Cambridge HSLAN Protocol Review. Position Paper.

David J. Greaves*
Ian D. Wilson†

March 1989

The Cambridge Fast Ring (CFR) local area network operates at 75 Mbit/s and forms the basis for much of our high-speed protocol research. The CFR mini-packet contains 16 bit source and destination addresses and 256 bits of data. Our primary block assembly/disassembly protocol is UDL (Unison Data Link). This carries higher level protocols such as Unity RPC, local protocols for file transfer, bootstrapping etc and TCP/IP. UDL is also being used experimentally for voice, real-time video and high-speed packet switching.

Unison, an Alvey collaborative project, together with Olivetti Research, has developed a CFR/UDL/RPC based protocol suite for wide-area site interconnection using 2 Mbit ISDN links and an architecture for address translation at domain boundaries. Without this, the sixteen bit addresses would be a limitation. Protocol performance over local and wide area conditions is presented.

Ongoing research is assessing UDL performance under multimedia traffic in areas such as MAC layer bridges, metropolitan area networks and workstation operating systems.

1 Introduction

The general aim of HSLAN protocol research in Cambridge, both at Olivetti Research Ltd and at the University Computer Laboratory, is to establish the benefits of a single mini-packet type for all classes of traffic, and to examine the advantages which result from the short MAC layer address field of 16 bits. The immediate

*Olivetti Research Ltd & University of Cambridge Computer Laboratory. Email: djg@cl.cam.ac.uk

†Olivetti Research Ltd.

advantages are that the short packet gives fine-grain sharing and hence very low network access delay and the routing field is small enough to serve as the index to a directly mapped array at bridges and gateways.

The mini-packet, or *cell* as the new terminology would have it, that we are using was first used for the Cambridge Fast Ring (CFR) local area network [1]. The CFR mini-packet is of fixed size and contains 16 bit source and destination addresses and 256 bits of data. This position paper reports on some of the projects that are making use of this mini-packet format and the protocols that are used. The projects are:

- Cambridge Fast Ring – Local Area Network
- Cambridge Fast Ring MAC layer Bridges
- Cambridge Backbone Ring – Metropolitan Area Network
- Project Unison – Wide Area / ISDN
- Cambridge Fast Packet Switch – Metro and Wide Area B-ISDN
- Project Pandora – Multimedia workstations
- Metrobridge Project – Transparent ethernet bridge

and the protocols are:

- UDL (Unison Data Link)
- Unity RPC
- MSNL/MSDL (Multi-service network/data link) see companion paper by McAuley
- Real-time voice protocol
- Real-time video protocol

2 Projects

2.1 Cambridge Fast Ring

The Cambridge Fast Ring (CFR) is a ring topology local-area network operating at 75 Mbit/second [1]. It uses the empty slot access protocol. Each slot can contain

a mini-packet (cell) consisting of a 16 bit source address, a 16 bit destination address and 256 bits of data. Stations transmit into a passing empty slot. The contents are then copied by the receiver and a response is marked. When the full slot returns to its sender, it is freed and passed on to the next station in order to ensure fair access. If there were a transmission error, detected by a CRC at the end of each slot, or if the response shows that the destination was unable to receive, then the transmitting station automatically schedules a retransmission, up to a programmable retry limit.

The length of a slot, including all control fields, is 38 bytes. The data portion is 32 bytes, giving a mini-packet efficiency of 84 percent. The ring is electrically extended at its monitor station in order to hold an integral number of slots plus a gap of 3 bytes. This is the minimum required for correct synchronization. Ignoring the gap, the 75 MHz implementation of the CFR offers a raw bandwidth of 63 Mbit/second. A typical, large ring with N=35 stations and 2 km of cable, would contain 1400 bits delay in the stations and 750 bits delay in the cables. This would carry 7 full slots and leave a gap of 3 bytes. With Q slots, transmitting stations are can transmit every Q+2 slot times, so the maximum point-to-point bandwidth available on such a ring is 7 Mbit/second. The guaranteed minimum bandwidth available to each station, even at maximum load, is 1.75 Mbit/second. At maximum loading, the utilization is $N/(N+1)$ which is 97 percent and the time before an empty slot arrives and services a waiting mini-packet is limited to $N+1$ slot times, which is $(N+1)/Q = 5.1$ ring revolution times.

2.2 CFR MAC layer bridges

The CFR access chip is able to operate in a mode designed to support MAC layer bridging between two CFRs. For the simplest type of bridge, one access chip sits on each ring and they are connected back-to-back via a packet-copying finite state machine. Bridge mode access chips look-up the destination addresses of passing mini-packets in a 64 K bit-map in order to decide whether to receive.

The current bridges use a combination of hardware and software to achieve approximately 10 Mbit/second full-duplex between two rings [2]. The software controls the routing table and other management functions and it also generates the buffer addresses required for each mini-packet transfer. The buffer addresses are generated in advance by the software and stored in two-level deep hardware FIFOs until required by the copying hardware. The software is written in C over Tripos [3] on 4 MIPs Acorn RISC machines (ARMs) [4]. One feature of the Acorn processor is very low interrupt latency (less than a microsecond) with dedicated CPU registers for interrupt routines. This enables the bridge software to take an interrupt for every mini-packet transfer in order to keep the FIFOs full.

The bridges can be programmed to operate in a mode where they implement a strict queue management policy. Although the buffers are allocated from a shared pool, there is an associative store, tagged by the 32 bit composite source/destination routing field, which holds the number of packets that are internally buffered for each possible path through the bridge. The bridge refuses to receive any further mini-packets with a given source/destination pair if the number of packets buffered with that pair exceeds a software controlled limit. The hardware response protocol of the CFR then exerts backpressure, possibly through multiple previous bridges, right back to the traffic source.

The bridges incorporate further hardware connected directly to the ring. This circuitry detects special scout mini-packets generated by other bridges which can be used for automatic generation of the routing tables. The bridge hardware uses two semi-custom chips, and a complete bridge (including the ARM computer, bridge chips, buffers and two CFR interfaces) fits onto three double-height eurocards.

These bridges operate in the MAC layer and can support all data-link and network level protocols. The hybrid hardware/software implementation offers considerable flexibility in the way errors and congestion are controlled and ongoing research will assess these issues.

2.3 Cambridge Backbone Ring

The Backbone Ring again uses the empty slot protocol [5]. Although not yet operational, it is designed to operate at 1 Gbit/second and cover an area 50 km in diameter. Owing to the several hundred slots present on such a network, the backbone stations are able to transmit in more than one slot per ring revolution. However, they still pass-on-free the slots that they have used to provide load balancing. The backbone mini-packet format is compatible with the CFR. Both offer the same interface to data-link layer protocols known as the M-access interface and specified in [6].

2.4 Project Unison

Project Unison is a collaborative project which is part of the British Alvey program. It proposed and implemented an architecture for interconnection of ATM networks (mainly LANs) between remote sites over ISDN links. At the lowest level, the basic unit of exchange was the CFR mini-packet. This was selected primarily for its size, but also so that the packet switch required at each site in order to provide concentrator-like local access to the ISDN links could be implemented with a CFR cluster. A CFR cluster is simply a complete CFR LAN with the station chips connected in a byte-wide parallel ring within a single 19 inch rack. The cluster was

termed the exchange ring and the one at Cambridge had two stations connected to ISDN 'ramps' to other sites and several 'portals' which provided access to the LANs. Three LAN were used: ethernet, CFR and the old 10 Mbit/second Cambridge Ring.

The Unison architecture was the first to implement translation of the CFR address field as CFR packets traverse the network. The translation was done by the ramps which divided the full network into separate addressing domains. The approach was to divide the 16 bit CFR address into two fields, domain and station, each of 8 bits. The distributed management software was then able to make any remote domain appear in a so-called 'window' at the local site whereby mini-packets addressed with the appropriate domain field passed through the window and eventually appeared at the destination site.

The switching nodes within the Unison architecture (ramps, portals etc.) were obliged to inspect the UDL headers of the mini-packets in order to offer expedited transfer of real-time traffic. This was because expedited transfer was indicated by a partition in the port identifier space. However alternative approaches are now being considered where the 16 bit routing fields do not have a rigidly partitioned domain field and the least significant bit, say, of the 16 bit field indicates expedited transfer. This would enable expedited transfer to be applied sparingly on the critical parts of the virtual circuit under control of the intermediate switches. The true CFR destination address would, as always, be inserted by the last translation of the path.

2.5 Cambridge Fast Packet Switch

The Cambridge Fast Packet Switch (FPS) is based on a self-routing multi-stage interconnection network with no buffering in the switch fabric [7]. This particular design has been selected as a basis for research owing to its simplicity and excellent simulated performance for all traffic types. The switch makes use of a priority field in the header of each packet to expedite the transfer of real-time messages. The switch is able to accept any amount of low priority traffic without significant degradation to the expedited traffic. Experimental gate-array implementations of a 4 by 4 element and an I/O controller have been constructed in order to test the feasibility, but no practical networks are yet operational. The simulation work has shown very good performance with multimedia traffic, a two-plane implementation offering close to the cross-bar ideal performance. The FPS simulations have emphasized the importance of short, fixed-length mini-packets. This area of research is rapidly expanding at Cambridge, and in time we hope to experiment with all of the CFR oriented protocols running over the Fast Packet Switch.

2.6 Project Pandora

Project Pandora is a joint Computer Lab/Olivetti Research programme. Its aim is to investigate multimedia integration for workstations, servers and networks. Particular emphasis will be placed on the structure of operating system kernels which support multimedia management. The Pandora workstation hardware consists of a Unix PC with CFR attachment and a Pandora Box. The Pandora Box contains peripherals, including a frame-store, video camera, video compression unit, loudspeaker and a standard telephone interface.

2.7 Metrobridge Project

The Metrobridge project is being undertaken by Olivetti Research to provide MAC layer interconnection of ethernets. The Metrobridge consists of a PC-type card frame which has a byte-wide parallel CFR running across the backplane. The backplane has four segments of PC bus, plugged into which can be up to four standard ethernet LAN adaptors. Adjacent to each ethernet card is a CFR/Transputer card which plugs into the parallel ring and drives the PC bus segment. The ethernet packets are segmented into UDL by the Transputers and transmitted over the CFR to their destination ethernet.

The Metrobridge contains two additional circuit boards: one is the CFR monitor station and the other is the expansion serialiser. The monitor board contains the CFR monitor, a ring extender which pads the electrical delay to an integral number of slots, and a master monitor Transputer. The expansion board optionally converts the parallel CFR to serial format so that a serial connection can be made between remotely situated Metrobridge units.

3 Unison Data Link Protocol (UDL)

The primary block fragmentation/reassembly protocol used over the CFR is Unison Data Link [8]. UDL provides an unreliable datagram service between peer entities, and is designed to be suitable for a wide variety of applications, real-time or otherwise. UDL operates over light-weight bidirectional virtual circuits known as 'associations', which encapsulate the necessary addressing, routing and priority information for peer communication. Associations are set up out-of-band by a 'secretary' service, of which there is one per addressing domain. In order to allow for associations to be created which cross addressing domains, secretaries communicate with gateway services and with each other, to ensure that the necessary translation tables are set up. UDL usually sits on the CFR/CBR M-access interface, although it is now being considered for other applications [9].

Protocol bit (1)	Start bit (1)	Port number (14)	RID (6)	Flags (2)	Seq Number (8)	Data: 0-28 bytes

Figure 1: The UDL mini-packet format

UDL allows collections of CFR mini-packets to be handled as a single block of data. The first 4 bytes of each mini-packet is taken up with UDL header information, the remaining 28 bytes being used for data. Encoded within the UDL header are a 'port number' used for identifying the mini-packet with a particular association, and a 'reassembly identifier' which identifies all the constituent packets of a single UDL block. Also encoded within each mini-packet is a 'sequence number' identifying its position within the UDL block. Valid sequence numbers are in the range 1 to 255, allowing UDL blocks to be from 28 bytes to 7K bytes in length.

Although UDL is general purpose, simple to implement and relatively efficient in terms of its bandwidth usage, it has drawbacks when used for certain types of application. Firstly, because there is no 'size' field within the UDL header, data is always handled in multiples of 28 bytes: an amount which is inconvenient when handling, say, FIFOs or disc blocks, whose size is invariably a power of two. Secondly, the reassembly identifier and sequence number fields in the UDL header allow for the detection of certain types of error (for example, lost mini-packet or mini-packet out of order), but because there is no low-level mechanism to ask for retransmission of missing data, erroneous blocks usually have to be discarded in their entirety.

For certain applications (particularly those implemented in hardware) a subset of UDL has been defined, which lessens the effects mentioned above, while still ensuring that the encoding of mini-packets is compatible with the CFR gateway services. Only the port field of the UDL header is used, thus allowing 30 bytes of user data. Not only does this improve bandwidth usage, but it allows the fragmentation and reassembly of blocks to be application specific, and omitted altogether if not required.

3.1 UDL performance: Local Area

CFR UDL drivers have been implemented on many systems, including those listed in table 1. The throughput performance depends mainly on the type of interface to the CFR access chip and the interrupt latency of the machine. The transputer and the VME interfaces present the CFR mini-packet as eight 32 bit words while the others only offer 16 bit access. A typical Unix host is not prepared to take an

Processor	Language	Project	Operating system
Transputer	Occam	Unison	-
68000/VME	Assembly	Unison	Tripos
ARM	C	Olivetti	Tripos
ARM	C	Olivetti	Unix
Sun 3	C	Olivetti	Unix
Olivetti PC	?	Ulster	MS-DOS
Transputer	Occam	Pandora	-
Transputer	Occam	Metrobridge	-

Table 1: UDL implementations

interrupt for every CFR mini-packet that arrives or departs and even in Tripos, which has comparatively light-weight tasks, such an overhead imposes too great a load. Two approaches are viable: either use a dally loop in the interrupt service routine so that interrupt return is normally delayed until the UDL packet has been completely reassembled, or use a separate network interface processor. The current generation of DMA controllers does not offer a viable solution since these are unable to accomplish the UDL specific action of looking up the port identifier of each mini-packet under circumstances of multiplexed receptions.

Figure 2 presents half and full-duplex UDL performance measured between two ARMs. These 4 MIPS processors were running Tripos with dally-loop style drivers written in C. The figures where generated with a trivial program at the transmitter which continuously queues UDL SDUs on the CFR/UDL driver. A similar program was required at the receiver to throw away the arriving packets. Two context swaps are required at both the transmitter and the receiver for each UDL SDU that is transmitted. Regression of the time per block against the number of mini-packets per block reveals that these points lie on straight lines. The half-duplex experiment has a gradient of 78 microseconds per mini-packet and intercept of 413 microseconds. The full-duplex experiment has a gradient of 147 microseconds per mini-packet and intercept of 606 microseconds. The network throughput is slightly greater for full-duplex operation than half-duplex. The operating system overhead of the dispatch and rescheduling was assessed by modifying the program so that it queued an unrecognised command on the driver. The time for the driver to spot that the command packet was in error and return it to the client task without transmitting it on the network was measured as 150 microseconds.

3.2 UDL performance: Wide Area

Some wide-area experience of UDL is available from the Unison project. Detailed measurements of delay, jitter and contention are presented in [10]. The Unison

	Half-duplex.		Full-duplex.		
Mini-packets per block	Blocks per second	Megabit per second	Blocks per second	Megabit per second	Equivalent MBit/sec
255	49	2.82	26	1.5	3.0
128	95	2.72	50	1.43	2.86
64	182	2.62	99	1.41	2.81
32	339	2.43	189	1.36	2.72
16	606	2.17	351	1.26	2.52
8	980	1.76	588	1.05	2.10
4	1429	1.28	870	0.78	1.56
2	1852	0.83	1163	0.52	1.04
1	2326	0.52	1351	0.3	0.6

Figure 2: ARM/Tripos to ARM/Tripos UDL performance

ramps introduced about 1 milli-second of delay per mini-packet at both the receiving and transmitting ends and could handle about 7000 mini-packets per second, which with 28 data bytes each, results in 1960 kilobits/second. These results are dominated by the interface to the ISDN links which operated at only 2 Mbit/second and not by the CFR or UDL. The main contribution of the Unison project was the overall architecture for wide-area interconnection and the protocol suite.

4 Higher level protocols and performance: Local area

Remote procedure call over UDL has been accomplished with the Unity RPC protocol [11]. This provides 'exactly once', 'at least once' and 'at most once' semantics to the client and is implemented using datagrams with 48 byte headers. UDL provides the datagram service. The header alone requires a minimum UDL SDU of two mini-packets, but if the marshalled user data fits into the remaining 8 bytes, then the complete RPC fits into the two mini-packets.

Figure 3 shows the time for a null RPC, that is, one with no arguments, for the Tripos/C ARM systems. In order to assess the network overhead the RPC speed was measured both over the CFR to a remote machine and to the local machine. The 'at least once' and 'at most once' semantics require the same amount of processing (in the absence of errors) while the 'exactly once' calls take longer owing to the overhead of the explicit termination. As with most layered protocols, there is a good deal of redundancy when Unity is run over UDL, and in the future we are hoping to collapse the protocol stack and make use of the UDL association

	Remote call	Local call (No network activity)
Exactly once	3.12 ms	2.8 ms
At least once and At most once	2.25 ms	1.9 ms

Figure 3: Unity null RPC performance: local and remote

to implement an RPC of equivalent power which, for a small number of arguments, will fit in a single mini-packet.

5 TCP/IP

TCP/IP runs over UDL. The UDL layer is required since the CFR mini-packet is not large enough to contain an IP datagram – it will not even hold the header. For Unix, mounting a new network interface requires a modified kernel. This has been done for several processors and the resulting TCP and NFS throughput has been comparable with the ethernet.

6 Real-time voice

Real-time voice can be transmitted over the CFR with great ease owing to the guaranteed minimum bandwidth available to all stations. Voice is sent with 16 bytes per UDL SDU resulting in one mini-packet UDL datagrams. This is 2 ms of speech which is about the quantity which can be tolerated if lost.

Real-time voice over the MAC layer bridges has not been tried in earnest, but may not be as easy since the CFR does not guarantee receiver bandwidth.

7 Real-time video

Real-time video is transmitted using modified UDL. The first two bytes of the standard four byte UDL header are still required since they contain the port number. However, in the experimental protocols, the second two have been replaced with the row and column offsets for the picture data which is contained in the remaining 28 bytes of the mini-packet.

Current real-time video experience over the CFRs has been limited to 128 by 128 four bit pixels transmitted at 25 frames per second. This corresponds to 1.6 Megabit/second and gives a grainy monochrome picture. The CFR can carry more than 10 full-duplex video connections at this rate and the quality is remarkably acceptable for videophone type applications but it is not sufficient for stills. Stills are sent as data using a transport level protocol over UDL at much greater resolution. The Pandora project will soon be onstream and much higher quality video is envisaged.

8 Dedicated network interface processor

The Pandora project has developed a network interface processor (NIP) to offload the UDL fragmentation overhead from the workstation's central processor. This contains two Inmos Transputers and presents the UDL service over the SCSI bus to host. The performance of the UDL NIP is over 5 Mbit/second half-duplex.

Acknowledgments

This position paper reports work by Joe Dixon, David Greaves, Andy Hopper, Ian Leslie, Derek (Mac) McAuley, David Milway, Roger Needham, Peter Newman, John Porter, Brian Robertson, Steve Temple, David Tennenhouse, Roy Want, Ian Wilson and many more. Project Unison is an Alvey collaborative project between Acorn Computers, University of Cambridge, Logica UK, Loughborough University of Technology and the SERC Rutherford Appleton Laboratory. Further Unison work was done by Olivetti Research. The CFR was developed by Acorn Computers, University Computer Laboratory and Olivetti Research. The Backbone Ring project was funded by SERC and Olivetti Research. The Cambridge Fast Packet Switch is funded by SERC. The Pandora hardware was funded by Olivetti Research.

References

[1] 'The Cambridge Fast Ring Networking System.' A Hopper and RM Needham. IEEE transactions on computers, Vol 37 no 10. October 1988.

[2] 'MAC layer interconnection of homogeneous LANS.' JD Porter. University of Cambridge Computer Laboratory, PhD Thesis. 1989.

[3] 'Tripos: A portable operating system for minicomputers.' M Richards, AR Aylward, P Bond, RD Evans and BJ Knight. Software Practice and Experience 9 no 7. 1979.

[4] 'The Archimedes User Guide.' Acorn Computers 1987.

[5] 'The Cambridge backbone network.' DJ Greaves and A Hopper. Proceedings European Fibre Optic Conference (EFOC / LAN 88) Amsterdam, June 1988.

[6] 'CFR M-access specification.' AM Chambers. Project Unison Document Reference UC021. 5th November 1986.

[7] 'The Cambridge Fast Packet Switch.' 'A fast packet switch for the integrated services backbone network' P Newman. IEEE J Selected Areas Communications, SAC-6 No 9. December 1988.

[8] 'The Unison data link protocol specification.' DL Tennenhouse. Project Unison document ref UC021. October 1986.

[9] 'High Speed Interneting.' Companion paper, this workshop.

[10] 'Site Interconnection and the Exchange Architecture.' DL Tennenhouse. University of Cambridge, Computer Laboratory PhD dissertation September 1988.

[11] 'Unity: An RPC mechanism.' DR McAuley. Project Unison document ref UC027. March 1987.