A Multiprocessor Communication Architecture
For High Speed Networks

Iyengar. S, N Apte, A.A.Roy, Victoria Jubilee Technical Institute, Bombay, India
S. Sanyal, Computer Systems and Communication Group,
N.M. Singhi, School of Mathematics,
TIFR, Bombay, India
Wu Geng Feng, Dept. of Electronics &Electrical Engg.
USTC, Hefei 230026, Anhui, Peoples Republic of China.

ABSTRACT
Over the years, communication speed of networks has increased from a few Kbps to several Mbps, as also the bandwidth demand. Communication Protocols, however have not improved to that extent. With the advent of Wavelength Division Multiplexing (WDM), it is now possible to “tune” protocols to current and future demands. The purpose of this paper is to evolve a High Speed Network architecture, which will cater to the needs of bandwidth-consuming applications, such as voice, video and high definition image transmission.

INTRODUCTION
Communication speeds of local-, metropolitan-, and wide-area networks have increased by about six orders of magnitude. Current technology and WDM networks offer bandwidths higher than peak electronic data-rates. At the same time, user-demand for bandwidth has increased. Given the fast-increasing utilization of networks for high definition image transmission, digital voice and video, future networks have to meet demands of over 1Gb/user.

Even though current media can sustain such a demand, existing protocol structures and implementations are still unable to exploit media to cater these needs. The discrepancy between the speed of communication processor and the channel, translates into bottlenecks in the communication process. We suggest an approach to exploit high speed media to deliver a high throughput using multiprocessors for faster packet-processing.

THE SWIFT ARCHITECTURE

1. Description
The SWIFT architecture as detailed by Chlamtac I. and Ganz A. forms the backbone of our design. The architecture works at three levels. At the physical layer it uses a multichannel physical configuration. It provides b, \(1 \leq b \leq N\), sub-channels of equal bandwidth. The data link layer protocol is slot-oriented with exactly b permissions in each slot, equal to number of sub-channels. Sub-channel allocation depends upon node address and a slot-allocation matrix. Multichannel network layer is responsible for routing packets and delay factors influence routing. The FINE algorithm deals with the delay calculations. The architecture favours broadcast networks.

2. Drawbacks
Though the architecture potentially improves system capacity compared to conventional single-channel architectures, it can be improved further with optical fibre technology and multiprocessors. SWIFT protocols do not permit full-duplex (FDX) communication since each node can transmit only to a particular node during its slot-time; it can not receive from the same node (Fig.1). Hence, they can only be of stop-and-wait type. Window or similar protocols used in most practical networks can not be mapped easily on SWIFT. Such an implementation would mean longer time-out periods and hence reduced performance. Thus SWIFT manages to utilize more bandwidth out of available bandwidth than other architectures, but it still leaves room for improvement in terms of FDX communication between nodes.

In this paper, we therefore propose an architecture for high speed networking which can sustain various traffic requirements. It essentially employs a fibre optic network and uses multiprocessors to remove the aforementioned bottlenecks. The architecture, termed, Multiprocessors In Networking (MIN), manages bandwidth-consuming applications using demand-oriented channel-allocation.

THE MIN ARCHITECTURE

1. Channel Allocation Policy
The main aim of MIN is to facilitate FDX communication for every node. In FDX communication we must allocate a pair of sub-channels between the two nodes. The connection requirements using WDM are as shown in Fig.2.

We note that for an N-node network, we require N sub-channels. These sub-channels are sorted in terms of time; in each time slot a node can communicate with only one other node which is distinct from the nodes it can communicate with, in the other slots. Thus the requirements (FDX and TDM-WDM) for allocation matrix A are:
A: Destination-conflict freeness.
B: Single node-pair allocation per pair of sub channels.
C: Non-overlapping pairing in each slot.
D: Unique destination (previous slots) in each slot.
E: Equi-spaced allocation (fixed cycle time) per node pair.

Cycle time is defined as number of elapsed slots between two slots with permissions for the same set of node-pairs. Any construction of A will favor an even number of nodes.

**Construction of the sub channel allocation matrix A:**

We note that A will have N rows (one row per node) and (N-1) columns, since \( N \) distinct pairs are possible; the cycle time for each node will be (N-1) slots. We number the nodes from 0 through (N-1) and the equal-duration slots from 1 through (N-1). We have, for elements of A:

- \( a_{ij} = 1 \) if \( i \neq j \) and \( i, j \in \{0, 1, 2, \ldots, N-1\} \)
- \( a_{ij} = 0 \) for \( i = j \) or \( i, j \in \{0, 1, 2, \ldots, N-1\} \)

FDX communication is possible between the nodes i and k in the slot j. The matrix A can be constructed from another \((N-1) \times (N-1)\) matrix B where B is a Latin square. B can be constructed as:

- \( b_{ij} = i + j \mod (N-1) \)

Matrix A can be constructed using the mappings:

- If \( b_{ij} = j \)
- If \( b_{ij} = 2i \), \( 0 < i < N-1 \)

Then \( a_{ij} = k \)

Choose N=N'+1 (N is even). Construct N X (N-1) matrix as above. Drop row number N' and replace all entries in A equal to N' by 'X' indicating idle state for that companionless node.

The cycle-time for N-node network can be given as:

\[ T = 2 \times \lceil \frac{N}{2} \rceil - 1 \]

This allocation follows FDX communication in every slot.

2. Routing

In MIN, we have three different levels of routing; with increasing complexity in operation. The concept of routing is based on the idea that a packet from i to k may be routed via j if the slots i-j and j-k occur before the slot i-k.

To formally describe the algorithm, we define waiting-time delay \( d_{ij} \) for a node to be the number of slots from the current slot – when a packet for a particular destination was received – to the slot in which the transmission actually occurs.

\[ d_{ij} = N-1-t-t' \]

where, \( t \) is current time slot and \( t' \) is the time slot when transmission actually occurred.

**Lemma 1:** The maximum waiting-time delay (in terms of slots) experienced by a node for transmission to a particular destination is N-2.

**Proof:** From (II) we see that

\[ d_{max} = (N-1) - t - t' \]

Such a condition occurs when a node just misses its designated slot and no other intermediate node is available for forwarding its packets. Hence,

\[ 1 + t - t' \]

which gives \( d_{max} = N-1 - 1 = N-2 \) when

**Level 0 Routing:** This level just follows the channel allocation matrix A for transmission. It does not allow any other transmission permission. As in (III), maximum delay experienced will be N-2. This type of routing may result in buffer overflow during large file transfers.

**Level 1 Routing:** Here we use other nodes for routing packets. Intermediate courier nodes are chosen according to a routing table RT, which can be constructed as follows:

- If delay is reduced when i transmits packet
- For j, in slot t, to current node

\[ r_{ij} = \{
\]

- 0 if i waits for an optimal slot or it uses
- level 0 or level 2 routing

An algorithm for constructing matrix RT (\( N \times N \times (N-1) \)) is given in Fig.4.

Initialize matrix RT to 0

FOR each source node i (0 \( \leq i \leq N-1 \))

FOR each destination node j (0 \( \leq j \leq N-1 \), \( j \neq i \))

FOR each slot t \( (1 \leq t \leq N-1) \)

BEGIN

\[ dw = \text{current waiting-time delay} \]

IF \( dw = 0 \) THEN exit

ELSE BEGIN

\[ d_{min} = dw \]

trans = 0

END

REPEAT

Probe packet walk-path for source node in allocation matrix to get next intermediate node. Replace source node with the intermediate node.

For every replacement, \( \text{trans} = \text{trans} + 1 \)

UNTIL destination is reached

IF \( \text{trans} < \text{d}_{\min} \) THEN BEGIN

\[ \text{d}_{\min} = \text{trans} \]

\[ t' = t \]

END

END

IF \( \text{d}_{\min} < N-2 \) THEN \( r_{ij'} = 1 \)

**Fig. 4 Level 1 Routing**

Thus if a source node misses its designated slot for a particular destination node it scans the time-axis of the corresponding entry in RT for an optimal slot. If there is none, then it uses Level 0 or Level 2 Routing.

**Level 2 Routing:** Here we require additional N sub-channels for proper synchronization of events and an extra set of transmitters and receivers at every node. Consider slot A-B. In case both A and B do not want to communicate with each other, this slot will be empty. We can make use of this slot. In the level 2 routing, A disconnects its transmit sub-channel with B and uses it to connect to C which is some other node A wishes to communicate with; A has missed the designated slot A-C. A will first convey to C its intention to transmit (Fig.5). The decision about continuing the session depends upon C, which will base its decision on the number of requests it has received. Requests for sessions from various nodes will be honored depending on the priority of the node which must be predefined (e.g. priority based on node addresses). Thus C will communicate its decision to the
proper node which then is eligible to transmit in this slot, to C. The connection between the node-pairs about their respective destination (i.e. in case both A and B wish to communicate with C) can be resolved based on a similar priority scheme. To minimize this overhead, node-pairs use status packets at start of the slot to decide the course of action. The resulting structure of the slot is shown in Fig.6.

This type of interaction however must be supported by the hardware – we need to have receivers (at least) for implementation of this routing. The processing power of the network interface will put an upper limit on the number of receivers, as the node now have to manage multiple sessions. All receivers used a fixed tone but transmitters are tunable, to allow transmission to different fixed receivers. To manage these protocols we have to use multiprocessors.

3. The role of multiprocessors

It was desired that system performance be comparable to the bandwidth of physical media. As speed and bandwidth of the media increase it is the software that influences overall packet delay. Efficient functioning of the system will be achieved only if hardware and software are properly matched.

An important observation is that protocols based on extensively layered architecture are inherently inefficient for high speed communication. Layered approach works satisfactory in normal cases; but, a strict adherence to it increases processing and hence queuing delays. Reasons for this degradation include replication of functions in various layers, overhead of control messages, evaluation of functions which are not necessary in all situations etc.

Multiprocessors can be used if packet-processing is parallelized. For networks with fibre links and digital switches the error rate is low. So success-oriented protocols can run in parallel. These functions use a shared memory buffer to store processed packets. A communication controller then uses this buffer to take appropriate packets and puts them in a queue. A slot-filter then chooses packets eligible for transmission during the current slot. A routing-filter then acts upon these packets for proper routing based on the routing policy.

CONCLUSION

In this paper, we have evolved an architecture which would cater to the needs of current and future bandwidth-consuming applications in high-speed networking. The changes in design of network interfaces are expected to improve the system throughput and average packet delay. Current concepts and implementations of protocols being inefficient for future networks, we have tried to eliminate the bottlenecks at various levels. These changes have strongly been influenced by WDM and related technology.

We have proposed a TDM-WDM channel allocation policy which can serve multiple users. Such an allocation will lead to optimal utilization of channel. The policy also facilitates FDX communication which will further help in improving performance. As this simple allocation will cause most of the bandwidth going wasted we have developed two more levels of routing. Such a routing mechanism can be made sensitive to the traffic and can adapt its behavior accordingly. At network interface level, a multiprocessor architecture emerges as the only practical solution.

We believe that the ideas presented in the paper will initiate a discussion in future networking, looking from this aspect.

REFERENCES

1. Hass Z., A Communication Architecture for High-Speed Networking, IEEE INFOCOM 1990. 433-441
2. Mukherjee B., WDM-Based Local Light-wave Networks Part I: Single-Hop Systems, IEEE Network. May, 1992, Vol.6 No.3, 12-27.
3. Kazovsky L.G. et al., WDM Local Area Networks. IEEE LTS, May, 1992, Vol.3 No.2. 8-15.
4. Chlamtac I. and Ganz A., Towards Alternative High Speed Network Concepts: The SWIFT Architecture, IEEE Trans Comm. Vol.38, No. 4, 431-439.
5. Taub H. and Schilling D.L., Principles of Communication Systems, 2nd Ed., McGraw Hill (1989). 137-138, 697-698.
6. Tannenbaum AS., Computer Networks, 2nd Ed Prentice-Hall of India (1991). 14-19, 223-239.
7. Reingold E.M. et al., Combinatorial Algorithms Theory and Practice, Prentice-Hall (1977).
8. Zitterbert M., High-speed Transport Components. IEEE Network. Jan. 1991.54-61
9. Zitterbert M., A Multiprocessor Architecture for High Speed Network Interconnections, IEEE INFOCOM 1989, 212-219.
10. Jolitz W.F., High-speed Networking. Dr. Dobb’s Journal, August, 1992. 86-93.
Fig 1: Subchannel Allocation matrix of SWFT

Fig 2: FDX Communication using WDM

Fig 3: Subchannel Allocation Matrix of MIN

Fig 5 Level 3 Routing

Fig 6: Structure of Slot

Fig 7: Architecture for communication subnet