A NEW LAN CONCEPT FOR LEP MACHINE NETWORKS

A STEP TOWARDS LHC

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

Abstract

The 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 fulfill 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 to the Prevessin control room (PCR) with single-mode fibres . From the bottom of the pits, the two adjacent alcoves will be cabled with multi-mode fibres. FDDI has been selected as the 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 of the 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 central to the CERN SL division controls system. When the LEP machine studies were made it was decided to adopt standard protocols. LAN evolution was not clear at that time. To cope with requirements for safety as well as machine performance and operation, delicate decisions had to be made, leading to the use of separate networks for the services (water, cooling, ventilation etc.) and for the machine operation and beam observation. These two networks are known as LEP SERVICES (LSV) and LEP MACHINE (LMA). It was decided to use the IBM Token-Ring protocol and the CCITT Time Division Multiplexing (TDM) as the transmission medium. 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. Preparation for the projects LEP 200 and LHC have led to fundamental changes.

This paper outlines these new ideas and concepts and gives a description of the present state of the LMA network after the first phase of improvements. The steps necessary to complete the project are outlined.

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 translation bridges. Although this worked satisfactorily for many years it bore some weaknesses which became critical with the increasing network load and the requirements for higher performance.

2.2 *RELATED PROBLEMS.*

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

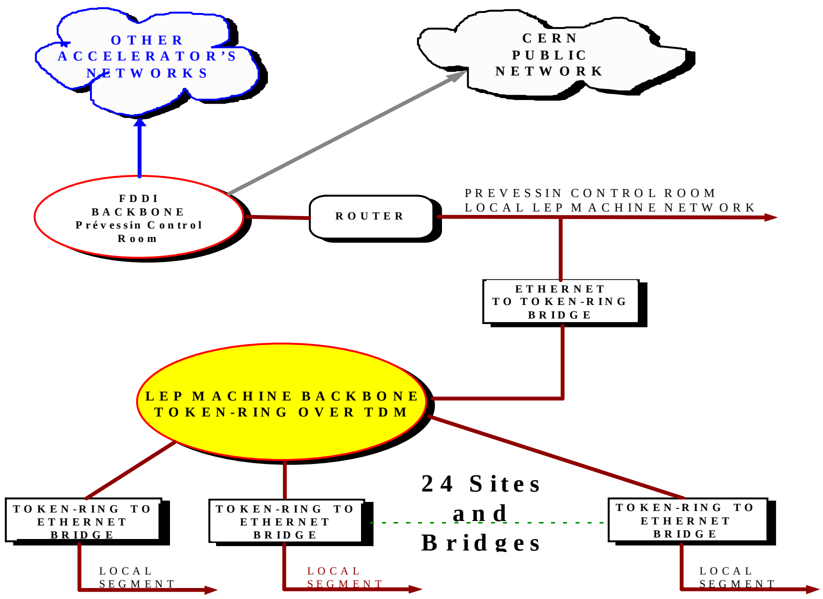

Fig. 1 LEP MACHINE LAYOUT AS IN 1994.

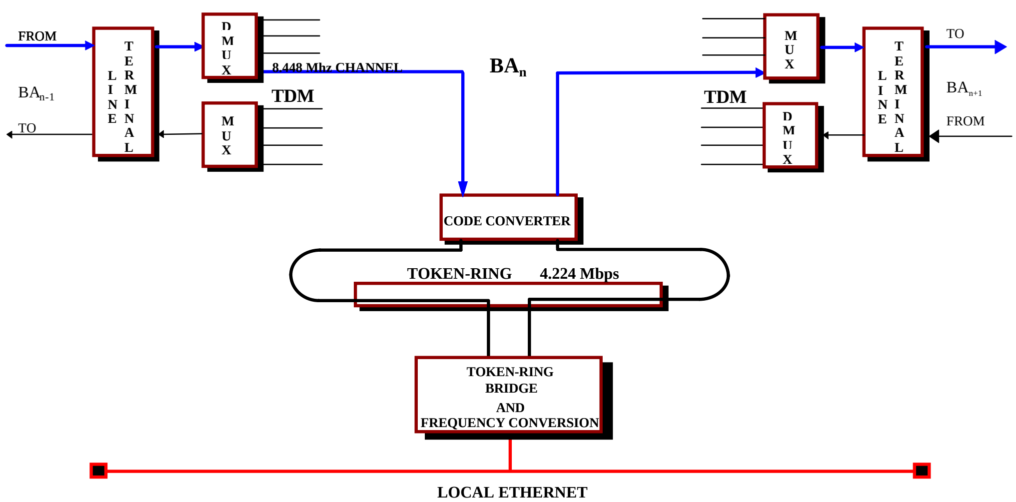

Fig. 2 LEP MACHINE LAYOUT IN A SITE AS IN 1994.

- *INCOMPATIBILITY BETWEEN TDM AND THE TOKEN-RING.*

Levels of incompatibility can be identified as :

1) We use an 8.448 Mb/s channel of the TDM to sample the Token-ring which has a 4 Mb/s bit rate These rates not being multiples of each other, some buffering is needed to enable the frequency conversion. We decided to make this conversion at the Token-ring to Ethernet bridge level. It has worked perfectly up to now but it is incompatible with industrial products.

2) The TDM transmission system uses the HDB3[1] code. The IBM Token-ring uses a biphase technique called Differential Manchester. The necessary conversion between these two codes was made by 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, but 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 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 synchronization errors on the TDM. As it is almost impossible to discriminate between these and real losses it becomes difficult to manage the TDM synchronization signals.

- *BANDWIDTH WASTING AND LIMITATIONS.*

Bandwidth is always a great concern in transmission systems and using an 8.448 Mb/s channel to transmit a 4 Mb/s protocol appears wasteful. It was imposed by the necessity to sample the Token-ring twice per bit, due to the Manchester encoding It also means that to increase the transmission rate of the Token-ring to 16 Mb/s would oblige us to raise the TDM rate to 140 Mp/s (a sampling rate of 34 Mb/s times four levels of multiplexing). This represents a real technical and financial limitation which eliminated the use the Token-ring as a means of improving performance.

- *TRANSLATION BRIDGES.*

In the 1994 topology, as a result of the historical evolution of the LMA network, a packet traveling 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 sometimes resulted in critical bridge congestion, often made worse because broadcast propagation could not be limited. In addition, the introduction of a great number (25) of bridges increased the potential of failure.

- *FAULT DEPENDENCY.*

Using the TDM as a medium for the Token-ring meant 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 has also severe repercussions on the TDM transmission. This interdependency made recovery from a fault and the management of both systems very difficult.

3. THE NEW CONCEPT

3.1 INTRODUCTION.

The introduction of diskless local computers, the generalised use of X terminals and of the SUN Network File System (NFS) protocol rendered the system described above 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 on it was not firmly decided: technology was again at a decision point and we were too early for the market.

3.2. THE FIBRE LAYOUT.

The installation of optical fibres started in 1985 to implement the TDM transmission over the long distances of LEP. At that time, the equipment available was essentially multi-mode so the cables were composed mainly of multi-mode (MM) fibres with only a few single-mode (SM) ones. We reinforced the single-mode infrastructure for LEP 200 in 1992. Thereafter, we had a comfortable number of SM fibres linking the PCR to each pit. In early 1994, we started the installation of SM fibres from the surface down to the equipment alcoves round the ring pits. The major problem to be

---

[1] High Density Binary code of the 3rd order, CCITT G703 recommendation.

solved in using fibres in the accelerator environment was to protect them from radiation degradation. Earlier tests had been made by putting some fibres in the central drain under the floor of the LEP tunnel [1] and they were found to be virtually unaffected after one and half years. The decision was made to apply the method to the whole network. Insertion, pulling and extraction of the fibre cables involved considerable engineering work around the drain-pipe. The fibres run in surface trenches to each pit and then go down to the underground site where they are connected to the electronics and then run from there through the drain-pipe to connect to both of the adjacent alcoves. In the drain, the cables are composed of 6 SM and 6 MM fibres. Elsewhere, the cable 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. The surface layout is shown in Fig. 3.

3.3 THE ARCHITECTURE.

THE CHOICE OF THE PROTOCOL.

In 1994, The ATM[2] implementations were proprietary solutions and ATM standards specifications had not yet reached stable levels. We decided to use FDDI[3] 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 required us to anticipate a migration towards ATM in the future.

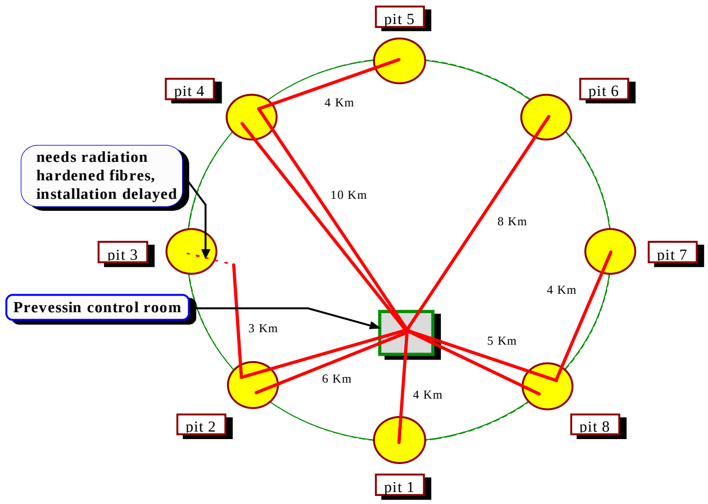

Fig. 3 LEP SINGLE MODE FIBRE LAYOUT.

THE LAYOUT.

Our first approach towards the new network architecture was again based on routing. It soon became evident that we should take advantage of the latest switching technology which provides an increase of bandwidth, bandwidth allocation and virtual LAN (VLAN) 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 links and eight peripheral hubs switching Ethernet links from each FDDI link. For reasons of economy we decided on a Single Attachment (SA) type of FDDI connection. Thus, two pits could connect to the same Double Attachment (DA) port of the central hub and share the same logical ring (see figure 5). Although this solution is less reliable than one based on the dual (counter-rotating) ring and DA type of connections, it saved a large amount of money (about 30%) for the electronics and the SM fibres. Another advantage of this layout will be the possibility of 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.

---

[2] ATM : Asynchronous Transfer Mode retained by CCITT and supported by ATM Forum.

[3] FDDI : Fiber Distributed Data Interface, ANSI X3T9 or ISO 9314 standards.

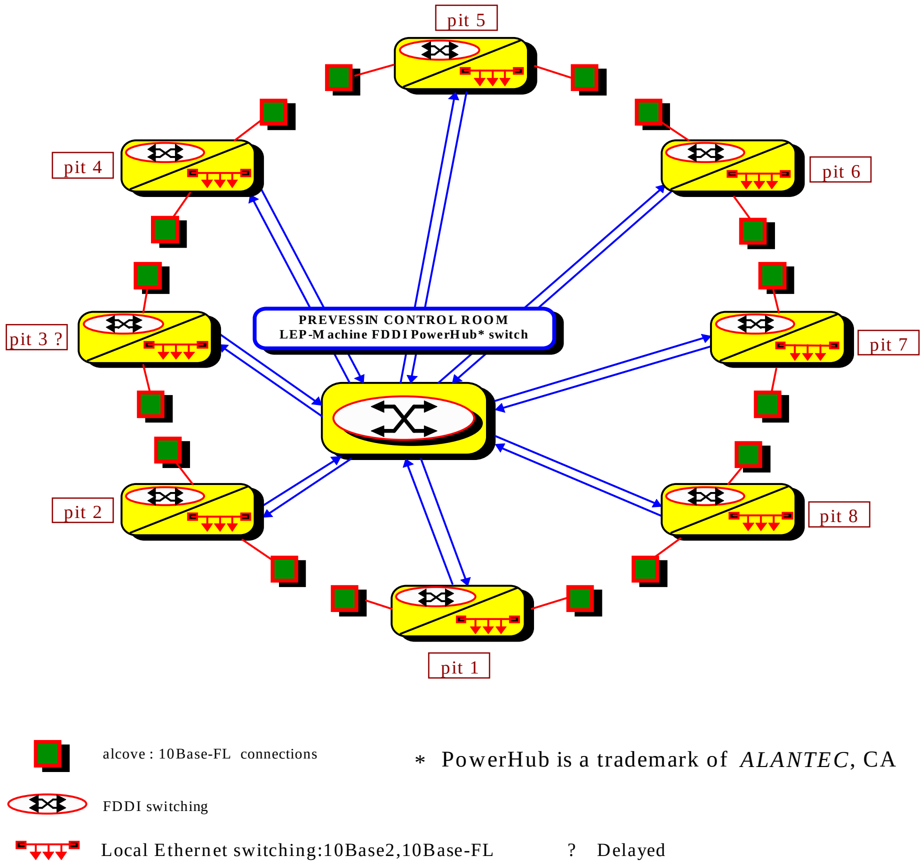

Fig. 4 NETWORK CONCEPTION BASED ON POWERHUBS* FDDI AND ETHERNET SWITCHING

THE ELECTRONICS.

The specification for the new network equipment called for:

- switch and bridge FDDI
- switch and bridge Ethernet from FDDI
- route multi-protocols, essentially the IP protocol suite and to 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

It also included several options. Among them were:

- full-duplex Ethernet
- TPPMD[4] FDDI
- 100 Mb/s Ethernet with a preference for 100VG-AnyLAN[5], our workstations being HP ones

---

4 TPPMD : Twisted Pair Physical Media Dependent.

5 100VG-AnyLAN is defined by IEEE 802.12 standard.

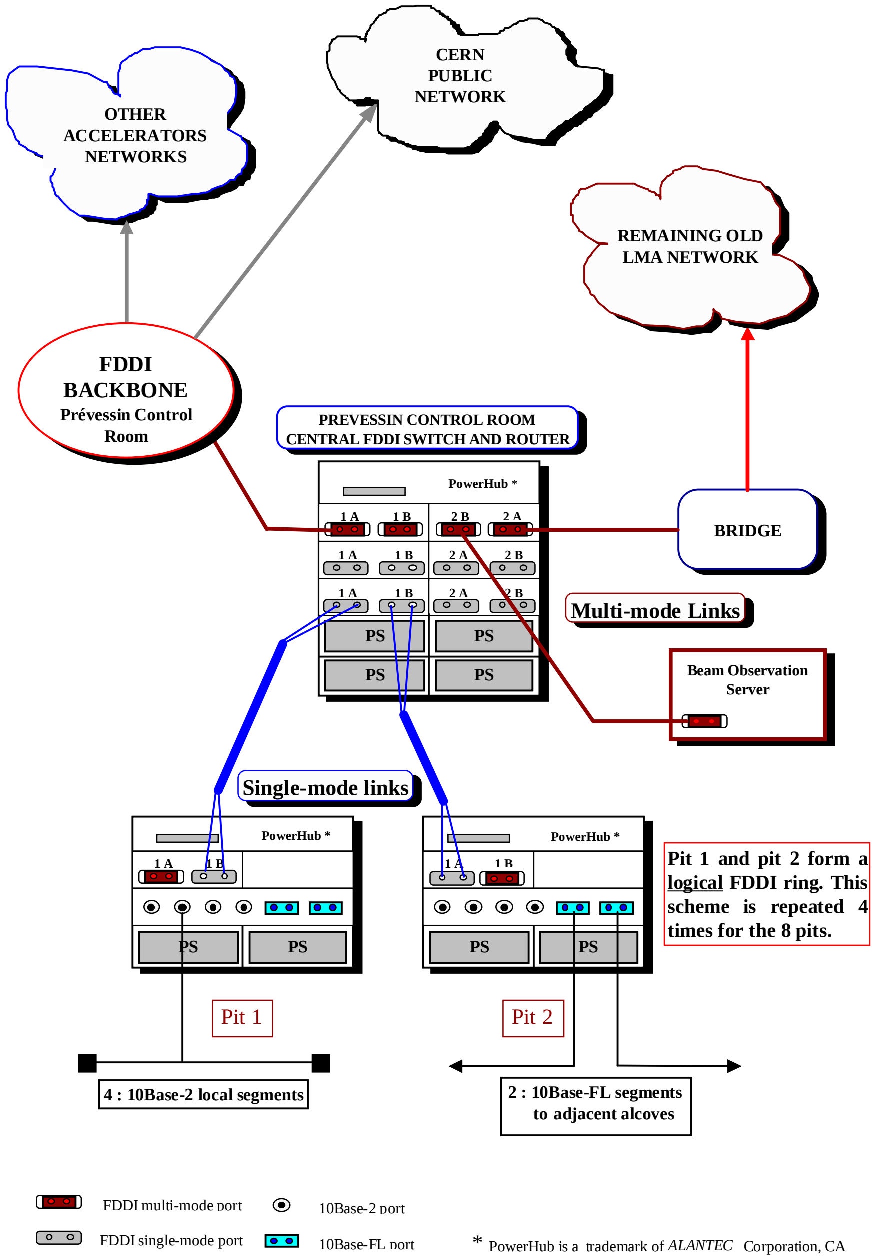

Fig. 5 LEP MACHINE LAYOUT AS FROM MARCH 1995.

All these features together led us to think that we had no chance of finding such a product on the market, but 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 firms with products closely following the specification.
PowerHub 7000 (PH) from ALANTEC was selected because it completely fulfilled the compulsory requirements and the optional features were also available, apart from the 100VG requirement.

The PH is expandable and scalable from 1.6 to 3.2 Gb/s total channel bandwidth. Its multiple-port shared memory provides and shares bandwidth at the chip level. This means that the PH software can forward packets using any bridging, routing or filtering method which can be expressed by a programmed algorithm. It 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.

The LEP network, like LEP Services and the SPS accelerator networks, uses SNMP[6] and RMON[7] protocols, via the HP OpenView Interconnect management product and the HP Metrics statistics and analyser tool, to control all the network equipment. These were all retained 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.

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 either to find radiation-resistant fibres or a special path to pull normal ones. Alcoves were linked to the pits with fibres in the drain-pipe, except between pits 1 and 4 where special engineering was needed to clear the drain.

ELECTRONICS.

PowerHubs were installed down all the pits, except pit 3, with the following configuration:

- 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 PowerHub in the 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 pit PH and using optical to copper converters to create the local Ethernet segment.

CONFIGURATION.

One of the MM ports on the PCR PowerHub was configured as the router port between the 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 using VLAN, unless one considers 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 :

- the power supplies did not comply with European standards
- some rare losses of a FDDI connection disturbed the PCR backbone
- some rare IP routing lock-out on the PCR port isolated the LMA subnet

---

[6] SNMP : Simple Network Management Protocol.

[7] RMON : Remote MONitoring protocol.

We have been testing some new power supplies proposed by ALANTEC and they have proved satisfactory, so we are now ready to install them. Upgrading the FDDI and OS software versions seems to have solved the problem of losses. We have now been running the latest 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 of three on the response time of a standard 64-byte ICMP[8] echo message crossing all the network. Statistics show extremely low error rates over long periods, for example:

PCR PH :

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

US25 PH :

FDDI : 69 million packets over 410 hours and 43% peak utilisation, zero dropped .
Ethernet port 13 : 7 million packets with 21% peak utilisation showed 21 Xmit buffer errors, 1 Frame Check Sequence error, 1 Frame Alignment error and 262 collisions. All other ports have 0 error counters.

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

4.2 COMPLETION OF THE FIRST PHASE

During the 1995-96 shutdown the drain-pipe preparation will be completed between pits 1 and 4 via pits 2 and 3. Then we will be able to pull the fibres in the drain. We now have a solution for the connection of pit 3 with normal fibres. The first phase will be completed by 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 in 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 meet LEP 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 installed in the PCR, switching will also provide us with a solid base for multimedia communications.

[8] ICMP : Internet Control Message Protocol.

7. ACKNOWLEDGMENTS

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.

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 1993. NIM Vol. A. 352, Nos. 1,2.
[2] B. Hall and D. Swoboda, The ACCI code conversion interface for the LEP Controls token ring, LEP Controls note 84, CERN, 25 February 1988.