The Kaleidoscope switch-a new concept for implementation of a large and fault tolerant ATM switch system

Dittmann, Lars

Published in:
IEEE ATM Workshop 1997. Proceedings

Link to article, DOI:
10.1109/ATM.1997.624766

Publication date:
1997

Document Version
Publisher's PDF, also known as Version of record

The kaleidoscope switch
- a new concept for implementation of a large and fault tolerant ATM switch system.

Lars Dittmann
Center for Broadband Telecommunications (CBT)
Department of Electromagnetic Systems (EMI)
Technical University of Denmark, Building 348
DK-2800 Lyngby
Tel : +45 4525 3851
Fax : +45 4593 1634
EMAIL : ld@emi.dtu.dk

Abstracts : *This paper describes a new concept for implementing large switch network based on smaller modules. The concept is based on an alternative selfrouting structure that due to a point symmetry allows the bit in the routing tag to be processed in random order. Among others this property provides an inherent fault protection and allows a simple implementation of broadcast and multicast.*
The concept has been implemented as a small prototype, that currently is used in a national experimental ATM network in Denmark

0-7803-4196-1/97/$10.00 © 1997 IEEE

1. THE KALEIDOSCOPE SWITCH

The KALEIDOSCOPE switch core represents an alternative way to implement an ATM switch, with the objective to fulfil many of the requested functionalities of an IBCN switch node but with limited technology requirement in order to be able to use the concept for high speed high capacity switch nodes.

The proposed switch structure demonstrates how different strategies are combined and used to the extent where they give an optimal performance. To allow a simple control it has been an objective to enable the use self-routing tags attached to the cells at the input, but with ability to exploit the multipath structure in case of failure or a need for load balancing.

The basic requirements of the KALEIDOSCOPE switch have been :

- Scalability to very large switches
- Broadcast and multicast support (with full bandwidth flexibility)
- Efficient usage of the available buffer capacity
- "Non blocking" possibility
- Ability to isolate and control the possible cell loss
- Minimised relation between size and technology

The requirements for the switch have been made very broad in order to be able to use the structure at different levels of the network - from access to the core switches.

Draw in a multi-dimension system the topology is a hyper-cube, but projected into a two dimensional system the structure has a point symmetric structure, as shown in Figure 1-1, which is the reason that it has been called the kaleidoscope switch. The Kaleidoscope switch has been implemented as a prototype and is currently use in the national Danish experimental network called BATMAN (Broadband ATM Access Network).

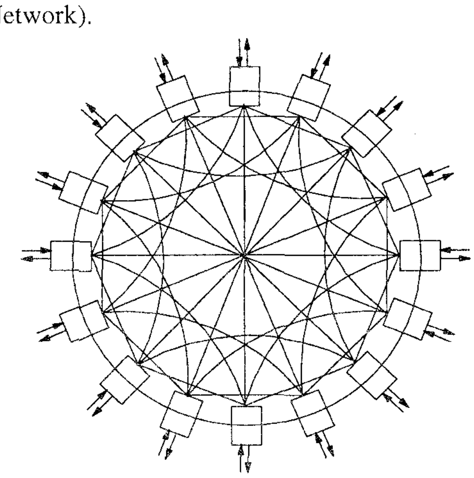

Figure 1-1

Point symmetric architecture of an 16x16 Kaleidoscope switch

2. THE NEXT GENERATION OF ATM SWITCHES

In the initial phase of ATM switch design there was a lot of debate concerning the location of the buffer and the degree of utilisation. The use of input buffer was commonly rejected due to the "head-of-line" blocking while the centralised buffer was elected as the most efficient, but difficult to implement, and therefore most designer went for the output buffer (in some shape) which also allowed a simple analysis because of the knowledge about the output process.

However as the definition of service and the general problems with resource management and specifically fairness became obvious, the buffer became more than just a simple contension buffer.

While size has been seen as the main different between the current generation of ATM switches and the next, it is believed that the ability to meet demands from new services (or service classes) is just as important. The latest example of how a service can have an impact on the switch structure is seen with the ABR service, which at least requires a separate service class buffering and preferably has a buffer/queue per channel.

Generally has the use of "per channel queuing" lately gained a significant interest as a way to guaranty fairness and restrict the mutual interference between channels. This way of implementing buffering totally reverse the traditional criteria as the term "head of line blocking" no longer has a meaning.

When considering large switches it is further believed that the traditional way of having a common type of i/o modules is to restrictive as individual demands (e.g. services, fault protection) should be fulfilled by replacing the i/o modules and to a very limited degree be dependent on the core system.

All these arguments speaks in favour of placing as many functionalities, including buffering, at the input of the switch and kept the actual switching function as simple as possible.

This has been the aim for the kaleidoscope switch presented in the following.

3. High Level Switch Structure

The basic concept for the Kaleidoscope switch is to connect input/output port in a ring oriented manner, but in addition to the basic ring a number of additional short-cuts are provided that allows a binary search or routing towards the destination port. Each node on the ring is intended to be buffered with a buffer for each of connected short-cuts and the output port. In addition to the data path between the nodes a flow control channels is available to be used for back pressure and performance monitoring. Except for the special routing tag handling and the flow control signals, the NxN Kaleidoscope switch can be seen as an interconnection of N output buffered switch elements with the size $(Log_2N+1)x(Log_2N+1)$.

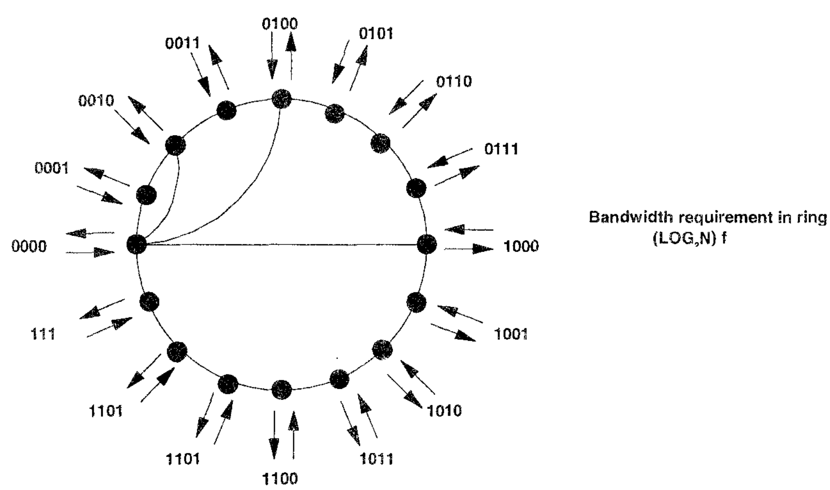

Figure 3-1

Ring switch with binary destination search using internal shortcuts

The 16x16 switch shown in Figure 3-1 illustrates the binary search for the given destination (using relative addressing). By the use of internal shortcut the route between the inlet (source) and outlet (destination) is continuously divided into two parts where the destination part is selected and further divided.

It can be shown that the load on each internal link is half the load of the external inlet based on a uniform load distribution. As each node is a $\log_2 N$ times $\log_2 N$ switch the required speed up in case of uniform load distribution is $\log_2 N$.

If the assumption of uniform traffic cannot be fulfilled and the switch have to be totally non blocking it can be shown that the switch is non blocking if :

Eq. 3-1

$$\text{SpeedUp} = \frac{\text{NodeCapacity}}{\text{LinkCapacity}} = 2 * \sum_{i=0}^{\frac{(\log_2 N)}{2}-1} 2^i \qquad \text{for } \log_2 N \in \{2,4,6,8,.\}$$

$$\text{SpeedUp} = \frac{\text{NodeCapacity}}{\text{LinkCapacity}} = \left[2 * \sum_{i=0}^{\frac{(\log_2 N)-1}{2}-1} 2^i \right] + 2^{\frac{(\log_2 N)-1}{2}} \qquad \text{for } \log_2 N \in \{3,5,7,9,.\}$$

In this it can be seen that even if the kaleidoscope switch is made non-blocking, then there is a large gain compared to e.g. a shared medium switches.

The above results are all valid for point-to-point traffic and full broadcast traffic. In case of multicast it is more difficult to ensure a non-blocking property as it depends on the rules for setting up connections.

3.1 Routing in the Kaleidoscope switch.

Due to the binary destination search, it is possible to make the Kaleidoscope switch selfrouting in a very simple way. By using routing labels with a relative destination address, the routing can be performed by searching for the most significant bit equal to 1. This bit is used to determine the outgoing route of the ring node. Before the data are sent to the next node, via the selected route, this bit is reset to zero.

Example : Routing from inlet number 3 to inlet number 14 in a 16x16 switch (Figure 3-2), i.e. each node has 4 inlets and 4 outlets (in addition to the external connections). The inlets/outlets will be called A, B, C and D with A being the outermost connection. The connection will give a relative routing label 14-3 = 11 (1011).

> As the first bit is 1 the data is first sent on the outlet D to the inlet D on node 11 and the routing label is changed to 3 (0011).
>
> In the new routing label the first bit equal to 1 is in the location that corresponds to outlet B, and the data is now sent to inlet B on node 13 and the routing label is changed to 1 (0001).
>
> Finally the data is sent to node 14 with the routing label 0 (0000), which indicates that the destination has been reached.

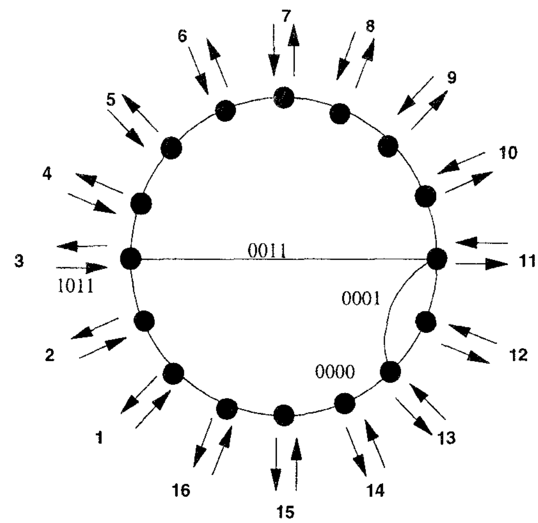

Figure 3-2

Self routing of data from inlet 3 to outlet 14 using relative addresses

Broadcast/multicast routing is handled by asserting a special bit in the routing label and then assigning the value of the label to the size of the switch (minus 1). Every time data passes a node the label is decremented with one and when the value gets to zero the data are removed from the ring.

3.1.1 Alternative routing.

The routing technique described above is the simples and most straight forward one for the Kaleidoscope switch. However, a little more advanced routing technique can be used with a potential gain in load balancing and fault protection.

In the simple routing approach the routing tag is evaluated from left to right, using the most significant non-zero bit first. However, as the order in which the routing bits are used is independent this rule can relaxed. If this is done, it will change the switch from a single path to (in most cases) a multipath switch, with a path for each non-zero bit in the routing tag.

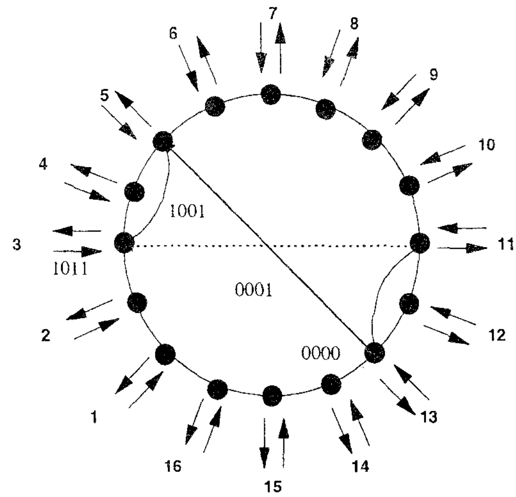

Figure 3-3

Self routing of data from inlet 3 to outlet 14 using relative addresses and alternative routing rule (dotted line shows basic routing rule).

To ensure that the cell sequence integrity is guaranteed, it will be a requirement that all cells belonging to the same connection are using the same routing rule. This can be ensured by adding an information in the routing tag, either with full indication of the order in which the individual routing bit shall be used or by a number that referees to one among a number of routing concept known by the system.

By allowing alternative routing rules and exploiting the internal flow control between each of the nodes, it will be possible to balance the load of the different internal links and avoid a blocking situation.

The alternative routes can also be used for fault recovery in case one of the nodes shows a malfunctioning behaviour. By extending the inter-node flow control with status information about the connected nodes, it will be possible to avoid to use a node that does not work correctly. As this kind of alternative routing is unique it does not require additional information in the routing tag. The routing rule simply has to be extended to:

> Search for the most significant bit equal to 1, that is connected to a well functioning node. If no such node exists the cell will be deleted.

This type of fault recovery cannot replace the fault protection obtained from redundancy, but can, in case of a distributed switch, allow a distributed implementation of redundancy. The concept can also be used for sub-equipping the switch, where a non active node position is handled similar to a faulty node. (This is further discussed in section 5)

4. DESIGNING A KALEIDOSCOPE SWITCH

The Kaleidoscope switch is a combination of two switch structures and represents a traditional way of making larger switches by interconnecting a set of smaller ones, similar to the way Clos did. The speciality of the Kaleidoscope switch is therefore not in the way the smaller switch elements are made, but in the way they are interconnected and the ability to operate the core as an autonomous unit due to the simple selfrouting .

To implement a switching system based on the kaleidoscope concept, the first thing to do is to select suitable switch element for the size of switch system requested. The element can in principle be any kind of NxN switch except that it has to be able to switch data to given outlet based on a routing label attached to the data and modify this label before it is forwarded to the next node.

In addition to the pure switching function the switch requires some pre-processing and post processing to insert the routing label on the inlet and remove it on the outlet.

The post processing also handles the multicast function, where the basic switch element just performs broadcasting by copying data, with the multicast/broadcast bit in the routing label active, to output number 1 and 2. In this way the switch is acting solely on the broadcast/multicast bit and does not distinguish between the two cases. In the postprocessing unit this type of data, recognised by the routing label, is either forwarded to the external output or deleted dependent on the configuration stored in the RAM.

5. FAULT PROTECTION IN A DISTRIBUTED SWITCH CONFIGURATION

To introduce fault protection of the Kaleidoscope switch node the general method of using a parallel redundant system can of course be used, but due to the distributed structure of the Kaleidoscope configuration, an alternative solution can be applied as well.

This alternative approach not only concerns the fault protection aspects but the entire structure of the switch node with the traditional subdivision into linetermation, (switch port) and switch core.

As the Kaleidoscope structure contains one ring node pair input output pair the proposal is to either include this function in the switch port or add a new functional unit - a switch node. The switch core will then be reduced to a passive interconnection board that most obviously can be integrated in the backplane.

Because of the various types of interconnection between switch nodes it is always possible to find alternative ways.

Due to the internal flow control using dedicated wires it is possible to use this system as a sense mechanism, and thereby easily detect a missing or malfunctioning i/o unit (ring node and/or switch port).

If the connection to a line unit is inactive, the cells to this specific outlet can clearly not handle correct, and must be deleted. Other cells that according to the binary routing strategy are supposed to pass the missing or malfunctioning node, must select an alternative route. The strategy could be as follows :

- Find first bit in routing label set to 1 : $rb_n = 1$
- If link that is associated with this bit is active
  - then use this link and invert the bit
  - else find next bit equal to 1 and use the link associated with this bit and reset the bit.
  - if no alternative routing bit is found
    - then delete cell (as destination is the inactive node)

This strategy could also be used for alternative routing in case of blocking. However, to ensure cell sequence integrity it has to be applied on a connection basis and not on a cell basis or by applying a mechanism at the enables to resequence the cells.

6. THE BATMAN NEWTORK

The first prototype of the KALEIDOSCOPE switch has been implemented for the Danish research and test ATM-network called BATMAN (Broadband ATM Access Network) under the code name ROBIN (Ring Organised Broadband Interconnect Network).

The BATMAN network is triangular network run by the Danish Operator TeleDanmark between two of the R&D centers of TeleDanmark and the Technical University of Denmark as shown in Figure 6-1.

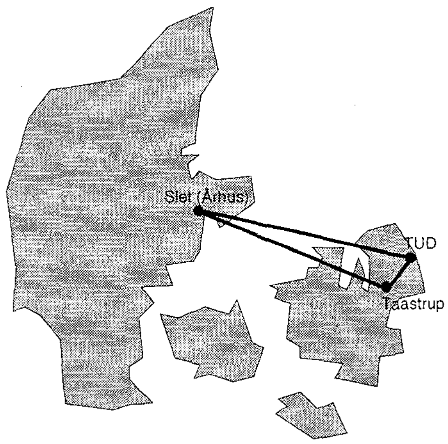

Figure 6-1
The 3 sites of the BATMAN network interconnected at 155 Mbit/s.

The interconnection between the different node has a capacity of 155 Mbit/s, with a parallel 2 Mbit/s network mainly for redundancy protection of the management and SS7 system.

The BATMAN network has been in continuos operation since November 1995 and is used for experimentation with different ATM services (such as LAN-emulation and high quality video conferencing) and new equipment and associated protocols (e.g. PNNI signalling and VOD servers).

7. CONCLUSION

This paper has described a new and alternative way to implement a large ATM switch system made up of smaller elements. Compared to existing proposals for selfrouting networks, that have a linear flow through the different stages, this proposal is point symmetric, which allows the routing tag to be processed in random order. This ability both allow for internal load balancing and inherent fault protection.

The concept, called the kaleidoscope network, has been evaluated in a prototype implementation for the Danish ATM test network called BATMAN.