A NEW LAN CONCEPT FOR LEP MACHINE NETWORKS,
A STEP TOWARDS LHC.

Louis GUERRERO, Patrick LIENARD, CERN, Geneva, Switzerland.

Abstract

LEP networks, implemented in 1987, are based on two Token-ring backbones using TDM as the transmission medium. The general topology is based on routers and on a distributed backbone. To avoid the instabilities introduced by the TDM and all the conversion layers it has been decided to upgrade the LEP machine network and to evaluate a new concept for the overall network topology. The new concept will also fulfil the basic requirements for the future LHC network. The new approach relies on a large infrastructure which connects all the eight underground pits of LEP with single-mode fibres from the Prevessin control room (PCR). From the bottom of the pits, the two adjacent alcoves will be cabled with multi-mode fibres. FDDI has been selected as the MAC protocol. This new concept is based on switching and routing between the PCR and the eight pits. In each pit a hub will switch between the FDDI LMA backbone and the local Ethernet segments. Two of these segments will reach the alcoves by means of a 10Base-F link. In a second phase implementation, this scheme will provide for workgroup organisation and bandwidth allocation. The technological choices make a future evolution towards ATM and 100Base Ethernet possible and allow us to preserve a large part of the investment. This paper describes the implementation of this scheme.

1. INTRODUCTION

Since 1974, starting with the SPS machine, networking has always been at the centre of CERN's SL division controls system. When the LEP machine studies came into force it was decided to go for standard protocols. LAN evolution was not clear at that time. To also cope with requirements for safety as well as machine performance and operation, delicate decisions had to be made. This led to separate networks for services (water, cooling, ventilation, ...), machine operation and beam observation. These two networks are known as LEP SERVICES (LSV) and LEP MACHINE (LMA). Decision was finally in favour of the Token-Ring[1] protocol and the TDM[2] as transmission media. During the past years these networks have undergone many changes [1] to reach the higher performances always required by users. For many reasons, these basic decisions were never revised. LEP 200 and LHC preparation make fundamental changes compulsory.

This paper outlines these new ideas and concepts, and states the situation of today's LMA network after the first phase of improvements. The steps to come until completion of the project will also be mentioned.

2. THE LEP MACHINE NETWORK IN 1994

2.1. LAYOUT AND CONCEPTS.

The LMA network is interconnected with the Prevessin control room (PCR) and the "rest of the world" by means of a router running the TCP/IP suite of protocols as well as the Apollo Domain one. As shown in figure 1, the LMA backbone was made of a Token-ring running over the TDM and interconnected with local Ethernet segments in alcoves by means of a translation bridge. Although this was a satisfactory scheme for many years it bore some weaknesses which became critical with the increasing network load and the requirements for higher performance.

[1] ISO 8802.5 standard.
[2] Time Division Multiplexing, specified in the CCITT G700 series recommendations.

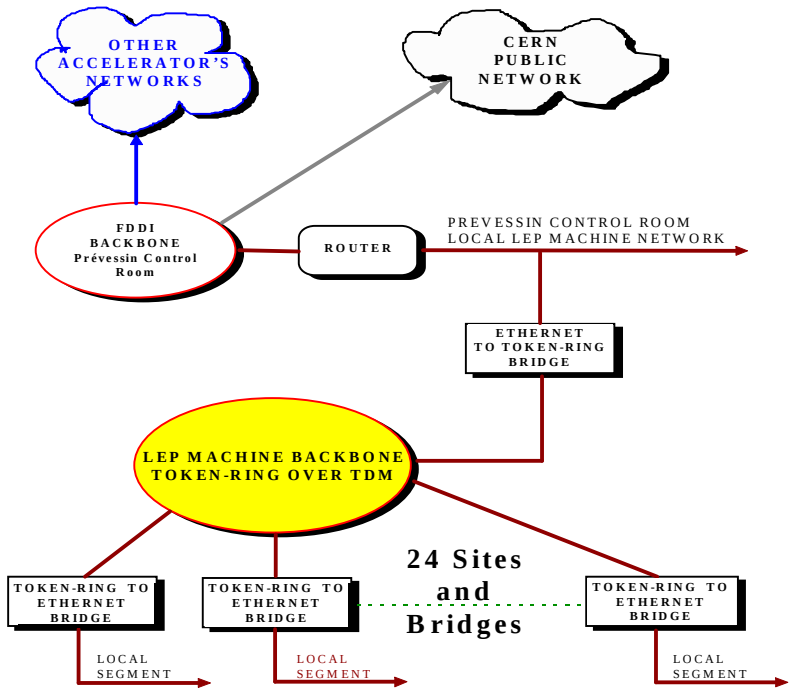


Fig. 1 **LEP MACHINE LAYOUT AS IN 1994.**

2.2. RELATED PROBLEMS.

Figure 2 represents the relationship between the Token-ring and the TDM at a site level. It clearly shows the boundaries of the two systems and where the problems mentioned can be identified.

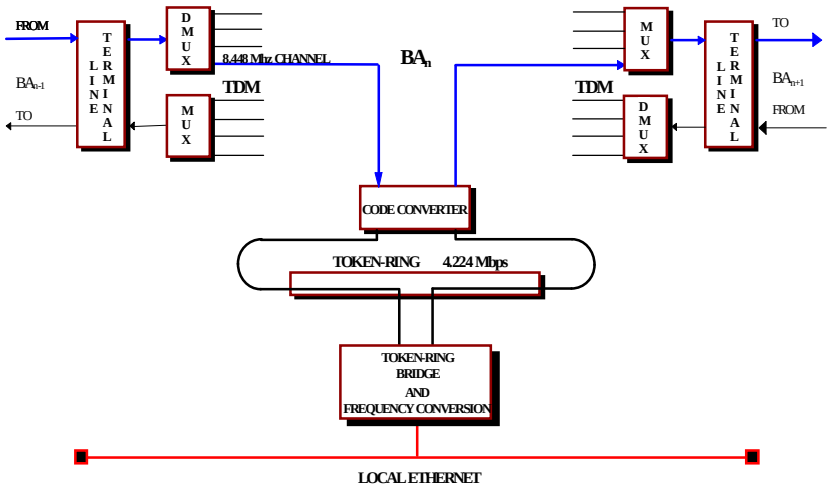


Fig. 2 **LEP MACHINE LAYOUT IN A SITE AS IN 1994.**

COMPATIBILITY BETWEEN TDM AND TOKEN-RING.

Levels of incompatibility can be identified as :

1) To correctly sample the Token-ring which has a 4 Mbps bit rate we use an 8.448 Mbps channel of the TDM. These rates not being multiples of each other, some buffering is needed to ensure the frequency conversion. Hence we decided to make this conversion at the translation (Token-ring to Ethernet) bridge level. It has worked perfectly up to now but is incompatible with industrial products.

2) Coding for the plesiochronous transmission systems is defined by the G703 standard which retains the HDB3[3] code. The IEEE 802.5[4] standard applies a biphase technique called the Differential Manchester[5]. The necessary conversion between these two codes was realised in a home-made code converter [2]. In fact, this was not a true conversion since it only considered the positive part of the electrical signals; nevertheless it allowed a simplified design. Some very exceptional patterns are not correctly translated. This introduces some frame errors and penalties on the bandwidth by creating unnecessary re-transmissions. Reliability of the whole system is obviously decreased by the failure potential of this additional interface.

3) During the insertion phase of a Token-ring station the insertion relay of the concentrator bounces. This introduces synchronisation errors on the TDM. As it is almost impossible to discriminate between them and real losses it becomes difficult to manage the TDM's synchronisation signals.

BANDWIDTH WASTING AND LIMITATIONS.

Bandwidth is always a great concern in transmission systems and using an 8.448 Mbps channel to transmit a 4 Mbps
protocol appears wastful. We have seen that it was imposed by the necessity of sampling the Token-ring two times per bit. It also means that increasing the transmission rate of the Token-ring to 16 Mbps would oblige us to raise the TDM rate to 140 Mbps (a sampling rate of 34 Mbps times four levels of multiplexing). This represents a real technical and financial limitation which left the Token-ring far behind as a means of improving performance.

TRANSLATION BRIDGES.

In the 1994 topology, as a result of historical evolution of the LMA network, a packet travelling from the PCR network to a computer attached to a local segment in an alcove had to be translated three times, introducing longer transmission delays and some penalty on the bandwidth. The traffic increase resulted sometimes in critical bridge congestion. Bridged topology also penalised throughput essentially because broadcast propagation could not be limited. Finally, it also introduced a great number (25) of bridges which obviously represented some potential of failure.

FAULT DEPENDENCY.

Using the TDM as a media for the Token-ring consequently implied that a fault on the TDM, either at the line terminal or at the multiplexer level, broke the ring integrity. An incident on the Token-ring had also severe repercussions on the TDM transmission. This interdependency, sometimes, made recovery from a fault and management of both systems very difficult.

3. THE NEW CONCEPT

3.1. INTRODUCTION.

Introduction of diskless computers, the generalised usage of X[6] terminals and of the NFS[7] protocol rendered the above model inadequate by the end of 1993. Fortunately this situation had been anticipated [1] and the infrastructure needed for the future was already well established. The new network had to eliminate all the above-mentioned problems, i.e., it had to be independent of the TDM. The decision to use an optical fibre-based network was made even if the protocol to be run was not firmly decided : technology was again at a decision point and we were too early for the market.

3.2. THE FIBRE LAYOUT.

Pulling fibres started in 1985 to implement the TDM transmission over the long distances of LEP. At that time, equipment was essentially multi-mode so cables were composed of a majority of MM[8] fibres and only a few SM[9]

[3] High Density Binary code of the 3rd order, CCITT G703 recommendation.
[4] IEEE 802.5 and ISO 8802-5 specify the Token-passing ring medium access protocol and its physical attachment.
[5] Differential Manchester code has the same properties as the normal biphase one but prevents the crossing of wires on the line.
[6] X or X Window : Graphic User Interface issued by the MIT in 1985.
[7] NFS : Network File System introduced by SUN in 1984.
[8] MM : Multi-mode.
[9] SM : Single-mode.

ones. We reinforced the single-mode infrastructure for LEP 200 in 1992. Thereafter, we had a comfortable number of SM fibres linking the PCR[10] to each pit. In early 1994, we started equipping from the surface down to the pits with SM fibres. The major problem to be solved, using fibres in the accelerator environment, was to protect them from radiation degradation. Pulling them in the central drain of the LEP machine [1] revealed itself to be very reliable. The decision was made to apply the method to the whole network. Insertion, pulling and extraction of the fibre cables involved an important engineering work around the drain-pipe. Fibres run, see figure 3, in surface trenches to each pit, then go down to the underground site where they connect the electronics and, finally, from there run through the drain-pipe to join both adjacent alcoves. In the drain, cables are composed of 6 SM and 6 MM fibres. Outside, compositions vary according to the sites but are never less than these figures. The types of fibre in the cables are 10/125 µm for the SM and 50/125 µm for the MM.

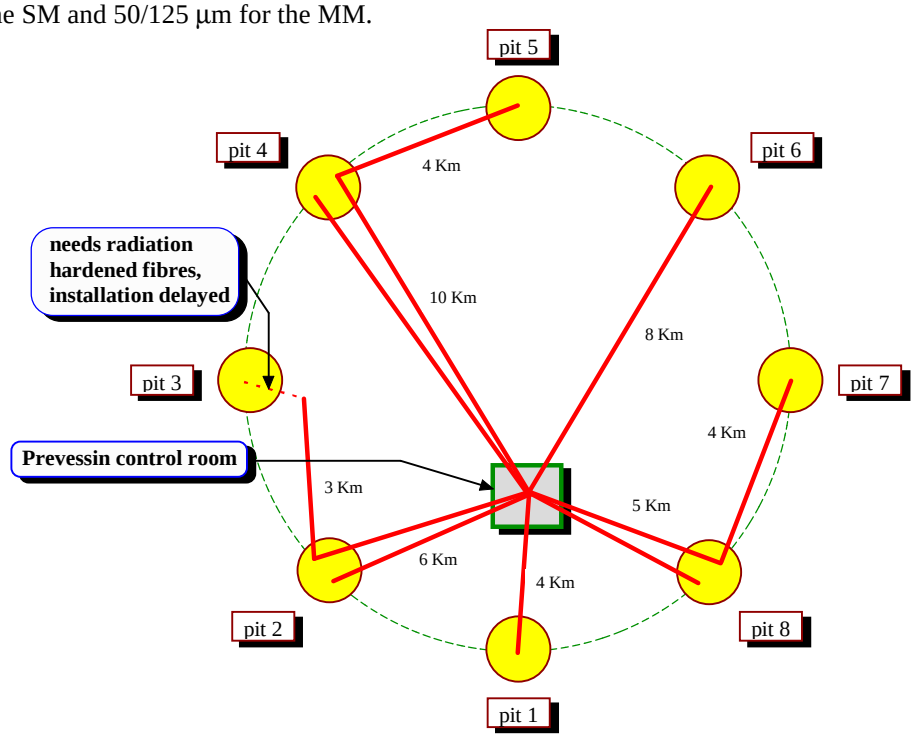


Fig. 3 LEP SINGLE MODE FIBRE LAYOUT.

3.3. THE ARCHITECTURE.

THE CHOICE OF THE PROTOCOL.

In 1994, ATM[11] implementations were proprietary solutions and ATM specifications had not yet reached stable standard levels. We decided on FDDI[12] which was offering, and still offers today, a more stable protocol for packet transfer. Nevertheless we had to choose a solution that prepared for the future and protected the current investment. This obviously anticipated migrating towards ATM in due time.

ROUTING OR SWITCHING ? THE LAYOUT.

Our first approach towards the new network architecture was again based on routing. Soon it became evident that we should take advantage of switching technology : increase of bandwidth, bandwidth allocation and VLAN[13] facilities. Preserving the investment should also be easier. This led to the topology of figure 4 which shows a central hub, in the PCR, switching FDDI and eight peripheral hubs switching Ethernet from FDDI. For economical reasons relative to single-mode transmission we decided on a SA[14] type of FDDI connection. Thus, two pits could connect the same DA[15] port of the central hub and share the same logical ring (see figure 5). Although this solution is less reliable than one

[10] PCR : Prevessin Control Room for SPS and LEP machines.
[11] ATM : Asynchronous Transfer Mode retained by CCITT and supported by ATM Forum.
[12] FDDI : Fiber Distributed Data Interface, ANSI X3T9 or ISO 9314 standards.
[13] VLAN : Virtual LAN.
[14] SA : Single Attachment, this type of connection uses only one physical port of type S to connect the network.
[15] DA : Dual Attachment, uses two ports of type A and B to connect the network.

based on the dual (counter-rotating) ring and DA type of connections it saved a large amount (about 30%) on the electronics and the SM fibres. Another advantage of this layout will be equipping the alcoves with FDDI instead of 10Base-F Ethernet. Finally, installation of dedicated servers on the LMA backbone was also foreseen at the PCR location.

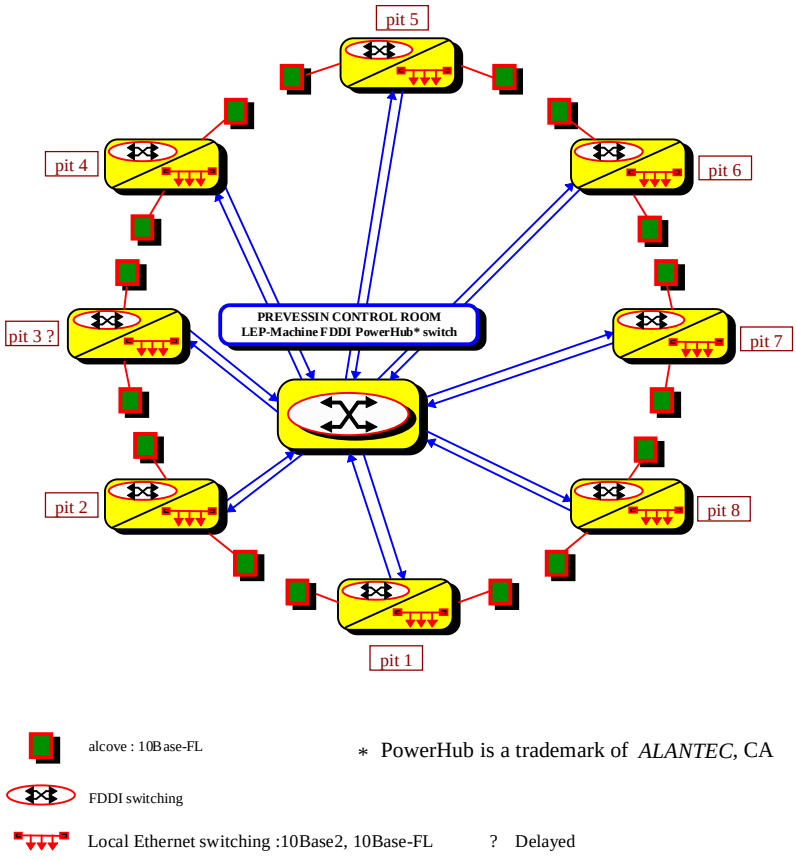


Fig. 4 NETWORK CONCEPTION BASED ON POWERHUBS* FDDI AND ETHERNET SWITCHING

THE ELECTRONICS.

Characteristics of the equipment foreseen were :

- switch and bridge FDDI
- switch and bridge Ethernet from FDDI
- route multi-protocols, essentially the IP protocol suite and be transparent to Apollo Domain
- have multi-mode and single-mode types of connection for FDDI
- deliver full bandwidth on each port
- allow bandwidth allocation
- feature VLAN facilities
- offer modules equipped with 10Base-2 and 10-F Ethernet ports
- are modular and scalable
- are compliant with SNMP
- are committed to ATM

We also asked for several options. Among them :

- full-duplex Ethernet
- TPPMD[16] FDDI
- 100 Mbps Ethernet with a preference for 100VG-AnyLAN[17], our workstations being HP ones

[16] TPPMD : Twisted Pair Physical Media Dependent.
[17] 100VG-AnyLAN is defined by IEEE 802.12 standard.

All these features together led us to think that we had no chance of finding such a product on the market. We knew several firms were ready to make some announcements in the year to come. We issued a call for tender and surprisingly found two societies with products closely following the specification.
PowerHub[18] 7000 (PH) from ALANTEC was selected because it completely fulfilled the compulsory requirements. Optional features were also met apart from the 100VG requirement.

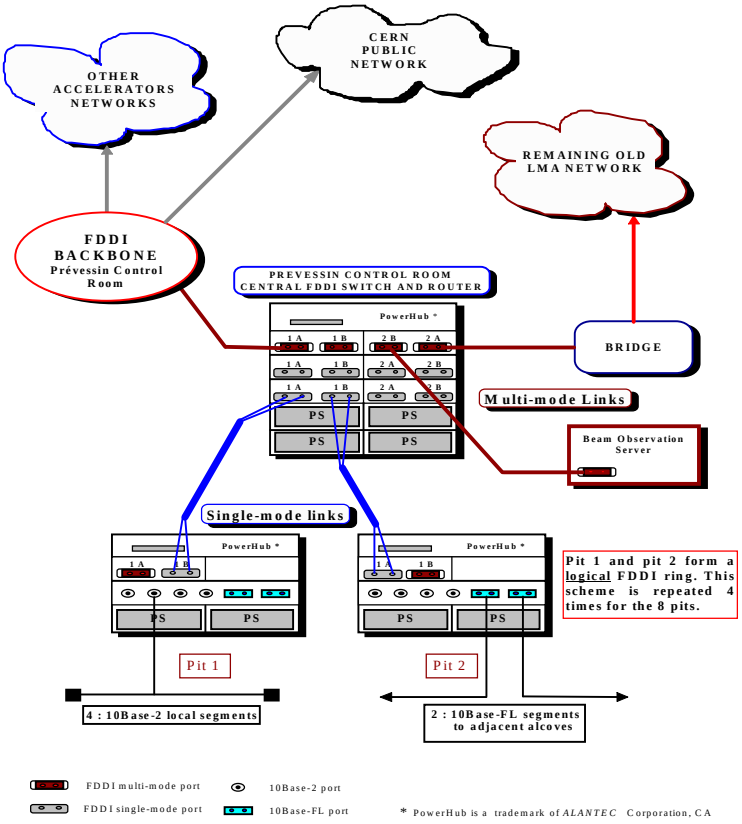


Fig. 5 LEP MACHINE LAYOUT AS FROM MARCH 1995.

The PH is expandable and scalable from 1.6 to 3.2 Gbps total channel bandwidth. Its multiple-port shared memory provides and shares bandwidth at the chip level (ASIC[19]). Thus, PH software cannot only forward packets using any bridging, routing or filtering method which can be expressed by a programmed algorithm, but also provides many network management features such as statistics gathering, security filtering and port monitoring. Furthermore, new features can be added at will, at any time and ***adaptation to new standards*** is easy.

NETWORK MANAGEMENT.

LMA, like LEP Services and the SPS accelerator networks, uses SNMP[20] and RMON[21] protocols, via the HP[22] OpenView[23] Interconnect management product, and the HP Metrics statistics and analyser tool, to control all the network equipment. This philosophy was, of course, kept for the new project. In addition, we needed and obtained an easy but powerful configuration tool for VLAN implementation. HP OpenView is a very flexible product which allows easy integration of other vendors' products. ALANTEC offers Power Sight, a network management utility, which is totally compatible with OpenView.

[18] PowerHub is a trademark of ALANTEC Corporation, SAN JOSE, CA.
[19] ASIC : Application Specific Integrated Circuit.
[20] SNMP : Simple Network Management Protocol.
[21] RMON : Remote MONitoring protocol.
[22] HP : Hewlett-Packard Company.
[23] OpenView : stands for Open View Network Node Manager, integrated in Open View Windows (OVW).

4. THE STAGE OF IMPLEMENTATION IN 1995

4.1. PRESENT STAGE.

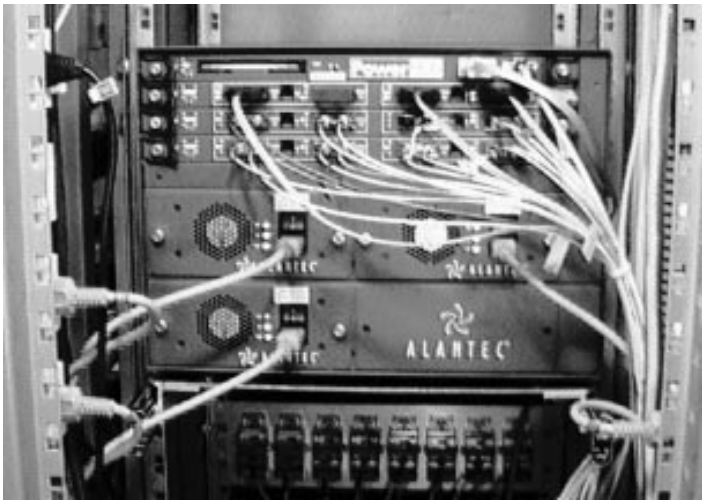
LAYOUT.

At the end of the LEP annual shutdown (end of March 1995) fibres had been installed down all the pits except pit 3. For the latter we had to find either radiation-resistant fibres or a special path to pull normal ones. Alcoves were linked to pits with fibres in the drain-pipe, except between pits 1 and 4 where special engineering was needed to clear the drain.

ELECTRONICS.

PHs were installed down all the pits, except pit 3, with the following configuration (see figure 5):

- one chassis with one power supply
- one packet engine
- one universal FDDI module equipped with one A or B type SM port (two sites sharing the same logical ring) and one B or A type MM port allowing local FDDI ring extension
- one universal Ethernet module equipped with four 10Base-2 ports and two 10Base-F ports



The PCR's PH was installed with the following configuration :

- one chassis with three power supplies
- one packet engine
- one dual FDDI MM module connecting the PCR backbone and offering an attachment for LMA local servers
- two dual FDDI SM modules connecting (switching) all the pits to the PCR backbone

The alcoves from pits 1 to 4 via 8,7... were all equipped with a 10Base-F link originating in the pits' PH and using optical to copper converters to create the local Ethernet segment.

CONFIGURATION.

One of the MM ports on the PCR's PH was configured as the router port between PCR and LMA subnets. All other ports of the PH are members of the LMA subnet and as such have received the same IP address. All ports are bridged. All the ports of a PH in a pit are bridged and have the same IP address per hub. Thus, only bridging is implemented nowadays without VLAN, unless it is to consider the whole LMA network as a workgroup.

FIRST IMPRESSIONS.

To date, we have accumulated six months of experience and had to face several small problems :

- power supplies were not compliant with European mains
- some rare losses of FDDI connection disturbed the PCR backbone
- some rare IP routing lock-out, on the PCR port, isolated the LMA subnet

Since we have been testing satisfactorily new power supplies proposed by ALANTEC we are now ready to install them. Upgrading FDDI and OS software versions seems to have solved the problem of losses. We have now been running the last version for over a month without any trouble.

At this stage, after a six months' operation, we have gained appreciably in reliability and performance. Time was too short to make measurements on the network but, in normal operation of LEP, we have gained about a factor three on the response time of a standard 64-byte ICMP[24] echo message crossing all the network. Statistics show extremely low error rates on integrated periods as compared to previously, for example :

PCR PH :

FDDI : 1 800 million packets over 605 hours with 45% peak utilisation showed 0 error on Xmit/Rcv buffers and 0 packets dropped.

US25 PH :

FDDI : 69 million packets over 410 hours and 43% peak utilisation, 0 dropped .

Ethernet port 13 : 7 million packets with 21% peak utilisation showed 21 Xmit buffer error, 1 FCS[25], 1 FA[26] and 262 collisions. All other ports have 0 error counters.

US25 is the worst case today in all the PHs installed.

4.2. WHAT IS LEFT TO BE DONE ?

During the 1995-96 shutdown the drain-pipe preparation will be completed between pits 1 and 4 via pits 2 and 3. Hence, we will be able to pull the fibres in the drain. We have now a solution to join pit 3 with normal fibres. So, the first phase will be completed for the end of the shutdown as we already have the electronics to equip these remaining sites.

5. FUTURE STEPS

The second phase will involve implementing VLAN over the network. Before doing this we need to gain confidence on the whole first phase, so it will certainly not take place before 1997. Studies could start in 1996.

During the 1996 shutdown we also plan to implement the same topology for the LEP Services network. Studies have already be undertaken for the SPS network.

6. CONCLUSION

Due to a lack of Token-ring components on the market and to a non-standard network implementation (the TDM being used as a communication media) the LEP machine network was inadequate for the requirements foreseen for LEP 200 and LHC. A new network conception had to be evaluated, based on today's standards and on an optical fibre infrastructure. It has been (and will be) a difficult job to install the fibre infrastructure but the effort was worthwhile as we have gained a solid base for many years. The decision to implement switching over FDDI and Ethernet appears to have been confirmed by market trends. Choosing the PowerHub brought, very quickly, more than the expected improvement in performance and reliability. The choice of ALANTEC offers us VLAN techniques today, good management tools and confirms that our decision on the HP network management product was also correct. Not only can we meetLEP machine users' requirements for many years but also we have prepared for future migration towards ATM. Associated with the category 5 cabling we had already implemented in the PCR, switching will also provide us with a solid base for multimedia communications.

7. ACKNOWLEDGEMENTS

The project has always been encouraged by L. Evans, director of the LHC project, P.G. Innocenti, head of CERN's Telecom board, K.-H. Kissler, head of SL division, and H. Isch, SL division planning officer. We also received a strong support from our group, particularly R. Lauckner, group leader, and R. Rausch, deputy group leader. We express here our gratitude.

[24] ICMP : Internet Control Message Protocol.
[25] FCS : Frame Check Sequence.
[26] FA : Frame Alignment error.

It is a pleasure to acknowledge the contributions of :
B. Amacker, K. Kohler, O. Olsen and D. Schamme for their tremendous work on the fibre installation, and
A. Francano and J.-L. Vo-Duy for the electronics installation.

8. REFERENCES

[1] P. Lienard, « Evolution of the SPS and LEP communication control network for the SPS and LEP accelerators », ICALEPCS 1994.
[2] B. Hall and D. Swoboda, The ACCI code conversion interface for the LEP Controls token ring, LEP Controls note 84, 25 February 1988.