Metro-Ethernet

Nicolas Guérin (guerinn@eurecom.fr)
Alexandre Villoing (villoing@eurecom.fr)

September 06

Index

1 General overview

1.1 Metro-Ethernet Definition

Today, Ethernet is the most used technology in Internet customers Local Area Networks (LANs). Hence, Internet Services Providers (ISPs) thought about extending the Ethernet LAN technology to Metropolitan Area Networks (MAN) and create Metro-Ethernet Networks, connecting millions of corporate and end-users networks, as depicted in Figure 1.

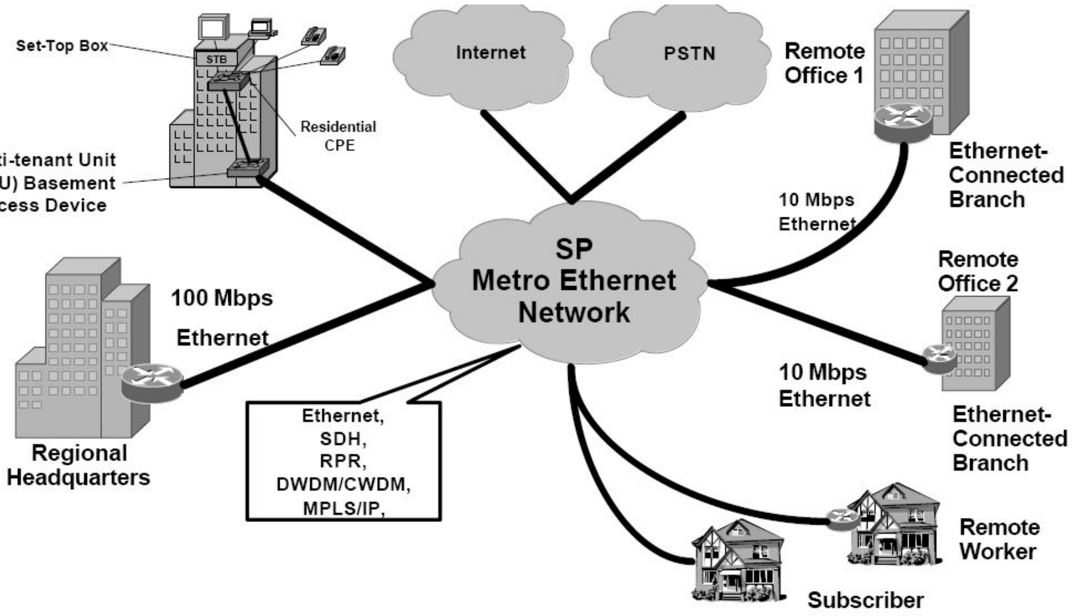


Figure 1: Overview of Metro-Ethernet Network

End-users attach their Customer Equipment (CE) to the SP network through the User to Network Interface (UNI) using a standard 10/100/1000/10000 Mbps Ethernet interface. From a customer perspective, the layer 2 technology used to connect its LAN to Internet is Ethernet.

1.2 Why Ethernet?

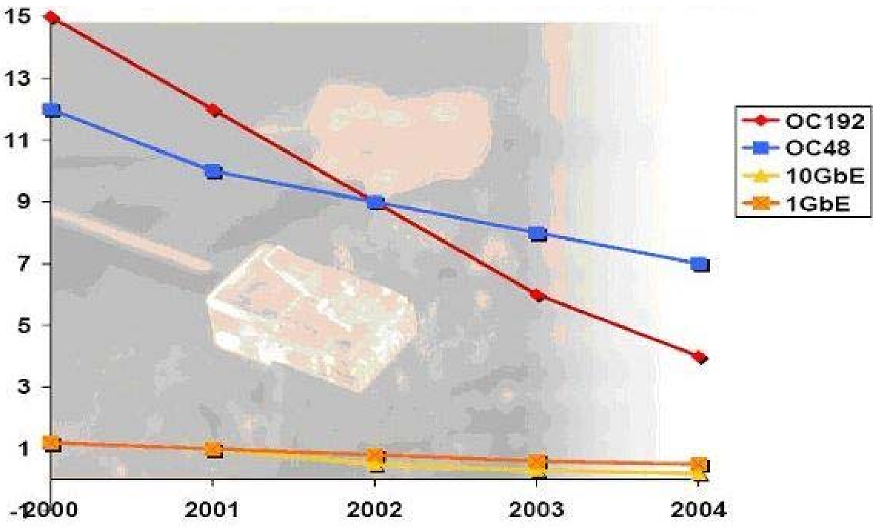
Extending Ethernet technology to Metro Area Networks is already a reality, because it has the best rate performance over cost. Furthermore, customers' LANs are already deployed using Ethernet technology. This part lists various advantages of deploying Ethernet beyond LANs.

Ease of use: Ethernet services are provided over a standard, defined by a joint-effort of IETF (Internet Engineering Task Force) and MEF (Metro-Ethernet Forum), widely available and well-understood Ethernet interface. Virtually all networking equipment and hosts connect to the network using Ethernet. Using an Ethernet service to interconnect such devices simplifies network operations, administration, management and provisioning.

Cost Effectiveness: Ethernet services can reduce subscribers' capital and operation expenses in three ways.

- First, due to its broad usage in almost all networking products, the Ethernet interface itself is inexpensive.
- Second, Ethernet services can often cost less than competing services due to cheaper equipment, service and operational costs.
- Third, many Ethernet services allow subscribers to add bandwidth more incrementally, e.g., in 1 Mbps increments. This allows subscribers to add bandwidth as needed so they only pay what they need.

Figure 2 show the decreasing price in dollars per Megabit of bandwidth of various Ethernet technologies between 2000 and 2004. OC192 and OC48 are optical norms.



Source: Dell'Oro Group

Figure 2: Decreasing Price of Ethernet Megabit bandwidth between 2000 and 2004

Flexibility: Many Ethernet services allow subscribers to build their business network in ways that are either more complex or impossible with other services. For example, a single Ethernet service interface can connect multiple remote locations to their Intranet through VPNs, or connect business partners via Extranet VPNs and provide a high speed Internet connection to an Internet Service Provider.

With managed Ethernet services, subscribers are also able to add or change bandwidth in minutes instead of days or weeks when using other access network services. Additionally, these changes do not require the subscriber to purchase new equipment and ask for a visit from a service provider technician.

Technically: Ethernet has many intrinsic technical advantages. It offers a great granularity for WANs ranging from a few megabits per second up to 10 Gigabits/s. So service providers easily satisfy customers' needs and tune their network.

Then Metro Ethernet provides a simple solution to multipoint connections support.

Last, since Ethernet is very widely used in customers LAN they already know the technology and almost no learning is required in order to switch to these new services.

Key Benefits for the Service Provider: MEF works with all the actors and vendors and also build specifications in order to ensure equipments compliance. So customers have the guaranty that all the devices are compliant with each other no matter what vendor it is issued from. This way, since customers do not need to test compliance of devices anymore they save money and time. Furthermore this interoperability maintains a wealthy competition between vendors.

Key Benefits for the Equipment Vendor: Since they will have to use the same specifications, vendors will also save money and time on approval processes by standardizing them. The installation time should also shrink because of this standardization.

1.3 Metro-Ethernet Services

1.3.1 Metro-Ethernet Services Types

By providing an Ethernet socket to the customer Service Providers (SP) are providing a full transparent connectivity that have many advantages in term of equipments. Hence, these services rely on a variety of transport technologies and protocols, such as SONET/SDH, DWDM, MPLS etc., to deliver the services (A widely accepted network model uses an MPLS or IP core).

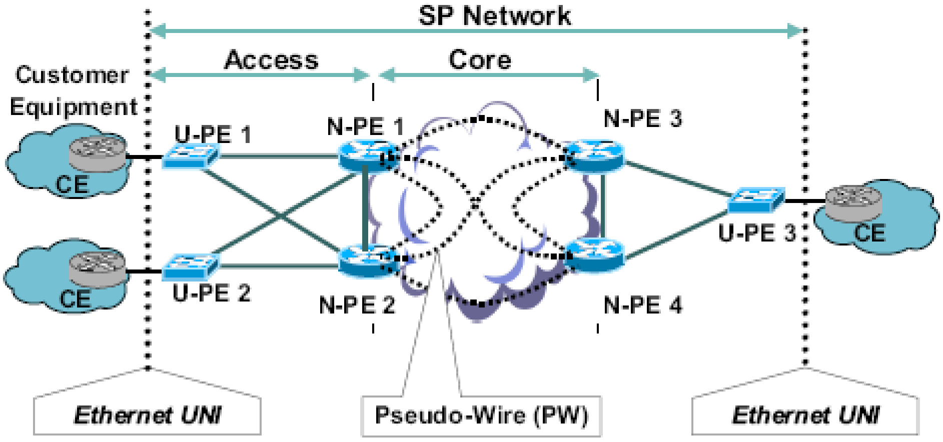


Figure 3: Metro-Ethernet Network Topology

Ethernet is the next natural evolution of customer UNI connection for both Layer-2 and Layer-3 VPN Services.

An Ethernet Service is an option among others to access a Layer-3 VPN. “Layer-3 VPN” services, especially “IP-VPN” using BGP/MPLS (according to RFC 2547) or IPSec are complementary to the Layer-2 Ethernet Services presented below.

An Ethernet Virtual Connection (EVC) is an association of two or more UNIs, where Service Frames can only be exchanged among the associated UNIs. A Service Frame sent into the MEN via a particular UNI *must not* be delivered out of the MEN via that UNI

Two basic EVC services can be identified: Point-to-point (PPP) and multipoint (MP or multipoint-to-multipoint: MP2MP) services. IETF, MEF, Service Providers and equipment manufacturers agreed to define all the different terms for Ethernet WAN Services.

The 4 following Ethernet WAN services may be coupled with higher layer services (at application layer, e.g. VoIP telephony, or at security layer, etc...) and corresponding service level agreements (SLA) that will form the actual services offering of the Service Providers.

The MEF defines two general types of services, called Ethernet Line Service and Ethernet LAN Service which are also declined into subcategories. The various extensions can be classified as shown on the following figure:

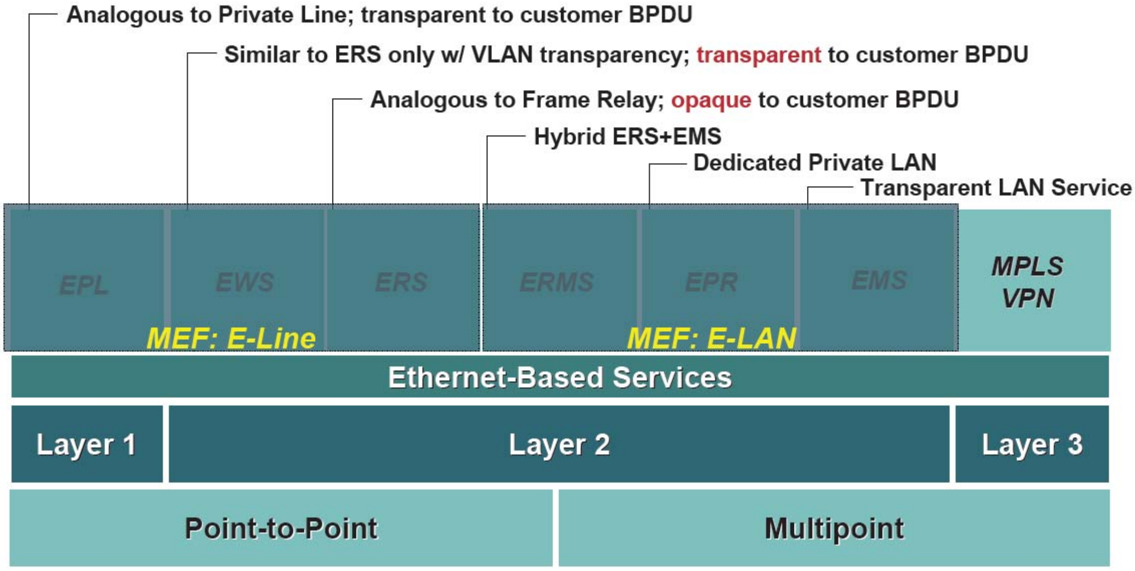

Analogous to Private Line; transparent to customer BPDU
Similar to ERS only w/ VLAN transparency; transparent to customer BPDU
Analogous to Frame Relay; opaque to customer BPDU
Hybrid ERS+EMS
Dedicated Private LAN
Transparent LAN Service
EPL
EWS
ERS
ERMS
EPR
EMS
MPLS VPN
MEF: E-Line
MEF: E-LAN
Ethernet-Based Services
Layer 1
Layer 2
Layer 3
Point-to-Point
Multipoint

Figure 4: Ethernet Services summary

1.3.2 Ethernet Line Services

Ethernet Wire Service (EWS) is the Ethernet analog of the private line service. Here, routers, bridges, or hosts may be used as customer edge devices that establish point-to-point connections between two remote sites. The attributes of a EWS are:

1. ***Integrity****:* A single remote site can be reached from a single physical connection to the SP network.
2. ***Reliability****:* A EWS ensures that all traffic received from the source Customer Edge (CE) is

delivered to the destination CE unaltered.
3. ***Transparency**:* All-to-One Bundling and VLAN transparency are employed at the expense of service multiplexing (more complex for the SP).
4. In the case of a bridge CE, customer's BPDUs (Bridge Protocol Data Units) are tunneled through the Service Provider's Layer-2 network.

In the case where the user wants to build a flat network with distinct remote sites, EWS can connect two CE bridges and allow the VLAN architecture of the service user to be extended between sites.

Ethernet Relay Service (ERS) is the Ethernet analog of the Frame Relay (FR) service. It is also often referred as to ***Ethernet Line Service***. Customer's routers or hosts establish PPP connections between two remote UNIs. Like Frame Relay, service multiplexing is a key in ERS as multiple logical, point-to-point connections are provisioned on a single physical interface that allows multiple remotes sites to be reached from a single UNI.

Instead of Data Link Connection Identifier (DLCI) used in Frame Relay, ERS uses VLAN IDs to identify the logical PPP Ethernet connections within a UNI and like FR, these values may be locally significant. Similarly, like the DLCI co-ordination between service provider and service users, co-ordination of VLAN IDs is required for ERS.

Layer-2 control protocols may be peered, tunneled or discarded by the network-side UNI, making ERS a non-transparent or "opaque" service. In the case of a bridge as CE, BPDUs may be dropped, due to IEEE 802.1q Spanning Tree operation. ERS is commonly used for:

Hub and spoke enterprise connectivity: At the hub location, a single high speed UNI is provisioned and multiple instances of ERS multiplex various services, one ERS to each branch location. This avoids using multiple physical ports on the CE at the hub and also managing and paying for multiple UNIs at the hub location.

Internet service provider (ISP) to customer connectivity: An ISP can sell Internet access in significant bandwidth increments, for example, 5–25 Mb/s, to many customers. Rather than consuming a Fast Ethernet port on the ISP router for each customer and underutilizing each port, multiple instances of ERS multiplex services via a Gigabit Ethernet UNI with each ERS going to a different ISP customer. The benefit to the ISP from the improvement in router efficiency is substantial.

Ethernet Multipoint Service (EMS) is the WAN analog of the multipoint Ethernet LAN capability, also referred to as ***LAN service***. EMS emulates a Layer-2 Ethernet Switch, operating as traditional Ethernet with ARP protocol, based on MAC addresses resolution. Broadcast/multicast frames or frames with unknown destination MAC addresses are replicated to all remote sites (except the originating one). The CE can be a router, bridge, or a host. Traffic from an originating CE may be received at one or more destination CEs unaltered.

Like EWS, all-to-one bundling and VLAN transparency are employed at the expense of service multiplexing. In the case of a bridge CE, customer's BPDUs are replicated and tunneled through the service provider's Layer-2 network to all UNIs. EMS is a good way to connect multiple remote sites of the same company together with CE routers. CE bridges can also be used but at the possible risk of suffering the ill effects of a large Layer-2 network, with MAC addresses table explosion.

1.3.3 E-LAN services in PPP configuration

An E-LAN service can be deployed to connect only two remote UNIs. Whereas this appears to be similar to an Ethernet Line Service, there are significant differences

With an E-Line service, when a new UNI is added, a new EVC must be added to all other sites to achieve full connectivity when using the E-Line service. The FR analogy would be to add a FR Private Virtual Circuit (PVC) between each site.

With an E-LAN service, only the new UNI needs to be added to the multipoint EVC. Hence, no additional EVCs are required for the new UNI to communicate with all other UNIs. An E-LAN service requires only one EVC to achieve multi-site connectivity.

To summarize, an E-LAN service can interconnect large number of remote sites with less complexity than meshed connections implemented using PPP networking technologies such as FR and ATM.

The service attributes differentiating the service types can be grouped under the following categories:

- Ethernet Physical Interface: 10BaseT, 100BaseT, 1000BaseSX
- Traffic parameters (bandwidth profiling)

The bandwidth profile is similar in concept to the traffic policing of Frame Relay, and corresponds to a characterization of the lengths and arrival times of ingress Service Frames at the UNI. The level of compliance with the Bandwidth Profile is assessed for each ingress Service Frame.

- Class of Service (CoS), Prioritizing and QoS (Quality of service)
- Performance parameters:
    - Availability: currently under definition by the MEF
    - Frame Delay

Frame Delay is a critical parameter that has considerable impact on QoS for real-time applications such as VoIP. It is defined as the maximum delay measured for a percentile of successfully delivered CIR-conformant service frames over a time interval.

    - Frame Jitter

Jitter is a measure of the variability over time of the latency across a network. A very low amount of jitter is important for real-time applications using voice and video.

    - Frame Loss
- VLAN tag support
- Service multiplexing
- Security filters

2 Challenges toward extending Ethernet to MAN

Compared to technologies like ATM and FR, Ethernet was not designed to be used in WANs. Initially thought as a simple shared medium access protocol, it has evolved into a full-duplex protocol for switched-networks. But still, its initial simple design has a great influence on its actual state, and if it is the most used protocol in LANs it needs to be adapted in order to suit MANs and WANs.

From a carrier perspective Ethernet has still to face some issues in order to be used.

First, scalability, Ethernet was initially designed for little networks with few nodes. Since it basically relies on broadcast in a network with a loop free topology and with a flat address plan it does not sound very scalable. Even though some improvements have been implemented, Ethernet still needs to be adapted to large networks in order to be reliable and provide services, like prioritization and load balancing.

Another issue is about operation, administration and maintenance (OAM), Ethernet has always had very limited features for this purpose so far. There are many reasons for that, but they are more or less all correlated to the fact that Ethernet products are targeted for LANs, in other words, networks that are not large enough to justify the use of sophisticated management tools.

Nowadays security is a very important issue in networking in general. Networks are carrying more and more critical information about your business or your private life. Furthermore original network design did not include security since the users where suppose to all be nice. Now it is one of the major concerns.

2.1 Scalability

Ethernet relies on spanning trees (ST) in broadcast domains and forwarding tables on switches. Currently, both mechanisms do not allow scaling Ethernet to MAN or WAN connections.

In the MAC address resolution protocol, the MAC table has to record each computer connected to a switch. This table cannot exceed a critical size, and causes a problem called the MAC addresses tables explosion issue.

In a bridged network the layer 2 switches periodically refresh their source MAC address tables. When millions of MAC addresses have to be refreshed, a new MAC address learning is painstaking, with the risk to create broadcast storms and latency problems. Even splitting the network into separate domains (using VLANs), some customers might have a huge number of Ethernet NICs spread all over the WAN. Therefore, efficient methods are required to optimize MAC address management.

By introducing VLAN tags in 802.1Q it was possible then to partition the problem into equivalent smaller units that could share the same physical links. But as there are only 4096 different VLAN ID (VID) available it's not enough to fill WANs needs.

What Ethernet mainly misses is hierarchy. It would reduce the problem enough to make Ethernet suitable for MANs and WANs. There are currently few solutions available to achieve this goal, all are based on encapsulation. This mechanism is well known and quite simple to use and to understand but implies a little overhead in traffic, the other advantage, is that devices that don't need to be aware of the encapsulation (core devices) still work, only the devices that deal with encapsulation and decapsulation (edge devices) need to be specific. The two most likely solutions are based on MAC encapsulation and Multi Protocol Label Switched (MPLS) encapsulation.

2.1.1 MAC Encapsulation

2.1.1.1 VLAN stacking

As we saw before, VID (aka Q-tags since they come from IEEE 802.1Q) are coded over 12 bits and so provide only 4096 different Ids at a time. Q-tag consist in an additional field of 4 bytes in Ethernet header.

- The 2 first bytes represent the Ethertype field, they indicate which protocol is used in the payload of the frame, for 802.1Q it's 0x8100.
- Then the two other bytes are used as:
  - bits are dedicated to implement prioritization,
  - 1 bit called Canonical Format Identifier (CFI) denoting whether MAC addresses in the frame are in canonical format
  - and the 12 bits of the VID.

The idea of VLAN stacking (standard IEEE 802.1ad) is just to add a Q-tag before the VID present in 802.1Q so that a “core device” only consider the first label and other labels are red as part of the payload. In this scheme there are two possible ways of interpreting the stacked Q-tags by the provider switches. In the first one, only the VID of the outer tag (inserted at an edge node) is used by the core Ethernet switches to identify the customer VLAN (C-VLAN), across the metro domain. The other variant is to combine both the VID fields of the stacked Q-tags (customer and provider assigned) and thus be able to support a larger number of C-VLANs (more than 4096). The first scheme is backward compatible with the IEE 802.1Q standard, the latter is not.

This method implies an overhead of only 4 bytes compared to 802.1Q and provide a simple solution to reduce the ST size but it doesn't always help to reduce forwarding table size. For example, in the case where only two ports of an Ethernet switch forward frames of a C-VLAN, the learning of individual MAC addresses of the C-VLAN can be avoided. The switch only needs to learn about the VIDs to forward frames. However, for the cases where C-VLAN frames arrive on more than two ports, all the end host MAC addresses of the C-VLAN have to be learned by the Ethernet switch.

2.1.1.2 MAC address stacking

Currently under consideration in IEEE 802.1ah MAC in MAC is not yet released. As shown on figure 5 it is an extension of the native Ethernet protocol.

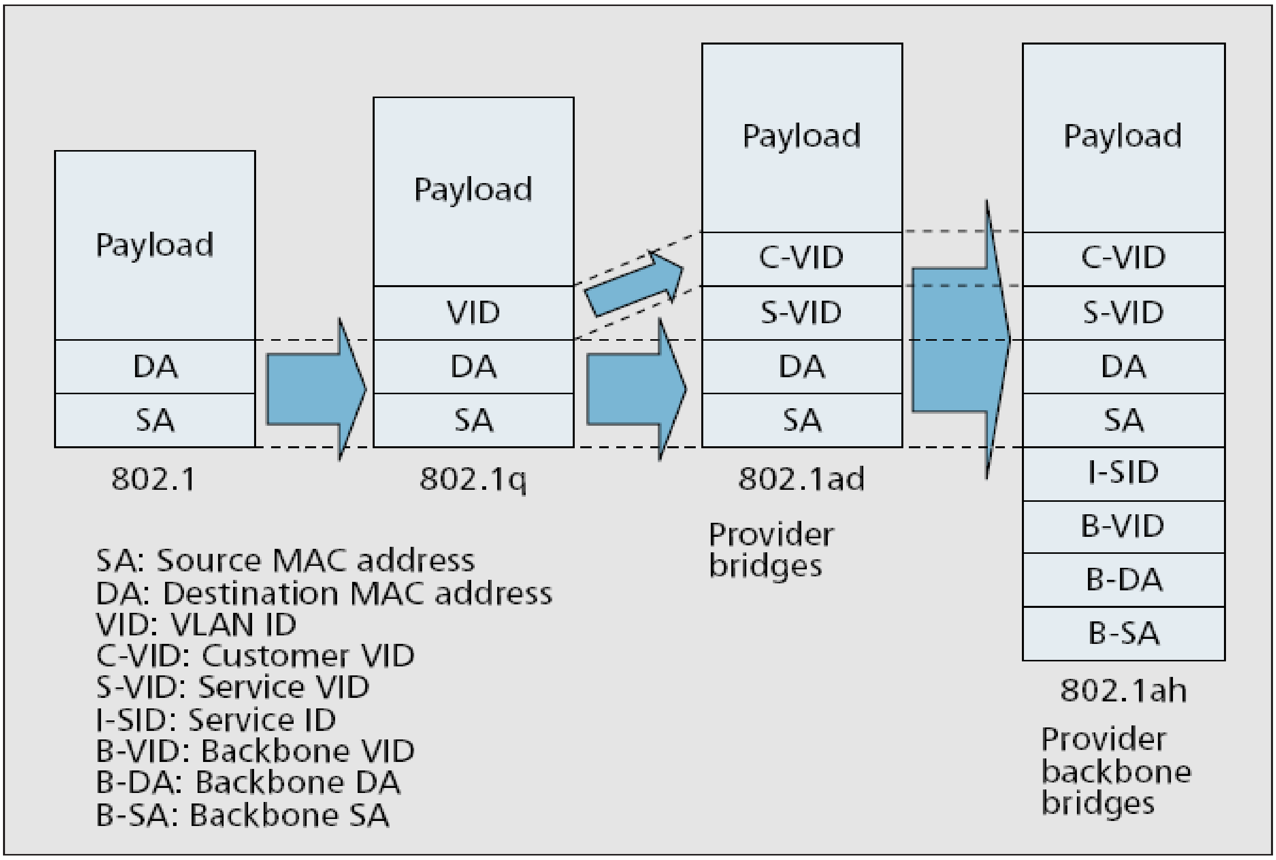


Figure 5: Evolution of Ethernet Hierarchy

It comes after the introduction of VID in IEEE 802.1Q and the VLAN stacking known as IEEE 802.1ad. Those two approaches simply partition the problem into identical units and don't really provide hierarchy. That's why the idea of 802.1ah is to work with a complete recursion such that the customer Ethernet sub-networks are completely encapsulated and isolated from the provider Ethernet network. All the field of 802.1ad (C-VLAN, S-LAN, Source Address and Destination Address) are replicated on the level above to get MAC in MAC.

This way the overhead is bigger (20 bytes) but there is a real isolation between the levels of the hierarchy and the devices in the top level don't need to know anything about the level under, and then the ST are still separated and the forwarding tables are much smaller.

2.1.1.3 Example

The figure 6 shows an example of a WAN in which the provider uses VLAN stacking. In this example the customer as three sites connected to the WAN by access points. The two dashed lines indicate communications between customer's sites A and B and between sites A and C. When a frame is sent by the customer's site A to B for example the frame goes through the customer access point where a Q-tag is added to it. Then it is forwarded along the ST to the access point of the site B where this same Q-tag is removed. This Q-tag is defined by the provider and usually identifies the customer and the kind of service he is using. Note that in this way Q-tags can be reused in separated networks.

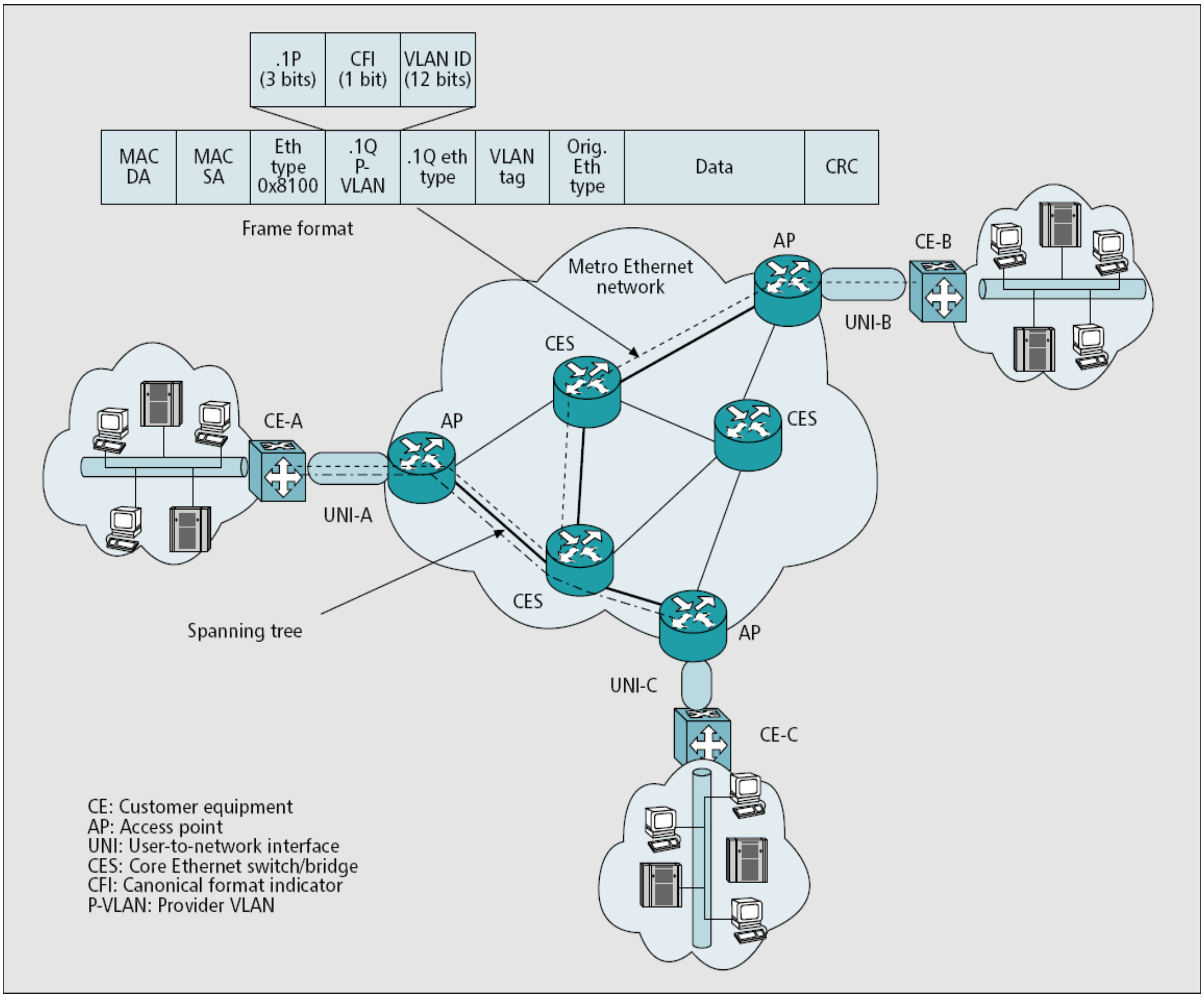


Figure 6: Example of a provider Network using VLAN stacking

2.1.2 MPLS for metro networks

Layer 2 service providers usually prefer circuit-switched protocol like ATM and Frame Relay in their WAN since there are supposed to be easier to manage and now many tools exists for those technologies.

MPLS (aka Martini encapsulation) is a carrying protocol that emulates some properties of a circuit-switched protocol over a packet-switch network. As we can see in figure 7, MPLS consists in adding labels before the header of the packet. These labels will be used by the switches and the routers in the provider network to forward the packet. The number of labels is not fixed and then makes this protocol very flexible.

Each label stack entry contains four fields:

- 20-bit label value.
- 3-bit field for QoS priority.
- 1-bit bottom of stack flag. If this is set, it signifies the current label is the last in the stack.
- 8-bit TTL (time to live) field.

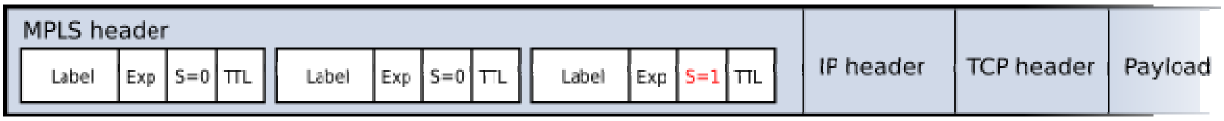


Figure 7: MPLS structure

The metro domain in this case is built a different technology and needs MPLS label switch routers (LSRs). From a customer point of view it's totally transparent he only sees Ethernet frames. Figure 8 shows an example in which two MPLS labels are added onto the customer Ethernet frames based on destination MAC address/port/Q-tag information at the access point. The first label at the top of the stack is the tunnel label, which is used to carry the frame across the provider network. The tunnel label is typically removed by the penultimate hop (i.e., the hop prior to the access point of the destination customer site). The second label at the bottom of the stack is the virtual circuit label, which is used by the egress label edge router (LER) to determine how to process the frame and deliver it to the destination network. Thus, in the case of MPLS encapsulation, at least two labels are necessary (tunnel and VC).

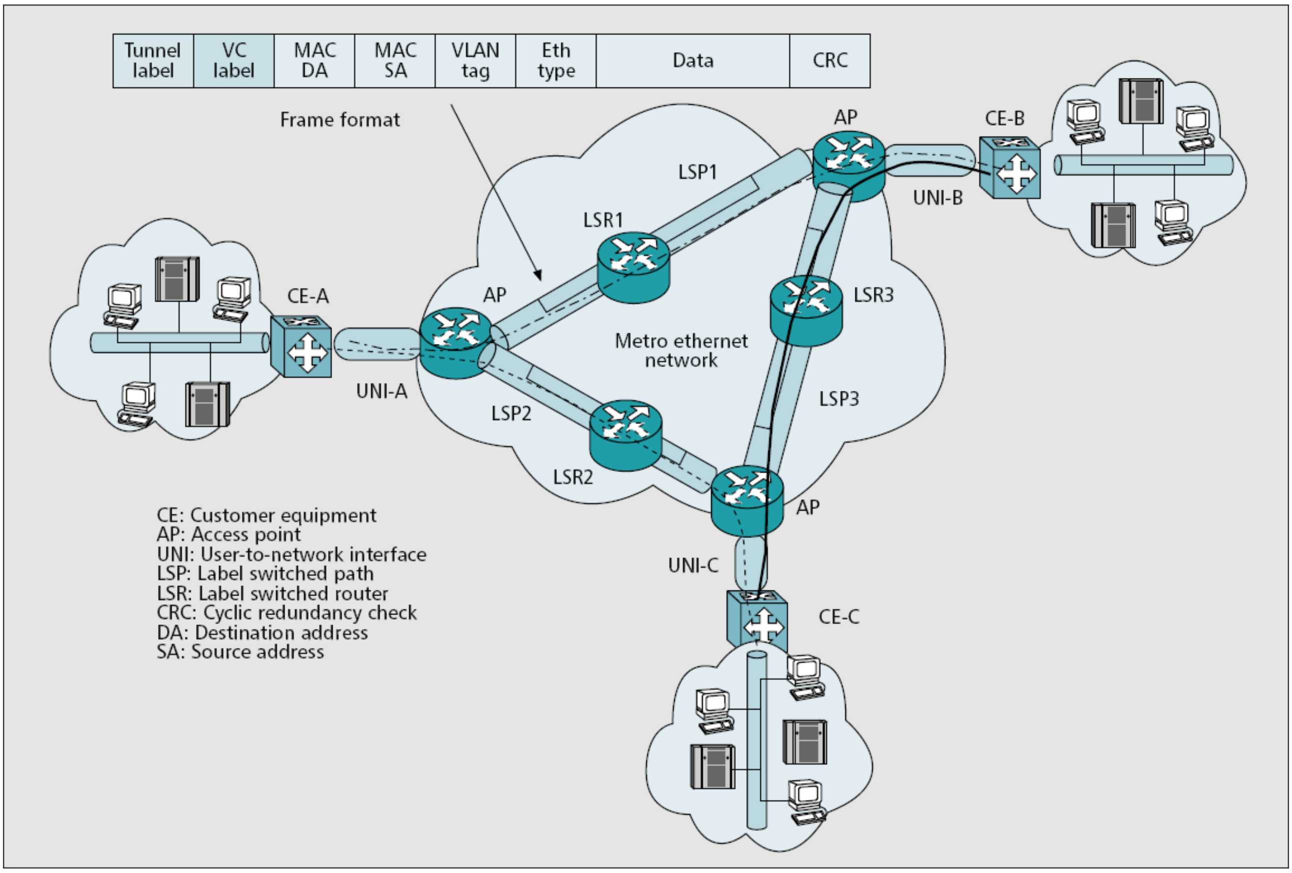


Figure 8: Example of a MPLS metro Network

2.1.3 Comparison

2.1.3.1 Overview of the aspect of the schemes

Table 1 summarizes the main differences between the three encapsulation schemes we have seen so far.

The first point is that MAC stacking and MPLS offer similar services while VLAN stacking seems less adapted and suffers from lacks like no MAC address table containment and provide only little scalability improvements. So even though it offers the smallest overhead, it doesn't appear as the best solution.

The main differences between MAC stacking and MPLS are that MPLS needs special equipments and so is less evolutive. On the other hand it offers a smaller overhead (only 8 bytes) than MAC stacking only (20 bytes).

We also have to keep in mind that MPLS provides a circuit-switched network while MAC stacking is packet-switched. This difference is very important since the actual protocols used in WANs are circuit-switched and there are already many tools to ensure traffic engineering and QoS, things that don't exist yet for packet-switched networks. This issue is discussed later in part 3 Traffic Engineering.

Encapsulation method	MAC address table containment	Scalability	Priority bits	Transition / Evolution	Localization of impact of duplicate MAC	Localization of impact of provider ST control	Signaling	Min. overhead (Byte)
VLAN stacking	No	Combin. of VID: Yes Provider: No	Yes	Yes	No	No	Needed	2
MAC stacking	Yes	Yes 16 million	Yes	Yes	Yes	Yes	Optional	20
MPLS	Yes	Yes 16 million	Yes	Partially (cost)	Yes	Yes	Yes	8

Table 1: Comparison of the schemes

2.1.3.2 New Forwarding modes

A spanning tree will not scale to large Layer-2 provider networks, nor can one use “routing” for Ethernet to provide loop-free Layer-2 networks. The generic idea to solve the problem of creating a loop-free network is to constrain the forwarding topology of the network.

From an abstract view, an interconnect medium is a system of real and/or virtual data paths and protocols that, taken together, emulate a single Ethernet LAN. Using this abstract view, the interconnect medium is not restricted to being the core portion of a VPLS, but could also be an emulated LAN using ATM LANE, a standard 802.3 network or a 802.17 resilient packet ring network.

Each island (group of remote MAC users belonging to the same LAN) is responsible for forwarding data within the island and avoiding local forwarding loops i.e. an island may run an instance of spanning tree. In addition to this, each island must ensure that only one port is used to pass customer frames between the island and the medium shared by all other islands. This does not imply that there is only one port, as there could be multiple ports for redundancy purposes, but that only one should be active at any one time. If more than one port were forwarding at any time, loops could occur, unless a global spanning tree were employed, which is not desirable.

These requirements lead to the definition of 5 general rules for providing a restricted topology that avoids loops and also obviates the need to employ a global spanning tree:

1. Each island is responsible for preventing internal forwarding loops.
2. Islands connect to other islands only through Interconnect Media ("no back doors").
3. Each island ensures that no customer data frame passes through more than one interconnect medium attachment into or out of the island.
4. Each island ensures that it attaches any given Customer Service Instance to no more than one interconnect medium.
5. An interconnect medium ensures that if an attached port can talk to any other attached ports, it can talk to all of the ports attached to that medium.

In summary the rules ensure that there is only a single active forwarding path such that redundant paths can be provided, loops cannot exist and reliability is maintained.

2.2 Operation, Administration and Maintenance (OAM)

Ethernet has been used as a Local Area Network technology for many years, and enterprises have managed these networks effectively, primarily with the use of IP protocols such as SNMP, ICMP echo (or IP Ping), IP Traceroute. However, Ethernet in the service provider space is an entirely different problem and requires new protocols to achieve OAM operations efficiency. This problem can be considered from two points of view, some Ethernet's characteristics are making easier the development but on the other hand we have to take into account the economic and practical issues.

2.2.1 Technical characteristics

Ethernet offers an inherently robust data plane when contrasted with other packet technologies. Use of link local-path identifiers such as MPLS labels, ATM virtual circuits' identifiers/virtual path identifiers, and so forth introduces a "level of indirection" into data-plane forwarding. The current trend for label swapping that has persisted for several generations of WAN technology actually has a detrimental effect on overall network reliability.

Significant progress has been made in the field of data-plane OAM with efforts in both the ITU-T (Y.17ethoam) and the IEEE (802.1ag). There is strong agreement as to the functionality required, and a comprehensive suite of tools is already emerging. This is a side benefit of the degree of rigor associated with the specification of Ethernet in terms of frame formats and transfer functions. When the data-plane forwarding and relay behavior is clearly understood and well specified, procedures for verification become a similarly well-understood problem. The end result will be an increasing degree of front-line reliability and much lower mean time to detect and repair when failures happen.

The addition of connection management to Ethernet via reuse of existing data-plane transfer functions means that currently specified OAM protocols will require no modification in order to be applied. Using configuration of existing forwarding addresses fundamental requirements for data-plane OAM and permits reuse of the standards in progress:

- Fault management OAM PDUs (such as continuity check, loopback, link trace, etc.) must utilize and exercise the same forwarding components in intermediate switches as the bearer path. For unicast forwarding based on known or configured VLAN/MAC, the transfer function is common regardless of how the forwarding table was populated.

- For real-time correlation of counts, performance management OAM PDUs must utilize both the forwarding components and specific queuing discipline at intermediate switches so that packets "in flight" can be correctly accounted for. Again, for unicast forwarding based on known or configured VLAN/MAC, the transfer function is common regardless of how the forwarding table was populated.

Note that those characteristics are intrinsic to Ethernet and in order keep advantage of this MAC stacking should be preferred to MPLS as a hierarchy mechanism. Thus MPLS can take advantage of the many OAM tools already existing for circuit-switched protocols like ATM and FR.

2.2.2 Practical point of view

Unlike ATM or FR, which are manufactured by a rather limited number of vendors, a very large number of manufacturers have commercialized their own Ethernet products, ranging from simple computer network Internet cards to backbone switches. Due to this very large number of vendors, Ethernet management might face some serious interoperability issues. Experience shows that multi-vendors OAM tools are inefficient as vendors often use proprietary OAM protocols. Moreover, bridged networks are increasingly being operated by independent entities, each with restricted management access to each other's equipment. Furthermore, the large number of Ethernet LAN, MAN, and WAN standards may cause some serious management information exchange problems. Hence, some specifications are needed to provide requirements for unified OAM protocols and tools.

Experience shows that the process of implementing an efficient and reliable management system is very slow and may take several years. Often, operators and providers focus on service delivery and are rather reluctant when it comes to management tools. All this means that it may take years before we can see a fully managed Ethernet network. Providers and carriers may then just want to leverage their existing management systems that are rather mature and proven. Thus, they may just want to stick to their existing WAN infrastructures. In addition, the claim that no learning effort is required is questionable, at least from a provider point of view, given the overwhelming number of new WAN and MAN Ethernet standards.

The IEEE is working on a set of fault management and network discovery standards such as 802.1ag, 802.1ab, and 802.1ak. However, it is not clear how and where to perform provisioning and configuration. Ethernet management tools and systems must be introduced with tangible added value over existing management tools and systems. One exciting feature would be the capability to support plug-and-play OAM functions to keep the Ethernet very simple and reduce operation expenditures. The customer must not feel a significant difference between the operation of the Ethernet technology deployed in his/her LAN and the technology he/she uses for WAN services. This requires a significant software development effort, essentially by vendors.

2.3 Security

2.3.1 Native Ethernet issues

Security is a prime consideration within any public switched network: one user should not affect any other user. Due to the “plug & play” nature of Ethernet, networks have to be designed with caution to provide the necessary degree of security among users. Irrespective of Architecture, Ethernet services use Ethernet technologies such as 802.1q, 802.1w/s, 802.1d, etc…

IETF, MEF and ITU created in a joint-effort the 802.1ah protocol, which has among others the ability to provide a full isolation among customers by frames encapsulation and also isolation of customers from provider.

But problems occur with encryption and the interoperability of protocols at same layer. For example, assuming that 802.1ah is used in the backbone, 802.1ad in the provider bridged network, and customers use both 802.1q as well as 802.1d. If the Ethernet frames need to be encrypted end-to-end, five levels of encryption are needed: the data field (IP packet), 802.3 header, .1q header, .1ad (q-in-q) header and .1ah (Mac-in-Mac) header, all at the same protocol layer!

Ethernet is natively a layer-2 protocol, based on MAC addresses resolution, which implies a lot of security issues. Then, classical attacks against ARP constitute a threat for Metro-Ethernet networks.

Some threats and possible counter-actions are discussed below:

- MAC attacks make the switch forwarding tables to overflow and flood users’ traffic. An answer is to administratively limit the amount of MAC addresses a device can learn per customer service instance / VLAN.
- ARP attacks such as ARP spoofing, misuse of Gratuitous ARP: Deploy dynamic ARP inspection, filter using wire-speed access lists, and isolation of customer ports from each other (employ so called “Private VLAN”) mitigate against these type of attacks.
- VLAN hopping, trunking protocol attacks: Careful configuration such as disabling of auto-trunking, use of dedicated VLAN-ID for trunk ports, disabling of unused ports etc. can effectively stop this type of attack.
- Spanning tree attacks such as hijacking the service provider’s spanning tree: Enable switch features that restrict which bridge may become root.
- DHCP Rogue Server Attack: Use DHCP snooping and differentiate trusted and untrusted ports.

Denials of Service (DoS) attacks based on DHCP protocol are possible, either due to a malicious activity if the user pretends to be the Network DHCP Server, or due to a misconfiguration, when the user configures router (DHCP server) incorrectly.

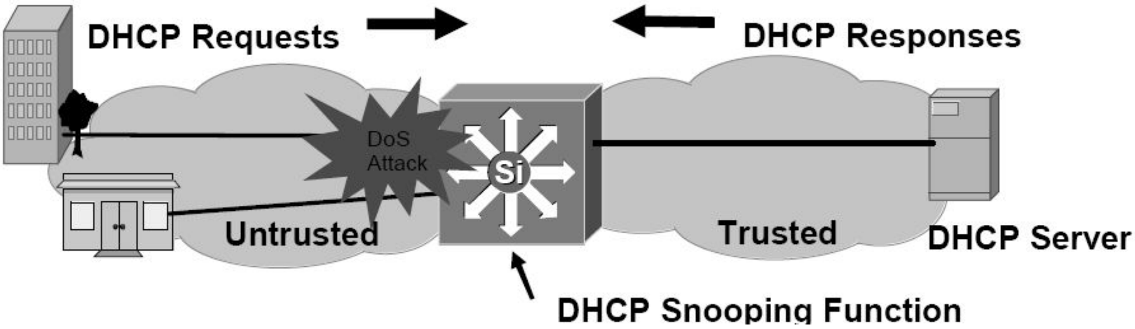


Figure 9: DoS attack on a switch

The access switch only forwards DHCP requests from access ports (untrusted ports). All other types of DHCP traffic from access ports are dropped. If the server is not local to the Switch, trust the uplink port. Then an attacker can perform a (D)DoS, sending millions of requests on the switch using untrusted ports as shown on figure 9.

DoS attack can be prevented by rate limiting the DHCP packets on access ports

DHCP snooping is not equivalent to Option 82 (DHCP Interface tracker), which performs subscribers management, assigning IP addresses based on switch ports and subscribers info. This is possible using port (Module/Port/VLAN) and local switch info added to DHCP Requests.

The above list is by no means exhaustive and could easily be extended. Additionally, to further secure the network, pro-active defense mechanisms such as IEEE 802.1x, MAC-level port security, wire-speed filtering access lists, server based VLAN policy management, etc… should be carefully considered.

2.3.2 The role of hierarchy

Hierarchy has a number of security advantages. Isolation of untrusted traffic via encapsulation and containment means that significantly less complexity is required at the network edge to police untrusted behavior, as by definition behavior limitations are rigorously enforced by the connectivity model.

It can also be said that auto-discovery (one of Ethernet’s strong suits) and security are polar opposites, yet from the point of view of operational complexity, much of Ethernet’s auto-discovery capability is still desirable. Hierarchy allows trusted domains to be established using auto-discovery. The transport layer interconnecting the network’s edge devices can operate in a *plug-and-play* mode, as the network edge is secured physically. The network edge encapsulates and constrains traffic received from outside the secure boundary.

3 Traffic Engineering & QoS

Traffic Engineering is a key which allows providing a reliable access to customers, dealing with network paths engineering, bandwidth management and load balancing.

3.1 Paths Engineering

Configured Ethernet-switched paths have a complete freedom to route datagrams from a source to a destination. The ability to define multiple paths to any end-system has a number of implications.

There may be numerous viable paths between any two points in the network, and spanning tree and bridging/auto-learning are only able to use one of them. The configuration allows more than one bearer path between any two points to be defined, and criteria beyond simply the shortest path to be used in selecting the routing of any individual path. Multiple metric techniques and graph-splitting algorithms exist to permit optimal sets of paths with specific end-to-end attributes and without common points of failure to be computed and instantiated. Then, paths can be engineered in MAN, and equilibrium between physical network build and offered load (a frequent occurrence) can be found.

For a network using both learning/bridging and configured behavior, the traffic associated with each forwarding behavior needs to be differentiated such that best effort traffic does not degrade “engineered” traffic, especially as the traffic matrix may dynamically change in response to failures and/or maintenance activities. This could be achieved via specifically instantiating “connection” state in transit switches, or can be achieved via intelligent use of class-based queuing and edge policing of class markings.

3.2 Classes of Services

Traffic emanating from customers needs shaping policies prioritizing communication flows for efficiency and reliability purposes.

Different classes of services (CoS) can be offered to customers, according to their SLA. Each ISP is able to create its own set of CoS, generally having between 5 and 7 different classes. In decreasing order of priority, we note especially the network management protocols (highest priority), then real-time applications such as VoIP, then urgent business data and finally data that does not require special QoS features.

3.3 Resilience

Spanning Tree Protocol was designed to optimize the tree’s reconstruction time, in case of link failure.

RSTP (Rapid STP) is an enhancement, which considerably speeds up the tree’s reconstruction. Topology changes resulting of a link failure are more precisely detected and rapidly propagated from the point of failure detection, rather than from the root node. RSTP uses additional port state definitions and faster transitions to the forwarding state.

3.3.1 Multicast

Multicast protocols impose lots of deployment issues and have to be handled in a smart way. Scalability, reliability, QoS are usual issues plus all problems arising from multicast trees apply: routing, scheduling at routers, and reliability.

MC protocols were designed toward multimedia content streaming to large crowds of users, using UDP at transport layer, which tolerates loss or corruption of the data. With new services such as VoD, reliability of multicast protocols must be ensured, imposing the use of acknowledgement mechanisms to correctly carry data packets to end-users. Then, if many losses occur at the leaves of a MC tree (i.e. for end-users), lots of ACKs will be sent back to the tree's root node, potentially resulting in a Denial of Service of the root node, which cannot handle all incoming ACKs. This phenomenon is called ACKs feedback implosion.

To avoid the problem of feedback implosion, members of the same MC tree can be hierarchically organized into sub-trees, with a local root node in charge of retransmitting packets in its own local tree when losses occur.

3.3.2 Broadcasting

Routing in intermediate routers consists essentially in finding an appropriate path to the next router and scheduling the data packets. Care has to be taken in order to meet the QoS demand of end-users, i.e. according to the service they require.

Scheduling corresponds to prioritizing streams to be forwarded by routers, and have to take into account the QoS needed by the application. For example, VoIP communications can tolerate neither loss nor delay. Flows priorities are commonly defined with 7 or 5 different service class, the most important being control commands, then VoIP and finally business data services.

4 In practice

4.1 Actors and realization

All major actors of the telecom industry have already launched the production of devices specially dedicated to Ethernet MAN/WAN communications: Cisco, Alcatel, Nortel Networks…

Cisco announced in February 2004 the orientation of Cisco products toward Metro Ethernet technologies, with the production of new devices such as the Catalyst 3750 Metro Switches Series (9000 euros), which allow a company composed of remote sites to segment a connection via VPNs with a bandwidth guarantee stipulated by the SLA.

4.2 SuperDemo 2003

To enforce this cooperation and the standardization of products, MEF organizes demonstrations where many companies come and build a network with all their devices. It is a good way to keep track of what is being done and show to other people how it goes and how it works. In this part we will consider some aspect of the SuperDemo 2003 that took place in Atlanta.

4.2.1 Actors

The MEF's SuperDemo 2003 network is built around a fully meshed 802.1Q Ethernet switched 10GigE core. The twenty-two vendors supplying aggregation devices are Alcatel, Appian Communications, Atrica, Cisco Systems, Coriolis Networks, Corrigent Systems, Ensemble, Extreme Networks, Fujitsu, Harmonics, Hatteras Networks, Internet Photonics, JDS Uniphase, Lantern Communications, Luminous Networks, Mahi Networks, Native, NEC, Nortel Networks, Riverstone Networks, Tpack, Vivace.

4.2.2 The Network and tests

Figure 10 shows the scheme of the network that was created for the demonstration.

First, there is a core network using devices from different actors (Cisco, Riverstone Networks and Extrem Networks). This core has 3 edges tagged with E-LAN group numbers: 92 in red, 94 in blue and 96 in green.

Then company sites are emulated. These sites are links with P2P connections. Each of these sites is represented by a box. Besides each box you can see the two Q-tags that are the C-VLAN and the E-LAN group. In each box there is a number representing the vendor of the device and his name. In brackets there is the number of the vendor of the device used on the other side of the connection. This way a single connection uses devices from many different vendors.

In order to demonstrate the interoperability of Ethernet services participants establish:

1. Three mandatory point-to-point E-Line services support webcam-enabled Netmeeting sessions as well as two test sets. Each service is configured between two partnering vendors. They are eleven such pairs on the network.
2. One mandatory E-LAN service supports an MPEG video stream broadcast to each one of three groups seven or eight vendors.
3. Vendors also have the option of establishing global E-LAN services to support four applications:

global video streaming, Yahoo Instant Messaging, theatre webcam and IP telephony.
4. In addition the Circuit Emulation Services vendors establish end-to-end connections through the core to demonstrate their ability to transport various types of TDM traffic over Ethernet.
5. Additional services and applications, described in the participant contributions section below, are established on a voluntary basis and can be viewed at SUPERCOMM 2003.
6. Service Level Agreements (SLAs) specifying Committed Information Rate (CIR) and Peak Information Rate (PIR) bandwidth profiles for each application, are applied to the E-Line and ELAN services.

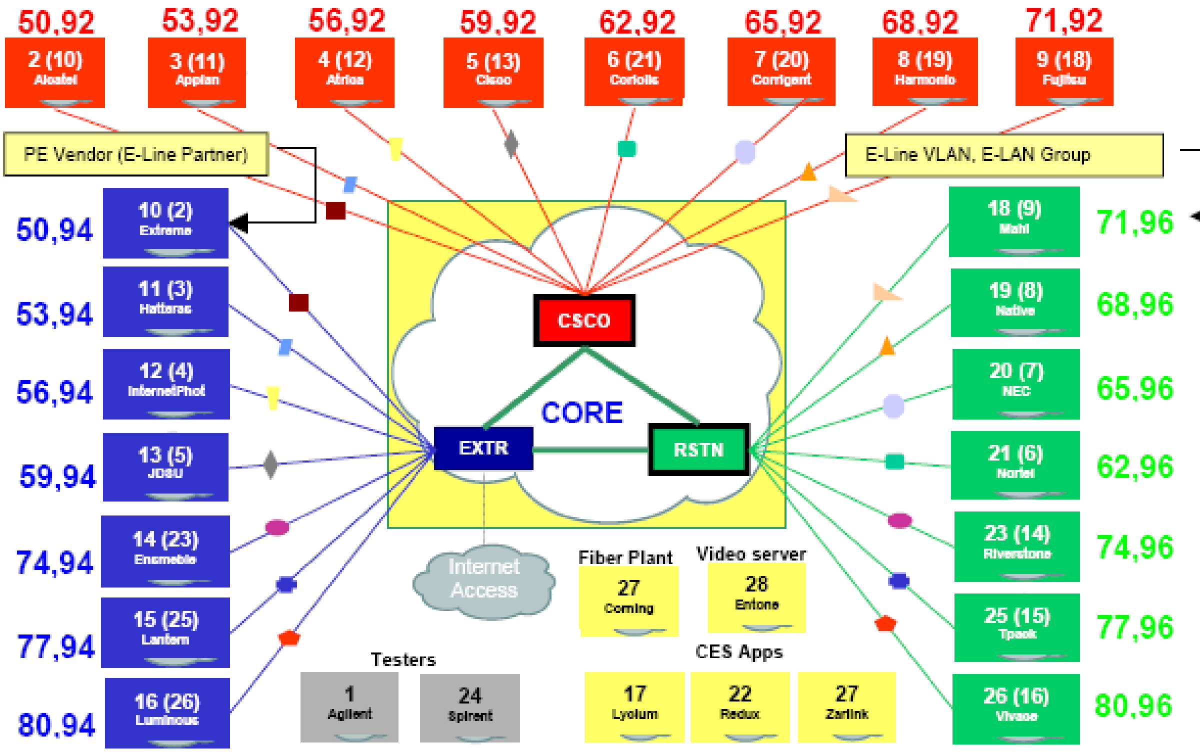


Figure 10: Network scheme of SuperDemo2003

The unprecedented scale and technical success of SuperDemo 2003 provides tangible proof to service providers that:

1. Metro Ethernet services can successfully extend packet delivery from the LAN to the provider networks' access, aggregation and core layers;
2. Metro E-Line and E-LAN services defined by the MEF are carrier-grade service options that provide seamless end-to-end connectivity to hosts running both point-to-point and multipoint-tomultipoint applications;
3. Metro Ethernet products from multiple vendors can interoperate at all layers of the provider network.

4.3 Competitor Technologies

4.3.1 Wireless

Wireless is partly a competitor to Ethernet, especially for LANs. It can also be thought as a complementary wireless technology to Ethernet.

But, due to the poor performance of current wireless technologies compared to Ethernet, this latter remains uncontested for LAN deployments.

The sole "real" advantage of wireless technologies over Ethernet today is the mobility it provides.

4.3.2 Optical Fiber (FTTH)

The Fiber-To-The-Home (FTTH) is a serious competitor to Ethernet for MAN/WAN communications. Its major competitive advantage over Ethernet is the bandwidth that can be offered to customers, providing up to 2.5 Gbps bandwidth to end-users, in latest tests realized in Paris in April 2006.
However, the deployment costs (cost of the fiber itself plus the installation/deployment) are still very expensive. But, with generalization, every technology tends to become cheaper.

As an example, the French ISP Free has announced it will deploy its own optical fiber network, providing subscription of 29.99 euros for a bandwidth of 50Mbps symmetrical.

Nombre d'abonnés FTTx à fin 2005 (en milliers)	
Pays	Nombre d'abonnés
Japon	4.640
Corée du Sud	1.620
Etats-Unis	858
Suède	321
Italie	257
Source : IDATE, "Déploiement FTTH - Quand et pourquoi ?", juillet 2006	

Table 2: Number of subscibers to FTTx technology

Internet access through FTTx represents less than 9% of the total of high-speed lines in the world, on the 31st march 2006, according to Point Topic.

4.4 Metropolitan Area Networks

If Ethernet is able to scale to large MAN/WAN networks, why don't we receive offers from ISPs yet?

According to Abdul Kasim, Vice President for Ethernet Business Development at ADVA Optical Networking, in an article for Converge Digest the 26th January 2006, the slow-start deployment of Metro Ethernet today is due to economical reasons.
"More than 120 carriers in the United States alone offer some Ethernet services to enterprise customers. Few, though, have deployed Ethernet services on a mass scale, even though the demand within the marketplace is clearly evident.
The dilemma is that, despite the compelling value proposition of Ethernet access to end customers,

carriers have generally been unable to offer these services at sufficiently attractive profit margins. Under these conditions, carriers understandably have been confined to nominal offerings, limiting Ethernet availability. And without economic incentives, a mass-market deployment of managed Ethernet services to the full gamut of enterprise customers will remain unlikely.”

However, the author is sure that Ethernet in the future will become the transport foundation for business services moving forward, because it proved to be scalable, routing-efficient, while remaining a low-cost technology.

References

[1] R. Santitoro, "*Metro Ethernet Services – A Technical Overview*", The Metro Ethernet Forum 2003

[2] M. P. McGarry, M. Reisslein and M. Maier, "*WDM Ethernet Passive Optical Networks*", in IEEE Communications Magazine 2006

[3] D. Armannsson, G. Hjalmtysson, P. D. Smith and L. Mathy, "*Controlling the Effects of Anomalous ARP Behaviour on Ethernet Networks*", in Proceedings of the 2005 ACM conference on Emerging network, 2005

[4] IEEE 802.1D/Q/p Working Group: "*Media Access Control (MAC) Bridges*" IEEE 802.1D, 1998

[5] K. Moerman, J. Fishburn, M. Lasserre and D. Ginsburg, "*Utah's UTOPIA: An Ethernet-Based MPLS/VPLS Triple Play Deployment*" in IEEE Communications Magazine November 2005

[6] G. Chiruvolu, A. Ge, D. Elie-Dit-Cosaque, M. Ali and J. Rouyer, "Issues and Approaches on Extending Ethernet Beyond LANs" in IEEE Communications Magazine March 2004

[7] M. Ali, G. Chiruvolu, and A. Ge, "*Traffic Engineering in Metro Ethernet*" in IEEE Networks March/April 2005

[8] Aref Meddeb, "*Why Ethernet WAN Transport?*" in IEEE Communications Magazine November 2005

[9] F. Brockners, N. Finn and S. Phillips, "*Metro Ethernet – Deploying the extended Campus using Ethernet Technology*" in IEEE LCN'03, 2003

[10] R. van Haalen, R. Malhotra, and A. de Heer, "*Optimized Routing for Providing Ethernet LAN Services*" in IEEE Communications Magazine November 2005

[11] D. Allan, N. Bragg, A. McGuire and A. Reid, "*Ethernet as Carrier Transport Infrastructure*" in IEEE Communications Magazine February 2006

[12] Metro Ethernet Forum, "*Super Demo 2003 white paper*", Metro Ethernet Forum, 2003