Metro Ethernet Networks (MEN)

Luis Velasco, Jordi Perelló, Gabriel Junyent
Optical Communication Group - Universitat Politècnica de Cataluya (UPC)
E-mail: luis.velasco@tsc.upc.edu, jper7775@alu-etsetb.upc.edu,
gabriel.junyent@tsc.upc.edu

Ethernet is the predominant technology in Local Area Networks (LAN) and it is also becoming a technology of reference in the access networks, specifically in Wide Area (WAN) and Metropolitan Area Networks (MAN). Its objective is to provide connectivity between geographically dispersed locations of clients, as if they were connected to a same LAN.

Adding to this the fast increase of the bandwidth demand for the transport of data and the availability of faster optical Ethernet interfaces at lower prices, we can think that it is possible to incorporate Ethernet technology to carrier networks.

On the other hand, limits between packet switching networks and circuit switching networks are disappearing and it is possible to provide similar services using both types of networks. For example, the new generation of SDH (with LCAS, GFP and VCAT) provides services of circuits and data in a flexible and trustworthy form.

From the economic point of view, it seems clear that the costs of implantation (CAPEX) and operation (OPEX) of the Ethernet technology are smaller than those of the SDH based networks. Nevertheless, to allow the Ethernet technology to be used in carrier networks, it is necessary to add it a set of essential characteristics that will allow it to offer services of quality.

This article reviews the mechanisms that will allow the deployment of Ethernet in carrier metropolitan networks, also referred to as *Optical Carrier-class Ethernet*.

1 INTRODUCTION

A Metropolitan Ethernet Network (MEN) is a network that connects geographically separated LANs directly or through a WAN, using Ethernet as main the protocol.

As seen in Figure 1, nodes of a MEN can be switches or routers, depending on their location in the network, the service that they provide and the protection needed. Links are point to point of any Ethernet speed (from 10Mbit/s to 10 Gbit/s).

MEN Networks are meshes of the necessary degree to provide the connectivity, services and the wished level of protection, and are interconnected with other MENs by means of WAN links.

Ethernet services [MEF-1] [MEF-6] can be classified in point to point (E-Line) or multipoint to multipoint (E-LAN) (See Figure 2):

- Ethernet Line Service (E-Line). It provides a point to point Ethernet Virtual Connection (EVC). It is analogous to use Frame Relay PVCs, or TDM Leased lines.
- Ethernet LAN Service (E-LAN). It provides multipoint connectivity. Sent information can be received by several points. Each end is connected to a

multipoint EVC and when a new location is added, it is only necessary to add the new site to the multipoint EVC.

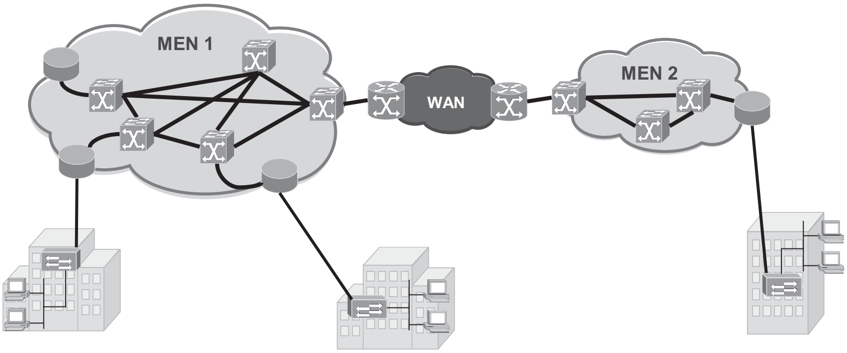

Figure 1 – Example of interconnected MENs.

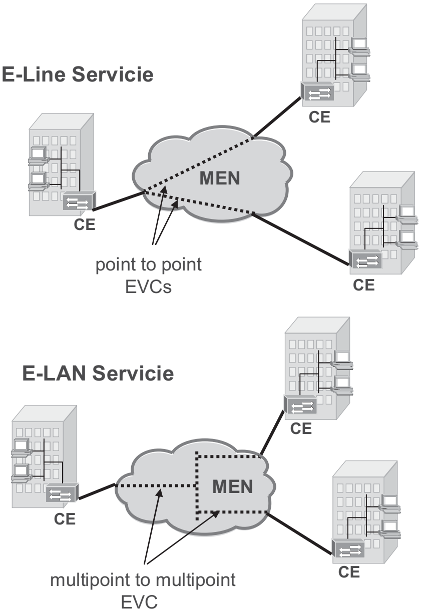

Figure 2 – E-Line and E-LAN Services.

Deployment of gigabit Ethernet is based on the following drivers:

- Cost. The cost of the GbE equipment is significantly smaller than the one of FR or ATM, due to its relative technical simplicity and the economies of scale. In addition the operational cost is significantly lower than the one of TDM (PDH and SDH) and with smaller cost of implantation [ROI].

- Fast on-demand provisioning. Ethernet services offer a big rank of speeds (1Mbit/s to 1Gbit/s) in increments of 1Mbit/s and they can be provisioned in a fast on-demand way.
- Packet based: Ethernet is an asynchronous technology based on frames that provides advantages by its flexibility over its more rigid competitors SDH and ATM.
- Ease of internetworking. It eliminates a layer of complexity (SDH and ATM) of the access, enabling simpler integration of client systems and networks and making the transport more efficient.
- Omnipresent adoption. Ethernet is the dominant technology in the LANs and there are standard interfaces for 10/100/1000/10000Mbit/s. Implications extend to benefits such as their training simplicity compared to ATM and SDH.

A set of limitations can be identified when using pure Ethernet as a transport protocol, respect ATM or SDH. The adoption of Ethernet as a universal transport layer in the metropolitan area will depend on the resolution of these limitations:

- Scalability and use of the resources of the network. Because VLAN id is 12 bits length, the maximum number of VLANs in a domain is limited to 4096.
- Protection mechanisms. The loss of a connection is handled via STP which takes several seconds in acting as opposed to the 50ms of SDH. This time is critical in applications of voice and video. In addition, it presents a low capacity of failure isolation. SDH has alarms like LOS, RDI, etc.
- Transport of TDM traffic. If it is desired to construct a multiservice network, it is necessary to transport TDM circuits, like E1, E3 or STM-1.
- End to end QoS guarantees. Ethernet needs the following mechanisms:
    - Planning, to assure that the service can be guaranteed in case of congestion.
    - Connection admission for new services requests.
    - Establishment of optimal path through the network. At the moment the spanning tree protocol (STP) is used.
    - Packet priorization.
- In service OAM. Ethernet does not have error rate monitorization capacity as SDH makes with bytes BIP-8 of the overhead.

The remaining of the article the mechanisms that will allow eliminating the previous limitations for the deployment of Ethernet as universal transport layer in the metropolitan area are reviewed.

2 ENCAPSULATION

To support technologies that allow scalable services based on Ethernet, like Transparent LAN Services (TLS) to connect several client locations by means of a MEN, it is necessary to have some encapsulation mechanism.

The more important aspects to deploy Ethernet in metropolitan networks are scalability, separation of clients and the limited size of MAC addresses table.

The encapsulation schemes insert/extract fields or additional labels in the client Ethernet frames at the edge nodes. In order to select a scheme over the others it is necessary to consider backward compatibility, performance and complexity.

2.1 SCALABILITY

2.1.1 MAC ADDRESSES

Ethernet switches learn MAC addresses of remote machines and they associate them with ports at which the Ethernet frames arrive.

If Ethernet switches were used in the core of the metropolitan network each switch would have to learn the MAC address of each remote machine of each client VLAN connected to the metropolitan network. This is known as *explosion of the MAC addresses table*.

2.1.2 VLAN

A VLAN is a logic LAN over a physical Ethernet shared network, as defined in IEEE 802.1Q. This standard defines a Q label: the VLAN identifier (VID), which is inserted in the Ethernet frames. The VID is 12 bits length, reason why the maximum number of different VLANs in a domain is 4096. Since the network is used by different clients, the VLAN identifier of each client has to be managed to assure that duplicated VIDs do not exist.

2.2 ENCAPSULATION MECHANISMS

2.2.1 VLAN STACKING: Q-IN-Q ENCAPSULATION

This mechanism inserts an additional Q label in the client frames arriving to the edge switch of the MEN (see Figure 3). Combining the client VID labels and the MEN labels, the space of VLANs is increased beyond the limit of the 4096.

This scheme is backward compatible and it has been introduced in the IEEE 802.1ad specification.

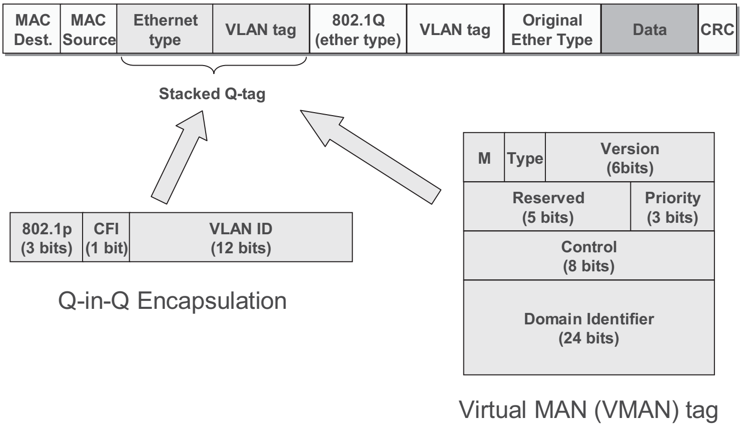

Figure 3 - VLAN Stacking.

2.2.2 VLAN STACKING: VIRTUAL MAN (VMAN) LABEL

A new 24 bits length label, called VMAN, is introduced in order to increase the number of client VLANs on the MAN (see Figure 3). This way it is not necessary to restrict the client VIDs and the number of VLAN transported on the MAN is increased.

Although the forwarding mechanisms, protocol stack, etc. are basically the same as in the IEEE 802.1Q architecture, this scheme is not compatible with existing switches.

2.2.3 LAYER 2 MPLS ENCAPSULATION

Layer 2 MPLS encapsulation (also known as Martini encapsulation [Martini]) facilitates the transport of Ethernet frames through MPLS domains (see Figure 4).

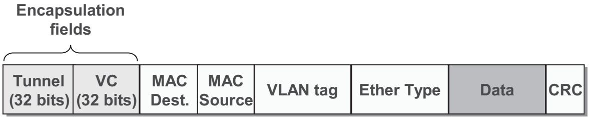

Figure 4 - Martini Encapsulation.

The ingress node (LER, MPLS Label Edge Router) inserts two MPLS labels in the client Ethernet frames, based on the destination information (MAC address, port and Q label).

- *Tunnel label*: It is used to transport frames through a MPLS domain. This label is eliminated by the penultimate LSR (MPLS Label Switch Router).
- *Virtual circuit label (VC)*: Used by the egress LER to determine how to process the frame and where to send it towards its destination.

The LER has to perform two functions: Ethernet bridge, learning client MAC addresses, and MPLS forwarding based on the LSP. The LER has to map learned MAC/VID addresses to a pre-established LSP transporting Ethernet frames through the MPLS domain.

The use of MPLS as an encapsulation mechanism provides additional advantages more over the scalability problem. For example, with the MPLS encapsulation, the Ethernet frames can be transported over any type of network. In addition, all the characteristics of OAM, protection mechanisms, traffic engineering and bandwidth guarantees of MPLS, are introduced automatically.

2.2.4 TECHNIQUES COMPARISON

Since MAC addresses must be learned only in the LERs of the domain, using MPLS the explosion of the table of MAC address is avoided.

From the perspective of the added overhead:

- Q-in-Q it introduces 4 bytes of overhead.
- VMAN introduces a minimum of 6 bytes.
- MPLS introduces a minimum of 8 (2*4) bytes, until a maximum of 30 bytes.

For small frames, about 64 bytes, it introduces an overhead of up to 46%.

Q-in-Q encapsulation provides scalability without adding a significant complexity, but MPLS provides a set of characteristics, like traffic engineering and reliability, desirable for carriers. Since both technologies are complementary, they can be used together; Q-in-Q in the access network and LERs and MPLS in the core network.

3 OPERATION, ADMINISTRATION AND MAINTENANCE (OAM)

With the introduction of traffic sensible to the real-time and to the quality of service (QoS), like voice and video services, it becomes necessary to control the switching, routing and the delivery of the packets corresponding to these services. In the last times an important effort is being made within the ITU and the IETF, to reflect the requirements of carriers in MPLS's OAM function [Y.1710] [Y.1730] [ReqOAM] and several standards with new and improved OAM functions have appeared.

- The mechanisms impelled by the ITU and the IETF, are:
- Connectivity Verification (CV) and Fast Failure Detection (FFD).
- Forward Defect Indication (FDI) and Backward Defect Indication (BDI).
- MPLS LSP ping.
- Bidirectional Forwarding Detection (BFD).
- LSR Self Test.

In the next paragraphs we will present these mechanisms.

3.1 CONNECTIVITY VERIFICATION AND FAST FAILURE DETECTION

The Connectivity Verification (CV) and Fast Failure Detection (FFD) mechanisms proposed by the ITU [Y.1711], allows detecting and diagnosing end to end connectivity defects in a LSP.

The flow of CV packets, with a periodicity of 1 packet/sec, has origin in the ingress LSR of the LSP and goes towards the egress LSR of this LSP. Its purpose is the diagnosis of possible errors in reception: Loss of packets, Reception of packets with another destination, etc, as it is observed in Figure 5.

FFD mechanism is identical to CV, except in that it allows the variation of the frequency of the packet flows, thus allowing fast detection of failures. The recommended value is 20 packets per second (a packet each 50ms).

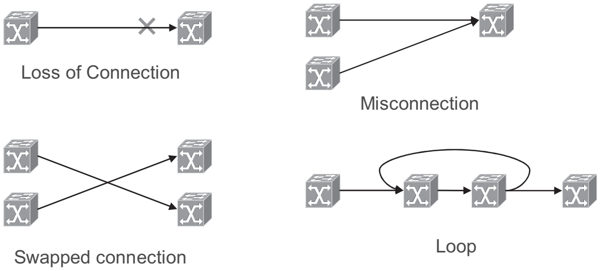

Figure 5 – Types of LSP defect.

3.2 FORWARD DEFECT INDICATION AND BACKWARD DEFECT INDICATION

The Forward Defect Indication (FDI) and Backward Defect Indication (BDI) mechanisms are proposed by the ITU [Y.1711].

The objective of FDI is to suppress the alarms produced in the client LSPs of a LSP affected a failure.

FDI packets have a periodicity of 1paq/sg and they are sent from the first node detecting the failure to the LSP end point. If the error has taken place in the server layer, it will be the first node following the failure. If the error has taken place in the MPLS layer it will be the end point of the LSP of the level in which the failure has been produced.

BDI mechanism informs, with a periodicity of 1paq/sg., to the source end of a LSP of a failure observed at destination. BDI needs a return path, which can be a dedicated LSP, a LSP shared by several LSPs, or a non MPLS return path.

Figure 6 briefly describes FDI and BDI mechanisms.

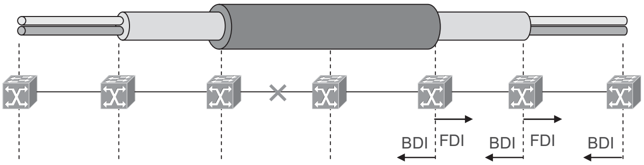

Figure 6 – FDI and BDI mechanisms.

FDI and BDI can be useful to measure the network availability or to trigger the switching event in protection mechanisms.

3.3 MPLS LSP PING

MPLS LSP Ping, proposed by the IETF [LSPPing], has the purpose of verifying that packets corresponding to a certain Forwarding Equivalent Class (FEC) end its MPLS path in the correct end node for that FEC. MPLS LSP Ping packets are sent by the ingress node towards the egress node, following the same path as the packets corresponding to the tested FEC. It has two operation modes:

- *Basic connectivity check*: the echo request packet arrives at the egress node and it is sent to the control plane to verify if the LSR is really the end node for that FEC. Once the verification has been made, the end LSR will send an MPLS LSP Ping echo reply reflecting the result of this verification.
- *Traceroute*: the packet will be sent to the control plane in each transit LSR. The transit LSP will verify that it is really a transit LSR for that FEC.

3.4 BIDIRECTIONAL FORWARDING DETECTION

As we have seen above, LSP Ping is a mechanism capable to detect failures in the data plane and to make an analysis of the data plane as opposed to the control plane. The Bidirecional Forwarding Detection (BFD) mechanism, also proposed by the IETF [BFDBase][BFDMPLS], is a mechanism designed to detect failures only in the data plane with a smaller computational cost compared to LSP Ping, allowing fast detection of failures (<1s compared to the several seconds of LSP Ping) and supports failure detection of bigger number of LSPs. In addition, thanks to its packet fixed format it is easier to implement in hardware.

3.5 LSR SELF TEST

The LSR Self Test mechanism proposed by the IETF [LSP-ST], defines a mechanism to allow an LSR test its label associations and the connectivity with the LSRs to which it is directly connected. LSR Self Test can be used in unicast LDP tunnels and in RSVP based tunnels.

4 PROTECTION MECHANISMS

SDH based transport networks provide traffic protection schemes with restoration times lower than 50ms. This characteristic allows that link connectivity losses, for example due to an optical fiber breakage or equipment card failures, do not have impact over the service provided to the clients.

On the other hand, pure Ethernet traditional solutions provide protection by means of standard mechanisms based on the *spanning tree* protocol, originally designed to recover failures in 30sg.

Requirements and protection mechanisms objectives for MEN are proposed and described in [MEF-2].

4.1 SPANNING TREE

Reaching a high availability is difficult using traditional bridging by means of the Spanning Tree Protocol (STP), defined in 802.1d. Spanning Tree prevents the appearance of loops and provides a back up mechanism in case of connection or port failure (see Figure 7).

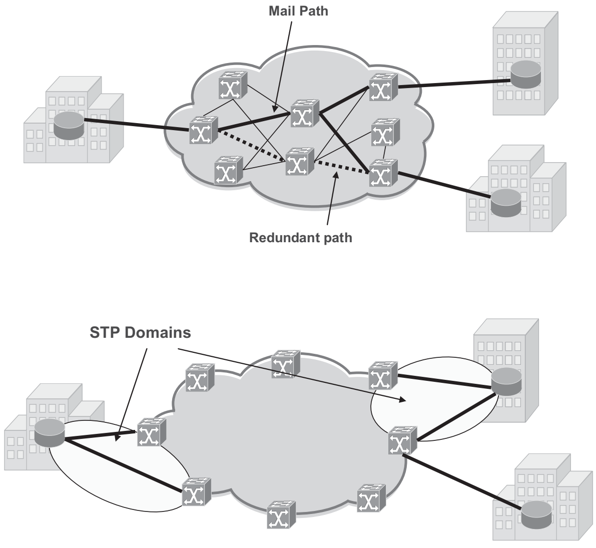

Figure 7 - Spanning Tree Protocol.

Nevertheless, the slow convergence time of STP is inadequate to support quality services. Depending on the network topology, the Spanning Tree Protocol can take between 30 seconds and several minutes to recover from a failure.

Although fast versions of STP exist, this protocol is not able to provide protection below 50 ms, threshold used by the carriers. Nevertheless, STP protocol can be supported in the access, as it is showed in Figure 7.

4.2 PROTECTION BY MEANS OF REDUNDANT CONNECTIONS

To provide protection times below 50 ms two ways of protection have been defined (see Figure 8):

- Aggregate Link and Node Protection (ALNP): It uses a detour LSP to avoid the resource with failure.
- End to End Path Protection: It uses an end to end protection path.

The detection of Loss Of Signal (LOS), Loss Of Link (LOL), Loss Of Frame (LOF) and Loss Of Sync in the Ethernet connection can be used to send protection events.

Moreover, Ethernet uses the 8B/10B line code for clock recovery and power balance. This code is also used to detect physical link degradation, measuring the error rate (BER). Thresholds of BER can be used to trigger protection events.

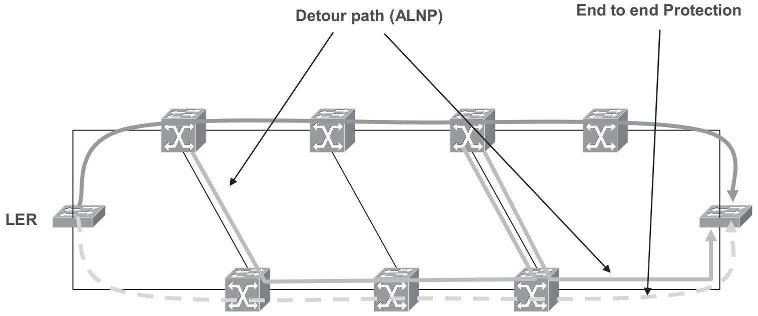

Figure 8 - Protection by means of redundant connections.

4.2.1 AGGREGATE LINK AND NODE PROTECTION (ALNP)

ALNP provides local protection of multiple links or nodes through the network, using detour LSPs created using disjoined resources of which main path uses (see Figure 9). When a failure is detected:

- the last element in the path before the failure resource adds an additional MPLS label to reroute the traffic from the primary LSP to the detour LSP,
- the failure resource is avoided and the following element in the main path subsequent to the resource with failure is reached,
- this element eliminates the additional label and sends the traffic by the main path.

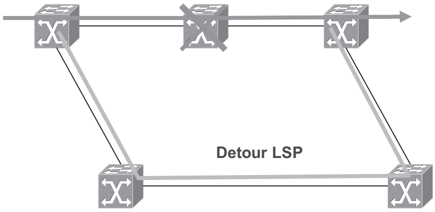

Figure 9 - ALNP Protection.

Bandwidth reserved for the detour LSP can be used for extra traffic in dedicated protection 1:1 or shared 1:n architectures when they are not used for protection.

4.2.2 END TO END PATH PROTECTION

The end to end path protection creates two or more redundant end to end LSPs between the ingress and the egress node [Y.1720]. The ingress and the egress node sent connectivity verification messages (CV or FFD) among them.

In the 1:1 protection architecture the ingress node sends the traffic through the main path. When a failure is detected, the reception endpoint sends a backward defect indication (BDI) packet to the transmission endpoint to switch the traffic to the protection LSP.

In the 1+1 protection architecture, the ingress node sends the traffic using both paths simultaneously to obtain a protection switching time lower than 50 ms.

5 CIRCUIT EMULATION

Circuit Emulation Services (CES) allows the transport of constant bit rate synchronous circuits, like E1 (2Mbit/s), E3 (34 Mbit/s), STM-1 or STM-4, over variable bit rate asynchronous networks. CES provides support to the traditional TDM voice applications and leased lines.

The Metro Ethernet Forum [MEF-3] has defined four general types of services:

- **TDM Access Line Service** (TALS). In this type of service, at least one of the end points finishes in another network (for example, PSTN), and it allows the transport of voice, Frame Relay and ATM circuits over Ethernet networks. The service is provisioned and managed by the MENs service provider.
- **TDM Line Service** (T-Line). In this type of service, the end points belong to a company. The service is provisioned and managed by the MENs service provider.
- **Service operated by the client**. The service is managed by the client.
- **Mixed way**: Any combination of the three previous services.

Figure 10 shows an example of T-Line and TALS services.

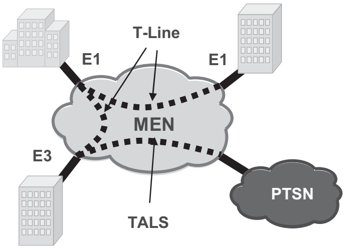

Figure 10 – Circuit Emulation Services.

5.1 T-LINE SERVICE OPERATION MODES

It is possible to provide multiplexed services, for example aggregating several E1 in a E3 or STM-1 link, creating configurations point to multipoint or multipoint to multipoint. This multiplexing service is made at the TSP optional block, that processes TDM services [MEF-3][MEF-4].

There are three possible operation modes, the first two are point to point and the third one allows multipoint configurations.

- **Unstructured mode**: Service between points with the same type of interface. The traffic is transported in a transparent way from one end to the other. For example, leased lines.
- **Structured mode**: Service between points with the same type of interface. The traffic is handled as overhead and payload. The overhead is terminated in the end points and the payload is transported transparently from one end to the other. For example, a STM-1 containing a VC-3.
- Multiplexed mode: Several lower rate services are multiplexed into a higher hierarchy level. Although service multiplexing is usually performed in the TDM domain, the emulation service is structured.

5.2 TALS SERVICE OPERATION MODES

TALS service is very similar to the T-Line multiplexed service. Both modes use the MEN similarly, unless in the TALS service, the final multiplexed service is handled by another network instead of by the end user. For that reason it has some additional performance requirements.

The MEN must maintain bit integrity, clock and other specific characteristics of the transported traffic format, without causing degradation that exceeds the offered service requirements. In addition, all the management, monitoring, etc. functions must be performed without affecting the transported service.

5.3 CES REQUIREMENTS

Some of the CES requirements are:

- **Packetization**: Process to convert a synchronous traffic flow into Ethernet frames. It requires the introduced delay to be constant and as low as possible. It is also possible to encapsulate multiple frames of synchronous flows to reduce the latency of the process.

  - **Latency (frame delay)**: Delay between the MEN entry point of the TDM flow, to the exit point. If it is very high it introduces the necessity of echo cancellation in telephone applications. MENs are able to provide latencies below 10 ms, which makes it possible to provide circuit emulation with no need of echo cancellation.
  - **Frame delay variation (jitter)**: The variable delay introduced by the MEN is due to the asynchronous nature of Ethernet switching and the variety of lengths of the frames crossing the MEN. The delay variation can be compensated using buffers (*jitter buffers*) at destination, with the cost of increasing latency.
  - **Frame loss and resequencing**: The frames may not arrive in the same order in which they were sent. The destination node must rearrange the frames, using the sequence number field present in the overhead of the frame. The *jitter buffer* must be able to verify the sequence number of the arriving frames and rearrange them if necessary. All this must be done maintaining the buffer size as smaller as possible in order to minimize the latency.

- **Clock recovery and synchronization**: Circuits transport clock information that is used to synchronize the transmitter and the receiver. If clock differences exist between the transmitter and the receiver, some information will be lost, causing a reduction of the quality of the circuit. A clock recovery mechanism, that resists latency, jitter and the loss of frames, has to be used.

6 CONCLUSIONS

The combination of high speed Ethernet at reduced prices with optical switching can fulfill the growth of the bandwidth demand, and it can represent one alternative, in the metropolitan scope, to the SDH technology traditionally used by carriers.

Optical Ethernet provides the platform to construct big Metropolitan Ethernet Networks, providing quality services incurring in total costs (TCO), implantation (CAPEX) and operation (OPEX) much smaller than those of the alternative technologies (New Generation SDH and Ethernet over WDM).

To allow that, is necessary to add a set of mechanisms (protection, OAM, emulation of circuits, Engineering of traffic, QoS, etc.) to pure Ethernet technology so it can fulfill the strict requirements that carriers have.

These requirements have been specified by standardization organizations, mainly ITU and IETF, and in the next months it is expected the appearance of equipment able to fulfill all these requirements.

REFERENCES

[MEF-1] Metro Ethernet Forum. Technical Specification MEF 1 “Ethernet Services Model, Phase 1”, Nov. 2003

[MEF-2] Metro Ethernet Forum. Technical Specification MEF 2 “Requirements and Framework for Ethernet Service Protection in Metro Ethernet Networks”, Feb. 2004

[MEF-3] Metro Ethernet Forum. Technical Specification MEF 3 “Circuit Emulation Service Definitions, Framework and Requirements in Metro Ethernet Networks”, Abr. 2004

[MEF-4] Metro Ethernet Forum. Technical Specification MEF 4 “Metro Ethernet Network Architecture Framework - Part 1: Generic Framework”, May. 2004

[MEF-6] Metro Ethernet Forum. Technical Specification MEF 6 “Ethernet Services Definitions - Phase I”, Jun. 2004

[ROI] Metro Ethernet Forum, “Comparison to Legacy SONET/SDH MANs for Metro Data Service Providers” July 2003.

[Martini] Martini, et al. “Encapsulation Methods for Transport of Ethernet Frames Over IP/MPLS Networks”, http://www.ietf.org/internet-drafts/draft-ietf-pwe3-ethernet-encap-08.txt, Sep 2004.

[Y.1710] ITU-T Y-1710, “Requirements for OAM functionality for MPLS networks”, Nov. 2002.

[Y.1711] ITU-T Y-1711, “Operation & Maintenance mechanism for MPLS networks”, Feb. 2004.

[Y.1720] ITU-T Y.1720, “Protection switching for MPLS networks” Sept. 2003.

[Y.1730] ITU-T Y-1730, “Requirements for OAM functions in Ethernet-based networks and Ethernet services”, Ene. 2004.

[ReqOAM] T. D. Nadeau et al., “OAM Requirements for MPLS Networks,” IETF Internet draft draft-ietf-mpls-oam-requirements-05.txt, Dec 2004.

[LSPPing] Kompella et al., “Detecting MPLS Data Plane Failures,” IETF Internet draft draft-ietf-mpls-lsp-ping-07.txt, Oct.2004

[BFDMPLS] R. Aggarwal et al., “BFD for MPLS LSPs,” IETF Internet draft draft-ietf-bfd-mpls-00.txt, Jul. 2004.

[BFDBase] D.Katz, D.Ward, “Bidirectional Forwarding Detection” IETF Internet draft draft-ietf-bfd-base-00.txt, Jul. 2004.

[LSP-ST] G. Swallow, K. Kompella, D. Tappan, “Label Switching Router Self-Test,” IETF Internet draft draft-ietf-mpls-lsrself-test-03.txt, Oct. 2004.

[RFC3469] V. Sharma, F. Hellstrand, “Framework for Multi-Protocol Label Switching (MPLS)-based Recovery” IETF RFC 3469, Feb. 2003.