Performance Evaluation of a Modified Cyclic Banyan based ATM Switch for B-ISDN

V.S. Tripathi, C.B Tripathi and S.Tiwari

1. Department of E & C E, M. M. M. Engineering College, Gorakhpur, India
2. Department of Electronics Engineering, M. N. I. T. Allahabad, India
Email: {vijay_s_t@yahoo.com, cbtripathi@rediffmail.com, sudarshantiwari114@hotmail.com}

Abstract

This paper focuses on designing a Fast Packet ATM switch suitable for B-ISDN. It is a Banyan network using cyclic interconnection among switching elements of the same stage for deflection routing. The proposed routing is as simple as that of the generic Banyan network, and all the switching elements (SE’s) have a uniform structure. To design the proposed network and to develop its self-routing property we observe that all the SE’s of the Banyan network are arranged in a regular pattern topologically. We, thus, present a high-performance switching fabric based on the topological properties of Banyan Networks. As a result, we show that the new network has a better performance than other networks.


1. Introduction

Fast Packet switches are used in Broadband Integrated Services Digital Networks (B-ISDN) to route the packets towards their respective destinations at a high speed of a few Giga Bits per second. They provide low bit-error rates, and are well supported for use on high-speed fiber links [1,2]. For services demanding quality and real-time delivery, connection-oriented Asynchronous Transfer Mode (ATM) is used. An ATM switch is shown in figure (1).

The core of a fast packet switch, which influences both the performance and the cost, is its switching fabric. Shared medium, shared memory and space-division approaches are used for the purpose. Among space-division switches a multistage interconnection network called Banyan network is a popular choice due to its suitability to VLSI implementation and its self-routing capability [3]. The switching fabrics based on Banyan Networks are self-routing, simple and modular but they are blocking type switches.

Several methods like increasing the bandwidth of internal links and providing internal buffers have been widely used but deflection routing is found to be a good alternative to treat the blocking. Deflection routing is used in high-speed networks, since it gives a good performance and is easy to implement [4 - 6].

Deflection routing in Delta networks was proposed by Park and others for implementation of a Cyclic Banyan Network [7]. They proposed that packets might move laterally if they are unable to move forward. This proposal requires complex routing decisions at switching element level. This increases delay. Another problem with this switch is the problem of Packet-sequencing.
We present a high-performance Modified-Cyclic-Banyan (MCB) Network based Fast Packet switch. It is a deflection self-routing Banyan network. Packets failing to get selected for the destined link are sent along different links within a cyclic group in the same stage so that they can find an alternate path to their destination.

This paper is organized as follows. In Section II, we investigate the topological properties of the Banyan network to derive the cyclic Banyan network and an MCB network based ATM Switch and its routing scheme. In Section III, we analyze the performance of the network using analytical model and simulation. Finally, concluding remarks are given in Section IV.

2. Architecture of an MCB Network Based ATM Switch

This section deals with the architecture of MCB network based Fast Packet switch.

2.1 Banyan Network

Banyan networks belong to the class of Multistage Interconnection Networks (MINs). They were defined in [8] and are characterized by the property that there is exactly one path from any input to any output. A simple example of an $8 \times 8$ banyan interconnection network is drawn in figure (2).

![Figure 2: Banyan Network](image)

This network consists of Switching Elements (SE) and links. Each SE is a $2 \times 2$ crossbar switch, which can receive packets at each of its two input ports and send them through each of its two output ports. The SEs are joined in a systematic manner to form stages of SEs. Each SE in stage $j$ is connected to SEs in stages $j+1$ through links. In an $N \times N$ Banyan network, there are $\log_2 N$ stages and each stage has $N/2$ SEs.

2.2 Cyclic Banyan Network

Topological properties of Banyan network were first studied in [9]. It was observed that all SEs of the Banyan network are arranged in a regular pattern in terms of topology. In other words, each stage of the Banyan network is composed of a sequence of the cyclic group, realized with SEs. Stages are connected symmetrically through the links between them. A Cyclic Banyan network was developed from the basic Banyan network by the addition of links chaining all SEs of a stage. These additional links required that each SE must be of size $4 \times 4$. On the basis of a fully adaptive routing control in the Cyclic Banyan Network a destination-tag based routing algorithm was also proposed.

This Algorithm is good for Delta network, which has a uniform pattern for links between successive stages. When link pattern is stage-dependant, as in Banyan network shown in figure (2), this routing algorithm will take huge amount of time. Moreover, a packet destined to last output port of an $n$ stage Banyan network can not reach there in a short time, if it is somehow de-routed to first SE at the $(n-1)^{th}$ stage. Therefore, there must be some method of restriction for packets to be de-routed, so that they don’t go to a location from where they could not be properly routed.

2.3 Modified Cyclic Banyan (MCB) Network Based ATM Switch

Modified Cyclic Banyan network based ATM Switch has following properties and advantages—1. Use of simple basic building blocks and modular construction allows the expandability of a switch, 2. The self-routing property eliminates the need of complicated central processor in path hunting/routing, 3. A patterned and systematic interconnection structure reduces the density of integration and packaging, 4. Simple links determine the feasibility in synchronizing the signals at bit, byte, and packet levels, 5. The capability of engineering the switch systematically to meet growing throughput and capacity requirements by incremental augmentation, 6. Robustness and reliability regarding the fault detection and ease of
replacement and 7. Overall hardware is simple to manage along with high throughput for multicast traffic [10].

The MCB network can be obtained from the basic Banyan network by addition of lateral links at all SEs of a cyclic group in the stage. The cyclic group of SEs in any stage is decided by a simple algorithm based on topological properties of the Banyan Network: starting from the first (0th) stage, kth stage has $2^k$ numbers of cyclic groups within it. The number of SEs in each cyclic group of $k^{th}$ stage is $(n/2^{k+1})$, where $n$ is number of ports.

In other words, the first stage has all $m$ SEs in a single group, where $m$ is number of switching elements in any stage. The second stage has $m/2$ SEs in each group. So, the second stage will have two cyclic groups. The next stage has $m/4$ SEs in each group. It has four cyclic groups. This way group size keeps reducing by a factor of 2 at each successive stage. Finally, the last stage has single SE in each group. Implementation of additional links requires augmented SEs, each having lateral-in links and lateral-out links along with the input and output links. SE is thus a 4 X 4 crossbar switch. Figure (3) shows an example of an 8 X 8 MCB network and the SE used.

2.4 Routing Algorithm

When the SE’s are grouped in the given way, the routing decision becomes very simple, as the packets are free to traverse laterally in their respective group at any given stage. Suppose a packet is blocked at any SE, it will try to start from any other SE situated in the same group. If a free path is obtained to the destination SE, the packet is transmitted during the same switching cycle; otherwise it tries to start from another SE in the same group. If a path is found, the packet is transmitted otherwise it tries another SE (may be original one) in the same group. This routing algorithm has two advantages – First, all SE’s have a single algorithm, independent to the stage, so its implementation in VLSI for all SE’s is universal and therefore, simple. Second, buffers are not essential within the switching fabric. However, there is a need of output buffer, as many packets may appear at the same output port in a single switching cycle.

3. Performance Evaluation

Following the analytical tool presented in [11] we can make some definitions as follows:

$p_0(k,t)$=the probability that the switching element buffer at stage $k$ is empty at the beginning of the $t^{th}$ clock, and $p_1(k,t)=1-p_0(k,t)$.

$q(k,t)$=the probability that a packet is ready to enter a switching element buffer at stage $k$ during the $t^{th}$ clock period.

$r(k,t)$=the probability that a packet in a switching element buffer at stage $k$ is able to move (forward) into the next stage during the $t^{th}$ clock period.

Further, we can assume that –

(i) Loading is balanced. The arriving packets are destined for each output with equal probability. The load on each input is $0 \leq q(1) \leq 1$. With a balanced load the state of each switching network in stage $k$ should be statistically the same.

(ii) The states of the two buffers within a switching element are statistically independent. This assumption is justified by taking note that packets entering the input of a switching element originate from disjoint and independent network inputs.
Now we can write a series of probabilistic equations, recursive in the stage number and in time for the above quantities.

Initially,

Case 1: if all the four buffers in previous stage are filled, then out of 9 there are 8 possible ways for a packet to go in desired buffer of this stage.

Case 2: if any three buffers in previous stage are filled, then out of 13 there are 10 possible ways for a packet to go in desired buffer of this stage.

Case 3: if any two buffers in previous stage are filled, then out of 12 there are 9 possible ways for a packet to go in desired buffer of this stage.

Case 4: if only one buffer in previous stage is filled, then out of 8 there are 4 possible ways for a packet to go in desired buffer of this stage.

So,

\[ q(k,t) = \frac{8}{9} \times p_1(k-1,t) \times p_1(k-1,t) \times p_1(k-1,t) \times p_1(k-1,t) \]

\[ + \frac{10}{13} \times p_1(k-1,t) \times p_1(k-1,t) \times p_1(k-1,t) \times p_0(k-1,t) \]

\[ + \frac{9}{12} \times p_1(k-1,t) \times p_0(k-1,t) \times p_0(k-1,t) \times p_0(k-1,t) \]

\[ + \frac{1}{2} \times p_1(k-1,t) \times p_0(k-1,t) \times p_0(k-1,t) \times p_0(k-1,t) \]

\[ k=2, 3, 4, ..., n \]

\[ \cdots \cdots (1) \]

This equation relates \( q(k,t) \) to \( p_{i_1}(k-1,t) \). The first term represents the case that all buffers in \((k-1)^{th}\) stage are filled. This term is followed by three terms corresponding to other combinations of full/empty buffers in the \((k-1)^{th}\) stage SEs.

Next,

\[ r(k,t) = \left[ p_0(k,t) + \frac{8}{9} \times p_1(k,t) \right] \times \left[ p_0(k+1,t) + p_1(k+1,t) \right] \times r(k+1,t) \]

\[ k=1, 2, 3, ..., n-1 \]

\[ \cdots \cdots (2a) \]

\[ r(n,t) = \left[ p_0(n,t) + \frac{8}{9} \times p_1(n,t) \right] \]

\[ \cdots \cdots (2b) \]

The first term of equation (2a) is the probability that there is no incoming packet to a \( k^{th} \) stage switching element buffer. The second term is the probability that the \( k^{th} \) stage switching element buffer is either empty or empties out.

This set of equations models the dynamics of the system. If there is an equilibrium solution, these quantities should converge to time-independent values for \( q(k), r(k), p_0(k), \) and \( p_1(k) \). The equations can be solved iteratively for the equilibrium values. The two performance measures of most interest, as usual, are throughput and delay. The normalized throughput, \( \bar{Y} \) or the average number of output packets per output link per slot is

\[ \bar{Y} = p(k), \quad k=1, 2, 3, ..., n \]

\[ \cdots \cdots (4) \]

The normalized average delay \( \tau_{\text{norm}} \), is

\[ \tau_{\text{norm}} = \frac{1}{n} \sum_{k=1}^{n} r(k) \]

\[ \cdots \cdots (5) \]

For instance, if \( r(k) = 0.5 \) for some stage \( k \), then the delay in moving through that stage is \((1/0.5) = 2\) slots. The normalized average delay is computed by summing each stage’s average delay and dividing by \( n \), the number of stages. The minimal delay is thus 1.0. This corresponds to a delay of one time slot for each stage.

The performance of a 16 X 16 MCB network with restricted-deflection self-routing has been evaluated and the results are shown in figures (4) and (5).

4. Conclusion

The graphs clearly depict that the MCB network has very good delay/throughput performance as compared to simple Banyan network. Moreover, the proposed network shows a substantial improvement in throughput and delay over Cyclic Banyan network of [9]. The basic building blocks are small modules, and can be used to construct a large-scale ATM Switch. This modular architecture can provide multicast services also. Routing is very simple and easy to implement. Physically, it can be realized as an extension of existing Banyan Networks. The switching modules are not completely cyclic, and this grouped-cyclic switch fabric provides a flexible distributed architecture, which is the key to simplify the operation of the whole switching
system. The modularity implies less stringent synchronization requirements, and makes higher speed implementation possible.

Specifically, we have described the following:

- We have described a method to exploit the hardware features of a multistage interconnection network (MIN) switch for use in fast packet switching.

- We have described the architecture of a new MCB Network based switching fabric-having advantages similar to those of Banyan.

- We have shown how the combination of these techniques results in efficient use of resources for B-ISDN.

References


Acknowledgements

The authors are thankful to AICTE for its support to this work under R & D scheme.