**European Union's
@LIS II Programme for Latin America**

Alice2 Project
RedCLARA2 Network
redesign

Periodical Progress Report

ALICE2 Deliverable: DO1.1-03(Código D Objetivo- Numero correlative) Título del Documento

Document Full Name	**DO1.1-03 RedCLARA2 Network redesign**
Date	**13/11/2009**
Activity	**To plan and design RedCLARA2 Network**
Lead Partner	**RNP**
Document status	**DRAFT**
Classification Attribute	**Internal to Project Partners**
Document link	

Abstract: This document presents the Technical Requirements for the Tender of Data Links to be procured for the implementation of the RedCLARA2 Network

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

COPYRIGHT NOTICE

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

DELIVERABLE ROUTE

	Name	Member/Activity	Date	Responsible
From				
Revised by				
Aproved by				

DOCUMENT DESCRIPTION

Subject	Date	Comment	Author

REGISTER OF CHANGES

Subject	Item	Reason for Change

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

TABLE OF CONTENTS

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

1.- INTRODUCTION

As the national research networks around the world are investigating - for some years now - technical and economical aspects and also developing tecnologies to deploy their next generation infrastructures to support advanced applications and offer better (i.e. faster, more transparent and more resilient) services to their users, in the next phase of the Alice2 project, CLARA is assessing the technical upgrades necessary for the RedCLARA2 backbone to support better collaboration services to its stakeholders, the client NRENs, providing Layer 3 (connectionless) services and also Layer 2 (connection oriented) circuit services. This plan is aligned to ESnet, GLIF, GEANT 3 and Internet2 initiatives.

Continental sized research backbones in the northern hemisphere have been evolving to architectures that support both the traditional IP based Internet over optical transport technologies for worldwide data exchange and, in parallel, some testbeds for infrastructure of static and dynamic optical/packet-switched circuit networks (DCNs). In the southern hemisphere there are few networks, if any, whose infrastructure have similar characteristics, since most of them do not own dark fibers or have access to multiple lambdas, despite the fact many already own some kind optical equipments to build their infrastructure over very expensive commercial providers fibers, or else lease also very expensive lambdas from those providers. To make things worse, there are not that many providers with enough optical capacity to cope with the research networks demands, at least with prices accessible to most NRENs budgets.

Searching for more ways of collaboration between the NRENs researchers, there are international cooperative initiatives as the Global Lambda Integrated Facility (GLIF) that aims to connect the research networks optical infrastructures to overcome the lack of long distance bandwidth capacity for data-intensive scientific research. For CLARA's members be able to participate in those collaborative initiatives, the RedCLARA2 must also evolve in terms of infrastructure, to have better technical support for a broader range of advanced Internet services.

This document presents technical aspects for network design and infrastructure for advanced services that may be of interest of the CLARA associated NRENs, and trying to show their respective advantages, drawbacks and impacts on the services and operational aspects of RedCLARA network and also on its associated NRENs networks.

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

2.- BACKGROUND

RedCLARA2 plans to have an infrastructure capable of offering Layer 3 IP based (connectionless) services and Layer 2 circuit services (connection oriented).

Over the years, many advances in networks technologies — including improved bandwidth, quality of service (QoS), multicast, and availability improvements — have taken place in the LANs, where the Ethernet has emerged as the dominant technology — due to its simplicity, cost advantages, ubiquity and its incremental speed advances. Service providers are looking to provide higher bandwidth, as well as enhanced services like QoS, and are now looking to Ethernet to scale the bandwidth offered to enterprise customers for WAN and MAN applications.

Until now, metro service providers have relied mostly on their SONET/SDH infrastructures to provide data services. Although SONET/SDH is clearly well understood and works as specified, it is not optimized for data traffic. As the bandwidth demands of the WAN and LAN have increased, it has become necessary to match LAN and WAN capacity and transmission speeds in the MAN.

Because of the availability, cost, and speed advances in Ethernet, many service providers are looking to deploy and offer their customers Ethernet as a connectivity option. Many commercial service providers are already offering Ethernet connectivity and others are considering or offering Ethernet as a Layer 1 private line service, as a pure Layer 2 transport mechanism, or to provide Internet Protocol (IP) and Multiprotocol Label Switching (MPLS) VPN services to complement their existing SONET/SDH, Frame Relay, or ATM services.

Ethernet is an alternative for increasing the capacity of the legacy networks technologies services. Service providers are looking to limit their spending on aging network technologies while offering like services with superior functionality, scalability, and lifetime cost ownership.

Trends of NRENs

The main service requested and delivered by NRENs in general is basic IP connectivity to worldwide Internet and secondly to access to other NRENs, at speeds and cost that service providers historically have been unable to provide [4],

| A project of the @LISII Programme | This project is funded by the European Union | A project implemented by CLARA |

neither in most well served locations, nor in the least served locations (e.g. away from metropolitan areas), which is a critical issue in third world countries, widening in large scale the digital divide gap. In recent years the maket dynamics changed and service providers are willing to provide fast basic IP connectivity for acessible cost, but in most cases, this does not apply to speeds of 10Gbps and above.

As the operation of IP networks is a well understood subject nowadays [4], NRENs (and Service Providers) shift their focus to other activities (or business opportunities). For NRENs - particularly in Europe and North America - the focus changed to newer trends: support for network research and support for big science, which can require high bandwidth and non-standard applications and protocols. The network research of the NRENs have been conducted in regional, national and international testbeds built on top of dedicated photonic layers, separated from their traditional basic IP connectivity service. The network virtualization research initiated by the GENI project proposes the implementation of network slices on top of a shared (IP or similar) infrastructure that supports production traffic as well as separated virtual networks running their own protocols on the same physical infrastructure as network research overlays like PlanetLab.

3.- CLIENT CONNECTIONS

In the new infrastrucure, the plan is to have the UNI (User to Network Interface) for the circuit service based on Ethernet technology, using VLAN ID as identifier and the transport from client NRENs premises to CLARA PoP may be provided by third party telecommunications providers, by way of IRU of optical fibers or lambdas to carry pure L2, MPLS or IP based alternatives, depending on the availability on each location.

4.- NETWORK SERVICES: CIRCUITS

Organizations and initiatives like ESnet, GEANT, Giga, GLIF, Federica, Internet2, KyaTera[1], Phosporous have identified communities of small users – astronomers, astrophysics, high energy physics, and biogenetic research, to name some - that have applications that have specific requirements from the network, which would

---

[1] http://www.kyatera.org.br

| A project of the @LISII Programme | This project is funded by the European Union | A project implemented by CLARA |

be supplied by connection-oriented circuits services. Some characteristics of these circuits services needs are: short duration (hours, days), dynamically established, latency, jitter and ability to run IPv4 / IPv6 based or another network protocol.

The circuits services to be provided by Ethernet access shall be mostly based on VLANs and can be established with granularity – will require traffic grooming - ranging from 155Mbps to 1Gbps.

Layer 2 Network Services

Internet users in general are used to have traditional IP Layer 3 services. NRENs around the globe and international initiatives are working for some years in enabling Layer 2 network services for the scientific communitiy that work with data-intensive applications and needs to perform big data exchanges amongst them.

For Layer 2 services, Ethernet offers a viable alternative for increasing the capacity of the now aging ATM and Frame Relay services still in use by Many commercial service providers are looking to limit their spending on their aging ATM and Frame Relay services while offering like services with superior functionality, scalability, and lifetime cost ownership.

Ethernet Services

Basically, the Metro Ethernet Forum (MEF) defined two types of Ethernet service types (EVC - Ethernet Virtual Connection). The MEF Ethernet services, expressed through the EVC construct, are defined in a manner that is agnostic to the specific technologies implementing them. However, the most common technologies used for supporting MEF services may not always provide full transparency with their transport services.

An EVC performs two functions:

- Connects two or more subscriber sites (UNIs) enabling the transfer of Ethernet service frames between them.

- Prevents data transfer between subscriber sites that are not part of the same EVC. This capability enables an EVC to provide data privacy and security similar to a Frame Relay or ATM Permanent Virtual Circuit (PVC).

The two types of EVC defined byt hte MEF are Point-to-Point and Point-to-Multipoint. An EVC can be used to construct a

The names of the Layer 2 services are different for the MEF and each equipment provider, but their functionality is the same. They are as follows:

- MEF E-Line Service / Cisco Ethernet Wire Service (EWS)
- MEF E-Tree Service / Cisco Ethernet Relay Service (ERS)
- MEF E-LAN Service / Cisco Ethernet Multipoint Service (EMS)

When discussing an Ethernet WAN (EWAN), the following terminology should be used (Figure 1):

- CE (customer edge): The customer device connecting to the service provider
- PE (provider edge): The service provider device connecting to the customer
- UNI: The connection between the CE and PE
- Multiplexed UNI: A UNI supporting multiple VLAN flows
- Pseudowire: A term used to indicate an end-to-end path in a service provider network

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

• Pseudowire: A term to indicate an end-to-end path in a service provider network
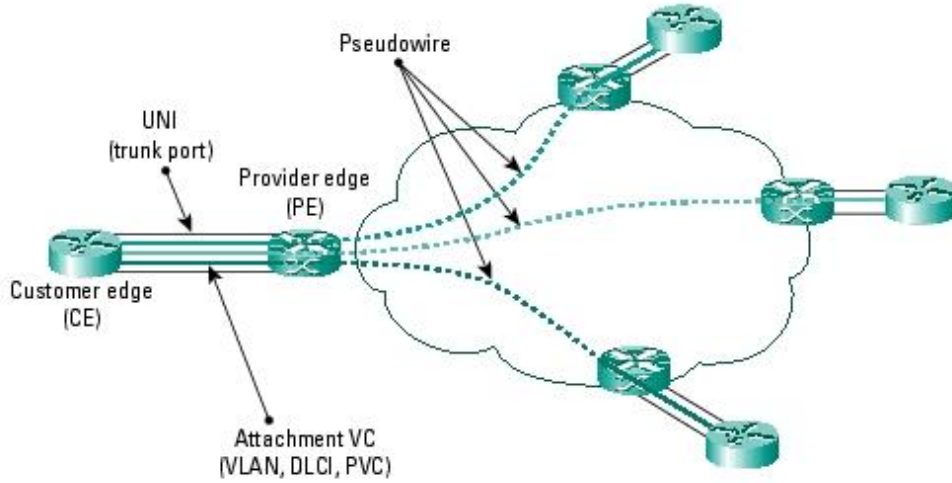
Figure 1 - EWAN Terms (Source: Cisco Systems, Inc.)

Ethernet Wire Service

An Ethernet Wire Service emulates a point-to-point Ethernet segment (Figure 2). This is similar to Ethernet private line (EPL), a Layer 1 point-to-point service, except the provider edge operates at Layer 2 and typically runs over a Layer 2+ network.

Figure 2 - EWS Example

The EWS encapsulates all frames that are received on a particular UNI and transports these frames to a single-egress UNI without reference to the contents

contained within the frame. The operation of this service means that an EWS can be used with VLAN-tagged frames. The VLAN tags are transp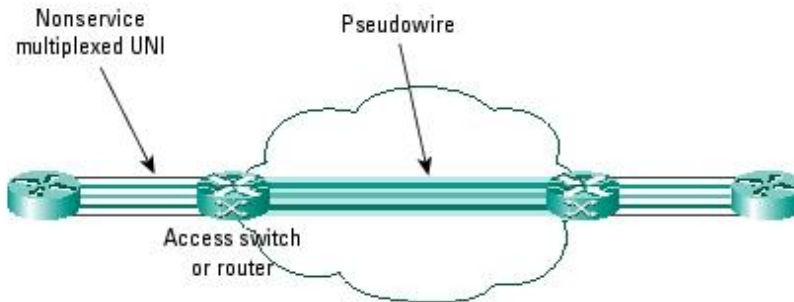arent to the EWS (bridge protocol data units [BPDUs])—with some exceptions. These exceptions include IEEE 802.1x, IEEE 802.2ad, and IEEE 802.3x, because these frames have local significance and it benefits both the customer and SP to terminate them locally. EWS is indifferent to VLAN tags that may be present within the customer Ethernet frames.

EWS subscribes to the concept of "all-to-one" bundling. That is, an EWS maps a port on one end to a point-to-point circuit and to a port on another end. EWS is a port-to-port service (Figure 3). Therefore, if a customer needs to connect a switch or router to n switches or routers it will need n ports and n pseudowires or logical circuits.
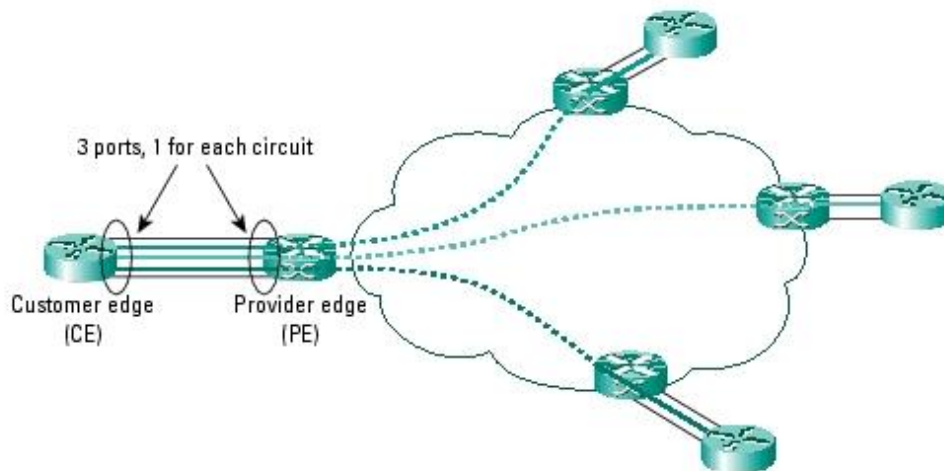
Figure 3 - Nonservice Multiplexing Example: Each Destination (Left) Needs Its Own Port (Right)
(Source: Cisco Systems, Inc.)

One important point to consider is that, although the EWS broadly emulates an Ethernet Layer 1 connection, the service is provided across a shared infrastructure, and therefore it is unlikely that the full interface bandwidth will be, or needs to be, available at all times. EWS will typically be a sub-line rate service, where many users share a circuit somewhere in their transmission path. As a result, the cost will most likely be less than that of EPL. Unlike a Layer 1 EPL, the SP will need to implement QoS and traffic engineering to meet the specific objectives of a particular contract. However, if the customer's application requires a true wire rate transparent service, then an EPL service—delivered using optical transmission

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

devices such as DWDM (dense wavelength division multiplexing), CDWM (coarse wavelength division multiplexing), or SONET/SDH—should be considered.

Ethernet Relay Service

Ethernet Relay Service is similar to EWS in that it offers point-to-point connectivity. The key differentiation between EWS and ERS is that an ERS uses a VLAN tag to multiplex several, non-same-destination pseudowires to one port. That is, unlike EPL and EWS, ERS is a "one-to-many," multiplexed service. Service multiplexing simply means that multiple pseudowires utilize a single access interface or UNI. These circuits can terminate within an L2VPN or on, for example, an Internet gateway. From the service user's perspective, this service multiplexing capability offers more efficient interface utilization, simplification of cable plant, and reduced maintenance costs associated with additional interfaces.

Using the same example as above, where a router connects to n other routers, the source router only needs one port for the service instead of n, as is the case with an EWS. The service need not be port-to-port, but can be logical-pseudowire-to-logical-pseudowire. In the case of an ERS, each circuit can terminate at a different remote location (Figure 4), whereas using EWS, all frames are mapped to a single circuit and therefore a single egress point.
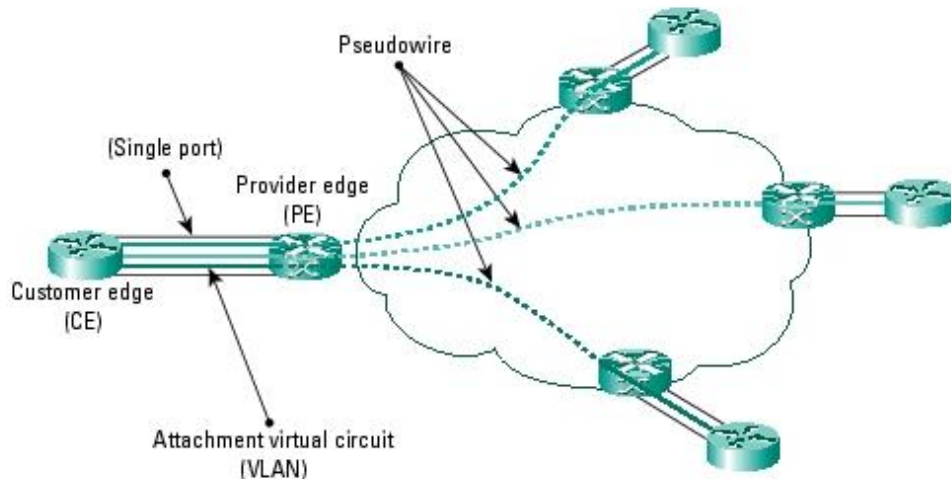
Figure 4 - ERS Service Multiplexing Example: One Port (Left) Can Be Used for All Destinations (Right) (Source: Cisco Systems, Inc.)

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

Like Frame Relay, ERS allows a customer device to access multiple connections through a single physical port attached to the service provider network. The service offered by ERS can be thought of as being similar in concept to Frame Relay, in that a VLAN number is used as a virtual circuit identifier in a similar fashion to Frame Relay data link connection identifier (DLCI) or an ATM permanent virtual circuit (PVC). Unlike EWS, ERS does not forward BPDUs, because IEEE 802.1Q (VLAN tagging) only sends BPDUs on a default VLAN. In a hub-and-spoke network, only one spoke at most would receive BPDUs, thus breaking the spanning tree in the rest of the network. Therefore, an ERS does not transmit any BPDUs and runs routing protocols instead of Ethernet Spanning Tree. The routing protocols give the customer and provider greater flexibility, traffic determination characteristics, and value-added services.

Ethernet Multipoint Service

An Ethernet Multipoint Service (EMS) differs from EWS and ERS in that an EMS provides a multipoint connectivity model. It should be noted that an EMS service definition is still under review within the IETF Virtual Private LAN Service (VPLS) working group. Although EMS uses a multipoint model, it can forward unicast packets to single destinations; that is, it also supports point-to-point connections. To the end user, the network looks like a giant Ethernet switch where each customer has their own VLAN or broadcast domain, rather than end-to-end pseudowire link(s) (Figure 5).
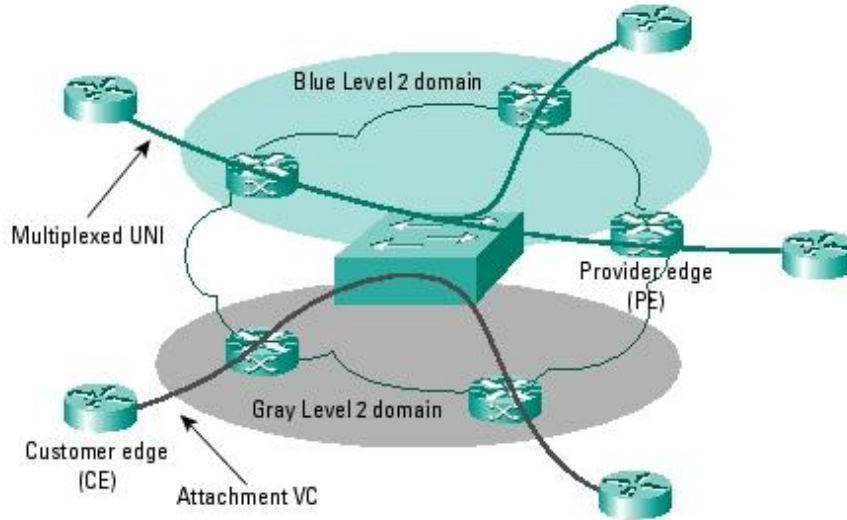
Figure 5 - EMS Example (Source: Cisco Systems, Inc.)

An EMS does not map an interface or VLAN to a specific point-to-point pseudowire. Instead, it models the operation of a virtual Ethernet switch: EMS uses the customer's MAC address to forward frames to the correct egress UNI within the service provider's network. An EMS emulates the service attributes of an Ethernet switch and learns source MAC to interface associations, floods unknown broadcast and multicast frames, and (optionally) monitors the service user's spanning tree protocol. One important point to note is that although the service provider may utilize spanning tree within the transport network, there is no interaction with the service user's spanning tree.

This service works similar to an MPLS VPN, except it functions at Layer 2 instead of Layer 3. While a VPLS EMS is a viable solution, its scalability and QoS control are suspect compared to that of MPLS VPNs. In addition, it is much more difficult, and may be impossible, for the service provider to offer value-added Layer 3 services (this is discussed later in the document).

Finally, emulating LANs in the metro requires a lot of overhead. EMS and protocols run the risk of turning into ATM LAN Emulations (LANE), which have shown their overcomplexity and inability to scale.

5.- PACKETS AND CIRCUITS

Where optical transport (lambda) is available, it will be needed a solution to share them to allow either packets and circuits services. As the availability of more than one lambda is unlikely anytime soon, the three main options to accomplish the sharing the transport media for packets and circuits services are:

1. L3 (IP) over L2 (Eth) over L1 (optical cross-connect) or L0 (lambda)

2. L2 (Eth) over L3 (IP and MPLS) over L0

3. L3 (IP) over L1, L2 (Eth) over L1 and L1 over L0

For CLARA, the recommended option is (2) which is the best option for transport of packets, which is the most part of the traffic, and the deployment would be simpler than the other options, like adding another service – in this case, MPLS/VPLS – on the traditional IP based network CLARA has been operating since its first deployment.
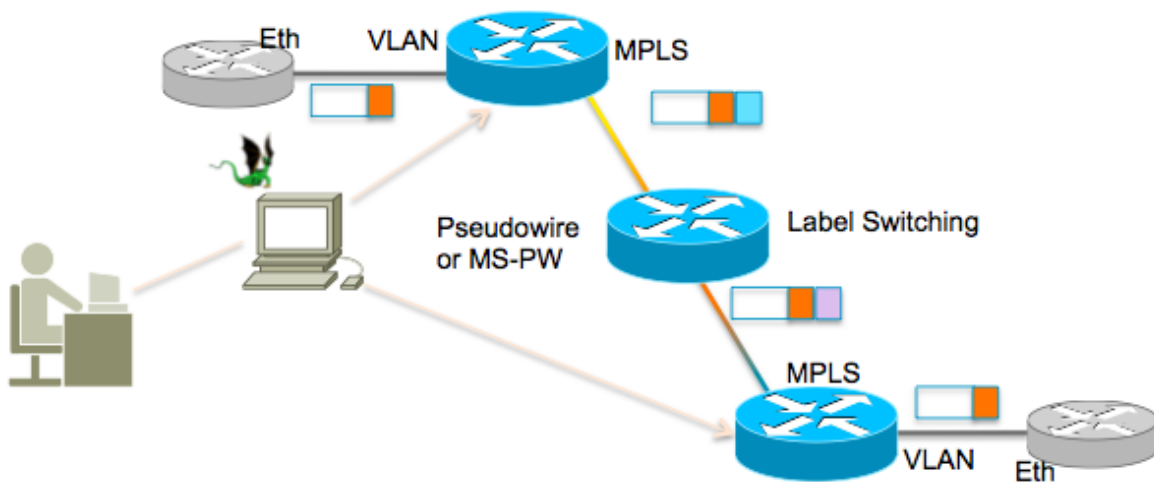
Figure 6 - Proposed model for CLARA: L2 over L3 over L0

For a matter of comparison, the model adopted by Internet2 is (3), where they have use multiple lambdas (10G) available, which permits the usage of L2 and L3 equipments for providing packets and circuits services separately. Thus, a generic cross-connect model for the Internet2 design would be:

Figure 7 - Cross-connect design for L3 over L1 and L2 over L1 using separate lambdas

Figure 8 - Same design, showing VLAN sub-interface to connect one neighbor router

The generic design in Figure 7 and Figure 8 exhibit the need for two 10G interfaces at the router to permit 10G routed traffic between neighbor PoPs. If used only one 10G interface, the inbound/outbound routed traffic would have only 5G of bandwidth.

6.- CONNECTIONS BETWEEN POPS

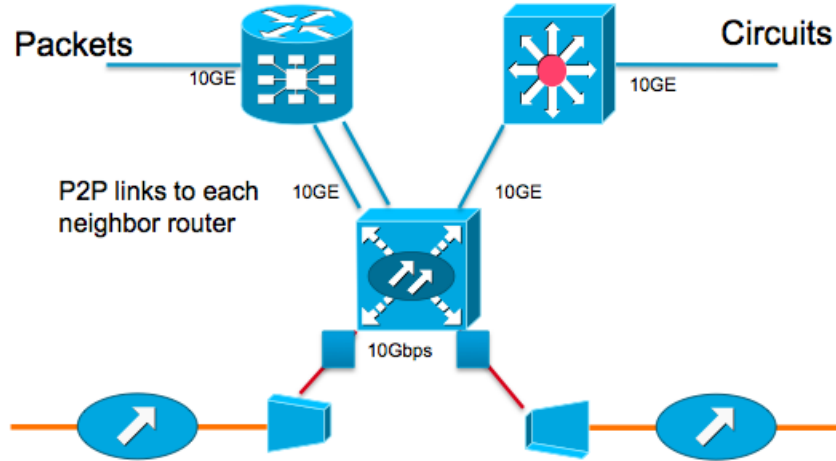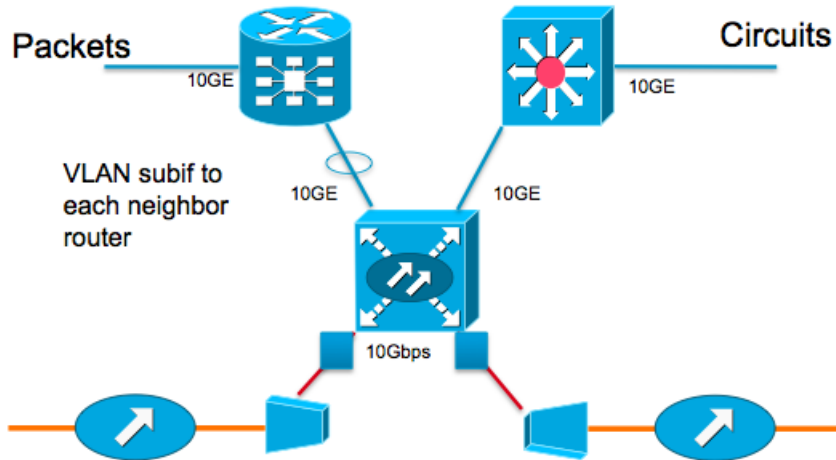CLARA needs to optimize the lambda usage that will receive from a telco or will light up itself using DWDM equipments using leased (IRU) fibers obtained on a tender process. There is also a need to build a cost-effective (CAPEX) network, based on Ethernet optimized equipments, also in a cost-effective way regarding management, troubleshooting and power consumption. The network needs to be resilient, with low convergence times and flexible for the future, being capable of supporting new applications, protocols, topology and bandwidth profile changes.

The proposed model for CLARA's PoPs allows for two basic designs. One with separate L3 and L2 equipments (Figure 9) and another design "collapsed" into one L3 equipment (Figure 10).
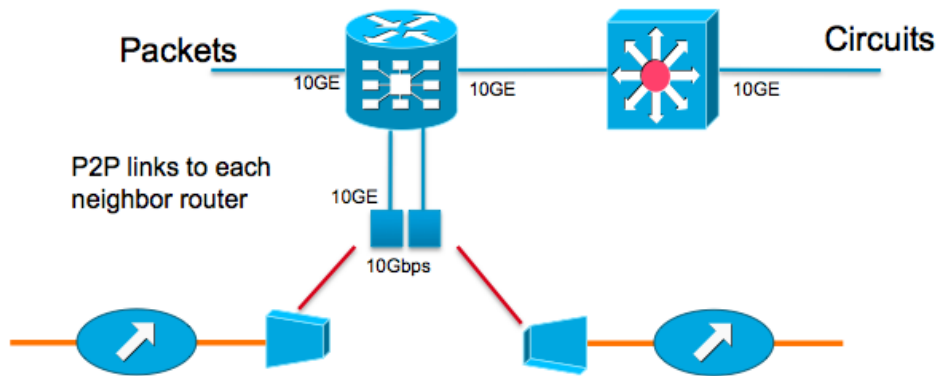

Figure 9 - CLARA PoP design proposal for 10G using separate L3 and L2 equipments

Figure 10 - CLARA PoP "collapsed" design proposal for 10G using one L3 equipment (L2 capable)

7.- NEW EQUIPMENTS FOR THE BACKBONE

The engineering proposal for the new CLARA backbone is based on some premises, acknowledged by mainstream network hardware manufacturers:

- Ethernet has become the standard technology for transport
- E-OAM (and MPLS OAM) available for the core
- New routers are Ethernet optimized (much lower cost per port)
- Statistical Multiplex allows a more efficient bandwidth usage
- L3 control plane is more robust than L2/L1 control plane (right now)
- A lot of legacy traffic is already running over L3 networks (Telcos)
- There is no evolution path for SONET/SDH gear
- Core BW availability growing from 1Gbps to 10Gbps to 100Gbps (2010)

There are some other premises that can be debatable depending on the available resources (e.g.: availability of fiber for deploying own 10G DWDM or lambdas by a provider), but were taken in account in the research and analysis for the technical proposals.

One of the considered premises is the price of 10G interfaces for routers and switches.

@LIS EUROPEAID CO-OPERATION OFFICE		CLARA
A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

Figure 11 - Price comparison of router and switch interfaces.
Source: Ethernet in Core Networks: A Technical and Economical Analysis – IEEE paper

Other assumptions:

- CLARA will not have control of Layer 1 optical network
- Will have a single lambda between PoPs (10G bandwidth)
- Bandwidth will have to be split between circuits and packets
- Protection (rings) will be available only to fiber cuts
- There will be damand growth for high bandwidth (1G) traffic for applications

Considering the design options in page 15 for a network composed of 24 nodes with 10G access, a total cost estimate by one mainstream hardware provider - considering street prices in Oct. 09, with very aggressive discounts on T/L1 and being conservative on L2/L3 hardware - would be:

XC1 – OXC – 1 if per router: USD 3.168M
XC2 – OXC – VLAN subif: USD 2.528M
MPLS1 – Router + Switch: USD 2.048M
MPLS2 – Router: USD 1.712M

Figure 12 - Estimate costs for a 24 nodes 10G backbone (based on street prices).

Where:

- XC = Cross-connect
- OXC = Optical Cross-connect
- T = Transport

For CLARA's purposes, both XC and MPLS design models would work. On the technical side, the MPLS is more similar to the current operational model than the XC. On the economical aspect, the MPLS would cost less than XC.

Also on the technical side, there are planned improvements for Ethernet as a carrier solution, so it is expected it to have better OAM[2] instrumentation – like SDH have – to make it better for transport on characteristics like manageability,

---

[2] * OAM - Operation, Administration, and Maintenance

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

reliability, scalability, simplicity and connected-oriented (operational) behaviour. If those characteristics become broadly available as standards on networking equipments, it is expected that Ethernet will be able to replace SDH as preferred transport technology.

The recommended protocols for the core of the network are MPLS and VPLS. When published as standard, the new MPLS-TP profile can be adopted, which will be able to enable MPLS (RFC3031) and pseudowires (RFC3985) to be used in a transport network and operated and supporting packet transport services in a similar manner and degree of predictability to today's most used existing transport technologies.

This design would give CLARA backbone the same capabilities as of the carriers that operate multiservice core backbones with access and aggregation based on Ethernet technology.

Comparing Access Switches and Routers

An important question regarding Layer 2 EWAN services is, "Is it better to attach to the service with a switch or a router?" Unfortunately, there is no one correct answer. Clearly there are come CAPEX advantages when one looks at the price tag of a switch versus the price tag of a router. Switches are almost always less expensive than routers, and for some networks, cost is the main or only issue. In that case a switch will be used, regardless of any networking issues. However, in most cases, a router is the better choice and will save money for the provider or enterprise in the long run, because they offer the following advantages:

- Flexible policing and traffic shaping
- Address structuring for traffic segmentation
- Fault isolation and traffic control
- A value-added service-friendly platform for service providers
- A futureproof solution

Flexible Policing and Traffic Shaping

"Policing" is the ability to look at packets, compare them to a traffic contract, and either pass them, drop them, or mark them as nonconforming. It is a common

misconception that policing alone provides complete traffic engineering—that is, that if the flow of packets is restricted into a network cloud, congestion will not occur. Although this may be true for a grossly underbooked, inefficient networks, it is not true for most "real" networks. When you simply restrict traffic into a cloud, important aspects—including traffic patterns, application-specific QoS issues, and time-of-day usage—must be considered. Even under the most thorough traffic analysis, many nondeterministic traffic patterns can still occur—especially with an EMS—any of which can cause a network element, or port, to congest and drop critical traffic. Since an EMS is a broadcast domain, its QoS characteristics are very unpredictable and can easily congest upon egress (Figure 13).
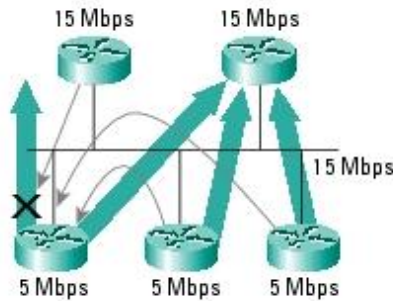
Figure 13 - Policed, but Congested, Network (Source: Cisco Systems, Inc.)

Even though policing in its own right does not constitute robust traffic engineering, it still plays a vital role in maintaining a congestion-free network. Although many switches can police, they do not have the same policing capabilities as routers. Many switches can police on a per-port basis, on IEEE 802.1P priority, and some can police on an IEEE 802.1Q VLAN. However, most routers can also do this, as well as police on IP ToS, DSCP, TCP port, UDP port, and IP address. Thus, with routers the granularity of policing can be based on IP level priorities, applications such as voice over IP (VoIP), and internal web applications, or even end stations, such as file servers or storage devices. This breadth of service enables the enterprise network to get the best use out of their expensive and critical wide-area infrastructure.

As stated earlier, policing at the edge cannot solve every problem, especially when you are trying to get every bit out of expensive WAN links. Traffic shaping adds another dimension to congestion avoidance and control. "Traffic shaping" is the ability of a router, under congestion or under traffic contract violation, to buffer and smooth traffic to an acceptable rate until the congestion or violation has abated.

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

This feature is common on routers but seldom found on switches. Even if a switch can shape, it has the same limitations, compared to routers, as policing does. Figure 14 shows a 50-Mbps contract over a 100-Mbps link. Traffic is first policed to conform to 50 Mbps, and then the excess bandwidth is throttled, or shaped, so it does not have to be dropped. Most switches would simply drop the extra traffic at the policer.
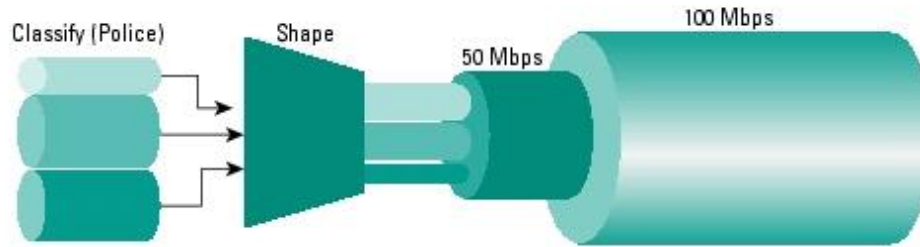

Figure 14 - Policing and Shaping (Source: Cisco Systems, Inc.)

Address Structuring and Traffic Segmentation

With IP, each end station and router has a configurable address. Although some switches and network adapters allow you to customize the MAC address, this has only a few uses, because these addresses cannot be structured or summarized like IP addresses. Summarization allows large multiples of IP addresses, in a structured system, to be stored in memory as a single-entry summary, rather than individually. This reduces memory sizes, reduces address lookup times, aids in debugging, and reduces failure and recovery time because the device needs only to relearn a summary or a group of summaries rather than a complete list of addresses. Switches do not enjoy this luxury, and since many switches cannot learn addresses at line rate, after network failure, traffic is flooded while a switch tries to relearn its forwarding table. This only exacerbates the congestion and QoS problems associated with the failure.

Summaries can also be hierarchical. For example, a common scheme is to hierarchically vary summarizations as they relate to the access nodes, distribution layer, and core of a network. However, VLANs, like MAC addresses, cannot be hierarchically summarized. Even a VLAN tag-stacking scheme is a one-to-one mapping of customer to tag and is only used by an service provider. Thus, Ethernet cannot duplicating an IP or MPLS hierarchy. By using hierarchies, you can segment broadcast domains. This means that broadcast storms, intentional and unintentional, can be contained to small communities of interest that, under strain,

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

do not affect the rest of the network.

Many protocols also help make Layer 3 multicasting even more efficient. Host-to-router protocols, such as IGMP, and router-to-router protocols such as PIM allow routers to create minimal tree multicast structures, ensuring that the multicast packets traverse only those links destined for valid destinations, rather than being broadcast (Figure 15). If this were done over a Layer 2 network, the multicast would be flooded throughout the Layer 2 domain (Figure 16). Furthermore, even when a switch performs IGMP and PIM snooping, there are issues regarding failure-recovery behavior and, more importantly, QoS, which limits these features' potential and predictability in large networks.)
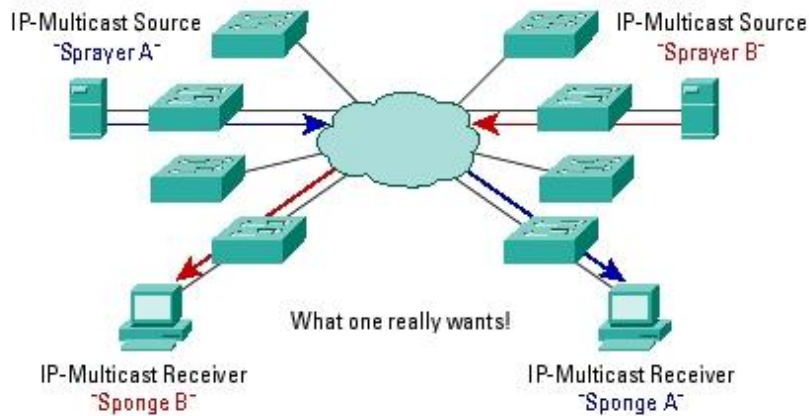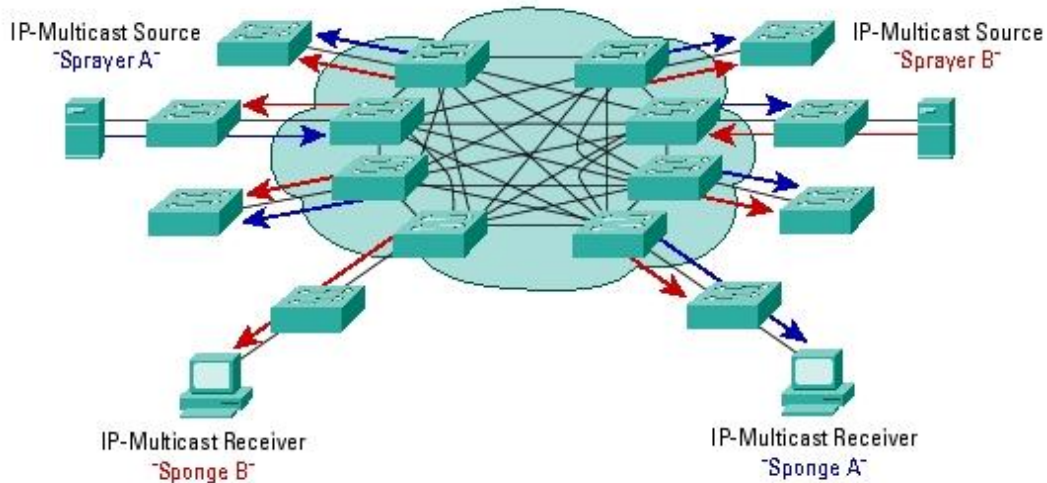

Figure 15 - Desired Effect When Multicasting (Source: Cisco Systems, Inc.)

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

Figure 16 - Adverse Effects of Multicasting over a Layer 2 Ethernet Switched Network (Source: Cisco Systems, Inc.)

Fault Isolation and Traffic Control

Another important quality of traffic segmentation is fault isolation. Since traffic can be highly segmented, when issues arise they are constrained to smaller areas, allowing them to be located, and thus fixed, more quickly—lessening the mean time to repair (MTTR). Also aiding MTTR is the structured nature of IP and the vast array of tools that take advantage of this structure. Two common tools often taken for granted are ping and traceroute. These simple tools allow you to determine if a host or router is reachable at the network level, pointing to potential application-layer issues. Traceroute also lists the path that a packet takes as it traverses the network layer, pinpointing the beginning of the failure. In addition to these basic yet effective commands are the vast array of proprietary or management-based software tools available on management systems and protocol analyzers. In contrast, even simple tools like ping and traceroute have no counterparts in the switching world. In addition, as packets traverse from one Layer 2 boundary to another, they pass through routers, and at that point any Layer 2 packet trace loses its end-to-end significance, rendering it useless. The availability of these Layer 3 tools makes it easier to debug a network-not to mention that it reduces the number of expensive ($50,000 and up) protocol analyzers you need to buy.

Finally, routing protocols offer greater flexibility and control over to the path a packet takes through the network and how the topology reacts to change (Figure 17). Spanning Tree Protocol, the most commonly used Layer 2 topology protocol, has several shortcomings when it comes to large-scale networking. It is slow to react, needs to block links (and in most cases many links) rendering them useless, has trunking support protocols (IEEE 802.17AD) that cannot share loads over multiple-nonparallel links, and it cannot forward based on policies. On the other hand, link-state-based routing protocols can quickly react to and repair large-scale networks and can forward and balance loads over any number of links based on policies such as source, destination, route priority, and congested transit network. All of this allows you to use your network efficiently. Another shortcoming of large-scale flat Spanning Tree Protocol networks is that, to keep the network properly utilized and maintaining proper QoS the service provider has to reengineer the network every time a new customer is added. In contrast, routing protocols self-adapt as new users come online.

STP Blocks Links   **Routeers Do Not Block Links— They Load Balance**
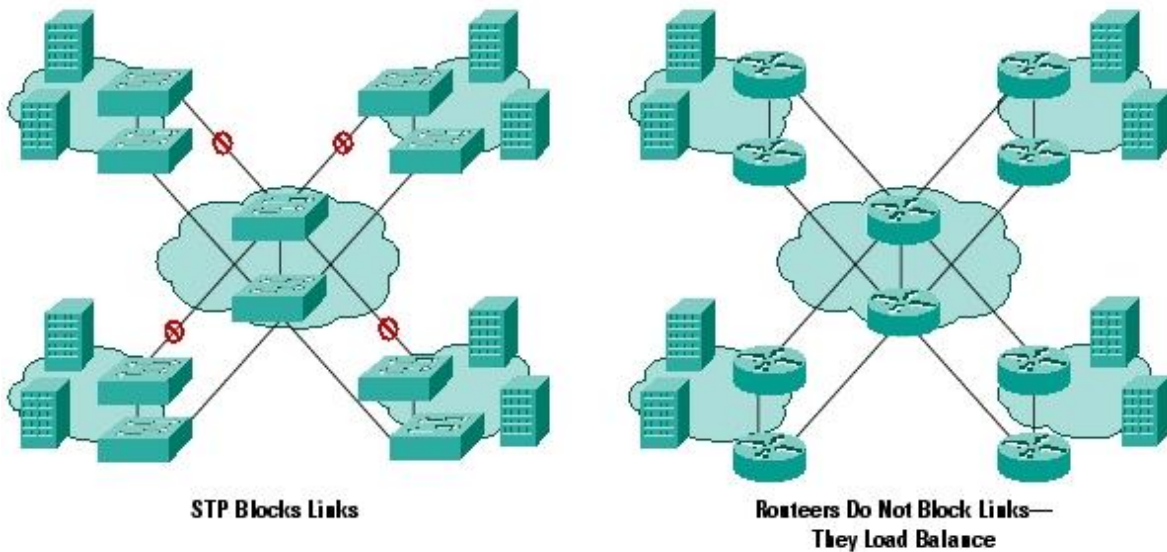
Figure 17 - Inadequacies of Spanning Tree Protocol Versus Routing Protocols (Source: Cisco Systems, Inc.)

SP Value-Added, Service-Friendly Platform

Another big advantage gained when choosing a router as an access device is that the service provider's ability to deploy value-added services grows. Simply put,

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

routers are feature rich. They are so because all the important features reside above Layer 2, allowing the enterprise to outsource high-touch services and save money. These services range from security to voice management to storage integration. This allows an enterprise to streamline its network needs by combining extranet resources and sending them over one service and one WAN port.

A common example of such a service is outsourced firewalling—an enterprise's frontline defense, which authenticates and controls outside access. IP Security (IPSec) VPNs offer additional security by authenticating sources and encrypting data before it passes over WAN links. IPSec is commonly used by financial institutions and government authorities to protect their data. In addition, a routed access network allows the service provider or enterprise to deploy intrusion-detection software to detect and locate hackers.

Voice-related services include Survivable Remote Site Telephony (SRST) and IP Centrex. SRST detects failures (unreachable destinations) in the network, and then takes IP telephony calls and reroutes them to the public telephone network, rather than dropping the calls because the destination is no longer reachable. For those who do not wish to manage their own IP telephony system but still want the cost savings associated with an integrated voice and data network, service providers offer IP Centrex, a remote service that offers call-management features including voice mail.

Another exciting area that can be integrated is IP storage. IP enhances traditional storage networks by allowing storage area network (SAN) traffic to be multiplexed along with voice and data traffic. In addition, IP storage can apply IP structuring and VLAN concepts creating virtual storage communities that allow an enterprise to better utilize its facilities. These virtual storage communities can either be managed by the enterprise or outsourced to the service provider.

Routers Are a Futureproof Solution

Routers are the right technology for the future, not just today. Whereas switches cannot act as routers, most access enterprise routers can also act as full Layer 2 switches. More importantly, these routers are based on IP and they run IP protocols. The primary purpose of an IP routing protocol is to scale—IP routing protocols were developed because networks became too large to run without them.

| A project of the @LISII Programme | This project is funded by the European Union | A project implemented by CLARA |

In addition, for service providers to continue to operate, they must continue to develop value-added features that provide real benefits to their customers. This requires routers. In addition, many routers now are application aware—and even when they are not, application developers bind their applications to TCP and UDP ports and make it possible for their applications to write to ToS and DSCP priority fields. These hooks are not readily available to switches, and without them a service provider's ability to provide today's or future value-added services becomes difficult and expensive. Thus, today's routers are "futureproof" in that they will continue to scale and provide new and beneficial services to meet tomorrow's needs.

When Is It Acceptable to Use an Access Switch?

Given that switches typically cost less than a router, using a switch as an access node merits consideration, even though there are very inexpensive routers on the market. There are two cases in which it is acceptable to deploy a switch as an access node. The first occurs over a point-to-point link using EPL. In this case, the EWAN link appears to be another segment in a LAN, and EPL is secure in that it does not terminate on any extranet services (Figure 18). The second case occurs when dark fiber when is used, for the same reasons as with EPL. If you use an access switch with EPL, or EWS a hub-and-spoke topology should be implemented; the spoke nodes are switches that hub back to a router. Furthermore, hub-to-spoke traffic is switched, while spoke-to-spoke traffic is routed, for security and traffic segmentation reasons. This network design is commonly found in school districts and remote medical clinics or offices. You should never use a switch with an EMS because QoS, security, and traffic patterns are unpredictable.

Figure 18 - Using Switches over EPL, Mapping Each Circuit to a Unique VLAN (Source: Cisco Systems, Inc.)

Can a L3 switch be a router?

Robust L3 switches forward on L3 addresses, so can they be considered routers, and must participate in routing protocols (OSPF, BGP-4 etc.). Many L3 switches do not support IS-IS, and many of them lack MPLS and VPLS features. Usually L3 switches are limited to what the hardware supports, but with forwarding based on TCAMs, many forwarding modes can be supported, and replacing the switching logic is cheaper (especially with centralized forwarding).

8.- SUMMARY: NEG PROPOSAL FOR CLARA INFRASTRUCTURE

The CLARA-NEG proposal for RedCLARA2 infrastructure is composed of the following topics:

- Network Design
- Network Services (Layer 2 and Layer 3)
- PoP infrastructure design
- PoP network equipments
- Client connections

Network Design

The proposed architecture for CLARA's PoPs, is a combination of one router (optimized for Ethernet) plus one Layer 2 (or Layer 3) switch. The router and the switch should be both equipments with hardware redundancy and interconnected through an 10Gbps multimode uplink.

The proposed architecture offers:

Wide range of services: Layer 2, Layer 3, L2VPN, E-LINE, E-LAN, E-Tree (multicast) etc.

- Full BGP routing, QoS, IPv6, Multicast IPv4 / IPv6 etc.
- Fully redundant router.
- Desirable features:
- Support for Ethernet OAM, MPLS OAM and Inter-AS VPLS
- Possibility of routing equipment with support for 100GbE in roadmap*

| A project of the @LISII Programme | This project is funded by the European Union | A project implemented by CLARA |

* This is optional and depends of availability dark fiber IRU offerings or CLARA having access to multiple lambdas from a carrier or to having its own resources deploy DWDM equipment.

According to [2], there are some common wisdom facts about backbone costs, saying that nowadays bandwidth costs are very low, bacause of DWDM technologies; routing hardware is very expensive, in particular, high-capacity (>= 10Gb/s) interfaces/line cards; and the only way to build cost-effective high-speed backbones is to build a DWDM network, use one lambda for "legacy" packet (IP) network, and use circuit switching to provide "end-to-end" connections to sites. The incremental cost for additional bandwidth can be low, by just adding more wavelengths.

Still according to [2]. the up-front investment in a DWDM network can be quite high, regarding infrastructure items like fibers (IRU/construction), Amplifiers, dispersion compensation etc., things like synchronization (for SDH/SONET); the overall operational expenditure (OPEX); fiber lease/maintenance; device maintenance and housing.

Network Services

CLARA updated infrastructure should allow integration with international research initiatives for provisioning of circuits services for advanced applications.

The proposed design and technologies wold allow CLARA to have support for a range of services and technologies, as listed in table.
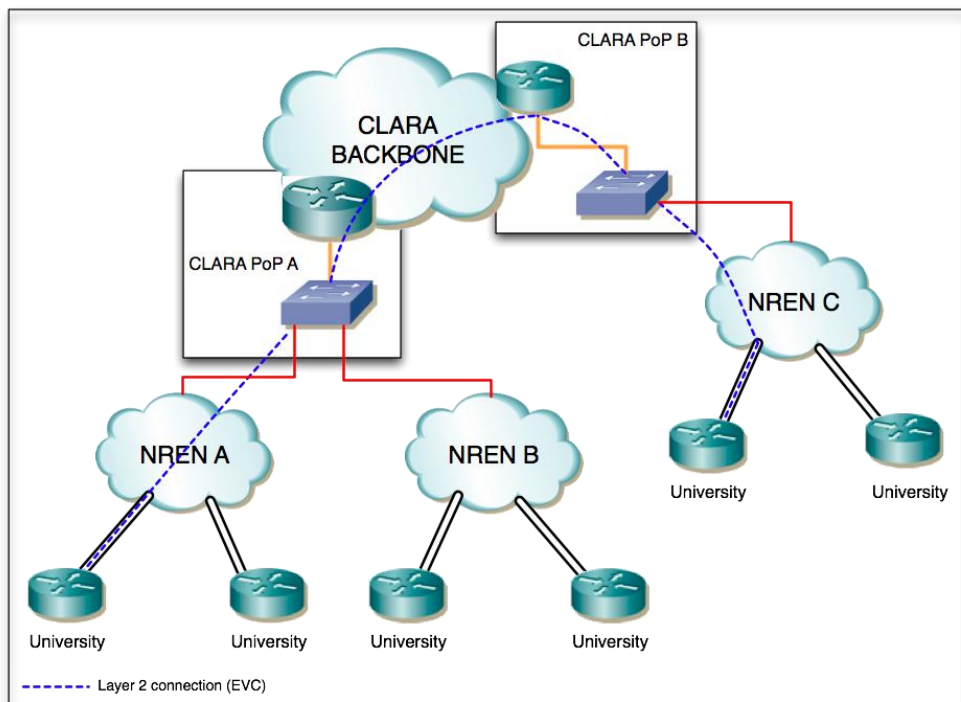
	Characteristic	**SONET / SDH**	**Optical OTN (ROADMs)**	**Electrical OTN**	**PBB-TE**	**MPLS-TP**	**IP/MPLS**
Ethernet	**Eline (10GE)**	Y	Y	Y	Y	Y	Y
	Eline (sub 10GE)	Y	N	Y	Y	Y	Y
	E-Tree	N	N	N	N	Limited	Y
	E-LAN	N	N	N	N	Limited	Y
Legacy	**Frame Relay**	N	N	N	Limited	Limited	Y
	ATM	N	N	N	Limited	Limited	Y
	TDM	Y	N	N	Limited	Limited	Y
IP	**L3VPN**	N	N	N	N	N	Y
	L3 Unicast	N	N	N	N	N	Y
	L3 Multicast	N	N	N	N	N	Y
	Content	N	N	N	N	N	Y
General	**Traffic Engineering**	Y	Y	Y	Y	Y	Y

50ms recovery	Y	Y	Y	Y	Y	Y
Multiplexing Technology	Time division	Wave division	Time division	Statistical	Statistical	Statistical
UNI processing	Limited	None	None	Typically rich	Typically rich	Typically rich
Granularity	VC-4	LAMBDA	ODU	Variable	Variable	Variable
Technology Maturity	Y	Y	N	N	N	Y

Table 1 – Comparison of transport technologies and supported Layer 2 and Layer 3 services today (may change in the future)

The recommended transport technology for the access (edge) of the network is Ethernet, which can allow deployment of Layer 2 circuit services and those will have a well defined client edge (CE) and provider edge (PE) demarcations, separated from the CLARA core equipments (P).

Figure 19 - Proposed CLARA EVC (Ethernet Virtual Connection) Service using pseudowires

PoP Infrastructure Design

The new CLARA PoP basic infrastructure should allow the connection of clients (NRENs) and PoP LAN servers with 1GbE and 10GbE bandwidths, trying to minimize the access connection cost, maximizing the bandwidth and services according to the clients requirements.

As the new CLARA PoP should have options for connections that offers both Layer 2 and Layer 3 services, both options are available in the proposed design for the CLARA PoP, as depicted in Figure 20.
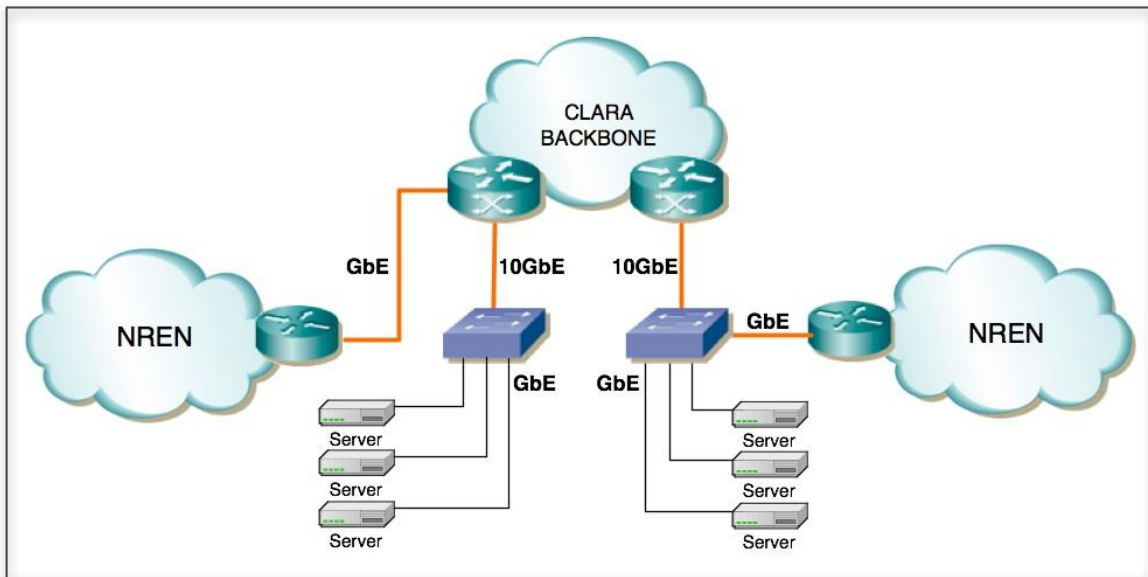

Figure 20 - CLARA PoPs with clients access attached to Layer 2 or Layer 3 equipments.

PoP Network Equipments

For the core of the CLARA backbone there were analysed some design options trying to fulfill the requirements of the required bandwidth, resilience, availability and costs for the services to be offered by the CLARA backbone.

One of the proposals of connections is through making use of Layer 2 - or Layer 2/3 - switches instead of routers. After an ample - but not exaustive - investigation,

| A project of the @LISII Programme | This project is funded by the European Union | A project implemented by CLARA |

the following conclusions were drawn based on the informations found.

It was not found – in a non-exaustive search – of a case or example of a production (e.g.: clients' only access to IP commodity and research networks) WAN backbone using this design, with a few exceptions of experimental testbeds, like GIGA[3], KyaTera[4], PREAMBULO and FEDERICA.

One first proposal for the design of the CLARA core backbone so it can be cost effective in terms of equipment investment and, at the same time, enabling pure Layer 2 services for the clients is to deploy Layer 2 switches (cheaper) at most PoPs and restrict the deployment of routers (more expensive) to a minimum of the PoPs. Nevertheless, this model is highly experimental and no similar design was found in the research for this proposal.

The most required service needed for the client NRENs would be the availability of two types of Layer 2 services - point-to-point and multipoint - besides the well-known Layer 3 services.

The proposed design for CLARA's PoPs gives the flexibility to have clients and PoP's LAN services connected to the Layer 2 switch or to directly to the Layer 3 router, with support for a wide variety of network services.

The PoP design would be two-tiered: aggregation / access (PE) and Core (P). This design gives a CLARA PoP a flexibility to connect clients to a L2/L3 switch offering multiple 1GbE interfaces, offering access to a MPLS/VPLS core, that would transport the clients' circuits, or establish the E-LAN (VPLS) – multipoint – service (Figure 20).

---

[3] http://www.giga.org.br/

[4] http://www.kyatera.org.br

@LIS EUROPEAID CO-OPERATION OFFICE		CLARA
A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

Equipments options for CLARA PoPs

	Option 1	**Option 2**	**Option 3**	**Option 4**
Hardware Types	L3 Ethernet optimized + L2 Switch (1G / 10GbE)	L3 Ethernet optimized	Layer 3 router (TDM) + Layer 2 switch (1G / 10GbE)	L2 switch (1G / 10GbE)
Supportes Technologies	L2VPN, L3VPN (MPLS, VPLS), QoS, Multicast, IPv6 etc.	L2VPN, L3VPN (MPLS, VPLS), QoS, Multicast, IPv6 etc.	L2VPN, L3VPN (MPLS, VPLS), QoS, Multicast, IPv6 etc.	L2VPN, (MPLS, VPLS), L2QoS, Multicast, etc.
Pros	Chassis and interfaces are cheaper than "traditional" L3 routers that support TDM interfaces - Optimized for Ethernet - Some equipments support (or will support) 40GbE and 100GbE	- Chassis and interfaces are cheaper than traditional TDM routers - Optimized for Ethernet - Some equipments support (or will support) 40GbE and 100GbE - Some models can support everything in one chassis without need of another Layer 2 box for 1GbE aggregation.	- Cheaper TDM interaces: SONET, SDH, ATM and FR interfaces - More variety of interfaces bandwidths	- Cheaper equipment with 1GbE and 10GbE interfaces - Cheap Layer 2 Ethernet services - Faster than L3 equipments
Cons	- Cheapest switches does are not fully redundant and are less realible	- TDM interfaces available, but are more expensive than Option 3 interfaces and there is limited support to interface variety. Some routers minimum bandwidth is STM-1.	- L3 equipments costs more than Ethernet optimized routers - Cheaper switches are not fully redundant, are more limited, are less realiable and does not have support for advanced features as MPLS and VPLS. - High cost for L2 equipment with redundancy and support for MPLS / VPLS	- Does not support any L3 services - Less flexible
Some Compatible models	Brocade (Foundry) CES 2000; Cisco ASR 9000, GSR and CRS-1 series; Extreme x450; Juniper series MX, T and EX	Brocade (Foundry) CES 2000, Cisco ASR 9000, GSR and CRS-1 series; Extreme x450; Juniper series MX, T and EX.	Cisco GSR and CSR Series; Juniper M, T Series	Brocade; Cisco 3700 Series; Extreme; Juniper EX series.

Table 2 - Characteristics of equipments options for CLARA PoPs

	L3 Services	L2 Services	10G Interface $	Points
Option 1	5	5	3	13
Option 2	5	5	3	13
Option 3	5	5	1	11
Option 4	0	5	5	9

Table 3 - Comparison of equipments options for CLARA PoPs

9.- CONCLUSION

This document presents the proposed basic design for the CLARA's PoPs, and for its new network equipments, along with information about the premises, assumptions, services, costs and investigations that were investigated to analyze and to support it.

The proposed architecture is based on a two tiered architecture composed of one Ethernet optimized Layer 3 router and, optionally, one Layer 2 switch with support for 10GbE interfaces, multiple 1GbE interfaces and a backbone core with support for MPLS/VPLS technologies. If used, the Layer 2 switch would be the provider edge (PE) demarcation and access aggregation, between the client and the provider (P) equipment, which is the CLARA's core router.

Alternatively, the architecture can be collapsed into only one Layer 3 Ethernet optimized router.

This design permits offering of flexible L2 and L3 services to for NRENs, being MPLS point-to-point or multipoint (VPLS) between CLARA's participants and/or with international collaborators.

10.- REFERENCES

1.Enterprise Connections to Layer 2 Ethernet Services Design
http://www.cisco.com/en/US/netsol/ns341/ns396/ns223/ns227/networking_solutions_white_paper09186a00801e1225.shtml

2.Myths about costs of circuit vs. packet switching
http://www.terena.org/events/tnc2006/programme/presentations/show.php?pres_id=214

3.Considerations for Implementation of MEF Services

A project of the @LISII Programme	This project is funded by the European Union	A project implemented by CLARA

https://www.ethernetacademy.net/index.php/20081012106/Ethernet-Academy-
Articles/considerations-for-implementation-of-mef-services.html

4.To Route or Not to Route?
http://www.terena.org/events/tnc2006/core/getfile.php?file_id=737

5.UKERNA - Network Design
www.ja.net/documents/communities/techinfo/NetworkDesignv3.1.doc

6.Juniper Networks - Characteristics of Switches and Routers
http://jnpr.net/solutions/literature/white_papers/200161.pdf

7.Layer 2 VPN experiences over a metro IPoDWDM network
http://www.it.uc3m.es/cgarcia/articulos/fvalera_eunice04.pdf