch-element concepts and operations

Our proposed switch-element operates as follow. A packet that cannot go out to its destination port would be written into a 1500-byte buffer if the buffer is empty, and wait until the destination is unoccupied, then the packet is read out. The process of accessing buffer deals with only one packet each time, as a result, the multiplexing stage and memory access speed issue, in normal shared-buffer switch, are not necessary to concern. In addition, the buffer management is simple because there are only 2 status of buffer, i.e. empty or not empty.

In case the buffer is not empty, and there is a new coming packet that cannot go to its destination, the coming packet will be added by misroute-tag and sent to the next Banyan plane in the similar way as being done in non-buffer switch-element.

Comparing to non-buffer switch, our switch reduces the number of misrouted packets, while still maintains more simplicity and requires less memory size than a normal shared-buffer switch does.

Implementation issues of switch-element

Our goal in this research is to propose a design of high capacity, scalable router. In order to confirm the idea, we plan to implement a scale-down model of Ring Banyan on the real hardware by starting from 64x64 scale, which we believe that it is large enough to show the performance improvement by our proposal.

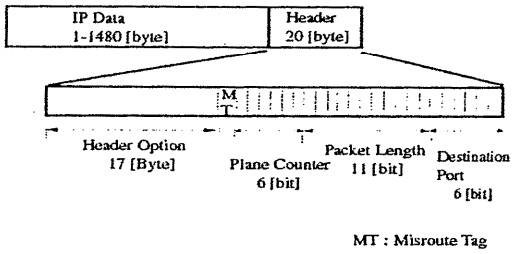


Figure 5: Packet Format for MRSB prototype

At the current state, we are implementing our prototype of MRSB switch on FPGA, and plane to implement 64x64 Ring Banyan Network on customized IC in the near future, in order to investigate its operation and to estimate the amount of hardware needed in the real system. We suppose that our information packet composed of 20-byte header, and 1 to 1480-byte data field, is transmitted to our prototype switch via RS-232 interface. The format of packet used with out prototype switch is shown in figure 5. Packet header field consists of 6-bit destination address, 11-bit packet length, 6-bit Banyan plane counter, 1-bit misroute field and 17-byte header option. When the packet faces internal blocking, its misroute field is set, the Banyan plane counter increases, then packet is transmitted to the next Banyan plane. If the counter is equal to the number of Banyan planes, the packet will be discarded at the

current switch-element before being transmitted out.

The structure of MRSB switch is composed of the following components.

Header Detection Unit (HDU)
HDU performs destination port detection, the passes the packet and its header to SIM.
Switch Input Module (SIM)
SIM is the most important part in MRSB switch. It receives packets and header information from HDU. Then SIM performs a decision making to transmit the packet to destination output port, or write the packet in buffer, or add a misroute-tag to the packet depending on the status flags from Switch Output Modules, write status from Write Unit, and read status from Read Unit. In addition, SIM also sends control-signal to inform Read Unit to read packet from the switch buffer. Another function of SIM is to check the Banyan plane counter field in packet header, and determine whether the packet should be lost in case the misroute-tag is added.
Switch Output Module (SOM)
SOM receives packets from each SIM, while transmits its status flag back. SOM also sends its status to Read Unit in order to let the packet in buffer be read. Normally, SOM receives only one packet from an input port each time.
Write Unit (WU)
WU receives a packet from a certain SIM when the packet is blocked and buffer is free, then performs writing the packet into buffer. The status of WU is sent to each SIM for being a factor in decision making.
Read Unit (RU)
RU performs reading a packet out of switch buffer. Then it transmits the packet to correspondent SOM, depending on the status-out signal from each SOM and control-signal from each SIM. In the same time, information of RU status is also transmitted to SIM for using in switching decision.
Buffer
Usage of buffer in MRSB switch aims to reduce packet-misroute rate in Ring Banyan, however, the buffer size in switch-element is an important factor that must be considered. To maintain small amount of buffer size and less complexity in buffer management, we propose to apply only 1500-byte buffer in MRSB switch. Only one packet is written, and read from buffer each time. In order to increase the switch performance, dual-port memory is expected to be used so that writing process and reading process can be performed at the buffer at the same time.

4. Performance Evaluation

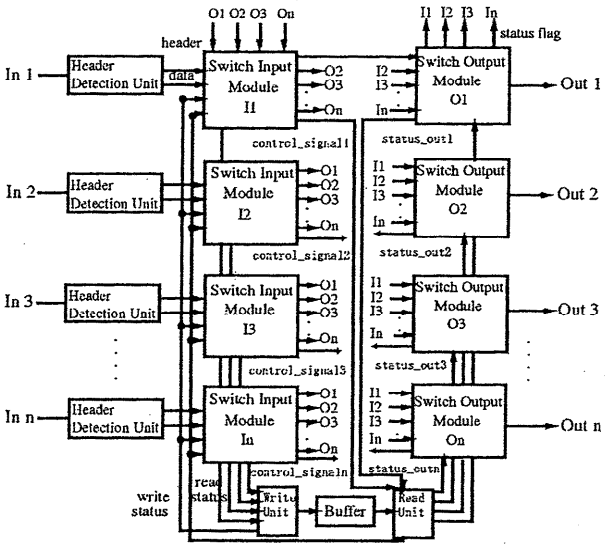


Figure 6: Structure of MRSB switch

In order to evaluate the performance of our MRSB switch, a simulation to investigate the packet-misroute rate of switch-element has been done first. Then we work out a simulation of Ring Banyan to study the suitable number of buffers in switch-element, and performance enhancement of Ring Banyan due to MRSB switch.

4.1 Performance of switch-element

4.1.1 Simulation Model and Condition

Switch-element model

- Number of ports 8 ports
- Number of buffers 0,1,2
- Buffer size 1500 bytes/buffer

Traffic Condition

- Packet length 20 bytes for header, 1-1480 bytes for data
- Packet interval exponential distribution
- Packet destination uniformly random

4.1.2 Simulation Result

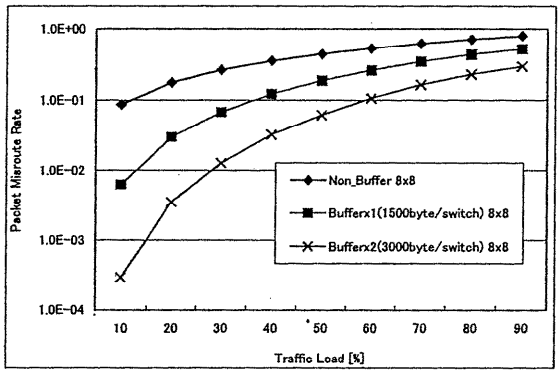


Figure 7: Packet-misroute rate of 8x8 switch-element

Figure 7 shows the packet-misroute rate of 8x8 switch-element. It is obvious that non-buffer switch results more misroute rate than our MRSB switch because some proportion of blocked packets are delayed in buffer for a while before correctly switching out, instead of adding misroute-tag to packets immediately as being done in non-buffer switch.

Comparing between switch with one buffer and switch with multiple buffers, it is normally known that increasing buffer size by adding the number of buffers in switch can reduce packet-misroute rate. However, in order to apply the MRSB switch to Ring Banyan, the optimal number of buffers used in each switch-element must be considered carefully. As a result, we investigate this issue in detail in the following subsection.

4.2 Performance of Ring Banyan

4.2.1 Simulation Model and Condition

Switch Model

- Switch Architecture Ring Banyan
- Number of total ports 64 ports
- Switch-element size 8x8
- Number of Banyan planes 4 planes
- Number of buffers 0,1,2 pieces
- Buffer size 1500 B/buffer

Traffic Condition

- Packet length 20 bytes for header 1-1480 bytes for data
- Packet interval exponential distribution
- Packet destination uniformly random
- Traffic load 60 %

4.2.2 Simulation Result

To determine the suitable number of buffers in one switch-element, we observe the result of packet loss in Ring Banyan using non-buffer switch, and MRSB switch. With non-buffer switch-element, it is obvious that packet-loss rate is much more than Ring Banyan using MRSB switch. This shows the superior of MRSB switch over the original non-buffer one.

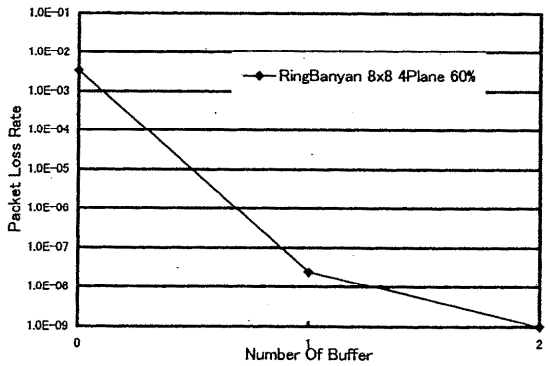


Figure 8: Packet-loss rate from Ring Banyan varying buffer number

Considering the effect of different number of buffers, the result shows that only one buffer in a switch gain packet-loss rate lower than 10^{-7}, which is low enough for this size of router.

Although the packet-loss rate of multiple-buffer switch is lower than that of a single-buffer one, implementing a single-buffer switch is less complicated, and hardware cost in Ring Banyan.

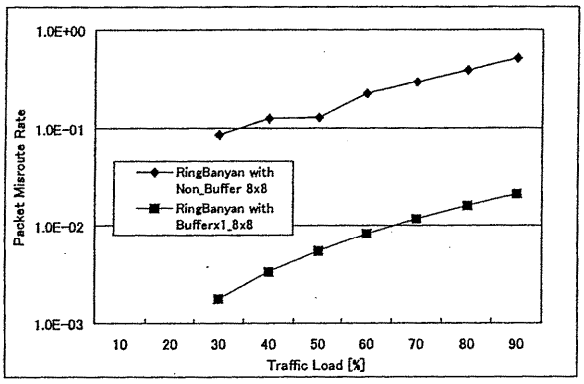


Figure 9: Total Packet-misroute rate from Ring Banyan

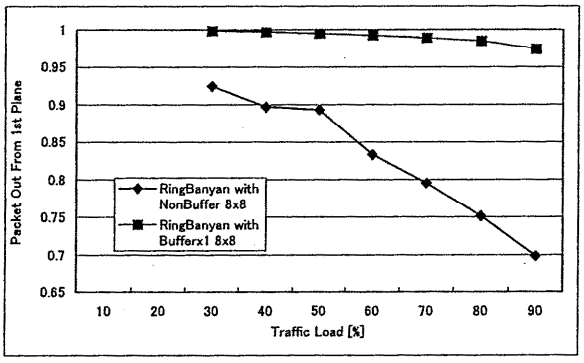


Figure 10: Proportion of packets leaving Ring Banyan after passing first Banyan plane

Considering total packet-misroute rate in Ring Banyan by varying traffic load, using MRSB switch result this feature nearly 20 times lower than using original non-buffer switch. As a result, comparing to non-buffer switch, MRSB switch enables a large proportion of packets to be switched correctly just after passing the first Banyan plane as illustrated in figure 10.

The comparison of packet loss in Ring Banyan, by using MRSB switch-element, and non-buffer switch-element, under 60% traffic load is shown in figure 11. Using MRSB switch requires only 4 Banyan planes in order to achieve packet-loss rate lower than 10^{-6}, while by originally proposed non-buffer switch, the number of Banyan planes is as large as 7. With MRSB switch, a larger proportion of packets are switched correctly after passing the first Banyan plane, as a result, less number of packets are switched to the next plane. That is why less Banyan planes are required when using our switch-element.

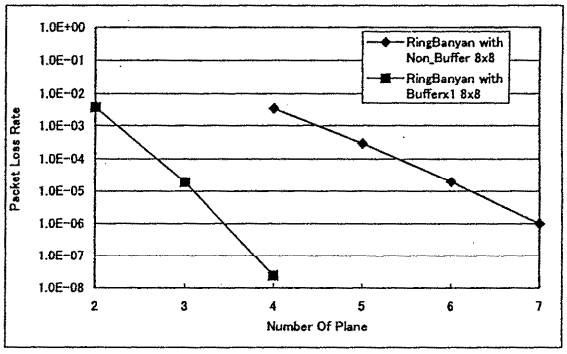


Figure 11: Packet-loss rate of Ring Banyan

The different numbers of Banyan planes also show higher performance and lower hardware cost in Ring Banyan using MRSB switch comparing to that using non-buffer switch.

4.3 Summary of Performance Evaluation

	Non-buffer Ring Banyan	1-buffer MRSB Ring Banyan	2-buffer MRSB Ring Banyan
Complexity	Excellent	Good	Bad
Packet loss	Bad	Good	Excellent
Total result	Good	Excellent	Good

Table 1:Comparison among switch-elements

From the simulation results mentioned above, the performance of Ring Banyan can be concluded as shown in table 1. Although non-buffer switch is the least complex because no buffer accessing parts are required, its packet loss rate is higher than that of others. While 2-buffer MRSB switch achieves the lowest packet-misroute rate and lowest packet-loss

rate, the complexity of implementing becomes much larger than that of 1-buffer switch. From these reasons, it is suitable to implement MRSB with one buffer.

5. Conclusion and Future Work

In this paper, we propose a new architecture of switch-element called MRSB, which can handle with variable-length IP packet directly, to enhance the performance of Ring Banyan switch. Comparing to the originally proposed non-buffer switch-element, MRSB switch-element reduces the number of misrouted packets by delaying packet into a temporary small shared-buffer. As a result, a large proportion of packets are switched correctly after passing their first Banyan plane, and this causes less number of Banyan planes are required. From this result, we expect that the total amount of hardware can be reduced from the original Ring Banyan. Taken into consideration on switch implementation issues, MRSB switch needs comparatively small buffer size, which is only 1500 bytes, while the memory management is not less complicated than a normal shared-buffer switch. The switch implementation has been working on Field Programmable Gate Array (FPGA) currently. The main components of MRSB switch includes a dual-port buffer with accessing units (WU and RU), decision making units (SIM), packet header detection units, and output selectors (SOM). Currently we are implementing a scale-down model of MRSB switch on Ring Banyan. By these architectures, we believe that up to 160 Tbps of switching fabric can be achieved.

Reference

[1] P. Newman, W. Edwards, R. Hinden, E. Hoffman, F. Ching Liaw, T. Lyon, and G. Minshall, "Ipsilon General Switching Management Protocol Specifiaction Version 1.0": Network Working Group, RFC 1987, May 1996.

[2] "Tag Switching Architecture", http://www.cisco.com/warp/public/732/tag/

[3] N. McKeown, M. Izzard, A. Mekkittikul, W. Ellersick, and M. Horowitz, "Tiny Tera: A Packet Switch Core", IEEE Micro, Jan-Feb 1997, pp26-33.

[4] N. McKeown, "Fast Switched Backplane for a Gigabit-Switched Router", http://www.cisco.com/warp/public/cc/cisco/mkt/core/12000/tech/fast_wp.pdf.

[5] C. Partridge et al., "A 50-Gb/s IP Router", IEEE/ACM Transactions on Networking, vol.6, No.3, June 1998, pp.237-247.

[6] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% Throughput in an Input-Queued switch", IEEE Transactions on Communications, 1998.

[7] T. Aoki, "Post-ATM Network Architecture", Ph.D. Thesis, The University of Tokyo, 1997.

[8] F.A. Tobagi, T. Kwon, and F.M. Chiussi, " Architecture, performance, and implementation of the tandem banyan fast packet switch", IEEE journal on Selected Areas in Communications, vol.9, October 1991.

論　文

多段接続網を用いた大容量IPデータグラムスイッチング方式

齊藤　忠夫[†]　相田　仁[†]　青木　輝勝[††]　西村　典久[†]
新保　健一[†]

A High-capacity Switching Method for IP Datagrams with Multistage Networks

Tadao SAITO[†], Hiroshi AIDA[†], Terumasa AOKI[††], Norihisa NISHIMURA[†], and Kenichi SHINBO[†]

あらまし　インターネットの急速な普及，インターネットを通じての動画像等の通信の一般化によってインターネットトラヒックは急速に拡大している．将来のマルチメディア通信のネットワークとしてインターネットが使われる可能性も高い．この場合には，インターネットのためのIPルータの大容量化が求められる．現在，IPルータとしては数Gbps程度の容量のものが実現できるが，これを超える容量を実現する手法としてはATM技術に依存する．IPトラヒックは可変長であり，パケット長も長いため，これをATMに適合することはできるが，可変長パケットを取り扱う大容量のスイッチを実現できれば，ATMによる大容量スイッチに比べ，性能を改善できる可能性がある．本論文では，1Tbps程度の容量を実現することをめざして，いくつかのIPデータグラムスイッチング方式を提案し，評価する．

キーワード　IPルータ，多段接続網，インターネット，ATM交換

1. まえがき

近年のインターネットの流行により，インターネット上を流れるトラヒックは急激に増加している．これに伴って，IPルータの高速大容量化が急務となっている．しかし，IPルータの高速大容量化は既存IPルータで用いられている原理を拡張してもその実現は難しいことから，新しい大容量パケットスイッチング技術の研究開発が必要な時期に来ている．高速大容量IPルータとしては，

- 伝送速度　155 Mbps
- 入出力ポート数　10,000

程度の大容量ルータの実現が期待されているが，このようなテラビットクラスの交換容量をもつ高速大容量IPルータを実現するにあたっては，4×4〜32×32程度のスイッチエレメントを多段接続することが考えられる．この多段接続網については既にATM交換の分野で広く研究が行われてきたが，多段接続網は大容量化が容易である反面，網内で内部ブロッキングが発生するため，高性能化を実現しにくい問題点がある．ATM交換では様々な手法により内部ブロッキングの低減を実現しているが，これらの手法をIPルータにそのまま適用することはできない．ATM交換とIPスイッチングでは固定長セルと可変長パケットの違いをはじめ，その性質が大きく異なるためである．本論文では，このような背景をふまえて，多段接続網を用いてIPデータグラムのスイッチングを行うときに生じる内部ブロッキングを回避するための二つの手法を提案する．なお，IPルータの基本機能には，

- 可変長パケット（IPデータグラム）のスイッチング機能
- IPヘッダのエラーチェック
- IPヘッダ内のTTL（Time-to-live）フィールドの更新
- IPヘッダのチェックサムの再計算
- （必要があれば）パケット（IPデータグラム）の分割
- （必要があれば）ネットワーク管理用統計データの取得
- （必要があれば）ルーチングテーブルの更新

などがある[1]が，本論文では上記のうち可変長パケットのスイッチング機能のみを検討対象とする．本論文

† 東京大学工学部，東京都
7-3-1 Hongo Bunkyo-ku Tokyo, 113–8656 Japan
†† 東京大学先端科学技術研究センター，東京都
4-6-1 Komaba Meguro-ku Tokyo, 153–8904 Japan

では，まず **2.** で既存 IP ルータ技術についてまとめるとともに，テラビットクラスの容量をもつスイッチを実現するためには多段接続網が必要であることを述べる．**3.** では，大容量 IP スイッチング方式として多段接続網を用いることの長所について述べ，更にこれまで ATM 交換技術で提案されてきた既存内部ブロッキング回避法がそのままでは IP スイッチングに適用できない理由を述べる．**4.**，**5.** では，多段接続網を用いた IP スイッチングにおいて，内部ブロッキングを回避する二つの手法（「疑似可変長交換機」及び「Down Loop 網」）の提案し，その有効性を示す．最後に **6.** で，本論文のまとめと今後の課題について述べる．

2. 既存 IP ルータ技術とその大容量化の限界

2.1 共通バス方式

共通バス方式 [2], [3] は現在最も主流となっている方式であり，ほとんどの商用 IP ルータで用いられている．この共通バス方式の簡略図を図 1 に示す．各入力ポートは round-robin 方式によって共通バスの使用権を獲得し，この共通バス上にパケットを送り出す（ブロードキャストする）．出力ポートは，このブロードキャストされたパケットのうち自分あてのものだけを受け取ることにより，スイッチングが行われる．

共通バス方式は，可変長パケットのスイッチングが容易であり，ルータを構成する回路素子数を少なくすることができるなどの利点があるものの，ルータ内のパケット損失を 0 にするためには，共通バスに（各ポートの伝送速度）×（ポート数）以上の動作速度が必要になるため，テラビットクラスの容量をもつルータを本方式で実現することは困難である．

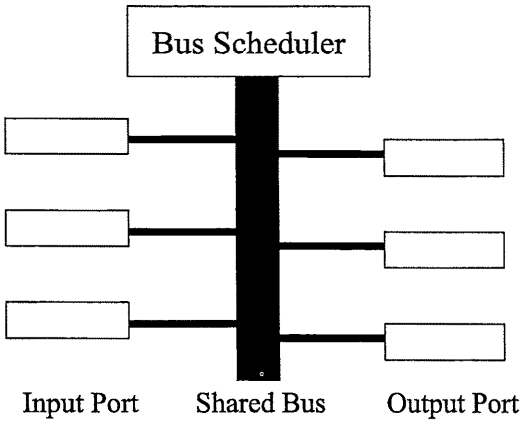


図 1　共通バス方式
Fig. 1　Shared bus method.

2.2 ATM 技術の適用

2.1 で述べた共通バス方式と並行して，近年，ATM の高速交換技術を IP ルータに取り入れる動きが生じており，現在までに IP スイッチ [4]〜[7] やセルスイッチルータ [8] などが製品化されている．しかし，ATM はもともと音声の転送を中心に考えられたためセルの長さが短いなど，インターネットにおけるデータ通信に対して適していない点も多い．更に，IP データグラムを ATM セルに分割することにより，以下のような共通バス方式の場合とは異なる新たな問題が生じる．

- IP データグラムの分割/組立てが必要 IP データグラムを ATM セルに分割/組立てする必要があるため効率が悪い．
- ヘッダ・オーバヘッドが大きい IP データグラムを ATM セルに分割することにより，20 octet の IP ヘッダに加えて各 ATM セルごとに 5 octet の ATM ヘッダを付加する必要がある．
- ATM セル損失時の再送問題 [9], [10] IP データグラムを ATM で転送する場合，1 個の IP データグラムは複数の ATM セルに分割して送られる．このうち，1 セルでも廃棄されれば，もとの IP データグラムは受信側で再構成できなくなるため，再送が行われるが，このとき損失セルだけでなく，パケット全体が再送されるため，トラヒックの増加を加速し，セル損失率が更に高くなるおそれがある．

以上のように，IP ルータに ATM 技術を適用することは有力なアプローチではあるが，IP ルータ大容量化におけるすべての問題点を解決しているわけではない．

3. 多段接続網の IP ルータへの適用

3.1 多段接続網の IP ルータへの適用

2. で述べたように，既存 IP ルータ技術を拡張してもテラビットクラスの大容量 IP ルータを実現することは困難である．そこで筆者らは，図 2 に示すように 4×4〜32×32 程度のスイッチエレメントを多段接続（Banyan 接続）することにより大容量化をはかることを考える．多段接続網において，入出力ポート数が 16×16 のスイッチエレメントを用いるとわずか 4 段で約 65,000 ポートの大容量を実現することができる．

3.2 多段接続網に対する IP トラヒックと ATM トラヒックの基本特性

本節では，最も基本的な多段接続網である Banyan

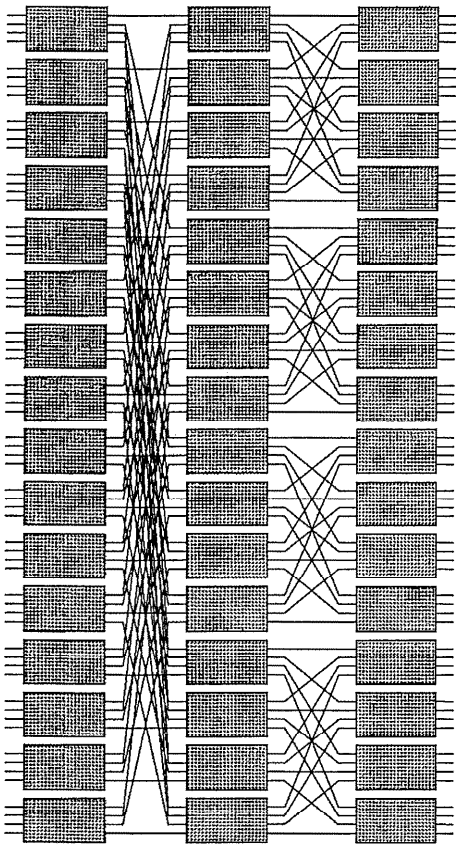


図2　4×4 スイッチエレメントを用いた Banyan 網
Fig. 2　Banyan network based on 4 × 4 switching elements.

網に対して IP データグラム及び ATM セルのスイッチングを行った場合の性能比較をシミュレーションによって行い，IP スイッチングにも多段接続網が適用可能であることを示す．

本節のシミュレーションにおけるパラメータ条件は以下のとおりである．

（1）　スイッチ条件

- スイッチ入出力ポート数 64
- スイッチエレメント入出力ポート数 4
- スイッチ内伝送速度 155.52 Mbps
- 入出力ポート伝送速度 155.52 Mbps

（2）　トラヒック条件

- 入力トラヒック量 80
- パケット長ヘッダ 20 octet（IP データグラム）データ 1～1,480 byte
- パケット長分布一様分布に従う
- パケット到着間隔指数分布に従う
- パケットあて先ランダム指定

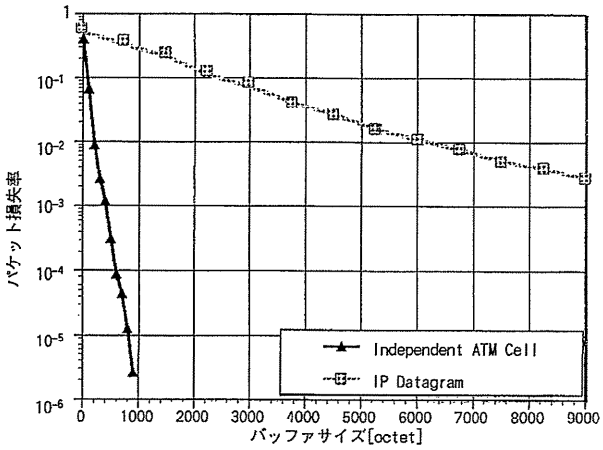


図3　IP と ATM の比較 (1)
Fig. 3　Performance Comparison between IP and ATM (1).

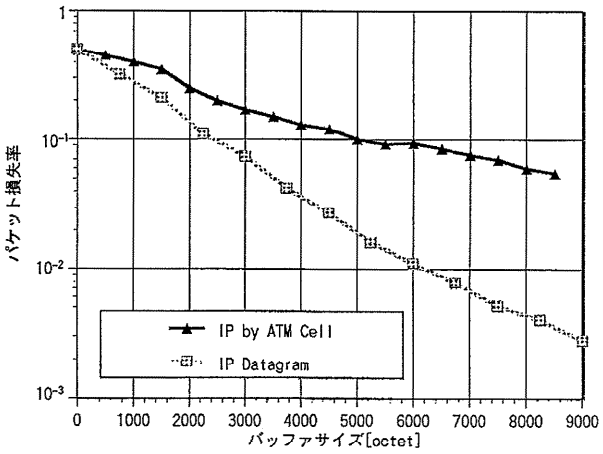


図4　IP と ATM の比較 (2)
Fig. 4　Performance Comparison between IP and ATM (2).

図 3，図 4 に IP と ATM の評価結果を示す．ただし，図 3，図 4 の横軸は各スイッチエレメント当りのバッファ量を表している．

図 3 は，IP データグラム及び ATM セルをそれぞれランダムあて先にスイッチングした場合の特性であり，この場合には確かに ATM のほうが優れた特性を示している．一方，図 4 は，IP データグラムを可変長パケットのままスイッチングする場合と ATM セルに分割してスイッチングする場合のパケット損失率特性を示している．この図 4 では，図 3 の特性とは異なり，IP データグラムをそのままスイッチングしたほうが良い特性になっている．これは，ATM セルに分割しても Banyan スイッチ内部では IP データグラム

の性質がそのまま残ってしまい，図3のような分散効果が期待できないためである．

4. 疑似可変長交換機の提案

4.1 疑似可変長交換機の構成

疑似可変長交換機は，可変長パケットを数種類のパケット長の異なる固定長パケット化することにより，可変長パケットを分割/組立てなしで固定長パケット交換機でのスイッチングを可能とする方式である．これにより，パケット同期が確立できるため，多段接続網の前段にBatcher回路を付加することが可能となり，内部ブロッキングを回避することができる．更に，**2.2**で述べたATM交換の問題点も同時に解決することができる．図5に疑似可変長交換機の構成を示す．疑似可変長交換機は，(1) セレクタ部，(2) 固定長パケットスイッチ（n平面構成），(3) 出力バッファから構成される．

（1） セレクタ部（図6） セレクタ部は，可変長パケットを固定長パケットに変換するモジュールである．スイッチ選択部は入力されたIPデータグラムのヘッダ部のTotal Lengthフィールドを読み，そのパケットを収容できる長さの固定長パケットを選択する．固定長パケット変換部は可変長パケットにダミー情報を付加して固定長パケットにする．

（2） 固定長パケットスイッチ部（図7） 固定長パケットスイッチ部は固定長パケット化された可変長パケットのスイッチングを行う．本交換機では固定長パケットを扱うので，スイッチエレメントを多段接続して大容量化をはかっても，Batcher回路を付加することによりノンブロッキングを容易に実現できる．スイッチ平面数はパケット長の分布から，各スイッチ面の負荷が均等になるように設計する．

例として，図8の実測インターネットトラヒック[15]に対しては，スイッチ平面数を3面にした場合，250, 750, 1,500 octestにすることにより，各平面に均等な負荷分散が行われることになる．

（3） 出力バッファ部（図9） 出力バッファ部は，固定長パケットからもとの可変長パケットへの変換と各固定長パケットスイッチから出力されるパケットの競

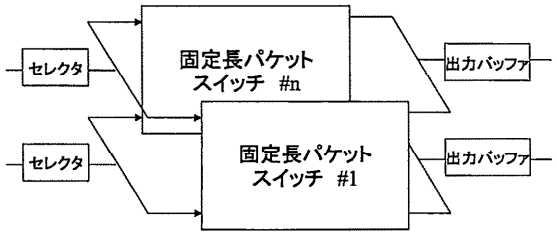


図5 疑似可変長交換機の構成
Fig. 5 Architecture of a quasi-variable-length packet switch.

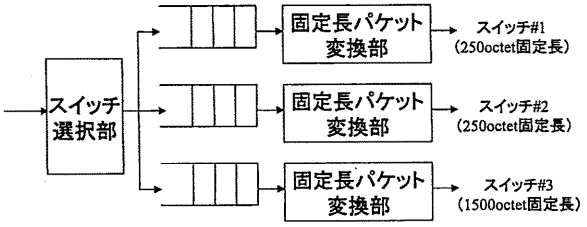


図6 セレクタ部の構成
Fig. 6 Architecture of a selector.

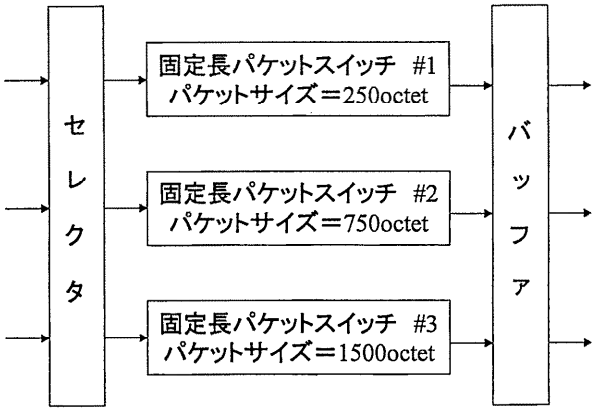


図7 固定長パケットスイッチ部の構成
Fig. 7 Architecture of a fixed-length packet switch.

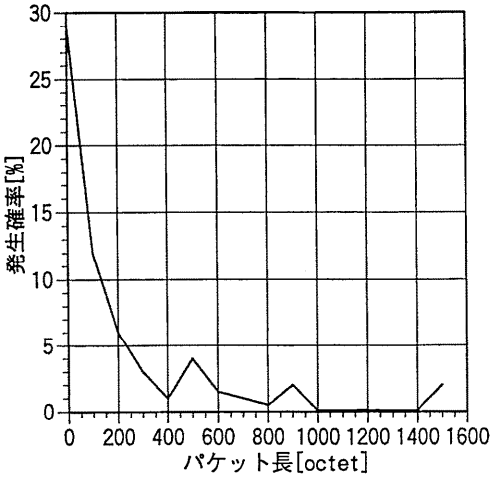


図8 インターネットトラヒック
Fig. 8 An example of the internet traffic.

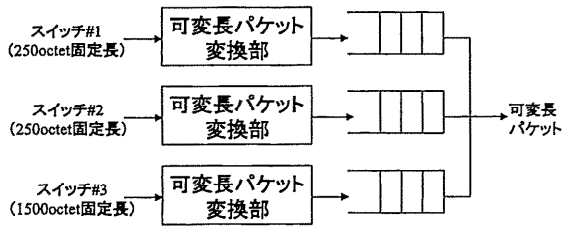


図 9　出力バッファ部の構成
Fig. 9　Architecture of an output buffer.

合整理を行う．可変長パケット変換部は固定長パケットスイッチから出力された固定長パケットを，もとの可変長パケット（IP データグラム）に変換する．図 9 中のバッファは出力ポートの競合制御を行うもので，出力ポートが空くまで可変長パケットを格納する．

4.2　疑似可変長交換機のスイッチング原理

以下に，パケット長 400 byte の IP データグラムが入力された場合の疑似可変長交換機の動作例を示す．疑似可変長交換機に入力された IP データグラムは，まずセレクタに入力する．セレクタでは初めにスイッチ選択部にて IP データグラムのヘッダ内の Total Length フィールドを読み，固定長パケット化する．固定長パケット化は，入力 IP データグラムが分割しない範囲で最小の固定長パケットを選択するため，この場合には 750 octet の固定長パケットに変換される．可変長パケットはまず固定長パケット変換待ちのバッファに入力されるが，固定長スイッチ変換部はバッファにある可変長パケットを固定長パケット時間（伝送速度 = 155 Mbps の場合，12.8 μs）ごとに読み出し，固定長パケット化して固定長パケットスイッチに入力させる．このパケットは固定長パケットスイッチでスイッチングされたのち，出力バッファ内の可変長パケット変換部にてもとの IP データグラムに変換される．そして出力ポート競合制御のための出力バッファを経由し，FIFO 順に出力する．

4.3　疑似可変長交換機の性能評価

4.3.1　シミュレーションモデル及びシミュレーション条件

本節では，疑似可変長交換機をシミュレーションにより性能評価した結果を示す．疑似可変長交換機は 3 平面構成としてあるため，比較対象として，図 2 の Banyan 網と 3 平面 Parallel Banyan 網 [16] を取り上げる．ただし，3 平面 Parallel Banyan 網では，入力トラヒックをランダムに 3 平面に振り分けることにする．また，疑似可変長交換機は Banyan 網の前段に Batcher 回路をもち，Banyan 網及び Parallel Banyan 網は直接可変長パケットを扱う．シミュレーション条件は **3.2** と同一とする．ただし，Batcher 回路の条件は，

- トラップ再帰ライン数 16
- トラップ内バッファサイズ 21,000 octet

と設定する．

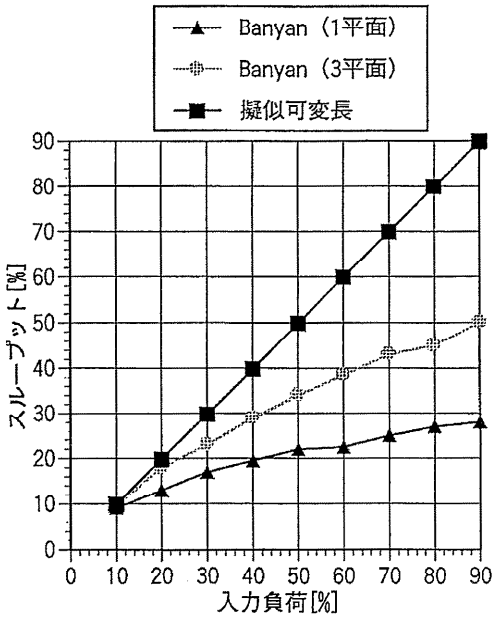


図 10　負荷—スループット特性
Fig. 10　Characteristics for offered load vs. throughput.

4.3.2　IP トラヒックにおける性能評価

図 8 に示したパケット長分布をもつトラヒックに対するスループット特性を図 10 に示す．図 10 では，疑似可変長交換機はノンブロキングが実現できるため交換機内部のパケット損失率は 0 であるのに対し，Banyan 網，Parallel Banyan 網では大きなパケット損失を生じていることが示されている．

これは，擬似可変長交換機では内部ブロッキングが発生せず，また，同一あて先のパケットが同時刻に複数個入力しても Batcher 回路のトラップによってバッファリングと再帰入力が行われるためである．

5.　DownLoop 網の提案

5.1　DownLoop 網の網構成

DownLoop 網は，経路設定を行うことなしに各入出力ポート間に複数経路を用意するための手法であ

る．Banyan 網は図 11 に示すように，小さなサイズの Banyan 網に分解することができる．小さなサイズの Banyan 網をサブ Banyan 網と呼ぶ．このサブ Banyan 網は，更に小さなサブ Banyan 網に分解することができる．それぞれのサブ Banyan 網は，Banyan 網と同じくセルフルーティングの性質をもつので，内部ブロッキングさえ起きなければ，パケットは入力された入口がどこであるかに関係なく，正しいあて先に到着する．

DownLoop 網は，この原理を利用して，各サブ Banyan 網の 1 段目のスイッチエレメントを補助リンクと呼ぶ接続線でつなぎ合わせることでループ（うかい路）を作り出す．以下に，図 12 に示す 2×2 スイッチエレメントを使用した 8×8 DownLoop 網を用いて DownLoop 網の動作を説明する．

ループは，1 段目の 4 個のスイッチエレメントをつないだものが一つと，2 段目（サブ Banyan 網の 1 段目）の 2 個のスイッチエレメントをつないだものが二つ構成される．スイッチエレメントに入力したパケットがその目的の出線に出られないとき，パケットはループの使用権を獲得し，その獲得したループを通って同じ段の下位のスイッチエレメントに入力する．図 12 では，あるスイッチエレメントで同一出線による内部ブロッキングが起きたが，あて先 “000” のパケットがループを通って次のスイッチエレメントに入力され，最終的に正しいあて先に到着していることを示している．

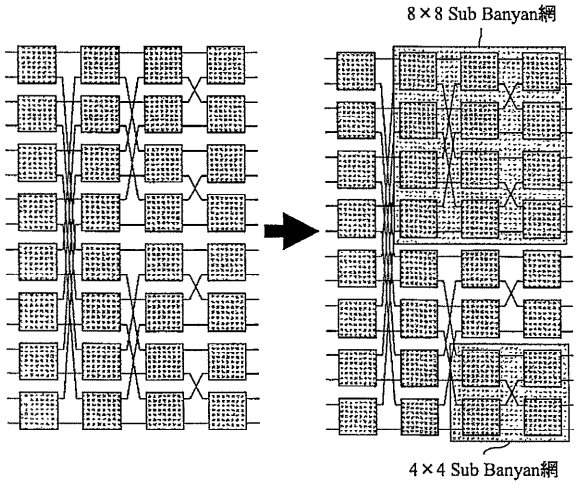


図 11　Banyan 網の性質
Fig. 11　Features of banyan networks.

このように，DownLoop 網はループによる負荷分散効果によって，内部ブロッキングを回避し，パケットの損失を防いでいる．$n \times n$ スイッチエレメントを用いる場合には，各スイッチエレメントのループを 1 本ではなく $n-1$ 本用意することによりパケット損失率の低減をはかる．この場合，$n-1$ 本のループをすべて一つ下のスイッチエレメントに接続すると，内部ブロッキングを起こした複数のパケットが再び一つ下のスイッチエレメントに入力することになり，連続して内部ブロッキングを引き起こす可能性がある．そこで，$n-1$ 本のループは図 13 に示すように接続間隔を変えてループを構成することにより，網全体の負荷分散効果を高めている．最終段では同一あて先のパケットによるブロッキングが生じるため，最終段の各スイッチエレメントには，適量のバッファを用意し，内部ブロッキング発生率が十分に小さくなるように設計する．

5.2　DownLoop 網の性能評価

5.2.1　シミュレーションモデル

本節では，DownLoop 網をシミュレーションにより性能評価した結果を示す．シミュレーションモデルは図 13 に示す 64×64 DownLoop 網を使用し，比較対象として図 2 の Banyan 網を用いる．また，シミュレーション条件は **3.2**，**4.3** と同一とする．

5.2.2　均一なトラヒックに対する損失率特性

図 14 に均一なトラヒックを与えたときの負荷–パケット損失率特性を示す．

図 14 において，DownLoop 網のパケット損失率は，Banyan 網と比較して小さくなり，特に低負荷時においては 10^{-5} 以下になる．これは DownLoop 網のループがもつ負荷分散の働きによって，内部ブロッキング

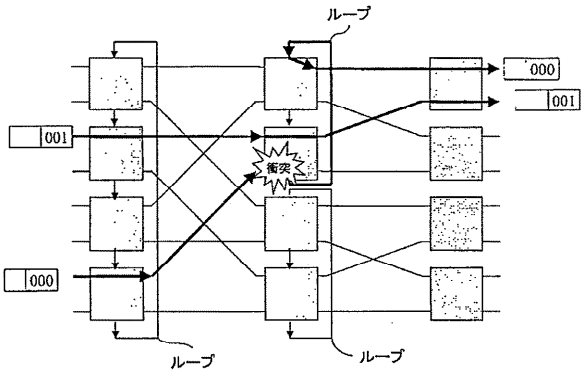


図 12　DownLoop 網の動作
Fig. 12　An example of packet routing in downLoop network.

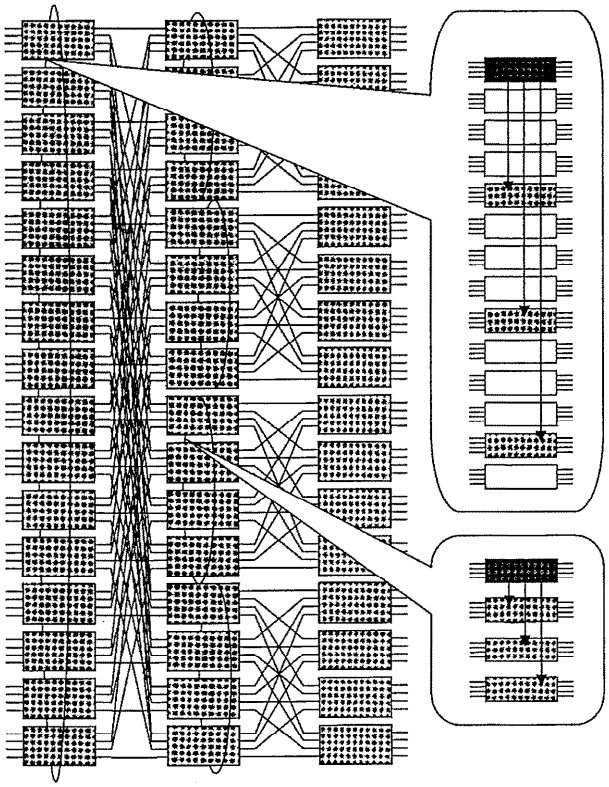


図 13　4 × 4 スイッチエレメントを用いた DownLoop 網
Fig. 13　DownLoop network based on 4 × 4 switching elements.

が回避されていることを示している．一方，入力負荷が高くなるとスイッチ内が混雑してくるため，低負荷時ほどループによる負荷分散の効果は得られない．しかしながら，

- 多段接続網内部の伝送速度を上げる
- 多段接続網を多平面用意する

などの手法 [16] により，交換機に入力するトラヒック量を相対的に下げれば，高負荷時にも良い特性が得られる．また，各スイッチエレメントにバッファ（最大パケット長 × 3 個分 = 4,500 octet）をもたせたとき，両者ともパケット損失率は小さくなるが，DownLoop 網のほうがパケット損失率の減少が大きく，Banyan 網に比べバッファリングの効果が大きいことがわかる．

5.2.3　トラヒックの入力に偏りがある場合の損失率特性

本項では，トラヒックを入力するポートを偏らせたときのパケット損失率特性として，全 64 ポート中 16 ポート（1〜16 ポート）のみにトラヒックを与えた場合の評価を行う．本シミュレーションの負荷–パケット損失率特性を図 15 に示す．

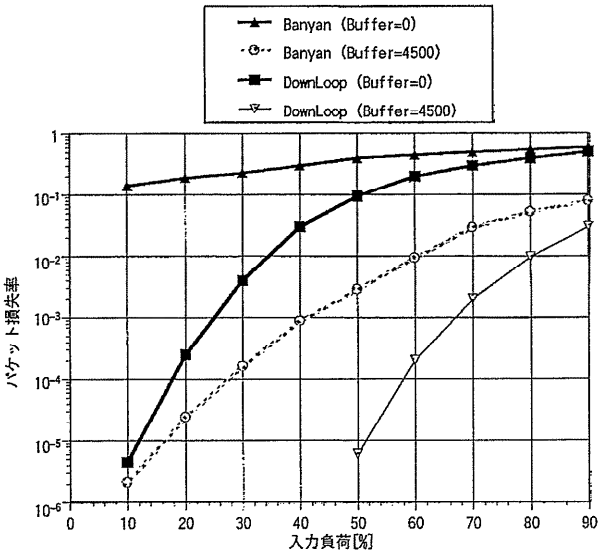


図 14　均一トラヒックに対する負荷–パケット損失率特性
Fig. 14　Characteristics for offered load vs. packet loss rate under uniform traffic.

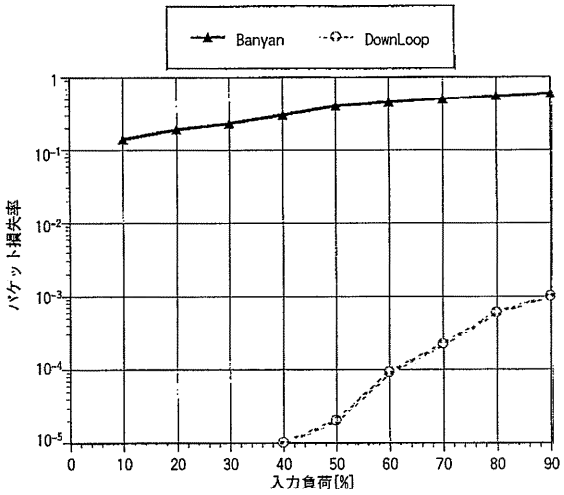


図 15　不均一トラヒックに対する負荷–パケット損失率特性
Fig. 15　Characteristics for offered load vs. packet loss rate under ununiform traffic.

トラヒックの入力に偏りがある場合，多段接続網内で使われていないスイッチエレメントが多いため，パケット分散の働きをもつ DownLoop 網が有利である．特に DownLoop 網は，高負荷時（90%）でもパケット損失率は Banyan 網の 1/100 以下になり，DownLoop 網が高い負荷分散効果をもつことがわかる．

6. むすび

本論文では，IP ルータ大容量化の一方策として，IP ルータのパケット交換部に多段接続網を適用する方法について検討を行った．多段接続網は，大容量 IP ルータを実現する方法として適しているが，パケットの同一入線によって内部ブロッキングの問題が生じるため，何らかの内部ブロッキング回避法を用いる必要がある．本論文では，この内部ブロッキングを回避するための手法として，疑似可変長交換機及び DownLoop 網を提案した．疑似可変長交換機は，可変長パケットである IP データグラムを疑似的に固定長パケットとして扱うことにより，Batcher 回路を付加して内部ブロッキング回避をはかる手法である．一方の DownLoop 網は，経路設定を必要とせずに入出力ポート間に複数経路を用意する手法である．今後は，IP ルータの交換機能以外の部分，特にルーチングテーブル参照部分の高速化についての検討を行う予定である．

謝辞 日ごろ御指導頂く東京大学工学部電子情報工学科斉藤・相田研究室の富山忠宏氏に，また，本研究を進めるうえで御討論頂いた学工学部電子情報工学科藤田博文氏，日高宗一郎氏，高橋由泰氏に深謝致します．

文　献

[1] A. Asthana, C. Delph, H.V. Jagadish, and P. Krzyzanowski, "Towards a Gigabit IP Router," J. High Speed Networks 1, 1992.

[2] I. Cidon and I.S. Gopal, "PARIS: An Approach to Integrated High-speed Private Networks," Int. J. Digital and Analog Cabled Systems, vol.1, 1988.

[3] R. Perlman, Interconnections: Bridges and Routers, Addison-Wesley, Reading MA, 1992.

[4] "The Intelligence of Routing, the Performance of Switching," Ipsilon Technical White Paper on IP Switching, http://www.ipsilon.com/produtinfo/techwp1.html.

[5] P. Newman, W. Edwards, R. Hinden, E. Hoffman, F. Ching liaw, T. Lyon, and G. Minshall, "Ipsilon Flow Management Protocol Specification for Ipv4 Version 1.0," Network Working Group Request for Comments, 1953 May 1996 http://www.ipsilon.com/protocols/rfc1953.txt.

[6] P. Newman, W. Edwards, R. Hinden, E. Hoffman, F. Ching liaw, T. Lyon, and G. Minshall, "Transmission of Flow labelled IPv4 on ATM Data Links Ipsilon Version 1.0," Network Working Group Request for Comments, 1954 May 1996 http://www.ipsilon.com/protocols/rfc1954.txt.

[7] P. Newman, W. Edwards, R. Hinden, E. Hoffman, F. Ching liaw, T. Lyon, and G. Minshall, "Ipsilon's General Switch Management Protocol Specification Version 1.0," Network Working Group Request for Comments, 1987 May 1996 http://www.ipsilon.com/protocols/rfc1987.txt.

[8] 勝部泰弘，永見健一，松澤茂雄，"セルスイッチルーター—基本コンセプトとマイグレーションシナリオ," 信学技報，IN95-145, 1995.

[9] D.D. Kandlur, D. Saha, and M. Willebeek-LeMair, "Protocol Architecture for Multimedia Applications over ATM Networks," ACM Comput. Commun. Rev., 1995.

[10] 阿野茂浩，長谷川亨，加藤聰彦，鈴木健二，"ATM 上の Available Bit Rate (ABR) を用いた TCP の実現法に関する一考察," 情処学第 51 回全大，1995.

[11] A. Huang and S. Knauer, "STARLITE: A Wideband Digital Switch," Proc. GLOBECOM'84, 1984.

[12] R.J. McMillen, "A Survey of Interconnection Networks," IEEE GLOBECOM'84, 1984.

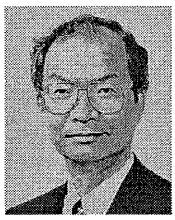
[13] G.J. Anido and A.W. Seeto, "Multipath Interconnection: A Technique for Reducing Congestion Within Fast Packet Switching Fabrics," IEEE J. Select. Areas Commun., vol.6, no.9, Dec. 1988.

[14] M. Alimuddin, H.M. Alnuweiri, and R.W. Donaldson, "The Fat Banyan ATM Switch," IEEE Infocom, 1995.

[15] A. Schmidt and R. Campbell, "Internet Protocol Traffic Analysis with Applications for ATM Switch Design," Comput. Commun. Rev., vol.23, no.2, 1993.

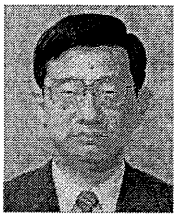
[16] U. Black, ATM: Foundation for Broadband Networks, Prentice Hall PTR, 1995.

(平成 10 年 3 月 11 日受付，11 年 6 月 17 日再受付)



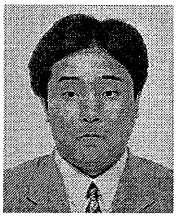
齊藤　忠夫　(正員)

昭 38 東大・工・電子卒．昭 43 同大大学院博士課程了．同大・工・講師，同大助教授，カリフォルニア工科大研究員を経て，現在，東大教授．工博．著書「電子回路入門」(昭晃堂，昭 52)，「ディジタル回路」(コロナ社，昭 57)，「計算機アーキテクチャ」(オーム社，昭 59)，「ディジタル通信網入門」(オーム社，平 1) など．昭 42 年度及び昭 61 年度本会論文賞，昭 47 年度同業績賞受賞．情報処理学会，電気学会，IEEE 各会員．



相田　　仁　（正員）

昭 55 東大・工・電気卒．昭 60 同大大学院博士課程了．同大工・電気・助手，昭 60 同大講師，SRI インターナショナル（アメリカ）客員研究員を経て，現在，東大助教授．工博．高並列推論計算機，分散オペレーティングシステム，知識処理などの研究に従事．著書「第 5 世代コンピュータ入門」（共著）．情報処理学会，人工知能学会，ソフトウェア科学会，IEEE，ACM 各会員．

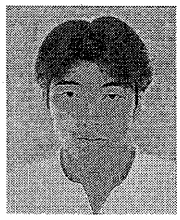


青木　輝勝　（正員）

平 5 東大・工・電子卒．平 10 同大大学院博士課程了．現在，東大先端研助手．工博．高速大容量パケット交換方式に関する研究，高速 LAN/MAN のアクセス制御方式に関する研究，高性能マルチメディアプロトコルに関する研究，超臨場感ビデオ会議システムに関する研究等にに従事．平 6 猪瀬学術奨励賞受賞．情報処理学会，IEEE 各会員．

西村　典久　（正員）

昭 62 山口県立萩工業高校電気科卒．同年富士通（株）入社．局用交換機向け試験診断ソフトウェアの開発に従事．平 8～9 東大・工・受託研究員．現在，高速大容量 IP スイッチング方式に関する研究に従事．



新保　健一　（正員）

平 5 北海道札幌工業高校電気科卒．同年（株）日立製作所生産技術研究所入社．大型計算機の電源配線設計技術の開発に従事．平 8～9 東大・工・受託研究員．現在，高速大容量 IP スイッチング方式に関する研究に従事．

４．　回線割当制御方式

○　衛星データ通信における順次予約方式
○　超高速ＬＡＮ／ＭＡＮのための高能率アクセス方式

UDC 621.394.346：621.396.946：629.783

衛星データ通信における順次予約方式

正　員　斉藤　忠夫†　　准　員　袴田　英則†

Ordered Packet Reservation System for Satellite Data Communication

Tadao SAITO†, *Regular Member and* Hidenori HAKAMADA†, *Associate Member*

あらまし　衛星を利用して比較的短いパケットの伝送を行う方式としてアロハ形の通信方式が知られている．この種の衛星通信方式ではパケットの平均伝送遅れが小さく，しかも衛星回線を能率良く使用することが望ましいが，いわゆる純アロハ方式では回線の能率が低い．本論文では回線上に予約部と情報部からなる可変長のフレームを設け，予約部ではパケットの衝突が生じないように，タイムスロットをTDMA的に割当てることを基本とした順次予約方式を提案する．この方式についてパケットの平均伝送遅れ時間，その分散，伝送誤りの影響，誤りからの回復方法などを検討し，従来提案されている各種方式との比較を行った結果，この方式は地上局数が比較的小数の場合には従来提案されている各種方式の中で最も効率が高く，しかも誤りに対しても安定であることを明らかにした．

1. まえがき

衛星を用いるデータ通信方式としてはチャネルを各地上局に固定的に割当てる固定割当て方式，制御局を設け，すべての接続要求をここに集めて回線の割当てを管理する中央管理方式なども考えられるが，メッセージが比較的短く，しかも地上局当りのトラヒックが比較的低い，コンピュータ通信のような応用においては，これらの方式では回線を能率良く使用することができない．これに対してN. Abramson らによって提案されているアロハ方式[3]では，各地上局は各地上局による回線の利用状況を知ることなく，パケット化されたメッセージをランダムに送出する．このようなランダムアクセス方式では，送出したパケットがほかの地上局で送出したパケットと衝突した場合には再送を行う．衛星を利用する場合には衝突の有無はすべての地上局でほかの地上局が送出したパケットを受信することができるために容易に検出される．この場合，静止衛星との距離を信号が往復するに要する時間（round trip time）は0.27秒であり，この時間後でなければ，衝突を知ることができないことが方式設計上の基本的な制約となる．このためアロハ方式では回線能率はたかだか18％にしかならないことが知られている．

衛星回線上でのタイムスロット同期をとるスロットアロハ[4]では回線能率は37％まで向上する．更に効率を改善するためにはタイムスロットを予約する各種の予約方式[6]～[8]があり，この場合にはいかに予約チャネルを割当てるかが問題になる．Roberts の方式[6]では予約チャネルはランダムに割当てられるため，予約チャネルでの衝突が生ずる．岩橋らの方式[7]では予約チャネルをTDMA方式で使用するが，フレーム長が一定であり，情報部のタイムスロットに無駄を生ずることがある．

本論文で提案する方式は，予約部をTDMA方式で使用してパケットの衝突を除去すると共に，フレーム長を可変として回線の利用効率を向上するものである．

本方式衛星チャネルを予約部と情報部からなるフレームに分割する．予約部には地上局数に等しい小タイムスロットがあり，各地上局に専用に割当てられる．予約部の要求によって情報部のチャネル割当てが決定される．このような原理に基づく予約形衛星パケット通信方式を順次予約方式と呼ぶ．本方式ではメッセージ長の長短によらず，比較的一様なサービスが可能であり，メッセージごとの伝送遅れ時間の分散も小さく，しかも回線の利用効率は地上局数があまり多くなければ，従来提案されているどのような方式よりも高くすること

† 東京大学工学部電気工学科，東京都
Faculty of Engineering, the University of Tokyo, Tokyo, 113 Japan
論文番号：昭54-156[D-24]

ができる．

2. 順次予約方式の概要

図1に本方式による多重アクセスの概念図を示す．本方式では，まず予約が行われ，それが完了したのちに情報が送られる．

予約部では，情報部のタイムスロットを予約する情報と先に与えられた信号に対するACK・NAK 情報の送受を行う．情報部では，この予約部によって予約されたタイムスロットを使用して，通常のパケットの送受を行う．

予約部では小タイムスロットが各地上局に割当てられているから，地上局数をN，小タイムスロットの幅を$1/V$スロット分とすると，予約部の長さはN/Vスロット必要となる．N/Vが整数とならない場合にはスロット同期を容易にするため，これよりも大きい最小のタイムスロットを予約部の長さとする．一つの予約部と情報部は一対一に対応しており，情報部の長さは，それに対応した予約部で予約された総スロット数に等しい．従って予約部の長さは常に一定であるのに対して，情報部の長さは可変である．通常，予約部と情報部は交互に送信されているが，すべての地上局において情報部の予約が全くなくなった場合には，予約部のみが続くことになる．

各地上局のスロット同期がとれており，衛星との間の伝搬遅延がすべての地上局で等しいと仮定すれば，情報を送信した地上局はその情報がすべての地上局で受信される正確な時刻を知ることができる．もし地上局の位置によって伝搬遅延が異なる場合には，伝搬遅延の調整を行う．

本方式の動作例を図1に従って説明する．各地上局が予約部R_0を使用するために$t=0$で予約パケットを送信し，1 round trip time 後の$t=10$でこれを受信したとする．R_0が受信されると，すべての地上局はすべての地上局からの情報部S_0のタイムスロットに対する要求を知ることができる．もし第ν局が第0フレームでP_ν個のスロットを予約したとすれば，第0フレームの情報部S_0では第j局はS_0の第$(\sum_{\nu=1}^{j-1} P_\nu+1)$スロットより第$(\sum_{\nu=1}^{j} P_\nu)$スロットが割当てられる．第0フレームの情報部$S_0$の長さは$(\sum_{\nu=1}^{N} P_\nu)$までとなる．

第0フレームの予約部R_0がすべての地上局で受信されると，第0フレームの情報部S_0の構成は決定されるが，これ以前のフレームの送信がまだ完了していなければ，その送信完了を待ち，次の予約部が完了してからS_0が始まる．図の例ではR_0はS_{62}の送信中に受信され，S_{63}，R_3のあとでS_0が始まる．この例ではフレームに64を周期とする番号が付けてある．この番号はフレームの予約部と情報部の対応を付け，同時に送出するACK・NAK情報の対応を付けるためのものである．この図ではフレームの並び方が$R_0 \cdot S_{61} \cdot R_1 \cdot S_{62} \cdot R_2 \cdot S_{63} \cdot S_0 \cdots$となっているが，この順番は常に一定であるわけではない．例えばR_7を受信完了した時点でR_4を受信していない場合には，次の送信はS_4ではなくR_8となる．このように同一フレームの予約部と情報部は分離して送信される．これらの間の時間差は，衛星との間の round trip time と先の予約部で予約されたタイムスロットの待ち時間との和である．

図2には予約部において各地上局が送信する予約情報の構成を示す．これはフレーム番号（6ビット），相手局番号（8ビット），予約パケット数（4ビット），又，ACK・NAK情報はこの部分が有効か無効かの表

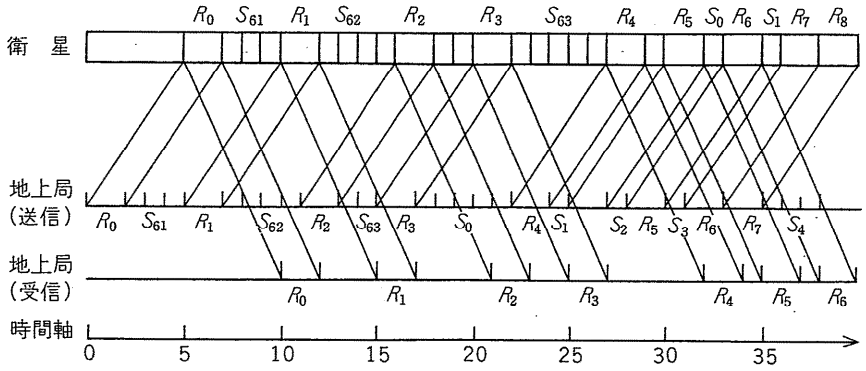


図1 順次予約方式による多重アクセスの概念図
Fig.1 -Principle of ordered packet reservation system.

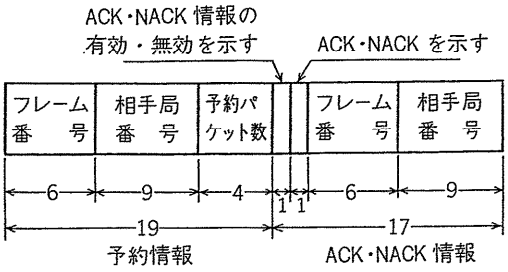


図2　符号構成
Fig.2－Contents of reservation signal.

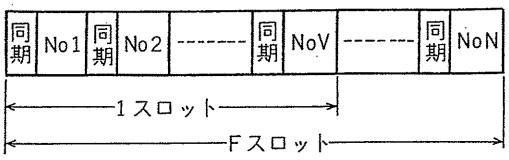


図3　予約部の構成
Fig.3－Structure of reservation part.

示（1ビット），それにACK・NAK情報に関するフレーム番号および相手局番号から構成されている．ACK・NAK部ではその相手局からそのフレームで送られてきた情報パケットのACK・NAKが示されることになる．このようにフレームごとに必ず送られる予約部を使ってACK・NAKを伝送することによって，受信確認のための短いパケットを送信する必要がなくなる．

このように予約情報は36ビットからなるが，その誤り保護のために誤り制御符号を付け，更に各々に同期用符号を付けて伝送を行う．同期用符号としては16ビット程度のパターンを用意すればよく，誤り訂正用には12ビット程度のチェックビットを持つたたみ込み符号[9]を考える．この条件で1パケット長を1,350ビットとすれば，小タイムスロット間の余裕を考慮して1スロット中に小タイムスロットは18個程度入る．すなわち$V=18$である．必要に応じて，予約情報の信頼性を向上するために予約情報を3度送りしてもよいが，この場合には$V=8$となる．

3. 伝送遅れ時間の評価

衛星パケット通信方式の重要な評価基準は，回線の使用効率とパケットの送信要求から送信完了までの伝送遅れ時間である．

これを求めるために，次の記号を使用する．

N：地上局数

V：タイムスロット長／予約用小タイムスロット長

B：1回に予約される平均パケット数

L：各地上局の呼びの生起率

σ_B^2：1回に予約されるパケット数の分散

X：平均フレーム長（s）

Z：回線容量（スロット／s）

R：round trip time（s）

F：予約部の長さ（スロット）

予約部の長さは一定であり，次のようになる．

$$F=\left[\frac{N}{V}\right]+1 \tag{1}$$

但し〔　〕はガウス記号

一方情報部の長さは，それに対応する予約により決まる．平均フレーム長は，

$$X\cdot Z=F+B\cdot N\cdot L\cdot X \tag{2}$$

であるから，次のようになる．

$$X=\frac{F}{Z-B\cdot N\cdot L} \tag{3}$$

以下，遅れ時間は情報部のスロット数を単位として求める．これを秒に換算するにはZで割算すればよい．

呼びが生起してから予約部が現れるまでの平均時間は再生理論の公式[16]から，フレーム長の分散をσ_f^2として，$X\cdot Z/2+\sigma_f^2/2X\cdot Z$である．しかし$\sigma_f^2$を理論的に評価することは困難であり，又，これによる誤差も後述のシミュレーション結果からも明らかなように小さいので，以下の評価ではこれを無視することとする．この仮定の下では，呼びが生起してから予約部に至るまでには平均して$X\cdot Z/2$スロットかかる．一つの予約情報では1個の予約しか行うことができないから，以前に生起した呼びの予約パケットが送信されていない場合には，待ち合せを生ずる．この待ち合せ時間はポラツェック・ヒンチン式により求めることができ[10]，

$$W_R \fallingdotseq \frac{S_R}{2\cdot V\cdot(1-S_R)}\cdot\frac{X\cdot Z}{1/V}$$

$$=\frac{S_R\cdot X\cdot Z}{2\cdot(1-S_R)} \tag{4}$$

となる．但しS_Rは予約用小タイムスロットの利用率であり，次のようになる．

$$S_R=L\cdot X \tag{5}$$

予約パケットを送信開始してから，それが受信完了するまでには$R\cdot Z+F$スロットかかる．従って呼びが生起してからそれに対応する予約が完了するまでの伝送遅れ時間D_Rは，次のようになる．

$$D_R \fallingdotseq \frac{X\cdot Z}{2}+W_R+R\cdot Z+F \tag{6}$$

予約が完了したのち，次の最初の情報部に至るまでには平均して$X\cdot Z/2$スロットかかる．この時点で以前に予約は完了しているが，まだ送信されていない情

報部があれば，この送信完了を待たなければならない．予約が完了して送信待ちに入る呼びの到着がポアソン生起すると仮定すれば，この待ち合せ時間は再びポラツェック・ヒンチン式により求めることができる．

$$E=\frac{B\cdot N\cdot L}{Z} \tag{7}$$

$$Y=\frac{B^2+\sigma_B^2}{B} \tag{8}$$

Eは回線効率，Yは1回に予約されるパケット数の2次モーメントを1回に予約される平均パケット数で除したものとすると，この待ち合せ時間は

$$W_S \doteqdot \frac{Y\cdot E}{2\cdot(1-E)}\cdot\frac{X\cdot Z}{X\cdot Z-F} \tag{9}$$

となる．先の予約に対応する情報部が始まってから，各地上局が情報パケットを送信完了するまでの平均時間は$(X\cdot Z-F)/2$スロットになる．従って予約完了後の伝送遅れ時間D_Sは

$$D_S \doteqdot \frac{X\cdot Z}{2}+W_S+\frac{X\cdot Z-F}{2} \tag{10}$$

となる．従って伝送遅れ時間の総和は，

$$D \doteqdot D_R+D_S \tag{11}$$

となる．

4. 誤り発生時の処理

本方式では本来パケットの衝突は生じないが，地上局間のフレーム同期がとれなくなった場合にはパケット衝突が生ずる†．これを放置すると始終衝突を起すことになり，その結果正常なほかの地上局をもつフレーム同期外れに巻き込んだり，回線効率の低下，伝送遅れ時間の増大などの悪影響が生ずる．従って，この状態からの回復方法を決めておく必要がある．

ある地上局のフレーム同期外れは，その地上局が情報部の長さの計算を誤ることにより生ずる．情報部の長さはそれに対応する予約情報により決定する．予約情報に誤りが生ずる原因は，雑音やパケットの衝突が主なものである．雑音により予約情報が誤る可能性は予約情報にバースト性雑音に強い符号を用いてあるので，十分に低いと考えられる．これに対して，フレーム同期外れを起している地上局はその地上局が送信を停止しない限り，連続して予約情報が衝突する可能性は十分に高い．従って，連続して2フレーム以上にわたり自局の送信した予約情報を正しく受信できなかった場合，その地上局はフレーム同期外れを起していると判定する．フレーム同期外れが生じたと判定すると，その地上局は自ら送信を停止する．

送信を停止したのち，その地上局は送信が継続している予約部を監視し，これに対応する情報部が解読された予約部のものと一致しておれば，自局に割当てられた次の予約用小タイムスロットから送信を開始する．この際には，送信を停止する以前に受信した予約情報はすべて捨てることにする．

地上局間のフレーム同期が全面的に失われた場合には，複数個の地上局が送信を停止する．もしすべての地上局が送信を停止することがあると，いずれの地上局もフレーム同期情報を得ることができないから，フレーム同期を回復することができなくなる．この場合にはあらかじめ指定された1局が予約信号を送り，ほかの地上局がそれに同期をとることによって，すべての地上局が同時に回復する．

以上を要約すれば，フレーム同期外れからの回復手順は次のようになる．

(1) 各地上局は，連続して2フレーム自局の送信した予約情報を正しく受信できなかった場合，送信を停止する．

(2) 送信を停止した局は，しばらくの間受信だけを続け，フレームの状態や予約の状態を完全に知ったのちに，送信を開始する．

(3) すべての地上局が送信を停止した場合，あらかじめ指定した1局が送信動作を開始する．

(4) 解読できなかった予約情報は，すべて無視する．

(5) 各地上局は，自局が予約部だとしている予約用小タイムスロットで受信された予約情報のみを解読する．但し，フレーム同期外れを検出したのちの送信停止状態では，この限りではない．

フレーム同期が失われても，この(1)〜(5)の手順を適用することにより，スロット同期が保たれている場合には回復することができる．

5. シミュレーションによる評価

3.で示した解析では伝送遅れ時間の平均値が求められたが，更に，伝送遅れ時間の分布，誤りが生じた場合の影響などを求めるためにシミュレーションを行った．

5.1 シミュレーションの条件

シミュレーション言語にはGASP Ⅱを用い，次のよ

† 衛星多重アクセスでの予約方式は，いずれの方式でも予約情報によって情報部に相当するものの存在位置が可変となるから，予約情報の誤りによってフレーム同期外れが生じ得る．この点に関しては予約方式では，いずれも特性に大幅な差はないと考えられる．

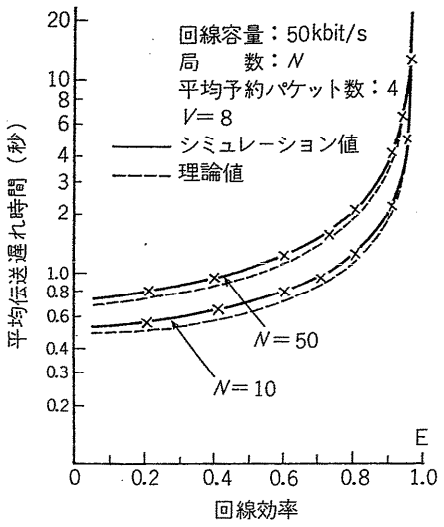


図4　平均伝送遅れ時間
Fig.4－Avarage transmission delay.

うな条件でシミュレーションを行った．

(1) シミュレーション言語：GASP Ⅱ
(2) 各地上局における呼びの生起：ポアソン生起
(3) パケット長：1,350ビット
(4) 予約用小タイムスロット長：1スロットの1／8（$V=8$）又は1／18（$V=18$）
(5) 1回に予約されるパケット数の分布：位相fのアーラン分布（$f=2$）
(6) 地上局数：10から50局程度
(7) 回線容量：50及び500 kbit／s
(8) 1 round・trip・time：0.27 s
(9) 各地上局からのトラヒック：均等

5.2　伝送遅れ時間の解析

図4は回線効率Eと平均伝送遅れ時間D/Zとの関係を示している．実線はシミュレーションの結果，破線は**3.**で示した理論式より求めた結果である．これは地上局数10，回線容量50（kbit／s），1回に予約されるパケット数4，$V=8$の場合について示してある．

図4より地上局数の増加に伴い，明らかに平均伝送遅れ時間が増大していることが分かる．図5は回線効率を0.7に固定した場合の地上局数と平均伝送遅れ時間の関係を示したものである．実線は予約情報の3度送り（$V=8$），破線は予約情報の1度送り（$V=18$）の場合について示してある．しかしながら，その差は小さなものであり，**3.**で示した理論式は妥当なものであるといえる．

図6は回線容量を変えた場合の平均伝送遅れ時間と回線効率の関係を示したものである．実線は回線容量が500 kbit／s，破線は回線容量が50 kbit／sの場合について示す．回線容量が大きくなれば，当然平均伝送遅れ時間は短くなる．この平均伝送遅れ時間をスロット時間で正規化して平均伝送遅れスロット数として考えると，平均伝送遅れスロット数への回線容量の影響は次のように考えられる．

回線容量が50 kbit／sから500 kbit／sになった場合にはスロット時間は1／10になるが，1 round trip timeは回線容量に関係なく一定であるため，単にほぼ1 round trip timeのスロット数の増加分だけ

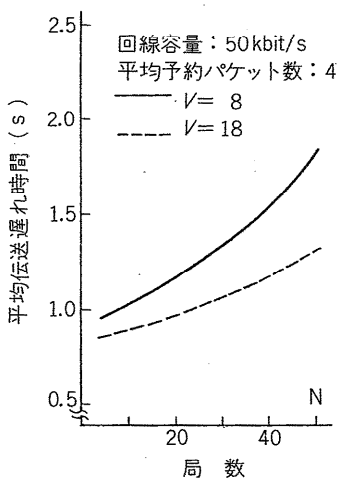


図5　局数の影響
Fig.5－Effect of the number of station.

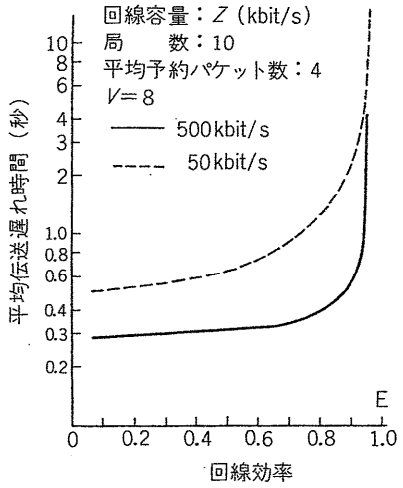


図6　回線容量の影響
Fig.6－Effect of transmission rate.

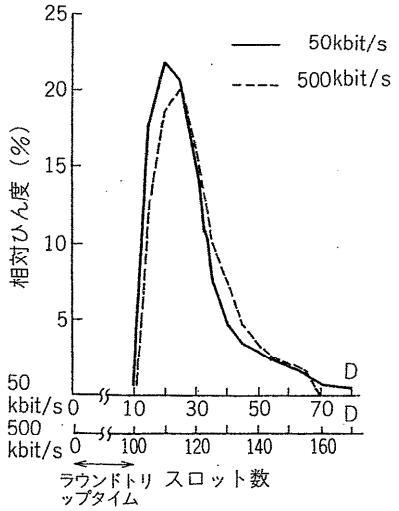


図7　伝送遅れ時間の分布
Fig.7－Distribution of transmission delay.

平均スロット数は増加する．図7は回線効率を0.6に固定した場合のスロット数の分布を示したものである．平均スロット数はほぼ1 round trip timeによるスロット数の増加分だけ増加しているが，スロット数で表した伝送遅れ時間のばらつきは回線容量によらずほぼ一定である．このばらつきはランダム予約方式[6]などに比べて非常に小さく[1]，本方式は比較的一定の伝送遅れ時間を持つことが分かる．

図8は1回に予約される平均パケット数Bの値を変えた場合について平均伝送遅れ時間と回線効率との関係を示したものである．1回に予約されるパケット数が増加すれば，明らかに平均伝送遅れ時間は増加しているが，その割合はそれほど大きなものではない．図9は回線効率が0.7の場合について，1回に予約される平均パケット数Bと平均伝送遅れ時間との関係を示したものである．

式(3)より平均予約パケット数Bが増加すれば，平均フレーム長が長くなるように思われるが，回線効率が一定であれば，平均フレーム長は一定である．式(3)と式(7)より平均フレーム長は，

$$X=\frac{F}{Z\cdot(1-E)} \tag{12}$$

となる．平均予約パケット数の影響は式(4)と式(9)に示される待ち合せ時間W_RとW_Sにだけ生ずるため，平均伝送遅れ時間と平均予約パケット数との関係は図9のような直線として示される．

図8の曲線で曲線の立上る所の回線効率は平均予約パケット数によって変るが，この理由は次のように考えられる．本方式では予約部の利用率が1を越えない限り，情報パケットの送信ができるため，

$$S_R=L\cdot X \tag{13}$$

で最大効率が制限される．この制限を考慮して，式(3)と式(7)より回線効率は，

$$E\leqq\frac{F}{1+1/B\cdot N} \tag{14}$$

となる．平均予約パケット数Bの値が大きくなれば，その最大効率は高くなることが分かる．

5.3 誤り発生時の伝送遅れ時間

4.で述べたように，フレーム同期外れが起った場合の動作は，

(1) ある特定の1局が送信停止

(2) 複数の地上局が送信停止

(3) すべての地上局が送信停止

という三つの事態のうち，どれか一つの事態が発生する．これら三つの事態に対してシミュレーションを行った結果，(1)と(2)については平均伝送遅れ時間に対する影響はほとんど無視できるほどであった．平均伝送遅れに影響を及ぼす事態は(3)の場合のみで，その結果を図10に示す．これは予約パケットの誤り率P_Eを10^{-2}及び10^{-3}とし，回線容量50Kbit/s，地上局数10，平均予約パケット数4として行ったシミュレーションの結果である．予約パケットの誤りは，ポアソン生起するものとしている．予約パケットの誤り率P_Eが10^{-2}の平均伝送遅れ時間は誤りが生じない場合のそれのほぼ2倍近いが，P_Eが10^{-3}になれば，平均伝送遅れ時間に対してほとんど影響を及ぼさないことが

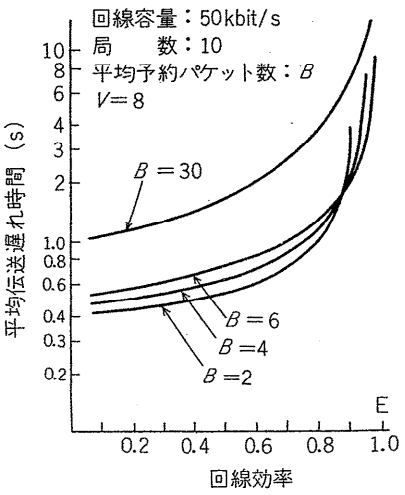


図8　平均予約パケット数の影響
Fig.8－Effect of block size.

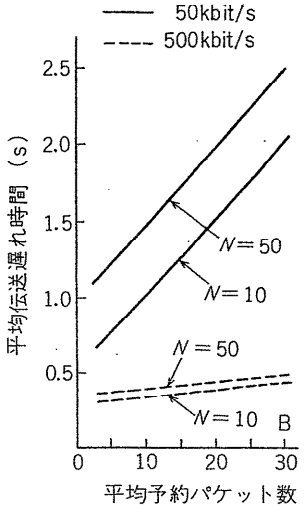


図9　平均予約パケット数の影響
Fig.9－Effect of block size.

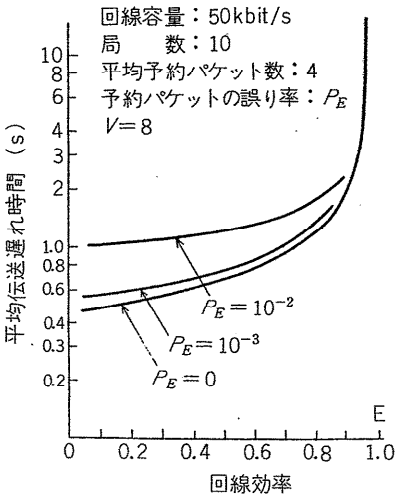


図10　誤り発生時の伝送遅れ
Fig.10－Effect of transmission error on transmission delay.

分かる．

ここで問題になるのは，予約パケットの誤り率P_Eが実際にどの程度になるのかということである．予約情報に8ビット程度のバースト誤りを訂正する能力を持つたたみ込み符号を用いた場合，予約パケットの誤り率をギルバートのモデル[12,13]を用いて計算した結果，予約情報の1度送りの場合でもビット誤り率よりも1けた程度改善される結果が得られた[15]．ビット誤り率は10^{-6}以下であることが知られているので，予約パケットの誤り率を10^{-7}以下にすることは容易である．予約パケットの誤り率は10^{-2}程度まで特性に影響を与えないから，予約パケットの誤りによる平均伝送遅れ時間の影響は無視できる．

6. 各種アクセス方式との比較

1.で述べたように，アクセス方式は各種の方式が提案されており，図11のように大きく三つに分類することができる[14]．ランダムアクセス方式はチャネルへのアクセスのアルゴリズムが簡単であるが，スループットが低く，伝送遅れ時間が大きい．無線チャネルのアクセス方式は，地上の無線通信用に考えられた方式である．衛星チャネルのアクセス方式は，衛星との伝搬遅延が大きく無視することができないので，この点に注意を払って考えられた方式である．これらの方式について，その最大効率および平均伝送遅れ時間を比較した．比較はできるだけ同一の外部条件の下で評価を行い，その条件下で最適のシステムパラメータを与えるようにした．

- ランダムアクセス方式
  - 純アロハ方式
  - スロットアロハ方式
- 無線チャネルのアクセス方式
  - CSMA（搬送波検知）方式
  - BTMA（ビジー・トーン・CSMA）方式
  - SRMA（分割チャネル）方式
- 衛星チャネルのアクセス方式
  - 順次予約方式
  - ランダム予約方式
  - スロット多重予約方式
  - SPSS（衛星パケット交換方式）

図11　各種アクセス方式の分類
Fig.11－Sort of multi－access.

表1　最大効率の比較

チャネルアクセス方式	最大効率
純アロハ方式	0.18
スロットアロハ方式	0.37
1－persistent CSMA方式	0.53
スロット1－persistent CSMA方式	0.53
non－persistent BTMA方式	
100kHz帯域幅	0.68
1MHz帯域幅	0.72
0.1－persistent CSMA方式	0.79
non－persistent CSMA方式	0.82
0.03－persistent CSMA方式	0.83
スロットnon－persistent CSMA方式	0.86
ランダム予約方式	0.83
スロット多重予約方式	0.83
順次予約方式	0.95

表1は最大効率の比較を示したものである．これは地上局数10，回線容量500kbit／s，平均予約パケット数4とし，ランダム予約方式[5]およびスロット多重予約方式[7]では5スロットごとに予約部がある（$M=5$）ものとして比較した．

図12はランダムアクセス方式と衛星チャネルのアクセス方式について，呼びの生起から情報パケットが相手局に受信されるまでの平均伝送遅れ時間を示したものである．これは最大効率の比較と等しい条件の下で比較した．順次予約方式は高負荷状態での特性が最も優れており，とりわけ回線効率が高い所まで安定した特性を示していることが分かる．ランダムアクセス方式はパケットの衝突による再送があるため，高負荷状態では平均伝送遅れ時間が大きくなっている．ランダム予約方式では予約パケットの衝突によって再送が行われるため，平均伝送遅れ時間が大きくなっている．スロット多重予約方式は順次予約方式とランダム予約方式との中間的な方式と考えられるため，特性も中間的なものとなっている．SPSSはメッセージ長によりチャネルの分割を行っているため，順次予約方式より特性が劣っている．このように順次予約方式はパケッ

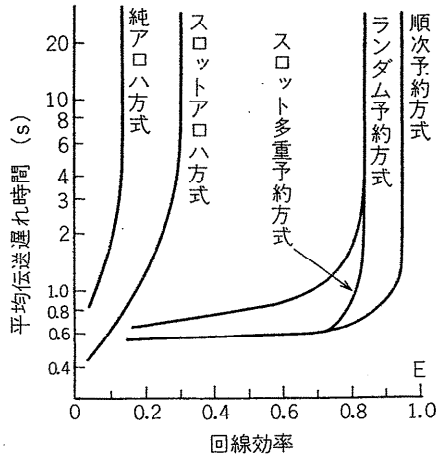


図12　平均伝送遅れの比較
Fig.12－Comparison of transmission delay.

トの衝突による再送がないこと，又フレーム長が負荷状態により自動的に最適な長さとなることなどから伝送遅れ時間の特性は最も優れている．

7. むすび

以上，予約用の小さいタイムスロットを各地上局に固定的に割当て，これによって可変長のフレームを形成して衛星チャネルを制御する順次予約衛星パケット通信方式について提案し，その原理，伝送遅れ特性，誤りに対する性質を示し，シミュレーションによってこれを確認した．この方式ではパケットの衝突が生ずることなく，又，タイムスロットの無駄を可変長のフレーム方式によって防止しているため，従来提案されている各方式に比べて極めて優れた特性を示している．又，符号誤りによって生ずるフレーム同期外れに起因するシステムの誤動作についても検討した結果，これに対する有効な対処方式があり，これによれば，予約部の誤り率が10^{-3}程度になってもほとんど特性の劣化がないことも明らかとなった．

この方式では当然のことながら純アロハ方式などに比べ地上局の装置は複雑となる．しかしその複雑さはこれより能率の低いほかの諸方式と比べ本質的には変りないものと考えている．今後，実際のシステム構成上の諸問題と共に，複雑化に伴うシステムのコスト増と利用効率向上による利益を比較するシステムの総合評価が必要である．

謝辞　本研究を進めるに当り，日ごろから御討論頂く東京大学猪瀬博教授，ならびに初期の検討に参加された上田考一，林秀一両君に感謝の意を表す．

文　献

(1) 斉藤，上田："衛星データ通信における多重アクセス方式"，信学技報，SE76-5 (1976).
(2) 袴田，斉藤："衛星パケット通信方式における順次予約方式の伝送誤りの評価"，昭52信学総全大，1458.
(3) Abramson, N. : "Packet switching with satellite", AFIPS N.C.C., p.695 (1973).
(4) Kleinrock, L. and Lam, S. : "Packet switching in a satellite channel", AFIPS N.C.C., p.703 (1973).
(5) Kleinrock, L. and Tobagi, F. : "Random access techniques for data transmission over packet switched radio channel", AFIPS N.C.C., p.187 (1975).
(6) Roberts, L. G. : "Dynamic allocation of satellite capacity through packet reservation", AFIPS N.C.C., p.711 (1973).
(7) 岩橋，岡田："通信衛星を用いたパケット交換方式に関する一考察"，信学技報，SE-5 (1974).
(8) 岩橋，中西，真田，手塚："衛星パケット通信方式"，信学技報，SE75-78 (1975-12).
(9) 宮川，岩垂，今井："符号理論"，昭晃堂 (昭50).
(10) 森村，大前："応用待ち行列理論"，日科技連 (昭50).
(11) Prisker and Kivit : "Simulation with GASPⅡ", Prentice-Hall Inc. (1969).
(12) Gilbert, E. N. : "Capacity of a burst-noise channel", Bell Syst. Tech. J., 39, 5, p. 1253 (Sept. 1960).
(13) 猪瀬，山本："データ通信"，産報 (昭46).
(14) 勅使河原："無線によるパケット通信"，情報処理，18, 12, p.1157 (昭52).
(15) 斉藤，袴田："順次予約衛星パケット通信"，信学技報，SE78-45 (1978-08).
(16) Kleinrock, L. : "Queueing System, 1, Theory p.173 (1975).

(昭和53年5月16日受付，8月24日再受付)

論　文

超高速LAN/MANのための高能率アクセス方式

学生員　森野　博章[†]　　学生員　青木　輝勝[†]
正　員　相田　　仁[†]　　正　員　齊藤　忠夫[†]

An Effective Access Control Method for High-speed LAN's and MAN's

Hiroaki MORINO[†], Terumasa AOKI[†], *Student Members*, Hitoshi AIDA[†] *and* Tadao SAITO[†], *Members*

あらまし　本論文では，伝送容量が100 Mbpsを超える超高速LANおよびMANのためのスループットの大きいアクセス方式を提案する．超高速LAN/MANのアクセス方式にはスロッテドリングが適しており，DQDB，ATMR等の方式が従来提案されているが，DQDB方式ではスループットが伝送容量を超えず，またATMR方式ではトラヒックに偏りがあるとスループットが低下するという問題がある．提案する方式では，リセット信号がリングを1周する間に各ノードが送信するスロット数をウィンドウサイズにより制限してアクセス公平性を保証する一方で，ノードの下流リンクが空いている場合は送信スロット数を制限しないため，トラヒックの状況によらず常に大きいスループットが得られる．

シミュレーションによる性能評価の結果，本方式は(1)サーバ等の特定のノードにトラヒックが集中する場合において従来方式に比べ約1.8倍のスループットが得られる，(2)映像等のバーストトラヒックに対して遅延時間を小さく抑えることができる，という点で有効であることが確かめられた．

キーワード　超高速LAN，MAN，広帯域ISDN，スロッテドリング

1. まえがき

広帯域ISDN (Integrated Services Digital Network)では音声やデータに加えて静止画像，動画像等のメディアを統一的に取り扱うことが可能であり，テレビ会議，ビデオオンデマンド等の新しい通信サービスが行われることが期待されている．各メディアの中でも特に動画像の伝送には広い帯域が必要であり，これに伴ってLAN (Local Area Network)，MAN (Metropolitan Area Network) の分野においてスループットの向上が求められている．

伝送容量が100 Mbpsを超える超高速LAN/MANのアクセス方式としては，DQDB(Distributed Queue Dual Bus)方式が知られているが，DQDBにおいては物理的伝送容量で正規化した各ノードの送信データ量の総和（これをスループットと呼ぶ）が1を超えない．これに対し最近では，ATMR，MetaRingのように最大のスループットが4程度になる方式が提案されている．しかし，これらの方式では最大のスループットはトラヒックが均質の場合にしか実現できず，特定のノードにトラヒックが集中する場合にはスループットが大幅に低下する可能性がある．スループットの低下を防止するには各ノードのトラヒックを予測して，それに適切なウィンドウサイズを割り当てればよいが，トラヒックの予測は一般に困難であり実現しがたい．

本論文では，特定のノードにトラヒックが集中した場合等の種々の場合について，従来知られている各方式より大幅にスループットを改善したアクセス方式を提案し，各方式との比較を示す．提案する方式のスループットはATMR方式等で各ノードに最適なウィンドウサイズを割り当てた場合と同じである．本方式はトラヒックの変動に応じてウィンドウサイズの最適割当てを自動的に実現できることに特徴がある．

まず，**2.**で従来提案されている方式の問題点を述べる．**3.**で新しいアクセス方式として半周内リクエスト方式を提案し，そのプロトコルを述べる．**4.**で，提案する方式をシミュレーションにより評価し，従来方式に比べて特にスループットの点で優れていることを示す．**5.**で全体についてまとめを行う．

† 東京大学工学部電子情報工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan

2. 従来提案されているアクセス方式とその問題点

従来提案されている方式はほとんどがスロッテドリングである．スロッテドリングは，

（i） スロットとしてATMセルを用いることにより広帯域ISDNと容易に接続できる．

（ii） スロットを着信ノードで解放して再利用することにより，1を超えるスループットを得ることができる．

といった特徴をもち，超高速LAN/MANに適したアクセス方式である．

しかし着信解放のスロッテドリングでは，要求スループットがシステムのスループットより大きくなると，公平性を保証するための制御を行わなければ一部のノードがスループットの大部分を占有し不公平性を生ずる場合がある．公平性を保証するための制御方式として従来提案されているものを大きく分類すると，IEEE 802.6標準案であるDQDB方式[1]に代表される分散キュー方式とATMR方式[2]，MetaRing方式[3]に代表される周期的リセット方式の2種類である．

（1） 分散キュー方式　IEEE 802.6に示された分散キュー方式は，双方向二重バスのトポロジーをもつ．送信すべきフレームをもつノードは，逆方向のバスを利用して上流ノードに送信要求を出す．この後，下流ノードからの送信要求の分の空きスロットが通過するのを待ってから自らのフレームを送信する．この方式は負荷が低いときは送信待ち時間が小さいという利点をもつ反面，

（i） 着信ノードでスロットを解放しないため，スループットが1を超えない．

（ii） 負荷が大きくなると，スロットジェネレータに近いノード程アクセスしやすく不公平が生ずる[4]．

という問題点が知られている．

（2） 周期的リセット方式　この方式は，ATMR方式やMetaRing方式に用いられている．これらの方式はリング型のトポロジーで用いられ，スロットを着信ノードで解放する．

ATMR方式では各ノードは送信を行う前にコネクションを確立し，その際に要求帯域に応じてウィンドウサイズ（WS）を決める．各ノードは自らの送信したセルの数を送信セルカウンタ（Cw）により数え，$Cw=WS$になるか，あるいは送信すべきデータがなくなった時点で送信を休止する．すべてのノードが送信を休止すると，いずれかのノードがこれを検出してリセットセルを発行し，これが周回することによりすべてのノードのCwがリセットされる．

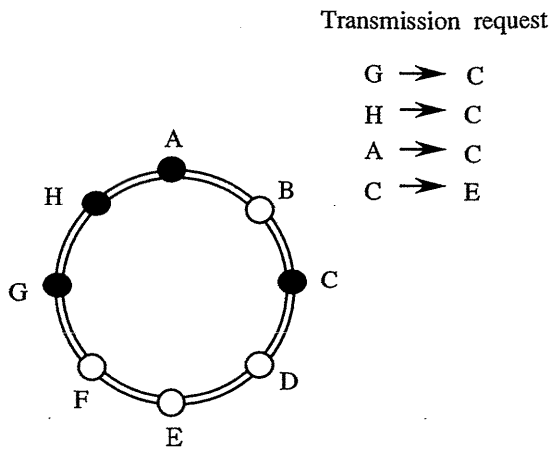


図1 リング型ネットワークにおける送信要求の例
Fig. 1 An example of transmission request in a ring network.

MetaRing方式ではデータの送信方向と反対の方向にSAT信号を巡回させる．各ノードにはSAT信号が1周する間に送信することのできるフレーム数の最大値kと最小値lが割り当てられている．各ノードはSAT信号を受信した時点で，前回SAT信号を受信したときから送信したフレーム数がl個以上の場合には直ちにSAT信号を次のノードに送信する．送信したフレーム数がl個より少なく，かつ送信すべきデータがあるときはSAT信号を保持（Hold）し，l個送信した時点でSAT信号を次のノードに送信する．

この方式では，伝送路が空いているかどうかによらず各ノードの帯域がウィンドウサイズ（またはk, l）で制限されてしまうため，特に多数のノードのトラヒックがリングのある部分に集中して，そこに多くのウィンドウサイズ（またはk, l）が割り当てられる場合に，伝送路の利用効率が小さくなってしまうという問題点がある．

例えば図1のように八つのノード（A〜H）をもつリングを考えてみる．この場合G→C，H→C，A→C，C→Eのそれぞれに常に送信すべきデータが存在するとする．このとき，単なる着信ノード解放のみで公平性制御を行わないスロッテドリングでは，図2(a)のようにGがすべての空きスロットを使用し，HとAは送信を行うことができない．Gの送信したスロットはすべてCで解放されるから，Cはやはりすべての空

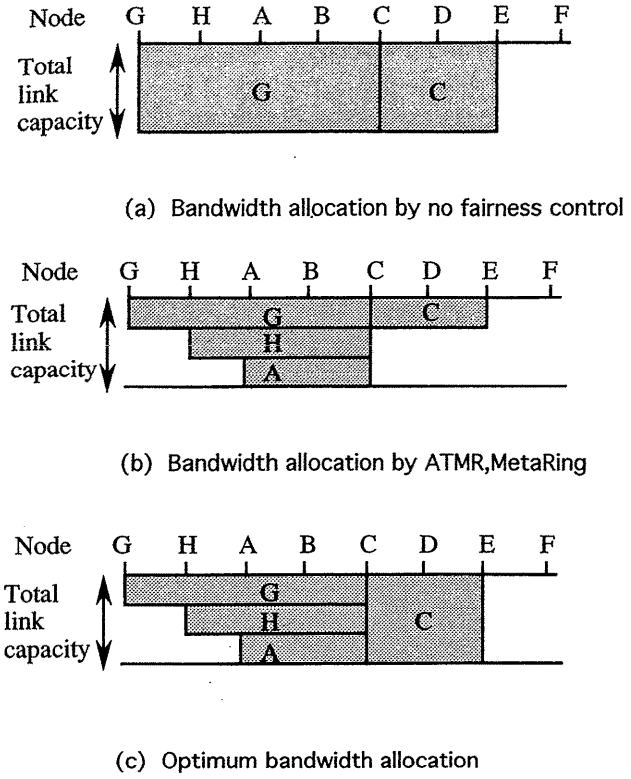


図 2　図 1 における各ノードへの帯域割当て
Fig. 2　Bandwidth allocation to the nodes in Fig. 1.

きスロットを使用できる．従って G と C だけが伝送容量に等しい帯域を使用することになる．

これに対し，ウィンドウ制御を行う ATMR および MetaRing では，すべてのノードに等しいウィンドウサイズ（または k, l）を割り当てた場合は図 2(b) のように G，H，A のそれぞれに伝送容量の 1/3 の帯域が割り当てられ，公平性が実現されるが，同時に C に割り当てられる帯域も 1/3 になる．理想的には図 2(c) のように G, H, A に 1/3 の帯域を，C に 1 の帯域を割り当てることが望ましく，これはあらかじめ C に対して G，H，A の 3 倍のウィンドウサイズを割り当てるか，または MetaRing 方式の場合には k を l に比べて大きく設定することにより実現できる．しかし，あらかじめネットワークのトラヒックを予測して C に大きなウィンドウサイズ（あるいは k, l）を割り当てる制御は困難である．

3.　半周内リクエスト方式の提案

本論文では，半周内リクエスト方式と呼ぶ新しいアクセス制御方式を提案する．上述した周期的リセット方式は，伝送路の空きにかかわらず各ノードの帯域がウィンドウサイズで制限されるのに対して，本方式では双方向二重リングの性質を利用し，あるノードから送信方向に向かってリングの半周先までのすべてのノードが送信要求をもたない場合はそのノードはウィンドウサイズの制限を受けずにスロットを送信できるため，伝送路の空きに応じてノードにダイナミックに容量を割り当てることができる．

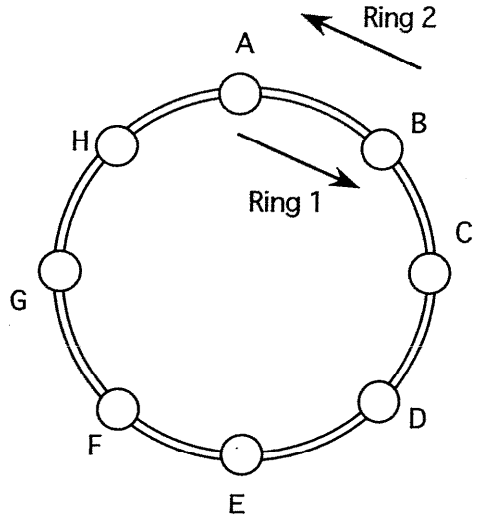


図 3　半周内リクエスト方式で考えるリング構造
Fig. 3　Ring architecture of half ring request protocol.

3.1　システムの概要

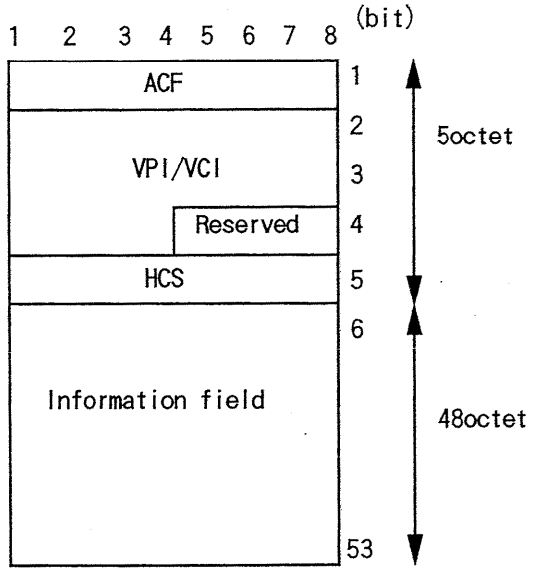
伝送路は図 3 に示すように双方向二重リングであり，時計回りのリングをリング 1 とし，反時計回りのリングをリング 2 とする．データの送信は，スロットがあて先ノードに到着するまでに経由するノードの数が少ないほうのリングを用いて行う．すなわち，図 3 のリングにおいては，A は B, C, D, E にはリング 1 を，E, F, G, H にはリング 2 を用いてデータを送信する．従って，あるノードが送信したスロットがあて先ノードに届くまでに経由するノード数（送信ノードとあて先ノードは含まない）の最大値 $H_{\max}$ は，リング上の全ノード数を N とすると，

$$H_{\max}=\lfloor N/2 \rfloor-1$$

である（但し $\lfloor x \rfloor$ は x を超えない最大の整数）．

各ノードは送信に先立ってコネクションを確立し，帯求帯域に応じてウィンドウサイズ（WS）を決める．

本方式で用いるスロットのフォーマットを図 4 に示す．スロットは ATM セルと同様に 5 オクテットのヘッダ部と 48 オクテットの情報部から成る．第 1 オクテットは ACF（Access Control Field）であり，後述するようにリングに対するアクセス制御に用いる．なお，スロットは着信ノードで解放することにする．



ACF : Access Control Field
VPI : Virtual Path Identifier
VCI : Virtual Channel Identifier
HCS : Header Check Sequence

図 4 スロットフォーマット
Fig. 4 Slot format.

3.2 公平性制御方式

3.2.1 送信要求信号の送信

各ノードは自分の送信したスロット数を送信スロットカウンタ (Cw) により数え，Cw の値に応じて次の二つの状態の間を遷移する．

READY：$Cw < WS$ であって，かつ送信すべきデータがある状態．

NOT READY：$Cw = WS$ であるか，または送信すべきデータがない状態．

各ノードはデータの送信方向と反対方向のリングのスロットの ACF を用いて，自らの状態を次のように上流に通知する．

(i) READY のとき　ACF に送信要求信号 REQ を上書きして送信要求を出す．REQ の値は，**3.1** で示した H_{max} である．

(ii) NOT READY のとき　ACF に記された REQ の値が 0 より大きければデクリメントする．

各ノードは ACF に書かれた REQ の値を見ることでデータの送信方向のリングに沿って H_{max} 個先までのノード（これを下流ノードと呼ぶ）の送信要求を知ることができる．すなわち，ACF に書かれた REQ の値が 0 より大きければ，下流ノードのいずれかが READY で送信要求を出しており，REQ の値が 0 ならば下流ノードがいずれも NOT READY で送信要求を出していないことがわかる．

3.2.2 リセット信号による送信スロット数の制御

リング 1 とリング 2 のそれぞれを使って，反対方向のリングによって伝送されるデータに対するリセット信号を一つ巡回させる．リセット信号は −1 の値をもち，ACF にこの値を上書きすることによって送信を行う．

各ノードは，リセット信号を受信するまでの間，$Cw < WS$ であればスロットを送信できる．また，$Cw = WS$ のときは反対方向のリングのスロットの ACF に書かれた REQ の値が 0 であればスロットを送信できる．$Cw = WS$ であって，かつ REQ の値が 0 より大きければ送信を休止する．リセット信号を受信したら，NOT READY の状態であれば直ちに Cw を 0 にリセットし，READY に戻って次のノードにリセット信号を送信する．READY の状態であれば，NOT READY になるまでリセット信号を保持する．そして，NOT READY になった時点で Cw を 0 にリセットし，READY に戻って次のノードにリセット信号を送信する．

本方式ではリセット信号はリングにつき一つであり，各ノードの Cw はリセット信号を受信するまではリセットされない．従って，例えば上流ノードの送信するスロットによって自ノードがウィンドウサイズ分のスロットを送信できないときは，受信したリセット信号を保持し，上流に送信要求信号を出し続けるから，上流の H_{max} 個のノードはウィンドウサイズ分のスロットを送信して休止する．双方向二重リングでスロットが着信解放されることから，この状態で自ノードに空きスロットが到着し，送信が可能になる．このようにして，各ノードはリセット信号を上流に送信した時点から次に上流に送信するまでの間に最低でウィンドウサイズ分のスロットを送信することができ，アクセス公平性が保たれる．

本方式では，各ノード i への割当帯域 BW_i と遅延時間 D は，リセット信号がリングを 1 周する時間（リセット周期）R_p と各々のウィンドウサイズ WS_i によって，次のように決まる．

$$BW_i \geq WS_i / R_p$$

$$D \leq R_p$$

リセット周期 R_p はトラヒックによって可変であるが，最大値は各ノードのウィンドウサイズの和によって決まる．すなわち，

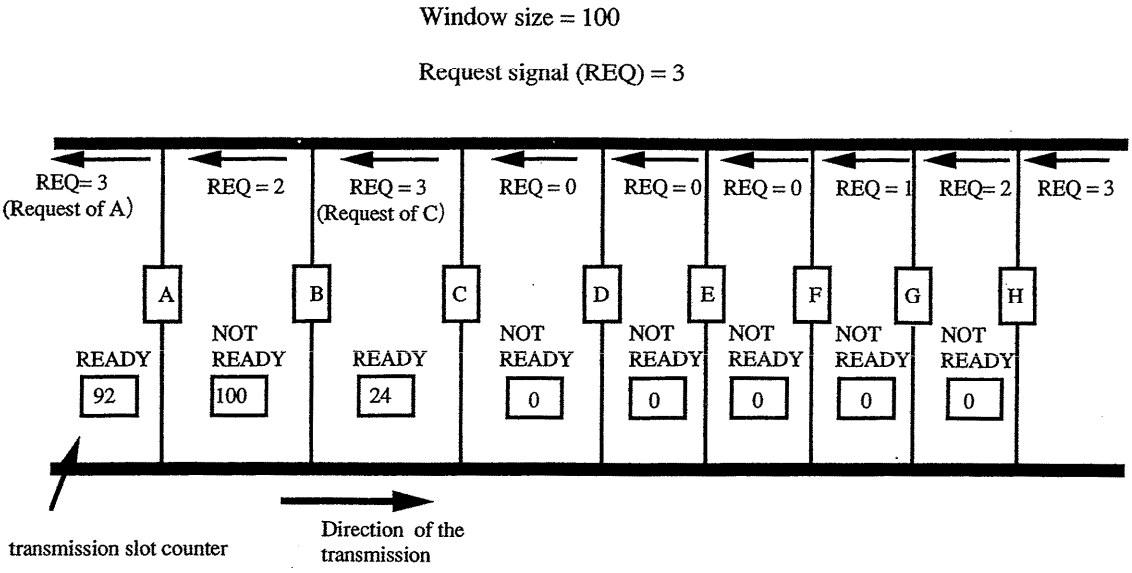


図 5　半周内リクエスト方式の動作：その 1
Fig. 5　Operation of half ring request protocol : case 1.

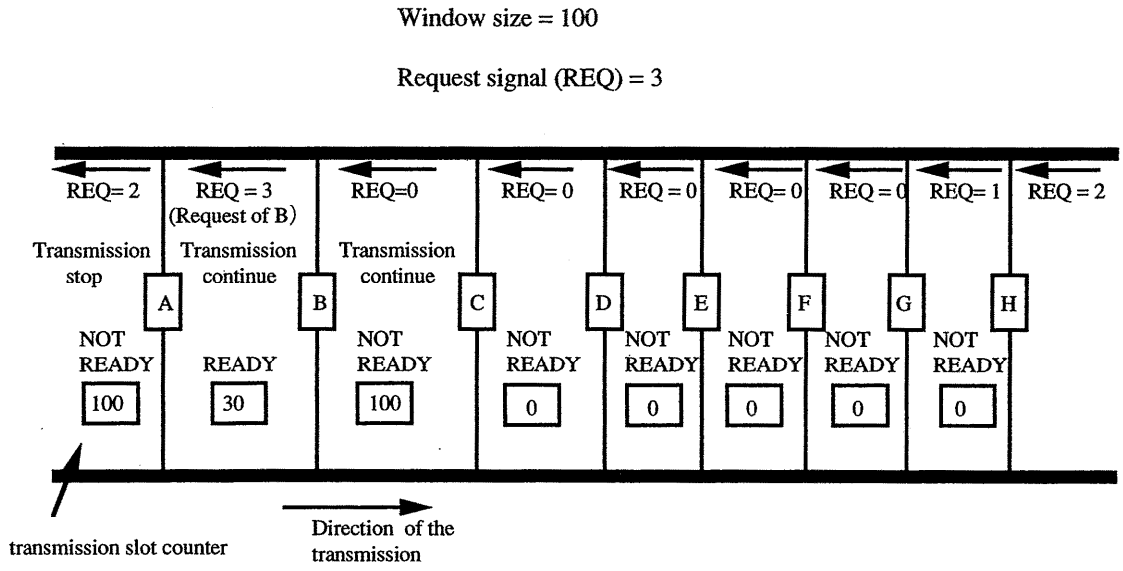


図 6　半周内リクエスト方式の動作：その 2
Fig. 6　Operation of half ring request protocol : case 2.

$R_p \leq \sum_{j=1}^{n} WS_j$　（n：ノード数）

動作例を，図 5 のノード数 8 のネットワークを例にあげて説明する．ノード A～H の中で A，B が C に，C が D に対してそれぞれ常に送信すべきデータをもち，他のノードは送信すべきデータをもたないとする．A, B, C のウィンドウサイズを 100 とする．ノード数は 8 であるから送信要求信号 REQ の値は 8/2−1=3 である．図 5 において A, C は READY であるから，スロットの ACF に送信要求信号 REQ=3 を上書きする．B は NOT READY であるため，ACF の REQ の値をデクリメントする．D～H はすべて ACF の REQ の値をデクリメントするから，1 周して C に届くスロットの ACF の REQ の値は常に 0 である．

つぎに図 6 の状況では A，C はともに 100 個のスロットを送信し終わって NOT READY の状態である，B だけが READY であって上流に送信要求を出している．A は ACF の REQ の値が 0 であり，下流ノードからの送信要求があるため送信を停止する．一方，C は ACF の REQ の値が 0 であり，下流ノードからの送信

要求がないため送信を続行する．A, C はリセット信号が自分に回わってきた時点で送信スロットカウンタを 0 にリセットし，再び READY の状態に戻る．

3.3 ノード数の通知機能

上述した制御は，リングを構成するノードの総数を各ノードが把握していることを前提としている．従って，リングの立上げ時，およびノードの追加・削除によってノード数が変化したときに，ノードの総数を各ノードに通知する機能が必要である．

ノード総数の通知は，リング内に一つ設けられたネットワーク管理ノードによって行う．リングの立上げ時には，管理ノードがノード総数を記した制御セルをブロードキャストして各ノードに通知する．ノードを追加する場合は，まず当該ノードがノードの追加を示す制御セルをブロードキャストし，管理ノードがこれを受信して新たなノード総数を記した制御セルをブロードキャストする．ノードを削除する場合も追加する場合と同様にして行う．

4. 性能評価

前述したように，従来の方式はトラヒックに偏りがあり，リングの一部に負荷が集中する傾向がある場合には性能上に問題が生ずる．本方式は，一般に従来方式と同等の性能をもち，トラヒックの偏り，負荷の集中に対しても良い性能を示す．これを明らかにするために本提案の方式，ATMR 方式，MetaRing 方式について，シミュレーションによる評価を行った．条件は各方式についてできるだけ平等になるようにしてある．

4.1 シミュレーションモデル

(i) ネットワークモデル

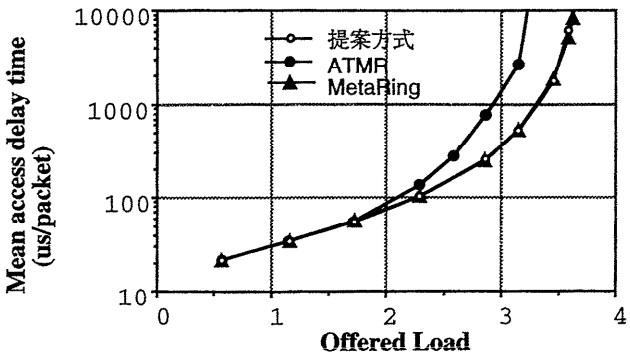
トポロジー	双方向二重リング
伝送容量	155.52 Mbps/リング
ノード数	25
ノード間距離	1 km
伝搬遅延	5 μs/km
スロットの長さ	53 オクテット
ノードでのバッファ遅延	5 オクテット

(ii) 各アクセス方式のパラメータ

提案方式 (HREQ)	ウィンドウサイズ=200
ATMR 方式	ウィンドウサイズ=200
MetaRing 方式	k=1=200

(iii) 平均転送遅延時間の信頼区間 98%

また，以下に示す結果では，負荷およびスループットは伝送路容量 (155.52×2=311.04 Mbps) で正規化している．



Maximum Throughput	
提案方式	3.82
ATMR	3.45
MetaRing	3.82

図 7 負荷・平均アクセス遅延時間特性
Fig. 7 Average access delay vs. offered load.

4.2 各ノードに均一な負荷を与えたときの負荷・遅延時間特性

負荷として，各ノードに超指数分布の間隔でパケットを発生させる．パケットの長さは，負荷の 20%が平均 50 byte，80%が平均 1,000 byte の Bimodal 指数分布に従い，あて先はランダムとする．

このときの負荷の大きさとパケットのアクセス遅延時間の関係を図 7 に示す．また，負荷の大きさを無限大にしたときのスループット（最大スループット）をその下に示す．

三つの方式を比べると，負荷の小さいところでの遅延時間はどの方式もほぼ等しい．負荷が大きくなると ATMR はリセット検出のオーバヘッドのために遅延時間が大きくなる．MetaRing および本論文で提案した半周内リクエスト方式はいずれも負荷が大きいところでも遅延時間が低く抑えられており，最大スループットも大きい．

4.3 バーストトラヒックに対する遅延時間特性

各ノードに図 8 に示したバーストトラヒックを与える．これは MPEG I で符号化した画像データの例について実データを 7 倍したものであり，33 ms ごとに 1 フレームが発生する．パケットのあて先は固定で，各ノードからリングの対角に位置する一番遠いノードに送る．このときのフレームのアクセス遅延時間の累積値を図 9 に示す．提案する方式では遅延時間の 99%が

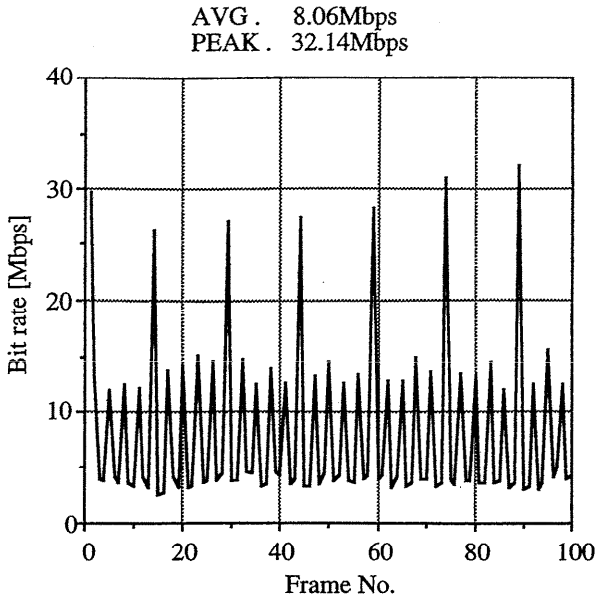


図 8 バーストトラヒックのモデル
Fig. 8 A model of burst traffic.

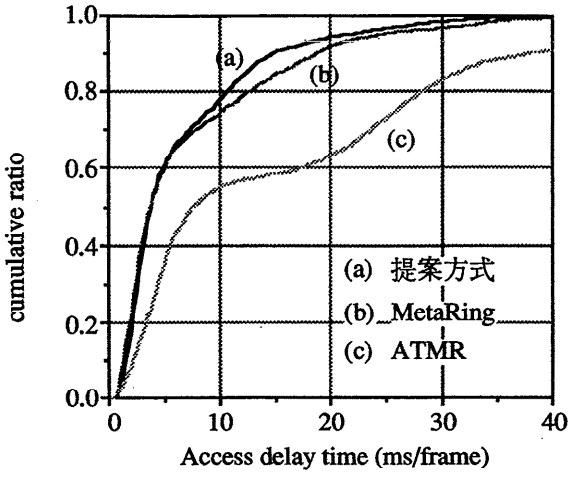


図 9 図 8 のトラヒックに対するアクセス遅延時間の累積値
Fig. 9 A cumulative access delay for traffic shown in Fig. 8.

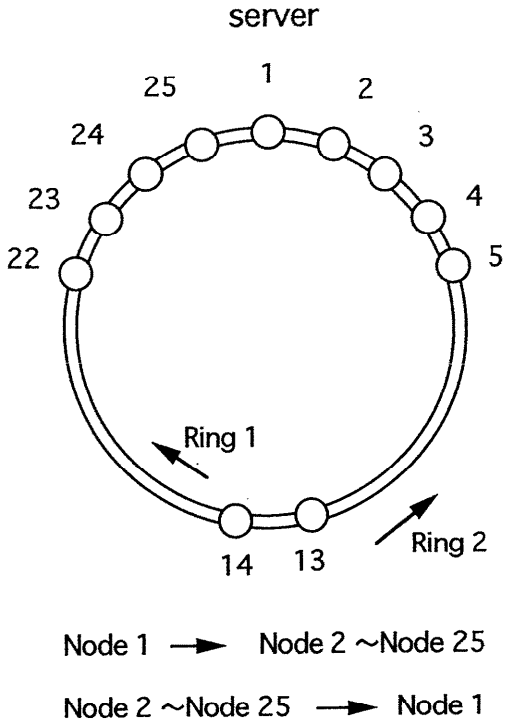


図 10 サーバ・クライアント間通信のモデル
Fig. 10 A model of communication between a server and client.

		Node No.		
		1	2～25	合計
要求帯域		1	0.5	13
割当帯域	ATMR,MetaRing	0.083	0.042	1.083
	提案方式	1	0.042	2

図 11 図 10 における各ノードのスループット
Fig. 11 Normalized throughput of each node on the condition of Fig. 10.

33 ms 以内に収まっており，従来方式のいずれよりも良い遅延時間特性を示している．

4.4 特定ノードにトラヒックが集中するときのスループット

図 10 のように，ノード 2～ノード 25 がノード 1 あてに送信し，ノード 1 がノード 2～ノード 25 あてに送信する場合を考える．各ノードには常に送信すべきフレームがあるとする．この条件でシミュレーションを行った結果を図 11 に示す．提案する方式により，ATMR および MetaRing に比べて，ネットワーク全体で約 1.8 倍のスループットが得られる．この条件は本方式の特徴が生かされる条件である．

この結果は次のように説明できる．この条件においては，リング 1 とリング 2 のトラヒックはノード 1 に対して対称であるから，リング 2 のトラヒックで説明する．全ノードのうちリング 2 を用いて送信を行うのはノード 1 から 13 である．ATMR, MetaRing 方式によってノード 1 からノード 13 のそれぞれに割り当てられる帯域 BW_i は

$BW_i=200/(200\times12)/2=1/24$

である（但し，伝送容量の 2 倍の値をを 1 とする）．ノード 1 は下流リンクが空いているにもかかわらず，リング 1 本当りでノード 2～ノード 13 と同じ 1/24 の帯

域しか使えない．

これに対し，本論文で提案する方式ではウィンドウサイズにより各ノードに上と同じ帯域がまず割り当てられる．更に，ノード1は下流ノード（ノード14〜ノード25）からの送信要求がないために通過する空きスロットをすべて使うことができ，ノード1の帯域は1/2となる．

このように，提案する方式ではトラヒックに偏りのある場合でも効率的な帯域割当てを行うことができる．

5. むすび

本論文では，超高速LAN/MANのためのスループットの大きいアクセス方式として半周内リクエスト方式を提案し，シミュレーションにより性能を評価してその有効性を示した．

提案する方式では，リセット信号がリングを1周する間に各ノードが送信できるスロットの数がウィンドウサイズによって保証され，アクセス公平性が保たれる．更に，下流ノードからの送信要求がないノードについては，ウィンドウサイズの制限を受けることなくスロットが送信できる．これにより，ネットワークの中でトラヒックに偏りがある場合でも効率的に帯域割当てが行われ，大きいスループットを得ることができる．シミュレーションの結果，本方式によりMPEG I等のバーストトラヒックに対する遅延時間を小さく抑えることができるほか，サーバやブリッジ等一つのノードにトラヒックが集中する場合に，従来知られている方式に比べ大きいスループットが得られることが示された．これらはLAN/MANにおいて極めて有効な特性である．

今後は，ノード間での優先制御，およびノードやリンクの障害が起こったときの動作について検討する予定である．

謝辞 本研究を行うに当り，御指導頂いた東京大学工学部電子工学科齊藤・相田研究室の金炳錫氏に感謝致します．

文　献

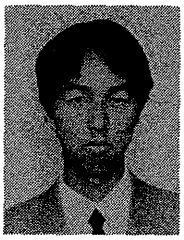
(1) IEEE Standards for Local and Metropolitan Area Networks: Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN) (1990).

(2) 伊東　匡，笠原英樹，森田直季，今井和雄："ATMリングのアーキテクチャと高速マルチメディアネットワークへの適用"，情処学会マルチメディア通信と分散処理研資，50-12 (1991-12).

(3) Cidon I. and Ofek Y.: "MetaRing—A Full Duplex Ring with Fairness and Spatial Reuse", IEEE Trans. Commun., **41**, 1, pp. 110-120 (Jan. 1993).

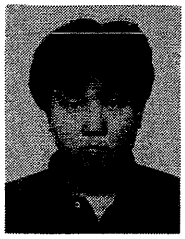
(4) Davids P. and Martini P.: "Performance Analysis of DQDB", IEEE Comput. Soc., International Phoenix Conference on Computers and Communications, 22, pp. 548-555 (1990).

(平成6年12月28日受付，7年3月30日再受付)



森野　博章

平6東大・工・電子卒．現在，同大大学院工学系研究科電子工学専攻修士課程在学中．LANのアクセス制御方式に関する研究に従事．



青木　輝勝

平5年東大・工・電子卒．現在，同大大学院修士課程在学中．次世代高速LANに関する研究に従事．平6猪瀬学術奨励賞受賞．

相田　仁

昭55東大・工・電気卒．昭60同大大学院博士課程了．同大・工・電気工学科助手，昭61同講師．SRIインターナショナル(米国）客員研究員を経て，平2東大・工・電子・助教授．工博．コンピュータネットワーク，不均質分散処理システム，超並列コンピュータ等の研究に従事．著書「第5世代コンピュータ入門」(共著)．情報処理学会，人工知能学会，ソフトウェア科学会，IEEE，ACM各会員．

齊藤　忠夫

昭38東大・工・電子卒．昭43年同大大学院博士課程了．同大・工・講師，同助教授，カリフォルニア工科大研究員等を経て，現在，東大教授．ディジタル通信方式，コンピュータネットワーク，分散処理等の研究に従事．工博．著書「電子回路入門」(昭晃堂，昭52)，「ディジタル通信網入門」(オーム社，平2)．昭42年度，昭61年度論文賞，昭48年度業績賞等受賞．電気学会，情報処理学会，IEEE各会員．

5.　プロトコル工学

○　オートマンモデルによるＨＤＬＣ　プロトコル製品検証の一方式
○　多層構造プロトコルをもつネットワークの伝送効率の評価
○　計算機網におけるウィンドウ方式　フロー制御の効果の一評価
○　トランスピュータによる多層プロトコルの性能評価用シミュレータ

UDC 681.324.078:681.325.001.42:519.713.4

オートマトンモデルによるHDLCプロトコル製品検証の一方式

正員 斉藤 忠夫† 正員 加藤 聡彦† 正員 猪瀬 博†

A Product Validation System for HDLC Protocol by Means of an Automaton Model

Tadao SAITO†, Toshihiko KATO† *and* Hiroshi INOSE†, *Regular Members*

あらまし 計算機網において，製品を網に接続する場合に網の定めた標準プロトコルに従っているかを確認することを製品検証と呼ぶ．本論文では，プロトコル機械を順序機械としてモデル化し，順序機械の同定の手法を用いてプロトコルの製品検証を行う方法をHDLC-BAプロトコルの例について検討した．プロトコル機械への入力には，外部入力，上位レベルからの入力，タイマ入力の3種類が存在する．製品検証に当っては，プロトコル機械を，テスト入力を発生するテスト機械に接続し，テスト入力に対する応答によって判定を行う．上位レベルからの入力を発生するには上位レベルのプログラムを作成しなければならず，検証を受けるプロトコル機械に対する追加のプログラミングの負担が生ずる．本論文では，これを考慮し上位レベルからの入力を与えることができる場合と，上位レベルからの入力を最小限に留めた場合の両者に分けて検証方法の検討を行った．

1. まえがき

コンピュータ間通信のためのネットワーク・アーキテクチャとプロトコルの標準化は多方面から注目されている．プロトコルの標準化に当っては，そのプロトコルが無矛盾であることを検証することは極めて重要であり，この意味でのプロトコル検証については多くの研究が行われている[(1)~(4)]．

プロトコルが一度標準化されると，それに従って製品が製造されるが，このとき製造者は製品のプログラムのテストを行う．これに加えて製品を網に接続する場合には，網サービス側がその製品が網の定めた標準プロトコルに従っているかを確認する必要がある．この確認は，製品製造側のプログラムテストとは違い，接続試験の形態をとる．これをプロトコル製品検証と呼ぶことにする[(5)]．

プロトコルの製品検証に当っては，原則として標準プロトコルのすべての場合について，製品が正しく動作することを確認する必要がある．本論文においては，プロトコルの製品検証をオートマトンモデルに従って系統的に実行する方法について検討し，これをリンクレベルプロトコルのHDLC-BAに適用した例について報告する．

2. プロトコル製品検証の形態

本論文においては，製品検証は，被検証プロトコルがインプリメントされている端末をプロトコルテスタと呼ぶテスト用ノードと接続して通信を行い，テスタからのテスト系列に対する被検証プロトコル機械の応答を観察することにより行われると考える．

プロトコル機械は順序機械とみなすことができるから，その製品検証に順序機械の故障診断の手法を用いることができるはずである．しかし，プロトコル機械への入力は，通信相手から送られるフレーム（パケット）のみではなく（これを外部入力と呼ぶ），上位レベルのプロトコルからの命令（これを上位レベルからの入力と呼ぶ）も存在する．このような入力による状態遷移は，通信相手からみると伝送する信号とは独立に自律的に生じるように見える．更に，プロトコル機械の入力には同一のレベルからの入力であるタイマ入力

†東京大学工学部電子工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan
論文番号：昭 55-416[D-94]

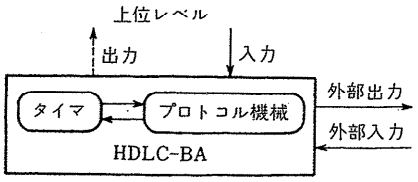


図1　プロトコル機械とその入出力
Fig.1-Protocol machine and its inputs/outputs.

がある．このタイマの起動はプロトコル機械からの出力によって行われる．このように考えれば，プロトコル機械への入力は外部入力，上位レベルからの入力およびタイマ入力の3種類が存在することになる．

ここで考えている製品検証の環境では，プロトコルテスタからの外部入力は容易に与えることができるが，上位レベルからの入力は被検証システム内に上位レベルのプログラムを作らなければ発生できないものが多い．

製品検証に当っては，この条件を満足できる場合とできない場合がある．すなわち，(i)上位レベルからの入力を与えることができる検証と，(ii)外部入力だけによる検証とに分けて，製品検証法を検討する必要があろう．いずれの場合も観測される出力は外部出力のみとする．タイマからの入力は，外部入力あるいは上位レベルからの2入力によって付随的に生ずることになるから，いずれの場合も考慮される．

(i)の検証では上位レベルからの入力をプロトコルテスタの入力と同期的に生じさせるために，図2のように検証用タスクと呼ぶテスト用プログラムをプロトコルプログラムにリンクさせる．プロトコルテスタから送信する信号のデータとして次に生ずべき上位レベルからの入力を指示し，検証用タスクがそれを解読し必要な場合は指示された上位レベルからの入力を起こすか，又は，はじめにどのようなテスト系列を用いるかをデータとして送信して被検証タスクに記憶させ，その後にテストを始め，検証用タスクはテスタからの入力を観察し必要な時点で上位レベルからの入力を起こすなどの方法をとることにする．

(ii)の方法では上位レベルからの入力を与えないから検証用タスクは与えない．

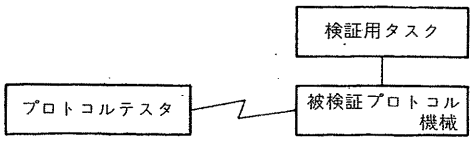


図2　上位レベルからの入力を用いる製品検証の形態
Fig.2-Validation scheme using input from higher level.

3. 検証用テスト系列の構成

プロトコル機械は上位レベルからの入力，外部入力，タイマからの入力が与えられたとしたときには，完全定義で強連結の順序機械となる．従って，上位レベルからの入力も含め，すべての入力が適切なタイミングで与えることができるとすれば，順序機械の一般論[6]に従って，次のようにして検証用テスト系列を作ることができる．

プロトコル機械を順序機械M

$$M=(S,I,O,\delta,\omega) \tag{1}$$

S：状態集合，I：入力集合，O：出力集合

δ：$S\times I\rightarrow S$　状態遷移関数

ω：$S\times I\rightarrow O$　出力関数

とみなし，次のようにテスト系列を構成する．但し，$|S|=n$，$|I|=m$とし，この順序機械の判定系列を$\tilde{x}_D$とする．又，S_iの$\tilde{x}_D$による遷移先を$S_i^{(l)}$とし，このときの出力を$\tilde{z}_i$とする．更に，S_iからS_jへ遷移させる遷移系列を$\tilde{x}(S_i,S_j)$で表す．

順序機械はS_1から出発するものとし，まず次のような入出力応答によりn個の状態を識別する（但し，出力は判定系列に対する部分のみを明記する）．

$$\left.\begin{array}{l}\tilde{x}:\tilde{x}_D\tilde{x}(S_1^{(l)},S_2)\tilde{x}_D\cdots\cdots\tilde{x}(S_{n-1}^{(l)},S_n)\tilde{x}_D\\ \tilde{z}:\tilde{z}_1\text{———}\tilde{z}_2\cdots\cdots\text{———}\tilde{z}_n\end{array}\right\} \tag{2}$$

次に状態をS_1に移す．

$$\left.\begin{array}{l}\tilde{x}:\tilde{x}(S_n^{(l)},S_1)\tilde{x}_D\tilde{x}(S_1^{(l)},S_n)\tilde{x}_D\tilde{x}(S_n^{(l)},S_1)\\ \tilde{z}:\text{———}\tilde{z}_1\text{———}\tilde{z}_n\text{———}\end{array}\right\} \tag{3}$$

更に，各入力x_iによる$S_1\rightarrow S_1^{(1)}(i)$という遷移と，そのときの出力を調べて状態を$S_1$にもどすという操作を続ける．

$$\left.\begin{array}{l}\tilde{x}:x_1\tilde{x}_D\tilde{x}(S_1^{(1)(l)}(1),S_n)\tilde{x}_D\tilde{x}(S_n^{(l)},S_1)x_2\tilde{x}_D\\ \tilde{z}:z_1(1)\tilde{z}_1^{(1)}(1)\text{———}\tilde{z}_n\text{———}z_1(2)\tilde{z}_1^{(1)}(2)\\ \quad\tilde{x}(S_1^{(1)(l)}(2),S_n)\tilde{x}_D\tilde{x}(S_n^{(l)},S_1)\cdots\cdots x_m\tilde{x}_D\\ \text{———}\tilde{z}_n\text{———}\cdots\cdots z_1(m)\tilde{z}_1^{(1)}(m)\end{array}\right\} \tag{4}$$

次に状態をS_2に移す．式(2)において$\tilde{x}(S_1^{(l)},S_2)$による$S_1^{(l)}\rightarrow S_2$の遷移が確認されているから，

$$\left.\begin{array}{l}\tilde{x}:\tilde{x}(S_1^{(1)(l)}(m),S_1)\tilde{x}_D\tilde{x}(S_1^{(l)},S_2)\\ \tilde{z}:\text{———}\tilde{z}_1\text{———}\end{array}\right\} \tag{5}$$

各入力による遷移に関しては式(4)と同様である．又，状態S_3以下についても同様である．

以上のような入出力応答は，状態数の分かっている順序機械の状態遷移および出力を一義的に決定するから，インプリメントされたプロトコルが規定された状態数以下の状態を有するという制限の下では，どのよ

うな誤りも検出可能である.

4. HDLC-BAの状態遷移表

この章ではHDLC-BAの検証用の状態遷移表を規定する.

HDLC-BAの仕様としては, その記述が明確に行われている電電公社のパケット交換網のための仕様書を利用した[7]. この仕様では, HDLC-BAについてその動作が状態遷移表とそれに続く処理状態遷移表で記述されている.

HDLC-BAプロトコル機械の状態は, 状態遷移表で規定されている状態(制御状態)とシーケンス番号に対応する送受信状態変数などの組と考えることができる. しかし一般にプロトコルでは, 制御状態はシーケンス番号の個々の値には影響されないから, 制御状態に関する処理とシーケンス番号に対する処理を分けて別々に検証する. 本論文においては, 制御状態に関する処理の部分の製品検証についてのみ考察することにする[8].

HDLCのプログラミングに使用する状態遷移表の中の状態は2大別できる. 第1の状態は, Pビット"1"受信後のように, フレームを受信してから, 受信されたフレームによって要求されたレスポンスを送信するまでの一時的な状態(この場合はFビット"1"のレスポンスを送信するまで)である. 第2の状態は, 安定な状態であり, これからは上位レベルあるいは外部からの入力がないときは他の状態に遷移することはない.

プロトコルの製品検証においては, 外部入力および上位レベルからの入力に対するプロトコル機械の応答を検査すればよいから, 一時的な状態を扱う必要はない.

本論文で使用するプロトコル機械の状態遷移表は, 文献(7)に示された状態遷移表から一時的な状態を除いて書き改めたものである.

プロトコル機械の検証では, 入力相互のタイミングにも問題がありうる. 例えば, ビジー発生後 RNR コマンド送出以前にビジーが解除された場合の扱いも検査する必要があろう. 又, 3者以上の入力がほぼ同時に生ずることも考えられる. このような場合は本論文の範囲から除いてある.

これについては, 状態遷移要因としてある入力の直後に他の入力が生ずるような入力の対を表の中に追加することにすれば, 検証項目の中に加えることもできよう.

表1にはこうして安定な状態となるHDLC-BAの状態の種類を, 表2に入力の種類を示す. 次に表3に製品検証用の状態遷移表の一部を載せる. この表には遷移先と出力が記述されている. 表の0印は状態が変化しないこと, 又, 斜線の欄はその入力を無視することを示す. この表には or で表記した非決定的な部分が存在するが, これは記述されている遷移のうち一つが生ずるものと解釈する.

又, Pビット"0"の情報フレームを受信した場合, その受信確認のためにRR, RNR などが送信される場合とされない場合とがあるが, これは(RR), (RNR)と括弧を用いて表している.

5. 上位レベルからの入力を用いる製品検証

4.で規定したHDLC-BAの上位レベルからの入力を用いる製品検証のためのテスト系列を求める.

表3の遷移表の判定系列として,

ビジー解除, IP($R \neq N(S)$), データ送信要求, REJF, DM$\bar{F}$, IP($R=N(S)$)を用いる.

表1 HDLC-BAプロトコル機械の状態の種類

状態	動作モード設定不可	DM受信待ち	UAレスポンス受信待ち	Fビット"1"受信後								Pビット"1"送信後								RNRコマンド未送信	DISCコマンド未送信	UAレスポンス受信待ち	再送で回復できない誤り
				相手局非ビジー				相手局ビジー				相手局非ビジー				相手局ビジー							
				自局非ビジー		自局ビジー		自局非ビジー		自局ビジー		自局非ビジー		自局ビジー		自局非ビジー		自局ビジー					
				$N(S)$ 正常	REJ 誤り	$N(S)$ 正常	REJ 誤り	$N(S)$ 正常	REJ 誤り	$N(S)$ 正常	REJ 誤り	$N(S)$ 正常	REJ 誤り	$N(S)$ 正常	REJ 誤り	$N(S)$ 正常	REJ 誤り	$N(S)$ 正常	REJ 誤り				
略称	N	AA	AC	B01	B02	B04	B05	B11	B12	B14	B15	B41	B42	B44	B45	B51	B52	B54	B55	CA	CB	CC	DA

表2 HDLC-BAプロトコル機械の入力の種類

入力	略称	入力	略称	入力	略称
通信開始	STAT	I$\bar{P}$($N(S)=R$)受信	$I=$	REJ$\bar{F}$受信	REJ
相手動作モード設定待ち	WAIT	I$\bar{P}$($N(S) \neq R$)受信	$I<>$	UAF受信	UAF
動作モード再設定	REST	IP($N(S)=R$)受信	$I=P$	DMF受信	DMF
動作モード終了	END	IP($N(S) \neq R$)受信	$I<>P$	DM$\bar{F}$受信	DM
強制動作モード終了	FEND	RRF($N(R) \geqq J$)受信	RRF	FRMRF受信	FRMRF
ビジー発生	BH	RRF($N(R) < J$)受信	RRF+	FRMR$\bar{F}$受信	FRMR
ビジー解除	BC	RR$\bar{F}$受信	RR	再送で回復できない誤り$F=1$	ERRF
タイムアウト	TOUT	RNRF受信	RNRF	再送で回復できない誤り$P/F=0$	ERR
N_2回連続してタイムアウト	N_2T	RNR$\bar{F}$受信	RNR	再送で回復できない誤り$P=1$	ERRP
データ送信要求	DATA	REJF受信	REJF		

表3　HDLC-BAの製品検証用の状態遷移表（部分）　但し，TS：タイマスタート．TP：タイマストップ

入力＼状態	N	AA	AC	B01	B02	B04	B05	B11	B12	B14	B15	B41
FEND		N		CC/ DISCP TS	CC/ DISCP TS	CC/ DISCP TS	CC/ DISCP TS	CC/ DISCP TS	CC/ DISCP TS	CC/ DISCP TS	CC/ DISCP TS	CB
TOUT			0/ SABMP TS	B41/ RRP TS	B42/ RRP TS	B44/ RNRP TS	B45/ RNRP TS	B51/ RRP TS	B52/ RRP TS	B54/ RNRP TS	B55/ RNRP TS	0/ RRP TS
DATA				0/I TS or B41/IP TS	0/I TS or B42/IP TS	0/I TS or B44/ I,RNRP TS	0/I TS or B45/ I,RNRP TS					0/ I TS
I<>P	0/ DMF			B02/ REJF	0/ RRF	B05/ RNRF	0/ RNRF	B12/ REJF	0/ RRF	B15/ RNRF	0/ RNRF	B42/ REJF
RRF												B01
REJ				0/I TS or B41/IP TS	0/I TS or B42/IP TS	0/I TS or B44/ I,RNRP TS	0/I TS or B45/ I,RNRP TS	B01/I TS or B41/IP TS	B02/I TS or B42/IP TS	B04/I TS or B44/ I,RNRP TS	B05/I TS or B45 I,RNRP TS	0/ I TS
DM		AC/ SABMP TS		AC/ CABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS

入力＼状態	B42	B44	B45	B51	B52	B54	B55	CA	CB	CC	DA
FEND	CB	CB	CB	CB	CB	CB	CB				CC/ DISCP TS
TOUT	0/ RRP TS	0/ RNRP TS	0/ RNRP TS	0/ RRP TS	0/ RRP TS	0/ RNRP TS	0/ RNRP TS	CB/ RNRP TS	0/ RNRP TS	0/ DISCP TS	0/ FRMR TS
DATA	0/ I TS	0/ I TS	0/ I TS								
I<>P	0/ RRF	B45/ RNRF	0/ RNRF	B52/ REJF	0/ RRF	B55/ RNRF	0/ RNRF				0/ FRMRF TS
RRF	B02	B04	B05	B01	B02	B04	B05	CB/ RNRP TS	CC/ DISCP TS		0/ TP
REJ	0/ I TS	0/ I TS	0/ I TS	B41/ I TS	B42/ I TS	B44/ I TS	B45/ I TS	0	0		
DM	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS	AC/ SABMP TS			N/ TP	AC/ SABMP TS

この判定系列によるテスト系列の一部を図3，4に示す．図3が状態を識別する系列であり，図4が各状態からの入力による遷移，出力を確認する系列である．ここで2列に並んだ系列の上が被検証プロトコルに加えるテスト用の入力系列であり，下がそれに対応する出力系列である．又，状態遷移表には非決定的な部分があるから入出力系列の対も幾とおりか存在する場合もあるが，ここではそれらを並べて表すことにする．但し，非決定的な部分を用いるのは判定系列，各入力を加える部分のみで，二つの状態間を遷移させる入力系列は非決定的な部分だけを使って構成する．タイムアウト，N_2回タイムアウトのタイマ関係入力は各状態に応じて発生方法を工夫する必要がある．このため，すべての状態について**2.**で示した方法による検証を終了してからタイマ関係の入力について検証するものとする．タイマ関係の入力は**2.**で述べた方法における判定系列，遷移系列としては使用しない．まず，状態N，AAにおいてはタイマを起動することができないためタイムアウトを発生できない．又，状態CAでは1回しかタイムアウトを発生できないため，N_2回連続してタイムアウトという入力は発生できない．

次に，状態AC，B41からB55，CA，CB，CCはその状態にいるとき必ずタイマが動作しているから，他の入力と同様にしてタイムアウトを発生できる．

更に，B01からB15の各状態ではタイムアウトが発生するとPビット“1”のコマンドが送信され状態がB41からB55に遷移する．状態をB01からB15にもどすためには，RRFなどを加える必要があるが，HDLC-BAの規定によるとFビット“1”のレスポンスを受信した場合は，連続タイムアウト回数を示す変数を0にするために，B01からB15の各状態においてはN_2回連続してタイムアウトを発生できない．

又，タイムアウトを発生するためには，状態B01からB05については各々の状態でデータ送信要求を発生させタイマを起動する．このときI，$\bar{P}$が送信された場合をテストに用いる．状態B11からB15については上述の方法でタイマを起動し，その後RNRを加える．

以上のようにして発生させたテスト系列は，全体で

```
X: B C   I<>P  DATA  REJF  DM    I=P  WAIT
Z:       DMF                    DMF
                                          N の判定

X: B C   I<>P  DATA  REJF  DM    I=P  UAF
Z:                         SABMP
                                          AA の判定

X: B C   I<>P  DATA  REJF  DM    I=P
Z:
                                          AC の判定

X: {B C  I<>P  DATA  REJF  DM    I=P  UAF  I<>P
Z:  }    REJF  I           SABMP          REJF
X: {B C  I<>P  DATA  REJF  DM    I=P
Z:  }    REJF  IP    I     SABMP
X: {B C  I<>P  DATA  REJF  DM    I=P
Z:  }    REJF  IP    IP    SABMP          B01 の判定

X: {B C  I<>P  DATA  REJF  DM    I=P  UAF  B H   RRP
Z:  }    RRF   I           SABMP          RNRP
X: {B C  I<>P  DATA  REJF  DM    I=P
Z:  }    RRF   IP    I     SABMP
X: {B C  I<>P  DATA  REJF  DM    I=P
Z:  }    RRF   IP    IP    SABMP          B02 の判定

X: {B C  I<>P  DATA  REJF  DM    I=P  UAF  B H   I<>   RRF
Z:  }RRP REJF  I     I     SABMP          RNRP  (RNR)
X: {B C  I<>P  DATA  REJF  DM    I=P
Z:  }RRP REJF  I     IP    SABMP          B04 の判定

X: {B C  I<>P  DATA  REJF  DM    I=P  UAF  RNR
Z:  }RRP RRF   I     I     SABMP
X: {B C  I<>P  DATA  REJF  DM    I=P
Z:  }RRP RRF   I     IP    SABMP          B05 の判定
```

図3 状態の判定のためのテスト系列
Fig.3-Test sequences for identification of states.

```
X: B C   I<>P DATA REJF DM   I=P        UAF
Z:

X: FEND       B C  I<>P DATA REJF  DM   I=P       STAT
Z: DISCP                                DMF       SABMP

X: B C   I<>P DATA REJF DM   I=P        UAF
Z:

X: {RR        B C  I<>P DATA REJF  DM   I=P
Z:  }              REJF I            SABMP
X: {RR        B C  I<>P DATA REJF  DM   I=P
Z:  }              REJF IP   I       SABMP
X: {RR        B C  I<>P DATA REJF  DM   I=P
Z:  }              REJF IP   IP      SABMP

X: B C   I<>P DATA REJF DM   I=P        UAF
Z:

X: RNR        B C  I<>P DATA REJF  DM   I=P
Z:                 REJF             SABMP
```

状態 B01 に関する遷移の確認のためのテスト系列

```
X: B C    I<>P  DATA  REJF  DM     I=P          UAF
Z:

X: {DATA          TOUT        B C    I<>P  DATA   REJF  DM    I=P
Z:  }I            RRP               REJF  I      I     SABMP
X: {DATA          TOUT        B C    I<>P  DATA   REJF  DM    I=P
Z:  }I            RRP               REJF  I      IP    SABMP
```

状態 B01 におけるタイムアウトに対する遷移の確認のためのテスト系列

図4 遷移の確認のためのテスト系列
Fig.4-Test sequences for validation of transitions.

12,411個の状態遷移要因から成る．このうち外部入力は8,166個である．これは網ら的なテストとして十分に実用的な長さである．

6. 上位レベルからの入力を考慮しない製品検証

6.1 上位レベルからの入力を考慮しない製品検証の問題点

上位レベルからの入力を用いない検証には次のような問題がある．

(1) 上位レベルからの入力を生ずることができないので，遷移先と出力を確認するのは外部入力を与えた場合に限られる．但し，検証を進めるに当り，どうしても必要な上位レベルからの入力（リンクの設定，切断の命令など）は最小限発生させる必要がある．後述のように実際のテスト系列では全体で3,274個の入力の内，上位レベルからの入力はデータ送信要求を除き3個のみである．

(2) 予想される上位レベルからの入力による状態遷移についてはそれが生じた場合の対処はできるが，予期しない上位レベルからの入力による遷移についてはそれが難しい．従って，上位レベルからの入力による遷移はその遷移，出力がすべて正しいと仮定する．

(3) ある状態に外部入力を加えたときの遷移先からもとの状態にもどす場合に上位レベルからの入力を用いなければならない場合は検証が難しい．

そこで，外部入力強連結という概念を定義することにする．状態 S_i と S_j が外部入力強連結であるとは，外部入力 $\tilde{x}$ と $\tilde{x}'$ が存在して，

$$S_i \xrightarrow{\tilde{x}} S_j \quad \text{且つ} \quad S_j \xrightarrow{\tilde{x}'} S_i$$

が成立することである．

プロトコルの状態集合を外部入力強連結な部分集合に分割する．そして同じ集合内の状態に遷移させるような外部入力についてのみ検証を行う．

(4) 判定系列は，ある状態からそれと外部入力強連結の関係にある状態にだけ遷移させるような入力の集合から選択しなければならない．すべての状態を識別できる前述の判定系列は，この条件を満足しないので，各状態ごとに判定系列を求める．この判定系列の集合を Dset と呼ぶ．Dset に含まれる判定系列は，ある状態を識別する判定系列をその状態に加えたときの固有の出力系列が他の状態にその判定系列を加えた場合，たとえ制御できない上位レベルからの入力があっても固有の出力系列と同一の出力系列を生ずることがないようにしておく必要がある．

(5) 上位レベルからの入力による遷移が非同期に生じた場合は，それが通信相手，すなわちプロトコルテスタで観測されなければならない．そうでなければどの状態に関して検証を行っているか不明になってしまう．

6.2 検証用テスト系列

HDLC-BAの上位レベルからの入力を考慮しない製

```
X:  I=P     RRF    DM     I=P    DM                UAF
Z:                                   ACの判定

X:  RR      I<>P   REJF           RRF
Z:  IP      REJF   IP                B01の判定

X:  RR      I<>P   REJF           I=        RNRF
Z:  IP      RRF    IP             (RR)
                                     B02の判定

X:  RNR     I<>P   REJF
Z:          REJF                     B11の判定

X:  RNR     I<>P   REJF           I=P       REJ
Z:          RRF                   RRF       IP   B12の判定
```

状態の判定のためのテスト系列

```
X:  DATA          RR     I<>P   REJF          DMF
Z:  IP            I      REJF   IP            SABMP

X:  I=P     RRF   DM     I=P    DM            UAF
Z:

X:  I=            RR     I<>P   REJF          DMF
Z:  (RR)          IP     REJF   IP            SABMP
X:  I=            RR     I<>P   REJF
Z:  RRP           I      REJF   IP

X:  I=P     RRF   DM     I=P    DM            UAF
Z:

X:  I<>           RR     I<>P   REJF          DMF
Z:  REJ           IP     RRF    IP            SABMP
X:  I<>           RR     I<>P   REJF
Z:  REJP          I      RRF    IP
```

状態B01に関する遷移の確認のためのテスト系列

図6　S_{B1} に対するテスト系列
Fig.6-Test sequences for S_{B1}.

```
X:  RR      REJF          I<>      RRF
Z:  IRNRP   IRNRP         (RNR)
                             B04の判定
X:  RR      REJF          I=       RNRF
Z:  IRNRP   IRNRP         (RNR)
                             B05の判定
X:  RNR     REJF          I<>P
Z:                        RNRF
                             B14の判定
X:  RNR     REJF          I=P      REJ
Z:                        RNRF     IRNRP
                             B15の判定
X:  RR      REJF          I<>
Z:  I       IRNRP         (RNR)  B44の判定
```

状態の判定のためのテスト系列

```
X:  DATA          RR     REJF          I<>     RNR
Z:  IRNRP         I      IRNRP         (RNR)

X:  RNR     REJF          I=       RRF
Z:          IRNRP         (RNR)

X:  I=            RR     REJF          I<>     RNR
Z:  (RNR)         IRNRP  IRNRP         (RNR)
X:  I=            RR     REJF
Z:  RNRP          I      IRNRP

X:  RNR     REJF          I=       RRF
Z:          IRNRP         (RNR)

X:  I<>           RR     REJF          RNR
Z:  (RNR)         IRNRP  IRNRP
X:  I<>           RR     REJF
Z:  RNRP          I      IRNRP
```

状態B04に関する遷移の確認のためのテスト系列

図7　S_{B2} に対するテスト系列
Fig.7-Test sequences for S_{B2}.

了，強制動作モード終了，ビジー発生，DAでは動作モード再設定，強制動作モード終了である．このうちB41からB52の各状態における動作モード終了，強制動作モード終了とDAにおける動作モード再設定を除き，RNR又はDISCが送信されるから検出可能である．B41からB52の各状態における動作モード終了，他については判定系列中のREJFにより観測される．又，DAにおける動作モード再設定入力に対してはSABMPが送出されるから検出可能である．

S_{B1} に対するテスト系列は**3.**で述べた方法で構成される．これを図6に示す．表記法は**5.**と同様である．

(4)　S_{B2} について

このグループの状態に遷移させるには，例えば，状態B01において上位レベルからの入力のビジー発生を生じさせなければならない．

S_{B2} の各状態に対する外部入力だけからなる判定系列は存在しない．これは，B04とB05などの$N(S)$正常とREJ誤りの状態を区別できないためである．これを区別することなしに検証を進めることはできる．B04,B05,B44,B45の各状態に対する判定系列として，

RR, REJF

をB14, B15, B54, B55の各状態に対する判定系列として，

RNR, REJF

を用いる．

この場合はDMF, DM, FRMRF, FRMR, ERRF, ERR, ERRP以外の外部入力とタイムアウト，データ送信要求について検証可能である．

又，S_{B2} の各状態からの遷移を生じる上位レベルからの入力は，動作モード終了，強制動作モード終了，ビジー解除である．このうちB04からB15の各状態における強制動作モード終了とビジー解除については各々DISCP, RRが送信されるから観測できる．動作モード終了，B44からB55の各状態における強制動作モード終了については判定系列中のREJFにより観測される．

S_{B2} に対するテスト系列も S_{B1} と同様に構成される．これを図7に示す．

(5)　S_{CA} について

状態をCAに遷移させるには，B41からB55の各状態のうちいずれかにおいて動作モード終了を生じなければならない．

状態CAの判定系列として，

$I=P$, RRF　（出力　RNRP）

を用いる．これは状態をCBに遷移させる．従って，状態をCAから遷移させない外部入力を加え，その後に判定系列を加えCAにいることを確認する方法をとる．

このようにして検証させるのは外部入力

$I=, I<>, I=P, I<>P$, RR, RNR, REJ, UA, DM

による遷移である．

この部分のテスト系列は，

$\{x_i\}$ x_i：上記，　$I=P$, RRF

となる．ここで最後にCBの判定系列がないのは，以下に示すようにCBの判定系列が状態をCCに遷移さ

せるからである．

(6) S_{CB} について

CBの判定系列として，

$I=P$, RRF　（出力　DISCP）

を用いる．これは状態をCCに遷移させる．従って，CA同様，状態を変化しない入力についてのみ調べる．検証できる外部入力はCAと全く同じである．

この部分のテスト系列は

$\{x_i\}_{x_i}$：上記，$I=P$, RRF, CCの判定系列

となる．

(7) S_{CC} について

CCの判定系列として

$I=P$, タイムアウト　（出力　DISCP）

を用いる．これは状態を変化させない．

この場合は，

$I=$, $I<>$, $I=P$, $I<>P$, RRF, RRF+, RR, RNRF, RNR, REJF, REJ

の外部入力とタイムアウトについて検証可能である．

この部分のテスト系列は

$\{x_i\}_{x_i}$：上記，$I=P$，タイムアウト

となる．

以上のように構成したテスト系列は全体で3,274個の入力から成り，このうち状態を遷移させるために必要な上位レベルからの入力は3個である．

7. 検証方法の評価

いままでに述べてきた製品検証方式の評価として，実際のプログラムのダイナミックステップ数，使用ひん度などを用いたものも考えられるが，ここでは簡単に次のように定める．製品検証においては，各状態にそれぞれの入力を加えて遷移先と出力を確認するが，この確認がどの程度の割合で行われているかを評価基準とする．これを検証率 R_V と呼ぶ．すなわち，

$$R_V=\frac{\text{遷移，出力を確認した状態と入力の組の数}}{\text{状態数}\times\text{入力数}}$$

と定める．

上位レベルからの入力を用いる製品検証では，状態N, AAでタイムアウトが生じないなどの発生できない状態と入力の組が13個存在する．この場合の検証率は，

$$R_{V1}=\frac{23\times29-13}{23\times29}=0.981$$

となる．

上位レベルからの入力を考慮しない製品検証では，検証できる組合せ数を計算して，検証率は

$$R_{V2}=\frac{381}{23\times29}=0.571$$

となる．

以上のように，上位レベルからの入力を考慮しない製品検証では，上位レベルからの入力を用いる製品検証の6割程度しか検証できないことが分かる．

8. むすび

以上のようにプロトコル機械を順序機械とみなし，オートマトンの同定の手法を用いてHDLC-BAプロトコルの製品検証のテスト系列を求めた．その際，被検証プロトコルの上位レベルからの入力の処理の方法により2とおりの検証方法を考え，その評価を行った．

本論文ではシーケンス番号に対する処理については考察を加えていない．これについては，本論文で述べた検証に先立って行う必要があるが，本質的な問題はないと考えている．

又，上位レベルからの入力を生ずるための検証用タスクの構成，そのプログラム量などについても検討を加える必要があり，これは今後の課題である．

文　献

(1) Zafiropulo, P.: "Protocol validation by doulogue-matrix analysis", IEEE Trans. Commun., COM-26, 8, pp.1187-1194 (Aug.1978).

(2) Stenning, N.V.: "A data transfer protocol", Computer Networks, 1, 2, pp.99-110 (Sept. 1976).

(3) Bochmann, G.V. and Gecsei, J.: "A unified method for the specification and verification of protocols", Proc. IFIPS Cong., pp.229-234, Toronto (1977).

(4) Danthine, A.: "Modelling and verification of end-to-end protocols", Proc. of NTC, Los Angeles (Dec.1977).

(5) 河岡，友永，高橋："プロトコルの記述方法と検証方法について"，情処学会ソフトウェア工学研究会資料8-5（昭53-11）.

(6) 当麻喜弘："順序回路論"，昭晃堂（昭51）.

(7) 電電公社："パケット交換サービスのインターフェース条件"（昭53-11）.

(8) 加藤，斉藤，猪瀬："オートマトンの縦続分解モデルによるHDLCプロトコルの検証の一方式"，信学技報，EC79-37（1979-11）.

（昭和55年3月18日受付）

論　文

多層構造プロトコルをもつネットワークの伝送効率の評価

正　員　斉藤　忠夫†　　正　員　白畑　厚志†

Performance Evaluation of Multi-Layered Protocol

Tadao SAITO † *and* Atsushi SHIRAHATA †, *Members*

あらまし　OSIの基本参照モデルにおけるプロトコルは，7層の階層化された体系として標準化されている．プロトコルは，リンクバイリンクに適用される場合，エンドツーエンドに適用される場合などがある．また，各層のプロトコルデータユニットが伝送されるタイミングは，ネットワークの形態，プロトコルパラメータの相互に関連している．このため，ネットワークの形態に応じて各層のプロトコルパラメータを適切に選択しなければ，ネットワークのスループットが低下する可能性がある．本研究では，OSIプロトコルのトランスポート層，ネットワーク層，データリンク層をもつネットワークについて，ネットワークのスループット評価の一つの尺度である伝送効率を取り上げて性能評価を行い，プロトコルパラメータとスループットの関連を求めている．プロトコルパラメータとしては，伝送速度，伝搬遅延時間，データ単位長，ウィンドウサイズ，中継ノードの数，ビット誤り率，プロトコル処理時間，トランスポート層多重度等を考える．これらの結果から，プロトコルパラメータの選択の方法が明らかになる．

1. まえがき

コンピュータネットワークのプロトコル標準化を目的として，OSI (Open System Interconnection：開放型システム間相互接続) の検討が進められ，一部は実装の段階に達している．OSIではプロトコルを論理的に7層に分割し，各層ごとに標準化している．OSIの検討は，従来機能面に集中されている．これを実システムで適用するには，今後その性能面の検討を進める必要がある．

本研究においては，OSIのトランスポート層以下の下位層のプロトコルパラメータと，その性能について検討している．通信する二つの端末が接続されているときに，実現されるトランスポート層サービスのデータ伝送速度は，伝送回線の速度のみならず，下位層のプロトコルパラメータによって影響される．

従来，プロトコルパラメータとスループットの関係は，1層のプロトコルについてのみ，例えばHDLCについては詳しく検討されている[(1),(2)]．また，共通線信号方式等の特定の応用についての検討もある[(3),(4)]．これに対し，本論文ではネットワークのスループット評価の一つの尺度として伝送効率を取り上げ，多層構造プロトコルを用いた種々のネットワーク形態について，伝送効率を一般的に評価し，プロトコル仕様の選定の目安を得ることを目的とする．但し，層としては，OSIモデルの論理レベルの下位3層，すなわち，トランスポート層，ネットワーク層，データリンク層を考える．

2. 性能評価の方法

図1に，論理レベルの下位3層による通信手順の例を示す[(5)]．この図は，二つの通信ノード間で一つのデータ単位を送受信する場合を示している．一つのトランスポートコマンド $T(C)$ が，送信局側端末Aから受信局側端末Bに伝送される場合，ネットワーク層コマンド $N(C)$ がネットワーク層内で送られる．ここで，各層において，それぞれのコマンドに対し一つのレスポンスが必要なので，ネットワークレスポンス $N(R)$ が必要となる．そしてまた，ネットワーク層のコマンドを伝送するために，データリンク層で1組のコマンドレスポンス対が必要となる．それ故，トランスポートコマンドを伝送して，それに対するレスポンスを受け取るためには，八つのデータユニットがデータリンク層内を行き来しなければならない．すなわち，プロト

† 東京大学工学部電気工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan

コルを多層化することによって，やりとりされるデータユニットの数が増大する．

このような通信に対して，種々のプロトコルパラメータがネットワーク全体の性能に影響を及ぼす．ネットワーク形態，誤り率，データ伝送速度，伝搬遅延時間，プロトコルの各層における伝送制御に要する処理時間

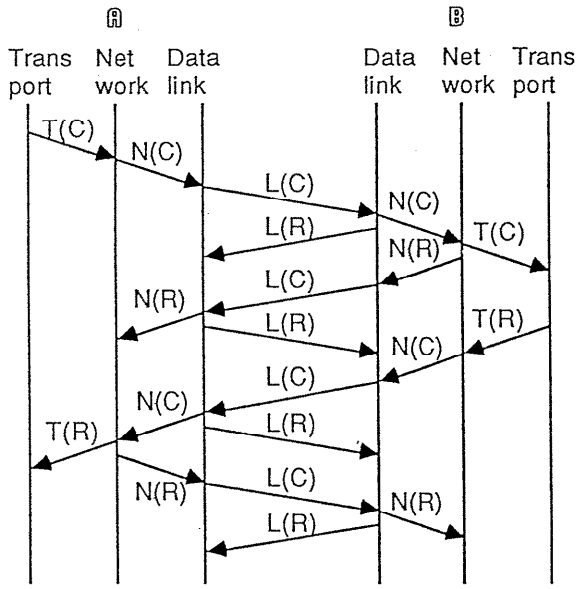


図1　多層構造プロトコルを用いたときの各層のコマンドレスポンスの一例

Fig. 1　An example of command response of each layer in multi-layered protocol communication system.

が与えられたとして，データ単位長，各層のウィンドウサイズ，各層における多重度がプロトコル設計のパラメータとなる．

本論文では，ネットワークを伝送効率によって評価する．これはスループットを正規化したものであり，その定義は以下に示すとおりである．

伝送効率＝スループット÷データ伝送速度

$$= \frac{\text{サービスデータユニット中の情報ビット数}}{\text{上記の情報ビットの伝送に必要な全時間}} \div \text{データ伝送速度}$$

図2に，伝送に必要な全時間の例を，トランスポート層とデータリンク層の2層からなり，各層のウィンドウサイズが1である場合について示す[(6),(7)]．送信側端末Aでは，伝送データがトランスポート層の送信処理の後データリンク層に与えられ，データリンク層の送信処理が行われる．送信処理を行うことにより，プロトコル制御情報（PCI）付きのデータが伝送される．

受信側端末Bでは，まずデータリンク層の受信処理が行われ，データが正しく受信されていればデータリンク層の応答が返送される．このデータリンク層の応答の返送と並行して，トランスポート層の受信処理が行われて，トランスポート層の応答が返送されるが，

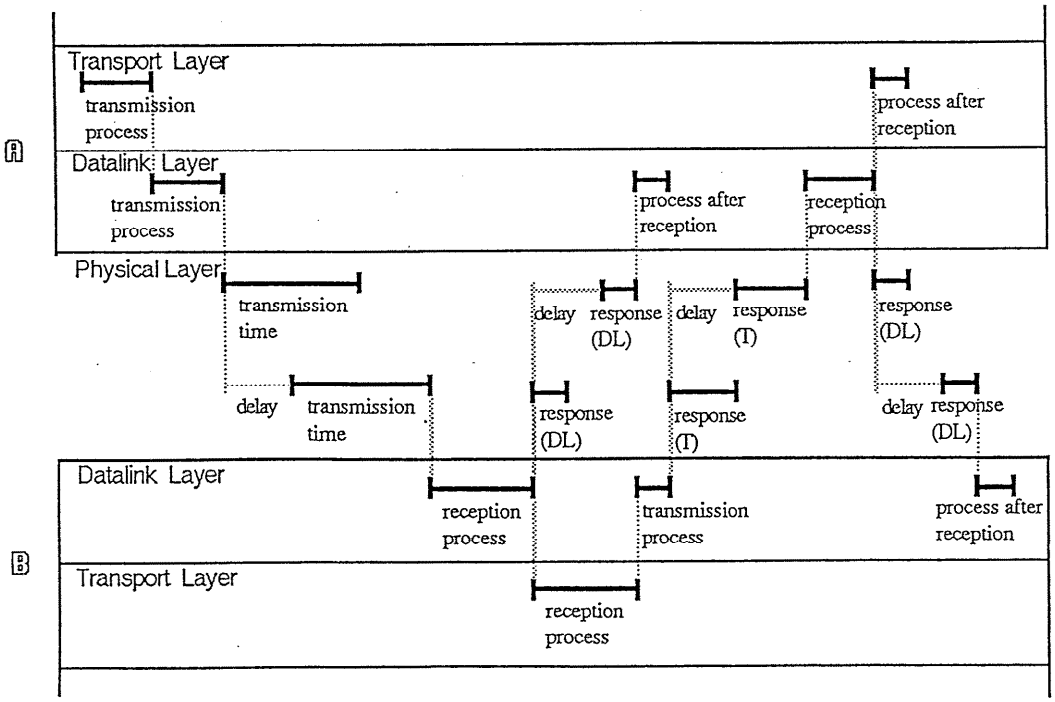


図2　多層プロトコルをもつシステムの動作例

Fig. 2　Operation example of a system having multi-layered protocol.

これはデータリンク層ではコマンドの送信として扱われ，そのための送信処理が行われる．

このようにしてトランスポート層での応答が送られ，これが端末Aで受信されると，端末Aのトランスポート層で後処理が行われる．この後端末Aでは，次のデータ単位を送信することができる．従って，図2でトランスポート層の送信処理開始から，後処理の終了までの時間が，伝送に必要な全時間である．

3. 評価対象モデル

多層構造プロトコルの性能評価に際して，評価対象のネットワークとしては，図3に示すように送信端末から受信端末に対して，1方向の通信が行われるモデルを想定することとしている．評価は，図3のモデルでコネクション設定後のデータ転送について行い，コネクションの設定，解放手順に関するオーバヘッドは評価していない．また，端末間に中継ノードが存在することもあり，この場合には，中継ノードはデータリンク層の機能のみをもつものとする．つまり，ネットワーク層の応答はエンドエンドシグニフィカンスをもつものとしている．

送信端末は，トランスポート層以下の機能を有するものとして，セション層以上の機能は評価の対象とはしない．セション層からトランスポート層に対して，データは連続して与えられるものとする．すなわち，送信側のセション層には連続して送信する十分なデータがあり，トランスポート層以下でデータ送信ができるときにはいつでも，直ちにセション層からデータが与えられ，受信側のセション層ではいつでも，受信されたデータを受理する用意が整っている．

データは1方向に送信されるから，受信側から送信側に対しては，確認用プロトコルに従って，それぞれの層での確認用の応答データユニットが返送される．この応答データユニットの大きさは，

トランスポート層……5バイト
ネットワーク層……4バイト
データリンク層……6バイト

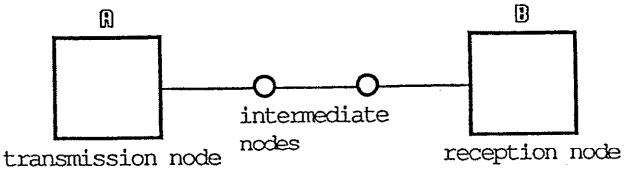


図3 本論文で評価の対象とするシステムのモデル
Fig. 3 System model evaluated by this paper.

とする．また，各層のプロトコル制御情報(PCI)の大きさは次のようにしている．

トランスポート層……5バイト
ネットワーク層……4バイト
データリンク層……6バイト

応答用データユニットについても，その下位層では上記PCIの長さに対応するヘッダが付加されることになる．

各層の処理のためには，それぞれのノードで処理時間が必要であるが，処理は層ごとに別のプロセッサで独立に行われるものとしている．各層の処理時間はシミュレーションのパラメータであるが，特に示さない限り以下のようにしている．

データ送信処理時間……2 msec
データ受信処理時間……3 msec
応答送信処理時間……1 msec
応答受信処理時間……2 msec
応答受信後処理時間……1 msec

各処理時間の関係は，図2に示したとおりである．データ送信処理，およびデータ受信処理は，データコマンドに対する各層の処理である．応答送信処理，および応答受信処理は，応答だけを含むデータユニットに対する各層での処理である．応答受信後処理とは，応答に対する各層における最後の処理である．

伝搬遅延時間もシミュレーションのパラメータであるが，特に述べる場合以外は，5 msecとしている．これは地上で1000 kmの距離を想定したものである．このほか，衛星系を想定した遅延270 msecの場合についても検討している．

伝送誤りの評価に関しては，誤りはデータに対してのみ生じ，プロトコル制御情報に関する誤りは無視している．誤りはデータリンク層で処理される．データリンク層の受信側で誤りが検出された場合は，次のフレームの受信後に応答送信処理が行われて，REJフレームにより回復処理を行う．データリンク層における残留誤りは無視するから，ネットワーク層，トランスポート層での誤り回復については評価していない．

4. シミュレーション結果

4.1 多層化による影響

二つの端末が直接1本のリンクで接続されたポイント-ポイント型のネットワークにおいて，プロトコルの層の数を変えた場合の伝送速度と伝送効率の関係を図4に示す．このグラフから明らかなように，プロトコ

ルの多層化をすれば伝送効率は低下する．すなわち，各層のウィンドウサイズが1，伝搬遅延時間が5 msecでポイント-ポイント型のネットワークの場合，プロトコルが1層から2層になると約10％，2層から3層になると15％程度効率が低下する．これは当然のことながら，多層化によりプロトコルの処理量が増すためである．また，どの場合にも伝送速度が増すにつれ伝送効率が下がっているが，これはプロトコルの処理に要する時間が，伝送媒体上でかかる時間に対して相対的に大きくなるからである．

4.2 ウィンドウサイズの影響

ネットワークの効率を改善するためには，応答を受け取らずに続いて伝送できるデータの数，すなわち，ウィンドウサイズを大きくすることが有効であると言われている．例えば，伝搬遅延時間が5 msecでトランスポートサービスデータ単位（TSDU）長が128バイトの場合，各層のウィンドウサイズをすべて7にすれば，伝送速度が50 kbit/s程度まで，そして，各層とも127まで広げれば400 kbit/s程度まで伝送効率を80％以上に保つことができる．

各層のウィンドウサイズに差がある場合の例として，図5に，トランスポート層のウィンドウサイズと伝送効率の関係を示す．この例では，ネットワーク層とデータリンク層のウィンドウサイズはともに7としている．トランスポート層のウィンドウサイズを変えていくと，7の倍数にあたるウィンドウサイズをとるときは，その前後の値をとるときに比べ，伝送効率が良い値をとる．更に，トランスポート層のウィンドウサイズが7以下では，ウィンドウサイズが大きくなるにつれ伝送効率がよくなるが，7より大きくしてもあまり改善しない．

図6に，ウィンドウサイズを変えた場合のノード数と伝送効率の関係を示す．図6にはウィンドウサイズが各層同一で，1，7および127としたとき(5)，(4)，

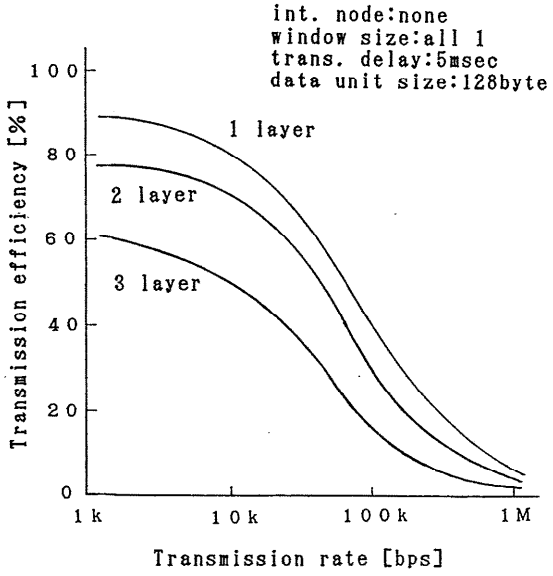


図4　プロトコルの多層化が伝送効率に与える影響
Fig. 4　Effect of number of layers on transmission efficiency.

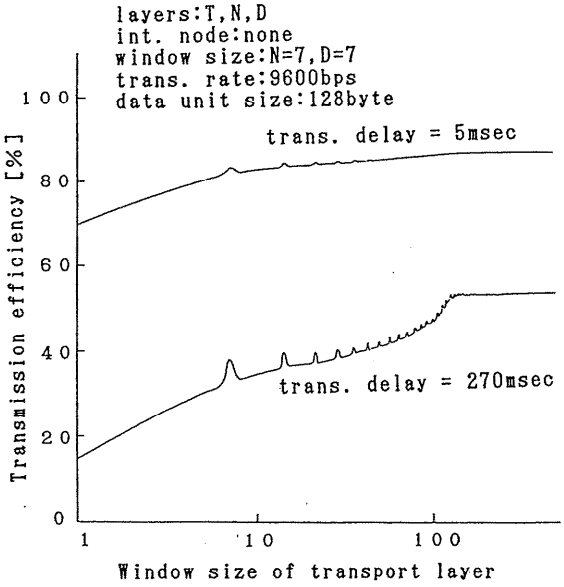


図5　ウィンドウサイズと伝送効率の関係
Fig. 5　Relation between window size and transmission efficiency.

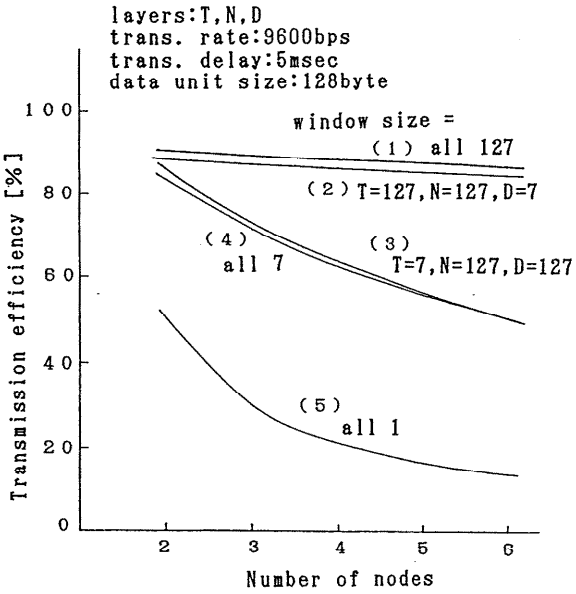


図6　ウィンドウサイズをパラメータとしたときのノード数と伝送効率の関係
Fig. 6　Efficiency versus number of nodes using window size as parameters.

(1), トランスポート層とネットワーク層のウィンドウサイズを127として，データリンク層を7としたとき(2), およびネットワーク層とデータリンク層のウィンドウサイズを127として，トランスポート層を7としたとき(3)の結果が示されている．この図では，両端のノード間にいくつかの中継ノードを含む場合を示している．これ以外にも，1層のみウィンドウサイズを広げた場合の評価を行った．トランスポート層のみで広げても，結果は図6(3), あるいは(4)の場合と同様になる．ネットワーク層，あるいはデータリンク層1層のみでウィンドウサイズを広げた結果も同様である．更に，トランスポート層とデータリンク層の2層で広げた場合も，同様な結果であった．これは，ネットワーク層の応答がエンドツーエンドで送られることから，トランスポート層のウィンドウを広げる効果は，ネットワーク層のウィンドウも併せて広げなければ現れないためと考えられる．逆に，データリンク層についてウィンドウを広げても，図6(3)の場合にはトランスポート層のウィンドウサイズがボトルネックになるため効果は現れないことになる．

すなわち，ウィンドウサイズは各層広げるのがよいことは明らかであるが，いずれかの層だけを広げるときには，(2)のように上位2層(トランスポート層とネットワーク層)を広げるのがよい．一方，中継ノードが増えるに従い，(2)と(3), (4)との差は顕著になり，反対に(3)と(4)の差がなくなる．これは，ノード数が増えるにつれ上位層のウィンドウサイズが小さいことがボトルネックであることがますます明確になるからである．以上述べた結果は，伝搬遅延時間が5 msecのときについてであるが，270 msecの場合も同様な結果が得られている．しかし270 msecの場合は，上位2層でしか広げなかったときの伝送効率(2)は，データリンク層での応答待ち時間の影響がでるため，すべての層でウィンドウサイズを大きくする場合(1)に比してかなり悪くなる．

4.3 ビット誤り率による影響

ビット誤り率と伝送効率の関係は，データ単位長によって支配され，ビット誤り率と最適データ単位長の関係については種々のデータリンク層プロトコルについてよく研究されている．ここでは，プロトコルの層数が増加したときに，従来知られている最適データ長が，どのように変化するかを求めた．図7には，二つのノードが直接接続されている場合に，プロトコルが1層であるときと3層であるときのデータ長と伝送効率の関係を示している．1層の場合は，従来の最適データ単位長として知られている値を示している．

図8には同様に，間に中継ノードがある場合の結果を示している．当然のことながらビット誤り率が大きいほど伝送効率は低くなる．最適データ単位長は，伝送速度が高いほど，また，誤り率が低いほど，そして，

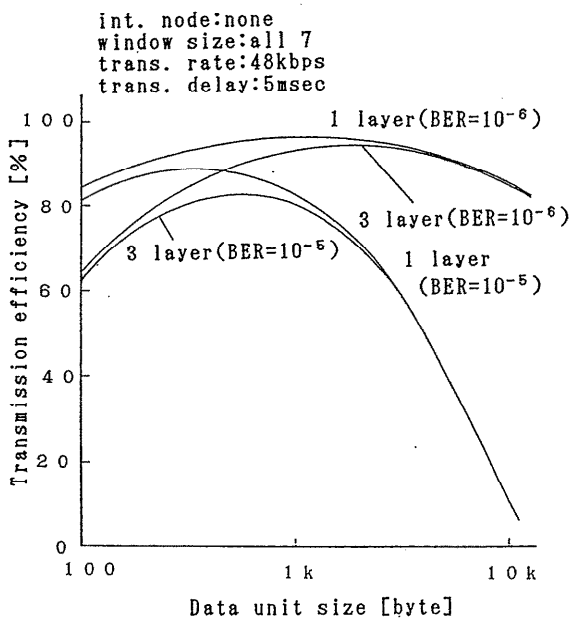


図7 誤り率を考慮したデータ単位長と伝送効率の関係
Fig. 7 Relation between data unit size and transmission efficiency with data error rate consideration.

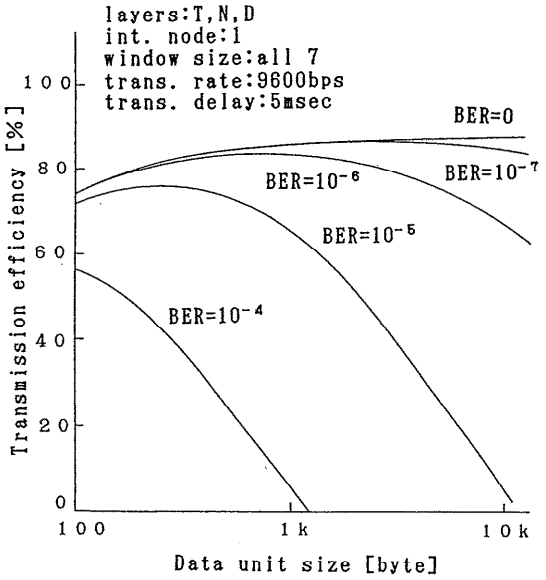


図8 中継ノードがある場合の誤りを考慮したデータ単位長と伝送効率の関係
Fig. 8 Relation between data unit size and transmission efficiency with data error rate consideration in a system having an intermediate node.

伝搬遅延時間が大きいほど長くなる．更に，最適データ単位長は，プロトコルが2層のときより3層のときの方が長くなるが，ノード数やウィンドウサイズが大きくなると短くなることも判明した．

データ単位長を長くすると誤りによるオーバヘッドが大きくなり，短くするとプロトコル制御情報によるオーバヘッドが大きくなるが，層数の増加によって後者のオーバヘッドがより大きい効果をもつために，最適データ単位長は層数の増加と共に長くなるものと考えられる．

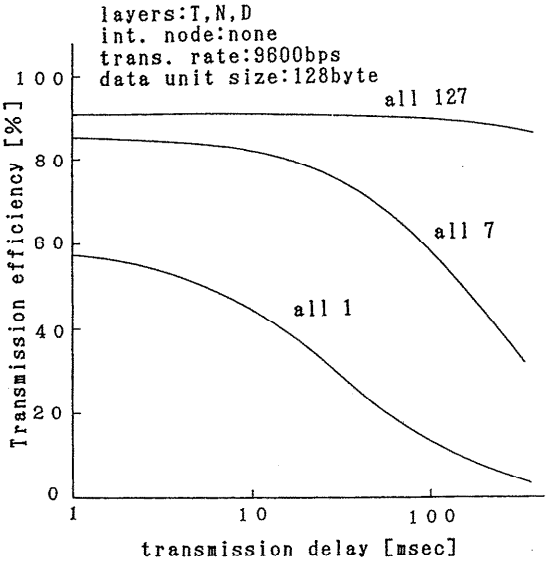


図9　伝搬遅延時間と伝送効率の関係
Fig. 9　Relation between transmission delay and transmission efficiency.

4.4　伝搬遅延時間の影響

図9に伝搬遅延時間と伝送効率の関係を示す．この図からわかるように，ウィンドウサイズを大きめにとれば，特に伝送効率が低いところでは，伝搬遅延時間による伝送効率の低下を防止できることがわかる．これは，伝搬遅延時間に比べデータ伝送時間が大きなものになり，伝搬遅延時間の影響が小さくなるからである．逆に言えば，ウィンドウサイズが大きくても伝送速度が高いところでは，伝搬遅延時間がデータ伝送時間に比べそれほど小さくはならないので，伝搬遅延時間による影響が出ていることになる．

5.　プロトコル処理時間の影響

以上，ウィンドウサイズを中心にして，**3.**で述べたプロトコル処理時間を仮定して，性能とプロトコルパラメータの関係を求めた．この結果，伝送速度が高い場合には伝送効率が比較的低い値にとどまることが判明した．このような効率低下の原因としては，プロトコル処理時間の影響が考えられる．そこで，プロトコル処理時間を短くした場合の，伝送速度と伝送効率の関係を調べた．図10では，プロトコル処理時間を**3.**で示した値から一定の率で減少した場合の効果を示している．プロトコル処理時間が減少しても，伝送遅延に起因する応答待ち時間はあまり減少しないため，伝送効

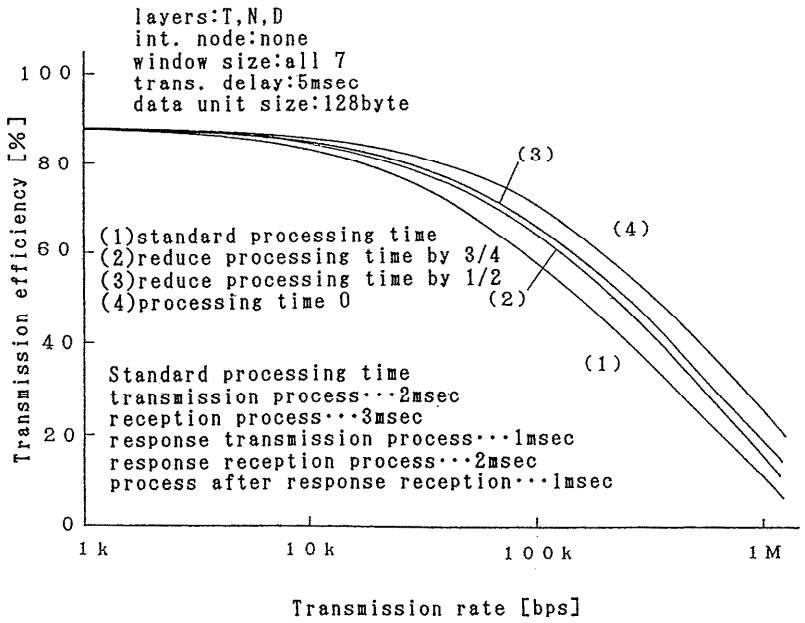


図10　プロトコル処理時間の短縮による伝送効率の改善
Fig. 10　Improvement of transmission efficiency by reduction of protocol time.

率の改善効果は，例えばプロトコル層数を減少した場合に比べてわずかである．伝送速度が小さい(数十 kbit/s 以下)ときには，上述の理由で特に改善効果は小さい．図には示していないが，当然ながら伝搬遅延時間が長いときにも，プロトコル処理時間の改善による伝送効率の改善は見られない．また，伝搬遅延時間が小さいときでも，ウィンドウサイズがそれほど大きくないときは，プロトコル処理時間の影響はそれほど見られない(すべてのウィンドウサイズが1のときで，最も違って8％程度，すべてのウィンドウサイズが7のときは15％程度)．これも同様な理由からである．

しかし，伝搬遅延時間が短く，ウィンドウサイズを127程度に広げた場合には，伝送速度が高いところにおいて，プロトコル処理時間が伝送効率に非常に大きな影響を与えていることが判明した(図10と同一条件でウィンドウサイズを127としたとき，1 Mbit/s で，処理時間を0としたときにはもとの場合に比べて伝送効率は2.5倍以上になる)．このことから，伝搬遅延時間が短く，伝送速度が高いときで，ウィンドウサイズが極めて大きい場合には，プロトコルの処理時間が伝送効率低下の主たる原因となり，効率改善のためにはプロセッサの能力向上が必要であると言える．しかし，そのほかの場合は，プロトコル処理時間は伝送効率に対してそれほど影響がないと言える．

6. データ単位長の影響

図11に，伝送速度が1 Mbit/s のときの，ビット誤り率をパラメータとした場合のデータ単位長と伝送効率の関係を示す．この図からわかるように，伝送速度が高いときは，データ単位長を長く(3.5 k バイトより長く)すれば，伝送効率が80％以上になる．伝送速度が高い場合の伝送効率低下の原因は，プロトコル処理時間，伝搬遅延時間等が挙げられる．しかし，これらのうちの一つを短くできたところで，伝送効率はいつでも改善されるとは限らないし，また，プロトコル処理時間や伝搬遅延時間は，簡単に短くすることはできない．従って，伝送速度が高いところにおいて，伝送効率を改善するために最適な方法は，データ単位長を長くすることである．しかしこの場合には，レスポンスタイムが長くなることに注意しなければならない．

7. トランスポート層の多重化による影響

OSI では，一つの〈N-1〉コネクションを複数の〈N〉コネクションで共用可能にするために必要な機能として多重化を定義している．多重化は，トランスポート層プロトコルのクラス2や，CCITT 勧告 X.25(ネットワーク層プロトコル)で実現されている．

トランスポート層の多重化による影響を調べるために，各層のウィンドウサイズを変えて伝送効率を求めたが，その結果の一つを図12に示す．伝送効率は多重度と共に増大するが，伝送速度が小さい場合は，伝送効率がすぐ飽和してしまう．これは，物理伝送路におけるデータ伝送の限界のためである．一方，伝送速度

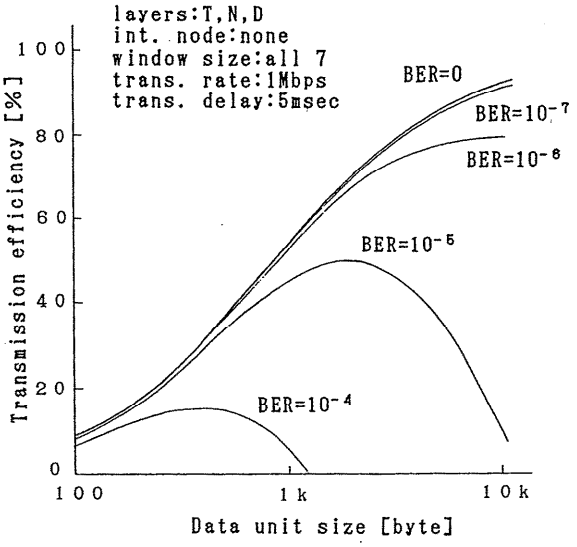


図11 データ単位長を長くすることによる伝送効率の改善
Fig. 11 Improvement of transmission efficiency by enhancing data unit size.

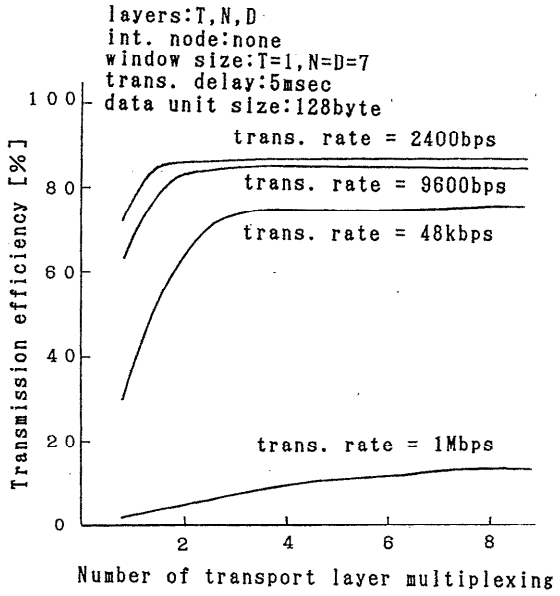


図12 トランスポート層の多重度と伝送効率の関係
Fig. 12 Improvement of transmission efficiency by transport layer multiplexing.

の高い場合は，多重化による伝送効率の改善があまり見られないが，これは，プロセッサの処理時間や伝搬遅延時間が支配的になるためである．以上まとめると，伝送速度が小さいとき，および，トランスポート層とネットワーク層のウィンドウサイズがほぼ等しいときは，トランスポート層多重度は2程度で十分である．そして，伝送速度が大きく，かつ，トランスポート層のウィンドウサイズがネットワーク層のウィンドウサイズより小さい場合は，これより高い多重度をとることが有効である．

8. あとがき

以上のシミュレーション結果をまとめると，多層化，ウィンドウサイズ，ビット誤り，伝搬遅延時間，プロトコル処理時間，データ単位長，トランスポート層の多重化などが伝送効率に与える影響は以下のとおりである．

・プロトコルの多層化による処理量が増すため，伝送効率が低下する．

・ウィンドウサイズを広げると，応答待ち時間による伝送効率への影響が減るので，伝送効率が改善される．

・ウィンドウサイズは各層同一であるのが最も良い．そうでない場合は，上位層を広げるのが良い．

・ビット誤りが生ずる場合は，伝送効率には最適データ単位長がある．この最適データ単位長は，他のプロトコルパラメータにより変化する．

・伝搬遅延時間がデータ伝送時間に比べ長いときは，伝搬遅延は伝送効率低下の主要な原因となる．

・多重化を考えない場合，プロトコルの処理時間は伝搬遅延時間が短く，伝送速度が高いときで，ウィンドウサイズが大きい場合のみ，伝送効率低下の主たる原因になる．

・伝送速度の高いところでの伝送効率の改善には，データ単位長を長くするのが有効である．

・伝送速度が高い領域でも，トランスポート層多重化による伝送効率の改善があまり見られないが，これは，プロセッサの処理時間や伝搬遅延時間の影響のためである．

9. むすび

ネットワークの性能に各種プロトコルパラメータが与える基本的な影響を知る上で，本論文が扱ってきたようなネットワークの性能評価の意義は大きい．しかし，実際のネットワークに更に近づけるためには，より複雑な構成のもの，例えば，分岐のあるネットワークやう回路のあるネットワークなどを扱っていく必要がある．また，各層における多重化や，トランスポート層でのデータの分割，組立等についても今後の課題として検討する．

文　献

(1) W. Bux, K. Kummerle and H. L. Truong : "Balanced HDLC procedure : A performance analysis", IEEE Trans. Commun., **COM-28**, pp. 1889-1898 (Nov. 1980).

(2) E. Gelenbe, J. Labetoulle and G. Pujolle : "Performance evaluation of the protocol HDLC", Computer Network, 2 (Sep. 1978).

(3) E. Vazquez, J. Vinyes and R. Colella : "Simulation of the X. 224 transport protocol over X. 25 network", 12th International Teletraffic Congress, Torino (June 1988).

(4) A. Pach, P. Brown and B. Delosme : "Performance analysis of the class 4 transport protocol", 12th International Teletraffic Congress, Torino (June 1988).

(5) T. Saito, H. Aida and A. Shirahata : "A performance evaluation of Computer network using multi-layered protocol", 12th International Teletraffic Congress, Torino (June 1988).

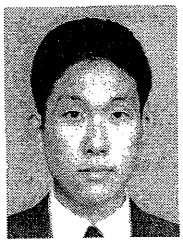
(6) 白畑厚志，相田　仁，斉藤忠夫："多層構造プロトコルを用いたネットワークの性能評価"，情処学第36回全大，2G-6 (昭63).

(7) 白畑厚志，相田　仁，斉藤忠夫："トランスポート層の多重化を考慮した多層構造プロトコルの性能評価"，情処学第37回全大，5E-9 (昭63).

(平成元年2月1日受付)

斉藤　忠夫

昭38東大・工・電子卒．昭43同大大学院博士課程了．東大工学部講師，東大助教授，東大教育用計算機センター主任，カリフォルニア工科大学研究員等を経て現在東大教授．ディジタル通信方式，コンピュータネットワーク，ディジタルシステムの信頼性などの研究に従事．著者としては電子回路入門(昭晃堂，昭52)，ディジタル回路(コロナ社，昭57)等がある．42年，61年度論文賞，48年度業績賞など受賞．電気学会，IEEE各会員．工博．



白畑　厚志

昭62東大・工・電子卒．平成元年同大大学院修士課程了．同年ＫＤＤ入社．在学中，コンピュータネットワークの評価に関する研究に従事．

UDC 681.324.078

計算機網におけるウィンドウ方式フロー制御の効果の一評価

正員　斉藤　忠夫[†]　　正員　小林　進一[††]　　正員　猪瀬　博[†††]

An Evaluation of Flow Control Capability by Window Scheme in Computer Network

Tadao SAITO[†], Shinichi KOBAYASHI[††] *and* Hiroshi INOSE[†††], *Regular Members*

あらまし　計算機網においては，通信サブネットに接続されている計算機や端末装置のデータ送受信の能力がまちまちであるために，フロー制御の機能が必要である．フロー制御の代表的な方式にウィンドウ方式があるが，これは，送信側と受信側であらかじめ連続して送信することのできる情報単位の総数に一定の上限値を取り決めておき，それを越える情報単位については，受信側から返信されてくる受信確認に応じて送信していく方式である．この上限値は，一般にウィンドウサイズと呼ばれ，フロー制御の効果を決める重要なパラメータのひとつである．CCITT勧告X.25では，網のエンドノード間で論理チャネルを設定し，この論理チャネル内でパケットに対してウィンドウ方式によるフロー制御を行うよう規定している．本論文は，エンドノード間で設定することを許される論理チャネル数とウィンドウサイズという2種類のフロー制御のパラメータの効果について評価したものである．待ち行列理論による解析の結果，並びに，シミュレーションの結果によって，この2種類の制御パタメータと，網内の伝送遅延やスループットなどの網のトラヒック特性との間に，ある特定の関係があることが明らかになった．

1. まえがき

計算機網は，通信回線やバッファや処理装置などの物理的な資源から構成されている．しかし，これらの資源には，限界となる容量があるため，その限界以上の負荷が与えられると輻輳状態となる．輻輳状態では，資源が浪費されることにより，網全体で運ぶことのできるデータの量は減少し，データ伝送に要する遅延は増大し，網の特性が著しく低下してしまう．この輻輳を防止するため，各種のフロー制御が実行されている．

現在，CCITT勧告X.25では，下位から順に，

〈1〉 物理レベル　(physical level)

〈2〉 リンクレベル　(link level)

〈3〉 パケットレベル(packet level)

の3階層のプロトコルが規格化されていて，このうちリンクレベルでは，隣接ノード間の伝送にHDLC手順を採用し，物理リンクごとに連続転送可能なフレーム数に上限を設けるフロー制限を行っている．さらに，パケットレベルにおけるバーチャルコール方式では，エンドノード間で論理チャネルを設定し，この論理チャネルごとに連続転送可能なパケット数に上限を設けるフロー制御を行っている．一般に以上のようなフロー制御の方式はウィンドウ方式と呼ばれ，上限値はウィンドウサイズと呼ばれている．パケットレベルでは，網のエンドノード間で設定することを許される論理チャネル数とウィンドウサイズが，フロー制御のパラメータである．本論文は，この2種類のフロー制御のパラメータの効果について評価したものである．

実際のウィンドウサイズの値は，リンクレベルでは7以下に規格化され，パケットレベルの標準値に関しては，DDX網では15に，TelenetやDatapacでは2になどと網提供者の側でいろいろに設定されている．ウィンドウサイズなどの制御パラメータの値をいくつに設定するかは，網全体のスループットや伝送遅延を決定する重要な問題であると考えられる．しかし現実の計算機網では，網の形態，ユーザから与えられる負荷，各層のプロトコルなどが一般に複雑であるために，

† 東京大学工学部電気工学科，東京都
†† 東京大学工学部情報工学科，東京都
††† 東京大学工学部電子工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan

論文番号：昭58－論144[D－38]

各種のフロー制御の効果に関しては解明されていない点が多く，フロー制御のパラメータを設定する際に，大きな問題となっている．

本論文で述べる手法によって，いくつかの限定条件が満足される場合には，網の形態や与えられる負荷や要求されるパケット遅延の保証値などに応じて，フロー制御のパラメータを設定することができる．

フロー制御に関する研究としては，Isarithmic方式[1]に関しては，岡田[2]らが研究している．また，リンクレベルのウィンドウ方式に関しては，Bux[3]，井上[4]らが単一リンクに全負荷（full load）のかかった場合のウィンドウサイズとスループットの関係を解析している．本論文で評価するパケットレベルのウィンドウ方式に関しては，Pennotti[5]，Takenaka[6]，松尾[7]らの研究があるが，これらは，直列にノードが連らなった形態の網のみを対象としたり，論理チャネルの外のトラヒックを適当に仮定するなど，一般性に欠ける面がある．これに対し，本論文では，2次元的広がりをもった網を対象として，網への負荷を変化させた場合について検討している．さらにフロー制御のパラメータとしても，ウィンドウサイズだけでなく，エンドノード間で設定することを許される論理チャネル数も評価の対象としている．

この2種類の制御パラメータの効果を明らかにするため，**2.**で説明する計算機網のモデルを対象として，**3.**で定義する遅延および **4.** で定義するスループットを評価する．ある程度網全体の機能を考慮して解析を進めるために，いくつかのかなり粗い近似が必要となっている．そのため，**6.** でシミュレーションによって，理論的解析の結果を確認している．

2. 評価の対象とする計算機網のモデル

2.1 モデルの概要

網の形態は，たとえば図1のように，複数のノードが全二重の物理リンク（physical link）によって接続されているものとする．ノードには，メッセージ(message)の生起するエンドノード（end-node）とパケットの交換を行う中継ノード（switching-node）とがある．メッセージは，エンドノード間にポアソン生起し，図2のX.25プロトコルによって転送される．

2.2 パケットレベルのモデル

論理チャネル（logical channel）は，メッセージがエンドノード間に生起すると1本設定され，そのメッセージの転送が終了すると解放される．ここで図3のように，エンドノード間に設定することを許す論理チャネルの数を m 本に制限し，メッセージが生起した時点で既に m 本の論理チャネルが設定されている場合には，このメッセージを呼損とする．図4のように各メッセージは固定長のパケット（packet）に分解されるが，メッセージ長は指数分布に従い，平均長がパケット L 個分に相当するものとする．また，エンドノード間でパケットの受信確認を行って，ウィンドウサイズが w のフロー制御を実行する．また，送信ノード（source node）からのデータパケット（DATA packet）が誤りなしに，受信ノード（destination node）に到着したことを示す受信確認の解析を簡単化するため，受信ノードで RR パケット（RR packet）を各データパケットに対して返信し，網内では固定式の経路選択を行うとする．

2.3 リンクレベルのモデル

パケットが隣接ノード間を伝送される際には図4のように1個のパケットに対し1個のフレーム（frame）が構成される．フレームの伝送誤り率が1％程度であ

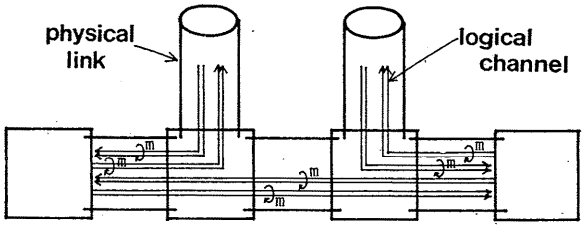


図3　物理リンクと論理チャネル
Fig.3−Pysical link and logical channel.

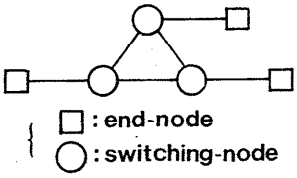


図1　評価対象の網形態
Fig.1−Network configuration of the evaluated system.

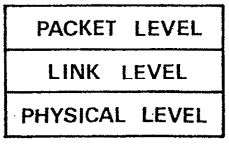


図2　X.25のプロトコル階層
Fig.2−Protocol hierarchy of X.25.

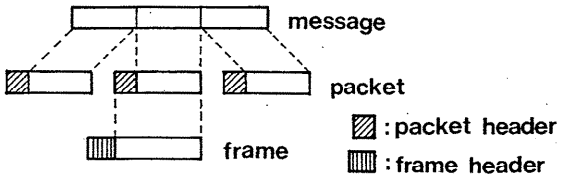


図4　メッセージ，パケット，フレームの関係
Fig.4−Relation between message, packet and frame.

る場合には，その影響は無視できるとの報告[8]もあるので，ここでは伝送誤りの影響を考慮しなかった．

2.4 ノードのモデル

パケットが網内を転送される際には，ヘッダの解析や付加などの処理とフレームの伝送という2種類のサービスを交互に受ける．図5のモデルでは，ノードにおける種々の処理（process）をPで表わし，リンクにおける伝送（transmission）をTで表わしている．また，この処理及び伝送において生ずる待ちを取り扱うために，FIFO のバッファを設ける．また，以下の解析では，各ノードのバッファのサイズはすべて無限大であると仮定している．

3. ウィンドウサイズと遅延の関係

本論文では，**2.**で述べたような計算機網を通して，メッセージを転送する場合のトラヒック特性について評価するが，まずトラヒック特性のひとつとして，次のメッセージ遅延とパケット遅延について考える．

3.1 メッセージ遅延とパケット遅延

メッセージ遅延（D_M）は，送信ノードでメッセージが生起してから，そのメッセージに属するすべてのデータパケットが宛先の受信ノードに到着し，その最後のデータパケットに対する RRパケットが送信ノードに返信されるまでの時間であると定義する．

パケット遅延（D_P）は，データパケットが送信ノードで交換の処理を待つバッファに入ってから，宛先の受信ノードで自局宛てであることが確認され，それに対しての RR パケットを中継ノードへ向けての伝送待ちのバッファに入れる処理が終わるまでの時間とする．

遅延の解析に当って，パケット転送の様子を簡単に把握するために，次の2つの仮定を設ける．

(a) $L > w$が成立している．すなわち，ひとつのメッセージ当りの平均パケット数は，ウィンドウサイズに比べて大きい．

(b) 送信ノードで，あるデータパケットを送出してからそれに対する RR パケットが返信されるまでに要する時間（ラウンドトリップ時間＝$2D_P$）は，同一メッセージに属するw個のデータパケットを連続して送出する時間に比べて長い．

これらの仮定は，ウィンドウ方式によるフロー制御が実際に効果をもつための条件である．また (b)は，ウィンドウサイズが小さい時や，網内伝送に要する時間が大きい輻輳時などに成立している．そうでない場合には，ウィンドウ方式によるパケット送出の制御を過大に評価していることになり，実際の遅延は，以下の解析で求めるものより小さくなると考えられる．

以上の2つの仮定の下では，送信ノードから受信ノードへ至るパケットの転送の様子は図6に示すようになる．RR パケットは各データパケットに対して返信されるが，第1番目のデータパケットに対する RR パケットが送信ノードに到達するまでに，仮定(b)によって，w個のデータパケットは送出完了済みである．このRRパケットの到着により送信ノードは（$1+w$）番目のデータパケットを送出することができる．このようにパケットはw個連続して網を転送されるが，このパケットバーストの数は$\left\lfloor \frac{L}{w} \right\rfloor$個である．$L$が$w$で割り切れない場合には，さらに$2D_P$の時間を要する．この余りのパケットの送出時間は$2D_P$に比べて充分小さいと考えられるので，第$L$番目の RR パケットが返信されるまでのメッセージ遅延は次式で与えられる．

$$D_M = 2D_P \left\lceil \frac{L}{w} \right\rceil \tag{1}$$

但し，$\lfloor x \rfloor = x$より大きくない最大の整数

$\lceil x \rceil = x$より小さくない最小の整数

D_MとD_Pは，比例していることがわかる．

3.2 ノードとリンクの呼量

ここでは，ウィンドウ方式による遅延抑制の効果を評価するための準備として，次のアルゴリズムにより，

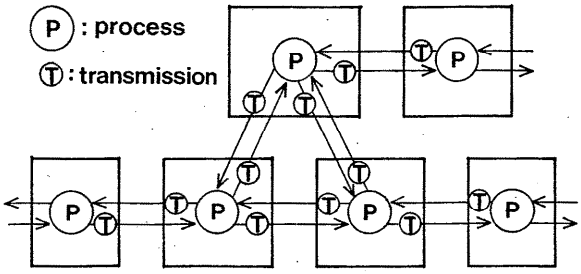


図5 ノードの処理とリンクの伝送のモデル
Fig.5-Model of process in node and transmission by link.

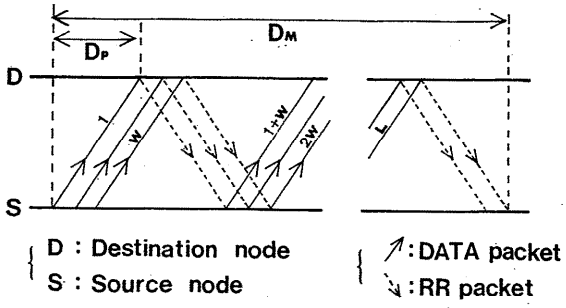


図6 パケットの転送と遅延
Fig.6-Packet transfer and delay.

図5の各ノードのPとTのバッファへの呼量を求める．

(i) 網内のノードのすべてのノードとリンクにあるバッファに対して，順番を全く無視して，1，2，…，j，… と適当に番号を付ける．

(ii) ノードでのパケットのヘッダ解析などに要する処理時間およびリンクでパケットを伝送するのに要する時間から，第j番のバッファのサービス率μ_j（packet／sec）を計算する．

(iii) エンドノード間でのメッセージ生起率やそのメッセージ長，網の形態，経路選択方式，受信確認の方式などから，第j番のバッファへの到着率λ_j（packet／sec）を計算する．

(iv) 各バッファの呼量は，$a_j=\lambda_j/\mu_j$ となる．

計算の一例として，このアルゴリズムを表1の状況に当てはめた場合を示す．まず，図1の形態の網には，図5のPおよびTで示す18個のバッファがある．また，

h_P： ノードでの処理時間 （sec／packet）

h_T： リンクでの伝送時間 （sec／packet）

と表わす．次に，網全体に対して，

λ（message／sec）$=\lambda L$（packet／sec）

の負荷が，図7のように3個のエンドノードに均等に与えられるとする．最短固定式の経路選択を行うことや，各データパケットに対してRRパケットを返信することなどを考慮に入れると，各バッファへのパケットの到着率は，図8のようになる．したがって，各のバッファの呼量は，図9のようになる．この図9より，

$2h_P \leqq h_T$ のとき，加入者リンクが最大呼量

$2h_P \geqq h_T$ のとき，中継ノードが最大呼量

となっていることがわかる．一般に，諸定数によって，最大呼量となる場所（隘路）が変化する可能性がある．

3.3 ウィンドウ方式フロー制御の効果

ここでは，以下に示すn段直列モデル[5]によって，ウィンドウサイズとパケット遅延の関係を解析する．

パケットは，図5のように網内を交換処理と伝送を交互に繰り返して転送される．ここで，1本の論理チャネルに着目すると図10のモデルとなる．次に，**3.2**で説明した全部のバッファの呼量のうち，この論理チャネルに沿って送信ノードの処理段から受信ノードの処理段までのものに，順に番号を付けて，第i段の呼量を，a_i（$1 \leqq i \leqq n$）と書くことにする．このn段のうちで，最大の呼量を持つ段の番号の集合を$\boldsymbol{I}$で表わし，$l \in \boldsymbol{I}$ に対しての第i段の相対負荷を，

$$\theta_i=\frac{a_i}{a_l} \tag{2}$$

と定義する．明らかに次の性質が成り立っている．

$$\left.\begin{array}{ll} i \in \boldsymbol{I} \text{ のとき，} & \theta_i=1 \\ i \notin \boldsymbol{I} \text{ のとき，} & \theta_i<1 \end{array}\right\} \tag{3}$$

次にウィンドウ方式の評価に当って，ウィンドウの影響を考慮するために，次のような仮定を設ける．

(c) 同一メッセージに属するパケットは，w個連続したパケットバーストとして，網内を転送される．

(d) パケットバーストは各段にポアソン到着する．

ここで，仮定(c)は，Lがwの整数倍で低負荷時という2条件の下では，ある程度満足されよう．前者の条件

表1 計算例の状況

網の形態	図1の6ノードの網
与えられる負荷	図7の均等な負荷
経路選択	最短の固定式

表2 計算例の諸定数

平均メッセージ長	L	21 packet／message
処理時間	h_P	3 msec／packet
伝送時間	h_T	20 msec／packet

$$\begin{bmatrix} 0 & \frac{\lambda}{6} & \frac{\lambda}{6} \\ \frac{\lambda}{6} & 0 & \frac{\lambda}{6} \\ \frac{\lambda}{6} & \frac{\lambda}{6} & 0 \end{bmatrix}$$

図7 エンドノード間に与えられる負荷の行列
Fig.7－Offered load matrix between end-nodes.

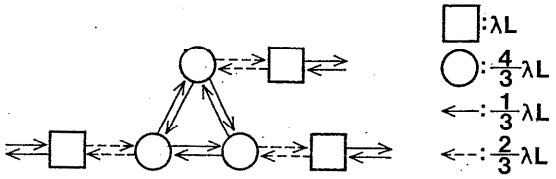


図8 ノードとリンクへのパケットの到着率
Fig.8－Packet arrival rate to nodes and links.

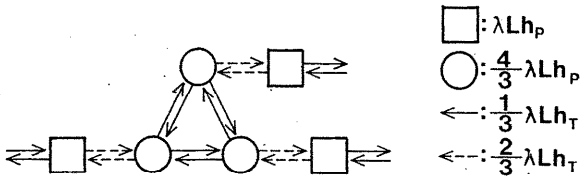


図9 ノードとリンクの呼量
Fig.9－Traffic intensity of nodes and link.

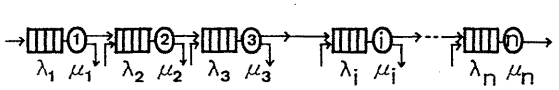


図10 n段直列モデル
Fig.10－n-stage tandem queuing model.

は L が w に比べて充分大きい場合には無視できると考えられる．しかし，後者の条件が成立しない高負荷時には，パケットバースト同志が衝突しあって，w 個のパケットはばらばらに各段に到着することになる．そのため，w を大きくすることによる遅延の増加は，以下に述べるほどには大きくならないと予想される．この点については，**6.** でシミュレーションにより考察する．次に，仮定(d)に関しては，実際にはパケットが固定長であるために，ある論理チャネルから各段への到着間隔は一定以上となり，それがいくつか集まっても全体としてポアソン到着とはならない．しかし，図5のように処理段への入力は一般に複数個存在すること，また伝送段については一般に伝送速度より処理速度の方が速いことによって，ある程度根拠のある近似であると考えられる．

以上の仮定の下で，パケットのバースト化による各段での待ちの増大を評価し，パケット遅延を解析する．パケットバーストに関する諸量に〜を付けて表わすと，

$$\tilde{\mu}_i=\frac{\mu_i}{w},\quad \tilde{\lambda}_i=\frac{\lambda_i}{w},\quad \tilde{a}_i=a_i \quad \text{となる．}$$

第 i 段におけるパケットバーストの待ち時間 (t_i) は，$M/D/1$ の公式を適用することにより，次式となる．

$$t_i=\frac{a_i}{2\left(\frac{\mu_i}{w}\right)(1-a_i)}$$

この t_i が第 i 段におけるパケットの待ち時間に等しいことを考慮すると，全 n 段でのパケット遅延 (D_P) は，サービス時間も含めて次式のように求められる．

$$D_P=\sum_{i=1}^{n}\left(\frac{1}{\mu_i}+t_i\right)=\sum_{i=1}^{n}\frac{1}{\mu_i}\left(1+\frac{wa_i}{2(1-a_i)}\right) \quad (4)$$

4. スループットの評価

本論文で解析しているモデルでは，エンドノード間で設定することのできる論理チャネル数 (m) を越えるメッセージの送信要求を呼損としている．ここでは，その呼損率とスループットについて理論的に解析する．

4.1 網のサービス率

論理チャネル数によるフロー制御の効果を評価する準備として，処理能力が刻一刻と変化している計算機網をマクロ的に捉えることによって，一般には固有のサービス時間の定まっているシステムに対して定義されるサービス率の概念を計算機網に対しても導入する．網のサービス率は，単位時間当りに網全体で転送できるメッセージの総数であると定義する．これは，**3.2** で求めた処理と伝送の各段の呼量 (a_j) のうちで，$\max(a_j)=1$ となるようなトラヒック状態に対して，すべてのエンドノード間でのメッセージ生起率の合計として求めることができる．これを次のように表わす．

μ (message／sec)

次に **3.2** と同様に表1の状況での μ の計算例を示す．諸定数を表2のようにとると，$2h_P<h_T$ が成立しているので，図9の呼量のうち最大のものは，加入者リンクの $\frac{2}{3}\lambda L h_T$ である．これを1と置くことにより，単位時間当りに網全体で転送することのできるデータの総量は，次式のようになる．

$$\lambda=\frac{3}{2h_T L}\ (\text{message／sec})$$

これが，図1の網のサービス率 (μ) である．

4.2 論理チャネル数とスループットの関係

論理チャネル数の解析を簡単に行うために，以下では，モデルを次の条件を満たすものに限定する．

(e) 網内のすべてのエンドノード間では，均等な負荷のメッセージが生起し，設定される論理チャネルの通過する中継ノードがすべて同数である．

たとえば，表1の状況は，条件(e)を満足している．条件(e)の下では，エンドノード間の状態がすべて対称となる．また，この条件が満たされない場合にも，個々のエンドノード間で以下に述べるような解析が可能であるが，ここでは，**5.** で網全体としての遅延の特性を総合的に評価する必要から，条件(e)を仮定する．

このとき，網全体に生起するメッセージの合計を，

λ (message／sec)

であるとする．すると網全体に対して与えられる負荷 (offered load) は，$a=\lambda/\mu$ となる．網の負荷(a)が与えられた時のメッセージの呼損率が (B) である時に網全体を実際に通過する負荷のことを，スループット (throughput) (γ) と定義すると，

$$\gamma=a(1-B) \quad (5)$$

が成立している．次に，網全体でエンドノード間の順列が e 通りあるとする（図1の網では，$e=6$）．あるエンドノード間でのメッセージ生起率は λ/e となり，サービス率は論理チャネル数 (m) にも反比例して μ/em となると考えられる．このエンドノード間に定常状態で k 本の論理チャネルの設定されている確率を $P(k)$ と書くと，次の統計的平衡の式が成立する．

$$\frac{\lambda}{e}P(k)=\frac{\mu}{em}(k+1)\,P(k+1)$$

これと，$\sum_{k=0}^{m}P(k)=1$ より $P(k)$ が得られ，呼損率は，

$$B=P(m)=\frac{(ma)^m}{m!}\Big/\sum_{k=0}^{m}\frac{(ma)^k}{k!} \tag{6}$$

となる．式(6)は，Erlang の呼損率の式に似た形であるが，前者では $1<a$ で過負荷になるのに対し，後者では $m<a$ で過負荷になる点が異なる．式(6)を式(5)に代入して，次のようにスループット (r) が求められる．

$$r=a\sum_{k=0}^{m-1}\frac{(ma)^k}{k!}\Big/\sum_{k=0}^{m}\frac{(ma)^k}{k!} \tag{7}$$

4.3 ウィンドウサイズとスループットの関係

一般に，全負荷状態にあるリンクでは，リンク内のウィンドウサイズとリンクを通過するスループットは関係することが知られている[(3),(4)]．しかし本論文で定義したスループットは，生起時に呼損とならずに受理されたメッセージ数によってほぼ決ってしまう．なぜなら，各メッセージに属するパケットの数は生起時に定まっているため，全負荷時と違って，最終的に転送されるパケット数が決っているからである．

といっても，w と r も微妙に関係する．例えば w を大きくすると，輻輳により D_P が若干増大するが，式(1)より D_M は減少する傾向になり，滞留メッセージが減少する．そのため高負荷時には呼損率が減少し r は増大すると考えられる．これは低負荷時には無視できる2次的な効果であるため本論文では解析しなかった．

5. 遅延の総合評価

ここでは，**4.**で求めたスループットが実際に網を通過するトラヒックとなる点に留意し，**3.**で述べた遅延の特性を総合的に評価する．これは式の上では，式(2)の最大呼量 (a_l) と式(7)のスループット (r) が等しいことに相当している．この2つの式から，図10の n 段直列モデルの第 i 段を実際に通過する呼量 (a_i) は，

$$a_i=\theta_i a_l=\theta_i r=\theta_i a(1-B)$$
$$=\theta_i a\sum_{k=0}^{m-1}\frac{(ma)^k}{k!}\Big/\sum_{k=0}^{m}\frac{(ma)^k}{k!} \tag{8}$$

のようになる．これを式(4)に代入すると，パケット遅延 (D_P) が，(m, w) の2種類の制御パラメータを考慮した形で，次式のように求められる．

$$D_P=\sum_{i=1}^{n}\frac{1}{\mu_i}\left(1+\frac{w\theta_i a(1-B)}{2(1-\theta_i a(1-B))}\right) \tag{9}$$

但し $\begin{cases}\theta_i & \text{：式(2)の相対負荷}\\ B & \text{：式(6)の呼損率（}m\text{により定まる）}\end{cases}$

次に，付録に示す解析によって，$a\to 0$ 及び $a\to\infty$ における D_P のふるまいを調べることによって，パケット遅延の増加率および絶対値と，制御パラメータとの間に，次のような関係があることがわかる．

低負荷時 $\dfrac{dD_P}{da}\propto m^0 w^1=w$ (10)

低負荷時 $D_P\propto m^0 w^0$ (11)

高負荷時 $D_P\propto m^1 w^1=mw$ (12)

また，式(1)の D_M と D_P の関係を用いることにより，メッセージ遅延に対しても，式(9)に相当する一般式や式(10)～(12)に相当する関係を求めることができる．

6. シミュレーション

本論文では，**2.**で記述したモデルに対して，**3.**～**5.**で，いくつかの仮定に基づいた理論的解析を行った．ここでは，解析結果を裏付ける意味で，**2.**のモデル以上に現実の計算機網に近い詳細なモデルに対してシミュレーションを実行し，両結果を比較した．

6.1 シミュレーションの実行

Gasp－II[(9)]を利用し，Fortran で作成したシミュレータによって，定常状態以後の数百個のメッセージ（$L=21$ として数千パケット）に対して統計をとった．

隣接ノード間の伝送には，ウィンドウサイズが7のフロー制御を行う．各ノードでフレームが占有することのできるバッファ数を合計100個とし，バッファの使用総数に応じて自局ビジーとして，RNR，RRの監視フレームを用いてバッファの管理を行う．理論値の計算例と同じく，表1の状況で，表2の諸定数で，図3の論理チャネルを設定する．監視フレームの伝送時間を4 msec とする．このようなモデルを対象にシミュレーションを実行し，以下に述べる評価を行った．

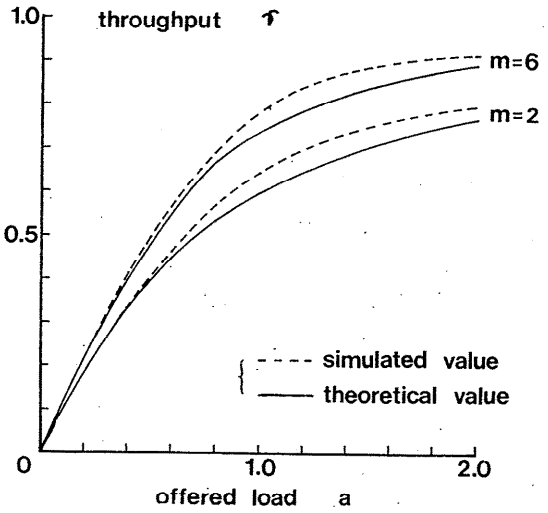


図11 与えられる負荷に対するスループットの特性
Fig.11－Throughput characteristics vs. offered load.

6.2 メッセージ遅延とパケット遅延

最初に，式(1)の D_M と D_P の関係について調べる．いくつかの α と m の値に対して，w を変化させて，結果を式(1)と比較し，平均と 95 %信頼区間を求めると，

$$D_M = 2\,D_P \left\lceil \frac{L}{w} \right\rceil \times (1.02 \pm 0.04)$$

となっており，理論とよく一致していた．

また，仮定(a)が成立していない場合，つまり平均メッセージ長 (L) がウィンドウサイズ (w) に比べてそれ程大きくない場合に，式(1)と比較すると，

$w = L$ のとき，$D_M \simeq 3\,D_P$

$w = 2\,L$ のとき，$D_M \simeq 2\,D_P$

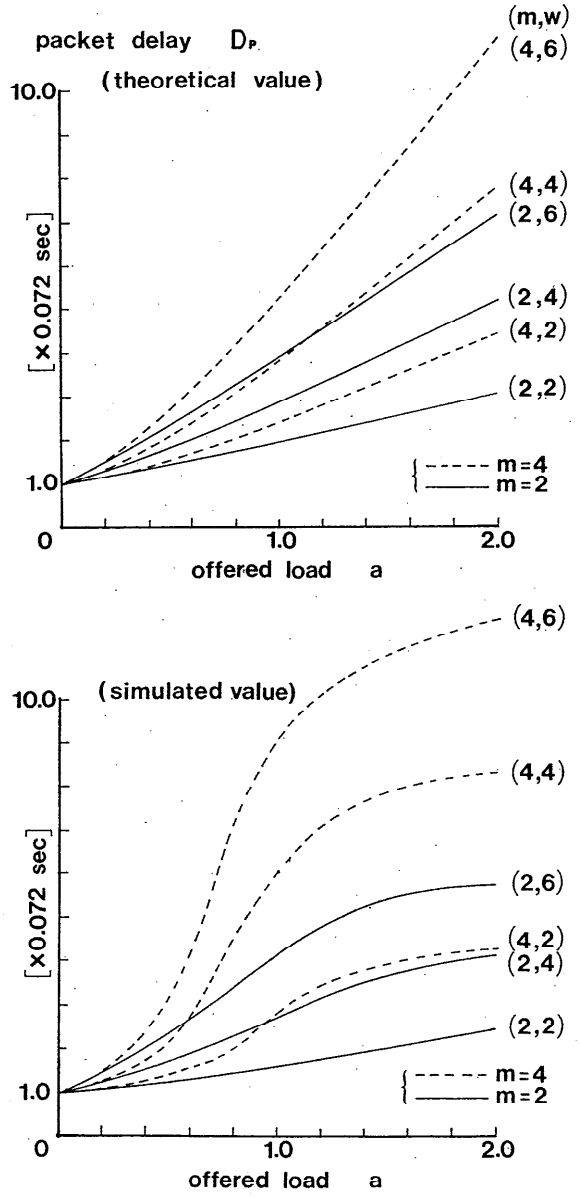


図12 与えられる負荷に対するパケット遅延の特性
Fig.12-Packet delay characteristics vs. offered load.

程度になっていた．ウィンドウが開いている機会が多くなって，データパケットが RR パケットを待たずに次々と送出されていることがわかる．

6.3 スループット

網に与えられる負荷 (a) に対するスループット (γ) の特性に，論理チャネル数 (m) の値がどう関与するかを調べる．図11に両結果を示したが，m や a を増加させた時の γ の増加率に関してはよく一致している．しかし，γ の絶対値に関しては，特に高負荷時に，理論値よりシミュレーション値の方が大きくなっている．これは，パケット交換により，物理リンクなどの網内の資源が一部の論理チャネルに独占されることなしに有効に利用されていることにより，**4.1** で述べた網のサービス率を越える転送能力が発揮されるためである．

また，**4.3** で予想した通り，負荷を一定として w を変動させたときの γ の値には有意な差は生じなかった．

6.4 パケット遅延

網に与えられる負荷に対するパケット遅延 (D_P) の特性に，論理チャネル数とウィンドウサイズ (m, w) の値がどう関与するかを図12に示す．低負荷時には式(9)の絶対値も式(10)と式(11)の比例関係も，両結果は非常によく一致している．また高負荷時には，シミュレーション値が頭うちになっているが，これは **3.3** で予想した通りである．高負荷時には式(9)は成立しないと考えられる．しかし，式(9)から導出された式(12)の比例関係 ($D_P \propto mw$) は，シミュレーション結果でもよく成立しており，D_P と $mw \sum_{i=1}^{n} (1/\mu_i)$ の比の値は，2.25 ± 0.21 であった．ここで 0.21 は標準偏差であり，平均値の 9 %に相当する．

7. むすび

エンドノード間でフロー制御を行う2種類の制御パラメータである論理チャネル数とウィンドウサイズの効果について評価した．待ち行列理論による解析の結果とシミュレーションの結果によって，ウィンドウサイズにも関係してメッセージ遅延とパケット遅延が比例する性質や，高負荷時にパケット遅延が2種類の制御パラメータに比例する性質などが明らかになった．

本論文で得られた結果によって，実際にウィンドウサイズの標準値を設定する手法の例を，以下に示す．

図1の形態の網で，0.5 の負荷が与えられエンドノード間で4本の論理チャネルが設定できる状態であるとする．また平均メッセージ長などの諸定数を表2のものとする．ここでパケット遅延をその下限値の 2.0

倍程度に抑えたい場合には，図12によるとウィンドウサイズの標準値を4以下に制限しなければならない．

このように，与えられる負荷や網の形態や要求されるパケット遅延の保証値などに応じて，フロー制御の制御パラメータを設定することができる．

また，式(10)で明らかになったウィンドウサイズとパケット遅延の比例関係などについては，メッセージ長がウィンドウサイズに比べて充分大きいという仮定 (a) が満足されれば，その他の仮定が満足されないような複雑な網にまで拡張できる結果であると考えられる．

しかし，本論文全体としては，評価の対象にできる網形態がかなり限定されている点，制御パラメータが網内ですべて同じであるとした点など，現実の網と必らずしも一致していない点も多い．さらに一般的な場合にも当てはまる評価が必要である．また，本論文では，パケットレベルのフロー制御について評価したが，今後は，上位のアプリケーションに近いレベルでのフロー制御や下位のリンクレベルでのフロー制御との相互干渉を評価することも必要である．

文　　献

(1) Davis, D.W. : "The Control Congestion in Packet-switching Networks", IEEE Trans. Commun., COM-20, 3, pp.546-550 (1972).

(2) 岡田，鈴木，須之内，手塚："パケット交換計算機網における等数パケットふくそう制御方式の解析"，信学論 (A), J59-A, 3, pp.208-215 (昭51-03).

(3) Bux, W., Kümmerle, K. and Truong, H. L. : "Balanced HDLC Procedure : A Performance Analysis", IEEE Trans. Commun., COM-28, 11, pp.1889-1898 (1980).

(4) 井上，中西，真田，手塚："HDLC手順を用いたデータリンクのスループット特性の解析"，信学論 (D), J65-D, 2, pp.202-209 (昭57-02).

(5) Pennotti, M. C. and Schwartz, M. : "Congestion Control in Store and Forward Tandem Links", IEEE Trans. Commun., COM-23, 12, pp.1434-1443 (1975).

(6) Takenaka, I., Kawasaki, K., Takeda, M. and Nomura, M. : "Evaluation of Flow Control Schemes for Packet Switched Network", Proc. ICCC, Kyoto, pp.141-146 (Sept. 1978).

(7) 松尾："ウィンドウ方式を用いたフロー制御の評価"，情報処理学会分散処理システム研究会資料，4-3 (1980-2).

(8) Gelenbe, E., Labetoulle, J., and Pujolle, G. : "Performance Fvaluation of the Protocol HDLC", Computer Network, 2 (Sept. 1978).

(9) Kiviat, P. K. : "Simulation with Gasp-II", Prentice-Hall Inc. (1969).

(10) Inose, H. : "An Introduction to Digital Integrated Communication System", Univ of Tokyo Press (1980).

(11) Inose, H. and Saito, T. : "Theoretical Aspects in the Analysis and Synthesis of Packet Communication Networks", Proc. IEEE, 66, 11, pp.1409-1422 (1978).

付　　録

網に与えられる負荷 (a) に対するパケット遅延 (D_P) の特性に (m, w) がどう関与するかを調べる．式(9)で

$$\lim_{a\to 0} D_P = \sum_{i=1}^{n} \frac{1}{\mu_i}$$

によって，式(11)が成立する．この値は，パケットが網内を転送される際に，待ちが全く生じていない場合に相当しており，D_P の下限値（理想値）となっている．図12は，D_P をこの下限値で正規化して示している．

次に，低負荷時の増加率について調べる．式(9)より，

$$D_P \simeq \sum_{i=1}^{n} \frac{1}{\mu_i}\left(1+\frac{w\theta_i a}{2(1-\theta_i a)}\right)$$

$$\therefore \frac{dD_P}{da} \simeq \sum_{i=1}^{n} \frac{w\theta_i}{2(1-\theta_i a)^2} \xrightarrow[(a\to 0)]{} \sum_{i=1}^{n} \frac{w\theta_i}{2}$$

これにより，式(10)が成立する．

最後に，高負荷時の特性について調べる．式(4)より，

$$D_P = \sum_{i=1}^{n} \frac{1}{\mu_i} + \sum_{i=1}^{n} \frac{w}{2\mu_i}\frac{a_i}{1-a_i}$$

$$= \sum_{i=1}^{n} \frac{1}{\mu_i} + \sum_{i\notin I} \frac{w}{2\mu_i}\frac{a_i}{1-a_i} + \sum_{i\in I} \frac{w}{2\mu_i}\frac{a_i}{1-a_i}$$

ここで，第1項は有限である．また，式(3)および $\lim_{a\to\infty} a_i = \theta_i$ であることから，

$$\lim_{a\to\infty}(\text{第2項}) = \sum_{i\notin I} \frac{w}{2\mu_i}\frac{\theta_i}{1-\theta_i} < \infty$$

第3項は，$a\to\infty$ で発散するが，式(2)，式(8)より，

$$\frac{a_i}{1-a_i} \xrightarrow[(a\to\infty)]{} \frac{a\sum_{k=0}^{m-1}\frac{(ma)^k}{k!}}{\sum_{k=0}^{m}\frac{(ma)^k}{k!} - a\sum_{k=0}^{m-1}\frac{(ma)^k}{k!}}$$

$$= \frac{a^m m^{m-1}(m-1) + a^m m^m}{a^{m-1}m^m + a^m m^m - a^{m-1}m^{m-1}(m-1) - a^m m^m}$$

$$= m-1+am$$

$$\therefore (\text{第3項}) \simeq \sum_{i\in I} \frac{w}{2\mu_i}(m-1+am)$$

これにより，式(12)の $D_P \propto mw$ が示された．

（昭和57年7月2日受付，10月18日再受付）

論　文

トランスピュータによる多層プロトコルの性能評価用シミュレータ

正　員　齊藤　忠夫[†]　　正　員　相田　　仁[†]
正　員　藤田　博文[†]　　正　員　ゾーカイ　サーダン[†]

A Simulator Using Transputers for Evaluation of Multi-Layered Protocols

Tadao SAITO[†], Hitoshi AIDA[†], Hirofumi FUJITA[†] *and* Zokaei SAADAN[†], *Members*

あらまし　データ通信の高速化に伴い種々プロトコルが提案されている．プロトコルの振舞いは複雑であり，その性能の評価について従来種々の研究が行われている．筆者らは従来ビット誤り率が無視できる場合についての理論およびシミュレーションによる評価を行ってきたが，ビット誤り率が無視できないが比較的低い場合の評価は困難であった．そこで，ネットワークノードおよびネットワークリンクにトランスピュータを用いたプロトコルシミュレータを作成した．本論文ではプロトコルシミュレータの構成と実現方法について詳細に述べ，また，シミュレータによって得られた結果について簡単に述べる．

キーワード　トランスピュータ，シミュレータ，多層プロトコル，伝送効率，高速通信

1. まえがき

近年，さまざまなアプリケーションにとって，100 Mbit/s から 1 Gbit/s 程度の非常に高速なデータ転送が必要となりつつある．データ通信においては，適切な誤り制御とフロー制御により，誤りのない通信を保証することが要求される．しかし，複雑なプロトコルでは，高速なデータ通信は困難である．

超高速のデータ通信を実現するためには，プロトコルの設計が重要であり，あるプロトコルを用いたときのネットワークの性能を評価しなければならない．このために，これまでさまざまな理論評価(1)~(3)やシミュレーションによる評価(4)~(6)が行われてきたが，こうした技術は複雑なネットワークにおけるプロトコルの性能を評価するには十分ではなかった．

超高速データ通信を想定した複雑なネットワークを評価するために，筆者らは現在，トランスピュータ(7)を用いて構成したネットワークシミュレータを作成している．トランスピュータは高速のデータ転送能力をもつマイクロプロセッサである．ネットワークノードとネットワークリンクをトランスピュータを用いて模擬することでネットワークのモデルを構成し，理論的評価では困難であった伝送誤りの発生する環境での性能評価を行う．

本論文は主にシミュレータの構成とその実現方法について詳細に述べ，更に，このシミュレータにより得られたシミュレーション結果の一部について簡単に述べる．

2. 背　　景

2.1 ハードウェアシミュレーションの必要性

一般に，OSI のような多層プロトコルにおいては，各層の遅延やプロトコル処理によるオーバヘッド等が複雑に影響するためモデル化が困難である．

文献(1)では，各層の処理を理想化して伝送効率を与える式を導出している．文献(2)では，各層の通信処理に対する命令数を測定することでスループットを推測している．文献(3)では，測定したプロセッサの通信処理能力に基づいて一つの層のスループットを計算している．文献(4)，(5)では，通信処理時間や伝搬遅延時間を加算することで解析的なシミュレーションを行い，伝送効率を求めている．こうした研究の結果，スループットの改善にはプロトコル処理時間の改善が不可欠であり，チェックサム計算がオーバヘッドの大きな要因となっているが，プロトコルパラメータを適切に選択することでスループットの改善が可能であることがわかっている．

しかし，実際のプロトコルの動作は複雑であり，ネットワークの構成や条件により影響されるため，上記

† 東京大学工学部電子工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan

のような理論的評価だけではエンドシステム間の動作を考察するには十分ではない．また，多層プロトコルをモデル化する際，各層を理想化して同一視し，プロトコル処理時間をデータ長によらずに固定としたり，各層のウィンドウサイズが等しいと仮定する場合には，現実性が乏しくなる．実システムのような複雑な環境を対象にして多層プロトコルの評価を行うためには，実際にプロトコル処理を行うハードウェアシミュレータが必要である．

実際に通信処理を行うプロトコル評価シミュレーションはワークステーションなどを用いても実現できるが，こうしたコンピュータで評価を行う場合，プロトコルそのものに起因する性能と，割込みやプロセス呼出し・システムコールなどの処理に伴うOSのオーバヘッドや，ワークステーションのハードウェア構成などの影響を分離して評価を行うことは困難である．

そこで本研究では，実際にプロトコル処理を行い，OSの影響をなるべく小さくし，伝送誤りなどのさまざまなネットワーク条件を考慮に入れた評価を行うことを目標にする．このために，単純なデータストリームを扱う10 Mbit/sの通信インタフェースを内蔵したトランスピュータを，通信用ノードなどのネットワークの基本要素として用いることによってシミュレータを構成することにした．

2.2 トランスピュータ

本論文では，トランスピュータ(T 425)を用いて各層のプロトコル処理および模擬伝送路を実現した．プログラムは並列記述言語occam 2[8]を用いて記述した．トランスピュータには

- 32ビットプロセッサ
- 優先度スケジューラ
- 高速メモリ
- タイマ
- リンクインタフェース

が内蔵されている．

実行待ちのプロセスはスケジューリングリスト上に保持される．このスケジューリングリストはプロセスの作業領域がリンクされたリストであり，図1のように，リストの先頭プロセスを指すレジスタ(Front pointer)と，最後プロセスを指すレジスタ(Back pointer)を用いて実現される．

プロセスは，入力待ち，出力待ち，あるいは遅延入力待ちとなるまで実行される．この間，プロセス・スイッチングは行われず，一つのプロセスが実行権を独占する．実行中のプロセスが上記の待ちに出会ったときは，このプロセスの命令ポインタが作業領域内に退避され，スケジューリングリストの先頭のプロセスが選択されて実行される．

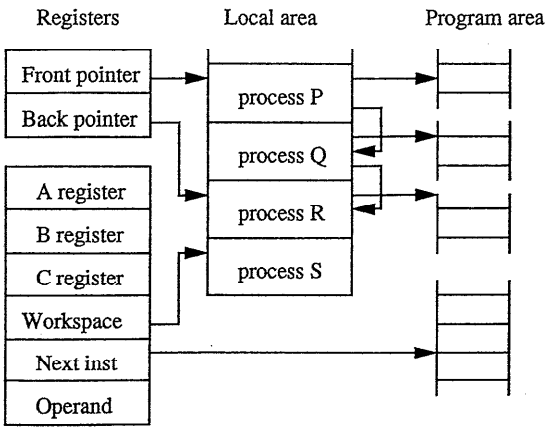


図1 トランスピュータでの並行プロセスの実行
Fig. 1 Execution of the concurrent processes in a transputer.

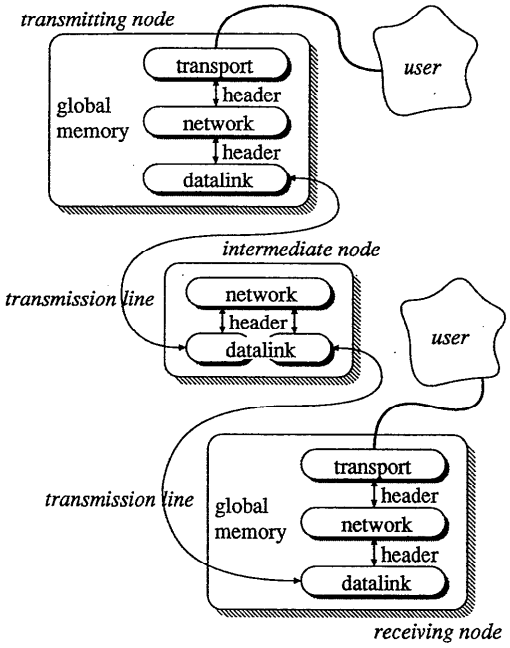


図2 シミュレーション対象のモデル
Fig. 2 Simulation model.

リンクインタフェースとプロセッサは，それぞれ独立に動作し，リンクを用いてメッセージが転送されている間，プロセッサは別のプロセスの実行を続けることができる．

2.3 シミュレーション対象のモデル

シミュレータは，図2のようなネットワークを評価の対象とする．送受信ノードはOSI (Open System Interconnection)参照モデルの論理レベル下位3層

(トランスポート層,ネットワーク層,データリンク層)の機能をもち，中継ノードやう回路を含むネットワークを介して接続されている．また，データコピーによるオーバヘッドを避けるため，層間ではヘッダ情報だけをやりとりし，実際のユーザデータは大域メモリ中に置かれるものとする．

3. シミュレータの構成とシミュレーションの詳細

3.1 システム構成

シミュレータシステムの構成を図3に示す．トランスピュータを組み込んだVMEボードをホスト計算機のバスに接続する．個々のトランスピュータは4Kバイトの内蔵メモリのほかに，それぞれ1Mバイトのメモリをもつ．

3.2 プロセスの割当て

本論文では中継ノードのないポイント-ポイント形態のシミュレータの構成について述べる．また，3層の処理を別々のプロセッサ上で並列実行するような構成も可能であるが，ここでは3層を同一プロセッサ上で実行する場合の構成について述べる．この場合，4個のプロセッサに対して図4のようにプロセスを割り当てる．それぞれのトランスピュータには次のような役割を与える．すなわち，送信ノードおよび受信ノードでは，3層の機能をもつプロセスが並行動作する．模擬伝送路では，両方向のそれぞれに遅延発生プロセスと誤り発生プロセスが並行動作する．テストユーザは，送信ノードにデータを与え，受信ノードからデータを受け取り，その間に費やされた時間を測定してスループットや伝送効率を計算する．

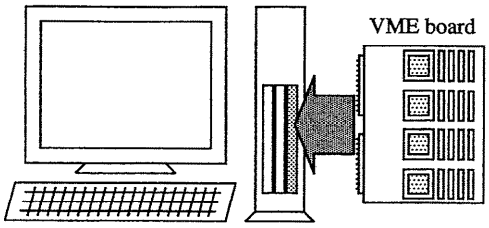


図 3 システム構成図
Fig. 3 System composition.

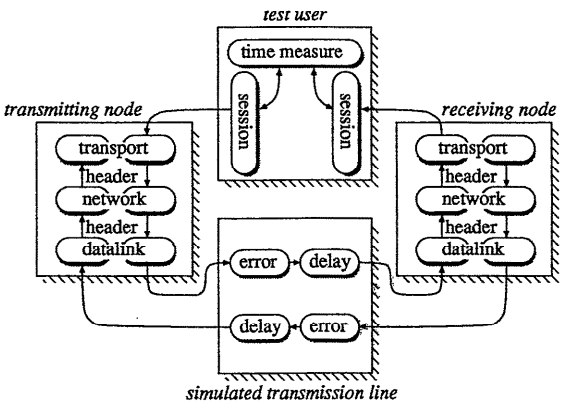


図 4 ポイント-ポイント形態でのプロセスの割当て
Fig. 4 Allocation of the processes in the point-point model.

3.3 ユーザデータの扱い

各層では図5のように，送信処理プロセスと受信処理プロセスが並行プロセスとして動作する．2.3の想定に従い，シミュレータは各層プロセス間ではヘッダ部分だけをやりとりし，実際のデータの移送は行わない．このため，図5のトランスポート層の上位とデータリンク層の下位に設けたプロセスで，ユーザデータの削除および追加を行う．また，このプロセスはリンクインタフェースとプロセッサとを同時に動作させる役割も果たす．但し，チェックサムなどの処理を行う場合は別に用意したバッファに対して行い，また，メモリコピーに要する時間は設定したメモリバンド幅とデータ長に従って計算する．

3.4 伝送誤り

伝送誤りは，伝送路を模擬するプロセッサの誤り発生プロセスで発生させる．誤り発生をシミュレートする場合，1ビットごとに乱数を生成し，このビットを誤りにするかどうかを決定するのが単純な方法である．しかし，伝送路をプロセッサ上のソフトウェアで模擬することを考えれば，この方法では伝送路の処理能力が極端(数百kbit/s程度)に低下してしまう．誤りをビットごとではなくフレームごとに発生させる方法

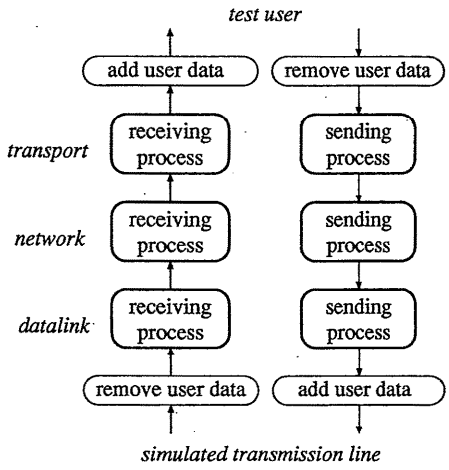


図 5 ユーザデータの削除と追加
Fig. 5 Adding and removing user data.

が考えられるが，あらかじめフレーム長がわかっていなければならない．

そこで，周期的に乱数を生成することで誤りを発生させることにした．例えば 10^{-6} のビット誤り率の場合，10^5 ビットごとに乱数を生成して参照し，1/10 の確率で誤りにする．

誤り発生プロセスでは誤りの発生したフレームは出力しない．従って，受信ノードではフレーム紛失による順序誤りとして扱う．誤りはデータリンク層またはトランスポート層で回復する．

3.5 伝搬遅延

伝搬遅延は模擬伝送路の遅延発生プロセスで発生させる．図6に示すように，受け取ったフレームをリングバッファの書込みポインタの指す位置に蓄え，そのときの時刻を記録して書込みポインタを一つ進める．読出しポインタの指す位置に蓄えたフレームが指定された遅延時間を経過すると，そのフレームを出力して読出しポインタを一つ進める．

このように，一方向当り約400Kバイトのバッファを用いて，10 Mbit/s の伝送に対して約300 ms までの遅延を実現できる．

3.6 伝送速度

トランスピュータ間のデータ転送速度(リンクの転送速度)は固定であり，VME ボード上で約10 Mbit/s である．このため，模擬伝送路の伝送速度もこの値に近いものになるが，伝送路を模擬するソフトウェアにより影響される．

そこで，模擬伝送路の伝送速度を以下のように定義する．

・図7のように，四つのトランスピュータに，送信プロセス，受信プロセス，伝送速度測定プロセス，模擬伝送路プロセスを割り当てる．

・模擬伝送路の各プロセスでは伝搬遅延時間とビット誤り率を0に設定する．

・送信プロセスは模擬伝送路にデータを連続して渡す．

・受信プロセスは模擬伝送路からデータを受け取る．

・伝送速度測定プロセスは，模擬伝送路が送信プロセスから最初のフレームを受け取り始めてから，受信プロセスに最後のフレームを渡し終えるまでの時間を測定する．

・転送したビット数を測定した時間で割ったものを，伝送速度とする．

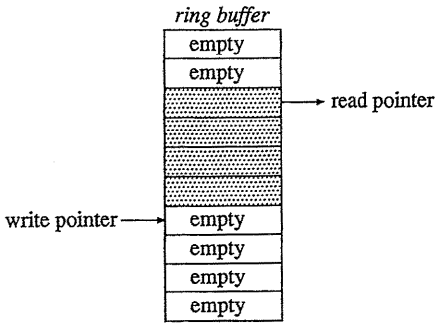


図 6　模擬伝送路における伝搬遅延の発生
Fig. 6　Generation of propagation delay in the simulated transmission line.

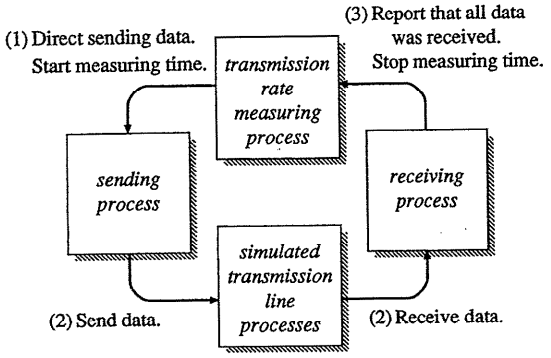


図 7　模擬伝送路の伝送速度の測定
Fig. 7　Measuring transmission rate of the simulated transmission line.

この伝送速度はシミュレーション中，一定の値であり，シミュレータで伝送速度を変化させることは不可能である．しかし，プロトコル処理時間と伝搬遅延時間を変化させることで，相対的に伝送速度を変化させたときの評価を行うことができる．例えば，伝送速度を 100 Mbit/s に設定する場合は，伝送速度が10 Mbit/s の場合に比べて，プロトコル処理時間と伝搬遅延時間を10倍にすればよい．プロトコル処理時間を仮想的に大きくするために，プロトコル処理の際，時間測定をして無駄ループを行う．この時間測定は，プロトコル処理のつど行われる．

時間調整をタイマ入力待ち(遅延入力待ち)によって行わないのは，プロセススイッチングが行われて別の並行プロセスが実行されてしまう可能性があるからである．また，送信処理・受信処理とも，入出力を行うところでプロセススイッチングが行われる可能性があるので，待機状態にあった時間を計上しないために，処理時間の測定は入出力開始の直前で一時停止し，入出力終了の直後に再開するようにした．

4. シミュレータによる性能評価例

シミュレータには任意のプロトコルを実装することができるが，本論文では評価するプロトコルとして，データリンク層はHDLC (High level Data Link Control)手順のBAC (Balansed operation ABM Class：平衡型非同期平衡モードクラス)の拡張モード，ネットワーク層は中継・分割・フロー制御の機能をもつプロトコル，トランスポート層はOSI/TP 4の拡張モードを用いる．

本章では上記のように構成したプロトコルシミュレータによるプロトコル性能評価について，評価例の一部を示す．

4.1 プロトコル処理時間

プロトコル処理時間を以下のように定義する．

[送信処理時間]　送信側で上位層からSDUを受け取ってから，PDUを生成して下位層に渡し，次のSDUを受け取る準備が完了するまでに費される時間．

[受信処理時間]　受信側で下位層からPDUを受け取ってから，SDUを生成して上位層に渡し，次のPDUを受け取る準備が完了するまでに費される時間．

[応答処理時間]　下位層からその層の応答PDUを受け取ってから，処理をした後次のPDUを受け取る準備が完了するまでに費される時間．

3.6で記したように，伝送速度を設定するためにプロトコル処理時間を調整する過程でプロトコル処理時間を測定するが，これを記録することで処理時間を得ることができる．図8は以下の条件で測定したときの処理時間である．

- ウィンドウサイズは3層とも127とする．
- 応答間隔は3層とも63とする．

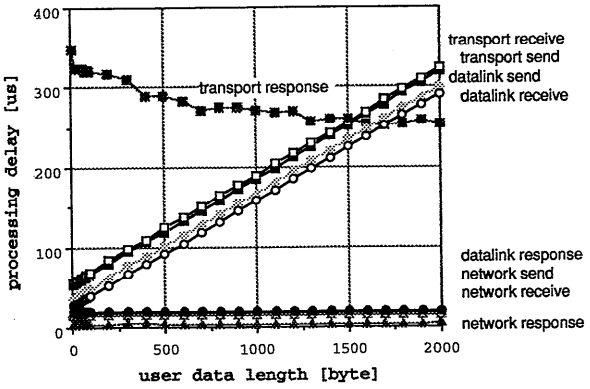


図 8　測定したプロトコル処理時間
Fig. 8　Measured protocol processing delay.

- メモリバンド幅は120 Mbit/sとする．
- 伝搬遅延時間は5 msとする．
- チェックサム計算は行わない．

ウィンドウおよび応答間隔の制御はバイト単位ではなく，PDU単位で行う．

トランスポート層とセション層の間，およびデータリンク層と物理デバイスの間で，データのメモリ間コピーをそれぞれ1回ずつ行うことを想定しているので，シミュレータではトランスポート層とデータリンク層でこれに相当する時間だけ無駄ループをしている．このときのメモリバンド幅が120 Mbit/sであるので，トランスポート層とデータリンク層の送信処理時間および受信処理時間については，図8における傾きの逆数が約60 Mbit/sになっている．ネットワーク層ではヘッダ部分だけを扱うので，送受信処理時間はユーザデータ長とは無関係である．

トランスポート層では各データパケットごとにタイマを用いているので，タイマの管理に時間がかかり，応答処理時間がほかの層よりも大きい．以下，トランスポート層の応答処理時間について考察する．ユーザデータ長が2,048バイトの場合，応答処理時間256 μsの内訳はおおよそ

(a)　タイマを停止する処理：205 μs
(b)　タイマを更新する処理：20 μs
(c)　その他の処理：28 μs

となっていた．(a)の処理では確認されたデータパケットの数のタイマを停止するので，この処理時間は応答間隔に依存する．測定した値は応答間隔が63の場合のものであり，応答間隔がnの場合は

$$\frac{205}{63}\cdot n \quad [\mu s]$$

であると考えられる．

(b)の処理は送出したデータパケットのうち未確認のデータパケットの数のタイマを更新するので，この処理時間は各層の送信および受信処理時間や伝搬遅延時間に依存する．データ長が大きいほど送受信処理時間が大きく，未確認のデータパケットの数は減少するので，(b)の処理に要する時間は小さくなる．図8でデータ長と共に応答処理時間が減少しているのはこのためであり，図よりユーザデータ長がUの場合の(b)の時間はおおよそ

$$20-0.0394\cdot(U-2048) \quad [\mu s]$$

である．

これより，応答間隔がn，ユーザデータ長がUの場

合，トランスポート層における応答処理時間は

$$\left\{\frac{205}{63}\cdot n\right\}+\{20-0.0394\cdot(U-2048)\}+\{28\}$$
$$=129+3.25\cdot n-0.0394\cdot U \quad [\mu s]$$

程度である．

4.2 スループットと伝送効率

本論文ではネットワークの性能をスループットまたは伝送効率によって評価する．伝送効率はスループットを正規化したものであり，その定義は以下に示す通りである．

伝送効率

＝スループット÷伝送速度

$$=\frac{\left(\begin{array}{c}\text{受信者が受け取る}\\ \text{情報ビット数}\end{array}\right)}{\left(\begin{array}{c}\text{上記の情報ビットが受け取られ}\\ \text{るために必要な全時間}\end{array}\right)\times\text{伝送速度}}$$

「受信者が受け取る情報ビット数」と「伝送速度」はパラメータとして与えられる値であるので，パラメータを設定し，「伝送に必要な全時間」を測定することでスループットと伝送効率を求める．

4.3 誤り回復方式の影響

高速通信では，選択的再送(SREJ)による誤り回復が適している．TP 4 による誤り回復は基本的にタイムアウトに基づいており，あまり効率の良いものではない[(9),(10)]．この欠点を改善するため，TP 4 の機能に，選択的に再送要求をする機能を付加したモデル(TP 4+SREJ と記す)についてもシミュレーションを行った．このモデルでは，受信側で受信できなかった PDU のすべてに対して選択的再送要求(Selective Reject : SREJ)パケットを返送することで，特定の PDU の再送を要求する．

図 9 において，通常の TP 4 と，受信側が再送要求を行う TP 4+SREJ とを比較すると，後者の方が全体的に効率が良い．ビット誤り率が 10^{-7}～10^{-6} の領域では，伝送効率の差は 30～40%程度である．再送要求によりネットワークのふくそうを招くことはほとんどないと考えられるので，タイムアウトだけによる再送開始よりも，再送要求による再送開始を併用する方式は優れていると言える．

また，誤り率が大きいところで HDLC (SREJ 方式)よりも TP 4 の方が有利になっている．これは，HDLC では SREJ 状態中に発生した誤りに対しては効率良く回復できないのに対し，TP 4 ではウィンドウ内のフレームはいくつでも保持しており，より強力な誤り回復を行うからである．

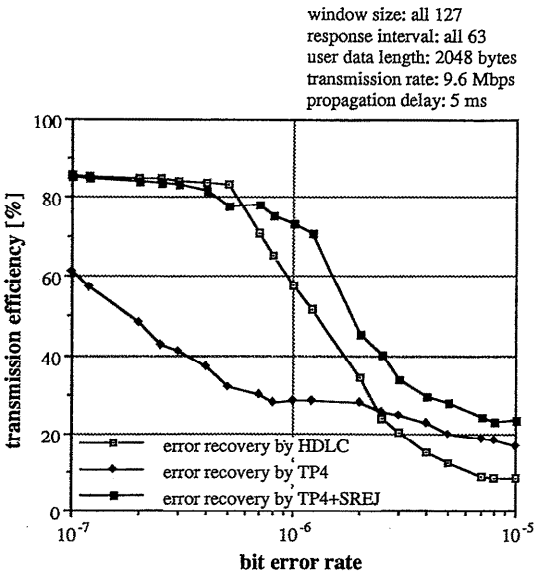


図 9 誤り回復方式による違い
Fig. 9 Effect of the error recovery method.

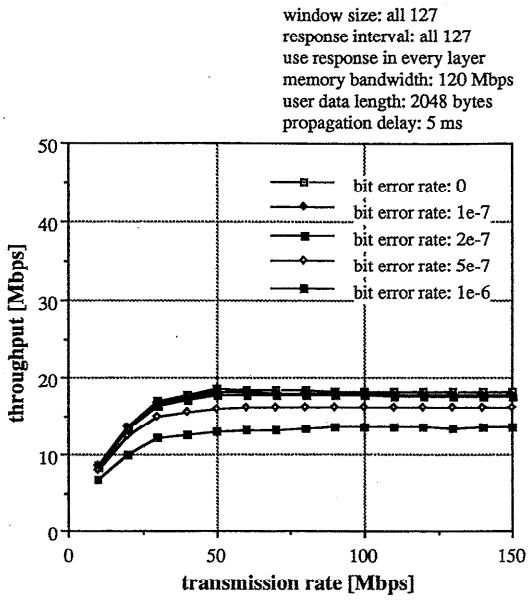


図 10 伝送速度の影響
Fig. 10 Effect of the transmission rate.

このように，TP 4 に選択的再送要求を行う機能を追加したプロトコルでは，ビット誤り率の広い領域で良好な結果を示した．

4.4 伝送速度の影響

前述した方法で伝送速度を変化させ，スループットを測定した(図 10)．伝送速度が大きくなるにつれ，伝送効率は低下しており，伝送誤りのない場合でもスループットは 18 Mbit/s 程度で飽和している．この飽和

の主な要因はプロトコル処理とメモリアクセスによるものと考えられる．スループットの向上にはこれらの改善が不可欠である．

5. むすび

多層プロトコルを用いたネットワークの性能評価を行うために，トランスピュータを用いて実際にデータを交換するシミュレータを作成した．このシミュレータを用いて，さまざまなプロトコルパラメータが伝送効率に与える影響について調査することができる．本論文にはシミュレーション結果の一部だけを示したが，ウィンドウサイズ，応答間隔，伝搬遅延時間，データ長，伝送誤り率，伝送速度など，多くのパラメータがスループットに影響を与える．また，中継ノードのある場合や，メモリバンド幅やチェックサム計算プロトコル処理の並列化などが与える影響についても評価している．作成したシミュレータでは，複雑なモデルや伝送誤りが無視できない場合など，理論的評価が困難な場合でも性能を評価できることは，実用化の上で意義が大きい．

今後はレート制御によるフロー制御について調査していくつもりである．

文　献

(1) 齊藤忠夫，福井俊之："多層構造プロトコルをもつネットワークの伝送効率の評価式"，信学論(D-I)，**J73-D-I**，4，pp. 424-431(1990-04).

(2) Clark D. D., Jacobson V., Romkey J. and Salwen H.: "An Analysis of TCP Processing Overhead", IEEE Trans. Commun., pp. 23-29(June 1989).

(3) Saulnier E. T. and Mitchell R. J.: "Impact of Implementation on XTP Throughput Performance", IEEE ICC '92, pp. 976-980(1992).

(4) 齊藤忠夫，白畑厚志："多層構造プロトコルをもつネットワークの伝送効率の評価"，信学論(D-I)，**J72-D-I**，7，pp. 562-569(1989-07).

(5) Saito T., Aida H. and Fu H.: "Protocol Parameter Selection and Throughput Evaluation for Gigabit Per Second High Speed Data Communications", 11 th ICCC, pp. 409-414(Sept. 1992).

(6) Meister B. W.: "A Performance Study of the ISO Transport Protocol", IEEE Trans. Comput., **40**, 3, pp. 253-262(March 1991).

(7) INMOS Ltd.: "Transputer Reference Manual", Prentice Hall(1988).

(8) INMOS Ltd.: "occam 2 Reference Manual", Prentice Hall(1988).

(9) La Porta T. F. and Schwartz M.: "Architectures, Features, and Implementation of High-Speed Transport Protocols", IEEE Network Mag., pp. 14-22(May 1991).

(10) Zitterbart M.: "High-Speed Transport Components", IEEE Network Mag., pp. 54-63(Jan. 1991).

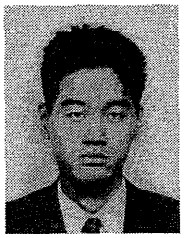
(平成5年11月22日受付，6年4月14日再受付)

齊藤　忠夫

昭38東大・工・電子卒．昭43同大大学院博士課程了．東大工学部講師，東大助教授，カリフォルニア工科大学研究員などを経て現在東大教授．ディジタル通信方式，コンピュータネットワーク，分散処理などの研究に従事．著書に「電子回路入門」(昭晃堂，昭52)，「ディジタル通信網入門」(オーム社，平2)などがある．昭42年度，昭61年度論文賞，昭48年度業績賞など受賞．電気学会，情報処理学会，IEEE各会員．工博．

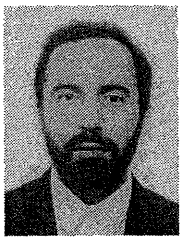
相田　仁

昭55東大・工・電気卒．昭60同大大学院博士課程了．同大工学部電気工学科助手．昭61同講師．SRIインターナショナル(米国)客員研究員を経て平2東大工学部電子工学科助教授．工博．コンピュータネットワーク，不均質分散処理システム，超並列コンピュータなどの研究に従事．著書に「第5世代コンピュータ入門」(共著)がある．情報処理学会，人工知能学会，ソフトウェア科学会，IEEE，ACM各会員．



藤田　博文

平4東大・工・電子卒．平6同大大学院修士課程了．現在同大大学院工学系研究科電子工学専攻博士課程．コンピュータネットワークの評価に関する研究に従事．電気学会，IEEE各会員．



ゾーカイ　サーダン

1976テヘラン大学電気工学修士了．1976から1978まで，パリ国立通信上級学校(ENST)コンピュータネットワークを学ぶ．1979から1989まで，テヘラン技術大学カジャナセルダントゥシ電気工学科で講師．1990から東京大学でコンピュータネットワークに関する博士課程に在籍．通信プロトコルの性能評価などの研究に従事．

6.　ハブ型ローカルエリアネットワーク

○ A star type local area network -MUITISTAR- having simultaneous multiple transmission capability

○ 同時多重伝送機能をもつスター形ＬＡＮ

ISS-87 MARCH 15-21, 1987, PHOENIX ARIZONA

A STAR TYPE LOCAL AREA NETWORK - MULTISTAR - HAVING SIMULTANEOUS MULTIPLE TRANSMISSION CAPABILITY

Tadao Saito, Hiroshi Inose and Tomoaki Kawai

Department of Electrical Engineering
Faculty of Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113 Japan

ABSTRACT

In the present paper, a new scheme for star type switching network called "MULTISTAR" is proposed in which a number of transmission media are provided in the central node to realize a number of simultaneous communication. Messages with headers applied to the central node contend for transmission media and messages which won contention are delivered to the destination station. The central node includes input side arbiters, switch modules and output side arbiters. Input side arbiters are used to arbitrate the contention to the switch modules and the output side arbiters are used to arbitrate the possible contention to the destination station. To cope with the corruption of data caused by collision of messages at either stage of arbiters, collision not detected (CND) signal is returned at media access level from the destination to the source. Detailed design of the system and the protocol of LAN as well as the result of performance evaluation are described in the paper.

1. INTRODUCTION

To create a flexible local communication system, variety of local area network schemes have been devised. The origin of local area network technology came from two traditional system,i.e., a bus interconnecting computer peripherals and a private branch exchange for telephone. The former technology is basically bus type system or ring type system which is the modified version of bus type system. The latter is basically star type system.

The principal features of bus or ring technology is the distributed control and header driven switching principle. Using the header driven switching in which data are routed according to the destination information included in the header, the destinations of series of messages consecutively transmitted from one station may be different for each message. In addition even if one station is transmitting a series of messages to another station, the station can receive a message from a third station. These flexibility of header driven switching is well suited for computer communication. On the other hand, local area network system based on PBX technology is operated by preset up principle. In the preset up switching system, one station can communicate to only one other station. In other words, consecutive messages can never be routed to different stations unless connection is changed through time consuming disconnection and connection procedures.

If packet switching principle is used in star type network, header driven switching suited for computer communication can be realized in local area network. In packet switching network, since each of packet is stored in the storage of switching node, the node requires fairly large capacity storage and it is usually difficult to realize high throughput and fairly long delay is caused by the switching node.

Therefore, local area network technology is featured by header driven principle and high throughput, using header driven storageless principle. However two problems can be identified in this technology. The first problem is the maintenance of the system. Since the origin of LAN technology is small scale experimental computer network, in some scheme of LAN technology, malfunction of one node may cause total system down. Since switching control function is distributed, if trouble happens, it is some times difficult to locate and fix troubles. Therefore, bus or ring type topology is only suited for small scale or well controlled environments. For large scale commonly used LAN, modification of ring topology called star type ring is frequently used.

Second problem of header driven storageless LAN is the capacity limit. In bus or ring type LAN total throughput is limited to the transmission speed of one transmission media.

Another types of star type LAN for large scale application are contention bus LAN including detakit LAN by ATT.[1],[3] In star type LAN, although arranged in star type, the capacity of the total system is limited to the capacity of one transmission media. On the other hand, in preset up type switching system, a number of simultaneous connection can be set up concurrently. Therefore, total capacity can be increased by increasing the possible number of simultaneous connections.

The paper proposes a new schemes for strageless header driven type LAN called MULTISTAR in which total capacity can be increased by increasing possible number of simultaneous connections. In the MULTISTAR LAN, capacity limit and maintenance problems inherent in conventional LAN technology are solved. In this paper principle, design and protocols for MULTISTAR LAN are described as well as the outline of capacity evaluation.

2 PRINCIPLE OF MULTISTAR LOCAL AREA NETWORK

The basic principle of many conventional local area network is broadcasting type communication. The header or a whole message from one station is broadcasted to all the other stations, and each of station receives the header and if the destination address contained in the header is identical to the address of the station the message is accepted by the station. Using this principle, it is impossible to provide simultaneous transmission capability because only one broadcasting is possible in one transmission media.

Therefore in the proposed MULTISTAR, a number of switch modules which have the capability to identify the destination address to route the message to the destination station are provided separately from stations. The central node is composed of a input side arbiter module(IAM), switch modules(SWM), and output side arbiter modules(OAM) as is shown in Fig.1.

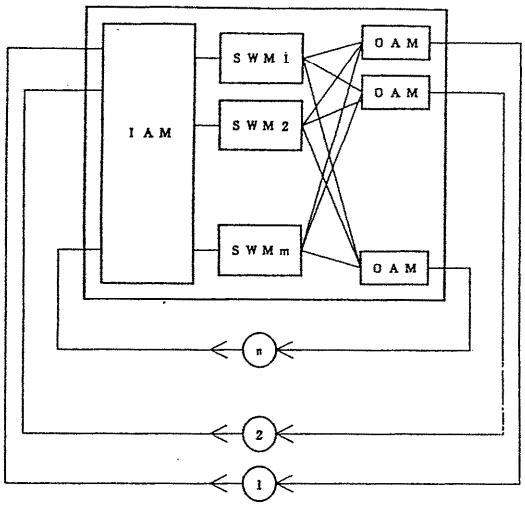


Fig.1 Structure of central node of MULTISTAR LAN

The input side arbiter is used to arbitrate the contention to switch modules. The switch module selects the destination to route the message. In a case that two or more switch modules select a destination station simultaneously, contention is solved by the output side arbiter.

In the example shown in Fig.1, n stations and m switch modules are provided. Therefore, in the central node, maximum number of simultaneous communication is m. Input side arbiter have n inputs and m outputs and select m out of n. Then each of m switch modules selects the destination and applies the message to the destination. Output side arbiters are provided for each of destination stations. An output side arbiter module has m input and one output.

The structure of a switch module is shown in Fig. 2. The switch module has a shift register having the capacity to store the destination address in the header of the massage. The address is detected by the address detector and applied to the decoder. The decoder controls the switch to route the massage to one of n destinations and hold the selected status for the duration of complete message. In this switch module a message is

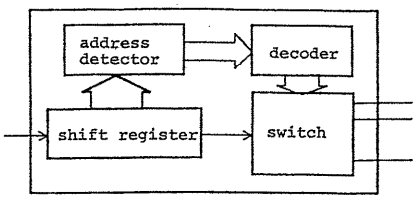


Fig.2 Structure of switch module

delayed by the duration needed to receive destination address in the header. The number of switch modules required is determined by the amount of traffic to be handled by the central node.

In this principle, if two switch modules apply a message to one destination, the output side arbiter pass the first message and the other message is discarded. In addition, messages are discarded in the following cases.

@ when the switch modules to be used is occupied.
@ when the destination station is receiving message from another station.
@ when some malfunction happens.

To confirm that message has been handed to the destination, some measure for acknowledgement is needed. For this purpose a media access control layer response called "collision not detected" (CND) response is to be transmitted back to the source station after the reception of the header of the message at the destination station.

CND response is to be transfered without collision, because if CND is also discarded by collision, message should be transmitted again and will cause some confusion. To provide guaranteed path for CND responses, we have following two possibilities.

1) Providing both way transmission capability in a switch module and hold the switch module selection status for a duration enough to return a CND response.

2) Providing a separate network for reverse direction and use the separate dedicated network for CND response.

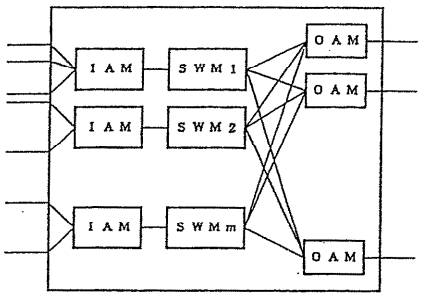
In the latter case, CND can be returned just after the reception of header by the receiving station. The dedicated network can be implemented much simpler than the network for message transfer. Therefore in this system, the latter method is adopted. In this scheme CND response can be received during the transmission of message. This means that higher layer protocol including logical link control is not affected by adopting the MULTISTAR LAN protocol in the media access control layer.

3. DESIGN OF MULTISTAR LOCAL AREA NETWORK

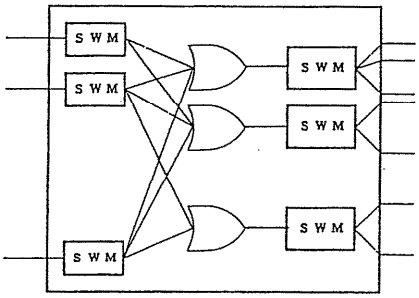
3-1. Design of the message and CND transfer route

The most important problems to implement the star type LAN is the design of arbiters. The arbiter must select one input quickly and must accommodate a large number of inputs. In general, arbiter design is the more complicated as the number of inputs increases.

Therefore, the n input and m output arbiter in the input side is reduced to m arbiters having n/m inputs and one output each. The structure of thus reduced network for message transfer is shown in Fig.3(a).



(a) Message path



(b) CND path

Fig.3 Structure of message path and CND path

Corresponding to this the network for CND transfer is shown in Fig.3(b). In the CND network, at the input, a switch module is provided for each of stations. This corresponds to the output side arbiter in the message transfer network. Corresponding to the switch modules in message transfer network, OR gates are provided in the CND transfer network. The first switch module decode the destination station included in the CND response and route the CND response to the OR gate which is in pair relation with the switch module in the message transfer network. The transfer of CND response will never encounters a collision because only one message is transfered through a switch module in the message transfer network. In pair relation with the input side arbiter in the message transfer network, m switch modules are provided to route the CND response to each of stations.

At each of the station, there provided a message/CND separation unit shown in Fig.4, to select one of signal transfer paths. If the signal is a message, it is routed through the network of Fig. 3(a) and if the signal is a CND, it is routed through the network of Fig.3(b). The overall network including message path and CND path is shown in Fig.5.

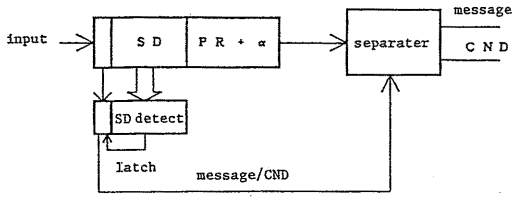
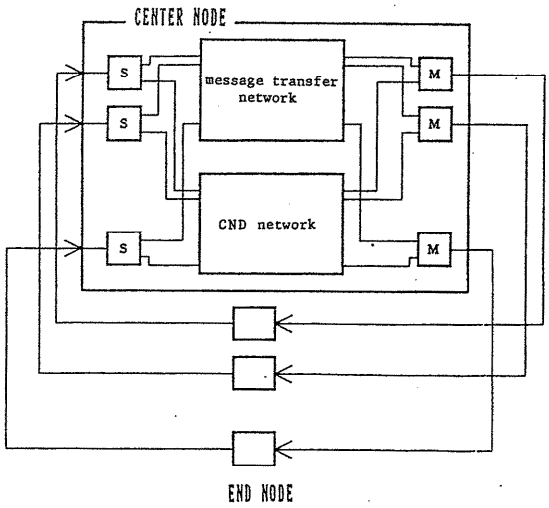


Fig.4 Message/CND separation unit



S: separater
M: mixer

Fig.5 Overall Structure of the MULTISTAR LAN

3-2. Physical signal code

In this system, to simplify the signal transmission, differential Manchester code is used for transmission interface between the center node and each of stations. Inside the center node, NRZ format is used.

All the stations are synchronized by the central node. For phase synchronization, preamble of one octet is used. The preamble pattern is "01010101"

3-3. Frame structure

Figure 6 shows the frame structure of a message. The header is provided with a destination address (DA), a source address (SA). The control field is just to identify message or CND.

In addition to them, a message data field is included. In CDN there is no data field. A start delimiter (SD), an end delimiter (ED) and frame check sequence (FCS) of 4 bytes are provided as LAN standard. In CND response there is no FCS field therefore the format is as shown in Fig.7.

P R	S D	CTR	D A	S A	D A T A	FCS	E D

PR : preamble
SD : start delimiter
CTR: control
DA : destination address
SA : source address
FCS: frame check sequence
ED : end delimiter

Fig.6 Frame structure of message

P R	S D	CTR	D A	S A	E D

Fig.7 Frame structure of CND

The CND response is transmitted just after the receiving station received the header of message. In other words, the result of error check by frame check sequence is not reflected to the transmission of CND response. Therefore the CND does not means that the message is received correctly in logical sense but just means that the message is physically delivered to the destination station. As the start and end delimiters, flag pattern of standard HDLC frame, that is 01111110 is used. Therefore "0" insertion and deletion circuits are provided at each station.

3-4. Design of arbiters

Fig.8 shows the structure of the arbiter. The

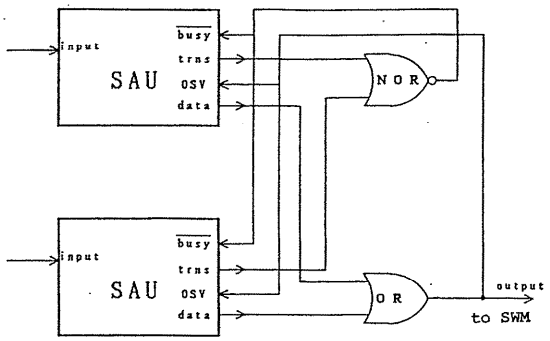


Fig.8 Structure of arbiter

arbiter is composed of plurality of subarbiter units (SAU) the number of which is identical to the number of inputs to the arbiter. The SAUs are interconnected by a OR gate and a NOR gate. Message signal is applied to the arbiter through "input". If a SAU won contention, the message is transmitted to the switch unit through the "output", which is the output from the OR gate. The message output from the SAU is called data. The busy lead to the SAU is "0" when message is passing through the arbiter else "1". In other words, the busy lead is used to inhibit SAU when the arbiter is occupied by other SAU. The lead trans means that SAU in transmitting message. OSV is the output supervisory lead to monitor the output lead of the arbiter. If the output of arbiter is not identical to the signal applied from the SAU, the SAU turns off the transmission.

The circuit diagram of the SAU is shown in Fig.9.

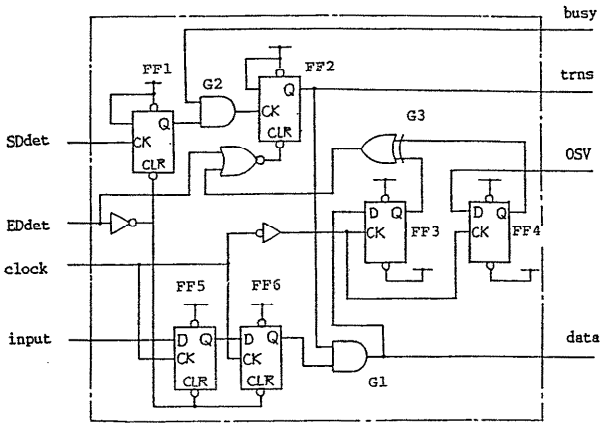


Fig.9 Circuit diagram of subarbiter unit.

The signal SD det and ED det mean the detection of start delimiter and end delimiter respectively. These signals are applied from the message/CND separation circuit. Here the inputs to the SAU are NRZ signal and clook (CK).

When SD signal is detected, the FF1 is set. If busy signal is "1", i.e. the arbiter is idle, the FF2 is set. This opens the gate G1 which applies the message to the data output delayed by two flip-flop stages i.e. FF5 and FF6. These two stages are provided to compensate the delay caused by FF1 and FF2. At this time trans signals is "1" and busy signal becomes "0" through the NOR gate of the arbiter.

When SD signal is applied, if arbiter is already busy, i.e. busy="0", then gate G1 is closed and FF2 is not set. After ED is detected FF1, 2, 5 and 6 are reset.

In the situation of collision, it is detected by the exclusive OR gate G3. The output of the arbitor is applied to the FF4 and the output of the SAU is applied to the FF3. the outputs are compared by the gate G3 and if not identical to the output from the gate G3 resets the FF2.

4. PROTOCOL FOR THE MULTISTAR LOCAL AREA NETWORK

The operation of each station which communicate through the LAN is very simple.

4-1. Transmission procedure

Each station has a transmission buffer having the capacity of at least one message. Transmission procedure starts after the reception of a message in the buffer in the following way.

1) A station can transmit a message when the station in not receiving other messages.
2) The station waits for a predetermined time The duration must be longer than the sum of the sound trip delay and time needed to CND.
3) If CND is received within τ, the source and destination addresses are tested to confirm that the CND is the response to the message which has been sent by the station. If it is confirmed, the station is reset to the initial state.
4) If CND is not returned within τ, transmission is terminated and the transmission procedure is repeated until success.
5) After transmission is started, the source station is not permitted to receive other messages before CND response is received.

4-2. Reception procedure

The reception procedure is as follows;

1) When message arrived, it is confirmed that the signal is the message and the destination address is correct.
2) If correct destination address is confirmed, CND response is transmitted, and the message is accepted in the receive buffer.
3) If destination address is not correct, the message is discarded.

5. PERFORMANCE OF MULTISTAR LOCAL AREA NETWORK

Denoting the number of stations by n and the message arrival rate from each station by λ_0, if the message generation rate from each station is equal, the total arrival rate λ is

$$\lambda = n\lambda_0 \quad (1)$$

The following notions are used in the analysis

- m : number of switch modules
- To : time needed to transfer a message through the central node.
- T_{rt} : round trip time between a station and the central node
- T_d : time needed to transfer a message when there is no collision
- : retransmission time for a message
- K : number of bits corresponding to delay at the central node
- v : transmission speed
- L_p : average message length
- t : time needed to generate CND after reception of message header

The following relations hold between parameters.

$$T_o = (K + L_p)/v \quad (2)$$

$$T_d = T_{rt} + T_o \quad (3)$$

and the value of τ should be

$$\tau > 2T_{rt} + 2K/v + t \quad (4)$$

Selecting the minimum value for

$$\tau = 2T_{rt} + 2K/v + t \quad (5)$$

If collision is neglected the usage of a switch module is

$$U_s = \lambda T_o/m \quad (6)$$

The usage of a switch module taking retransmission into account is

$$U = U_s \{ 1 + \eta(m-1)/m \} \quad (7)$$

Where η is the average usage of a station.
The total efficiency of a switch module is

$$S = \lambda L_p/vm \quad (8)$$

The value of η is

$$\eta = mS/n = \lambda L_p/vn \quad (9)$$

Denoting the success probability of transmission by P_o average transmission time t_f for a message is

$$t_f = T P_o + (T_d + \tau)(1 - P_o)P_o + (T_d + 2\tau)(1 - P_o)^2 P_o + \text{-----} \quad (10)$$

Since $P_s = 1 - U$ (11)

$$t_f = T_d + \tau U/(1-U) \quad (12)$$

Normalizing the value of t by average transmission time,

$$T = L_p/v \quad (13)$$

the normalized delay is

$$D = \{ T_d + \tau U/(1-U)\} \; v/L_p \quad (14)$$

As a numerical example assuming that the transmission speed is 1 Mbps, average cable length is 2km, number of stations is 50 and average message length is 1000 bits, delay at the central node is 100 bits, and the delay to generate CND is 100 bits, the relation normalized delay versus message arrival rate is given by Fig.10. As the

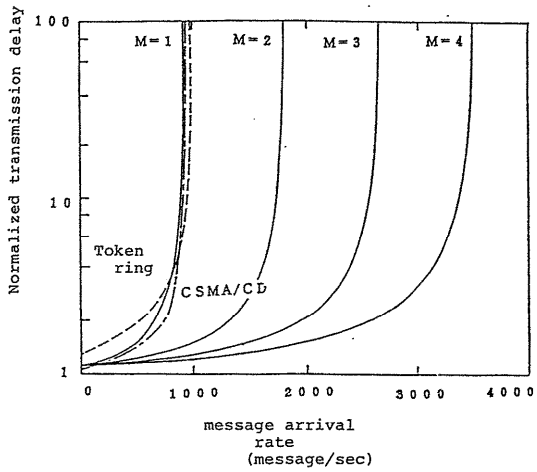


Fig.10 Normalized transmission delay versus message arrival rate

reference, the similar relations for CSMA/CD and token ring system is also shown in the figure.

This indicates that the proposed system is a little less efficient when only one switch module is used compared with conventional LAN schemes, however, the capacity can be increased roughly proportional to the number of switch modules.

6. CONCLUSION

This paper proposed a new scheme for header driven storageless switching system called MULTISTA which is adequate for local area network application. The scheme has capability to expand the capacity proportional to the number of switch modules. This scheme is adequate for large scale local area network and if needed to more general application.

REFERENCES

(1) A.S. Acampora, M.G. Hluchyj and C.D. Tsao : A Centralized Bus Architecture for Local Area Network, Proc. ICC pp.932-938 (1983)

(2) W. Bux : Local Area Network : A Performance Comparison, IEEE Trans on Commun., Vol COM-29 No.10, pp.1465-1873 (Feb. 1981)

(3) A. Albanese : Star Network With Collision-Avoidance Circuits. ATT Bell Syst. Tech Jourl Vol.62 No.3 pp.631-638 (March 1983)

論　文

同時多重伝送機能をもつスター形LAN

正　員　斉藤　忠夫[†]　　正　員　相田　　仁[†]　　准　員　井上　一郎[†]

A Star-Type Local Area Network Having Simultaneous Multiple Transmission Capability

Tadao SAITO[†], Hitosi AIDA[†], *Members and* Ichiro INOUE[†], *Associate Member*

あらまし　従来のヘッダ駆動ストレージレス(HDSL)形のLANでは，一般的に同時に複数の伝送を行うことができず，システムの規模拡大などによるトラヒックの増加に対応することが困難であった．本研究ではセンタノードに複数の交換機能を設置することで同時多重伝送を可能にしたスター形LAN（マルチスターLAN）を提案している．マルチスターLANはセンタノードの交換機能により，HDSL形の通信を行うが，交換機能の使用権の付与には，グループ化されたエンドノードからのデータユニット同士の競合，調停による方式をとっている．本論文ではこうしたマルチスターLANの構成と動作の概要を述べた後，負荷・遅延特性の理論的，実験的評価を行う．これにより，マルチスターLANが，低負荷時の遅延時間が小さく高負荷時の特性の悪化も少ない上，負荷の増加に対して柔軟性をもっていることが示される．

1. まえがき

多数のノードの間で通信を行うとき，情報のあて先を決定する方法には，回線交換のように情報が与えられる前にあらかじめ回線を設定しておく予設定方式と，パケット交換のように情報のヘッダによってあて先が規定されるヘッダ駆動方式に分類される．ヘッダ駆動方式は更にストレージ形とストレージレス形に分類できる．ストレージ形では設定されるべき回線もしくは相手ノードがふさがりであるときを配慮して交換機能の中に蓄積能力をもつ．ストレージレス形はこれをもたない．

ローカルエリアネットワーク（LAN）の技術は，代表的にはヘッダ駆動ストレージレス（header driven storageless: HDSL）形で構成されている．これにおいてはいずれも，ある発信元から送られたデータユニットが，他の全ノードに対して放送され，あて先ノードが自分あてのものを選択する方式をとっている．このためシステム全体の中で，同時には一つのデータユニットしか取り扱えないことになる．しかし逆に，ネットワークは同時には一つのデータユニットしか取り扱わないから，あて先ノードがふさがりのため通信が失敗することはなく，従ってストレージレス形でよいことになる．

これに対して予設定形の交換方式では同時に多数の接続を設定することによって同時に多数の伝送を行うことができる．

従来のLANでは，どのようなトポロジーをとってもHDSL形の通信を行う限り，一般的に同時に複数の伝送を行うことができない．例えばスター形であって，中央のノードにすべての回線が集中している場合でも，HDSL形の制御を行えば，同時には一つのデータユニットしか伝送できないのである．

こうしたHDSL形の制約があるため，LANにおいては，エンドノードが増加してトラヒックが増大した場合，これに対応するためには高速化によらなければならない．光LAN等の技術によれば高速化はさして困難でなくても，エンドノードの入出力回路の高速化はコスト増を伴う．しかもエンドノードそのものの瞬時伝送速度が高くなっても，エンドノードの数が増加すれば，個々のエンドノードで取り扱える情報量は本質的には増加しない．

このようなことを考えれば，エンドノードの数を増やすために高速化しなければならないのは従来のHDSL形のLANの本質的な欠点であり，できれば複数の伝送を複数の伝送路を用いて同時に実現する方式が望ましい．

† 東京大学工学部電気工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan

近年 HDSL 形の LAN においても，CSMA/CD，すなわち ISO 802.3 規格互換性のあるスター形 LAN が注目されている．また AT & T ベル研では，センタノードにアービタを設けて衝突を防止したスター形 LAN が提案されている．これらはスター形であるから伝送路そのものは基本的にはエンドノードの数だけ用意されている．しかし，こうした従来のスター形 LAN では先に述べたような放送形の通信を行うために，同時には一つのデータユニットしか伝送できない．

複数の伝送を同時に実現して LAN の容量を増大する方法としては，複数の伝送路を用意して，各エンドノードがそれらにアクセスできるようにする方式がある．この方式では，各エンドノードに複数のアクセスユニットが必要になること，同一エンドノードに同時に複数のデータ着信があったときの対応が複雑になることなどの問題点が残る．また，複数の CSMA/CD 形 LAN を高速の LAN で接続する方法も工夫されている．しかし，この方法でもトラヒックのローカリティを仮定しなければ，個々の LAN の容量でシステム全体の容量が制約される点は解決されない．

本論文においては，スター形構成で，センタノードに必要な多重伝送の数に対応する伝送路を設置し，同時多重伝送が可能なスター形 LAN を実現する方法を提案し，マルチスター LAN と名づける．

従来の LAN では，先に述べたような容量の制約から，個々の伝送路の速度をエンドノード当りの伝送速度と比べて格段に高く設定する傾向にあり，このために種々の実現上の困難やコストの増大を生じがちである．マルチスター LAN の方式では，個々の伝送路の伝送速度をあまり高くしなくても，必要な容量を得ることができる．本論文では，このような性質を利用して，伝送速度 1 Mbit/s 程度でシステムを実現することを想定して評価する．この場合，同軸ケーブルではなく電話用に布設されている回線でマルチスター LAN を実現することも可能になり，コストが大幅に低下する．しかし，マルチスター LAN の原理は，要求に応じてより高速な（例えば 10 Mbit/s 程度）LAN に適用することも可能である．

以下，**2.** でマルチスター LAN の構成と動作の概要の説明を，**3.** で動作の詳細と使用するフレーム構造の説明を，**4.** ではバックオフアルゴリズムの検討，**5.** では負荷・遅延時間特性の理論式導出，**6.** ではその計算結果とシミュレーションによる結果の検討を行う．

2. マルチスター LAN の構成

図 1 にマルチスター LAN の全体構成の概念図を示す[2]．センタノード（CN）には複数のスイッチモジュール（SWM）を設置し，これに対して，各エンドノード（EN）は 1 対の受信線，送信線を経由して接続される．各 EN は CN との間でスター形に接続され，各 SWM に分散されて収容される．

マルチスター LAN では同時に複数の伝送が行われるから，放送形の通信を行うことはできない．このため，図 2 に示すようにデータユニットのヘッダを解読してあて先を選択する機能を設ける（SWM）．図 3 に示すように，SWM の構成は原理的には単純であり，各データユニットのヘッダのあて先アドレスを記憶する

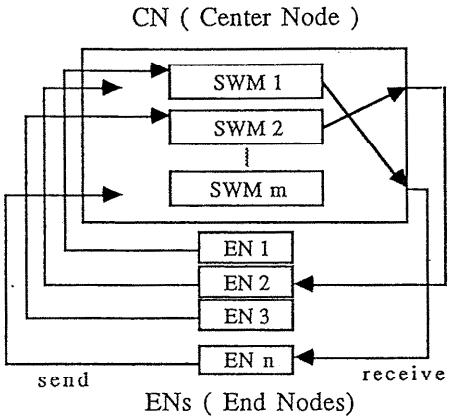


図 1 マルチスター LAN の構成
Fig. 1 Structure of MULTISTAR LAN.

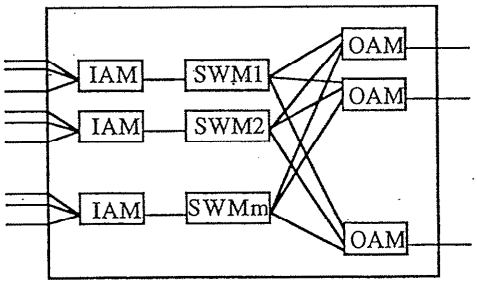


図 2 データパス（DTP）の構成
Fig. 2 Structure of Data Path (DTP).

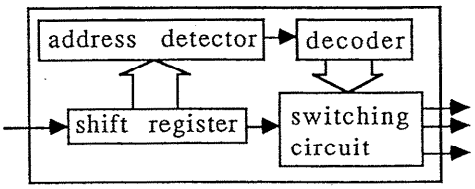


図 3 SWM の構成
Fig. 3 Structure of SWM.

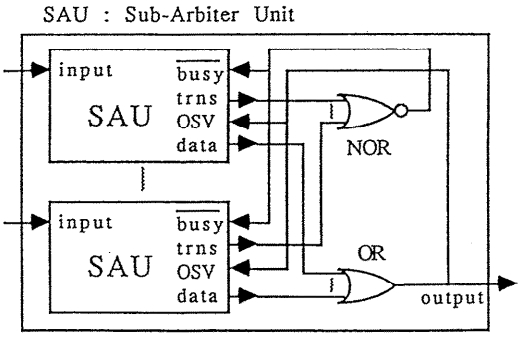


図 4　アービタモジュールの構成
Fig. 4　Structure of arbiter module.

シフトレジスタと，あて先アドレスに応じてルーチングを行うスイッチ回路からなる．CN には，EN からのデータユニット同士の競合を調停し，そのうち一つが SWM を利用できるようにするための入力アービタを設ける．入力アービタは全 SWM に共通に設置して全 EN を収容することもできるが，ここでは各 SWM ごとに設置し，EN を分割収容するものとし，これを入力アービタモジュール（input arbiter module : IAM）と呼ぶ[5]．一方，SWM の出力があて先で衝突することを避けるため，出力側には各 EN に対する受信線ごとに出力アービタモジュール（output arbiter module : OAM）を設ける．これらをまとめてデータパス（DTP）と呼ぶ．図 4 にアービタモジュールの構成を示す．サブアービタユニット（SAU）と，相互結合のための OR, NOR 回路からなる．必要となる SAU の数はアービタモジュールの入力数に等しい．SAU における信号の意味とアービタモジュールの動作を説明する．

入力：データユニット入力

出力：アービタ出力．競合に勝ったデータユニットが出力される．

data : SAU 出力．

$\overline{\text{busy}}$：データユニットがアービタを通過中であれば“0”，そうでなければ“1”．SAU は $\overline{\text{busy}}=0$ のときはデータユニットを通過させず，$\overline{\text{busy}}=1$ のときは通過させる．

trns：その SAU がデータユニットを出力中であることを示す．

OSV : 出力監視用入力．出力が自 SAU 出力と異なっていれば，SAU は自分の出力を中止する．

SAU にデータユニットが到着すると，アービタモジュールが空きであれば，すなわち $\overline{\text{busy}}=1$ であれば，trns=1 にして，data からデータユニットを送出する．trns=1 になることによって，全 SAU の $\overline{\text{busy}}=0$ になり，アービタモジュールが閉そくであることが示される．もし，$\overline{\text{busy}}=0$ の状態から新たにデータユニットが到着しても，SAU は data から何も出力せず，破棄することになる．

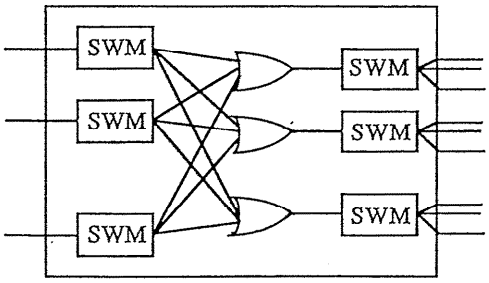


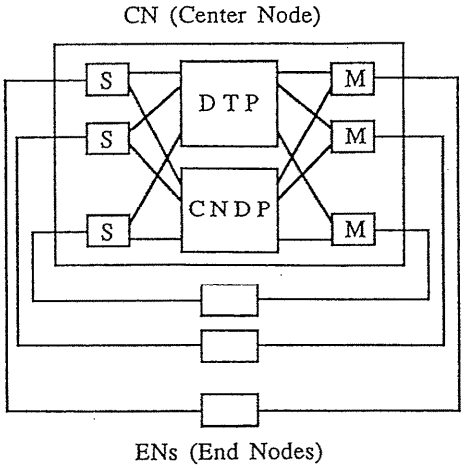
図 5　CND パス（CNDP）の構成
Fig. 5　Structure of CND Path (CNDP).

trns が 0 から 1 になるまでの遅延時間内に到着したデータユニットは，SAU で破棄されずに通過してしまい，複数のデータユニットが OR 回路に到着することになる．ここで，各 SAU は，OSV によって OR 回路の出力を監視し，自分の出力と比較して，異なる場合は自分の出力を直ちに中止して，入力中のデータユニットを破棄すると共に，trns を 0 に戻す．

このように，アービタユニットは競合する複数の入力データユニットのうち一つを選択して出力する機能をもつことになる．

EN が送出したデータユニットは，IAM, SWM, OAM を経由してあて先 EN に伝送される．いずれかのアービタで衝突が生じたときにはそのデータユニットは捨てられる．従来の放送形の LAN では，発信側の EN は自分が送出したものとは異なるデータユニットが放送されたことで失敗を検知することができたが，マルチスター LAN ではこの方式をとることができない．そこで，EN は正しいデータユニットを受信したとき，直ちに発信側の EN に対して物理層の肯定応答として，CND（collision not detected）と呼ぶ確認信号を返送するものとする．この目的で，CN にはデータパスとは独立に CND パス（CNDP）を設ける．CND パスの構成を図 5 に示す[1]．CND パスにはデータパスの OAM に対応して EN ごとに SWM があり，データパスの IAM に対応してまた別の SWM がある．

データパス，CND パスを含む CN の構成を図 6 に示す[1]．EN との間の伝送路はデータ，CND で共有される．これはデータを送信中はほかからのデータは受信しないことを前提とし，従って CND を伝送する必要があるときには送信線は空いていると考えているた



S : Data/CND separater
M : Data/CND mixer

図 6 全体の構成
Fig. 6 Structure of whole system.

めである．このため CN にはデータ/CND 分離，混合回路が設けられる．

マルチスター LAN では，トラヒックに応じて CN 内の設備を増設することにより，容量を増加させることができる．例えば，IAM に用いるアービタは図 4 に示すように各 EN に対応して OR 回路に接続される SAU を追加することによって増設できる．SWM を増設するときにも，対応する OAM について同様の増設が可能である．その他のモジュールについても，並列的に回路を付加することによって増設でき，複雑さの増加はたかだか比例的であるからシステムとしての増設性は高いと言うことができる．しかし，実装上，増設用のスロットをどのように残しておくかなどの配慮は十分にしておく必要がある．

3. マルチスター LAN の動作

図 6 のように構成したマルチスター LAN において，CN および EN は次のように動作する．

まず EN の送信について考えてみる．EN は送信に対して CND を受信する必要があるから，送信中は他の EN からの着信があってもこれを受け付けてはならない．このために EN は送信を開始するとその EN に対応する OAM をビジー（ふさがり状態）にする．

OAM をビジーにするタイミングとして，送信したデータユニットが IAM に到着したときとする方法と，送信したデータユニットが IAM を通過して SWM を使用し始めたときとする方法が考えられる．後者の場合には，IAM を通過できなかった（すなわち送信に失敗した）データユニットの送信元 EN にも他の EN からのデータユニットの着信を許すので，明らかに高負荷に対応できる．これに伴い，各 EN は送信中でもデータユニットが到着すればこれを受け付けるようにする．前者を基本形[1]，後者を改良形と呼ぶ．

基本形では第 1 の EN から送信されたデータユニットが IAM を通過できなくてもその EN の受信側 OAM をビジーにするため，もし IAM を通過できなかった原因となった相手の送信要求が，第 1 の EN をあて先としているときには，IAM を通過した送信要求もあて先には到着できないことになり，トラヒックの集中に対してネットワークの特性が劣化する原因となり得る．改良形はこのような劣化を回避している．

また，図 2 に示すように，マルチスター LAN では EN が IAM に分割収容されているが，一部の IAM に負荷が集中すると特性の劣化が生じる．これに対する解決策として，一部の IAM に負荷が集中しないように収容を工夫する必要がある．IAM を分散して構成するのではなく，すべての入力を同時に扱えるようなアービタを構成することが望まれるが，これは将来の研究にゆだねたい．

以下に改良形の場合の EN と CN の動作を示す．

[EN の動作]

(送信)

① 受信中でないことを確認する．受信中ならば，CND 以外の送信は行わない．データユニット受信中に送信を行うと，それに対応する CND と受信中のデータユニットとの衝突が生ずる場合があるためである．

② データユニットを 1 回だけ送出する．

③ 一定時間 τ だけ待つ．この τ は
(EN－EN 間の往復時間)＋(CND 生成に必要な時間)
より長くなければならない．

④ τ 時間内に CND が返送されたかどうか監視する．この間にデータユニットが到着したら，自分の送信は失敗したと判断し，直ちに送信を中断して受信を開始する．

⑤ CND が返ってきたら，その発信元およびあて先のアドレスを調べ，自分の送出したデータユニットのあて先，および自アドレスと一致しているかどうかを確認する．もし，一致していれば送信成功であり，送信終了後アイドル状態に戻る．

⑥ タイムアウトになる前に（τ 時間内に）自アドレスと一致する CND が返送されない場合には，まだ送

信中であれば送信を中止し，バックオフアルゴリズムに基づく時間だけ待った後，①に戻って送信を繰り返す．バックオフ待ち中には，受信が可能である．

(受信)

① データユニットが到着すると，それがデータ用であり，あて先アドレスが自アドレスと一致するかどうか調べる．

② 一致していれば，そのデータユニットを入力バッファに入れ，CNDを返送する．

③ 一致していなければ，そのデータユニットは捨てる．

[CNの動作]

① データユニットが到着すると，対応するIAMがアイドルであればビジーにし，そのパスを使用する．そのとき，発信元のENが別のパスを通ってきたデータユニットを受信しないように発信元ENに対応するOAMをビジーにする．IAMが既にビジーであれば，そのデータユニットは破棄されCNの状態に変化はない．

② データユニットが通過したことを検知すると，発信元IAMおよびあて先OAMをアイドルにし，そのパスを解放する．

③ データユニットに対応したCNDがCNを通過したことを検知すると，発信元OAMをアイドルにし，ほかからのデータユニットを受信できるようにする．

図7(a)，(b)にはそれぞれデータ用とCND用のデータユニットのフォーマットを図示している[2]．このフォーマットは基本的には標準的なLANのフォーマットに従っているが，データ/CND分離回路の動作を迅速化するために，データとCNDを区別する情報を含むコントロールフィールド（CTR）をスタートデリミタ（SD）の直後に付してある．

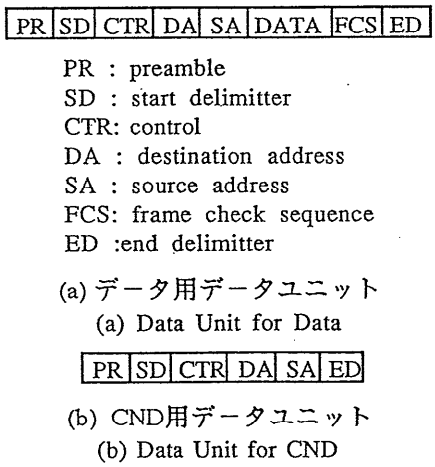


図7 マルチスターLANのデータユニット構造
Fig. 7 Structure of data units for MULTISTAR LAN.

4. バックオフアルゴリズム

マルチスターLANにおいては，ENが送信を開始してから期待される時間以内にCNDが返送されないときには，再送を行う．これは送信したデータユニットがIAMもしくはOAMのいずれかで破棄されたために，再送が必要となるからである．

このようなバックオフについてのアルゴリズムはCSMA/CDの場合によく研究されている．問題は複数のENで同時に送信に失敗すると，それらのENは再送信を行うため，再び衝突を生ずる可能性があることである．従来のCSMA/CDの場合には衝突を生じたデータユニットはすべて破棄され再送されるから，再衝突の可能性が高いことが問題となり，それを回避するための工夫が重要であった．このためバイナリエクスポネンシャルのように，再衝突を生ずると，再送までの待ち時間を次第に長くしてゆくなどの工夫がなされている．

これに対してマルチスターLANでは，複数のデータユニットが衝突しても一つのデータユニットの送信には成功すること，キャリアセンス機構が設けられていないため衝突が生ずる原因は既に送信中のデータユニットが存在するためである場合が多いことなどの理由から，バックオフアルゴリズムの評価もCSMA/CDの場合などとは異なる．

マルチスターLANの場合には，送信に失敗したときには，競合して勝った相手は送信に成功しているわけであるから，バックオフの回数を多くしないためには，その相手データユニットの送信が終了すると予想される時点で再送できれば有利である．

1データユニットの送信に必要な時間を T とすると，送信に失敗したとき，通信中であった相手の残余通信時間は $0 \sim T$ の間で一様に分布するはずである．従って，最初の送信から T 時間以降に再送すれば，先の送信の失敗の原因となった相手の通信は完了しているはずである．

これより短い再送間隔を選択したとしても，相手の通信は終了していることもあるから，再送に成功する可能性もある．もし失敗すれば，再びバックオフを行うから，この場合はバックオフの回数が増加し，CNに到着するデータユニットの数も多くなる．しかし，各

SWMではこれに妨害されることなくデータユニットの処理を行えるから，全体のスループットが低下することはない．

一方，バックオフ間隔が短ければ，SWMが利用可能になったとき，直ちに送信を行うことができるから，バックオフ待ち時間を短縮できる．また，何回目のバックオフであっても，その成功率は最初の送信要求からの時間で決まる．

このように考えれば，バックオフの回数を制約しなくてもよい場合には，バックオフ間隔を短くした方が平均の待ち時間を短くする上で有利になる．これについては後述のシミュレーションによって更に検討する．

5. 負荷・遅延特性評価式の導出

簡単のためにEN数を無限大とする．また，SWM数を m（有限）とし，CNにおけるデータユニット通過率（送信に成功したもの）を λ（有限）とする．また，

T_c：データユニットのCN通過時間

T_{rt}：伝搬遅延時間

t：データユニットを受信してからCNDを返送するまでの時間（CND生成時間）

T_0：衝突のない場合の，全伝送時間

τ：送信失敗判断時間

K：CNでの遅延ビット数

v：伝送速度

L_p：データユニット長

とすると，

$$T_c=(K+L_p)/v$$

$$T_0=T_{rt}+T_c$$

$$\tau \geqq 2T_{rt}+2K/v+t$$

となる．

今，先の λ の定義から，SWMを通過すれば必ず送信に成功するものとすれば，各SWMの利用率 U は，

$$U=\lambda T_c/m \qquad (1)$$

となる．現実のシステムの場合にはSWMを通過したデータユニットがあて先のOAMで衝突を生ずることがあるが，これはあて先ENがデータユニットを受信中もしくはデータユニットを送信中（改良形では送信中であり，そのデータユニットがSWMを占有している場合）にのみ起こることであり，今全システムのデータユニット通過率が有限でEN数が無限大を仮定しているから，こうした状態になる確率は0と考えることができる．それ故，各SWMの利用率は一般に式(1)で与えられると考えてよい．

平均値 T_b のランダムバックオフアルゴリズムの場合，平均伝送時間 t_f は，1回の送信につき，それが成功する確率を P_s として，

$$t_f=T_0P_s+(T_0+T_b)(1-P_s)P_s+(T_0+2T_b)(1-P_s)^2P_s+\cdots \qquad (2)$$

となる．

一方，U はあるENがデータユニットを送信しようとしたとき，IAMで阻止される確率すなわちその送信が失敗する確率であるから，

$$P_s=1-U \qquad (3)$$

である．

式(2)，(3)から，

$$t_f=T_0+T_bU/(1-U)$$

となり，これを

$$T=L_p/v$$

で正規化すると，正規化遅延時間 D は，

$$D=\{T_0+T_bU/(1-U)\}v/L_p \qquad (4)$$

となる．

現実のシステムにおいては，EN数は有限であるから，SWMを通過したデータユニットがOAMで阻止される可能性がある．このため，特に高負荷時には理論値よりも遅延時間が増加すると考えられる．

6. 負荷・遅延特性の計算結果とシミュレーションによる検証

本章では，前章で導出したモデルによる負荷・遅延特性と，シミュレーションによる結果を示す．シミュレーションに際しては，EN数は有限としなければならないので，OAMでのデータユニットの衝突が生ずる．このために前章で導出した理論式とシミュレーションでは評価条件が異なっており，遅延時間の理論値はシミュレーションによる結果より小さめになると思われる．

理論値の計算，シミュレーションに当って，各定数は以下のように定めてある．

伝送速度　v: 1 Mbit/s

CN-EN間距離：無視

EN数　n: 48

データユニット長　L_p: 1,000ビット

CN遅延　K/v: 90 μs

CND生成時間　t: 100 μs

送信失敗判断時間　τ: 350 μs

バックオフ待ち時間の平均値　T_b: 1,150 μs

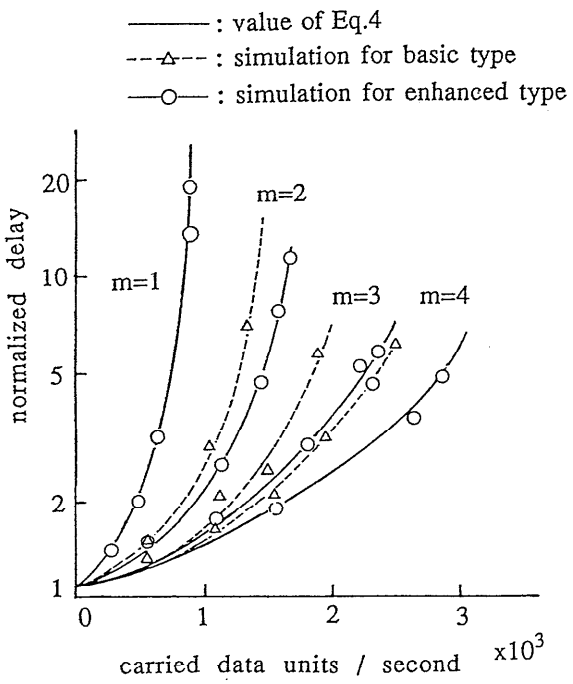


図 8　マルチスター LAN の負荷遅延特性
Fig. 8　Traffic-delay characteristics of MULTISTAR LAN.

(送信失敗判断時間 τ を含む)
式(4)から求めた理論値と，シミュレーションの結果とを併せて図 8 に示した．

この結果わかることは次のとおりである．

① 低負荷時には基本形と改良形の性能に大きな差異はない．

② 負荷の増加に伴い正規化遅延時間が増加するが，その度合は改良形の方が基本形より小さい．これは，EN 数が無限大のときには無視できた OAM でのデータユニットの衝突が，特に高負荷時には増加して，改良形の利点が強調されたからである．但し，それも m が 2 以上のときであるが，これは $m=1$ のときには IAM (SWM) を通過したデータユニットは OAM で阻止されることがないからである ($m=1$ のときには基本形と改良形の特性が一致する)．

③ SWM を 2，3…と増やすことによって，限界となるデータユニット通過率をほぼ 2 倍，3 倍…と増やすことができる．

その動作原理から，$m=1$ のマルチスター LAN は，低負荷時には CSMA/CD と同様の特性を示すと考えられる．これを検証するために，CSMA/CD，更にトークンリング方式についての負荷・遅延特性をマルチスター LAN の特性と共に図 9 に示した．図 9 で横軸は限界スループットで正規化されたスループットである．なお，CSMA/CD とトークンリング方式の特性は文献(4)によっている．これから，次のことがわかる．

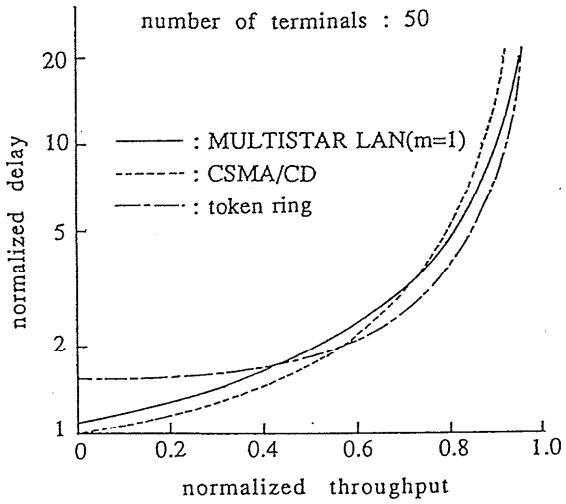


図 9　各種 LAN の負荷遅延特性の比較
Fig. 9　Comparison of traffic-delay characteristics of three LAN systems.

④ $m=1$ のマルチスター LAN を CSMA/CD と比較すれば，低負荷時には，CN の通過遅延のためにマルチスター LAN の方が遅延時間が大きいが，高負荷時には，衝突したデータユニットのうちの一つは送信に成功するという性質からマルチスター LAN の方が遅延時間が小さくなることがわかる．

⑤ $m=1$ のマルチスター LAN をトークンリング方式と比較すれば，低負荷時には，トークン処理のためにトークンリング方式の方が遅延時間が大きくなるが，高負荷時にはマルチスター LAN の方が大きくなることがわかる．

また，先に述べたようにマルチスター LAN では，平均バックオフ待ち時間を短くしても，CSMA/CD に見られるような性能の悪化は生じないと思われる．しかし，平均待ち時間を著しく短くすることは，端末における処理時間を考えると現実的ではない．そこで，シミュレーションによって，固定的な送信失敗判断時間 350 μs に一様分布の待ち時間を加えたものをバックオフ待ち時間として，バックオフ待ち時間と正規化遅延時間の関係を求めた．図 10 には正規化スループットをパラメータとしてランダムバックオフの場合の平均バックオフ待ち時間と正規化遅延時間の関係を改良形のマルチスター LAN について示している．

この結果わかることは次のとおりである．

⑥ バックオフ待ち時間を短くすると，衝突を起こしたデータユニットが同じデータユニットに再衝突す

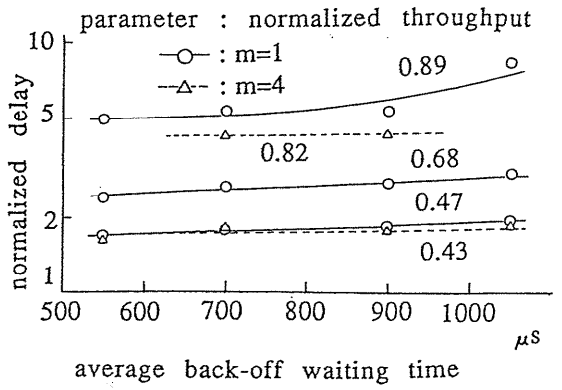


図 10 改良形マルチスターLANのバックオフ待ち時間と正規化遅延時間の関係

Fig. 10 Back-off waiting time-normalized delay characteristics of enhanced type MULTISTAR LAN.

る確率は大きくなるが，各データユニットはアービタにより順次処理されるので結果として平均の遅延時間を短くできる．

7. むすび

本論文では，通常のスター形LANを改良することで同時に複数の通信を実現できるようになったマルチスターLANを提案した．このマルチスターLANは，

・SWMの増設によりほぼそれに比例して容量を増やすことができるため，さまざまなトラヒックに対して柔軟に対応でき，負荷の増加に対して拡張性に富む．

・コンテンション方式のため低負荷時の遅延は少なくできる上，アービタ方式のため高負荷時の特性の悪化も少ない．

・スター形形状のため保守性に優れている．

・通常の放送形のスター形LANと異なり，1対1通信であるため，プライバシー保護が容易である．

・CNとENの間に中継，分岐が不要なため，伝送媒体として光ファイバを導入しやすい．

などの特徴を有する一方で，

・高負荷時に遅延時間がばらつく

・同報通信が実現しにくい

などの点は，今後解決されるべき問題点として残っている．

謝辞 本研究を進めるにあたり，御尽力頂いた河合智明氏，宮尾健氏をはじめとする研究室各位に深く感謝致します．

文献

(1) T. Saito, H. Inose and T. Kawai: "A star type local area network —Multistar— Having simultaneous multiple transmission capability", ISS-87 (Merch 1989).

(2) 河合智明，川本広昭，斉藤忠夫，猪瀬 博："同時多重接続機能を持つスター型LAN"，信学技報，IN85-98 (1985-12).

(3) 河合智明，増田俊也，斉藤忠夫，猪瀬 博："同時接続が可能なスター型LAN"，第31回情処全大，2Q-8 (昭60-09).

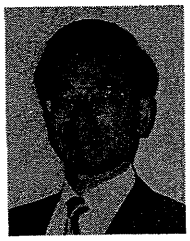
(4) 所真理雄，田丸喜一郎："ローカルコンピュータネットワーク(2)"，信学誌，**65**，6，pp. 624-628 (昭57-06).

(5) A. Albanese: "Star network with collision-avoidance circuits", AT&T, Bell Sys. Tech. Jounal, **62**, 3, pp. 631-638 (March 1983).

(平成元年2月13日受付，8月18日再受付)

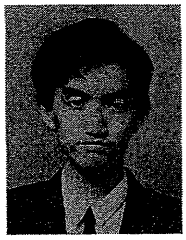
斉藤 忠夫

昭38東大・工・電子卒．昭43同大大学院博士課程了．工博，東京大学工学部講師．東京大学助教授，東京大学教育用計算機センター主任，カリフォルニア工科大学研究員などを経て現在東京大学教授．ディジタル通信方式，コンピュータネットワーク，ディジタルシステムの信頼性などの研究に従事．著書に「電子回路入門」(昭晃堂)，「ディジタル回路」(コロナ社) などがある．昭42，昭62本会論文賞，昭48本会業績賞など受賞．電気学会，情報処理学会，IEEE各会員．



相田 仁

昭55東大・工・電気卒．昭60同大大学院博士課程了．工博．東京大学工学部助手．昭61同講師．現在，日本学術振興会海外特別研究員としてSRIインターナショナル(米国)に滞在中．高並列推論計算機，分散オペレーティングシステム，知識処理などの研究に従事．著書に「第5世代コンピュータ入門」(共著) がある．情報処理学会，人工知能学会，ソフトウェア科学会，IEEE，ACM各会員．



井上 一郎

昭63東大・工・電子卒．現在，同大大学院修士課程在学中．コンピュータネットワーク，LANの研究に従事．情報処理学会，IEEE各会員．

７．　分散処理・コンピュータネットワーク

○　ネットワーク・オペレーティング・システムにおけるプロセス管理方式
○　分散形データフロー計算機の構成と実験的評価
○　ＦＩＦＯ付きブロードキャストメモリを用いた共有メモリの実現とその評価

論　文

ネットワーク・オペレーティング・システムにおけるプロセス管理方式

正員 中島　周[†*]　正員 斉藤 忠夫[†]　正員 猪瀬　博[†]

A Process Management Scheme for a Network Operating System

Amane NAKAJIMA[†*], Tadao SAITO[†] *and* Hiroshi INOSE[†], *Members*

あらまし　ローカル・エリア・ネットワークを通信サブネットとするコンピュータ・ネットワークでは，ネットワークの意識なしに資源を利用できるようにするネットワーク透明性の実現が重要な課題である．ネットワーク透明性を実現する対象にはファイル，周辺装置などがあり，これを実現したシステムの例もすでにある．本研究ではプロセスが実行されるプロセッサに関してのネットワーク透明性をプロセスに関するネットワーク透明性と呼ぶが，この実現問題は解決されていない．本論文ではプロセスに関するネットワーク透明性の要求条件，実現方法および試作システムによる実現性の確認について述べている．ここに示す方式では，集中管理ノードと複数のプロセッサの間でプロセス管理機能の機能分散を行い，プロセッサ間でプロセス管理機能の負荷分散を行う．本論文ではこのような環境下におけるプロセス管理の詳細を述べ，さらにプロセスのプロセッサへの割り付け方法についても検討を行っている．

1 まえがき

高性能のマイクロコンピュータの発達により，マイクロコンピュータをＬＡＮのような通信手段で相互に接続し，高性能のコンピュータに匹敵する能力を持たせるネットワークの構成方式が多様に検討されている．特に，多数のマイクロコンピュータの複合体を形成し，その利用者が，あたかも単一コンピュータをＴＳＳで利用できるようなシステムは，実現性も高く，また多様な用途が期待できる．

このようなシステムでは，まずネットワーク内のどこからシステムにアクセスしていても，同一の方法でネットワークの存在を意識することなくすべての資源を利用できるという性質が望ましい．これをネットワーク透明性と呼んでいる．

ネットワークは一般に多数のプロセッサを含む．ネットワーク内のいずれかで自分のプロセスを実行し，ネットワーク内のファイルを使用する．ファイルがネットワーク内のどこにあってもそれを意識しなくてもよいことをファイルに関するネットワーク透明性と呼ぶ．一方プロセスがどのプロセッサで実行されていてもその場所を意識しなくてもよいことをプロセスに関するネットワーク透明性と呼ぶ．

このようなネットワーク透明性を実現するには，ネットワーク全体をネットワークOS（Operating System）によって管理する必要がある．ネットワークＯＳには，従来から種々のものがあり，各ノードにどのような機能があるかで分類される．各ノードが機能的に完結した構成を持ち，１ノードだけでも動作可能なネットワークＯＳを独立型ネットワークＯＳと呼んでいる．これに対してノード間で機能分散を行い，複数ノードでなければ動作できないネットワークＯＳを一体型ネットワークＯＳと呼ぶ．この２種は２つの典型であるが，この中間にも種々の考え方がある．独立型ネットワークＯＳの制御動作は，ホストーホスト接続形式のネットワークの各ノードで実行される．一体型ネットワークＯＳでは，ノードの種類ごとに分散カーネルが実行され，ネットワーク全体で論理的に１つのＯＳとなる．一体型ネットワークＯＳで資源の管理体系が１つであれば，ネットワーク全体をより詳細により有機的に管理することが可能となる．

現在までに発表されたネットワークＯＳには，

† 東京大学工学部電気工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan
* 現在，日本アイ・ビー・エム株式会社

AEGIS[1], LOCUS[2], V-kernel[3], AMOEBA[4]などがあるが，AMOEBA以外は独立型ネットワークOS，あるいは独立型ネットワークOSに近いネットワークOSである．これらのネットワークOSでは，ファイルに関するネットワーク透明性はサポートしているが，プロセスに関するネットワーク透明性は実現していない．AMOEBAではプロセスの生成，移動を行うプロセス・サーバを提案しているが，具体的な実現方法は示されていない．プロセスに関してのネットワーク透明性を実現するには，各プロセッサ・ノードをネットワークの資源として管理し，その集合体を全体として1つの論理的なプロセッサとして見せるようにすることが必要である．

本論文では，標準OSのひとつであるベル研究所のUNIXと同等の環境を多数のプロセッサが存在する環境で実現することを通してプロセスに関するネットワーク透明性に関して検討し，それを実現した結果を述べる．

2. 一体型ネットワークOSの構成

著者らは1983年より一体型ネットワークOS，COSMOS[5]の研究，開発を行ってきた．

COSMOSはLOCUSと同じく，UNIXを基本とし，ユーザにUNIX上位互換性を保証する．COSMOSでは，ネットワーク透明・ネットワーク均質な資源管理を実現するために，ネットワーク・マネージャ(Network Manager)という集中管理ノードを導入する．ネットワーク・マネージャはCOSMOSシステム内に1つだけ存在し，各ノードに分散されたネットワークOSの統合処理を行う．COSMOSでは異種ノード間で機能分散を行い，同種ノード間で負荷分散を行う．図1にCOSMOSシステムのネットワーク構成を示す．COSMOSではネットワーク内の諸資源，つまりファイル，デバイス，プロセス，プロセッサをまとめてオブジェクトと呼ぶ．COSMOSにおけるオブジェクトとは，外部からの操作が定義されたモジュールのことである．COSMOS内のすべての資源をオブジェクトとみなし，すべての処理をオブジェクトに対する要求の送出とその要求の実行として実現する．このオブジェクト概念の導入により，ネットワーク内資源の統合的管理が可能となる．オブジェクトをCOSMOS内で一意に指定するために，オブジェクトにはネットワークで一意なオブジェクト識別子を与える．オブジェクト識別子の管理はネットワーク・マネージャが行う．

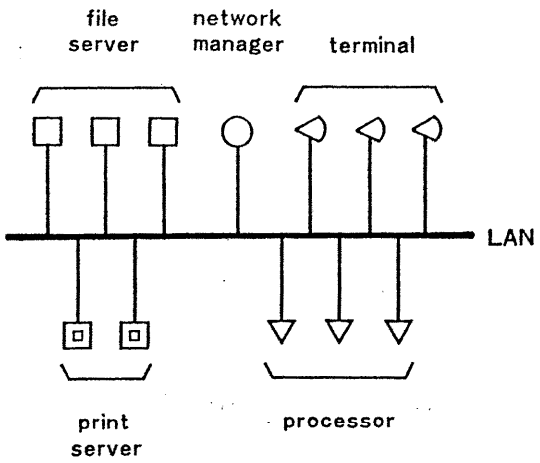
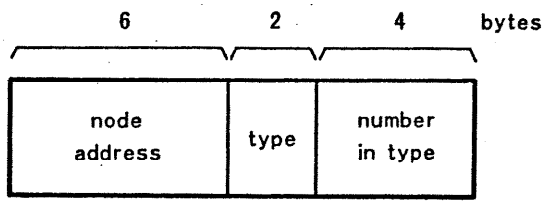


図1 COSMOSシステムのネットワーク構成
Fig. 1 Network configuration of COSMOS system.




```
struct oid
 {
  unsigned short o-nadr[3];
  unsigned short o-type;
  unsigned int   o-num;
 };
```

図2 オブジェクト識別子の構成
Fig. 2 Structure of object identifier.

図2にオブジェクト識別子の構成を示す．オブジェクト識別子は12バイトであり，ノード・アドレス6バイト，オブジェクト・タイプ2バイト，タイプ別番号4バイトから成る．ノード・アドレスとは，そのときそのオブジェクトが存在するノードのイーサネット・アドレスである．オブジェクト・タイプは現在はファイル，デバイス，プロセス，プロセッサの4種類であるが，将来の拡張のために2バイトとなっている．タイプ別番号は各オブジェクト・タイプで自由に使用できる．

COSMOSでは，ファイル，デバイスの論理名は基本的にUNIXと同じ方法で管理する．UNIXではツリー状のディレクトリにより，ファイル，デバイスの名前を管理している．COSMOSでは1つのファイル・サーバ上のファイルの集合をサブツリーとし，この

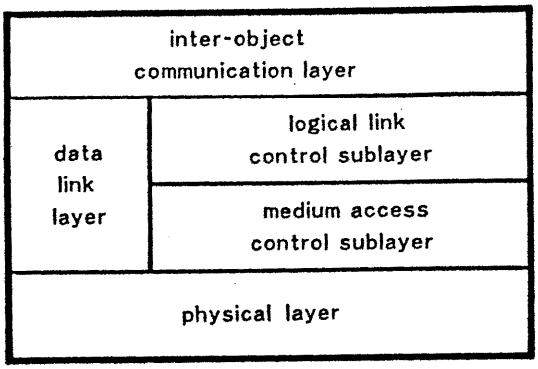


図3　COSMOSの通信プロトコル階層
Fig. 3 Communication protocol layer of COSMOS.

サブツリーをまとめ，ネットワーク全体のファイル，デバイスで1つのツリーを構成する．UNIXでは，*inode*というファイル制御ブロックを用いてファイルを管理する．COSMOSのファイル型オブジェクトではタイプ別番号に*inode*の番号を指定し，これによりネットワーク内のファイルを一意に識別する．

図3にCOSMOSの通信プロトコルの階層を示す．オブジェクト間通信層はオブジェクト間の通信をサポートする．

3. プロセスの分散管理

3.1 プロセスの実行環境

COSMOSではUNIXと同様のインタフェースでユーザがプロセスを実行できる環境を形成する．しかし，図4に示すように一体型ネットワークOS，たとえばCOSMOSとシングル・プロセッサ用のOS，たとえばUNIX，ではOSの管理する範囲が異なる．図4ではOSの管理する範囲を点線で示した．実線で囲まれた領域は1つのノードの処理の範囲を表現している．シングル・プロセッサ用のOSではプロセッサは1つしかないので，OSはその1つのプロセッサとそこで実行されるプロセスを管理する（図4(a)）．一体型ネットワークOSではネットワーク全体でOSは論理的に1つであり，この1つでネットワーク全体の資源を管理する（図4(b)）．そのためプロセスについては，ネットワーク内の全プロセッサで実行される全ユーザ・プロセスを管理する．

一体型ネットワークOSの環境ではネットワーク内の全プロセスを複数のプロセッサに割り付けて実行する．また，プロセッサではマルチプログラミングにより複数のユーザ・プロセスを実行するので，ユーザ・プロセスの実行は2段階に多重化されているといえる（図5）．

3.2 プロセス管理機能の分担

一体型ネットワークOSではユーザ・プロセスは各プロセッサで実行される．各プロセッサでは，そのための管理を実行しなければならない．一方ネットワークOSで，プロセスに関するネットワーク透明性を実現するために，ネットワーク・マネージャにおいて全体としての管理を行う必要がある．プロセス管理をネットワーク・マネージャと各プロセッサでどのように分担管理するかが，ネットワークOSにおけるプロセス管理の主要な問題である．一般的に，ネットワーク・マネージャでネットワーク内の全プロセスを対象とした統合的なプロセス管理を行い，各プロセッサでは

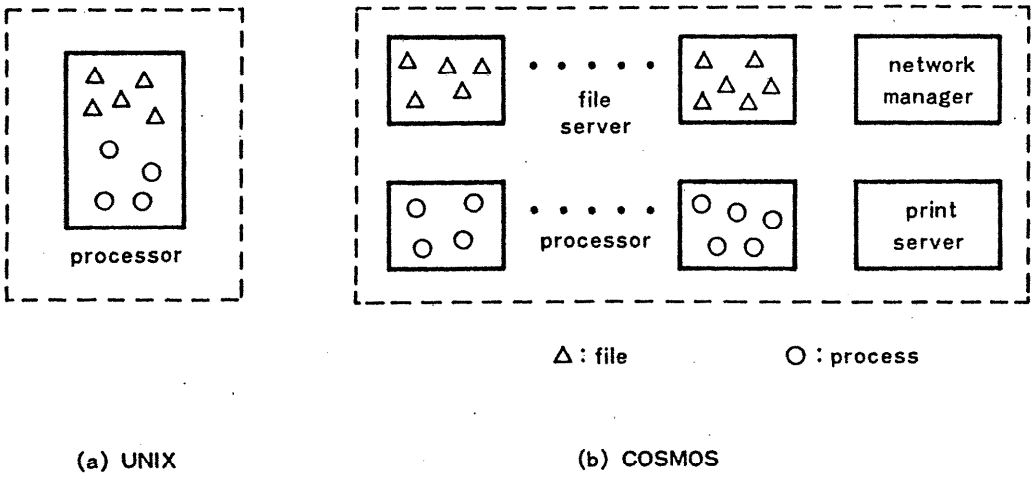


図4　OSとプロセス，プロセッサ
Fig. 4 OS, process, and processor.

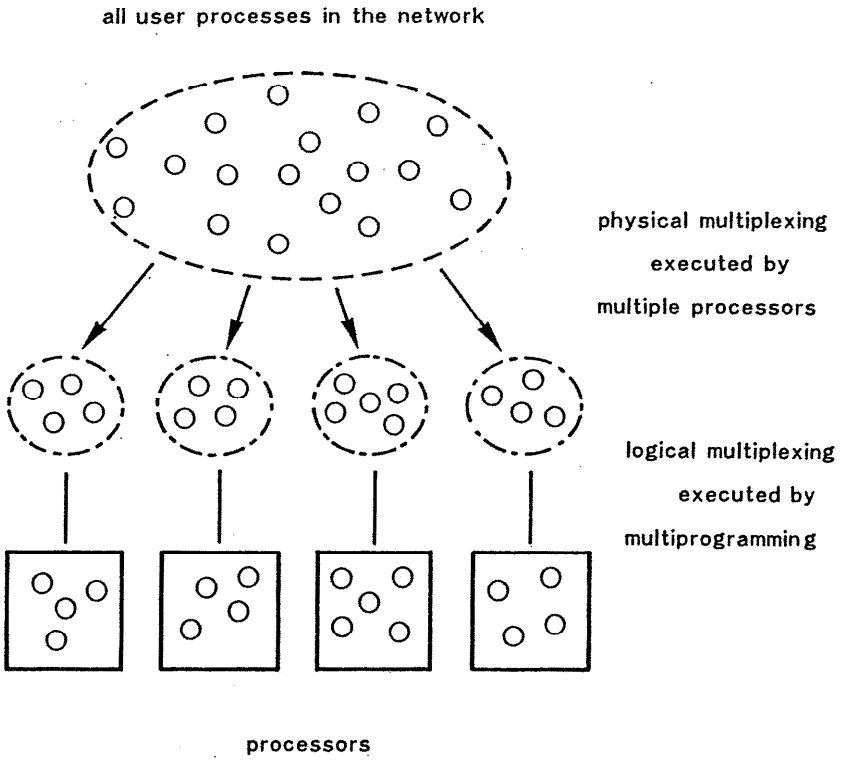


図5　ユーザ・プロセス実行の2段階の多重化
Fig. 5 Multiplex execution of user processes in two phase.

各プロセッサに割り付けられたプロセスを対象とした分散されたプロセス管理を行うようにするのがよいと考えられる．

プロセスの管理機能にはプロセスの名前管理，プロセスのプロセッサへの割り付け，プロセス・スケジューリング，プロセス・ディスパッチならびにプロセスに関するメモリ管理がある．

メモリ管理およびプロセス・ディスパッチはプロセッサのハードウェア，メモリ容量などに依存するので，各プロセッサごとに実行する必要がある．

プロセス・スケジューリングは純粋に論理的な処理であるので，プロセッサでもネットワーク・マネージャでも実行可能である．しかし，プロセス・スケジューリングをネットワーク・マネージャで行うことにすると，割り込みなどを介してプロセッサの種々の状況をネットワーク・マネージャで把握することが必要となるためにその通信オーバヘッドが大きくなる．さらに，プロセス・スケジューリングの対象がネットワーク内の全プロセスとなり，大規模なシステム構成では膨大な数になるので，プロセス・スケジューリングは頻繁に行う必要があり，ネットワーク・マネージャにとっては重い負荷となってしまう．そこで，プロセス・スケジューリングは，各プロセッサが自分に割り付けられたプロセスを対象として実行するのが適切である．

ネットワーク・マネージャはプロセッサとは異なり，集中管理ノードという性格からシステム内の全プロセスを対象とした管理を行う．システム全体を参照しなければならない処理，プロセッサで行うプロセス管理の上位に位置づけられる処理はネットワーク・マネージャで行うことになる．

まず，プロセスもオブジェクトの一種なので，プロセスの識別子としてのオブジェクト識別子の管理，存在場所の管理を行わなければならない．このネットワーク・ワイドなプロセスの識別子の統合的な管理により，どのプロセッサで実行されているプロセスも名前（プロセス番号）に関してすべて同等に扱うことができるようになり，プロセスとそのプロセスの実行ノードはユーザには意識されなくなる．つまり，プロセスの名前に関してネットワークが透明になる．

次に，ネットワーク・マネージャは各プロセッサの負荷バランスをとり，プロセッサという資源が有効に使われ，プロセスの実行が迅速に行われ，システム全

表1　COSMOSにおけるプロセス管理機能の分担

	ネットワーク・マネージャ	プロセッサ
対　　象	ネットワーク内の全プロセスを管理	各々のプロセッサに割り付けられたプロセス群を管理
管理のレベル	高次の管理	低次の管理
プロセスの実行	プロセスの実行は行わない	プロセスの実行を行う
管理の内容	・プロセスの識別子と存在場所の管理（プロセスに関するネームサーバ） ・プロセス生成時の実行プロセッサの決定 ・全プロセッサの状態の定期的なモニタリング ・プロセス間の同期	・メモリ管理 ・プロセス・ディスパッチ ・プロセス・スケジューリング

体のスループットを向上させ，レスポンス・タイムを安定化するためにプロセスのプロセッサへの割り付けを行う．これにより，ユーザはプロセス実行時に実行プロセッサを指定しなくてよくなり，プロセスの実行場所についてもネットワーク透明性を実現することが可能となる．

以上のプロセス管理機能の分担を表1にまとめた．

4. ネットワーク・マネージャでのプロセス管理

4.1 プロセス番号の管理

UNIXではプロセスに番号をつけ，これをプロセスの識別子として用いている．プロセス番号は4バイトの大きさで，カーネル内だけでなく，ユーザ・プログラムで，あるいはコマンドの引数として使われる．COSMOSはUNIXと互換であるので，プロセス番号によってネットワーク内のプロセスを一意に指定できなければならない．また，COSMOSではプロセスもオブジェクトの一種なので，前述のオブジェクト識別子も与えられる．COSMOSではこの2つのプロセスに関する識別子をまとめなければならない．そこで，プロセス型オブジェクトでは，タイプ別番号としてプロセス番号を採用する．これにより，ネットワーク内のプロセスで一意なプロセス番号と，ネットワーク内のオブジェクトで一意なオブジェクト識別子がうまく関係づけられる．プロセスを指定するときは，コマンドの引数，ユーザ・プログラム，カーネルではプロセス番号を用い，オブジェクト間の通信ではオブジェクト識別子を使うことになる．

```
struct nmproc
 {
  struct oid      *p-poid;
    /* pointer to oid struct */
  unsigned int     p-pid;
    /* process identifier */
  unsigned int     p-cpu;
    /* process cpu time */
  unsigned int     p-ltime;
    /* process life time */
 };
```

図6　*nmproc* ストラクチャ
Fig. 6　*nmproc* structure.

ネットワークの全プロセスに一意な番号を与え，それらを集中管理すれば，プロセスの名前に関してネットワークは透明になる．そこでネットワーク・マネージャに図6の *nmproc* ストラクチャを導入し，ネットワーク内の全プロセスに1対1で *nmproc* ストラクチャを割り当てることにする．*p_poid* は図2に示した *oid* ストラクチャへのポインタであり，*p_pid* はネットワークで一意なプロセス番号である．*p_cpu* はそのプロセスが現在までに使用したCPU時間，*p_ltime* はプロセスが生成されてからの経過時間を表す．これにより，プロセス番号からのオブジェクト識別子の検索といったプロセスのネーム・サービスや，ネットワーク・マネージャによるプロセスのスケジューリングが可能となる．

4.2 プロセスのプロセッサへの割り付け

一体型ネットワークOSでは，ユーザにはプロセッサの集合があたかも1つのプロセッサであるかのように見え，プロセスの実行プロセッサを指定することは行わないので，ネットワークOSがそれぞれのユーザ・プロセスをどのプロセッサで実行するかを決定しなければならない．本論文ではこれをプロセスのプロセ

```
struct prcsr
 {
  unsigned short r-freem;
    /* size of free core memory */
  char           r-pnum;
    /* number of processes */
    /* in this processor */
  char           r-ldav;
    /* load average */
    /* of this processor */
 };
```

図7 *prcsr* ストラクチャ
Fig. 7 *prcsr* structure.

ッサへの割り付けと呼ぶ．このようにプロセスをOSによってプロセッサへ割り付けることができるようになっているために，一体型ネットワークOSでは独立型ネットワークOSに比べてより高度なプロセス，プロセッサの管理が可能となる．

プロセスのプロセッサへの割り付け法には2つの方法がある．1つの方法は，プロセスの生成時にプロセスを実行するプロセッサを決める静的な割り付けである．もう1つの方法では，生成時に割り付けを行うが，さらに実行中のプロセスを対象として実行プロセッサの再決定を行い，これを動的な割り付けと呼ぶ．動的な割り付けでは実行中のプロセスの移動，つまりプロセス・マイグレーション(Process Migration)を行うことになる．

ネットワーク・マネージャは各プロセッサの負荷を測定することによってプロセスの割り付け先を決定する．各プロセッサの負荷はロード・アベレージ(Load Average)で表す．ロード・アベレージとは，実行可状態にあるプロセスの数の時間平均である．各プロセッサはこの値をネットワーク・マネージャに数秒ごとに定期的に報告する．ネットワーク・マネージャはプロセッサの管理のために各プロセッサに対応して図7の*prcsr*ストラクチャを持ち，*r_ldav*に各プロセッサのロード・アベレージを記入する．ネットワーク・マネージャはプロセッサのメモリの空き容量を示す*r_freem*が割り付けるプロセスのプロセス・イメージよりも大きく，かつ*r_ldav*が一番小さいプロセッサに新しいプロセスを割り付ける．

プロセス・イメージをファイル・サーバ，あるいは今まで実行していたプロセッサからこれから実行するプロセッサに送信することをプロセス・イメージの転送という．静的割り付けとは異なり，動的割り付けを実行する場合は，実行しない場合と比べて，プロセス・イメージの転送分だけはオーバヘッドとなる．静的割り付けを行っていれば，プロセスの生成時にプロセッサの負荷バランスを実現しようとする力が働く．通常のプロセスのCPU時間は1秒以下と短いので，プロセスの生成は頻繁に行われる．このような状況ではプロセスのプロセッサへの動的割り付けは必要なく，静的割り付けだけで充分である．

4.3 COSMOSにおけるプロセスのプロセッサへの割り付け

静的割り付けは，プロセスの生成時に実行される．UNIXでは，プロセスの生成は*fork*と*exec*という2つのシステム・コールによって行われる．*fork*では親プロセスと同じ内容の子プロセスを作る．*exec*では自分のプロセス・イメージを別のプロセス・イメージに変更する．通常は親プロセスが*fork*で新しいプロセスを作り，作られたプロセスが自ら*exec*を行って希望のイメージのプロセスとなる．

静的割り付けを*fork*時に行うか，*exec*時に行うかにより処理内容が異なる．*fork*時に割り付けを行うと，その後の*exec*時にもう1度プロセス・イメージの転送が必要となり，計2回のプロセス・イメージ転送となってしまう．普通は*fork*の後に*exec*を行うので，*exec*時に割り付けを行えばプロセス・イメージの転送は1回でよい．よって静的割り付けは*exec*時に行うのが有利である．

また，**4.2**で述べたように動的割り付けはオーバヘッドが大きいので，COSMOSではプロセスのプロセッサへの割り付けは*exec*時の静的割り付けだけを行っている．

5. プロセッサでのプロセス管理

5.1 プロセス・イメージ

COSMOSではUNIXとのユーザ・プロセス・レベルの互換性を保つために，UNIXと同じプロセス・イメージを採用する．このプロセス・イメージを図8に示す．プロセス・イメージは，テクスト・エリア，パー・プロセス・データ・エリア，データ・エリア，スタック・エリアに*proc*ストラクチャを加えたものである．*proc*ストラクチャについては，**5.3**で述べる．

UNIXでは実体が1つであるプロセスが，プロセッサのユーザ/カーネルというモードにより，論理的にはユーザ・プロセス，カーネル・プロセスという2つ

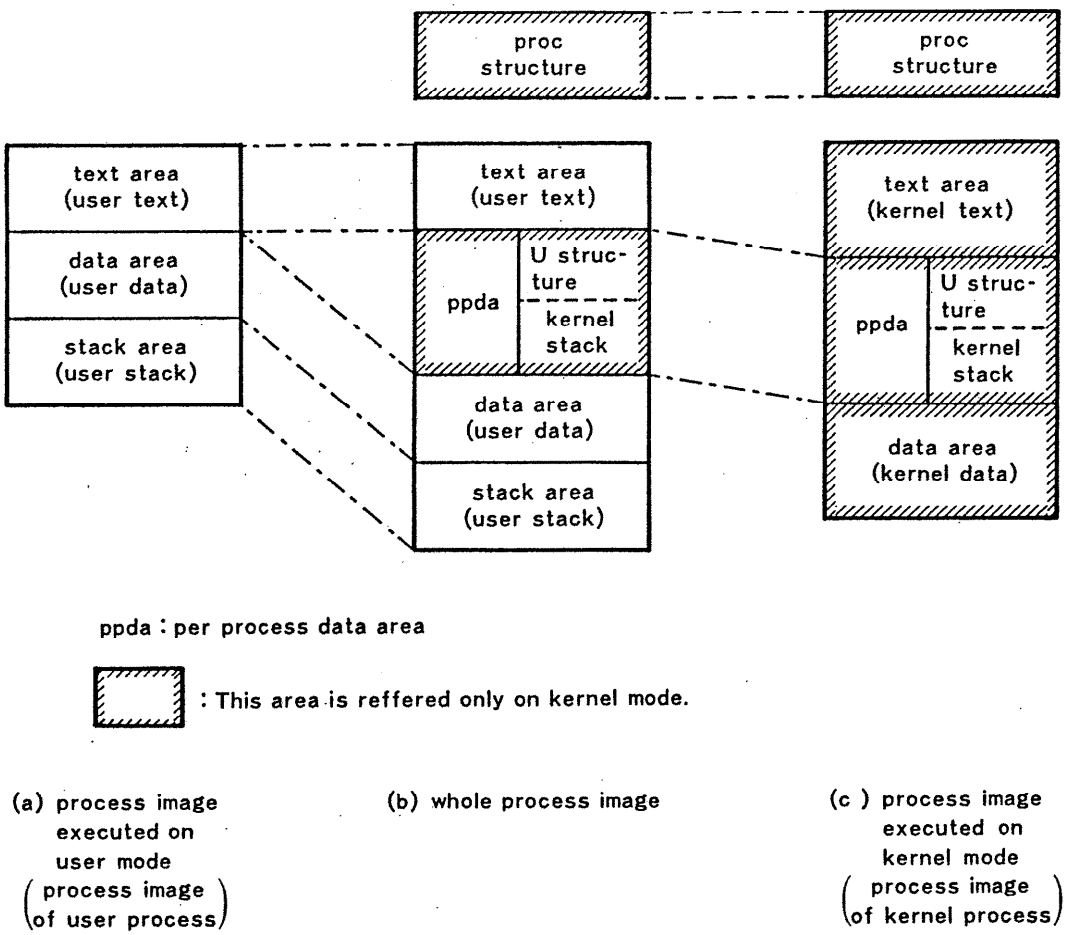


図8　UNIXのプロセス・イメージ
Fig. 8　Process image of UNIX.

のプロセスになる．プロセス・イメージのエリアのうち，テクスト・エリア，データ・エリア，スタック・エリアはユーザ・モード時のユーザ・プロセスの実体となる．カーネル・プロセスのテクスト・エリア，データ・エリアはカーネル・プロセス全体によって共有される．各プロセスのパー・プロセス・データ・エリアだけがプロセスごとの固有のエリアとしてカーネル・プロセスの実体の一部となる．

5.2　プロセス・ディスパッチ

メモリ管理については**6.3.1**で述べる．

プロセス・ディスパッチはカーネル・モード用（カーネル・プロセス）と，ユーザ・モード用（ユーザ・プロセス）に2段階に実行される．最初に，カーネル・アドレス空間のパー・プロセス・データ・エリアを切り替える．これによってカーネル・プロセスが切り替わることになる．次に，ユーザ・アドレス空間のテクスト・エリア，データ・エリア，スタック・エリアを切り替える．その後ユーザ・モードに移行することによってユーザ・プロセスのプロセス・ディスパッチが完了する．これらの処理はMMUのレジスタの内容を変更することによって行われる．

5.3　プロセス・スケジューリング

プロセス・スケジューリングは，プロセス・イメージの一部である*proc*ストラクチャ内の情報に基づいて行われる．*proc*ストラクチャは各プロセスと1対1に対応し，常時メイン・メモリに存在する．*proc*ストラクチャにはプロセスが生成されてからの経過時間，使用したCPU時間，プロセス・イメージの大きさなどが記述されている．ユーザ・プロセスの優先度は，定数にプロセスの優先度を示す値と使用したCPU時間を加えて計算される．この値の小さいプロセスほど優先度が高い．以上はUNIXと同じ方式を採用した．

ただし，UNIXと異なる点が2つある．1つは，プロセスが他のオブジェクトにメッセージを送った後に

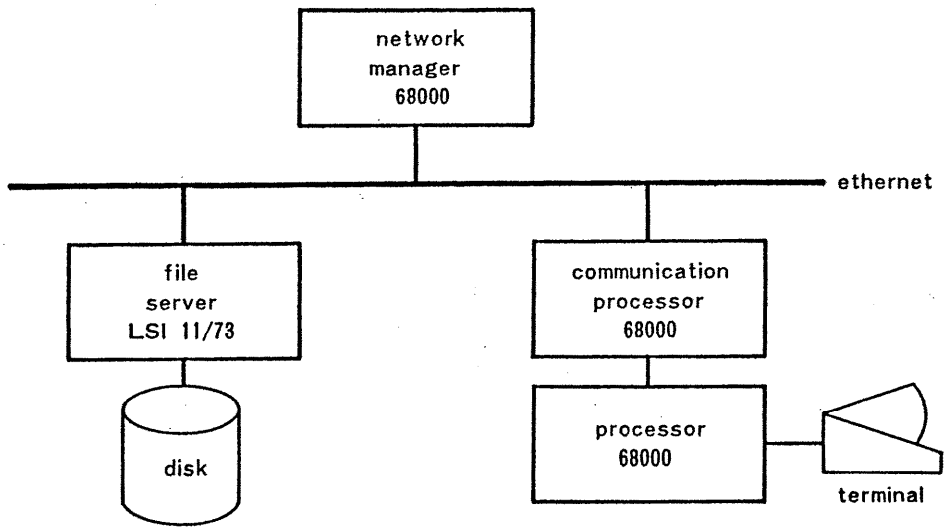


図9　COSMOS試作システムのハードウェア構成
Fig. 9 Hardware configuration of experimental COSMOS system.

入出力待ち状態になり，送信したメッセージに対応する応答を受け取ると実行可状態にもどることである．もう1つは，ネットワーク・マネージャからのメッセージにより，プロセッサ内のプロセスの状態（入出力待ち，実行可）を変更できることである．

6. 試作システムと評価

6.1 試作システムの概要

図9に著者らが製作したCOSMOS試作システムのハードウェア構成を示す．試作システムでは，ネットワーク・マネージャ，ファイル・サーバを各1ノードと，プロセッサとターミナルが一体化したプロセッサ・ノードを持ち，これらが10Mbpsのイーサネットで接続されている．CPUは主として68000であり，これを8MHzのクロックで使用する．

通信プロトコルの論理リンク制御サブ層ではAcknowledged Connectionless Services方式を採用した．試作システムは通信プロセッサにおける処理時間を入れて，68000通信プロセッサ間で746Kbpsの通信速度を得た．オブジェクト間通信層では，Unacknowledged Connectionless Services方式とConnection Oriented Services方式を採用し，伝送するメッセージに従って選択的に使用するようになっている．

COSMOSのオブジェクト間通信では，イーサネットの1パケット（1518バイト）で行えない通信はファイルの多量読み出し/書き込みだけである．1パケットで完了する通信をコネクションを設定して行うことはオーバヘッドが大きいので，これらの通信はUnacknowledged Connectionless Services方式で行い，多量の情報を一度に伝送するファイルの多量読み出し/書き込みをConnection Oriented Services方式で行う．

6.2 ネットワーク・マネージャ・カーネル

ネットワーク・マネージャで実行されるCOSMOSの分散カーネルをネットワーク・マネージャ・カーネルと呼ぶ．ネットワーク・マネージャ・カーネルは，オブジェクト識別子の管理，4.で述べた高次のプロセス管理，サブツリーを構成するファイル・サーバ上のファイルの統合的管理，ユーザの登録/抹消，login/logoutなどのユーザ管理，各ノードの時刻が同一であることを保証する時刻管理，ノードの健全性を調べるノード管理などを行う．

現在，試作システム上のネットワーク・マネージャ・カーネルはプロセス管理の部分が完成し，他の部分は実装中，あるいは未実装である．プロセス管理のプログラムはC言語で記述し，約200行である．

6.3 プロセッサ・カーネル

6.3.1 メモリ管理

プロセッサで実行されるネットワークOSの分散カーネルをプロセッサ・カーネルと呼ぶ．プロセスへのメモリの割り付けは，プロセス・イメージのエリアごとに行う．そのため，プロセッサ・カーネルは，メモリの空き領域を，先頭アドレスと大きさの対で表現してメモリの使用状況を把握し，メモリの割り付けを行う．

COSMOS試作システムのプロセッサではCPUとして68000，MMU(Memory Management Unit)と

表2　プロセス・ディスパッチに関するルーチンの大きさ

単位：ステップ

ルーチン名	機　　能	UNIX	COSMOS
save〔アセンブラ〕	カーネル・プロセスのディスパッチ	13	22
resume〔アセンブラ〕	カーネル・プロセスのディスパッチ	15	20
sureg〔C言語〕	ユーザ・プロセスのディスパッチ	28	12
estabur〔C言語〕	ユーザ・プロセスのディスパッチ	106	0

として68451を2個持っている．68451は1個につき32のディスクリプタを持つ．ディスクリプタによってメモリ・セグメントを管理し，これらを用いて論理アドレスと物理アドレスの変換を行う．68451MMUではメモリ・セグメントは2のべき乗の大きさにしか設定できないので，プロセス・イメージのうち可変長のエリアには複数のディスクリプタを割り付け，メモリの使用効率が低化するのを防いだ．

このメモリ管理プログラムはC言語で記述した．大きさは約900行である．PDP-11用のUNIXバージョン7では，この部分はアセンブラ約100行，C言語約200行である．プログラムの大きさは3倍になったが，UNIXバージョン7では64Kバイトに制限されていたプロセス・イメージが，COSMOSでは論理的には2Mバイトとほとんど無制限になり，また，**6.3.2**で述べるようにプロセス・ディスパッチが迅速に実行できるようになった．

6.3.2　プロセス・ディスパッチ

COSMOSでもUNIXと同じアルゴリズムでプロセス・ディスパッチを行う．対応するルーチンには同じ名前をつけた．しかし，メモリ管理機構の違いのため異なる点がある．それは*estabur*というルーチンの有無である．UNIXバージョン7では，ユーザ・モード用のMMUの16個のレジスタに設定する値のプロトタイプを*estabur*で各プロセスのパー・プロセス・データ・エリアに格納する．これに対してCOSMOSでは，ユーザ・モード用のMMUのレジスタは2つだけであり，しかもこれらのレジスタに設定する値はメモリ割り付け時にパー・プロセス・データ・エリアに格納されているので，*estabur*は行わなくてよいのである．

UNIXとCOSMOSのプロセス・ディスパッチに関するルーチンの大きさを表2に示す．カーネル・プロセスのプロセス・ディスパッチは退避するレジスタの数が多いため，COSMOSの方がやや処理量が多くなっている．一方，ユーザ・プロセスのプロセス・ディスパッチはCOSMOSの方が断然少ない．これは，メモリ管理でプロセス・ディスパッチに関する処理を一部行っているためである．

7. むすび

本論文では一体型ネットワークOSのプロセス管理方式について述べた．プロセスを実行するプロセッサと，システムを集中管理するネットワーク・マネージャでプロセス管理機能の分担を行い，ネットワーク・マネージャでネットワーク内の全プロセスを対象として，プロセスの番号管理，プロセスのプロセッサへの割り付けを行えば，プロセスに関するネットワーク透明性が実現できることを述べた．

複数のプロセッサ・ノードを持つシステムで本方式を実現し，実際のTSS環境で定量的な評価を行うことは今後の課題である．

謝辞　本研究に関し，御討論頂いた東京大学文献情報センターの橋爪宏達氏に感謝致します．

文　　献

(1) アポロ社，日本ラッド(株)，丸紅(株)：“アポロDOMAIN情報処理アーキテクチュア”(1982).
(2) B.Walker, G.Popek, R.English, C.Kline and G.Thiel：“The LOCUS Distributed Operating System”, ACM Operating Systems Review, 17, 5, pp.49-70(Oct.1983).
(3) D.R.Cheriton：“Local Networking and Internetworking in the V-System”, ACM SIGCOMM Computer Communication Review, 18, 4, pp.12-20(Oct.1983).
(4) A.S.Tanenbaum and S.J.Mullender：“An Overview of the AMOEBA Distributed Operating System”, ACM Operating Systems Review, 15, 3, pp.51-64(July 1981).
(5) 橋爪宏達：“ネットワーク資源共有のための統合型ネットワーク・オペレーティング・システムの研究”，東京大学大学院工学系研究科電気工学専門課程学位論文(1984-01).

(昭和60年6月25日受付，10月21日再受付)

中島　周

昭58 東大・工・電子卒．昭60 同大学院修士課程了．同年日本アイ・ビー・エム(株)入社，サイエンス・インスティチュート勤務．在学中ローカル・エリア・ネットワークを用いたシステムに関する研究に従事．情報処理学会会員．

斉藤　忠夫

昭38 東大・工・電子卒．昭43 同大学院博士課程了．工博．同年東大講師，昭44 同助教授．昭53～56 同教育用計算センター主任．昭49～50カリフォルニア工大研究員．PCM統合通信，コンピュータ・ネットワーク等の研究に従事．42年度論文賞，48年度業績賞，55年度電気学会論文賞等受賞．

猪瀬　博

昭23 東大・二工・電気卒．昭30 工博．昭36 東大教授，現在に至る．ディジタル通信などの研究に従事．功績賞および業績賞，日本学士院賞，マルコニ国際学術賞など各受賞．IEEEフェロー，米国科学アカデミーおよびアメリカ哲学協会外国人会員．

分散形データフロー計算機の構成と実験的評価†

大山敬三†† グェン・ニュット††
ビノッド・プラサッド・スレスタ††
斉藤忠夫†† 猪瀬 博††

データフロー計算機は，従来のフォンノイマン型計算機にかわり，高度の並列性が要求される情報処理の面で，重要な役割を果たすものと期待されている．データフロー計算機の実現にあたっては，命令セルからの命令を処理エレメントへ転送する裁定回路網と，処理エレメントの出力を命令セルへ転送する分配回路網の，効率のよい構成を見いだす必要がある．本論文では，この点に着目して，分散形データフロー計算機の新しい構成を提案し，縮小モデルの試作を通じて，その実現性を実験的に確認している．提案しているシステムにおいては，裁定回路網と命令セルを処理エレメントに分散配置している．また分配回路網については，可能な諸構成を検討した結果，高速共通バスを採用することとし，処理能力 10 MFLOPS をもつ分散型データフロー計算機のシステム設計を行った．さらにこのシステムの実現性を確認する目的で，縮小実験モデルを試作し，評価した結果，提案システムの実現性の確認ができた．

1. まえがき

電子計算機の歴史において，いわゆるフォンノイマン型の構造は現在まで一貫した思想である．しかし，フォンノイマン型の構造に対しては，ハードウェアの点からも種々の制約が指摘されている．

最近の情報技術は，マイクロエレクトロニクス技術にもとづく大量生産によって急速な価格低減を続けている複雑な論理機械を中心として発展してきている．しかし，いわゆるフォンノイマン型の構造で大型計算機を構成しようとするとマイクロエレクトロニクス技術の量産性の特質を十分享受することはできないといえよう．

たんなる計算能力からすれば，安価なマイクロプロセッサを多数使用するほうが，単一の大型計算機よりも安いコストで計算を実現できることは一般的に理解されている．もちろん，小型計算機を多数集めただけでは，大規模な処理を一体として実行する大型機の能力は実現できない．このため，小型のプロセッサを多数有機的に結合して，大型機に匹敵する能力を実現する技術開発が求められている．

本論文では，小型のプロセッサを用いた演算処理装置を多数使用し，これらを有機的に結合することによって，高度の並列処理を行うことにより数値計算を主対象として 10 MFLOPS 程度の処理能力をもつ，データフロー計算機を実現することを目的として，その構成法を具体的に明らかにするとともに，縮小モデルを試作して，実現性の検証，性能の評価，問題点の抽出を行っている．

2. システムの基本的構成

データフロー計算機の構成方法は基本的に次のような視点から分類できよう．

（1） 集中形構成と分散形構成
（2） アクタの処理レベル
（3） 一様構成と階層構成

従来，データフロー計算機においては，データフロー原則の自然な実現形態として，おもに図1のような集中形構成が研究されてきたが[1]，性能に対する考察はあまり行われていなかったといってよい．筆者らは高性能データフロー計算機を実現するという立場でまず集中形構成の検討を行った結果，以下のようなボトルネックが明らかになった．

（1） 裁定回路網における処理装置の選択の高速化が困難である．
（2） 命令メモリへのアクセスの競合が生じ，性能低下の原因となる．
（3） 命令メモリの管理処理が大規模化する．

すなわち，裁定回路網においては，データフローグラフで記述されたプログラムの処理単位であるアクタが実行可能となるたびに動的に処理装置を選択せねば

† System Design of a Distributed Data Flow Computer and Its Experimental Evaluation by KEIZO OYAMA, NHUT NGUYEN, VINOD PRASAD SHRESTHA, TADAO SAITO and HIROSHI INOSE (Department of Electrical Engineering, Faculty of Engineering, University of Tokyo).
†† 東京大学工学部電気工学科

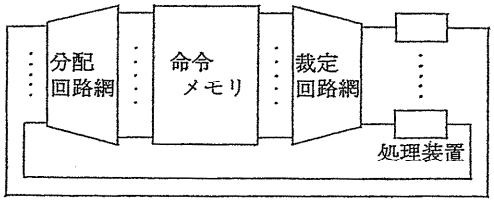


図 1 集中形データフロー計算機の構成
Fig. 1 Structure of a concentrated data flow computer.

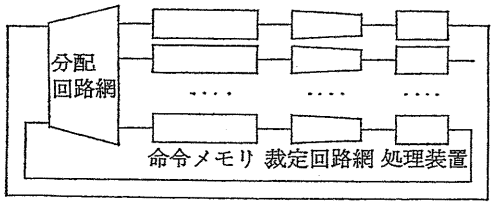


図 2 分散形データフロー計算機の構成
Fig. 2 Structure of a distributed data flow computer.

ならず，非常に多数の処理装置の管理を高速で行う必要があり，高性能化の障害となる．

命令メモリには，分配回路網と裁定回路網からのアクセスが頻繁にあり，競合が起こる．また，大規模なデータフロープログラムを実行するには，命令メモリの容量も非常に大きくなり，したがって，この管理が大規模化し，同種の回路を多数用いて実現することがむずかしくなる．

以上の問題を解決するため，筆者らは先に**図 2** に示すような分散形構成を提案した[2),3)]．すなわち，命令メモリを分割し，そこに記憶されているアクタを実行する処理装置に位置的な制限を加えることにより，裁定回路網の負荷を軽減し，同時に命令メモリの規模を小さく抑える構成である．

データフロー計算機は，対象とするデータフロープログラムの処理単位であるアクタに含まれる処理内容に応じて低，中，高レベルに分類できる．低レベルとは機械語のレベルをいい，高レベルとは通常の手続きやプロセスのレベルをいう．本システムは，マイクロプロセッサを用いるという立場から，四則演算にして数個から数十個程度の処理を単位とする中レベルデータフロー計算機とする．

本システムでは，分散形構成を採用したことによりアクタの割付問題が生ずる．筆者らは先に分配回路網に環状回路網を用いた予備実験システムを構成し[2),3)]，割付問題を検討したが，この場合には隣接した処理装置間では通信を短時間で実行できるが，遠い処理装置間の通信に長時間を要するため，データフロープログラムのアクタの割付が複雑になることが明らかになった．

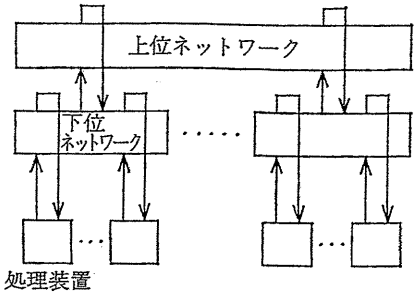


図 3 階層構成のデータフロー計算機
Fig. 3 Data flow computer having hierarchical architecture.

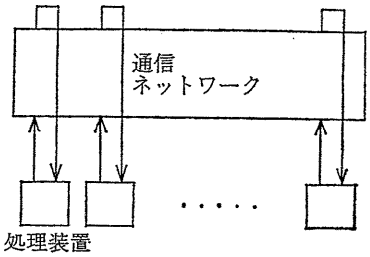


図 4 一様構成のデータフロー計算機
Fig. 4 Data flow computer having uniform architecture.

図 3 に示すような階層性のあるアーキテクチャでもプログラムの局所性を抽出する必要があるため，割付は複雑となる[3),4)]．これに対して**図 4** に示す一様構成をとれば，割付問題は単純な負荷配分の問題に帰着でき，階層構成のデータフロー計算機の場合より，はるかに容易となる．

以上の検討をまとめ，分散形データフロー計算機の利点と欠点を示せば，次のようになる．

利点としては，

（1） 裁定回路網，命令メモリの構成が容易である，

（2） システムの拡張性が得やすい，

（3） ハードウェアの融通性・柔軟性が高い，

（4） 同種の部品，回路を多数用いることで構成が可能である，

（5） 静的割付を行うため，入出力装置等の機能の追加が，全体的構造の変更なしに行える，

等が挙げられる．

一方，欠点としては，

（1） 割付により実行速度が大幅に変化し，割付が重要課題となる，

（2） 割付の動的変更が困難である，

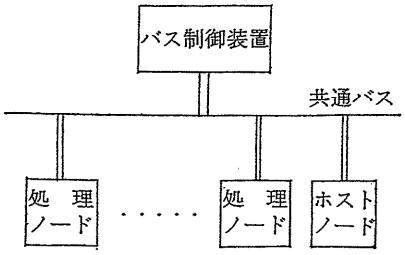


図 5 分散形データフロー計算機の全体構成
Fig. 5 Overall system configuration of the distributed data flow computer.

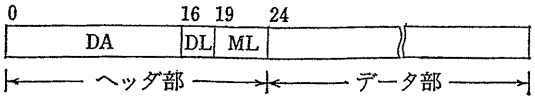


図 6 メッセージ構成
Fig. 6 Message format.

(3) 資源の利用効率が下がる，

等が挙げられる．

マイクロプロセッサのコストは急速に低減しているので，個々の利用効率がある程度低くとも，これらを多数使用した分散形構成をとることにより，システム全体としての性能価格比の向上が期待できる．

以上をまとめると，本システムの全体的構成は**図5**に示すようになる．分配回路網には，データ転送容量，遅延，一様性，拡張性，実現の容易さ等を考慮し，高速の共通バスを使用することとする．

本システムでは 50 kFOPS 程度の処理能力をもつ演算装置を含む処理ノードを 256 台程度用いて 10 MFLOPS の性能を実現することを想定している．そのほかに，データフロープログラムの作成，コンパイル，ロードおよびデータの入出力のために，ホスト計算機として1台のホストノードを置く．

3. システムの詳細な構成

各部のより詳細な構成は次のようになる．

3.1 共通バス構成

分配回路網には，同期式の共通バスを用いる．システムの規模によっては，他の結合回路網を用いたほうが適切な場合もあり，構成に応じて決める必要がある．ノード数 256，全体の性能が 10 MFLOPS 程度では共通バスが最も実現性がある．

バス幅は，データバス 128 本，アドレスバス 8 本とし，その他に必要な制御線数本を張る．

メッセージは**図6**に示すように，ヘッダ部とデータ部とから成る．ヘッダ部はメッセージに関する制御情報であり，DA は目的アクタ番号で 16 ビット，DL は目的リンク番号で 3 ビット，ML はメッセージ長で 5 ビットとする．データ部はオペランドデータを運ぶ部分で，13〜409 バイトである．

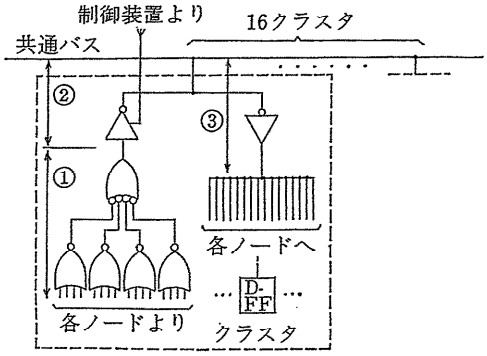


図 7 クラスタ化による共通バス接続
Fig. 7 Common bus connection by means of clustering.

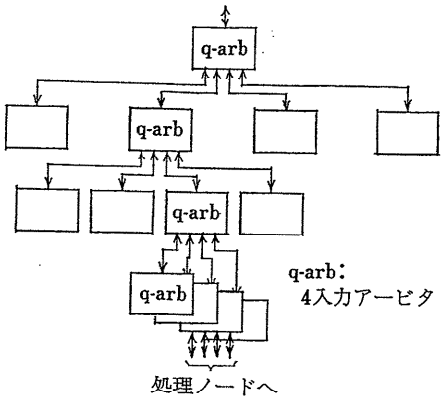


図 8 競合制御回路の構成例
Fig. 8 Example construction of competition control circuit.

共通バスに直接 256 ノードを接続するのは困難だが，たとえば 16 ノードごとにまとめてクラスタ化すれば，直接接続されるのは 16 クラスタとなる．データ線 1 ビット分に注目すると，図 7 のような構成となる．この構成でも論理的にはクラスタ化を行わない場合と等価であり，結合の一様性は保たれている．**図7**の回路の信号遅延を実測したところ，① が 7 nsec，② が 11 nsec，③ が 9 nsec であった．制御信号遅延は ②+③ と同程度で，ターンアラウンド遅延は 47 nsec となる．現在，より高速な TTLIC も入力可能であり，50 nsec 程度のバスサイクルも可能である．

これにより，共通バスの通信容量は 320 MB/sec となり，10 MFLOPS の処理を実現するには十分である．

競合制御には，**図8**に示すように 4 入力アービタを

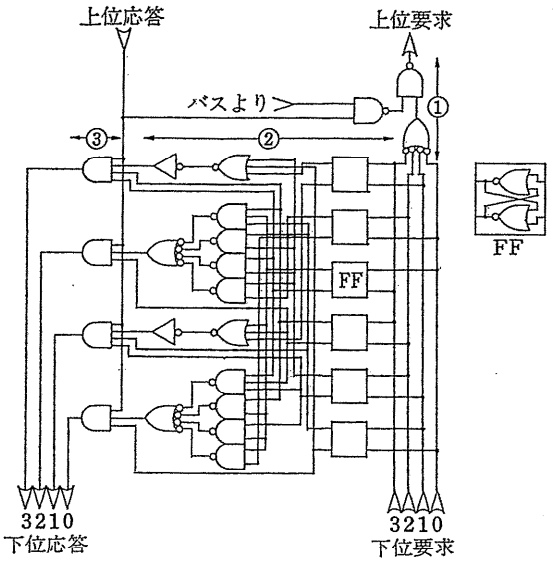


図 9 4入力アービタの回路図
Fig. 9 Circuit diagram of 4-input arbiter.

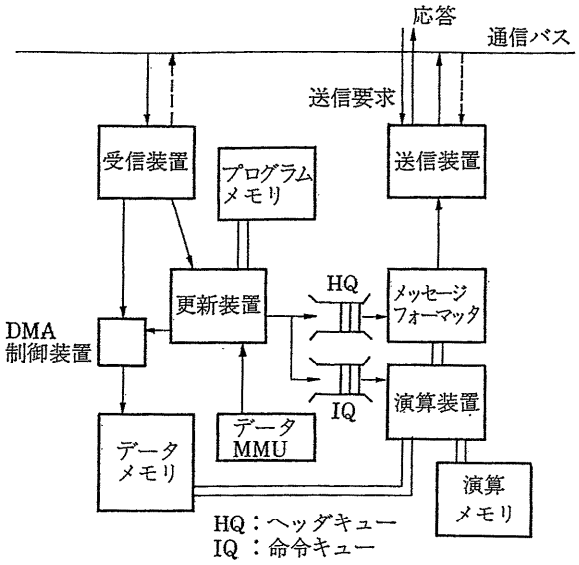


図 10 分散形データフロー計算機の処理ノード構成
Fig. 10 Processing node configuration of the distributed data flow computer.

階層構成にした競合制御回路を用いる．4入力アービタの構成を図9に示す．各要求入力間の発生順序を記憶するRSフリップフロップと，これらの出力から応答出力を作る組合せ論理回路とによって構成される．各部の信号遅延を実測したところ，①が6nsec，②が9nsec，③が3nsecであった．競合制御回路全体の遅延は①×3+②+③×3で，36nsecとなるので，50nsecのバスサイクルに対して十分高速である．

3.2 処理ノード構成

図10に処理ノードの構成を示す．以下では各部の機能と構成について述べる．

3.2.1 演算装置

演算装置は命令キューから命令を取り出して実行する．命令はアクタの演算内容を指定する命令コードと，そのオペランドのデータメモリ中の所在情報より成る．

命令の実際の内容は演算メモリに格納されており，命令コードで指定された処理を，データメモリに蓄えられたオペランドを参照しながら実行する．処理の実行中に得られた結果データは順次メッセージフォーマッタのバッファに書き込み，処理が終了するとメッセージフォーマッタに知らせる．

これらの機能を実現するために，演算装置には高性能のマイクロプロセッサを用いる．要求される処理能力については5章で検討する．オペランドデータはデータメモリから受け取り，結果データはメッセージフォーマッタに渡すため，これらは演算装置の主記憶のアドレス空間の一部として構成する．

3.2.2 メッセージフォーマッタ

メッセージフォーマッタは，ヘッダキューからメッセージヘッダを取り出し，バッファに格納された結果データに付加して送信装置へ転送するハードウェアである．

3.3.3 送信装置と受信装置

送信装置はバス送信手順に従い，バス制御装置に要求(REQ)を出し，応答(ACK)が返されたらバスにアドレスとデータを送出する．受信装置は，アドレスが自ノードの場合，データを受信し，受信応答(RPLY)を返し，バスサイクルを終了する．受信可能でないときは受信応答が返されないので，送信装置は再び要求からやり直す．送信および受信装置には，高速のバスと比較的低速の処理装置との間の速度差を吸収するために高速のFIFOを置き，制御もハードウェアにより実現する．

3.3.4 更新装置とプログラムメモリ

更新装置は受信装置からメッセージのヘッダを受け取り，メッセージ長を解析し，データメモリ管理装置(データMMU)からデータメモリの空きブロック番号を受け取り，DMA制御装置にこれらを与えて，受信装置とデータメモリ間のDMA転送を起動する．次に，ヘッダから目的のアクタとリンク番号を解析し，プログラムメモリに蓄えられた命令セルを更新する．命令セルの構成を図11に示す．タグビットはメッセージの到着を示すフラグであり，全部オンになる

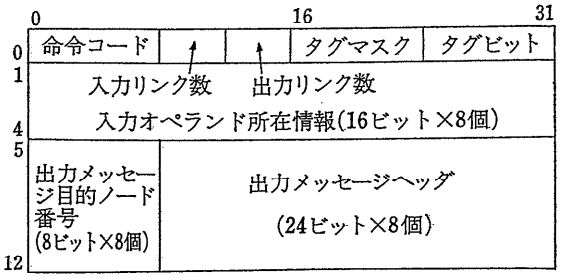


図 11 命令セルの構成
Fig. 11 Instruction cell format.

とこの命令が実行可能となる．タグマスクはタグビットの初期値である．入力オペランド所在情報は到着したメッセージデータを格納したデータメモリのブロック番号である．

更新により実行可能となった場合は，命令コードと入力オペランド所在情報を命令キューに入れ，また，出力メッセージ目的ノード番号と出力メッセージヘッダをヘッダキューに入れる．

これらの機能を実現するために，更新装置には高性能のマイクロプロセッサを用いる．要求される処理能力については5章で検討する．

3.3.5 データメモリ管理装置

データメモリ管理装置は，データメモリを数種類の大きさのブロックに分割管理し，更新装置からの要求により空きブロックを与える処理と，演算処理が終了して不要となったブロックの回収の処理を行う．マイクロプロセッサとハードウェア FIFO により実現する．

3.3.6 データメモリ

データメモリは DMA 制御装置からの書込みと演算装置からの読出しを処理するため，高速スタティック RAM と競合制御回路により，等価的に2ポートメモリとして実現する．

4. 縮小モデルデータフロー計算機 EDAC の構成

本章では，3章で述べた分散形データフロー計算機の構成に従い，この実現性の検証と詳細な検討のために試作した縮小モデルであるデータフロー計算機 EDAC の構成について述べる．

EDAC は，処理ノード2台とし，共通バスのサイクルは 280 nsec，アドレスバス幅は2ビット，データバス幅は 32 ビットの規模とした．したがって，バスの通信容量は提案システムの 0.045 倍となる．

ホスト計算機の OS は，米国ベル研究所開発の UNIX 第6版であり，プログラム開発，システムプログラムの初期ロード，データフロープログラムロード，実行時の入出力制御等に用いている．

表 1 EDAC の共通バスの信号線
Table 1 Signal lines of EDAC's common bus.

種類	信号名	内容
データ	DA 0～DA 31	メッセージヘッダ，データ
アドレス	ADR 0～ADR 1	メッセージ目的ノード番号
制御線	REQ 0～REQ 2	トランスミッタ→バスコントローラ，リクエスト
	ACK 0～ACK 2	バスコントローラ→トランスミッタ，ACK
	DOUT	トランスミッタ→バスコントローラ，データ出力あり
	DIN	バスコントローラ→レシーバ，データ入力
	CK 1～CK 4	制御用クロック
	MS	メッセージスタート
	ME	メッセージストップ
	RS, HRS	リセット

以下に共通バスと処理ノードの構成について述べる．

4.1 共通バス構成

バス線の構成を**表1**に示す．EDAC ではノード数がホストを含めて合計3台であるので，REQ 線，ACK 線とも3本であるが，ノード数を増やす場合はこれも増える．しかし，N台のノードを一次元的に配置しても，4入力アービタを最適に分散配置すると，REQ と ACK 線の総配線面積は $O(N \log_4 N)$ となり，256 ノードでも問題ない．

メッセージ構成は，図6において DA を9ビット，DL を3ビット，ML を4ビットとし，データ部を2～62バイトとした構成である．

バス制御装置は図9の4入力アービタを縮退させた3入力アービタと，同期のためのクロック発生器より成る．バス送受信手順を**図12**に示す．

送信装置から送信要求 (REQ) が出されると，バス制御装置は，アービタにより選択された一つの送信要求に対し，送信許可の応答 (ACK) を返す．メッセージ送出中は ACK は出し続ける．送信装置は，ACK が返ってくるとメッセージをバスに送出する．受信応答 (RPLY) が返されない場合は送出をやめ，一時 REQ を下げて再び送信手順を開始する．受信装置は，メッセージ開始 (MS) が出されたときにアドレス解析を行い，自ノード宛ての場合は，受信バッファである FIFO に1メッセージ分の空きがあれば RPLY を

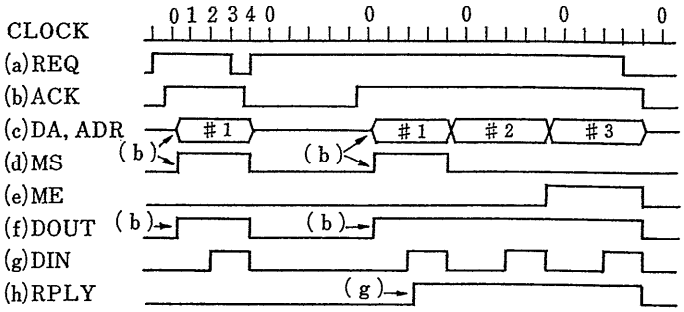


図 12 バス送受信手順
Fig. 12 Send and receive protocol of bus.

返す．メッセージ終了 (ME) が出されるまで RPLY は出し続ける．

この手順により，非同期に発生したメッセージを，任意の目的ノードへ相互の衝突なしに転送することが可能となる．

4.2 処理ノード構成

EDAC の処理ノードの構成は，図 10 に示した提案システムの処理ノード構成において，データ MMU を取り除き，その処理を更新装置に含めた構成であり，その他については，基本的構成は等しい．

更新装置には DEC 社の 16 ビットマイクロプロセッサ LSI 11/2，演算装置には同じく LSI 11/23 を用い，その他は LSI, MSI, SSI によるランダムロジックで実現している．

演算装置の処理の流れを図 13 に示す．これは命令キューから命令コード，オペランド所在情報を取り出し，メッセージフォーマッタが入力可能であると，命令実行を開始する．終了すると，メッセージフォーマッタに終了信号を与え，次の命令実行に移る．試作システムのため，システムプログラムの保護機能等はない．

メッセージフォーマッタではハードウェア制御によるダブルバッファリングを行っており，演算装置からの書込みと送信装置への転送が並行して実行できる．

更新装置の処理の流れを図 14 に示す．メッセージ受信処理は到着したメッセージの先頭の 1 ワード (16 ビット) を占めるヘッダを取り込み，メッセージ長を解析し，データメモリの空きブロックを確保して，DMA 制御装置を起動し，メッセージデータの格納を行い，次に，ヘッダから目的のアクタ番号とアクタのオペランドに対応するリンクの番号を解析し，プログラムメモリ内の命令セルを更新する．更新によりその命令が実行可能となると，実行可能命令処理を行う．この処理では，命令の種類を示す命令コードと，その実行に必要なオペランドデータの所在情報を命令キューに入れ，命令の演算結果をメッセージに組み立てるために必要なヘッダをヘッダキューに入力する．

命令セルは，図 11 において，入力オペランド所在情報を 8 ビット×8，出力メッセージ目的ノード番号を 2 ビット×8，出力メッセージヘッダを 8 ビット×8 とした構成である．データメモリ管理処理で扱うブロックは固定長とした．

4.3 システムソフトウェア構成

これまで，EDAC の実行時の動作を中心にその構成を述べてきた．この構成において，データフロー計算機を実現するためには，更新装置と演算装置に制御のための制御プログラムが必要である．また，ホストノードにおいては，処理ノードにおいてハードウェア

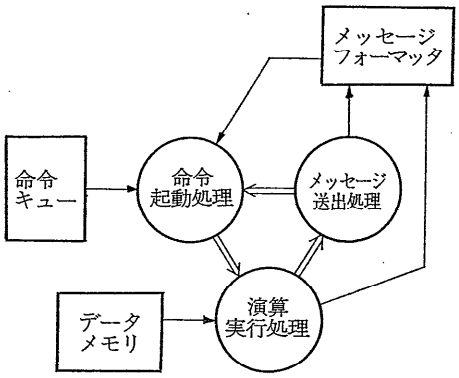


図 13 演算装置制御プログラムの構成
Fig. 13 Configuration of operation unit control program.

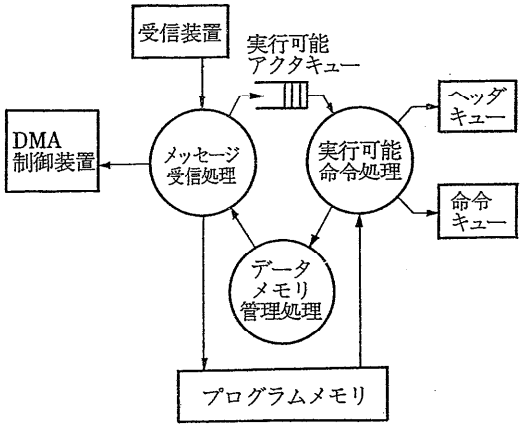


図 14 更新装置制御プログラムの構成
Fig. 14 Configuration of updater control program.

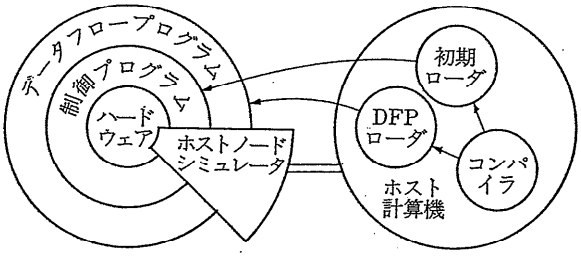


図 15 EDAC のソフトウェア構成
Fig. 15 Software configuration of EDAC.

で実現している処理を代行するソフトウェアのシミュレータが必要である．ホスト計算機としても，その他にいくつかのプログラムが必要である．

これらのソフトウェアの全体構成を図15に示す．制御プログラムはホスト計算機の外部記憶装置に格納されており，システムの起動時に初期ローダにより通信バスを介して処理ノードへロードする．実用システムではこれらの制御プログラムは ROM 化して処理ノード中に常駐させるべきであるが，EDAC は実験システムであり，制御方式の変更に伴うプログラムの変更等が容易に行えるよう，このような方式を採用した．

EDAC 上で実際にデータフロープログラムを実行するには，ホスト計算機のコンパイラでコンパイルされた機械語表現のデータフロープログラムを処理ノードへロードする必要があり，これも DFP ローダにより通信バスを介して行っている．

5. EDAC の評価と提案システムの実現性の検討

4章に述べた構成により EDAC を試作した．このシステムにおいて，演算装置と更新装置のシステムソフトウェアのオーバヘッドを測定した結果，演算装置の命令起動処理とメッセージ送出処理の合計の処理時間は 0.25 msec，また，更新装置の実行可能命令処理時間は 0.4 msec，1メッセージ当りの受信処理時間は 0.85 msec であった．

数値計算等を考えると，1命令当りの入力メッセージ数は3程度となることが多く，更新装置と演算装置の負荷のバランスをとるためには，演算装置の全処理の和が，更新装置の実行可能命令処理と受信処理の合計の処理時間，すなわち $0.4+0.85\times3\simeq3$ msec と等しくなることが望ましく，1アクタ当りの演算時間は約 2.7 msec とすればよい．これは，演算装置の処理能力からすると，アクタの処理内容を約20浮動小数演算程度とするのが最も効率のよいアクタのレベルとなることを示す．

表 2 メッセージ受信処理ステップ数
Table 2 Number of steps in message receiving process.

処理内容	命令ステップ数
初期化，後処理	9
ヘッダ読込・解析	5
空ブロック確保	7
DMA セット	6
命令セル更新	4
実行可能アクタ処理	7
合　計	38

実際の数値計算等を考えると，アクタのレベルを5ないし10浮動小数演算とするのがよい場合が多いと思われる．

EDAC で演算装置と更新装置に用いたマイクロプロセッサは，基本的な命令の実行時間が，それぞれ約 3 μsec と 8 μsec であり，上記の結果を考え合わせると，両プロセッサは同程度の処理能力をもつべきである．

提案システムの実現性を考えると，256 台の処理ノードで 10 MFLOPS を実現するためには，各演算装置の1浮動小数演算実行時間は 20 μsec 程度以内でなければならないから，1アクタ当りの演算時間は約 200 μsec となる．これは EDAC の場合の10分の1以下であり，したがって，この実現には，処理能力を演算装置で約3倍，更新装置で約13倍にする必要がある．

演算装置については，図10において，命令キューを命令コードと各オペランド分をすべて並列に置くことにより，3倍程度の処理能力の改善が期待できよう．したがって，演算装置の基本的処理能力は変えずに，浮動小数演算回路部だけを高速化することで対処できる．

更新装置については，演算装置と同じプロセッサを用いると，約 2.5 倍の性能となる．また，EDAC のシステムソフトウェアはC言語で記述しているが，これを最適化することにより 40% 程度の改善が期待できる．これにより，実行可能命令処理時間が 100 μsec，受信処理時間が 200 μsec 程度となり，入力メッセージ数を3とした場合の1アクタ当りの更新装置の処理時間は 700 μsec で，さらに 3.5 倍の性能改善が必要である．

EDAC の受信処理の内容と処理ステップ数の内訳

を表2に示す．データメモリ管理機能のハードウェア化と，実行可能命令処理の機能分散化により，初期化，後処理，空ブロック確保が不要となり，さらにヘッダ解析と DMA 起動を組み合わせてハードウェア化することにより，メッセージ受信処理の処理量を3分の1程度にすることができるから，入力リンク数を3とすると，提案システムに要求される処理能力を得ることができる．

以上，縮小モデルの試作データフロー計算機による処理ノードの構成に関する評価を通じて提案システムの実現性に関する検討を行った．

通信バス，およびシステム全体の評価については，別に計算機シミュレーションによって行っており，目標性能の実現性の確認をしている．その詳細については別途，ご報告する予定である．

6. むすび

データフロー計算機の分類を行い，諸形態の得失を検討し，その結果にもとづき，10 MFLOPS の演算能力を実現するための分散形・中レベル・一様構成に基づくシステムを提案し，その全体構成を示した．本システムにおいては結合回路網に共通バスを採用し，処理ノードを多数置くことにより並列処理を行っている．また，処理ノード内の各種処理をパイプライン化し，処理能力の向上を図っている．

この構成に基づき，提案システムの実現性の検証のため，縮小モデルの試作データフロー計算機 EDAC を試作し，性能評価を行った．この結果，EDAC では小規模の試作モデルであるため更新装置の受信処理がボトルネックとなっていることが見いだされたが，機能分散とハードウェア化により提案システムの実現が可能であることは明らかになった．

謝辞 なお本研究は文部省科学研究費補助金一般研究A「分散形データフロー計算機の構成に関する研究」の補助を得て行われたものであり，ここに記して感謝の意を表す．

参考文献

1) Dennis, J. B. and Misunas, D. P.: A Preliminary Architecture for a Basic Data-flow Processor, *Proc. 2nd Annu. Symp. on Comput. Archit.*, pp. 126-132 (Jan. 1975).
2) 浅田，グェン，堀，斉藤，猪瀬：データフロー計算機における故障検出の一方式，情処学会分散処理システム研究会資料，1-4，pp. 1-10 (1979).
3) グェン，浅田，大山，斉藤，猪瀬：環状回路網を用いた分散型データフロー計算機の一方式，信学論，Vol. J65-D, No. 12, pp. 1528-1535 (1982).
4) Gostelow, K. P. and Thomas, R. E.: Performance of a Simulated Data-flow Computer, *IEEE*, Vol. C-29, No. 10, pp. 905-919 (1980).

(昭和58年3月11日受付)
(昭和58年7月19日採録)

論　文

FIFO付きブロードキャストメモリを用いた共有メモリの実現とその評価

正　員　齊藤　忠夫[†]　　正　員　相田　　仁[†]
正　員　小口　正人[†]　　非会員　黒岩　　実[†*]

Implementation and Evaluation of a Shared Memory Using Broadcast Memory with FIFO

Tadao SAITO [†], Hitoshi AIDA [†], Masato OGUCHI [†], *Members and*
Minoru KUROIWA [†*], *Nonmember*

あらまし　複数のプロセッサ間で共有メモリを構成する場合，メモリアクセスのボトルネックにより，システムの性能が大幅に低下してしまうことが問題となる．このボトルネックに対処するための方式の一つに，ブロードキャストメモリがある．ブロードキャストメモリは，各プロセッサにメモリ内容のコピーをもたせた方式であり，メモリ読出しの際のアクセス競合の低減に効果がある．しかし一方で，メモリ書込みの際，結合バスにおいてアクセス競合が存在する．そこで本論文では，ブロードキャストメモリにFIFOバッファを設け，メモリ書込みの際にもアクセス競合を低減する方式を提案する．論文中では，まずこのFIFO付きブロードキャストメモリを用いて共有メモリを構成する方式について検討する．そしてこの方式と他の方式との比較を，シミュレーションにより行った結果で示す．更にこの方式により市販のワークステーション3台を用いて構成した実験システムの構成について述べ，数値計算のアプリケーションを用いて行った実験システムの性能測定結果を示す．

キーワード　共有メモリ，ブロードキャスト，FIFOバッファ，並列処理，ワークステーション

1. まえがき

複数のプロセッサ間において共有メモリを実現する場合，検討すべき問題がいくつかある．まず一つにはメモリに対するアクセス競合が起こりプロセッサがブロックされることにより，システムの性能が大幅に低下する可能性がある．また特にプロセッサの台数が多くなると，すべてのプロセッサから物理的に近い位置にメモリを配置することが難しくなり，アクセス時間そのものが大きくなる．これらの問題は，プロセッサの性能が向上するほど，相対的にシステム全体の性能に与える影響が大きくなる．

これらの問題点を解決すべく提案された方式の一つに，ブロードキャストメモリ(1)~(4)がある．ブロードキャストメモリは，各プロセッサにメモリ内容のコピーをもたせ，読出しは各自のメモリにおいて行い，書込みを行うときにはその内容をブロードキャストしてすべてのメモリに同時に書き込むという方式である．この方式ではメモリからの読出しを独立かつ同時に行うことができるため，メモリへのアクセス競合を低減させることができる．また各メモリをプロセッサから物理的に近い位置に配置することができるため，アクセス時間そのものを小さくする効果もあると考えられる．

ブロードキャストメモリでは，メモリからの読出しは独立に行うことができるのだが，書込みは全メモリに対して同時に行う必要があるため，アクセスの競合が起こる．そこで本論文では，ブロードキャストメモリにおいて各プロセッサにFIFOバッファをもたせることにより，書込みの際のアクセス競合も低減させることを目的とした方式を提案する．本論文では，まずこのFIFO付きブロードキャストメモリの構成法について述べ，次に他の方式との比較をシミュレーションを用いて構成したFIFO付きブロードキャストメモリ

† 東京大学工学部電子情報工学科，東京都
Faculty of Engineering, The University of Tokyo, 113 Japan
* 現在，日本電気株式会社C&Cシステム研究所

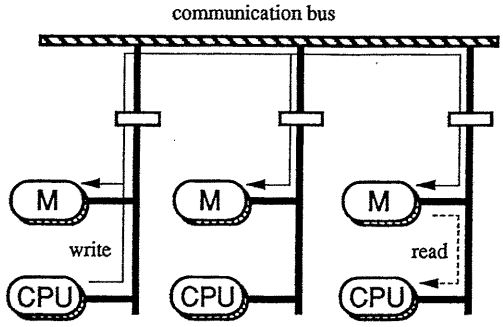


図1 ブロードキャストメモリの概念図
Fig. 1 Conceptual diagram of a broadcast memory.

の実験システムを紹介し，その評価について述べる．

2. FIFO付きブロードキャストメモリを用いた共有メモリの実現方法

図1に通常のブロードキャストメモリの概念図を示す．各プロセッサはコピーされた同じ内容のメモリをもっており，読出しは各々が自分のメモリより行う．一方，書込みについては，1箇所で書込みが行われるとそれが結合バスを介して全プロセッサのメモリへブロードキャストされて書込みが行われ，メモリ内容が同一に保たれる．すなわち，各プロセッサはメモリ内容を共有しているが，メモリ自体は共有していない．このような方式を実現することにより，メモリ読出しのアクセス競合を低減できるだけでなく，アクセス時間そのものの短縮を図ることができる．

このようにブロードキャストメモリではメモリからの読出しは独立かつ同時に行うことができるが，しかし一方で，書込みに関しての競合は避けることができない．すなわち1箇所で書込みが行われるときには各プロセッサからのメモリに対するすべてのアクセスを待たせなければならない(以降においてこれをバスロック形と呼ぶ)．

そこで本論文ではブロードキャストメモリにFIFOを付加し，書込みの際にも競合を低減させる方式を提案する．これはブロードキャストを行う結合バスの待合せ用と各プロセッサのメモリの待合せ用にそれぞれS-FIFO，R-FIFOと呼ぶFIFO (First In First Out memory) を設けるものである．これにより，書込みを実行するときに他のプロセッサがバスに対して書込みを行っている場合には，FIFOにより待ち合わせることができる．このFIFO付きブロードキャストメモリを図2に示す．

プロセッサからの書込みに対しては，そのアドレス

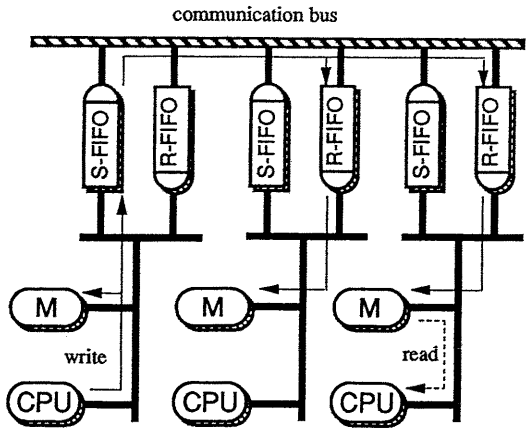


図2 FIFO付きブロードキャストメモリの概念図
Fig. 2 Conceptual diagram of a broadcast memory with FIFO.

とデータがまず送信バッファS-FIFOに取り入れられ，プロセッサから見た書込みサイクルは，S-FIFOに書き込んだ時点で終了する．次にこれらのアドレスおよびデータは結合バスを介して他のプロセッサエレメントにブロードキャストされ，すべてのプロセッサエレメントの受信バッファR-FIFOに取り入れられる．最後にR-FIFOから各プロセッサごとのメモリに対して書込みが行われる．読出しについては，それぞれのプロセッサエレメントのメモリから，結合バスを介すことなく独立に行うことができる．

このようにFIFOを備えることにより，書込みの際にも基本的に競合を回避することができる．但しFIFOを用いていることにより，書込みサイクルの終了から実際に他のプロセッサエレメントのメモリに書き込まれるまで遅延が生じることに注意しなければならない．このためこのメモリ上でTest-and-Setなどの方法で同期をとることはできず，同期のためには別の機構が必要である．

3. FIFO付きブロードキャストメモリの性能評価

本章ではプロセッサ間の同期を必要としないアプリケーションを仮定し，シミュレーションによってFIFO付きブロードキャストメモリの性能評価を行った結果を示す．比較対象としては，バスロック形ブロードキャストメモリを考え，更にメモリとして，読出しと書込みを同時に行える1-read 1-writeのマルチポートメモリを用いた場合と通常のシングルポートメモリを用いた場合の比較も行う．

すなわち比較検討するモデルは，図3に示す4方式

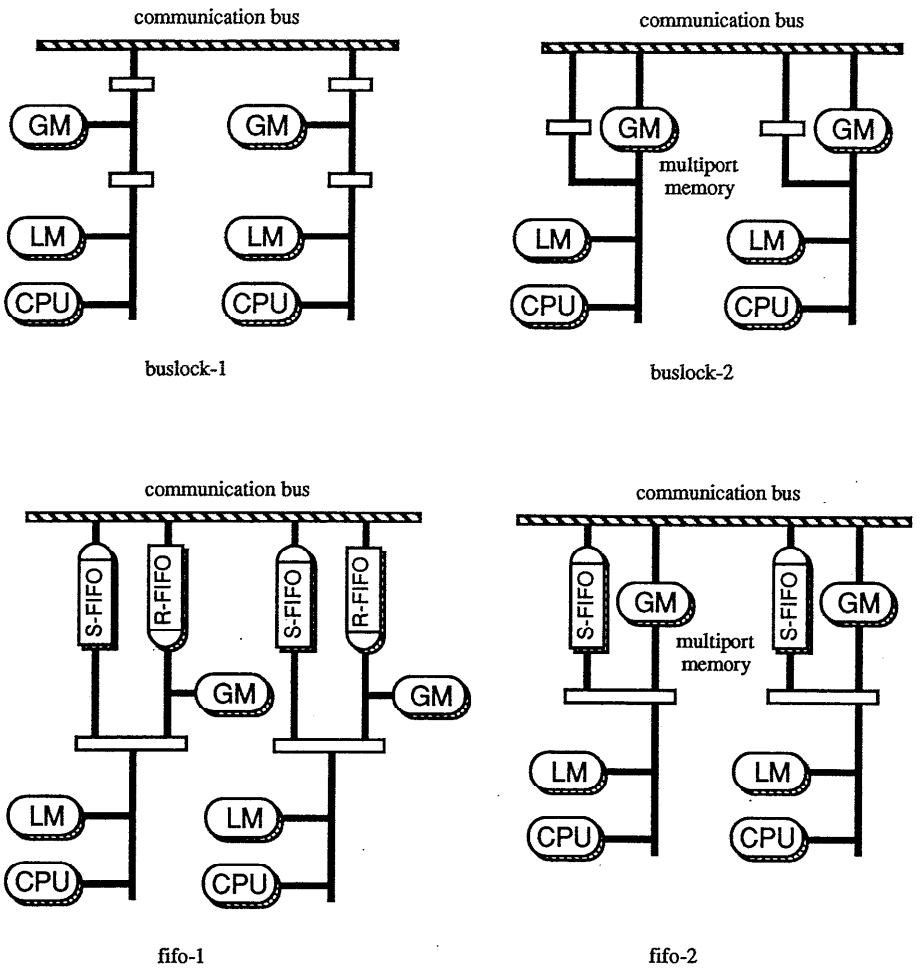


図3　シミュレーションで比較した4方式
Fig. 3　Four systems compared by simulation.

である．

・buslock-1：シングルポートメモリを用いたバスロック形ブロードキャストメモリ

・buslock-2：マルチポートメモリを用いたバスロック形ブロードキャストメモリ

・fifo-1：シングルポートメモリを用いたFIFO付きブロードキャストメモリ

・fifo-2：マルチポートメモリを用いたFIFO付きブロードキャストメモリ

各プロセッサエレメントは，結合バスによって接続される．結合バスの制御方法は要求到着順とする．すべてのモデルにおいて，各プロセッサエレメントは，ブロードキャストの対象となるグローバルメモリとは別に固有な内容をもち，独立してアクセス可能なローカルメモリをもつ．ローカルメモリへのアクセスは，他のプロセッサエレメントのグローバルメモリへのアクセスと同時に行うことができる．

またバスロック形では，一度には一つのプロセッサしかグローバルメモリへの書込みができない．マルチポートメモリを用いているモデルでは，このとき，他のプロセッサがグローバルメモリの読出しを行うことができるが，シングルポートメモリを用いたモデルでは，読出しも競合により待たされる．一方FIFO付きにおいては，バスの使用状況にかかわらずグローバルメモリへの書込みを行うことができ，グローバルメモリへの書込みが待たされるのは，FIFOが飽和した場合だけである．またシングルポートメモリを用いた場合には，プロセッサからのグローバルメモリの読出しとR-FIFOからの書込みが競合する．このときには，メモリからの読出しを優先するものとする．

評価のために仮定したパラメータは，以下のとおりである．

・バスの転送速度：500 ns/access

・FIFO長：50

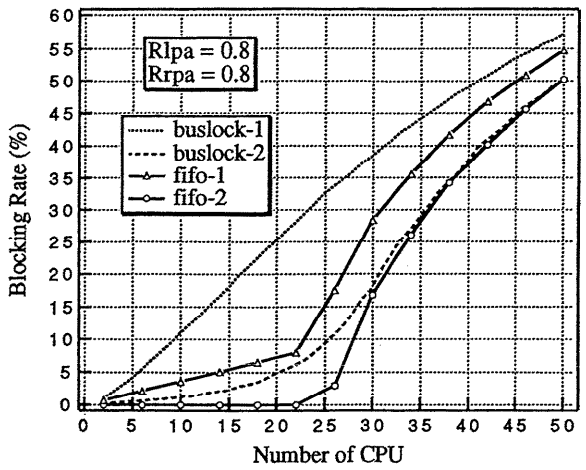


図4 各方式のブロック率
Fig. 4 Block rate of each systems.

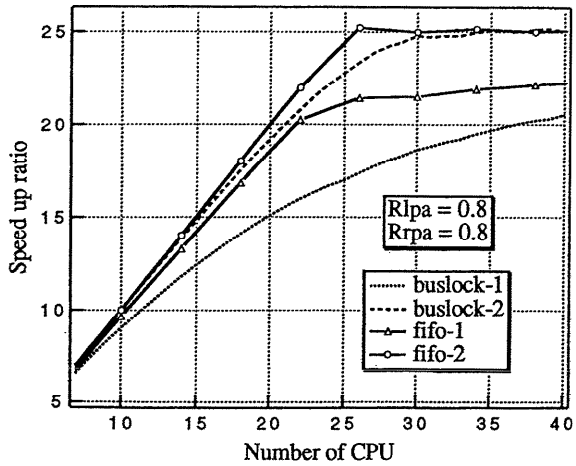


図5 各方式の並列度
Fig. 5 Comparison of performance of each systems.

・ローカル・グローバルアクセス比：4：1
・リード・ライトアクセス比：4：1

各方式の性能を評価，比較する尺度としては，プロセッサのブロック率および並列度を用いた．ブロック率とは，プロセッサがアクセス競合，あるいはS-FIFOの飽和によって待たされた時間を全シミュレーション時間で割った値である．また1-(ブロック率)はプロセッサが有効な動作を行った時間，すなわち稼動率と考えられ，この稼動率にプロセッサの台数を掛けた値は，単一のプロセッサで処理をしたときの値を1とした，マルチプロセッサシステムとしての性能であり，これを並列度とする．

シミュレーションの結果として図4，図5を得た．図4にブロック率，図5に並列度を示す．プロセッサ台数が22〜25台以上になると，プロセッサのグローバルメモリアクセスが結合バスの能力を超えて飽和するため，ブロック率が急激に増加している．バスが飽和するときのプロセッサ台数は，次の式で求められる．プロセッサの台数をN，ローカルアクセス比をR_{lpa}，リードアクセス比をR_{rpa}とすると，buslock-2，fifo-2において結合バスの飽和条件を表す式は

$$1/500\text{ns}=N\times(1/500\text{ns})\times(1-R_{lpa})\times(1-R_{rpa})$$

これより

$$N=\frac{1}{(1-R_{lpa})(1-R_{rpa})}$$

となり，$R_{lpa}=R_{rpa}=0.8$のときに$N=25$となる．一方buslock-1とfifo-1においては，グローバルメモリに接続されたバスにおけるアクセス競合によりシステムの性能が飽和するが，このときの飽和条件を表す式は

$$1/500\text{ns}=N\times(1/500\text{ns})\times(1-R_{lpa})\times(1-R_{rpa})$$
$$+(1/500\text{ns})\times(1-R_{lpa})\times R_{rpa}$$

これより

$$N=\frac{1-(1-R_{lpa})R_{rpa}}{(1-R_{lpa})(1-R_{rpa})}$$

となり，$R_{lpa}=R_{rpa}=0.8$のときに$N=21$となる．

プロセッサ台数がこの値以下の範囲で各方式を比較してみると，FIFO付きが有利な値を示していることがわかる．特にfifo-2は，バスが飽和しない領域ではブロック率がほとんど0であり，最も優れたシステム構成法である．またbuslock-1とfifo-1を比較しても，やはりFIFOの効果が顕著に現れている．これは，グローバルメモリへの書込みの際に，すべてのプロセッサのグローバルアクセスの終了を待つ必要がないということと，グローバルメモリからの読出しをR-FIFOからのメモリへの書込みに優先して行えるということによるものであると思われる．

またバスが飽和した領域においても，同じ形のメモリを使用したモデル同士を比較すると，FIFO付きの方がブロック率が低くなっている．このことにより，バースト的なグローバルアクセスが生じた際の性能も，FIFO付きが有利であると言える．

並列度を表す図5から，buslock-1以外の方式では，プロセッサの台数を増やしていったときの性能低下をかなり軽減できていることがわかる．またこの図より，マルチポートメモリを使用すると，並列度の飽和値を高くできるということがわかる．

4. 実験システムの構成

前2章において，FIFO付きブロードキャストメモリの構成と特徴，および基本的な性能評価について述べ

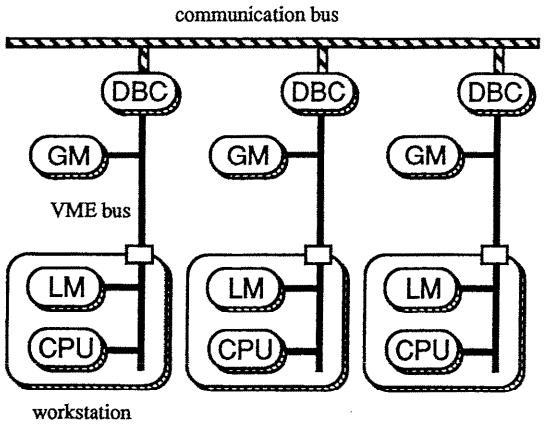


図6　FIFO付きブロードキャストメモリの実験システム
Fig. 6　Conceptual diagram of the experimental system of broadcast memory with FIFO.

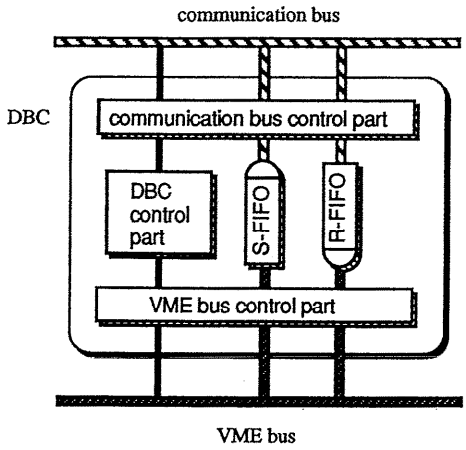


図7　DBCの構成
Fig. 7　Block diagram of DBC. (Distributed broadcast controller)

た．これによりシステムの基本的な性能を予測することができたが，実際のアプリケーションにおいては，アクセスのバースト性やリード・ライト比の時間的変動などがあり，シミュレーションによる結果だけではなお不十分である．また並列処理を行う場合に，同期が必要となるケースはかなり存在すると考えられ，そのようなアプリケーションを実際に走らせてみたときに，どの程度システムの性能が引き出せるのか確かめる必要がある．

このような理由により，市販のワークステーションを用い，実際にFIFO付きブロードキャストメモリの実験システムを構築した．実験システムは図6に示すように，3台のワークステーションにより構成されている．基本的には前章のfifo-1方式によっている．

ワークステーションには，モトローラ社のMC68020またはMC68030をプロセッサにもつSONY製NEWS 3台を用いた．ワークステーションの外部バス（VMEバス）にグローバルメモリとして用いるためのメモリボード（市販）とブロードキャスト機能を実現するためのDBC (Distributed Broadcast Controller) ボードを接続して，このDBC同士を結合バスで接続することによりシステムを構成した．なおワークステーション内のメモリはローカルメモリとして，ブロードキャストの対象にはならない．

DBCは図7に示すように，S-FIFO，R-FIFO，DBC制御部，VMEバス制御部および結合バス制御部より構成され，全体がダブルハイトより少し大きいボード1枚の上におさまっている．FIFOはともに1024 word (1 word=4 Byte) のアドレスとデータを格納する容量をもっている．VMEバス制御部は更にアドレスデコーダ，タイマ，VMEバスリクエスタ，VMEバスマスタなどの機能に分割され，制御部全体でおよそ9個のPALによって実現されている．

DBC同士を接続する結合バスには，アドレスとデータを1サイクルで転送する同期方式のバスを使用した．最大の転送速度は，300 Mbit/sである．結合バスの調停方法にはラウンドロビン方式を採用し，バスの使用権を獲得したらS-FIFO内のすべてのデータをブロードキャストしてから次のDBCに使用権を渡すようにして，アービトレーションによるオーバヘッドを削減している．R-FIFOがあふれた場合には，バス上に信号が出され，送信が停止される．

5.　実験システムの性能評価

ここでは前章で紹介した，3台のワークステーションにより構成した実験システムの性能測定結果を示す．

測定にあたっては，実際にアプリケーションを走らせ，実行時間の測定を行って，ワークステーションの台数を増やしたときの速度向上比を求めた．アプリケーションの例として，ここではヤコビ法によって連立方程式を解く問題について説明する．ヤコビ法とは，例えば係数行列が3×3の連立方程式

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

であれば，これを以下の式に変形し，計算を繰り返すことによって収束値を求める反復法である．

$$x_1^{k+1} = (b_1 - a_{12}x_2^k - a_{13}x_3^k)/a_{11}$$
$$x_2^{k+1} = (b_2 - a_{21}x_1^k - a_{23}x_3^k)/a_{22}$$
$$x_3^{k+1} = (b_3 - a_{31}x_1^k - a_{32}x_2^k)/a_{33}$$

このヤコビ法で，係数行列が30×30，60×60，90×90，120×120の連立方程式を解かせ，実行時間を測定した．評価条件としては，$x_1 \sim x_n$ $(n=30,60,90,120)$ に初期値0を与え，反復計算を実行してノルム $\left(\sum_{i=1}^{n}(x_i^{k+1}-x_i^k)\right)$ がある一定値 ε 以下になったときを実行終了とみなすこととし，計算を実行するワークステーションの台数を変えたときの実行時間を求めた．このアプリケーションにおけるデータの取扱いでメモリに対するローカル・グローバルアクセス比，リード・ライトアクセス比はともにほぼ7.5：2.5程度であった．ワークステーションへの負荷の分配方法としては，連立方程式を3等分して3台のワークステーションで解く場合，2等分して2台で解く場合，および1台で全部解く場合の3通りについて測定した．1台で解く場合には，逐次処理との比較のために，並列処理の際に他のワークステーションと共有する必要のある変数をグローバルメモリ上にとって計算を行う場合と，すべてローカルメモリだけで計算を行う場合について，測定した．

また複数のワークステーションで連立方程式を分担する場合には，ワークステーション間で同期をとる必要が生じる．ここでは計算を行っているワークステーションのうちの1台に同期管理も受けもたせ，このワークステーションがループの実行開始，実行終了の合図をグローバルメモリに書き込むことにより同期を実現した．

測定により得られた実行時間から，ワークステーションの台数に対する速度向上比を求めた結果を図8に示す．2本の破線はそれぞれワークステーション1台の場合にローカルメモリだけを使用した場合と，グローバルメモリを使用した場合と比較した理想直線である．

グローバルメモリ使用時の理想直線と比較すると，特に連立方程式の次数が大きくなった場合，速度向上比は理想直線にほぼ等しくなっている．一方ローカルメモリ使用時の理想直線と比べて速度向上比が少し落ちているのは，ローカルメモリとグローバルメモリに対するアクセス時間の違いによるものであり，これは市販のワークステーションをそのまま使用しているということを考えると，満足できる結果である．

6. むすび

本論文では，複数のプロセッサ間で共有メモリを実現する際のアクセス競合を解決する方法として，FIFO付きブロードキャストメモリを提案した．シミュレーション

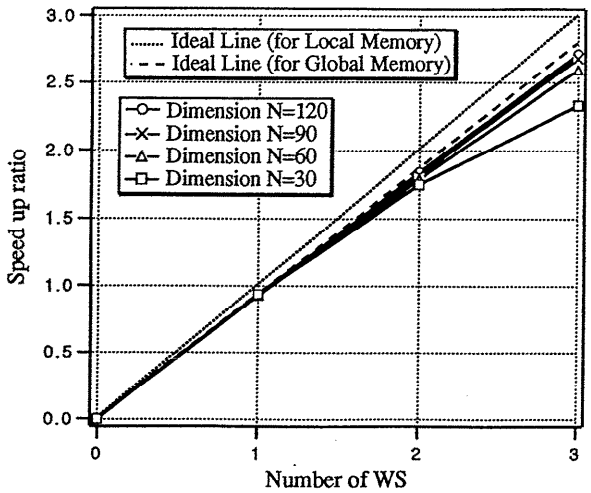


図8 ヤコビ法による連立方程式解法の場合の速度向上比
Fig. 8 The speed up ratio in execution of Jacobi method for simultaneous equation.

により，FIFO付きのブロードキャストメモリがバスロック形のブロードキャストメモリより有利であるという結果が示された．また3台のワークステーションを用いて実際に試作したマルチプロセッサシステムの構成を紹介し，その性能測定結果について述べた．その結果このシステムは単に市販のワークステーションの外部バスにボードを追加した構成であるにもかかわらず，ワークステーションの台数にかなり近い速度向上比が得られることが示された．

筆者らはこのFIFO付きブロードキャストメモリが，分散環境における共有メモリの実現に有効であろうと考えている．しかし本論文に示したようなバス接続方式では分散環境に拡張することは難しく，光ファイバリングのような他の方式を用いる必要がある．この場合にはプロセッサ間の同期やメモリ内容の一貫性の維持が大きな問題になると考えられ，現在これらの事柄について，更に検討を進めている．

謝辞 研究を行うにあたり御指導や御協力を頂きました東京大学電子情報工学科の齊藤・相田研究室の皆様に感謝致します．

文　献

(1) 金田悠紀夫，小畑正貴，前川禎男："マトリクスブロードキャストメモリ結合形並列計算機による n 元連立一次方程式の $O(n)$ 時間計算"，情処学論，**23**, 5, pp. 570-575 (1982-09).
(2) 小畑正貴，金田悠紀夫，前川禎男："ブロードキャストメモリ結合形マルチマイクロプロセッサシステムの試作"，情処学論，**24**, 3, pp. 351-356 (1983-05).
(3) 天野英晴，地川淳二，吉田隆一，相磯秀夫："マルチリー

ド・メモリを用いた並列計算機の性能解析”，信学論(D)，**J67-D**, 9, pp. 1028-1035 (1984-09).

(4) Amano H., Boku T. and Kudoh T. : “(SM)²-II : A Large-Scale Multiprocessor for Sparse Matrix Calculations”, IEEE Trans. Comput., **39**, 7, pp. 889-905 (July 1990).

(平成4年7月8日受付)

齊藤　忠夫

昭38東大・工・電子卒．昭43同大大学院博士課程了．工博．東京大学工学部講師，東京大学助教授，東京大学教育用計算機センター主任，カリフォルニア工科大学研究員等を経て現在東京大学教授．ディジタル通信方式，コンピュータ・ネットワーク，ディジタルシステムの信頼性などの研究に従事．著書としては「電子回路入門」(昭晃堂)，「ディジタル回路」(コロナ社)などがある．昭42，昭62通信学会論文賞，昭48通信学会業績賞など受賞．電気学会，IEEE各会員．

相田　　仁

昭55東大・工・電気卒．昭60同大大学院博士課程了．同大工学部電気工学科助手．昭61同講師．SRIインターナショナル(米国)客員研究員を経て平2東大助教授．現在同大教育用計算機センター主任．工博．高並列推論計算機，分散オペレーティングシステム，知識処理などの研究に従事．著書に「第5世代コンピュータ入門」(共著)がある．情報処理学会，人工知能学会，ソフトウェア科学会，IEEE，ACM各会員．

小口　正人

平2慶大・理工・電気卒．平4東大大学院・工・電気修士了．現在同大大学院・工・電子博士課程在学中．計算機システムの研究に従事．

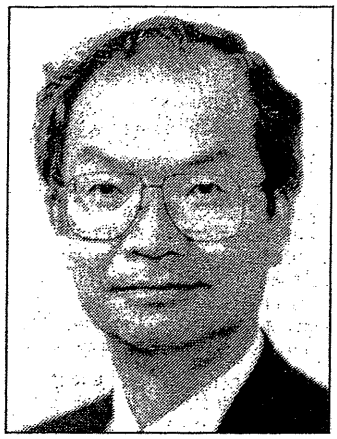
黒岩　　実

平1・東大・工・電気卒．平3同大大学院・工・情報修士了．同年日本電気㈱入社．並列処理システムに興味をもつ．

8. 標準化

○ A proposal for dejure standardization by implementation

○ Regional activities for telecommunication standardization



A PROPOSAL FOR DEJURE STANDARDIZATION BY IMPLEMENTATION

Tadao Saito
University of Tokyo
(Japan)

7.3

Abstract

N I I projects planned in many countries are based on drastic cost performance improvement of electronics. Similar to the cost performance improvement of computers, cost performance of communication networks will also be improved in near future. This means that video communication cost of near future will be essentially identical to voice communication cost of present day. Variety of new applications and new businesses are to be created. New intellectual society will be created through open exchange of information through the network.

Standardization is the key to guarantee free exchange of information. The scope of standardization will be extended and quick standardization is requested to realize new networks in synchronism with technology improvement. On the other hand, standardization of software related technology is more important. Interoperability of all the aspect of network is more and more requested. In some part of the network defacto standard is to be included in N I I standard. A new framework is needed to realize quick, efficient and fair standardization processes.

1. Introduction

Remarkable advance in electronics characterizes the worldwide activity of last two decade of the 20th century. The technology changed a dream of the past into a reality. Cost of advanced electronics has been drastically reduced by the order of 10^3 or more in many applications which enjoyed mass production. New products has been put into market much faster than the past. In communication systems, too, application of electronics made it possible to realize rapid development of products, however, standardization is essential for product realization. Time needed for standardization tend to delay w systems and new services. In worldwide development of next generation information infrastructure represented by information superhighway or N I I / G I I, efficiency in standardization is requested.

Dependence on software in implementation of standard is an another important trend of technology. Therefore, a standard is described as a specification of software. Standardization of function of software is very much different from standardization of hardware. The software standardization request a new scheme for standardization.

2. Plan for Next Generation Information Infrastructure

2-1 Background Technology

To improve cost performance of telecommunication network by the order of 10^3, what is needed first is the technology to improve the performance of the network components by the order of 10^3. Telecommunication system is composed of variety of component including subscriber lines, switches and long line transmission facilities. If technology having performance improved by the factor of 10^3 is realized with essentially the same order of cost as the present technology, cost performance improvement of 10^3 is realizable.

For subscriber lines, switches and long line transmission the most promising candidates are optical fiber subscriberlines, ATM switches and coherent optical transmission systems.

2-2 Objective of N I I

In modern society we have many social problems which are different country by country. Industrialization was a key to solve the problem of social poverty in past. For regional development, to encourage factory construction, construction of transportation facilities including sea port, highway and railway was efficient means which led to welfare of the region.

In recent years, however, construction of heavy industry was not welcome because values added by heavy industry is lower and pollution problem became more serious. Because of this reason, in many developed countries effectiveness of traditional means for regional development became less and less.

Other problems like care of senior citizen, improvement of K-12 education, improvement of medical services, congestion of urban automobile traffic, etc, are less and less effectively solved by traditional policy means.

For improvement of medical services, variety of highly specialized medical doctors are needed. To have all these variety of medical doctors in one hospital could be too costly. Therefore, number of such hospital is limited and it becomes difficult to prevail such services to all people. Conventional solution for improvement of such problem will need infinite investment.

If a system by which highly specialized doctors are accessible through a broadband network and if proper diagnostics is possible through high resolution picture and other instruments which can be used through telecommunication, high level medical services are available for majority of people. This is an example of social problem solving in information era taking advantage of information technology. In some particular field of diagnostics such service have been proved to be useful by several experiments.

Followings are also important social issues which could be solved by the next generation information infrastructure.

1. Support of aged population
2. Over concentration of population in metropolitan area
3. Change structure of economy

Issues to be solved by the N I I are real social issues which were difficult to be solved by traditional policy means. Issues are different country by country, however, N I I could be an efficient means to solve these problems.

3. Layered Structure of Next Generation Information Infrastructure

3-1 Implementation Aspects of Next Generation Information Infrastructure

Although such services for social problem solving are not possible without broadband telecommunication networks, broadband network infrastructure itself has no meaning in the sense. Variety of social services must be developed using broadband network infrastructure.

Figure 1(a) shows implementation aspect of broadband network infrastructure. Layer 1 is the network infrastructure including optical fiber local access system, broadband switches and broadband long line transmission facilities. To use the network infrastructure, central service facilities and end user terminals for broadband application must be equipped. This is layer 2. Variety layer 2 facilities must be developed depending on needed applications.

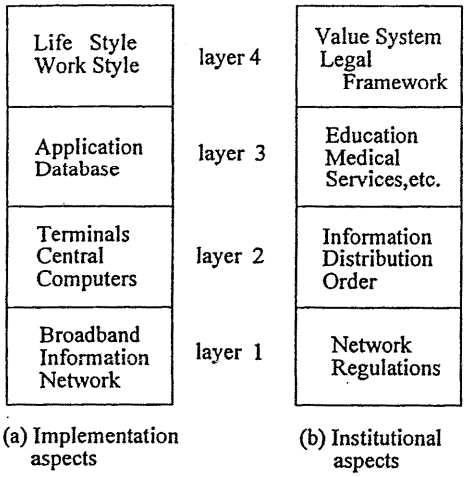


Fig.1 Four layered Architecture of the Next Information Generation Infrastructure

Using layer 2 facilities, variety of services including entertainment, education, health services, shopping services, telecommuting etc can be developed. Softwares for the purpose as well as databases should be developed. Variety of contents for entertainment are also needed. These are layer 3 implementations.

Technical contents of information network can thus be understood by the 3 layer architecture. Layer 1 end-to-end optical communication network can be constructed by economical reason by CATV service providers as well as telecommunication carriers, however, variety of investments for layers 2 and 3 are needed by various service providers as new businesses or businesses which replace present service styles. Some of investment including K-12 education and health care should be done by government. The motivation of the government to invest to create new service is to improve governmental services without increasing burden to people. For example, health care program can be improved using broadband networks extensively, much less costly than increasing hospitals and medical specialists.

However, even if the service level is improved by broadband telecommunication facilities, if people is not satisfied by the service, the network infrastructure will not be appreciated. It may happen that if the symptom of a patient is determined to be not very serious by telemedecine, the patient may not believe the result of diagnostics by telemedecine. In tele-education, also attitude of students must be adjusted to new education style. The fourth layer of the implementation of next generation information infrastructure is people's life style and work style suited for information infrastructure based services.

3-2 Institutional Aspects of Next Generation Information Infrastructure

Figure 1(b) shows another aspects of next generation information infrastructure. Present day laws and ordinances do not suppose information network based services. For example medical treatments using network are not permitted by law in many countries. There can be similar obstacles in variety of services including education and social welfare. These laws and ordinances should be modified to permit new way of services. These are layer 3 problems.

In layer 2, new social orders are needed related to multi-media information distribution. The issue includes intellectual property right and security issues. Intellectual property right is important to encourage creative artists and scientists as well as businesses to supply a plenty of informations. At the same time intellectual property right must be flexible enough to permit to create various new information products taking advantage of existing informations.

In layer 1, to encourage communication carriers to construct next generation information networks including optical fiber subscriber line, regulations related with convergence of telecommunication and broadcasting changed recently in many countries.

In layer 4, institutional problem is more complex. Change of life style and work style depend on value systems of people and organization. Basic legal framework is also related to layer 4 of the institutional aspects. As an example, if office workers can efficiently work by means of telecommuting, if their director cannot appreciate their work, telecommuting will not be accepted by workers. Employee evaluation system of private company is to be changed to prevail telecommuting based on next generation network infrastructure.

4. Interoperability Requirements for N I I

4-1 Layered Structure of N I I Standard

To realize interoperability in N I I, standardization in all the four layers are needed. Figure 2 shows the four layer structure from the view point of standard. Layer 1 is network standard which guarantee network connectivity. In layer 2 various standards related to computer hardware software and data are included. Layer 3 is related to various contents related standards.

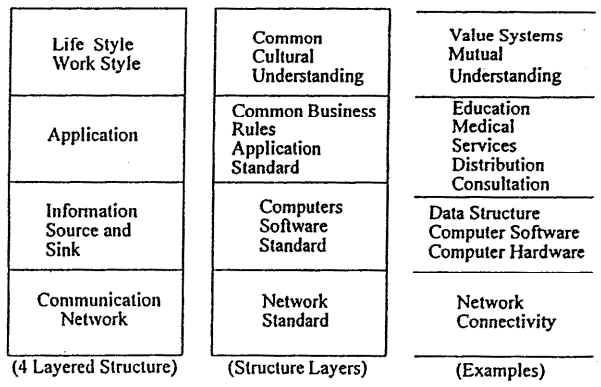


Fig.2 Layers fo NII/GII Standardization

In medical application, for example, measurement standard for medical data is important. In addition, business convention on medical services are also to be based common understanding to realize smooth network services. In layer 4 mutual common understanding of people is the key.

In the layered architecture of standard, some have been subjects of conventional dejure standard process. Some have been subjects of defacto standard, however, some have never been subjects of standardization.

In all the layers of standards, to realize N I I, fast developments of standards are requested. However, variety of factors including extension of scope of standards, software related standards, quick movement of products made standardization more and more difficult.

4-2 Structure of standard

Figure 3 is recent structure for standardization. In telecommunication systems, variety of standards are discussed based on the OSI reference model. Majority of ITU-T standards for digital systems are based on the reference model. The OSI reference model have 7 layer structure and the basic principle in each layer has also been standardized. These are the second level standards.

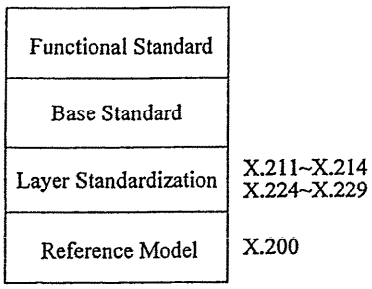


Fig.3 Hierarchy of standards

Actual ITU-T standards are arranged based on the standard of each layer. Standards for N-ISDN are agreed based on layer by layer standards. These are the third level standards. The third level standards are called base standards. Majority of ITU-T or ISO standards for telecommunication are the third level standards.

In past, actual telecommunication products are produced based on the base standards. Base standards, however, have many possibility of variations caused by options and parameters. Therefore, if different options or parameters are adopted by different manufacturers, the products can not be interoperable.

In telecommunication product, one major reason of incompatibility is differences in parameter and option selection. The fourth level standard called functional standard is needed to solve the problem. In ISO (International Organization for Standardization) standard procedure, the functional standard is a recent important subject for standardization. In telecommunication field, the issue to realize interoperable system have been solved by the telecommunication authorities of each country. Because of this reason, agreement to internationally harmonize functional standards is far from realization.

Therefore, we frequently have difficulty to interconnect systems purchased from different countries.

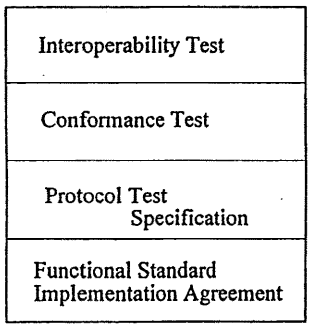


Fig.4 Framework for interoperability realization by ISO

Even if, products are constructed based on a functional standard, interoperability is not automatically guaranteed. Figure 4 shows a frame work for interoperability realization by ISO. Once functional standard is established, a test specification to test products is developed. Test specification is also to be internationally harmonized. Based on internationally standardized test specification, protocol testers can be constructed. The protocol tester is used to test the conformance of each product to the functional standard. Finally the interpretability of products must be guaranteed by interpretability test.

4-3 Interoperability of Network

Figure 5 shows a simplified diagram of networks including terminals. Networks A and B may be networks of different countries. End systems A and B connected to networks A and B respectively want to be interconnected. End systems have seven layers of protocols and networks have lower 3 layers of protocols.

From the view point of terminal users, the first requirement is to realize communication between end systems. If upper layer protocols of end systems are interoperable, end to end communication may be realized. However, in many cases, two networks A and B may have different lower layer protocols, and end systems A and B are requested to have different protocol implementations.

Equipments developed by user initiative are subjects of international trade. In some occasions, a terminal which has no problem in the network of one country may not be used through the network of another country, by incompatibility of network to terminal interface. Conformance of network interface is needed for further development of network applications.

In past, majority of telecommunications are confined to domestic regions, however, since many people want to internationally travel bringing portable terminals, differences of lower layer specifications by countries become more serious problems especially in near by countries. From the view point of network users, conformance of lower layer protocol of network becomes more important request. However, we do not have any framework to test protocol conformance of networks at present. I expect this will be a more important issue in future.

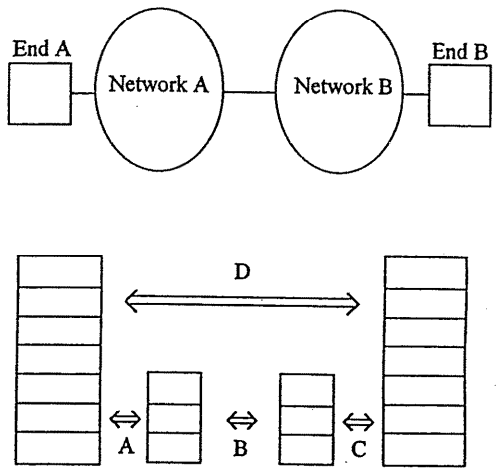


Fig. 5 Conformance test of end systems and networks

5. Software Related Standards

Concept of standardization was originally created for hardware related products. Especially in mechanical products standardization can be fitted into usual production processes. For example to produce a screw, specification of production is given as a drawings and figures included in the drawing. If the figures are standardized, identical screws can be produced in an integrated manner in usual industrial production.

In software production, specification is also needed, however, identical software function is not expected to be realized by specifying the products in usual manner. Much more detailed specification is needed to produce identical software function. In some cases, time needed to describe software by specification is more than time needed to implement the software based on specification. In many cases, specification is not clear enough to produce interoperable system. Lack of interoperability is frequently caused by this reason.

To guarantee interoperability costly development of test system is requested as shown in Figure 6. Since test procedure is not included in usual production, overhead of test tend to increase the cost of standard product.

To have identical software for interoperable environment, simple copy of program is enough. If implemented programs are released in source codes, programs can be fitted into any hardware without major difficulty.

In recent competition of IPS (internet protocol suit) and OSI, many specialists recognized better functionality in OSI. Although OSI is much superior to ISP since OSI implementation is not available easily, OSI was not successful in the competition against IPS.

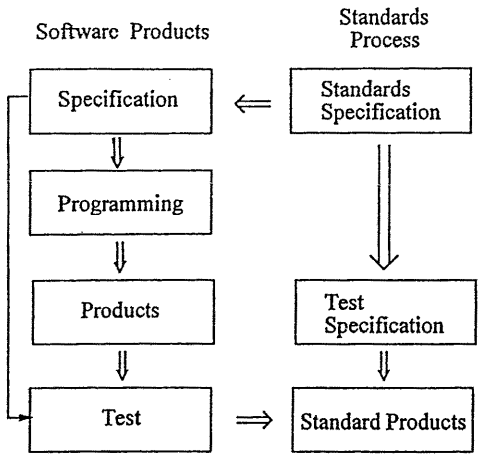


Fig.6 Software related standardization process

Specification of IPS is available in a series of documents called RFC (request for comment). However, in usual case, systems using IPS are installed by copying appropriate software implementation available essentially free of charge. Because of this reason, specification is not requested to be detailed enough. Costly test procedure is not needed. The comparison seems to give an important suggestion for successful software related standardization.

6. Realization of Reference Implementation

Reference implementation can be realized simultaneously or sequentially with standard documentation.

6 -1 Standardization Accompanied by Reference Implementation

To make reference implementation available simultaneously with standardization by document, implementation activity must accompany with standardization activity. Figure 7 shows the outline of flow of standardization.

When a new work item is proposed, the scope and the objective of the work item must be clearly defined. Then the environment and the platform for reference implementation are defined. The definition includes software platform like an operating system and a language. After the platform is defined, contribution for standard document is to include the code to implement the specification proposed. The code is computer executable under the agreed software platform. The editor group of the standard must have a function to debug and test the code which was realized by combining contributions. The debugging must be made under the close cooperation with contributors. Some project team type activity may be needed to test and debug a large scale implementation. Specification is to be discussed in parallel with reference implementation development. In this case standard document and standard code (reference implementation) will be available simultaneously.

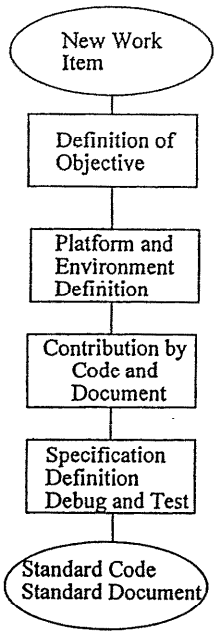


Fig.7 Flow of Parallel Development of Document and Code

6 -2 Sequential Realization of Reference Implementation

Although simultaneous realization of reference implementation is preferable, to provide a reference implementation in existing standard, realization must be sequential. Figure 8 shows the flow for sequential realization. In this case implementation is to be made after functional standardization. Since a plurality of implementations may be realized by different bodies, the implementation must be authorized by international agreement. To escape from a situation where a plurality of different implementation exist having slight differences international division of development effort in recommended.

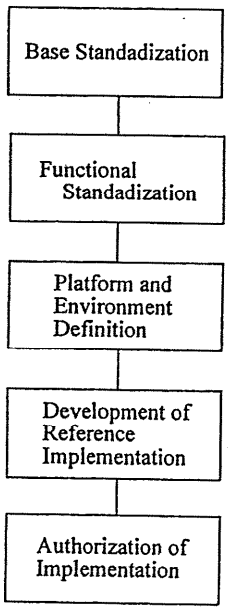


Fig.8 Flow for Sequential Realization of Reference Implementation

7. Product Realization

The standard reference implementation can be used in variety of ways. Commercial software may be realized just copying the reference code, however in majority of cases, adaptation of code to fit the code in more efficient commercial product is recommended. Optimization is usually needed to realize commercial product. Even if reference code is free of charge, manufacturers can compete on the efficiency and performance of the optimized product. Interoperability of the commercial product can be tested using the standard reference implementation. The usage of reference implementation is shown in Figure 9.

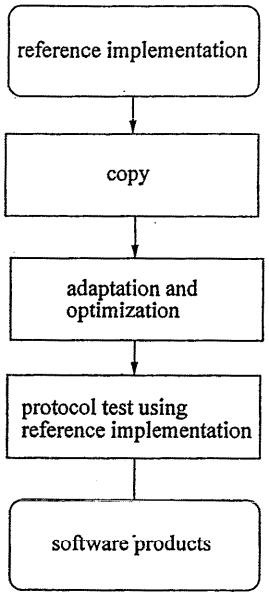


Fig.9 Production of Commercial Product using Reference Implementation

8. Efficient and Fair Standardization

To realize N I I /G I I , efficient quick standardization in needed. For effective standardization, market driven mechanism is to be respected.

To smooth implementation, a reference implementation of standardization is advantageous. If the reference implementation is efficient enough to be used in actual product, smooth proliferation of standard will be accelerated.

Considering the expansion of scope of standardization, in upper layers of standards, many defacto standards are expected to be included in N I I /G I I standardization.

Defacto standard, is especially suited for software related standardization. In implementation, defacto standard is distributed by common software implementation. Interoperability test is not needed. Specification can be formulated quickly and market related. Some aspect of N I I/G I I standardization will be realized through defacto standardization process. If defacto standard is open enough, it will be useful especially in higher layers. However, to realize fair standard, what is " open " is to be defined in more detail. Scope of openness could range:

open dejure standard
open available specification
open available software implementation
open interconnection

Some open systems are proprietary. To guarantee fairness and quickness in standardization, N I I /G I I standard is to be combination of defacto and dejure standards.

However, if many different defacto standards are used competitively, it will harm interoperability.

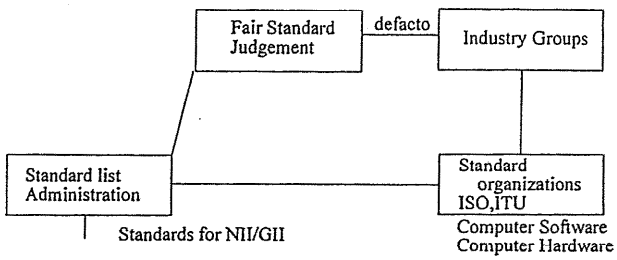


Fig.10 A possible framework for NII-GII standardization

Figure 10 shows an possible framework for combined defacto / dejure standardization. A list of standardization on G I I item is to be maintained by an international body. For national matters related to language and cultural difference, standard list administration body may be needed at each country. Dejure standard should first be included in the list. Additionally defacto standard from industry groups can be included after evaluation to guarantee fairness and openness.

9. Conclusion

To convey technical achievements to real communication service quickly, efficient and fair standardization process is requested. Modification to traditional standardization frameworks and methodology are requested from the view point of efficiency. Without the effort, effectiveness and fairness of traditional dejure standardization will be lost, especially in software related standardization. This presentation described problems of traditional standard frameworks and proposed an idea to improve the situation. Future looking discussion in a new standardization framework is essential for realization of next generation information infrastructure.

References

(1) Reforms toward the Intellectually Creative Society of the 21st Century : Ministry of Post and Telecommunication of Japanese Government (May 31, 1994)

(2) T. Saito : An Evolving Scenario of Communication Network towards B-ISDN:IBCN&S Conference Proceedings,Copenhagen, Denmark (April 19-23, 1993)

(3) T. Saito : Regional Activities for Telecommunication Standardization, APT meeting Doc. No. 11 (May 1994)

(4) T. Saito : Next Generation Information Infrastructure Plan in Japan , Proceedings of SEACOM '94 pp.153 - 162 (Oct. 1994)

(5) A. Maisel : A Model for a Barrier Free National Information Inftastructure, Sun - Microsystems Inc. (March 15, 1994)

Biographies

Tadao Saito received the Ph.D degree in electronics from the University of Tokyo in 1968. Since 1986 he is a professor of Information and Communication Engineering Department of the University of Tokyo. He is interested in digital communication and computer networks especially in next generation communication networks and services. He wrote two books on electronics circuitry, four books on computer and two books on digital communication and multimedia. He received the paper prize in 1967 and 1986, and the achievement prize in 1972 from Institute of Electronics, Information and Communication Engineers of Japan (IECEJ). He is a member of IEEE, IECEJ and Information Processing Society of Japan.

Meeting on Regional Co-operation in Standardization
25 - 27 May, 1994 : Phuket, Thailand

Doc.No. STANDARD-11

REGIONAL ACTIVITIES FOR TELECOMMUNICATION STANDARDIZATION

by

Dr. Tadao Saito
Professor
The University of Tokyo
Japan

Organized by
Asia-Pacific Telecommunity

Regional Activities For Telecommunication Standardization

Tadao Saito (The University of Tokyo)

Abstract

Due to recent developments of high performance electronics, telecommunication systems are more and more complex and performances are more and more improved. Because of complexity of systems, standardization process became more complex, however, functions which was very costly in past can be realized by reasonable cost. In past because of limit of technology, functions of telecommunication system were tend to be limited by manufacturing technology. In recent years to realize user friendly systems, participation of users in standardization processes became more important, because systems can be designed without limit of technology. In Asian area, variety of character sets used for communication systems is one of common issues to be realized by telecommunication systems. Social similarity of each country may request standards suited for the society To understand and solve common problems related with standardization, regional activities are expected to be important in future. Exploration to establish a frame work for regional standardization activity is essential and timely effort.

I. INTRODUCTION

Recent telecommunication technologies are characterized by digitalization. By employing digital technologies, more and more advanced functions are realized by telecommunication systems. As an example, in past telecommunication systems can handle only roman alphabetic character set. In recent years variety of Asian character sets are handled by telecommunication systems. This is technical freedom given by technical improvements.

Standardization is a key to realize complex system with reasonable cost. To provide country by country variations in character set, a common standardized framework is needed. It may also be possible to realize systems suited for society by setting standards based on social requirement.

Development of technology made it possible to provide user friendly environment, however, to realize truly good systems, participation and commitment of all related bodies are important.

Geographically near countries generally have common close cultural background cultivated by history. Asian countries potentially have common issues related with standardization. It is well known that European countries have common cultural background. Regional standardization activities of European countries has history of success to realize user friendly systems in European counties.

The technical freedom given by technical improvement gives us a new issues for standardization.

II. STRUCTURE OF STANDARDS

Standard is the essential tool to realize interoperability, however, we have many experiences to find products which are constructed based on one international standard are not interoperable. Since products become the more complex, such experiences become the more frequent.

Figure 1 is recent structure for standardization. In telecommunication systems, variety of standards are discussed based on the OSI reference model. Majority of ITU-T standards for digital systems are based on the reference model. The OSI reference model have 7 layer structure and the basic principle in each layer is also standardized. These are the second level standards.

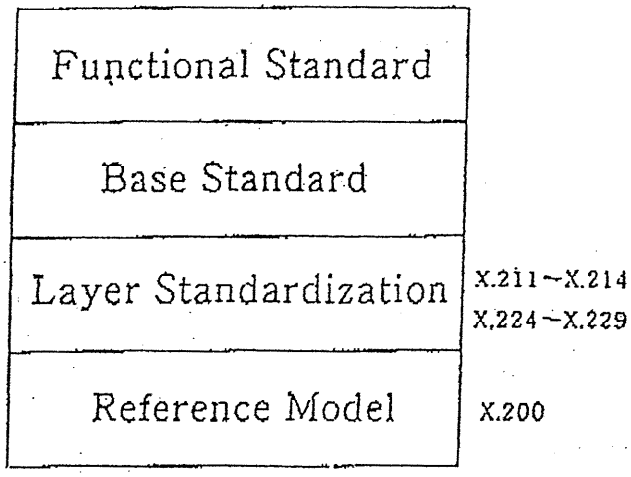


Fig. 1 Hierarchy of standards

Actual ITU-T standards are arranged based on the standard of each layer. Standards for N-ISDN are agreed based on layer by layer standards. These are the third level standards. The third level standards are called base standards. Majority of ITU-T or ISO standards for telecommunication are the third level standards.

In past, actual telecommunication products are produced based on the base standards. Base standards, however, have many possibility of variations caused by options and parameters. Therefore, if different options or parameters are adopted by different manufacturers, the products can not be interoperable.

In telecommunication product, one major reason of incompatibility is differences in parameter and option selection. The fourth level standard called functional standard is needed to solve the problem. In ISO (International Organization for Standardization) standard procedure, the functional standard is a recent important subject for standardization. In telecommunication field, the issue to realize interoperable system have been solved by the telecommunication authorities of each country. Because of this reason, agreement to internationally harmonize functional standards is far from realization.

Therefore, we frequently have difficulty to interconnect systems purchased from different countries.

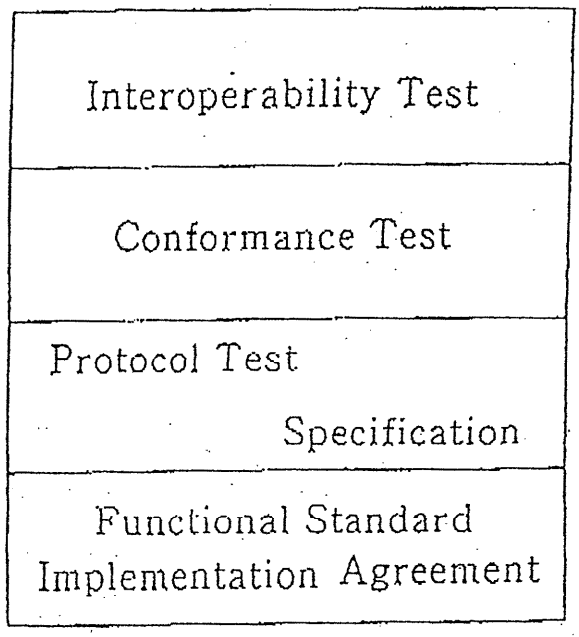


Fig.2 Framework for interoperability realization by ISO

Even if, products are constructed based on a functional standard, interoperability is not automatically guaranteed. Figure 2 shows a frame work for interoperability realization by ISO. Once functional standard is established, a test specification to test products is developed. Test specification is also to be internationally harmonized. Based on internationally-standardized test specification, protocol testers can be constructed. The protocol tester is used to test the conformance to each product to the functional standard. Finally the interpretability of products must be guaranteed by interpretability test.

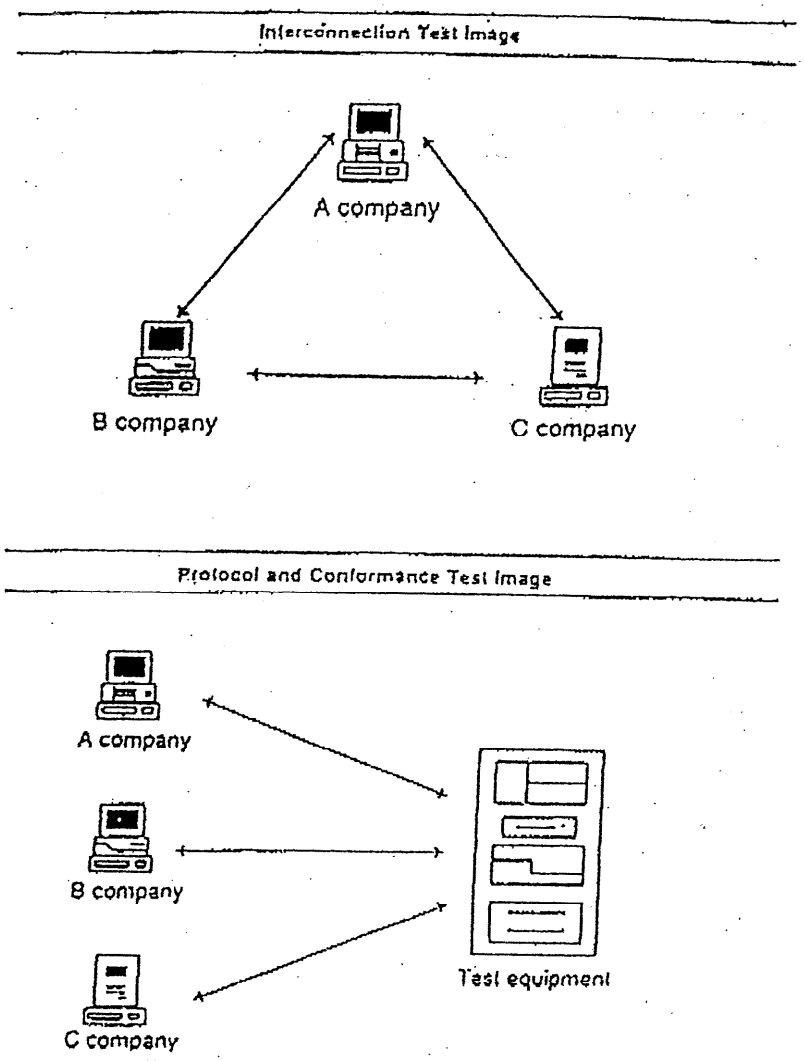


Fig.3 Concepts for interconnection test and conformance test

For some type of products like telecommunication terminals, interoperability test which test interconnectivity by direct interconnection is possible. Figure 3 shows simplified schematical diagram of the interconnection test and the conformance test. HATS (harmonization of advanced telecommunication system) activity chaired by myself is an successful example of interconnection test. Although all the product to be tested must be collected at the same time to test locations, the interconnection test is a practical way to guarantee interpretability.

III. SCOPES OF STANDARDIZATION

Figure 4 shows a simplified diagram of networks including terminals. Networks A and B may be networks of different countries. End systems A and B connected to

networks A and B respectively want to be interconnected. End systems have seven layers of protocols and networks have lower 3 layers of protocols.

From the view point of terminal users, the first requirement is to realize communication between end systems. If upper layer protocols of end systems are interoperable, end to end communication may be realized. However, in many cases, two networks A and B may have different lower layer protocols, and end systems A and B are requested to have different protocol implementations.

Equipments developed by user initiative are subjects of international trade. In some occasions, a terminal which has no problem in the network of one country may not be used through the network of another country, by incompatibility of network to terminal interface. Conformance of network interface is needed for further development of network applications.

In past, majority of telecommunications are confined to domestic regions, however, since many people want to internationally travel bringing portable terminals, differences of lower layer specifications by countries become more serious problems especially in near by countries. From the view point of network users, conformance of lower layer protocol of network becomes more important request. However, we do not have any framework to test protocol conformance of networks at present. I expect this will be a more important issue in future.

Differences in functional standard or implementation specification in networks by countries also cause problems in procurement of products by network operators. If common specification can be agreed by telecommunication carriers of Asian region, procurement procedure could be much simpler.

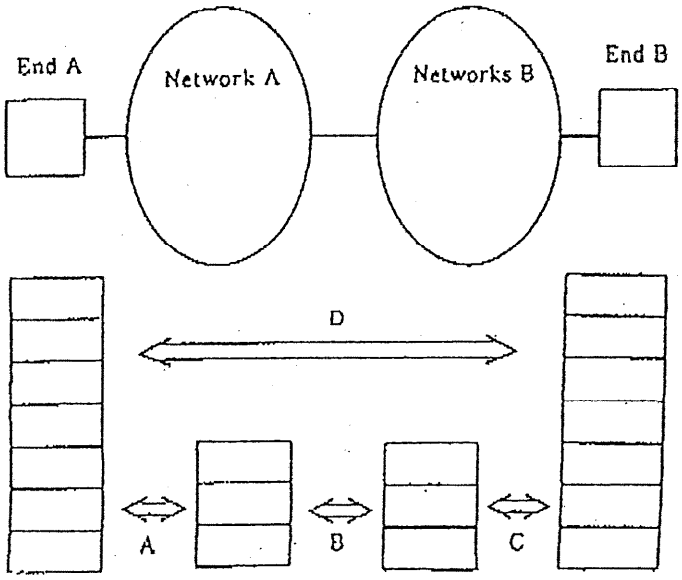


Fig.4 Conformance test of end systems and networks

IV. COMPULSORY STANDARD AND VOLUNTARY STANDARD

To interconnect a terminal to the telecommunication network, the terminal is requested to conform to the standard of the network. Some kind of certification is needed for this purpose. In past, all the terminals connected to the network are owned by the particular network operator. This scheme is possible when request for network is simple. In recent years, since the variation of requirement from network user is so complex that variety of terminals owned by users are connected to the network. Therefore, it is more and more difficult to guarantee complete interoperability between terminals by testing terminals to be connected to the network from the network side. Even in this case, terminals connected to the network should not give adverse effect to the network or other communications. Certification of terminals to permit network connection is now confined to minimum requirement which related to protection of networks and other communications. Standards related with this minimum requirement is called compulsory

standard. By separating compulsory and voluntary standards, it became possible to use variety of terminals designed by network users. Recently promotive force of development of advanced application of network tend to come from terminals developed by network users. To complete communication, another level of standard is needed. The standard is called voluntary standard in contrast to compulsory standard.

V. STANDARDIZATION PROCESS

In recent communication systems, because of increased functionality and complexity, standardization is more essential than past. Freedom of design was also increased by information technology. To guarantee communication, unification of design by standardization is essential.

At the same time, cost and time needed for standardization are increasing. If a product is once-developed, it is more difficult to change design. Therefore, standardization is requested to be completed before the first product. Standardization activity for future product is important.

It is also important to guarantee interoperability by standardization. Therefore, functional standardization is more stressed rather than base standardization. In functional standardization, some experimental products to test feasibility is needed. Therefore work items near to products are selected for functional standardization. In the standardization framework of ISO, functional standards are not drafted by ISO but drafted by regional workshops composed mainly from industries. The following 3 workshops:

EWOS : European Workshops for Open Systems
OIW : OSE Implementers Workshop
AOW : Asia Oceania Workshops

are main contributors to draft functional standards of OSI/OSE subjects. These three workshops works in downstream standardization. Also, close cooperation is ongoing in protocol test suit development among 3 workshops.

Workshops are mainly composed of industry sectors. Work items are selected basically based on industry interest. Industry can only afford expenses based on near future sales of products. The selection by industry is important to avoid standardization just for standardization.

In this type of standardization activity, regional cooperation is assumed as the framework of activity. This framework is especially useful when character set problems are handled.

In the standardization activity of ITU, Global Standardization Collaboration Group (GSC) started recently. It will be useful for Asian area to find common interest for standardization and to cooperate to help each other to contribute for worldwide standardization.

VI. DEFACT STANDARDS

A trend featuring recent standardization is success of defact standards. Especially in complex computer based system, OSI based systems and IPS (Internet Protocol Suite) based systems competed. In spite of tremendous efforts on OSI, near future win of IPS is expected. Dejure standards have difficulties in many aspects.

The main reason of success of IPS is the fact that IPS is distributed not in the form of specification but in the form of software implementation. In the latter case interoperability is automatically guaranteed and complex procedures as shown Figures 1 and 2 are not required. Interoperable system can be realized easily.

Standard is a essential way to realize compatibility in manufacturing industry. To get identical screws to fix doors to poles, standard design of screws is needed. This is because, industrial product cannot be copied with no cost.

In the field of software, to produce compatible softwares based on standard written by natural language is not efficient way. Software product can be copied without any cost

by proper licensing. To realize interoperable system distribution of implemented software is much more efficient.

To realize effective dejure standard, reason of success of defact standard must be carefully studied.

VII. CONCLUSION

Recent trends of standardization is very much influenced by complexity of system caused by intelligence and digitalization of modern electronics. Freedom of design obtained by modern electronics made it possible to design a system more suited for local culture represented by character sets or social situations. It is also possible to design a system suited for development status of each area. Although we have variety of potentials, actual realization is only possible through standardization. Countries having cultural and social similarity are requested to cooperate each other to have standards to reflect the needs of the society.

Standardization process is complex, however, standardization is the only means to effectively realize telecommunication networks. I hope every entities in APT region watches the new trends in standardization and find a way to cooperate to satisfy regional requests for standardization

Document Name:

Regional Activities For Telecommunication Standardization

Speaker:

Tadao SAITO

Professor, The University of Tokyo

Profile:

Engineering Dr..

Leading person concerning the telecommunications technology in Japan.

Member of Telecommunications Technology Council of the MPT(, which determines the telecommunications technology principle and guidelines in Japan).

Chairman of Standardization policy committee in the Council(, which determines Japanese standardization guidelines).