MIN

A Metropolitan Interconnection Network

P. Coltelli, M. Mannocci, L. Rizzo, F. Tarini, P. Zini

Technical note C87-6[1]

© CNUCE - Pisa, March-April 1987

[1] Work funded by C.N.R. as a part of the "Computer Networks Strategic Project" with contributions by Olivetti Italia. Minor contributions have been given later on by Honeywell Information Systems Italia and by Siemens Italia.

Preface

MIN is a Metropolitan Area Network which tries to extend to the environment of a whole town services presently available only in a local environment; the design is aimed to allow also future developments and improvements of such services to be extended to the same environment.

The MIN network is being developed as a subproject of the "Computer Network Strategic Project", funded by C.N.R. with main contributions by Olivetti Italia.
A prototype implementation of MIN is being carried on in Pisa; its completion depends on whether C.N.R. and Olivetti will continue funding the project. Up to now, neither the C.N.R. has yet decided to continue funding the "Strategic Projects" after 1986, nor it is clear if similar activities can be continued with some other type of funds; moreover, the Olivetti has not yet officially approved its informal agreement with the C.N.R., in the context of which past contributions have been given.

The subproject is co-ordinated by CNUCE and co-operating research units presently exist at the following locations:

- Centro Studi Telecomunicazioni Spaziali, Milano, engaged for 18 man-month per year, directed by Prof. L. Fratta;
- CNUCE-C.N.R., Pisa, engaged for 27 man-month per year, directed by Dr. F. Tarini;
- Dipartimento di Scienze Fisiche, Universita' di Cagliari, engaged for 6 man-month per year, directed by Prof. G. Pegna;
- Istituto di Elettronica e Telecomunicazioni, Universita' di Pisa, engaged for 5 man-month per year, directed by Prof. G. Frosini.

This technical note contains the pre-prints of two papers to be presented at "EFOC/LAN.87", which will be held in Basel, Switzerland, June 3-5, 1987. The first paper ("A LAN approach to MANs") outlines the architecture of the network and its prototype implementation; the second one ("High Performance Bridge for Ethernet-like Networks") describes caracteristics and implementation of the bridge designed to connect the metropolitan backbone to the local networks.

A LAN approach to MANs[1]

by
P. Coltelli - M. Mannocci - F. Tarini - P. Zini

CNUCE - An Institute of C.N.R. - Pisa - Italy

Abstract

A metropolitan network mostly built with technology and components typical of local networks is presented. The shared bus architecture, the bit rate and the packet format are taken from the IEEE 802.3 standard. High level protocol implementations and network services available for Ethernet environments are used too. On the contrary, a different channel-access protocol and possibly different communication media are used.
The main purpose is a transparent interconnection among local networks distributed all around a town, so that services typically available in local environment can be extended allover a town. A prototype of the network here presented is being implemented in Pisa.

1. Introduction

At present and/or in the near future, most computers (spanning from Personals to Mainframes) are conveniently connected to each other through LANs, mainly when they are installed in the same building or in contiguous ones. The performances of LANs and suitable software packages often lead to the definition of Distributed Systems, whose advantages over the collection of single computers are well known.
We feel that the same advantages can be gained even when computers to be connected are physically distributed in a wider area, such as a town; in other words, we feel that computer communications in a metropolitan environment can grow to levels of transparency and feasibility comparable to local ones; and this is the goal of the work here presented.

2. Requirements

Considering the goal, requirements are quite obvious and can be summarized saying that the resulting Metropolitan Network must behave as far as possible like a LAN, even spanning over longer distances and even interconnecting LANs.

1 This work is a part of the "Progetto Strategico Reti", supported by the C.N.R. ("Consiglio Nazionale delle Ricerche"), and had a partial support by Olivetti - Italy.

Besides the low cost/performance rate, the following characteristics are desired:

- feasible interconnection of LANs and possibly of single hosts;
- transfer rate and delay comparable to those of a single LAN;
- transparent addressability, that is no need of special addressing, regardless of the mutual position of sender and receiver, either on the same or on two different interconnected LANs, or directly on the MAN;
- distributed control and graceful degradation;
- generality, both as multi-vendor open architecture and as feasible interconnection of different types of LANs;
- inexpensive, easy to install communication media.

3. Architecture

The above requirements suggest the overall architecture of the MAN to be based on a two level structure made by a metropolitan backbone network, to which single LANs are connected (fig.1). Communications across the boundary LAN/MAN are managed by suitable BRIDGEs or GATEWAYs; every host on each connected LAN is then automatically in MAN, even if (in special cases) a single host can be directly connected to the backbone.

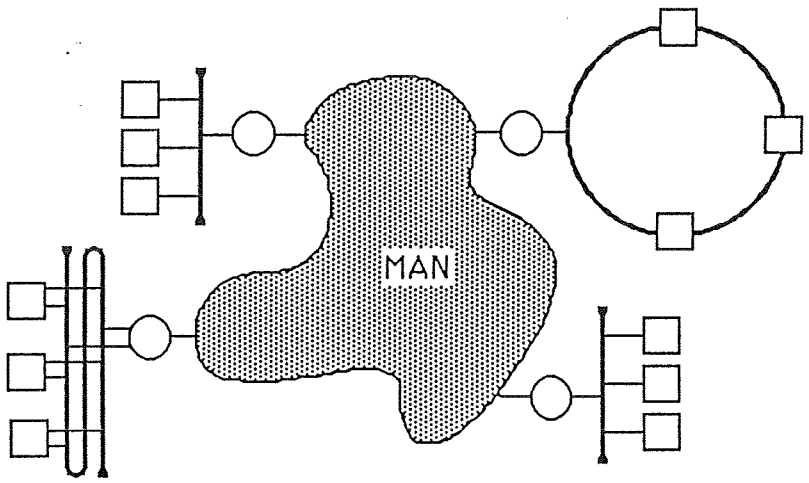

Fig. 1

By this structure, already existing LANs can remain unchanged and the MAN is not choked by local traffic. Moreover, a single connection on the MAN can be used by many hosts, thus reducing costs of connection and increasing the total number of possibly connected hosts.

Because of its physical structure, this MAN has been called **MIN**, standing for Metropolitan Interconnection Network; MIN stands also for **min**imizing resources to

gain desired results.

For the sake of reliability, simplicity and feasibility in the target environment, MIN is a BUS network (fig. 2) and its physical interfaces specifications, bit rate, packet format and addressing structure comply with Ethernet[5] specifications. However, a different channel access protocol is used to guarantee good performances even when operating on longer distances than supported by the CSMA/CD protocol.

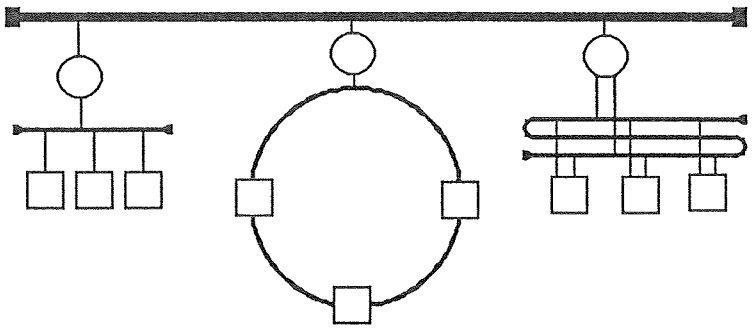

Fig. 2

3.1 Bridges and gateways

Generally speaking, the choice between BRIDGE and GATEWAY can vary for each different type of LAN to be connected. According to a widely accepted agreement, a GATEWAY acts at level 3 of the ISO-OSI reference model (NETWORK), while a BRIDGE is limited to level 2 (DATA LINK): the boundary between them can be seen in whether or not routing is performed; a selective repeater is meant to be a bridge, even if packet selection is based on destination addresses.

Cost/performance rate and maximum feasible performance are obvious advantages of bridges over gateways, provided that interconnected nets are "compatible" enough, e.g. about packet formats and maximum packet lengths. This means that such characteristics of the MIN has to be chosen in order to maximize its compatibility towards existing LANs, thus maximizing use of bridges. Unfortunately, despite international activities about standardization, existing LANs belong to several different types and their compatibility is often poor. Therefore, characteristics of the MIN have been chosen to maximize its compatibility versus the most widely used LANs, i.e. those complying with the IEEE 802.3 standard and possibly with the 802.5 one.

3.2 Physical media

The physical bus must comply with given specifications, such as attenuation, trip delay and bit rate; on the other hand, in such a complex environment as the target one, no single type of physical communication media can guarantee its availability in all the different subsets of the area to be served. As an example, high costs and/or restricting regulations could make practically impossible to put cables with given characteristics in specific routes.

For these reasons, MIN's bus is made by a set of segments, each possibly made by a different medium, connected to each other by repeaters (fig. 3). Up to now, several media have been considered, including coaxial cables, fiber optics, laser-beam bridges and microwave bridges.

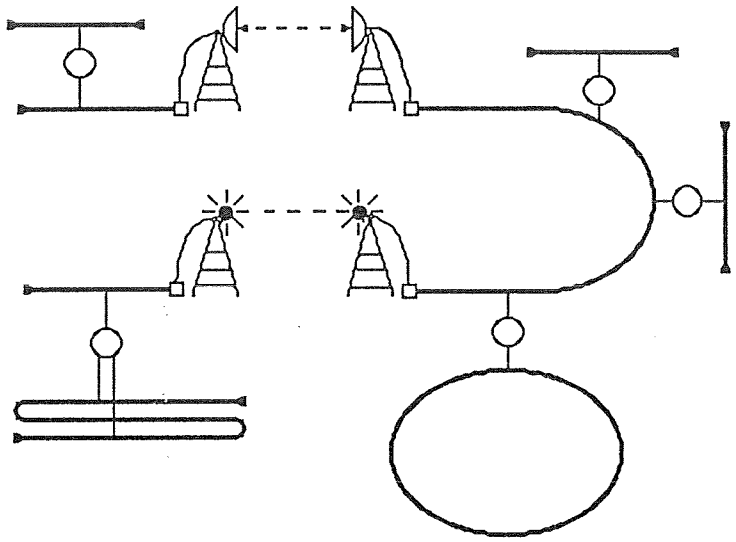

Fig. 3

Due to the wide diffusion of Ethernet-like LANs, bus segments made of coaxial cable just need to be bought and installed.

Microwave digital links are also widely used, even if not exactly for MAN communication purposes; in most weather conditions their reliability is very good, but it must be noted that ether is strongly crowded in the most interesting frequencies, restrictive regulations exist and the directionality of the communication is poor.

Laser-beam links are not prone to these limitations and offer interesting characteristics about bit rate, cost and easy installation, but their utilization is much more limited by weather conditions, mainly by fog.

A reasonable conclusion is that both of these technologies must be available while building a given MAN; as the microwave's behavior for data communications is well known, the laser-beam technology is experimented in the first prototype of the MIN.

The operating behavior of laser beams in fog conditions depends on the ratio between the used wavelength and the size of fog drops; when this ratio is near to one, the beam is strongly attenuated by diffraction and a communication with acceptable error rate can be led to distances not longer than two or three times the corresponding optical visibility, as stated by meteorologic services.
Available laser systems span from ultra-violet to far infra-red; as an example, a CO_2 laser operating at 10.6 μm can be used, providing that the detector is kept at very low temperature; but, in heavy fog, water drop sizes can reach 50-100 micron. Any laser system up to now available is thus prone to fog. The maximum distance on which a given system is reasonably usable depends on local weather conditions and, of course, on project requirements about service continuity. When requirements are strong, an optimal solution can probably adopt a twin system, made by both a laser-beam and a microwave link, so that they can backup each other in different weather conditions. Such twin links could be limited in power, as each component is not expected to operate in its worst weather conditions.

3.3 Channel access scheme

The choice of channel access scheme must comply with network physical architecture and to the required compatibilities.
For use on bus networks, mainly two families of channel access schemes are available, respectively based on ordered and contention (or random) access.

Contention-based access schemes are not actually usable in our target environment, since their bandwidth utilization strongly decreases with the ratio between duration of a single packet and the maximum signal propagation time. As both parameters are bound to the context, it results from simple calculations that a MIN can not be a CSMA/CD network, even if this choice would greatly help compatibilities.

Among ordered access schemes, choice is practically between token and virtual token. In this choice, complexity of the target environment must be taken into account, mainly because a network possibly prone to system errors due to lost or corruption of tokens would be very difficult to manage and maintain. Thus, a virtual token scheme has to be adopted; among them, the **L-Express** channel access protocol has been chosen.

L-Expressnet[2] was first proposed, verified and prototyped as the communication subnetwork of the **C-net** loosely connected distributed system, which was developed as a part of the **P.F.I.** (Progetto Finalizzato Informatica), funded by C.N.R.
Its access mechanism is based on measures of silence time intervals on the channel; stations are numbered in ascending order according to their position on the channel and the silence time intervals are divided in short slots, each lasting a few bit-transmission

times, say **k** seconds. Each station is assigned a given time slot, according to its ordering number, **i**; when the **i**-th station has to send a packet, before beginning transmission it counts silence time intervals up to **i** times **k**.
The overall effect is that a sequence of packets is formed on the bus, ordered according to their source and separated from each other by one or more silence slots; this sequence propagates in the "ascending" direction (that is towards left or right, this depends on an arbitrary initial choice) maintaining its consistency, namely the order among packets and their distance, thus forming a "train". Of course, each packet propagates in the reverse direction, too, and it can reach all stations, including "previous" ones; in the reverse direction distances between packet are not conserved, since they also depend on (variable) inter-station propagation delays, but this does not cause collisions because a station is not allowed to send a new packet during the same train.
A new train is started once the round is complete, that is after each station has got its access right. The end-of-train event is detected by measuring an uninterrupted silence gap lasting longer than **k** times the maximum number of active stations, plus maximum propagation delay. After detecting this event, each station **i** waits a time proportional to **i** before starting a new train. Providing that suitable constants are chosen, it can be proved that only the lowest-ordered active station will actually start a new train.
The cold start algorithm acts in a similar way.

Although the plain use of L-Express is in a bus with a linear shape, (the name itself indicates this fact, **L** standing for Linear) it has been shown[4] that networks with a tree topology can easily adopt this access protocol. This fact can allow optimal solutions for complex configurations, possibly occurring in a metropolitan environment.

Delay analysis, simulation results and measures made on prototype networks show for L-Express a very good bandwidth utilization and throughput-delay behavior, even when the total channel length is much longer than allowed by Ethernet specifications or by any other CSMA/CD access protocol operating at the same bit rate.
Furthermore, the ordered access guarantees both a fair distribution of bandwidth among stations and a computable maximum delay per packet; this last characteristic allows L-Express to be used for time-dependent distributed applications. Moreover, thanks to the completely distributed control, in case of fault of a segment of the channel, (a link via ether or a repeater) the parts in which the net is divided can continue operating separately without any external intervention.

However, L-Express has a moderate complexity and prototypes have been implemented by using Ethernet standard components and adding a simple device which performs manipulations of carrier sense signals.

3.4 Net Services and Communication Protocols

A common communication protocol must be adopted in MIN, to guarantee communication among all possible users. This does not prevent single user groups from using private protocols for special purposes, but provides a basic tool available in all the environment served by the network.
Basic services must include: File Transfer, Electronic Mail and Virtual Terminal services. Remote File Access, Distributed Graphics, Telefax, and even Digital Voice Transmission services ought to be included, too, at least for given subsets of hosts.
Due to the two-level structure of the MIN, the gateway services towards external services such as public or private networks are automatically obtained from connected LANs which have already implemented them.
For the same reason, tasks such as user directory maintenance service do not need to be global for the whole MIN because they can be performed acting on the single bridges; therefore, they can be committed to the single LAN managements.
Traffic data to be used for global MIN maintenance and administration can be obtained from the bridges as well.

4. Implementation

Many institutions in Pisa are possibly interested to services provided by a MAN: besides Institutes of the University and of C.N.R., there are scientific centers and/or research departments belonging to most of the main firms operating in Italy about computers; many minor firms in the same field operate in Pisa, too.
Therefore, Pisa is a qualified environment to test a prototype of the MIN.

At present, a MIN prototype is under development; it reaches the CNUCE itself, the I.E.T. ("Institute of Electronics and Telecommunications", belonging to the University of Pisa) and the I.L.C. ("Institute for Computational Linguistic", another Institute of C.N.R.).
In each of these locations, a segment of the bus is made by coaxial cable to which at least a bridge is connected, hooking LAN(s) spanned in the same building and possibly in contiguous ones.
Bus segments between locations (which are separated by distances not greater than one Km) operate via ether, as explained in the following. All of the bus segments operate at the same bit rate, that is 10 Mbps.

CNUCE's location, besides being the main prototyping environment, has a particular importance for the network, because many other networks can be reached from CNUCE (and thus from the MIN).
At present the following networks can be accessed from CNUCE:
-EARN, together with its American counterpart BITNET, which give services for file transfer, net job entry and electronic mail and gateways for the same services toward

ARPANET, CSNET etc.
-ARPANET, for remote logon TELNET, file transfer FTP and electronic mail SMTP;
-PASSTHRU, a network connecting institutions belonging to italian Universities and to the C.N.R., with remote logon service.

4.1 Communication Channel

A laser-beam link is being developed, based on infrared laser units and receivers made by Elicam, an Italian firm. The transmission units feature 20 mW light power output at 840 nm of wavelength and come with a suitable focusing system; the beam is modulated in amplitude according to a TTL input. The receiver, based on an APD (Avalanche Photo Diode) detector, outputs an amplified TTL signal. An interference light filter cuts out unwanted wavelengths; an optical system for the receiver has been obtained from a low-cost telescope.

As the laser link behaves similarly to a remote repeater for Ethernet, interfacing presents similar problems. However, the use of L-Express access scheme leads to specific requirements; first, as the scheme is collision-free, a "half-duplex" link is sufficient; second, as the scheme is based on silence time slot counters, the link must transfer silence slots in their integrity; furthermore, as these slots can be affected by errors while transmitted or received, the link must restore the integrity of corrupted silence slots, at least in the "ascending" direction.
Interfaces between laser units and the transceivers for coaxial cables used by MIN are under development according to these requirements.

At the same time, a commercial laser-based Ethernet remote repeater is being bought by a companion research group (directed by Prof. G. Prati from C.S.M.D.R. - Pisa) and will be experimented also by the MIN. A further study is being carried on by a third group (directed by Prof. G. Pegna, from University of Cagliari) about possible use of low pawer microwave links based on spread-spectrum technologies with such a bit rate requirement.

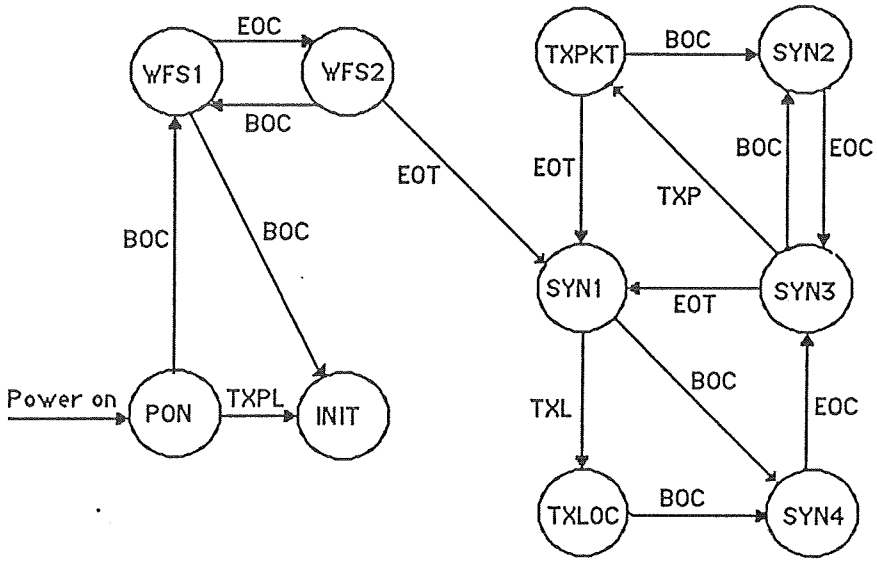

fig. 4

4.2 Channel Interface

The L-Express channel access protocol has been adopted for the MIN, as mentioned in 3.2.

The MIN interface module is based on standard transceivers and on Intel 82586 LAN controllers, which by themselves implement the IEEE 802.3 standard protocol.

In order to let them behave according to L-Express, their internal access mechanism has been disabled and a **LEXA**[6] (**L-Express Adapter**) device has been added.The LEXA device implements a finite-state automaton (fig. 4) whose main task is to enable the access to the common channel; in order to determine the access right, the CS (Carrier Sense) signal is monitored and suitable time counters are started/stopped according to it; fig. 5 depicts a logic scheme of LEXA.

A first implementation of this device, made during the P.F.I., has a chip count of about fifty TTL components. A second version is under development; using also some standard LSI components (Intel 8253 timer/counter), it has a chip count of about ten components.

Any Ethernet commercial card using an Intel 82586 chip can be converted to L-Express, just providing a plug-in unit (containing a LEXA automaton and a 82586) to substitute the 82586 itself. All Ethernet software, including low level drivers, can be adopted without modifications. Implementation of bridges between the MIN and Ethernet LANs is described in [3].

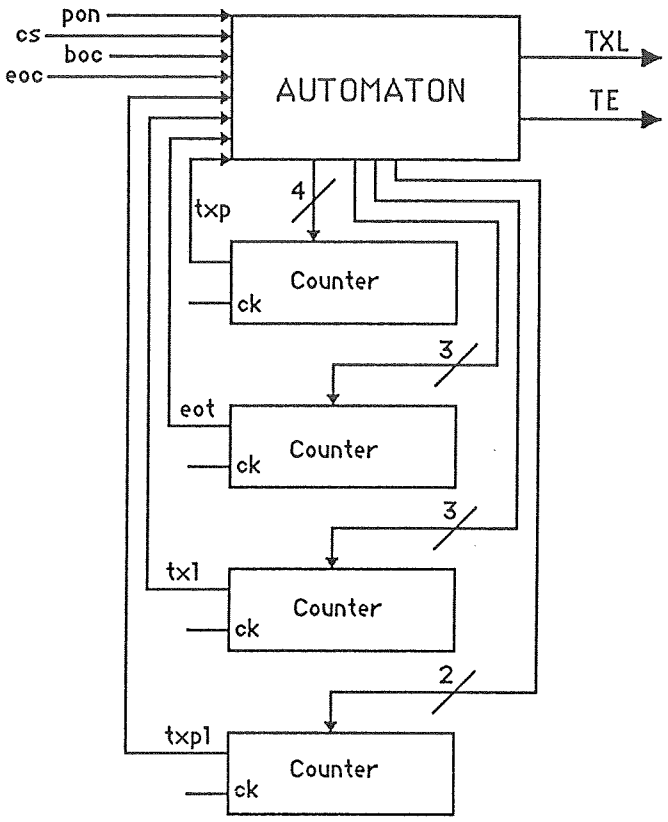

fig. 5

4.3 Communication and service protocols

The TCP/IP communication protocols are adopted by the MIN, even if they are not official standards, mainly because their implementations are already available for all possible nodes.
They operate at levels 3 (network) and 4 (transport) of the ISO-OSI reference model, but they do not comply with the model itself, having been designed previously. The distribution software package usually includes services such as: TELNET (virtual terminal), FTP (file transfer) and electronic mail.
These services meet the basic service requirements of the MIN. Their level of performance is quite poor, since they derive directly from packages designed for networks with lower reliability, such as wide area networks. Nevertheless, in this project no man power is devoted to improve their performances, as many efforts are made elsewhere to solve this problem and the MIN network will be able to adopt any new implementation complying with international standards.

On the contrary, work is done to extend services. About Distributed Graphics, a distributed version of the GKS (Graphic Kernel System) has been defined and is being

implemented. The idea is to split the software implementing GKS in different layers, so that the workstation can offer by itself given subsets of GKS functions, while the host is reached through LAN or MAN to perform heavier tasks. Another companion group (directed by Prof. L. Fratta from Politecnico - Milano) is dealing with voice-data integration.

Conclusion

A prototype MAN is being developed in Pisa, connecting at present only three different locations. The main purpose is not simply to extend to a metropolitan environment services presently available in local environments, but to enlarge the environment covered by LANs, including possible developments and improvements.

References

[1] IEEE Std 802.3, Carrier Sense Multiple Access with collision Detection (CSMA/CD) Access Method and Physical Layer Specifications.

[2] F. Borgonovo, L. Fratta, F. Tarini, P. Zini: "L-Expressnet: The Communication Subnetwork for the C-Net Project", IEEE Transactions on Communications n. 7-33, July 1985.]

[3] L. Rizzo, F. Tarini, P. Zini: "High Performance Bridge for Ethernet-like Networks", EFOC/LAN 87.

[4] A. Canalicchio, A. De Carlo: "Estensioni al Protocollo di Rete Locale L-Express", Master Degree Thesis in Computer Science, University of Pisa, 1985.

[5] Digital Equipment Corporation, Intel, Xerox:
The Ethernet, a Local Area Network: Data link Layer and Physical Layer Specification

[6] L-EXpress Adapter (LEXA): specifiche di implementazione. P.F.I. - C-Net internal document.

HIGH PERFORMANCE BRIDGE FOR ETHERNET-LIKE NETWORKS*

L. Rizzo F. Tarini+ P. Zini+

+ CNUCE - An Institute of C.N.R. - Pisa - Italy

ABSTRACT

This paper describes a bridge component designed to connect Ethernet compatible Local Area Networks to a Metropolitan Area Network. The main feature of the system is the use of a hardware automaton (SCANNER) for packet filtering, and the almost complete software transparency above the DATA LINK layer. The SCANNER cooperates with a commercial LAN controller, giving it the ability to analize all the network traffic without disturbing the CPU for non interesting packets.

1. INTRODUCTION

Generally speaking, the connection between networks can be made by three kinds of devices: repeaters, bridges and gateways, the first being only usable between segments of the same network. Each of them operates at a higher level in the ISO/OSI protocol, with increasing processing work for each transferred message. In order to increase both the transparency of the interconnection and the inter network transfer rate, when the various systems are highly compatible it is advisable the use of bridges.

As bridges work at the DATA LINK level, they are invisible to software, in that no difference appears between communications with local and non local stations. High transfer rate can be obtained due to the simple tasks bridges have to perform. No restriction is imposed by the medium access protocol, provided the bridge has the appropriate network interfaces. For instance, the prototype bridge here described connects a CSMA/CD[3] network to a virtual token one (L-Express[1]).

The packet transfer rate is limited by the operations that the CPU must execute on each of them, and a great overhead is caused by the analysis of non interesting packets. The use of a hardware filter has great benefits on performances: it permits the CPU to operate only on packets to be trasferred, which are a small fraction of the total network traffic.

---

*This work is part of the "Progetto Strategico Reti", supported by C.N.R. ("Consiglio Nazionale Delle Ricerche"), and partially funded by Honeywell Information Systems Italia.

2. BRIDGE REQUIREMENTS

2.1. TASKS

Our purpose is to realize a bridge connecting two networks with full compatibility at the ISO/OSI data link level, that is with the same addressing and packet format. Under these conditions, no transformations are needed for the packets, and the bridge only operates a selective retransmission of them. This lets the same addressing rules be used for local and external (through the bridge) communications, without the need for an inter network protocol (level 3 of ISO/OSI).

The bridge must operate at the network full speed, guaranteeing a given number of transferred packets per second. It is important for the bridge to analyze each packet on the network, but it is not necessary to be able to transfer the maximum network load. In fact, the inter network traffic is usually a small fraction of the total one, and peaks of traffic are likely to be caused by a bulk transfer. As longer packets are used to this purpose, packet frequency is lower than the maximum, and the bridge can support it. Possible bursts of short packets are easily absorbed by the internal queues of the bridge, in which received packets are stored before being processed.

Bridge performances should not be influenced by the network load, at least on the receive side. It is obvious, however, that overloads on the transmit channel may significantly reduce the transfer of packets.

The bridge must operate a selection on the received packets. To identify its two sides, let 'local' and 'city' be the names of the connected networks. Each bridge divides network stations in two groups: the knowledge of the private addresses of each station on one network, say the local one, is sufficient to the purpose. Local station addresses can be stored into an internal table of the bridge itself. Packets coming from the local network are sent on the other side if their destination address is not local. Similarly, retransmission on the local network occurs for all packets coming from the city network with a local destination address.

Packet filtering is usually performed with software techniques: for a generic non meshed network, in fact, the number of stations on each side of the bridge may be very high. When the network has a two level structure[4], the number of stations on the local side is limited (1024 for Ethernet[2] networks), and makes possible the development of hardware techniques for packet filtering.

2.2. SPEED REQUIREMENTS

Though the task performed by the bridge is very simple, speed problems make it hard to execute by software. As an example, simulation has shown that an 8 MHz 8086 CPU takes about 80uS to recognize a 48 bit element in a 1024 word array. On Ethernet, the minimum time between two packets is 67.2uS, so the only analysis of all packets would need at least a 4 times faster processor (it has to deal with two networks). The execution of

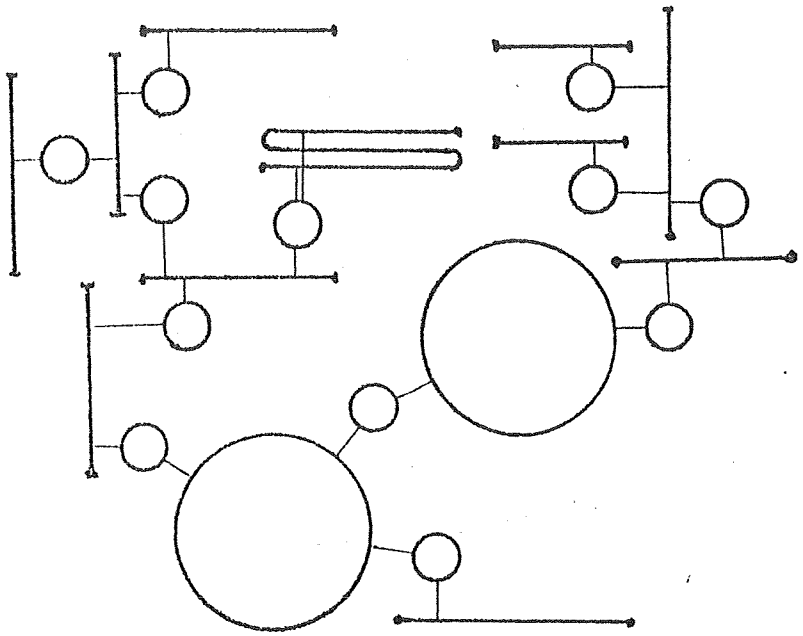

figure 1: A non meshed network can be realized using bridges, and an arbitrary number of stations may reside on each side of a bridge.

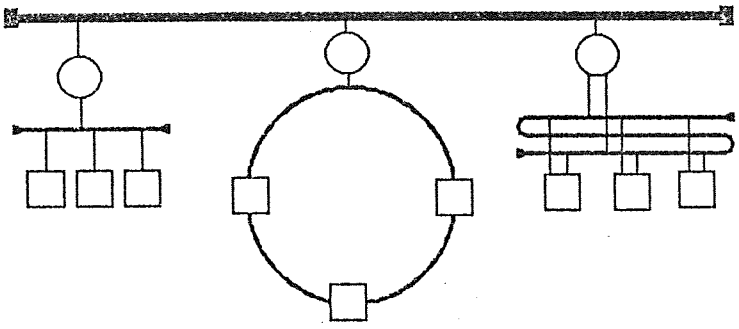

figure 2: In a two level network, the number of stations on the 'local' network is limited to a given maximum.

---

other tasks (buffer management, communication with the network controllers) may take a further 100 to 200uS for each packet, requiring an even faster processor.

It is easy to understand that most of the processing time is lost in discarding non interesting packets, resulting in a poor use of the processor's power. This overhead can be avoided by receiving only the packets that have to be transferred; it

requires the network interfaces to be able to receive (or to discard) all and only the packets with a local destination address.

2.3. SOFTWARE TRANSPARENCY

Each station on the interconnected networks is assumed to have a six bytes unique address. Our implementation expects the destination address to be the first field of the packet, according to the Ethernet standard. No other assumption is made on the content of the packets the bridge receives. Bridging is performed without any alteration of packets, allowing complete software transparency.

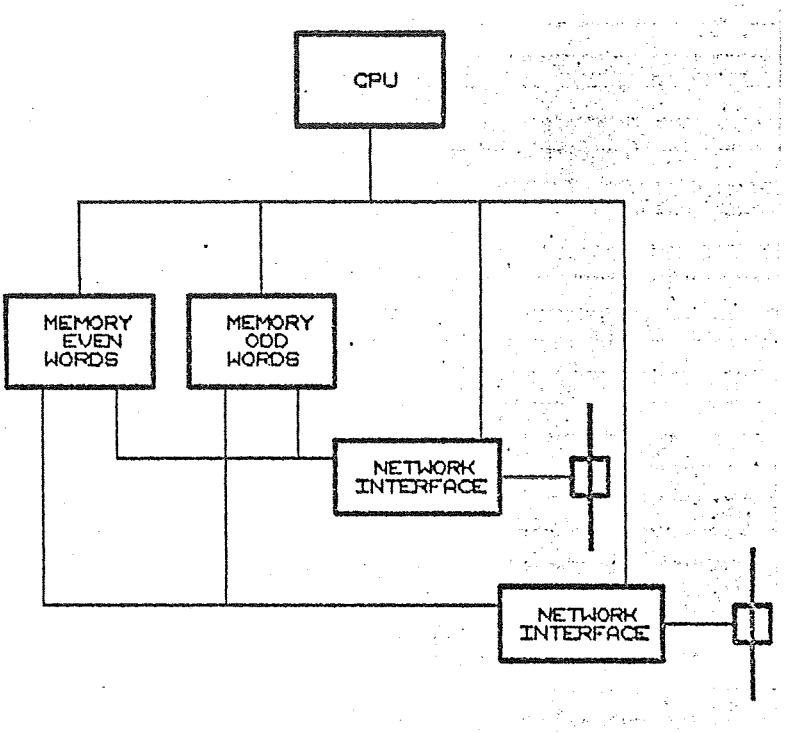

figure 3: The bridge architecture. Memory is realized with two interleaved banks to increase bandwith. Modularity allows an easy substitution of any module without modifying the others.

Interfacing other networks with the same address length may at most require a simple rearrangement of fields, with no influence on transparency. However, complete software transparency might be limited if we want to communicate with the bridge to add and remove station addresses from its internal tables. In practice this is a real need, so the bridge should intercept some particular packets for its own purpose rather than transfer them.

3. ARCHITECTURE

3.1. HARDWARE

The architecture of the bridge is quite simple. Four modules are needed and are present in our realization: a CPU that manages the entire system, two network interfaces and a common memory for packet buffering and communication between the three masters. Interfaces between modules are very sharp, allowing easy substitution of any of them to change network protocols or to increase CPU or memory performances.

The most engaging task (address filtering) must be executed by the network interface. Unfortunately the commercial VLSI controllers (Intel, AMD, etc.) for Ethernet-like networks can only recognize a limited number (64 or less) of addresses. Of course it would be impractical to build from scratch an entire Ethernet controller, even with the use of bit slice elements. The solution here described uses a standard Ethernet controller sided by a special device (SCANNER) which filters packets according to their destination address. The hardware filter cooperates with the LAN controller, giving it the ability to recognize all the 1024 possible local addresses, with no restriction on their value. The SCANNER's operation is completely invisible to the CPU.

Network controllers usually share the bus with the CPU, but in our case such a solution is not usable because of the high bandwidth request on the bus. Instead, a three port memory has been used to give each master a sufficient bandwidth. In particular, memory has been realized by means of two interleaved banks in order to use standard speed DRAMs.

3.2. THE SCANNER

The SCANNER must recognize all the packets with (or without) a local destination address. The SCANNER operates in a similar way to a Successive Approximation digital voltmeter. The operations of the SCANNER may be better understood with a two level description.

3.2.1. OPERATIONS In principle the algorithm is executed by a simple Successive Approximation Register (SAR). An internal 1K*48 bit table contains the private addresses for all the station on the local network. Local station addresses are written in ascending order into this table, and the destination address of each packet is searched in the table with a binary search technique. When a packet is present on the network, its reception is started and the destination address is copied in a compare register. The SAR is then started, and when it finishes its operations the EQ output of the comparator gives information on the presence or absence of the address in the table. This output is used to abort reception when necessary. To ensure correct operation, the search must finish before the end of the shortest packet (51.2uS on Ethernet).

3.2.2. REFINEMENTS The SCANNER needs some refinements with respect to the previous description, in order to improve performances and reduce hardware complexity.

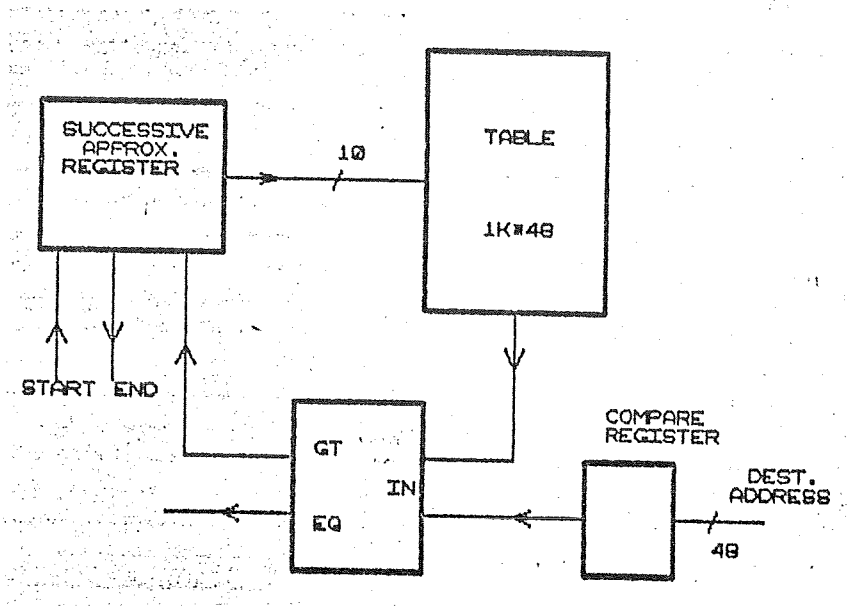

figure 4: A simplified version of the SCANNER, as described in 3.2.1.

Table management may require some milliseconds, causing the bridge to loose some packets: infact, during manipulation of the table the SCANNER cannot work correctly, and reception must be stopped. A simple way to achieve continuous operation is to duplicate the table, and swap them when the channel is inactive. At any given time a copy of the table is used by the automaton, and the other one can be accessed by the CPU for insertion and removal of new station addresses. The table switching requires a few nanoseconds and can be made in the interval between two subsequent packets.

The data paths are 16 bit wide: this eases implementation and interfacing with both the CPU and the LAN controllers. The original 48 bit comparisons are split in 16 bit operations, without reducing performances: the longest search takes only 19uS, largely below the minimum packet duration. The reduction of data path size requires a slightly more complex automaton, but the implementation benefits are still sensitive.

The table size is 2048 elements, due to the availability of 8K*8 memories; the additional entries can be used to store multicast addresses in addition to individual ones.

Address reception and extraction from the serial data stream is performed by the network controller, relieving the SCANNER from this boring task.

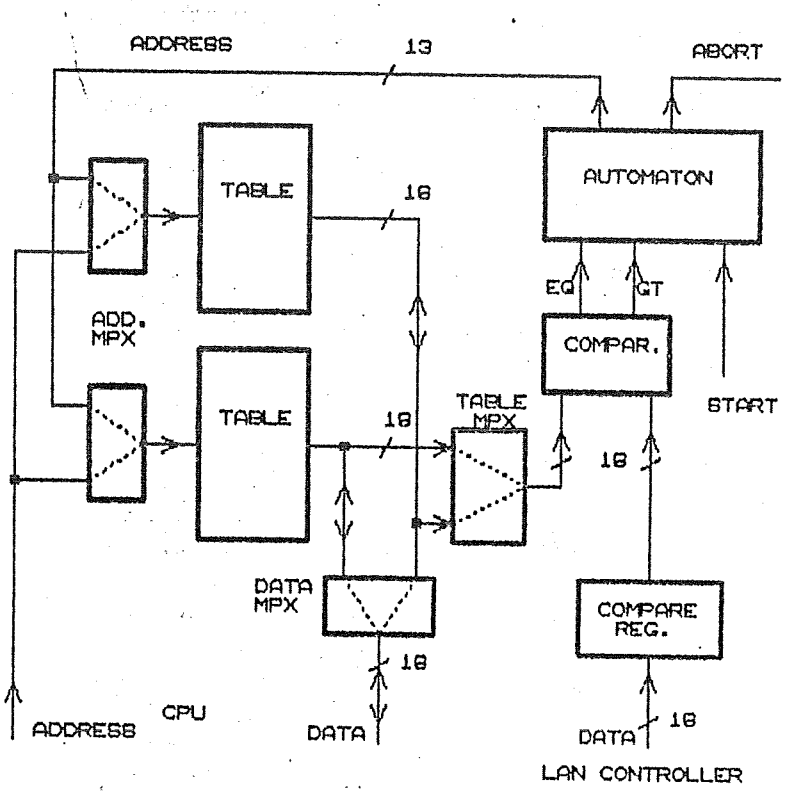

figure 5: The SCANNER's prototype, with two tables and 16 bit data paths. The automaton uses a SAR, a PAL and some random logic.

3.3. SOFTWARE

As the bridge receives from each side only its own control packets and data packets to be transferred, software is very simple. Control packets are quickly recognized because they have the private address of the bridge itself. Upon their reception the required operations are performed and an acknowledgement is

sent back to the sender. No care is needed for lost packets (commands or acknowledgements), because the possible operations (insertion or removal of an address) are idempotent.

Any other incoming packet must be retransmitted on the other side of the bridge: this requires only some list manipulation to issue the appropriate command to the network controller. As no elaboration is needed for data or control fields, each packet transfer requires only a couple of hundredths of CPU instructions, taking well below 500uS if an 8086 CPU is used.

4. IMPLEMENTATION

Some choices have been forced by time limits and by the availability of components and of a particular software development environment. However, a great modularity has been given to the prototype, in order to allow successive improvements.

4.1. NETWORK INTERFACES

Intel 82586[8] have been chosen as network controllers because of their compatibility with the L-Express protocol. The 82586 is a coprocessor communicating with the CPU by messages exchange; it manages the network traffic on a packet base, and can autonomously deal with error conditions (essentially bad frames on the network).

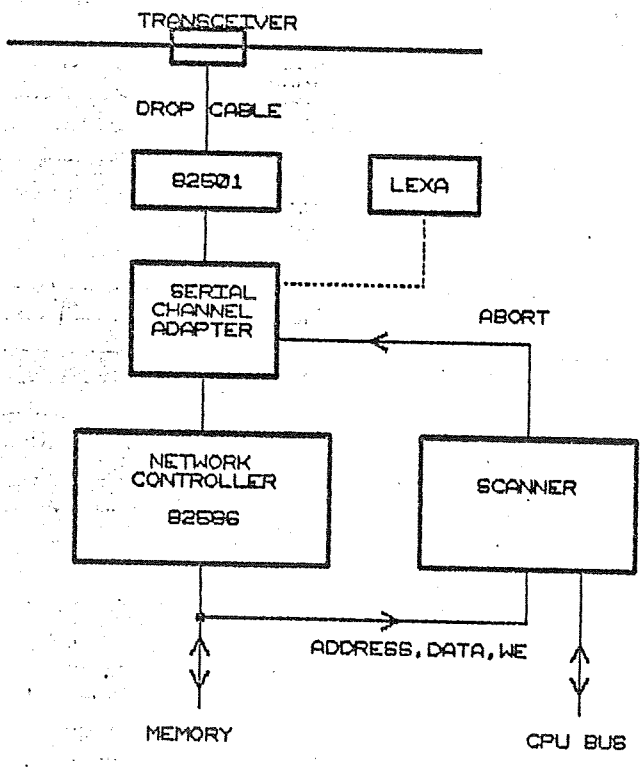

figure 6: Several blocks are present on each network interface. Some jumpers are used to choose the abort condition and the medium access protocol (Ethernet or L-Express).

Each network controller is sided by the SCANNER. When it decides a packet must be discarded, an error is introduced on the serial data stream, causing the 82586 to abort reception. This solution is needed because of the lack of an 'ABORT' input on the controller. The kind of error we simulated is well managed by the 82586, being similar to the Serial Interface Adapter (Intel 82501[8]) behaviour when the receive clock is temporarily lost. This insures the correct working of this solution on different versions of the 82586, and allows performing packet filtering without disturbing the CPU at all.

The error is simulated by the Serial Channel Adapter, which also provides interfacing with the L-Express protocol adapter[6] (LEXA). The abort condition (abort if found, abort if not found) as well as the channel access protocol (Ethernet, L-Express) are chosen by means of some jumpers.

4.2. SCANNER

The SCANNER complies to the description given in 3.2.2. The automaton is build with a TTL Successive Approximations Register and a single PAL device, operating at 2MHz. Another PAL and some random logic interface the LAN controller, while the CPU interface consists mainly in data and address multiplexers. The total chip count for the SCANNER is under 40.

Cooperation between the network controller and the SCANNER requires a particular programming of the 82586. First, the internal address filter of the LAN controller must be disabled. Second, all packet buffers must have the same size, not smaller than the maximum packet size. These features enable the SCANNER to read the destination address packed in 16 bit groups, a more suitable form for its purpose than a serial stream of bit.

4.3. THREE PORT MEMORY

Memory is realized with two interleaved banks, each of them with independent arbitration. The size of each bank is 512K, for a total of 1M of memory, and provides enough room for buffering packets. The two banks solution allows the use of cheap and large capacity 150 nS DRAMs, avoiding the need for fast static memories. Memory size can easily be expanded, if needed, because the network controller has a minimum addressing space of 4 Mbytes. However, 1Mb of memory is sufficient to store over 200 packets for each receive queue, giving enough ability to absorb load peaks. The queue is filled by the 82586 at the network full bit rate.

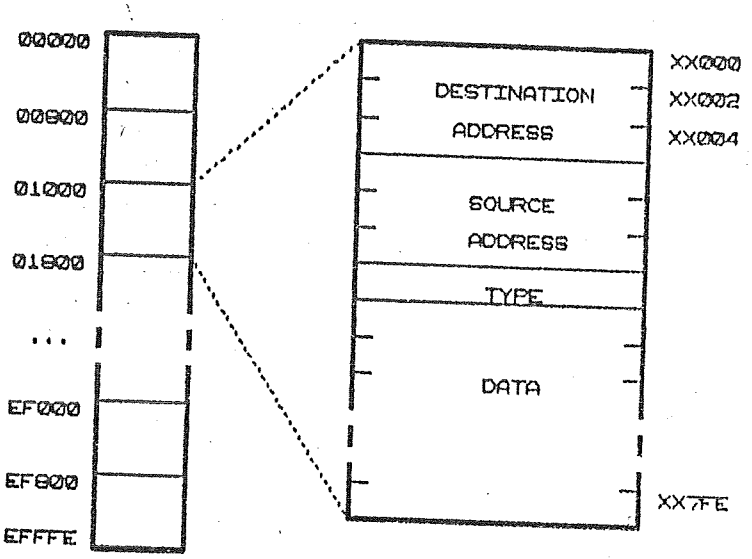

figure 7: Buffers are allocated in the common memory on 2K boundaries. A simple circuitry is used to detect the writing of the destination address of the packet currently being received, and to copy it into the SCANNER's registers.

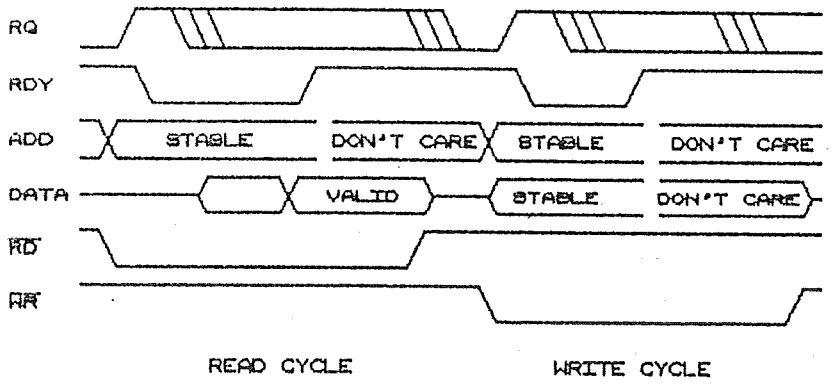

figure 8: Communication with memory uses a fully asynchronous two wire handshake (RQ, RDY).

Communication with memory is fully asynchronous, so that each module (CPU, network interfaces) can be changed with no modifications to the rest of the bridge. Memory itself can be changed if faster devices are available or a greater bandwidth is needed for bridging faster networks. Memory is internally synchronous, and uses a rotating priority arbitration to guarantee the worst case access delay. Arbitration and RAM timing logic are built with PALs and standard TTL gates, and work with an 8 MHz clock. The total chip count for each memory card is 40, and most of the devices are used for data and address multiplexing.

4.4. CPU

As packet filtering is performed by the SCANNER, particularly fast CPU are not needed. The CPU speed influences only the transfer rate, not the ability to analyze all the network traffic. As stated in 2.1, the transfer rate can be a fraction of the maximum network speed without introducing a significant loss of performances. A value of 2000 packets/sec should be adequate, allowing the use of an 8086 CPU. Faster processors are usable to increase transfer rate, if needed, thanks to the great modularity of the system.

For software development reasons the CPU card has been obtained interfacing an 8086 based PC. A later version of the bridge may use an 80186 card with ROM written software. The 80186 is fully SW compatible with the 8086, and includes some useful peripheral on chip (DMA and interrupt controllers, timers and programmable decode logic).

As the addressing space of the 8086 is limited, the CPU can see, at a given time, only one of the 16 64K blocks of the shared memory. The selection is made by a 4 bit relocation register mapped in the I/O space.

The CPU sees in its memory space the SCANNERs' tables to perform their manipulation. A register on each network interface provides control lines to swap tables, enable/disable the SCANNER, sense the interrupt line and issue Channel Attention to the LAN controller.

More details on this implementation can be found in [5].

REFERENCES

[1]-F. Borgonovo, L. Fratta, F. Tarini, P. Zini
L-EXPRESS NET: The Communication Subnetwork for CNET Project
IEEE Transaction on Communications N.7/33, July 1985

[2] Digital, Intel, Xerox
The Ethernet: Data Link Layer and Phisical Layer

Specifications

[3]-IEEE Std 802.3-1985, Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Phisical Layer Specifications.

[4]-M.Mannocci, F.Tarini, P.Zini
A LAN approach to MANs
EFOC-LAN 87

[5]-L. Rizzo
Progetto e realizzazione di un modulo di collegamento tra reti locali
Masted Degree Thesis, University of Pisa, 1987

[6] L-Express Adapter (LEXA): specifiche di implementazione
P.F.I. - CNET Internal document

[7] Intel Microsystem volume 1
Intel 1985

[8] Intel Microsystem volume 2
Intel 1985