

A DISTRIBUTED ROUTING ALGORITHM IN CLOS NETWORK OF VARIABLE BIT RATE TDM SWITCH

Hiroaki Morino, Hitoshi Aida, Tadao Saito

Department of Information and Communication Engineering

Faculty of Engineering, University of Tokyo

7-3-1 Hongo Bunkyo-ku Tokyo 113-8656 Japan

Tel: +81-3-3814-4251 ext. 6762

Key words: Variable bit rate, TDM switch, Video traffic, Clos network

Abstract: In this paper, a new routing method is proposed in Clos network of variable bit rate TDM switch. This method is dynamic routing method, and it has following two features; (1) In a TDM frame period, all timeslots applied to the switch can reach their destination ports without loss due to internal blocking. (2) Buffer for queuing is not required in switches of the first stage. Performance of the method is evaluated by software simulation and it is indicated that it can be controlled by commercially available microprocessor in real-time when size of Clos network is within 16×16 .

1. INTRODUCTION

In recent years, there has been an increasing demand for video communication, including video conferencing and video-on-demand system. With advancement of real-time video encoding technology, cost effective variable-bit-rate (VBR) transmission will have widespread use in the future.

As a suitable architecture for variable bit rate video switching system in the future, we have already proposed variable bit rate TDM video switch (VTDM video switch) architecture [1]. The fundamental characteristics of the switch is that end-to-end delay jitter of timeslot is constant regardless of

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35581-8_35](https://doi.org/10.1007/978-0-387-35581-8_35)

offered load. Moreover, large timeslot size and intelligent priority control enables both of efficient transmission and guarantee of traffic quality for variable-bit-rate video traffic.

Today, bit rate of encoded video of standard TV-quality is several Mbit/sec, and to cope with large amount of video connections in backbone network, required switch capacity is estimated at least at Tbit/sec. For design of switch fabric of this capacity, introduction of MIN (Multistage Interconnection Network) is indispensable.

As an appropriate type of MIN for connection oriented video traffic, Delta network[2], Benes network[3], and Clos network[4] are candidates. To maintain switch performance for large size switch, it will be desirable that switch elements have buffer within them and that arbitration of routing of timeslots are done in each switch element independently. In addition, number of stages should be small to minimize switching delay. Though Delta network has advantage among three configurations in these viewpoints, speed up of internal links is necessary to avoid internal blocking at high load rate in this network, and it will cause considerable increase of hardware. In this paper, we focus Clos network as an alternative configuration, and routing method to avoid internal blocking is proposed.

The most important feature of VTDM switch is timeslots in the same input TDM frame period are transmitted in the same output TDM frame period. Therefore, for routing method of Clos network, it will be required that all of timeslots in one input TDM frame period arrive at their destination output ports without internal blocking within corresponding output TDM frame period.

Routing methods of Clos network are classified into two groups; static routing and dynamic routing. In static routing, required characteristics for VTDM switch described above can be easily achieved if appropriate connection admission control and traffic shaping is adopted. On the other hand, in dynamic routing, there is also an existing scheme [6] to meet the demand which is designed for ATM. This method works in framed ATM switching system [7], which has similar switching principle to VTDM switch architecture, it can be simply applied to a VTDM switch.

For both of these types of routing, internal buffer is necessary in a switch element of the first stage since timeslots that are routed to the same middle stage switch should be queued to avoid blocking. Therefore total switching delay of Clos network is 1.5 times larger than that of Delta network, which is not appropriate for real-time traffic.

If routing method is devised so that only one timeslot is routed to each switch of middle stage at the same time, no buffer will be needed in switches of the first stage, and then switching delay of Clos network will be the same as Delta network.

In this paper, a new routing method in Clos network of VTDM switch to satisfy this condition is proposed. Since the proposed method is devised so that only one timeslots is routed to each switch of middle stage at the same time, and internal buffer in switches of the first stage is not required.

In the rest of this paper, operation of the proposed method is described and feasibility of the method is discussed by evaluation of processing time.

2. VARIABLE BIT RATE TDM SWITCH ARCHITECTURE

In this section, features of VTDM switch architecture are described as background of the research.

2.1 Overview of video-specific network

Figure 1 shows overview of video-specific network located in multimedia backbone network where VTDM switch architecture is applied. In this configuration, backbone network consists of traffic-specific networks for voice, video, and data communication and access network accommodates various types of traffic by integrated TDM or WDM system and provides a unified network interface to users.

2.2 Frame configuration

VTDM frame format is designed as shown in Figure 2, assuming 622 Mbit/sec transmission. To enable flexible time slot assignment for variable bit rate video traffic, header information is added at the top of the frame. Frame header has information for each time slot about connection number, priority level, and tolerable delay etc.

To perform efficient priority control for video traffic, length of VTDM frame is 1/300 (sec), which is a common divisor of frame time interval of NTSC video signal 1/30 (sec) and PAL video signal 1/25(sec).

By one timeslot is assigned for one connection for every TDM frame, the connection gets bandwidth of $(16 \text{ (kbits)} * 300 \text{ (frame/sec)}) = 4.8\text{Mbps}$

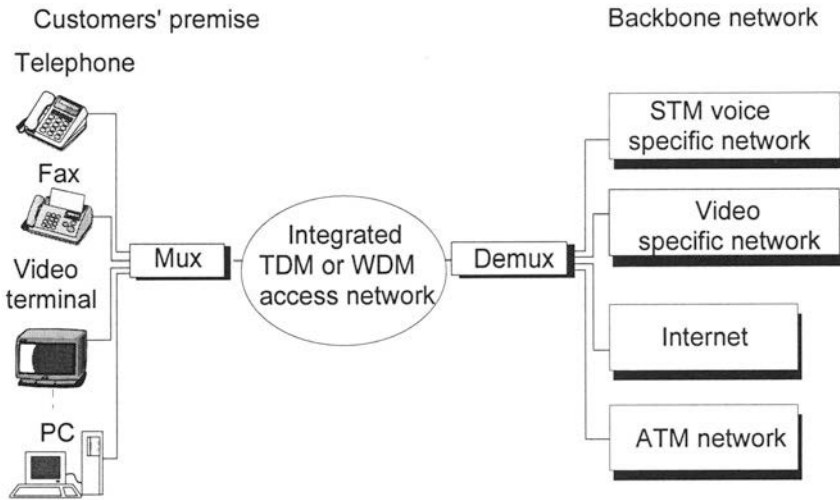


Figure 1. Overview of multimedia network system consisting of traffic-specific networks

2.3 Switch architecture and operation

There are various types in VTDM switch configuration. Figure 3 shows crosspoint buffer type configuration. At each crosspoint, a buffer and an address filter (AF) are provided. An address filter reads frame header of input TDM frame and timeslots to be routed in corresponding output port is stored in the crosspoint buffer. At each output port, timeslot scheduler and content analyzer are provided and these perform priority control according to video data contents of timeslots. When all timeslots are stored in crosspoint buffers, output TDM frame is generated by frame generator, and timeslots are assigned in the frame and sent to the next switch stage.

To guarantee timeslot switching delay, each crosspoint buffer has double buffer structure shown in Figure 4, where the length of each buffer is the same as TDM frame length. all timeslots of input TDM frame are stored in buffer (a) or buffer (b), and they are transmitted in the following TDM frame. Timeslots that cannot be transmitted at that time are discarded inevitably.

In this operation, once several timeslots are assigned in the same TDM frame at the access network of source terminal, these always arrive at a receive terminal in the same TDM frame as shown in Figure 5. That is, end-to-end delay jitter is bounded within TDM frame.

Denoting TDM frame length by T , and delay jitter by D , the following condition is satisfied.

$-T < D < T$

This delay jitter boundary is constant regardless of offered load to a switch, and it is effective for switching real time traffic including video traffic.

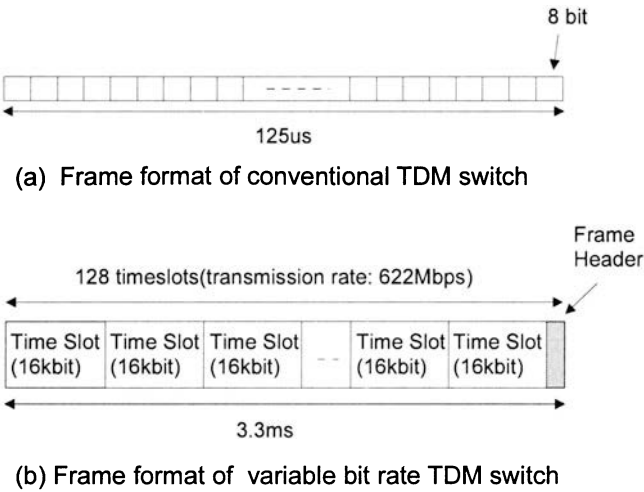


Figure 2. VTDM frame format

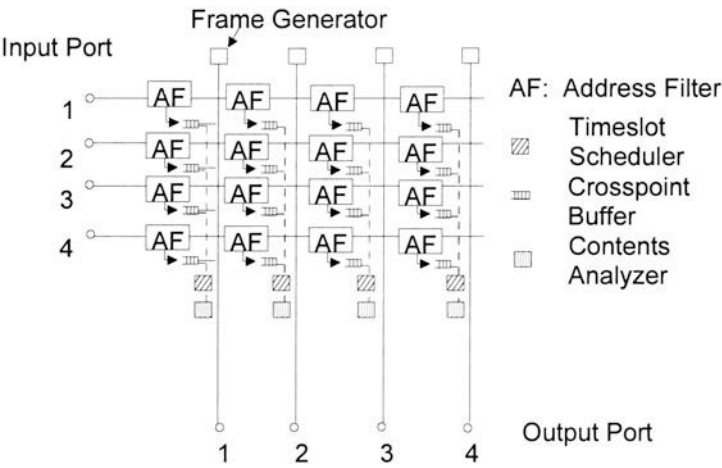


Figure 3. Crosspoint buffer type switch

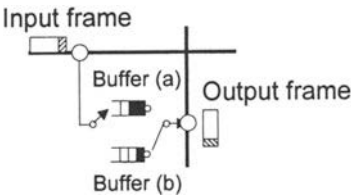


Figure 4. Double buffer structure

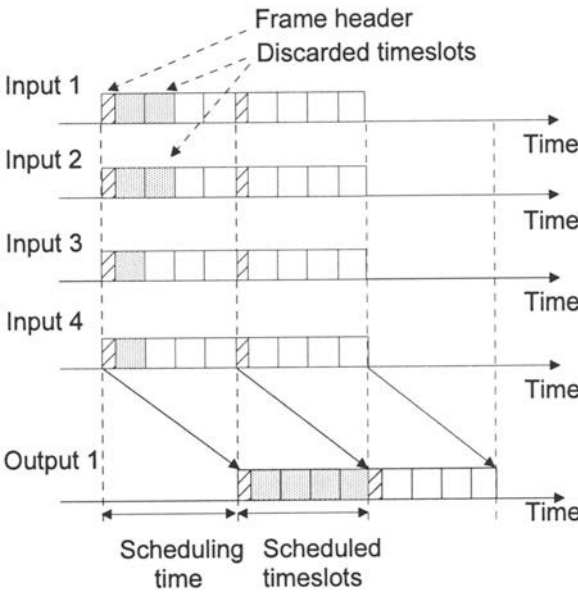


Figure 5. Timechart of switch operation

3. PROPOSAL OF A NEW ROUTING METHOD OF CLOS NETWORK

According to reasons described in Section 1, we focus Clos type network as a target to be examined, and design of large scale switch of VTDM switch using multistage interconnection network is examined.

3.1 Target configuration

In this paper, a configuration in Figure 6 is examined, where number of row and size of switch element is the same. To realize switch fabric whose capacity is Tbit/sec by VTDM switch where port speed is 622 Mbps, size of

switch must be at least 4096 x4096. Corresponding with this size, size of switch element must be at least 64 x 64.

3.2 A proposed routing method

In Clos network of VTDM switch shown in Figure 6, there is possibility that some timeslots are applied to one switch of the first stage and they are destined to one switch of the third stage. An aim of the proposed routing algorithm is that internal paths for these timeslots are distributed among all switches of the second stage to avoid internal blocking. To realize this characteristics, each switch of the first stage has a middle stage switch assignment matrix, and records results of middle stage switch assignment for timeslots in the past time. For example, usage of a matrix of the switch 11 is explained in Figure 7. Each column of this matrix corresponds with a switch of the third stage to which an input timeslot goes. On the other hand, each row corresponds with middle stage switch to which the timeslot is routed. The way to give numbers to switch elements is as shown in Figure 6. For example, if a timeslot is going to the third stage 31, and this timeslot is routed to middle stage switch 21, then the switch 11 increases a value of an element of (1,1) of the matrix by one.

When a new timeslot is applied to this switch, the switch refers all elements of a column that correspond with a switch of the third stage to which a timeslot goes. Then a middle stage switch to be routed . i.e. an output port from the first stage switch is assigned so that all values of elements of this column are statistically equal to achieve paths distribution.

Here an operation of the method is described for an example of timeslot pattern to the switch 11 shown in Figure 8. It is assumed that middle stage switch assignment for timeslots in the first timeslot position number is already finished and we focus on processes for timeslots in the second timeslot position number. At this time, initial status of the matrix before assignment is done is as shown in Figure 9 (a).

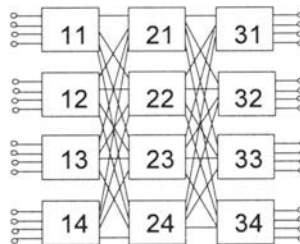


Figure 6. A configuration of Clos network examined in this paper

		1	2	3	4	Column number
		31	32	33	34	Corresponding number of switch of the third stage
1	21	0	0	0	0	
2	22	0	0	0	0	
3	23	0	0	0	0	
4	24	0	0	0	0	

Row number	Corresponding number of routed middle stage switch
---------------	---

Figure 7. A middle stage switch assignment matrix of the switch 11

- Phase 1 Selection of candidates of middle stage switch to be assigned

When a new timeslot is applied in a switch element of the first stage, the switch element refers elements of one column of a middle stage switch assignment matrix. This column corresponds with a switch of the third stage to which a timeslot goes. By comparing values of the elements of the column, rows in which minimum value are written are searched. Then middle stage switches that correspond with these rows are selected as candidates to be assigned to the timeslot.

For example, in case that candidates are selected for a timeslot whose ID number is (1,2), the second column of the matrix of Figure 9(a) is referred since the timeslot goes to switch 32.

- Phase 2 Assignment of middle stage switch

a) Step 1

Among all of applied timeslots to the switch, numbers of candidates of middle stage switch are compared with each other. Then a timeslot that has the least number of candidates is selected. In this case, two timeslots whose ID numbers are (1,2) and (2,2) have the least number of candidates. An order of assignment is determined as shown in the fourth column of Figure 9(b).

b) Step 2

According to the order determined in Step 1, a middle stage switch is assigned for timeslots.

c) Step 3

Going back to Step 1.

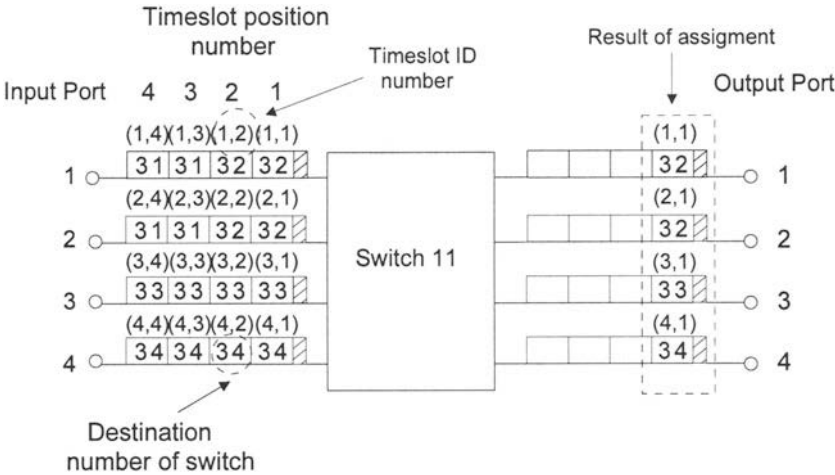


Figure 8. Timeslot input pattern to the switch 11

In this method, a timeslot that has fewer candidates of middle stage switches to be routed has priority about assignment of middle stage switch. In this way, probability that middle stage switch is assigned among candidates is improved, and efficiency of paths distribution is enhanced.

4. PERFORMANCE EVALUATION

In Clos network of VTDM switch, proposed middle stage switch assignment algorithm requires substantial amount of processing time, and a sequence of these processes must be finished within a VTDM frame interval. To demonstrate feasibility, estimation and quantitative evaluation of processing should be performed. To prepare for these evaluations, following notations are defined on Clos network of VTDM switch.

- N: number of port of Clos network
- n: number of port of a switch element of Clos network
- h: number of maximum timeslots in a TDM frame

Between N and n, we assume relationship of $N = n^2$.

	31	32	33	34
21	0	1	0	0
22	0	1	0	0
23	0	0	1	0
24	0	0	0	1

(a) State of the matrix of the switch 11

Timeslot ID number	Destination switch	Selected Candidates			Order of assignment	Assigned number of middle stage switch
(1,2)	32	23	24		1	23
(2,2)	32	23	24		2	24
(3,2)	33	21	22	24	3	21
(4,2)	34	21	22	23	4	22

(b) Assignment of middle stage switch

Figure 9. Middle stage switch assignment for timeslots in the second timeslot position

4.1 Estimation of order of amount of process

We define $C(nh)$ as amount of process of middle stage switch assignment for all timeslots applied in the switch in one TDM frame period. Moreover, we define $C(n)$ as amount of process for middle stage switch assignment for timeslots that belong to the particular timeslot position number. There are relationship of $C(nh) = hC(n)$. In this section, estimation is done for $C(n)$ in phase 1 and phase 2 respectively.

At first, In phase 1, following two processes are required to select candidates.

1. A process to select minimum value among elements of one row of the matrix.
2. A process to compare minimum value with other elements, and to select candidates for middle stage switches to be assigned.

For one of applied timeslots, both of two processes require processing time that is proportional to size of switch element. Therefore, when n timeslots are applied to a switch element, order of amount of process is calculated as $O(n^2)$.

In phase 2, sort process is also required to compare numbers of middle stage switch candidates of timeslots that are selected in phase 1. Whenever

middle stage switch is assigned to one timeslot, the sort process requires amount of process which has order of $O(n)$. Therefore, sum of amount of process of this sort for all timeslots in one timeslot position number is $O(n^2)$.

From these analysis, order of $C(n)$ is $O(n^2) + O(n^2) = O(n^2)$.

4.2 Quantitative evaluation

The amount of process of the proposed routing algorithm is also evaluated quantitatively by simulation. A routing algorithm is described by C language. A hardware platform to run simulation is Sun enterprise server (Processor: UltraSparcII processor, 300Mhz). Processing time of the algorithm is measured by internal function of UNIX named `gettimeofday()`.

The result is shown in Figure 10. Processing time is measured under condition that offered load is 100%. In the figure, both of measured values and approximation line are shown. Though this approximation line is polynomial function, it is expected to be approximated by linear function where number of ports of switch element is larger than 100, and this result shows that estimation in section 4.1 is adequate. To use this algorithm in an actual system, processing time must be less than TDM frame time of 3.3ms. In this condition of simulation, processing time is still over 3.3ms at any size of switch element. However, reduction of amount of process of the proposed method will be expected by introduction of parallel processing. For example, it is known that amount of sort process of n elements is reduced from $O(n)$ to $O(\log n)$ by introduction of parallel processing. Effects of reduction of phase 1 and phase 2 by introduction of parallel processing is calculated as follows.

– Phase 1

Sort process for each timeslot can be processed in parallel, and sort process itself can be also processed in parallel. Therefore, order of amount of process is reduced from $O(n^2)$ to $O(\log n)$.

– Phase 2

Sort process to decide to which timeslot middle stage switch is assigned can be processed in parallel. Therefore, order of amount of process is reduced from $O(n^2)$ to $O(n \log n)$.

Total amount of process will be dependent on ratio of amount of process between phase 1 and phase 2, and here it is estimated as $O(n \log n)$.

When N is 16 and n is 4, amount of sort process will be reduced by 1/2 of the original, and absolute processing time will be within 3.3ms. Switch element of 4x4 is about one twentieth of a target size of switch element of 64x64 described in Section 3. To reduce processing time farther, both of introduction of parallel processing in algorithm and devices in hardware

logic to realize this algorithm will be indispensable. The latter subject is a future work.

5. CONCLUSION

In this paper, a new dynamic timeslot routing algorithm in Clos network is proposed as an appropriate method for variable bit rate TDM switch. The most important feature of the method is that timeslot routing in switches of the first stage can be controlled without internal blocking, and operation to realize this characteristics is explained in detail. Processing time of the proposed method is evaluated by software simulation, and it is indicated that it can be controlled in real-time when switch element size is 4×4 . As a future work, design of specific hardware to reduce amount of process is planned.

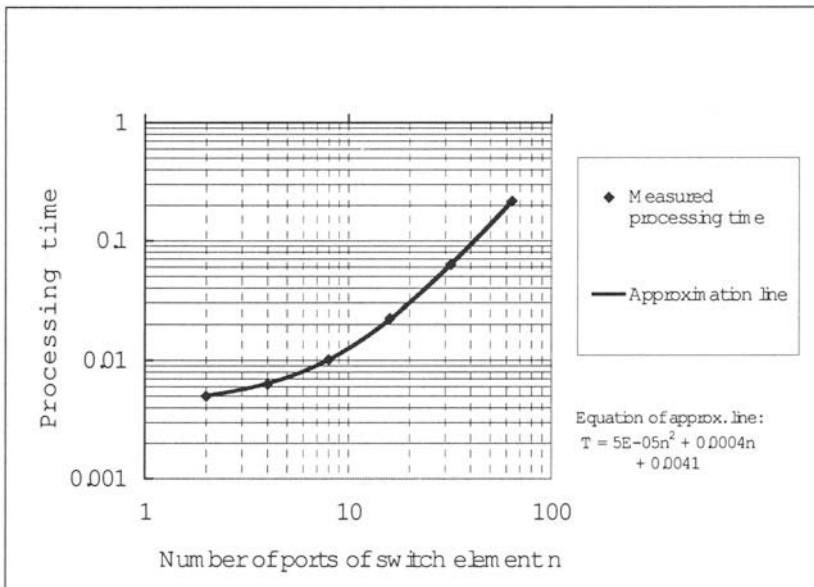


Figure 10. Evaluation of processing time by software simulation

REFERENCES

- [1] Tadao Saito, Hitoshi Aida, Udomkiat Bunworasate, Takayuki Muranaka, Terumasa Aoki. (1999). VTDM: A variable bit rate TDM switch architecture for video stream. *IEEE GLOBECOM '99*.

- [2] Patel J. (1981) Performance of processor –memory interconnections for multiprocessors. *IEEE Trans. Comput*, 70, (10), pp.771-780.
- [3] V. E. Benes. (1964). Optimal rearrangeable multistage connecting networks. *The Bell System Technical Journal*, Vol.43, No.4 pp.1641-1656.
- [4] C. Clos. (1953). A study of non-clocking switching networks. *The Bell System Technical Journal*, Vol.32, pp. 406-424.
- [5] Satoru Ohta, and Haruo Yamaguchi. (1987). A Non-Congestion Self-Routing Control Algorithm of Packet Switching Networks. *Transactions of IEICE J70-A*, 2, pp.312-319. (In Japanese).
- [6] Hideki Satoh, Yoshihiro Ohba. (1992). Switching Algorithm and Network Control in Framed ATM Networks. *Workshop of IEICE SSE92-22*. (In Japanese).
- [7] S.J.Golestani. (1991). Duration-Limited Statistical Multiplexing of Delay-Sensitive Traffic in Packet Networks. *IEEE INFOCOM '91*, pp.4A.4.1-4.A.4.10.