The FDDI Pilot Project in Computer Science at the University of Calgary

David Hankinson, Bruce MacDonald, Jon Rokne, Brian Scowcroft

Computer Science Department, The University of Calgary
2500 University Drive NW, Calgary, Alberta T2N 1N4
Ph: (403) 220-6015, Fax: (403) 284-4707, Telex: 03821545 UNIV OF CGY
{dwh,bruce,rokne,bps}@cpsc.ucalgary.ca

Keywords: FDDI, networks, fiber optics, token rings

Abstract

Continued demands on the Computer Science Department's resources forced an unsatisfactory network topology. Sun Microsystem's implementation of FDDI was selected to replace the backbone, and is the first such installation in Western Canada. The paper reviews the hardware, software and topology of FDDI systems, explains the configuration chosen for the department, and discusses the installation. Each ring node requires a VME bus FDDI controller board. Physically the ring is star connected, providing convenient patch panel reconfiguration for maintenance and testing. All department file/client servers were attached to the ring, while groups of client workstations are controlled over sub-Ethernets. The resulting topology is a central backbone ring, with sub-Ethernets radiating from each FDDI node. FDDI's simplicity enables performance improvements in addition to the increased speed, so long as the network topology and systems configuration are carefully designed. Reliability is achieved by duplication of file systems on all ring nodes, so that any subnetwork may operate independently of the ring. We conclude by recommending that machine crashes should cause the FDDI board to *pass through* the data, rather than wrapping the dual rings and possibly causing unnecessary ring fragmentation. The paper discusses the relative merits of concentrators and optical bypass switches for fragmentation protection.

1 Introduction

The Computer Science Department maintains and operates a sophisticated UNIX computing system consisting of 9 fileservers, 75 workstations, 150 terminals and a variety of other research machines, previously distributed solely over an Ethernet network. In 1989 communication between the 8 existing fileservers was recognized as critical to the efficient operation of the facility. By then system growth necessitated by increased demands in teaching and research, dictated a fileserver connection topology that was necessarily unbalanced and included internal gateways. The failure of one server would often cause the eventual failure of the entire network. In addition Ethernet does not perform well when heavily loaded.

Forseeing further demands — such as continued enrollment increases, growth in research computing needs, an emerging trend for many machines distributed throughout the Department, and the resulting need for a better backbone network — a Department investigation selected the new Fiber Distributed Data Interface (FDDI) standard of bidirectional ring, fiber-optic interconnection as most suitable to provide a stable network configuration for the early 90's.

A small feasibility study with potential vendors found the most mature FDDI implementation to be provided by Sun Microsystems of Mountain View, California. Funding was already committed for a new fileserver and additional funds for FDDI were obtained within the University. The negotiations were successfully concluded and an order placed for the new fiber-optic network, with delivery and commissioning slated for the final quarter of 1990.

As the project has university-wide ramifications, the Department involved other units interested in digital fiber-optic communications: the office of the Vice President (Research), the Communications Department, Academic Computing Services, and Information Services, all of whom gave financial support.

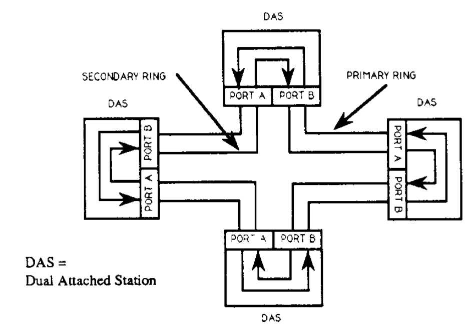

Figure 1: Dual Counter Rotating Ring.

This paper reviews the characteristics of FDDI, and discusses the hardware (cable, connectors, interface board, testing, and bypass options) and software (machine allocation, addressing, file system configuration, and routing issues).

2 Hardware

FDDI provides a high bandwidth (100Mb/sec) general purpose transmission link among computers and peripheral equipment using a 62.5 / 125 micron fiber as the transmission medium. An FDDI network is typically configured as a dual, counter-rotating ring with branch and tree links joined to the ring at concentrators. One primary use for an FDDI ring is a backbone local area network supporting heavy traffic flow. Typically the ring will interconnect local segments of Ethernet or other slower networks. FDDI may also be used as a front-end network supporting high speed graphics workstations or fileservers. Typically FDDI is made available at the desktop, perhaps in the future over twisted pair telephone cables.

The dual ring requires four fibers for each station, two for each ring (see Figure 1). If a cable or node fails the two rings can be joined or "wrapped" at the nodes on either side of the break to provide a single ring, as show in Figure 2. If two non-adjacent breaks occur then two isolated single ring networks are formed when each break is wrapped out. A concentrator, described below, centrally controls the ring and can effectively bypass any number of breaks while maintaining ring integrity. Optical bypass switches or board pass-through prevent ring fragmentation when a node's host ceases functioning.

2.1 Components

Controller board The Sun FDDI controller board, shown in Figure 3, is a triple height (9U) VME card with a 68020 processor, 256 KB data RAM, 256 KB of buffer space, an RS 232 port for monitoring the power up diagnostics, and an LED display for fault analysis.

During initialization neighboring stations exchange a stream of line status control symbols over the bidirectional links of the connection. This

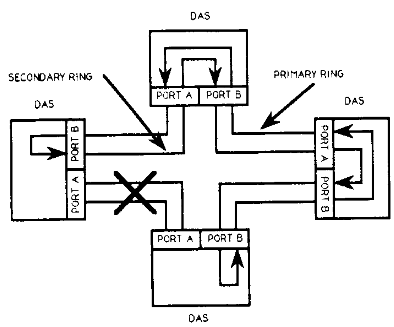

Figure 2: Dual Ring Exhibiting Self-Healing Wrap.

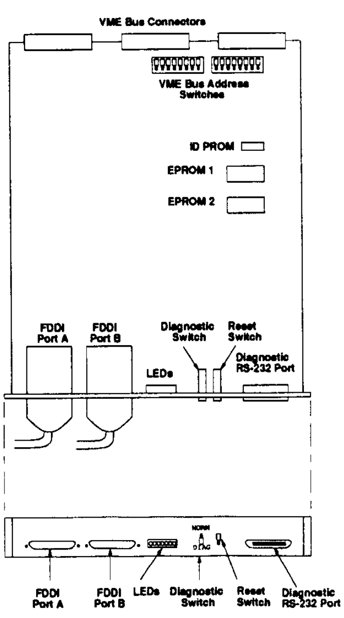

Figure 3: FDDI Controller Board.

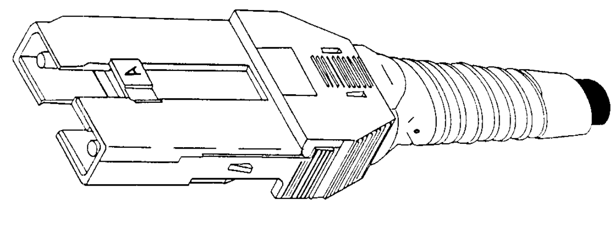

Figure 4: FDDI Connector.

confirms the quality of the interface before the board goes to an operational state. Data transmission is controlled by a Timed Token Protocol. Only one station generates data at a time, and the other stations repeat it. The destination station stores a local copy and passes the data on with a few bits changed to inform the sending station that the data was received. The sender then removes the data from the ring. When a station wishes to send a new token, or resend a lost one, it initiates a "claim" by sending a special "bid" data frame which proposes a value for the Token Rotation Time. When the next station receives the claim frame, it compares the received bid value with its own proposed bid. If the received bid is lower, the station repeats the claim frame. Otherwise it replaces the claim frame with one containing its own lower bid. If a station receives its own claim frame, it knows that it has won the claim, and may initiate a data token. If a claim frame does not arrive within an allotted time the station will assume there is a fault and begin sending a stream of beacon frames. All stations will repeat the beacon. If the originating station receives its own beacon it assumes the fault was transient and it will then begin the claim process. Otherwise, the last station sending a beacon knows the fault is between that station and its upstream neighbor. In this case, the sending station will time out and begin further recovery action.

FDDI cable The FDDI specification requires a 62.5 / 125 micron fiber with a modal bandwidth of 500MHz-km at 1300 nm (Table 1). The total system attenuation must be less than 11dB between active devices (each controller board is active). The recommended maximum attenuation for a typical cable is 2.5dB per km and the intent is that stations be placed up to 2km apart, allowing for connector losses. Our installation uses "Siecor" cable. The Media Interface Connector (MIC) is the standard FDDI connector (see Figure 4), and the specification controls the geometry across manufacturers. The connector is polarized and keyed to distinguish send/receive and dual/single attached connectors. There are four key types for the receptacle — A, B, S, and M — and six for the connector — A, B, S, M, AM, and BM — as shown in Table 2. The FDDI connectors are not user-installable and are generally purchased with both ends factory-connected, or as a single ended "flying lead" which can be terminated with an ST connector in the field. The ST connector shown in Figure 5 has become the industry standard fiber connector, as it is easy to install, cheap, durable and has excellent optical characteristics. No keying is available, however the connector is ideal for terminations at a patch panel where only skilled users are normally rerouting cables. The main advantages in our network are cost and user installability.

Wavelength	Maximum Attenuation		Minimum Bandwidth
	Loose Tube Cables	Tight Buffered Cables	All Cables
850 nm	3.5 dB/Km	3.75 dB/Km	160MHz·Km
1300 nm	1.1 dB/Km	1.75 dB/Km	500MHz·Km

Table 1: Optical Performance.

Connector tool kit Since the MIC connectors must be factory installed, the main field requirement is the termination of the trunk cables and possibly splicing, which may be mechanical or, more reliably, by fusion. The fusion splice tool kit is expensive and for applications where

Adjacent Port

Local Port		*A*	*B*	M	S
	A	X	V	VC	X
	B	V	X	VC	X
	M	VC	VC	P	V
	S	X	X	V	V

V Valid FDDI Connection
X Invalid FDDI Connection
VC Valid FDDI Connection where Port B is the Primary Connection taking preference over Port A
P Prohibited

Table 2: FDDI Connection Rules Matrix.

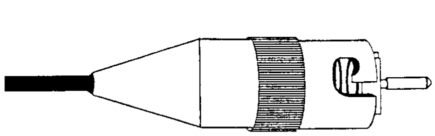

Figure 5: ST Connector.

loss is not a critical factor the mechanical splice will suffice, with a typical loss of about 0.2dB. As the Computer Science network did not require splicing, the only special tooling needed was for terminating the ST connectors, and a Siecor TKT-012 ST connector kit was purchased. This comprises cable strippers, a scribe, a cable polishing die, lapping films, a microscope and UV light curing equipment. Our technicians installed each connection in less than 10 minutes, with an optical loss less than the specified 0.2dB–0.5dB in all but one (rejected) case (it is sometimes impossible to detect a crack in the cable cut before the connector is attached).

Test equipment Siecor suggested testing the cable by shining a high intensity flashlight at one end and observing the result at the other! This simple method proved satisfactory for our installation. A Siecor CME 1000 optical attenuation test set was purchased as a University resource since the additional expense can be justified for larger installations and for fault finding once the cable has been commissioned. The set is a dual wavelength (850/1300 nm) LED/LASER transmitter and receiver with an LED display.

Optical bypass These switches are activated by an FDDI board in the event of a host failure or power loss. A high speed pivoting, spherical mirror causes the optical signal to be passed through passively on both rings (and thus prevents ring fragmentation should two or more non-adjacent hosts fail). The main disadvantages are the cost ($1,350 each), and the 2.5dB loss in the active state, which usually limits the number of adjacent active switches to three. Also, board bypass is ineffective against cable failures.

Concentrators Also called *collectors*, these are active controllers that connect large numbers of devices to the ring. A typical configuration has the concentrator dual-attached to the ring and provides multiple single attached branches. The main advantage is *not* the reduced cost of the optical cable, but the increased tolerance to cable or host failures that would normally cause ring fragmentation. They would also be useful for supplying FDDI connections to individual office workstations. There is no need for bypass switches in this configuration, single attached workstations may be bypassed electronically by the concentrator. The main disadvantage is the high capital cost and the catastrophic loss of the complete subnetwork in the event of concentrator failure.

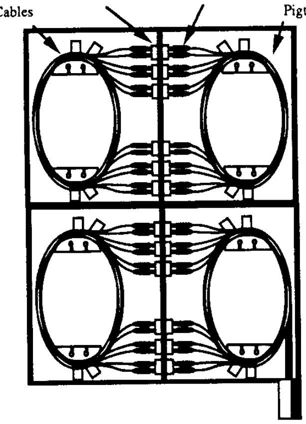

Figure 6: FDDI Patch Panel.

2.2 Physical Layout

In a typical campus, the FDDI backbone would be configured as a physical ring extending throughout the campus. When buildings or rooms are close to each other, as in our case, the ring can be physically configured in star form. This provides flexibility for cable management, future expansion and ease of fault isolation, and allows building-level isolation of inactive machines which would otherwise require optical bypass switches. In our installation there are two machine rooms, four floors apart. There is a patch panel in each room, with a 12 strand fan-out fiber cable (see Figure 6). Only four fibers were required to connect the two rooms, but the low incremental cost of a 12 fiber cable allowed provision for spare lines and future expansion, such as future Ethernet connections. All cables were installed by Computer Science technicians with technical expertise provided by the Communications department. Closely buffered cables were installed throughout and no splices were required.

3 Software

FDDI software is straightforward to install. A new kernel must be built with the proper device driver, the many utility programs must be installed, and manual entries reformatted. While the vendor-supplied automatic procedures eliminate much of the work, our specialized machine configuration has long since dictated a policy of inhouse system installation. So FDDI software was manually installed, but this was still an easy task.

The controller board configuration optionally allows for a user selected MAC (Media Access Control) address, which may be considered equivalent to an Ethernet address. These 48 bit numbers are assigned by the IEEE and we felt there was no need to change them. Sun Microsystems recommends the IDROM addresses, which are the default.

Less straightforward is the required reorganization of the network. Important points include:

1. Which machines go onto the FDDI ring?
2. Internet Protocol (IP) addressing and subnetwork issues.
3. How will the Network Information Service (NIS) be configured, if at all?
4. Routing/gateway issues.

3.1 Machine Allocation

The selection of machines for direct connection to the FDDI ring is crucial for overall system and network performance. Computer Science has several categories of machines:

Diskless client Has no disks and must be attached to a network in order to function.

Dataless client May have several disks but stores only system files (a complete or partial set), and does not store user data.

Diskfull client A standalone client that stores user data.

Fileserver A complete standalone machine that may serve other client machines, and contains the complete system environment.

Only the fileservers are connected to the FDDI backbone, as currently the only controller boards are 9U sized VME bus modules and fit only the larger machines. Also, each Ethernet subnetwork is controlled from a fileserver, so a *client/server* model is strictly followed. The fileservers are allocated in the central logical ring shown in Figure 7. In comparison, the previous fileserver connections were a series of sub-Ethernets.

The FDDI ring is well-balanced, there are no breaks nor extra gateways when communicating between fileservers and much more intelligent use of the on-board Ethernet controllers is achieved. Previously, extra gateways split the network in an attempt to alleviate performance problems. In practice network splitting is detrimental in our environment because all machines share a common name-space. The subnetworks became unbalanced, resulting in an exceptionally high load on one subnet and negligible load on others. Machines from one subnetwork have difficulty transmitting data to another as they must rely on the gateway machine, causing congestion. If the gateway crashes the network becomes fragmented, and in general NFS does not function well through gateways.

Other advantages of the new scheme include:

- The Timed Token Protocol evenly distributes load over the entire network reducing congestion commonly associated with Ethernet or bus type backbones. The physical bandwidth (100 Mb/sec) contributes enormously to reducing delays when accessing data.
- The dual-attach nodes dynamically configure themselves in the case of software failure (one node maximum). Single fileserver failures affect only the subnetworks which attach directly to them, *not* the entire network.
- Each fileserver acts as a boot server only for the attached subnetworks, reducing network hops.

Computer Science machines are carefully configured so that all locally maintained software is duplicated on each fileserver. While this may seem wasteful of disk space, the redundancy has several advantages.

- Local subnetworks receive their data from the parent server. Only in very specialized cases (mail and home directories) is this not followed.
- Network stability and error prediction is greatly simplified. If a parent fileserver crashes, then all its client workstations are rendered unusable until the machine is brought back on line. A crash of a non-parent fileserver will not bring down another subnetwork.
- Bandwidth is conserved throughout the backbone and Ethernet subnetworks, as traffic is as localized as possible.
- The entire network gains more independence when the ring topology is used. Since the ring is more reliable the network itself is more robust.

3.2 IP Addressing

An IP address refers to a software configurable rendezvous point for a particular *interface*. Two physical interfaces on the same machine will have different IP addresses. FDDI presents itself to the UNIX kernel as just another interface, and it appears just below the "link-layer" in the OSI protocol. So FDDI and Ethernet may transparently coexist within the networking framework. There are three differences, the interface names must be different, the FDDI topology is a closed ring and the maximum transmission unit is 4500 bytes rather than Ethernet's 1500 bytes. Otherwise FDDI software administration is *exactly* the same as for Ethernet. A machine with two or more interfaces is a gateway, and these are discussed fully in Section 3.4.

IP addresses are assigned to each node as the ring is commissioned. Each interface has a *hostname* which maps directly to its underlying IP address, and mappings are stored in **/etc/hosts**. A unified naming convention is desirable when configuring a network with many gateways, and we chose a simple scheme in which a host's FDDI interface has the same name as the host, while the Ethernet subnet is named "<host>**-gate**." This is very effective in reducing potentially catastrophic administration errors at the interface level, for example when mounting filesystems and building diskless clients, and is important in presenting a simple network model to users. A *netmask* is required for subnets, but as most large installations already have subnetting, this is not a problem.

When FDDI is initially installed, the IP addresses may change, meaning that the system administrator must be exceptionally careful with the **/etc/hosts** file and the use of **ifconfig(8)** (interface configure). Otherwise entire subnetworks may be rendered unreachable and the only resort is to reboot machines. It cannot be overemphasized that if IP addresses change *all machines must receive a new copy of the host database*, so they know who they are on reboot. MAC address clashes result in single ARP (Address Resolution Protocol) entries for duplicate stations, causing disaster.

3.3 NIS Configuration

The Computer Science UNIX machines provide a homogeneous user environment. No matter where a user logs in, the same environment is presented. Users find this is a considerable advantage, but it is not without technical difficulty. Proper configuration of the NIS is essential for such a level of transparency. Since NIS is a packet or datagram oriented facility, information does not pass through gateways automatically. Instead **ypserv** is run on all machines through which information is to flow, and **ypbind** on all machines that wish to interface with NIS databases. There is one *master* machine from which all source databases originate.

The proper solution in our department is to run the NIS master and slave servers only on the FDDI ring. This requires a subtle but relatively painless change to the boot procedure's interface configuration order. All fileservers on FDDI run **ypserv** and thus have their own set of NIS databases.

3.4 Routing/gateway issues

Routing under FDDI is the same as any other local area network although the topology is a closed ring, which is irrelevant in routing, except indirectly in that it will influence routing by reducing hop counts. Examining the complete network topology in Figure 7, it is easy to see that, in the absence of failures, the maximum number of hops required between any two machines is two. Further, the standard routing protocols used by the dynamic routing daemons choose routes based on hop counts. So as traffic passes across our network from place to place *all* inter-fileserver communication is via FDDI. Only the lesser populated Ethernet subnetworks utilize gateways to route to their destination. The FDDI ring has the tremendous benefit of making the network more symmetrical, which also aids in calculating how information will travel. We do not use the standard dynamic routing program, **routed**, but instead **gated**, a program more adaptable to multiple gateway protocols.

It is also easy to see from Figure 7 that *static* routing could be used for all non-fileserver machines. Static routing is just a "default" route pointing to a predetermined destination. It requires no daemons and is set up at boot time. For all diskless clients, static routing is enabled by default, conserving resources for those machines with less capacity. It may be implemented for all dataless workstations on the Ethernet subnetworks if performance requirements warrant it.

4 Discussion

No significant problems were encountered during the installation of the hardware. Two 12 conductor cables were pulled and terminated without mishap. Care was taken to ensure that the cable was not stressed and the bend radii were kept to a maximum. The cables between the patch panels and FDDI boards were purchased terminated from Siecor and were installed under the machine room sub-floor. The main problem foreseen for the present hardware is the possible fragmentation of the network if two or more non-adjacent servers crash. At present, manufacturers set their boards into wrap mode if the host and FDDI board fail to communicate. It is our belief that in this case the preferred default action would be for the board to remain in the ring and pass on all data, although specific actions should be site-configurable.

5 Conclusion

The FDDI configuration has resulted in a complete, robust and transparent networking solution, shown in Figure 7. FDDI is an excellent solution to modern networking problems created by more powerful, more numerous and more widely distributed servers and desktop workstations.

Acknowledgements

Tim Bliek, Gerald Vaselenak, Mark Stadel, Ken Stauffer, Robert Fridman, Theo DeRaadt, and Mark James also contributed to the installation of FDDI.

References

[1] Siecor, "FDDI Network Cabling Design Guide," (draft) Siecor, Hickory, NC.

[2] Sun Microsystems, Inc., "SunNet FDDI/DX Controller Administration Guide," Mountain View, CA.

[3] Sun Microsystems, Inc., "SunNet FDDI/DX Controller Card Configuration and Installation Manual," Mountain View, CA.

[4] Swastek, Mary Rose, Vereeke, David J. and Scherbarth, "Migrating to FDDI," Sun Microsystems, Inc., Mountain View, CA.

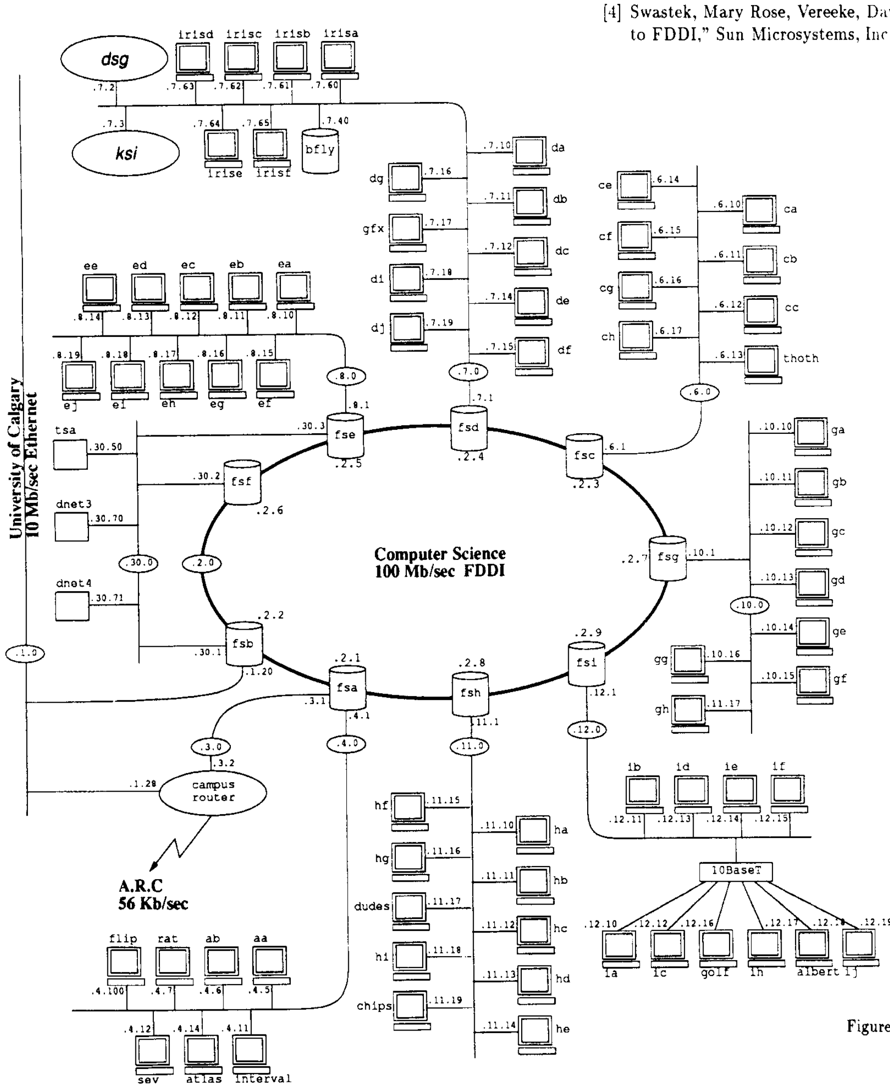

Figure 7: Complete Network Topology.