"P802.1d MAC Bridges Draft Standard"
IEEE Project 802 Committee P802.1d/D8, (March 89)

[2] —,
Remote Bridge Draft Standard

[3] —,
"Synchronous Optical Network (SONET) Transport Systems: Common Generic Criteria"
Bellcore Technical Advisory TA-NWT-000253, Issue 6, (Sept 1990)

[4] —,
"P802.6 DQDB Subnetwork of a Metropolitan Area Network Draft Standard"
IEEE Project 802 Committee P802.6/D15, (Oct 90)

[5] —,
"Proposed 802.6 SONET PLCP Mapping"
IEEE Project 802 Committee P802.6_SONET_PLCP/D3, (Sept 90)

[6] Taft, Nina
"Bridge Protocols for an SMDS-FDDI Remote Bridge"
Bay Bridge Project Report 5, (June 1991)

[7] Le, My T & McKeown, Nick
"Project Proposal for SMDS-FDDI Bridge" *Bay Bridge Project Report 0*, (Aug 1990)

[8] Le, My T & Kao, Ting
"Overview of DQDB MAC Chip Architecture"
Bay Bridge Project Report 3, (Oct 1990)

[9] McKeown, Nick
"Segmentation and Reassembly for Cell based Mis-ordering Networks"
Hewlett-Packard Technical Report

Address Learning. The remote bridge gradually learns the addresses of all stations on the local and remote networks. On the FDDI interface, the source address (IEEE 48-bit format) of all passing FDDI packets are stored in the Address Table. When FDDI packets are removed from SMDS messages from the SMDS interface, the source address of the remote FDDI station is stored in the Address Table, along with the associated remote bridge address (E164).

Forwarding and Filtering. The destination address of all packets on the FDDI ring are compared against the Address Table. Only packets whose destination is not known to reside on the local FDDI ring are forwarded to the Host or to SMDS. If the FDDI destination address is in the Address Table, the packet is forwarded to the single, known remote bridge. Otherwise it is multicasted to all other remote bridges in this bridge group.

A.2 Multiple Spanning Tree

The multiple spanning tree algorithm would only be different when determining the topology.

Each Bay Bridge "declares" itself as the local root. This is done by broadcasting BPDUs on the local FDDI ring with a bridge identifier with a value guaranteed to be lower than the bridge identifiers of all other local bridges on the tree. ('0' would be an obvious choice). BPDUs are not forwarded or transmitted over the SMDS network. This is seen by some as an advantage of the multiple tree, as the SMDS traffic (and hence the user cost) would be reduced. However, it is arguable that the BPDU traffic would be extremely low and would contribute negligible extra cost to the SMDS service.

If a local tree has two remote bridges attached, as in figure 2, *both* bridges may become roots. This is because the local tree would split into two trees, each connected to the other via SMDS. The split would be determined by the local bridges — each local bridge will choose the closest (or lowest cost) root and not forward packets to the other. Hence, loops and packet duplication is avoided. The "split" point between the two trees may be moved by changing the cost count transmitted by the roots. This could be used by the network manager for load-balancing.

References

[1] —,

rameters, such as bridge group addresses, may be set locally by the network manager.

Running FDDI SMT The Host runs standard FDDI station management software.

Monitoring and Statistics Gathering The Host monitors the performance of each board collecting information on: the traffic on each network, packet errors, address conversion hit ratio etc. This information will be presented to the user and to network manager via the local network management agent.

Running Network Management The Host will also run SNMP network management code to allow remote monitoring of the bridge.

A Spanning Tree Topology

Two different spanning tree topologies will be investigated and implemented on The Bay Bridge:

1. A single spanning tree over the whole network, and
2. Multiple disjoint spanning trees with each remote bridge as the root of the local tree.

Once a topology has been determined, both algorithms operate almost identically. Both methods are summarised below and are described more fully in [6].

A1 Single Spanning Tree

The Single Spanning Tree algorithm performs 3 basic functions:

Negotiate the Topology. The Bay Bridge exchanges BPDUs with all stations and local bridges on the local FDDI ring and with other remote bridges. A root would be chosen based on the Bridge Identifier in exactly the same way as in a standard spanning tree. Topology loops and packet duplication would be avoided in an identical way to that described in the IEEE802.1d Standard.

2.3.1 Address Filtering

As mentioned in section 2.2, the Bridge Board must copy to the host or SMDS interface all packets from the ring with a destination address that is *not* in the Address Table. To perform this function, the FDDI MAC copies *all* packets from the ring. The Protocol Converter then decides whether the packet should be forwarded or filtered.

2.3.2 Bridged Packet Stripping

It is the responsibility of all stations on an FDDI ring to remove their packets from the ring. For a normal FDDI interface, this is straightforward — the interface strips all packets with a source address that matches its own.

This is more difficult for a bridge, as each packet transmitted onto the ring by the bridge will have a different source address.

There have been a number of solutions proposed to this problem, the most popular solution is to use MY_VOID Stripping. The current AMD FDDI Solution does not allow MY_VOID Stripping to be implemented; so as an interim solution we use a technique known as Source Address Table Matching. The source address of each packet transmitted onto the FDDI ring is stored in local associative memory (MUSIC Semiconductors 64x1k LANCAM). Each incoming packet is matched against the addresses in the CAM. If a match is found, the packet is removed from the ring. This technique has the advantage that no extra VOID frames are transmitted, at the expense of extra complexity. The technique will fail if more than 1024 packets are transmitted by the bridge each packet to a different destination during one token hold time (THT). For our demonstration network, this is not possible.

2.4 Host and Host Interface

The Host (Sun SPARC II with SBUS backplane) acts primarily as a monitor and manager; it is *not* involved in processing data packets between the two networks.

The Host will be will be responsible for:

Initialising the bridge This involves resetting cards and device drivers, initialising statistics counters and allocating addresses. User-definable paof 50ns enabling over 100,000 packets per second to be processed. During the processing of a packet, the converter will search for patterns in the protocol headers (e.g. MAC headers, LLC, SNAP, IP etc.) and consult local address tables to make routing decisions. When a decision has been made, the packet may be forwarded to a number of different output channels. These channels may be connected to another network interface (perhaps at different priority levels), host buffers for incoming data packets or for the logging of performance statistics or call-logging data by the Protocol Converter that are passed onto the host. Address conversion, checksum calculation and encapsulation/decapsulation is also carried out by the Protocol Converter as the data streams through.

Learning, Forwarding and Filtering

The Bridge Board gradually *learns* the addresses of all stations on the local and remote networks. On the FDDI interface, the source address (IEEE 48-bit format) of all passing FDDI packets are stored in the Address Table. When FDDI packets are decapsulated from SMDS messages from the SMDS Interface, the source address of the remote FDDI station is stored in the address table, along with the associated remote bridge address (E164).

For *forwarding and filtering,* the destination address of all packets on the FDDI ring are compared against the Address Table. If they are not in the table, they are copied from the ring. If the FDDI destination address is in the Address Table, the packet is forwarded to the single, known remote bridge. Otherwise it is multicasted to *all* remote bridges in this bridge group.

FDDI packets with group addresses may be selectively multicasted or filtered. A static entry in the Address Table determines whether the multicast address should be forwarded or filtered. In the absence of an entry in the table, the packet will be forwarded.

FDDI packets decapsulated from incoming SMDS messages are checked against the Address Table to check that they are destined for a station on the local FDDI ring.

2.3 FDDI Board

The FDDI board will use the AMD FDDI Supernet chipset to control access to the FDDI ring and to buffer packets between the FDDI MAC and the Bridge Board.

Apart from the normal FDDI MAC operation and packet buffering, the FDDI board must perform the following functions:

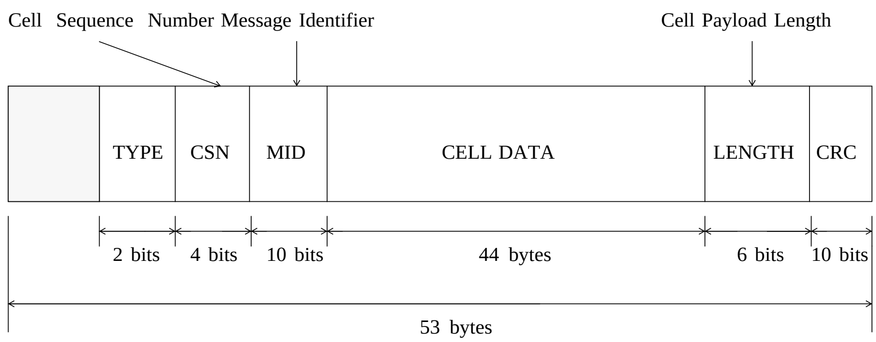

Figure 7: Structure of an SMDS Cell.

the message does not complete, it will timeout and will not be delivered. The message timeout value may be defined by the network manager.

Up to 128 interleaved SMDS messages (L3_PDUs) may be reassembled at the same time up to a total data size of 512 kbytes. Within each message, cells may be misordered up to a maximum misordering of ± 8 cells. This is an implementation of the algorithm described in [9].

2.2 Bridge Board

The Bridge Board is responsible for the forwarding and filtering of packets as well as building a consistent topology. The topology will be determined by the distributed spanning tree algorithm running as a process on the local host. Two alternative spanning tree protocols will be implemented as described in Appendix A.

All forwarding and filtering of packets will be carried out in hardware. The decision making and header manipulation will be carried out by a Protocol Converter.

Protocol Converter
This is a custom built unit, optimised for fast conversion between network protocols. It is programmable allowing many different protocols to be handled by the same unit. This will enable the Protocol Converter to provide significant hardware assist for high speed routing. Furthermore, the instruction set used by the Protocol Converter is programmable and readily upgraded should new operations be required.

The present implementation of the Protocol Converter runs on a clock cycle

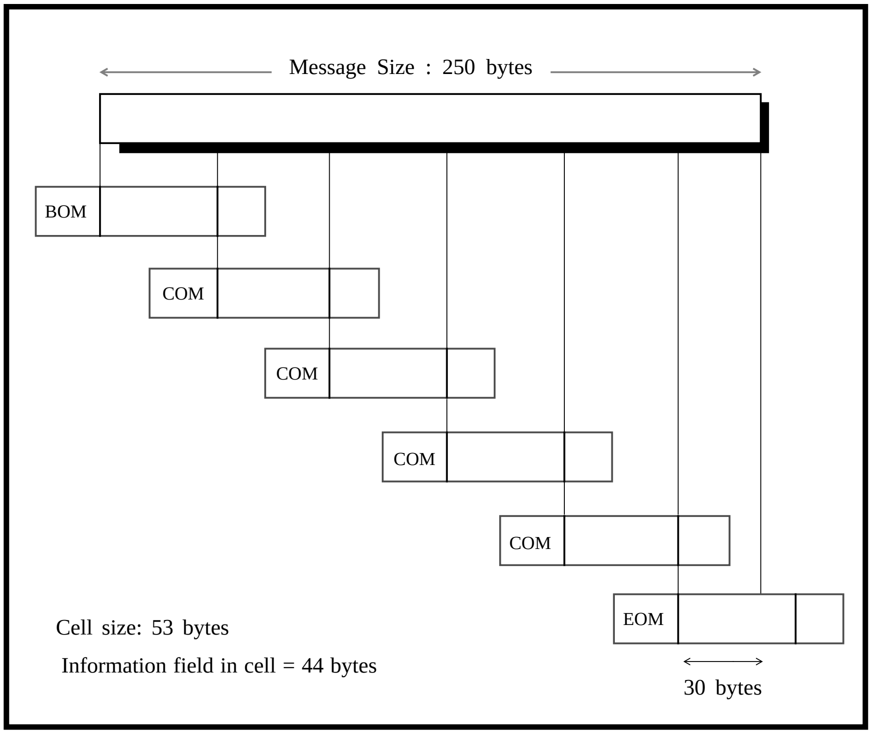

Figure 6: Segmentation of an SMDS Message.

Length, CSN (Cell Sequence Number) and MID (Message Identifier), figure 7.

Cell-sequence numbers are maintained for each message, starting at zero for the BOM cell.

The cell-length field will always equal 44-bytes, except for the EOM cell. The SAR Board calculates the cell-length for the EOM cell.

Reassembly

The SAR Board accepts all cells from the SMDS Board synchronous to the SMDS line clock. Incoming cells are reassembled into SMDS messages before being handed to the Bridge Board.

A MID table is maintained indicating which messages are currently being re-assembled. Cells arriving with new MID values are added to the table.

All incoming cells are stored in a cell buffer until a complete message is received. When the message is complete, the message is handed to the Bridge Board. If

Identifiers (MIDs) corresponding to messages in progress. The DQDB MAC maintains a table of current MIDs and alerts the Host of exceptions. The MAC does not check for correct cell ordering as this is handled by the SAR Board. The MAC checks the CRC of all incoming cells and hands cells directly to the SAR Board for reassembly, indicating whether the cell has passed the CRC.

Outgoing cells:
Cells are handed to the SMDS Board by the SAR Board. When the SAR Board requests a transmission, the DQDB MAC will negotiate for access to the SMDS medium. When access is granted, the SAR Board passes a cell to the SMDS board. The CRC checksum is calculated by the MAC and added to the cell trailer. The cell is placed in a SONET/DS3 frame by the PLCP and transmitted onto the medium by the Physical Layer.

The Bay Bridge will operate with either a STS-3 or DS3 physical layer and PLCP. These will be interchangeable via a daughter board to the main SMDS Board. The clock for the SMDS Board is derived from the Physical Layer.

The PLCP:
For incoming cells, the Physical Layer Convergence Protocol (PLCP) is responsible for accepting cells from the DQDB MAC, packing them into a SONET STS-3 or DS3 frame structure. The STS-3 mapping is shown in figure 4 and the DS3 mapping is shown in figure 5.

For outgoing cells, the PLCP places the cells into the frame structure before handing the payload to the Framer. The system will use the Bellcore Framer for SONET STS-3 rate and the TranSwitch Framer for the DS3 rate.

The PLCP is currently implemented using a number of finite state-machines (FSMs) in programmable logic devices. This may be further integrated in the future.

2.1.2 Segmentation and Reassembly Board

The SAR Board operates synchronously to the SMDS line clock, segmenting outgoing messages into cells and reassembling incoming cells into messages.

Segmentation
When the Bridge Board has an SMDS message to transmit over the SMDS network it hands the message to the SAR Board. Messages are segmented into 44-byte cells with 9 bytes of header and trailer information added. As an example, figure 6 shows how a 250 byte message is segmented. The SAR Board is required to fill in the following fields of the cell-header and trailer: the Type,

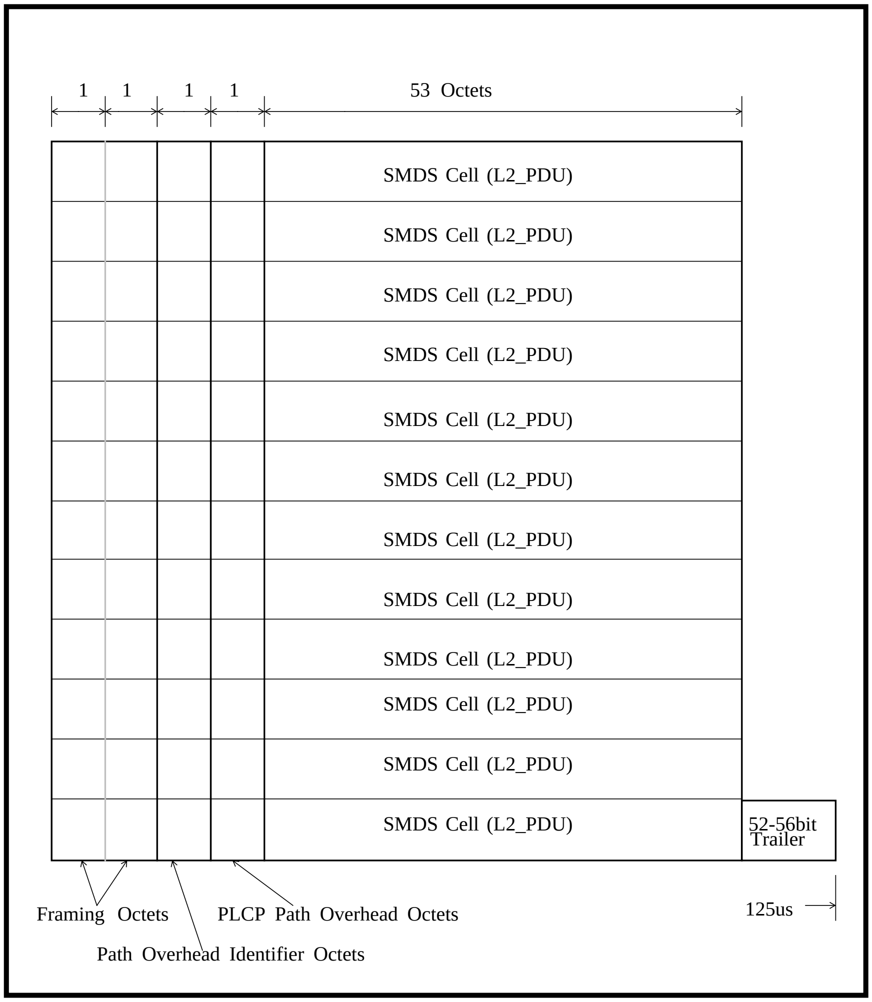

Figure 5: Mapping of SMDS Cells into a DS3 frame.

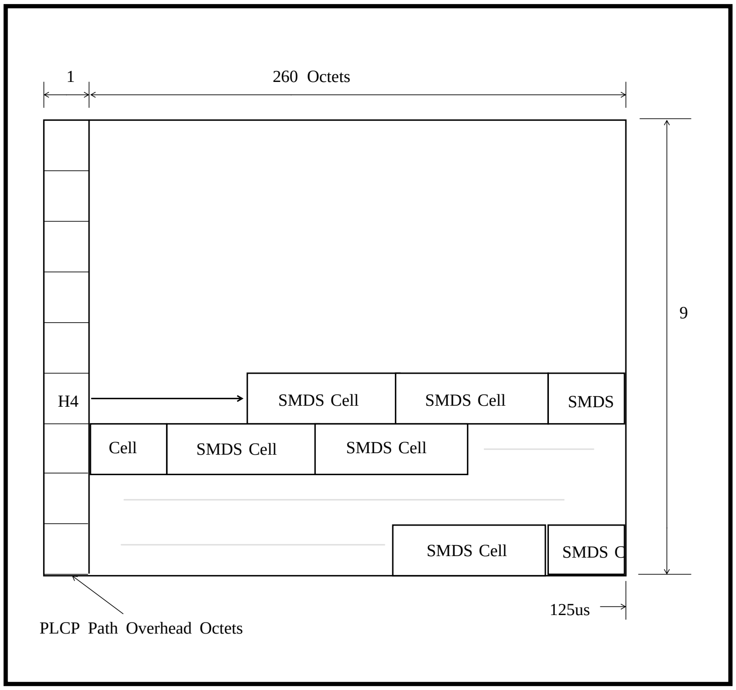

Figure 4: Proposed mapping of SMDS Cells into a SONET STS-3 frame [5].

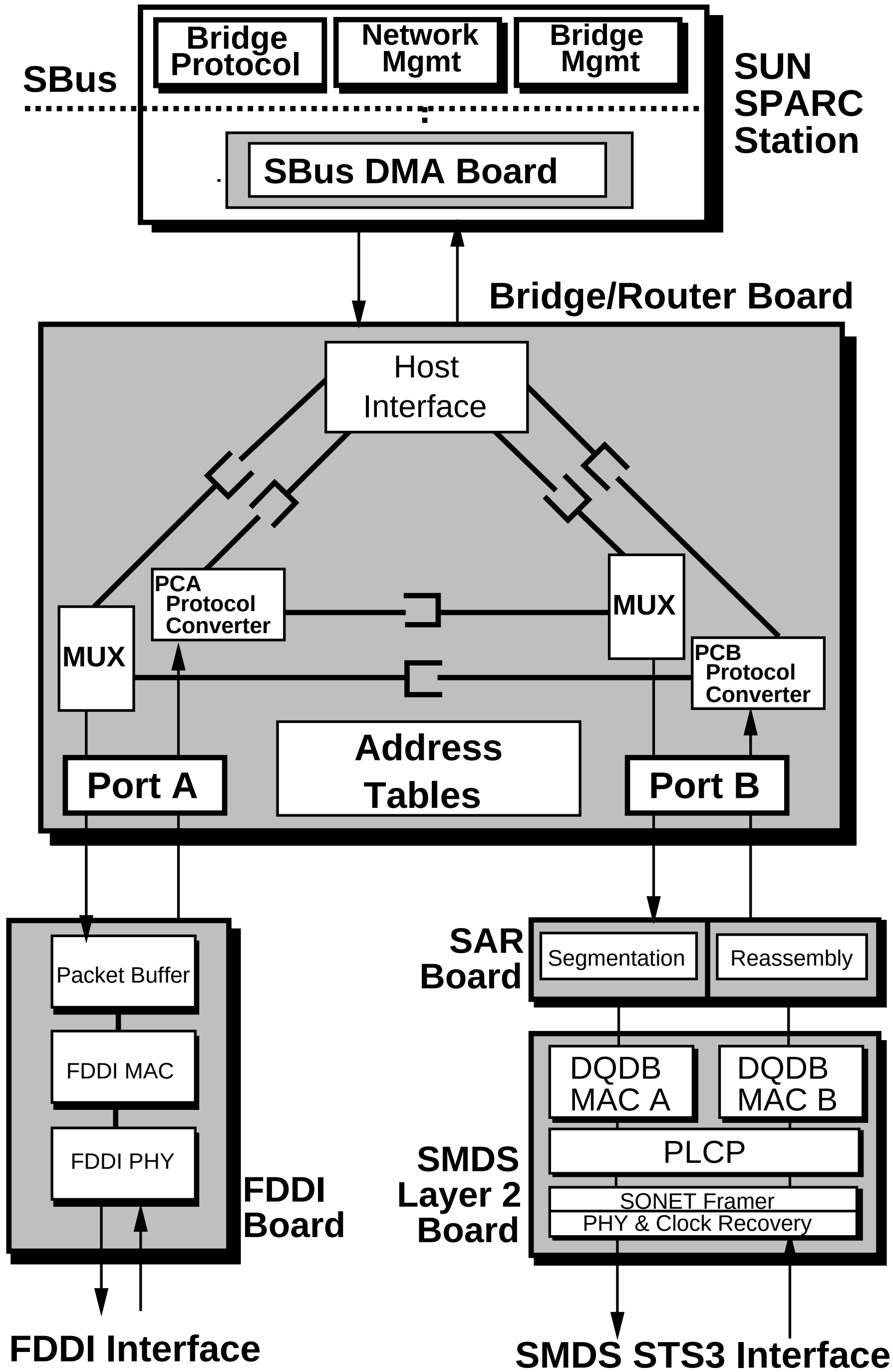

Figure 3: System Architecture showing main functional blocks.

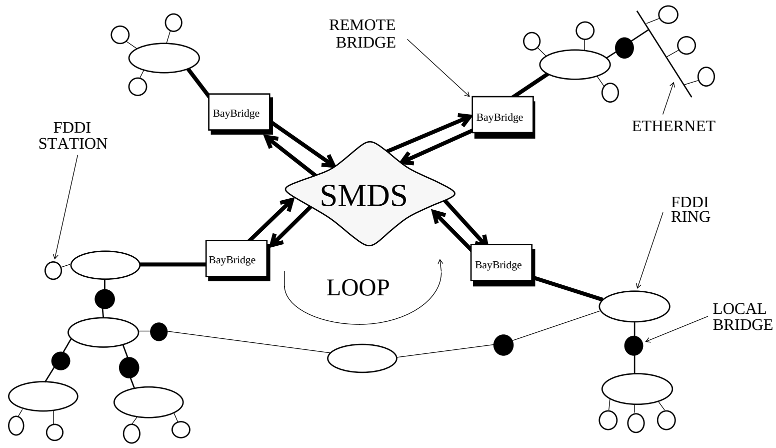

Figure 2: FDDI rings may be connected together — even if local loops exist in the topology.

2.1.1 SMDS Board

The main features of the SMDS board are:

- Operates at either DS3 or SONET STS-3 line rate.
- Supports Single CPE attachment to an SMDS network.
- Uses DQDB MAC chip developed as part of this project [8].
- Processes cells[2] in both directions at the maximum line rate.

The bridge will communicate with the local exchange using the standard DS3 or STS-3 framing format and standard DQDB cells conforming to the IEEE 802.6 Standard [4]. Figure 4 illustrates the framing format for SMDS operating at SONET STS-3 and figure 5 for DS3.

Incoming cells:
Cells received from the Physical Layer are first removed from the SONET or DS3 frame structure by the Physical Layer Convergence Protocol [4] (PLCP) and handed to the MAC. The DQDB MAC will accept all incoming BOM (Beginning of Message) cells that match any external MAC address[3], cells with Message

[2] In this project, we use the term "cell" to describe an SMDS L2_PDU (DQDB DMPDU), the term "message" to describe an SMDS L3_PDU (DQDB IMPDU) and the term "packet" to describe an FDDI packet.

[3] Default number of external addresses is 1024

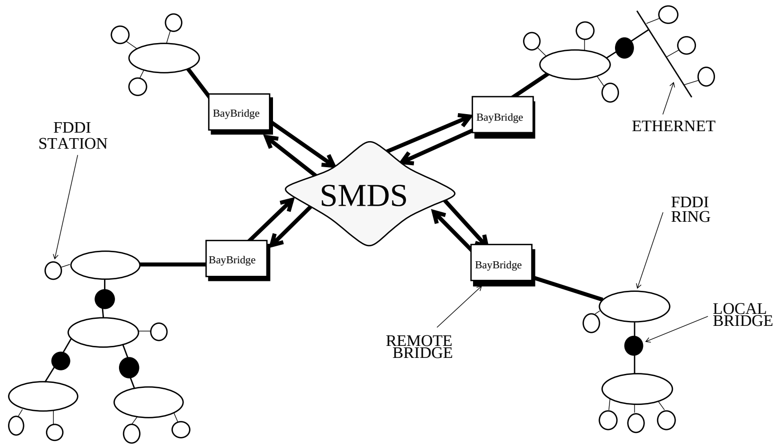

Figure 1: A typical topology interconnecting FDDI rings. Each FDDI ring may contain local bridges conforming to the IEEE 802.1d Standard.

and simple — the datapath consists simply of a 32-bit wide bus for transmit and receive and a "tag-bit" to delimit packets. There is an additional control bus driven from the host to set up and gather parameters from the system and for interrupts to report error conditions.

The bridging protocol will allow an arbitrary topology of remote bridges and FDDI rings. This allows two or more remote bridges to connect one local tree of FDDI rings to the public network for load-balancing or redundancy. Such a topology is shown in figure 2.

2 Overview

Figure 3 is a block diagram of the bridge architecture. The bridge consists of four main blocks: the Bridge Board, the SMDS Interface, the FDDI Interface and the Host Interface.

2.1 SMDS Interface

The SMDS Interface will consist of two circuit-boards: the SMDS Board and the Segmentation and Reassembly (SAR) Board.

1 Objectives

High Throughput. Bridging over 100,000 packets per second. Routing is expected to approach this rate.

Application. Intra-company, inter-campus communication.

SMDS rates. SONET STS-3 155Mbps, DS3 45Mbps.

Expandable Address Tables. The internal address tables hold 4096 address lookups by default, but may be expanded arbitrarily. Similarly the bridge may communicate with an arbitrarily large number of other bridges.

Flexible Configuration. The interfaces between the Bridge Board and each network card are identical. The bridge may be reconfigured as an FDDI-FDDI bridge, or other network cards may be readily added to the architecture.

Programmable Protocol Conversion. A custom designed protocol converter is used to convert between different protocols: this is the heart of the bridging/routing function of the system. The converter may be readily programmed to handle multiple MAC layers and routing protocols.

The aim of this project [7] is to build a high-throughput encapsulating remote bridge/router between an FDDI ring operating at 100Mbps and the public SMDS network operating at STS-3 SONET (155Mbps) [3] or DS3 (45Mbps) rate. Where possible, the bridge will comply with the proposed IEEE 802.1g Remote Bridge Draft Standard and the proposals of the IEEE 802.6 Multiport Bridge Committee [2, 1]. Although the system will support high speed hardware routing, this will not be the subject of this report and will be described in a later revision[1].

A typical application and topology is shown in figure 1. FDDI packets will be encapsulated into a single SMDS message and routed to one or more remote bridges. The bridge will be capable of operating at a sustained maximum throughput in excess of 100kpps. A host interface will provide an FDDI and SMDS interface to a local workstation. The workstation will also run the Spanning Tree Bridge Topology Protocol, the Network Management agent, the FDDI Station Management Process and will enable a network manager to set local and remote addresses as well as monitor network performance parameters.

It is important to note that the system does *not* use an internal shared bus. This is to reduce the amount of bus contention and to remove the need for a single very high speed bus. Instead, each interface to the Bridge Board is identical

[1] For more details of the routing capabilities of The Bay Bridge please contact the authors.

Contents

FDDI–SMDS Bridge Architecture

Nick McKeown

Abstract

The Bay Bridge project is implementing a high throughput bridge/router capable of bridging over 100kpps between FDDI and SMDS as a platform for the investigation of very high speed bridging and routing in hardware.

This document describes the architecture of the first implementation: an encapsulating remote bridge between an FDDI network operating at 100Mbps and an SMDS network operating at the SONET STS-3 or DS3 rate. Plug-in replacement SMDS interfaces will be bilt for each rate.

The architecture is split into four main blocks: the Bridge Board, the SMDS Interface, the FDDI Interface and the Host Interface.

Each architectural block is described and the bridging protocol summarised.

The Bay Bridge

Project Report: 1 Revision: 2.0

Department of Electrical Engineering
and Computer Sciences
University of California at Berkeley