Data Handling Division
DD/87/12
April 1987

CERN's approach to MAC-level bridging of LANs

J. Joosten and C. Piney

(Presented at FOC/LAN '87, 3-5 June 1987, Basel)

DD-vf

CERN's approach to MAC-level bridging of LANs

Abstract:

In this paper we explain how a system of transparently bridged Ethernets has been built, using our existing high performance network (Cernet) as a backbone.

The rationale for this approach is that the CERN site is too large to be covered by a single Ethernet and that many different protocols are being used. The use of filtering MAC-level bridges solves these problems, with the advantage that local traffic on a segment is not passed on to other segments.

Address resolution is controlled by dynamically updated tables in the bridges; the routing of packets between segments is done by the backbone. The look-up of the tables has to be done very efficiently since they are consulted twice for every packet. This has been achieved by doing the look-up in hardware using a binary search algorithm.

Currently 13 LANs are bridged via Cernet, thus giving all the users the impression that they share one large local area network.

Traffic is growing rapidly and Cernet will be saturated in the not too distant future, requiring the installation of a new backbone, which will probably be based on a ring in the 100 Mbit/s range.

1. INTRODUCTION

When the first Ethernet was installed at CERN in 1983, networking of computers was generally done with Cernet [1]. Cernet is a site-wide, mesh-structured, packet switching network, designed and built at CERN in 1975. The communication hardware allows for line speeds of upto 8.5 Mbps, which is one of the CCITT G703 transmission standards and allows for a practically unlimited span.

The communication subnet provides a datagram service, and software installed on the connected machines provides the services for Transport and higher levels. Thus connectivity was obtained between mainframes (IBM, Cyber) and over 100 minicomputers (VAX, PDP, HP, Norsk Data).

The trend to personnel computers, workstations, terminal concentrators etc., together with the emergence of international standards for Local Area Networks (LANs) [2] to which these devices tend to be connected, forced us to solve the problem of internetworking physically distinct LANs (because of distance) and the use of different protocols on these LAN's (TCP/IP, DECNET, XNS etc.).

Since no commercial products were available to solve our problem, it was decided to use the existing Cernet infra-structure as a vehicle (backbone) for interconnecting the LANs in a transparent (protocol indepedent) way.

This article will describe how this has been achieved by making use of the concept of bridges.

2. MAC-LEVEL BRIDGING OVER A NETWORK

A bridge provides a means of propagating a basic information frame from the source to the destination network, in such a way that the information content is unchanged. This operation is constrained by the OSI model to being performed in the lower sublayer of the Link layer (level 2 in the OSI model), whence the complete term "Medium Access Control (MAC)-level Bridge".

This approach to intercommunication allows any mix of higher-level protocols to operate within the resulting bridged area network in the same way as within a single local area network. In addition, a bridge can carry out "filtering" on the destination address. This entails forwarding to each segment only the frames whose destination corresponds to an address on that segment.

Work on the MAC-level service definition is being done within the IEEE project 802 [3] and is expected to be incorporated into ISO 8880 and ISO 8802/1.

The transparent interconnection of LANs with bridges implies generally the use of a Network level service (level 3 in the OSI model). Only in very special cases can a set of segments be interconnected with bridges, without resorting to a network level service. See figure 1.

2.1 Header manipulation

In order to support communication over a network, the header of the packet has to undergo some manipulations. The most obvious approach in our case would be to encapsulate the Ethernet frame in a Cernet Network level packet and forward it to Cernet. However, one of our design choices was to minimize the dependence on actual Cernet and Ethernet characteristics. For this reason, it was decided to carry the LAN data by means of the ISO connectionless Network protocol [4]. In addition, to ensure that the data part of the Network packet carries only the data part of the LAN frame, the information of the Ethernet MAC-level header is copied into the ISO Network header. Thus, changing the backbone network and mapping between different LANs is made as straightforward as possible.

With the conventions described above, the bridge has to perform the following operations:

- On input from Ethernet the MAC-level header gets stripped off and the source address, the destination address and the type identifier are incorporated in the ISO Network header as well as the local segment identifier (see figure 2). The ISO header is prefixed to the data field of the packet. This again is prefixed with a Cernet Network header, which contains a Cernet address, corresponding to the destination Ethernet segment. The resulting packet is forwarded to Cernet.

- On reception from Cernet, the Cernet header is discarded and the the ISO Network header is replaced by the Ethernet MAC-level header, putting into it the relevant information extracted from the ISO header. Then the packet is forwarded to the Ethernet segment.

2.2 Address resolution

In order to decide if a packet is local or to be forwarded over the network, the bridge has to obtain the correspondence between the address of a station and the segment to which it is connected. This correspondence is obtained from a table inside the bridge which is maintained dynamically.

2.2.1 Table structure

The address table is used to give the correspondence between LAN addresses and LAN segment identifiers. Each entry is 64 bits wide, being composed of the 48 bit LAN address and a 16 bit segment qualifier (see figure 3). The top bit ("indicator bit") of the segment qualifier is zero if the remaining 15 bits hold the actual segment identifier, otherwise it is set to identify a special entry in the table (see appendix A). The Ethernet addresses in the table are ordered by increasing weight and searched using a binary search algorithm. Thus, a table with 8k addresses is searched in at most 13 steps.

2.2.2 Table updating

The address information both in the Ethernet packet header and in the ISO Network header of the packets carrying the data across the backbone is used to update, when necessary, the address table. Appendix A explains how this is done.

3. FRIGATE: THE MAC-LEVEL BRIDGE IMPLEMENTATION

The site-wide interconnection of Ethernets over Cernet as a backbone has been achieved within the context of the "Frigate" project (Flexible, Reconfigurable Internet GATEway). The Automatic Internet Datagram (AID) service, installed in a Frigate, provides the functionality of a filtering MAC-level bridge. The hardware and software used for this application consists of a mixture of commercial and home made modules.

3.1 Frigate hardware

The Frigate hardware is based on the VME standard [5] and consists of the following modules (see figure 4):

- A commercially available processor board with a Motorola 68000 and 512 kbyte of dual-port memory, bus arbitration logic and a terminal port.

- A BInary Search Engine (Bise), which allows for a hardware look-up of addresses as descibed earlier. The Bise accepts from the cpu the three words forming the 48 bit MAC-level address, and returns, in the next read cycle, the 16 bit segment qualifier (see figure 3). The search is generally terminated before this read cycle is executed. If the search is not succesful, i.e. the looked for address is not present in the table, an all zero qualifier and a pointer are returned to the cpu, which then has the task of inserting the address in the right position in the table. This is a time consuming process, since all the entries in the table beyond the pointer have to be shifted down, but it happens very infrequently. There is space for 8k 48bit addresses.

- A commercially available Ethernet interface with a dual-bus DMA capability (VME and VMX), capable of running in universal transmit and promiscuous reception mode.

- A Cernet interface, via which the Internet packets are transmitted to and received from the Cernet backbone. Coupled to the Cernet interface is a transmission line driver, which allows for transmission at 2.5 Mbps over 1.5 km of good quality twisted pair cable. It is also possible to transmit at 2 and 8.5 Mbps using the G703 standard, giving a virtually unlimited transmission distance.

- A DMA interface, which looks after the transmission of data between the Cernet interface and the memory.

Not all of the hardware descibed above is required to perform the AID service. The system can function without the Bise or DMA, but will of course be less performant. At start-up, the software will detect which modules are missing and configure so as to be able to perform the required application.

3.2 Frigate software

The MAC-level bridge service for Ethernets, using Cernet as a backbone, connects to Cernet as a normal subscriber and to the Ethernet segment as a station with specific characteristics (promiscuous reception and universal transmission).

The software is basically composed of an Ethernet driver, a Cernet driver and a dispatcher. The operating system used is RMS68K [6].

Apart from handling the I/O of the Ethernet interface, the Ethernet driver also does the address filtering and the destination identification (by looking up the Bise table). The reason is that a large proportion of the packets received from the Ethernet do not have to be forwarded and can therefore be rejected directly by the driver. This approach optimises the performance of the system, which must be very efficient in discarding of received packets.

The Cernet driver looks after the transfer of packets to and from Cernet.

The dispatcher task controls the correct operation of the drivers and manages the overall bridging functions. The main actions are: buffer allocation for Cernet input; buffer deallocation; address table maintenance; packet header manipulation; Cernet Link control; transit of packets between the the LAN and the Cernet backbone.

As explained before, "packet header manipulation" implies preparing frames input from Ethernet for transmission across Cernet, and processing packets input from Cernet for presentation to the local Ethernet.

4. CURRENT STATUS AND PLANS

There are currently 13 bridges installed on as many Ethernet segments. Thus we have succesfully created a site-wide "virtual Ethernet", over which many different protocols are being used (DECnet, TCP/IP, Cernet, XNS, LAT, etc.) by as many as 130 active stations.

Per second over 2000 frames can be filtered and 30 frames can be forwarded to, as well as received from Cernet. The switching capacity of a Cernet node is 170 packets/s, indicating that the AID software can still be optimised. Steps in order to optimise the code are being taken and we expect a considerable performance increase. Another possibility to augment the performance is the introduction of a 68020 based cpu.

However, with the limited capacity of Cernet and the progressive introduction of clusters, which will produce a lot of traffic, steps have to be taken to prevent the flooding of the system. The first measure is to make sure that clusters are connected on the same physical segment and therefore only load the filtering capability of the Frigate, but not the backbone. A next step is to filter on the "Type" field, specific to certain clusters. This filter can already handle 7000 frames per second. If the filtering capacity is still not adequate, clusters might have to be isolated with commercially available inter-Ethernet bridges, which can filter and forward at the full Ethernet speed.

It is obvious the the capacity of Cernet will be the biggest limitation and therefore steps have to be taken to remove that bottleneck. The intention is to replace Cernet by a new high performance backbone, progressively and in a way transparent to the user. The solution could be the use of a fast ring such as the 100 Mbs FDDI standard, proposed by ANSI X3T9.5. However, it is currently unclear wether chips and products will be available in time. We therefore acquired an existing commercial product in the same speed range and do a pilot project to see if sufficient throughput can be obtained. Work on this project has been started.

5. ACKNOWLEDGEMENTS

The following persons gave a major contribution to the Frigate hardware and software development: J. Anthonioz-Blanc, S. Brobecker, D. Davids, P Dunon, A. Ludtke and J. Rochez.

6. CONCLUSION

By implementing MAC-level bridges, using the existing Cernet packet switching network as a backbone, we were able to provide a sitewide "vitual Ethernet" for a large variety of computers and protocols.

Although the filtering capacity of the bridges is probably adequate, the limiting factor will be the switching capacity of Cernet.

However, since the system has been designed to be as independent as possible from the current technology, the progressive replacement of Cernet by a new, high performance backbone should be possible without affecting the end users.

We expect that the future 100 Mbs FDDI token ring will be able to fulfill the requirements.

APPENDIX A

ADDRESS TABLE UPDATE

Below are described all the cases in which the address information in the Ethernet header and the ISO Network header is used to update the address table.

A.1 Source address handling

On input from the backbone, the source LAN address is known to reside on the segment corresponding to the value of the LAN segment identifier given in the ISO Network header (see figure 2). On input from the LAN, the source address is obviously "local" – i.e. this address can be accessed through this bridge.

A.2 Destination address handling

– > *Group addressing*

A destination address may represent a single destination or it may specify a group of destinations (i.e. broadcast or multicast address). One specific bit is reserved by ISO in the destination address, to indicate group addressing [2]. Packets with such addresses must be forwarded across the backbone to all connected LANs – via their bridges. This also has as effect that the address of any station sending to a group-address will become known to all active bridges.

– > *Known destination*

An individually addressed destination can either be known or not in the address table. If this destination is known and "local", the frame is ignored. If the destination is remote, the corresponding Network level packet can be created and forwarded as described earlier.

– > *Unknown destination*

If the address is unknown, the same procedure is used as for a group-addressed destination, i.e. forwarding to all connected LANs. The addressed destination, if it exists, will receive the message; its response will inform the bridge corresponding to the initiating segment as to the location of this address.

This whole procedure is only valid if the addressed station does respond. It is ineffective in the case of a communications protocol which is unacknowledged at all levels. Luckily, this is not the case for any of the protocols currently running on the site.

Since it is possible that an addressed destination is missing or inoperable, the procedure has an additional safeguard: an entry is made in the address table when a frame for an unknown destination is received on the LAN. This entry has the indicator bit (figure 3) set and contains the time at which the information was forwarded. While this indicator is set, no more information will be forwarded to this destination for ten seconds. This entry is simply overwritten on receipt of a response from the corresponding address.

REFERENCES

[1] J. M. Gerard, "CERNET – A High-speed Packet-Switching Network" CERN 81–12, Dec. 1981

[2] International Standards Organisation, International Standards IS 8802/3, IS 8802/4, IS 8802/5.

[3] IEEE Computer Society, "MAC and MAC-Bridge Service Definition", document H12, Feb. 1985

[4] ISO "The Protocol for Providing the Connectionless Network Service", International Standard IS 8473

[5] Motorola "The VMEbus specification".

[6] Motorola "M68000 Family Real-Time Multitasking Software User's Manual", M68KRMS68K/D8

CONTENTS

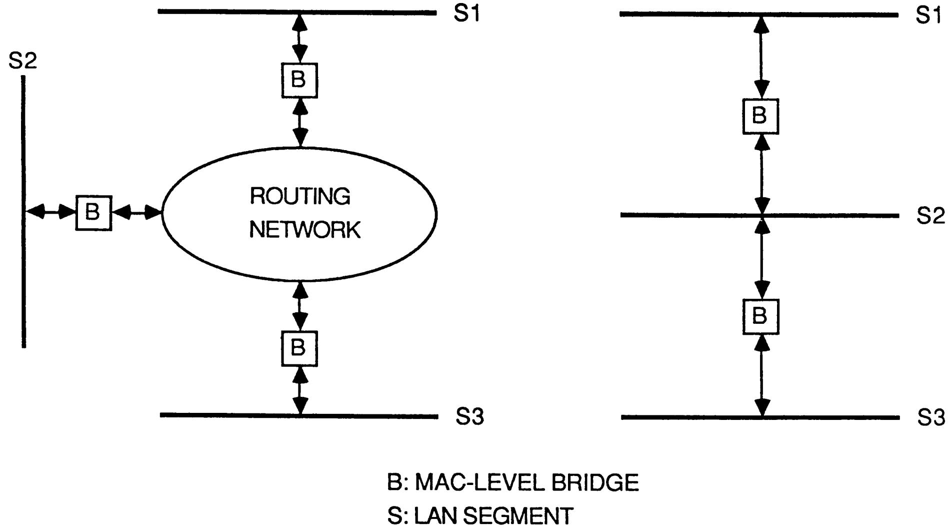

FIGURE 1: BRIDGING WITH OR WITHOUT NETWORK

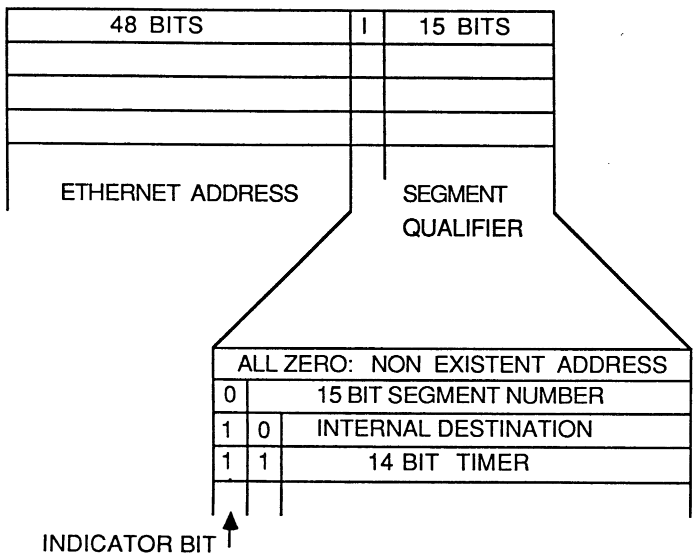

FIGURE 3: ADDRESS LOOK-UP TABLE LAYOUT

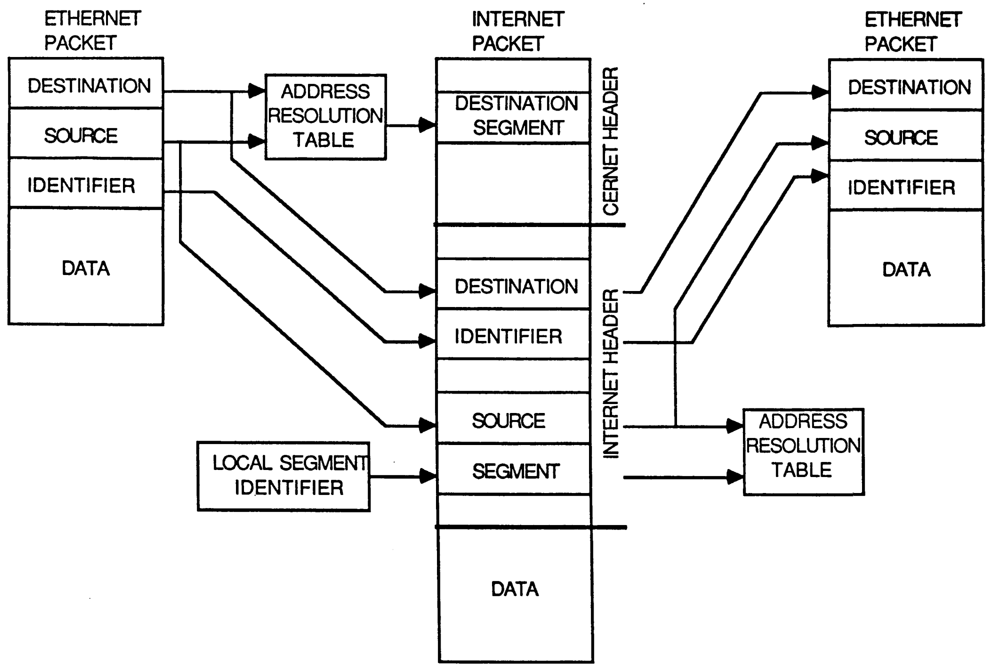

FIGURE 2: PACKET HEADER MANIPULATION

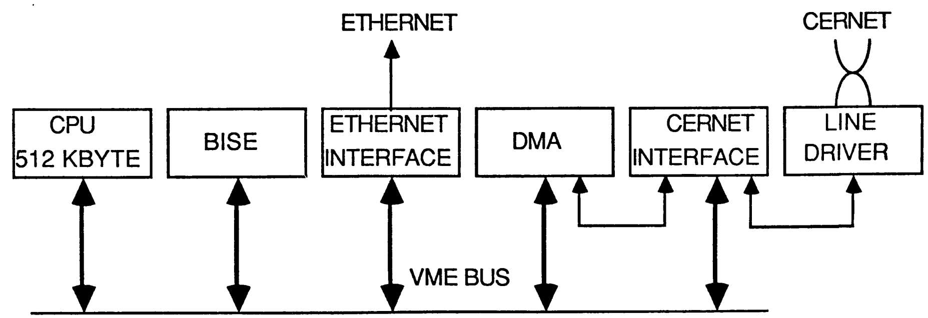

FIGURE 4: FRIGATE HARDWARE CONFIGURATION