

IEEE INFOCOM 2006 High-Speed Networking Workshop: The Terabits Challenge

HIGH-SPEED NETWORKING WORKSHOP: THE TERABITS CHALLENGE

In Conjunction with IEEE INFOCOM 2006

Barcelona, Spain, April 24 th , 2006

TCHSN

Workshop General Chair

Thomas Ndousse

Office of Science United States Department of Energy
(DOE)

Phone: (301)-903-9960

Email: tndousse@er.doe.gov

Technical Program Chair

Nasir Ghani

Department of Electrical and Computer Engineering Tennessee Tech University Cookeville,
TN 38506 Phone: (931) 372-3046

Email: nghani@tntech.edu



WORKSHOP PROGRAM

7:30 – 8:15 a.m. Registration

8:15 – 8:30 a.m. Chairs Opening Welcome



13:15 - 14:30 p.m. Lunch

14:30 - 15:45 p.m. Session 3: IP Networks & Switch Design

- “Hierarchical Broadcast Ring Architecture for High-speed Ethernet Networks”, *H. Jang, H. Kim (Carnegie Mellon University)*
- “Multi-Terabit/s IP Switching with Guaranteed Service for Streaming Traffic”, *M. Baldi (Politecnico Di Torino), Yoram Ofe (Università di Trento)*
- “Packet Classification at Ultra High Speeds”, *J. Van Lunteren (IBM Zurich)*
- “Ethernet Fabric Routing (EFR): A Scalable and Secure Ultrahigh Speed Switching Architecture”, *J. Morales Barroso (L&M Data Communications), G. Ibañez Fernandez (Universidad Carlos III Madrid)*
- “Parallel Firewall Designs for High-Speed Networks”, *E. W. Fulp (Wake Forest University)*

19:00 – 21:00 p.m. Dinner & Plenary Session

Ethernet Fabric Routing (EFR): A scalable and secure ultrahigh speed switching architecture.

José Morales Barroso, L&M Data Communications jmb@ieee.org
Guillermo Ibáñez Fernández, Univ. Carlos III Madrid gibanez@it.uc3m.es

A new ethernet based switching architecture for the Universal Ethernet Telecommunications Service (UETS) is proposed, capable of performances up to one Terabit per second. It is based on hardware switching of Ethernet frames using topological and hierarchically assigned standard local MAC addresses. The architecture solves Ethernet's scalability and security problems, has low cost/performance ratios and is compatible with existing Ethernet and IP networks, providing a non disruptive migration path to high performance networks and overcoming TCP/IP performance limitations.

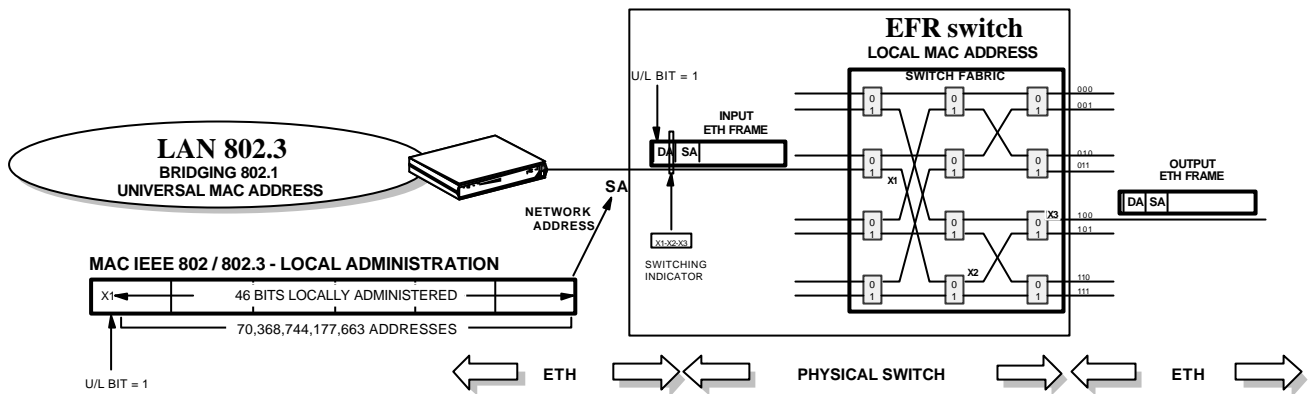


Fig. 1. Ethernet Fabric Routing Architecture

The availability and good cost/performance ratio of Gigabit and 10 Gigabit Ethernet switches has led to increased sizes of Ethernet domains, till now fragmented by multilayer switches and the like, to complete campus and enterprise networks, Metropolitan and even Wide Area Networks.

However the current use of Ethernet universal MAC flat addresses does not scale due to their lack of hierarchy that prevents aggregation of routes. The schemes currently being standardized, based on Ethernet encapsulations, IP Tunnels or MPLS transport, are both complex and expensive.

This paper summarizes the UETS/EFR architecture. More information is available at [1][2] and successive papers.

Operation

The proposal is based on a new node concept, which uses hierarchical local (U/L bit) addresses to perform routing and switching of Ethernet frames using HW switch fabrics like the ones used in Fibre Channel, Infiniband, ATM and many others. These switches build up an EFR Domain that may cover a local campus network, an ISP area, or a metropolitan area network. An Ethernet DNS service (EDNS) is used at the domain to perform translation from URLs and IPs to UETS addresses. The hosts at the network may be either standard hosts or dual stack terminals. Standard hosts operate in the standard way using TCP/IP. Dual stack (L2 and L3 stacks) terminals, named TUE, use the Ethernet over LLC stack for intra UETS domain communication and TCP/IP for external and (if desired) internal communications. The switching principle is shown at figure 1 and the TUE dual stack at figure 3. Hosts use standard IP addresses to connect to hosts outside the EFR

domain and may use just Ethernet or both Ethernet and IP inside the EFR domain.

To set up communication to a host or server in the UETS domain, the destination URL or IP address is resolved to its MAC address. This is done via EDNS. In the LLC/Eth stack, TCP connections are replaced by LLC type 2 (connection oriented service). This allows congestion control at each switching node and flow control end to end. UDP protocol is replaced by LLC type 1 (connectionless unconfirmed service). At network terminations interfacing with 802.1 LANs, encapsulation or translation of standard MAC addresses is required to allow interoperability. In the UETS domain, MAC addresses are physically dependent and controlled by the ISP or network owner, allowing hardware based switching and enhancing network security by preventing layer two address spoofing. Local (UETS) addresses may be hidden to service users for enhanced security.

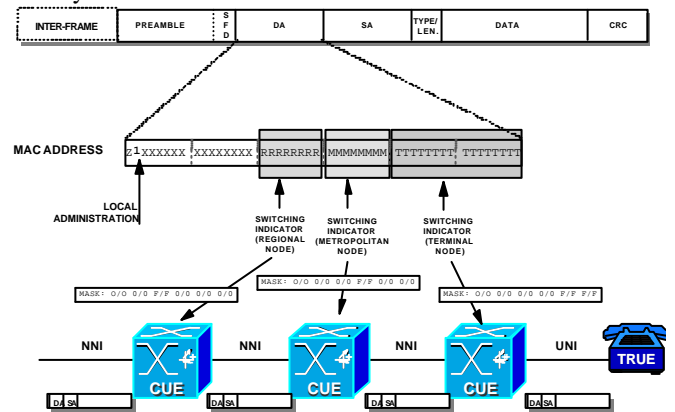


Fig. 2. Addressing and switching principle. Example.

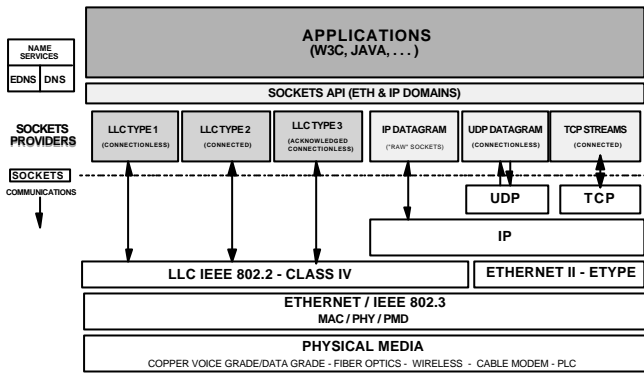


Fig. 3. TUE Dual stack LLC/Eth + TCP/IP/Eth

UETS/EFR RM: simple LLC/Eth stack

The UETS system reduces the complexity of the Network to an astonishing simplicity: Just two protocols (LLC/Eth) combined in different ways provide all the required service types: Ethernet/ 802.3 transports the information, and LLC / 802.2 performs the control as shown in Fig. 4.

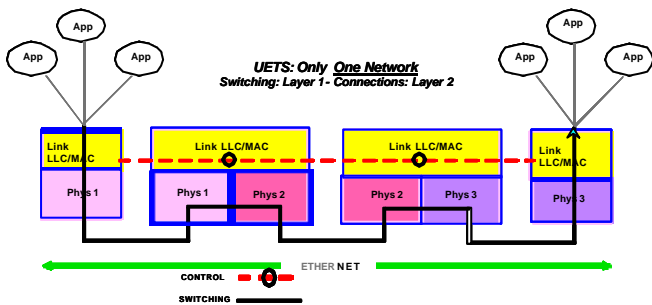


Fig. 4. LLC over Ethernet control and switching.

Congestion control

Standard congestion control is currently performed end-to-end by TCP/IP congestion control mechanisms. TCP/IP mechanisms were designed to obtain reliability and controlled flow through an unreliable network. In the context of increased performance and reliability of today's networks, it starts to show its limitations. Congestion control performed at layer two is currently under discussion at IEEE Congestion Management [4] interim meetings and a PAR [5] is under preparation.

As Ethernet currently satisfies TCP/IP's requirements, the goals of congestion management at layer two include [4]: make Ethernet capable of supporting Fibre Channel applications and using Ethernet as a backplane connect within a system. UETS may achieve these goals allowing high performance Ethernet frame routing through simple HW based frame switching.

EFR Routing and Ethernet Fabric Switching.

When describing layer two networks, some ambiguity exists regarding the precise meaning of the terms *switching*

and *routing*. In the architecture described, *routing* is performed between switches for transit of frames toward their destination area (i.e. frames that are outside the UETS address domain of the switch). This is named *microrouting* and will be detailed in another paper. *Switching* of frames is performed across the switches located inside an address domain. Both switching and routing are performed without label swapping at nodes.

Switch designs

EFR switching is designed to perform frame switching at Layer 1. So it is well suited to Banyan type switches where selection of the switching path is performed by successive decoding of the destination address field. However other switch designs can be used, like those oriented to tag switching or label switching like ATM based cell switches or MPLS switches. Figure 5 illustrates this option. Label swapping is not needed, eliminating memory accesses and frame modification. The switch does neither need internal tables to map tags or labels to output ports. Fibre Channel fabric uses 24 bit addresses and can be used directly, applying a bit mask to the MAC destination address.

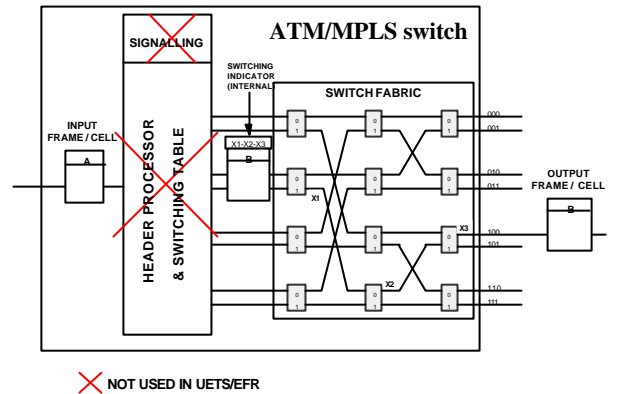


Fig.5. Alternative EFR switch

Scalability and domain size

EFR switching architecture is able to scale from small domains up to international domains. UETS domains may be as small or as large as desired. Each UETS domain requires an E-DNS service to translate URLs to UETS addresses of endnodes working in UETS mode, with an LLC/Eth stack instead of or in addition to the TCP/IP stack.

Scalability of UETS and EFR derives from two facts: its inherent hierarchy and the 48 bit MAC address length that can be used for up to 2^{48} addresses (more than 70 trillion unicast addresses). See Figure 2 for a hierarchical addressing example.

If even more addresses are needed, using UETS plus MAC-in-MAC and VLANs for addressing provides a total of 16 bytes are available for addressing, equivalent to the addressing capability of IPv6.

VLANs and encapsulation

The architecture is fully compatible with IEEE 802.1Q VLANs and other Ethernet encapsulations like MAC in MAC. MAC in MAC encapsulation is used at UETS domain border switches to standard 802.1D bridged areas.

QoS and CoS

EFR switches use the standard Ethernet frame structure; the only difference is the U/L bit usage for addresses. They can implement the same priority mechanisms as standard 802.1Q switches regarding priority levels in order to provide different classes of service inside the switch, according to the explicit priority tag (802.1p) in the frame. If traffic prioritization is performed per frame by traffic classification instead of by explicit tagging, the same mechanisms can be applied as those used by standard Ethernet switches.

The proposed LLC/Eth protocol stack allows service differentiation. Considering the four basic classes of traffic: premium service, mission critical, stop and go, and best effort. Traffic like real time voice, that is critically sensitive to delay (both Constant Bit Rate rt-CBR, and Variable Bit Rate, rt-VBR), may be transported using the unacknowledged, connectionless, LLC-1 protocol. Mission critical traffic may be transported using LLC-2, which supports connections and flow control. Stop and Go services may use the acknowledged, connectionless LLC-3 service.

MAC address configuration

EFR switching and routing is based on the correlation of MAC addresses to the physical hardware architecture. Due to its locally controlled address structure, configuration of addresses may be coordinated and arranged hierarchically. Configuration of MAC addresses may be performed either as an standard Network management activity via SNMP over Ethernet (RFC1089) or via automatic address assignment protocols executed by EFR switches intercommunicating with the Root UETS switch and address server.

Security

The UETS/EFR architecture provides inherent additional layer two security. Due to the relation of MAC addresses to the point of attachment to the network, MAC addresses can not be spoofed, as they are controlled by the switch owner ISP, and physical access to the point of attachment is required. Additional security measures using 802.1X may complement this security.

As there is no spanning tree protocol used in the switched network, no spanning tree protocol attacks are possible. As the switches do not learn addresses, no attacks are possible to the switches caches.

Node Mobility

Node mobility can be handled by UETS. A specific range of local MAC addresses may be reserved for mobile nodes. Mechanisms similar to mobile telephony (GSM) or mobile-IP can be devised to support mobility.

E-DNS performance requirements

The operation and requirements for the E-DNS service is in the worst case similar to the conventional DNS service. However, if there exists a numerical correspondence between UETS domains and geographical domain names (e.g. .es .us .it), EDNS load is reduced drastically.

Applications

The architecture can be applied to high speed networks (LAN/MAN/WAN), Storage Networks (SAN/NAS), High Performance Computing (HPC) and Networks of Workstations (NOW), Layer 2 VPNs, Video HDTV distribution, home networking and military secure networks. The main advantages of the architecture are its inherent security, wire speed performance, TCP/IP/Ethernet compatibility and interoperability with existing IP networks and applications, and lower cost/performance ratios. The architecture proposed is currently under consideration for several evaluation projects in different application areas such as Access Networks, HDTV distribution and Computer on Net architectures. With the architecture proposed, networks based on the Ethernet paradigm are now able to evolve toward higher performance and security, while maintaining compatibility with the IP architecture.

Acknowledgements

This paper was partially done with the support of Ministerio de Educación y Ciencia of Spain through Project CAPITAL (TEC2004-05622-C04-03/TCM). Thanks to Matt Hutton, who reviewed the manuscript.

Conclusions

Although apparently radical, the new and simple UETS/EFR Ethernet based switching architecture described in this paper is compatible with TCP/IP networks and exhibits outstanding advantages: It performs as highly as the LLC protocol permits, up to one Terabit per second; provides inherent security at layer two; exhibits lower latencies than TCP/IP protocols; allows maximum hardware switching speeds and full link utilization and scales to all network sizes. The architecture is suited to High Performance Networks in LAN, Enterprise and MAN/WAN transport networks. Applications specially suited to this architecture are: Internet Providers (ISP) and Quadruple Play, HDTV Distribution, Home Networking, Secure Networks, L2VPNs, High Performance Computing and Storage Networks.

References

- [1] José Morales Barroso, "From Computer Networks to the Computer on Net", IEEE Communications Magazine / Global Communications Newsletter, October 2005, pp. 2-4.
- [2] José Morales Barroso, "A new Communications Architecture and Switching Paradigms", [online] <http://www.lmdata.es/uets/uets-cm1.pdf>
- [3] Tanmay Gupta, Manoj Wadekar, Jeff Wise, "Congestion Management Capabilities of Various Fabrics", [online] <http://www.ieee802.org/1/files/public/docs2006/new-cm-capabilities-of-various-fabrics-0106.pdf>
- [4] Mick Seaman "Congestion Notification", [online] <http://www.ieee802.org/1/files/public/docs2006/new-seaman-cm-congestion-notification-0206-01.pdf>
- [5] IEEE 802.1 "Congestion Notification: Draft PAR", [online] http://www.ieee802.org/1/files/public/docs2006/new_cm_wadekar_draft_PAR_congestion_notification.pdf

Is it time to simplify the protocol stack where possible?

Ethernet Fabric Routing (EFR): A scalable and secure ultrahigh speed switching architecture

IEEE INFOCOM 2006
High-Speed Networking Workshop: The Terabits Challenge

*José Morales Barroso, L&M Data Communications
Guillermo Ibáñez Fernández, Univ. Carlos III Madrid*



<http://www.LMdata.es/uets.htm>

Ethernet issues today (IEEE 802.1/802.3)

- **Scalability: Flat (universal) MAC addresses & Spanning Tree**
 - MAC address explosion when tens of thousands end nodes
 - Standard Encapsulations: MAC-in-MAC, Q-in-Q, C-MAN
 - Increased complexity of configuration, overhead & frame length
 - Links blocked to avoid loops, limits strongly network size
 - Convergence speed (up to 2 seconds maximum RSTP)
- **Configuration complexity of networks**
 - Complex configuration when Multiple Spanning Trees (VLANs) used to improve infrastructure usage
 - IP addresses need to be configured/administered
- **Inherent lack of security at Layer 2 more important than ever**
 - MAC spoofing, host cache corruption, bridge tables saturation (switch goes to hub mode), spanning tree attacks (false root bridge)

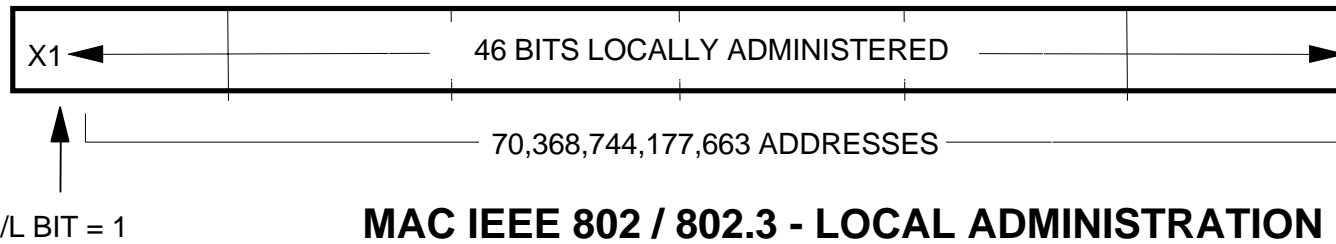
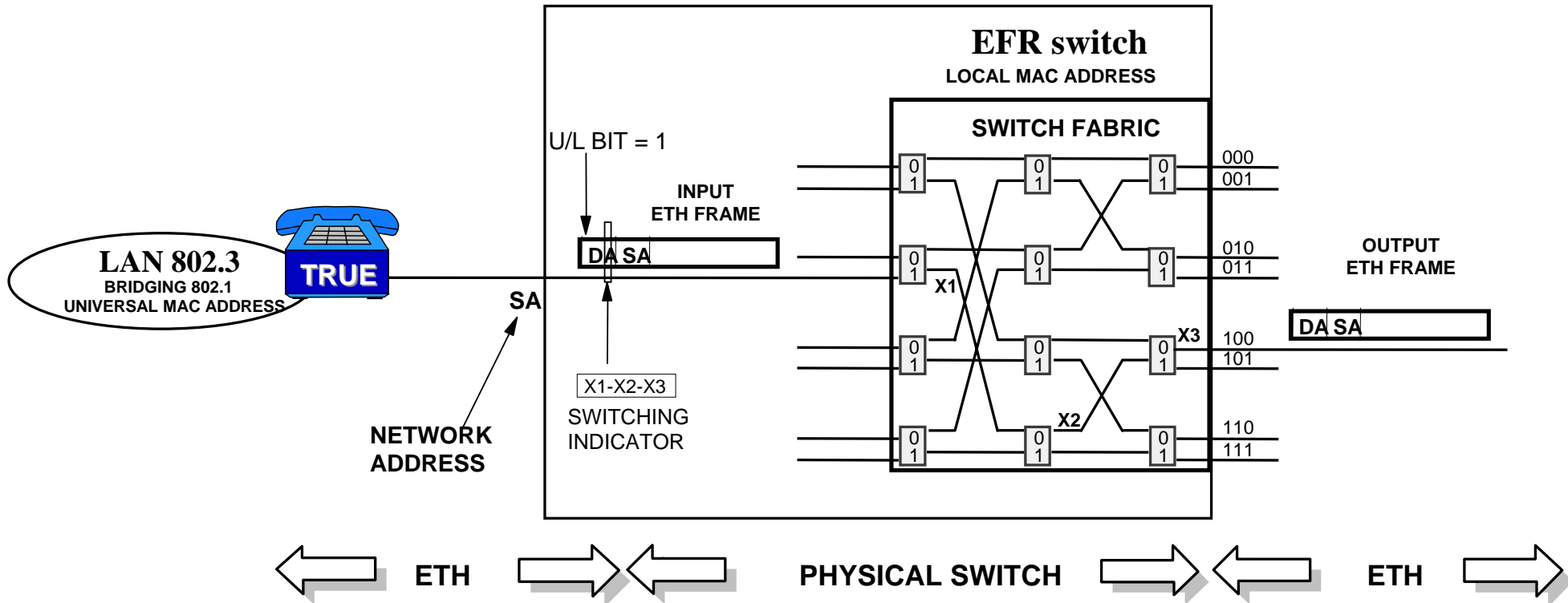
Ethernet & TCP/IP performance issues today (IEEE 802.1/802.3)

- **Effective congestion management**
 - “It is an absolute must for carrier grade systems”
 - Congestion Management implementations should be in Hardware
 - Reactions required in end stations
- **“Rate Control” performed at end-points based on congestion information provided by L2 network**
 - “Reactive”: congestion information from network nodes to the edges
 - “Proactive”: significantly reduce packet drops & buffer requirements
- **Reduce end-to-end latency and latency jitter**
- **Reference solutions today: Layer 2 operation**
 - Infiniband:® fabric of choice for clustering - IPC
 - Fibre Channel: dominant SAN technology
- **TCP/IP server performance limits**

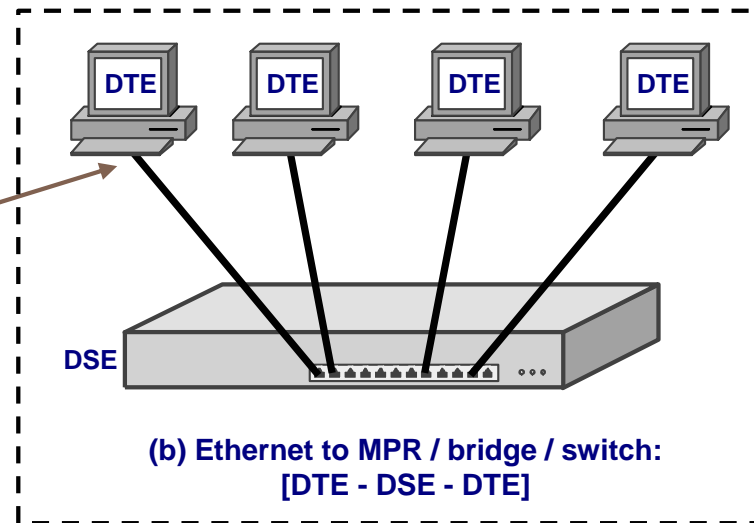
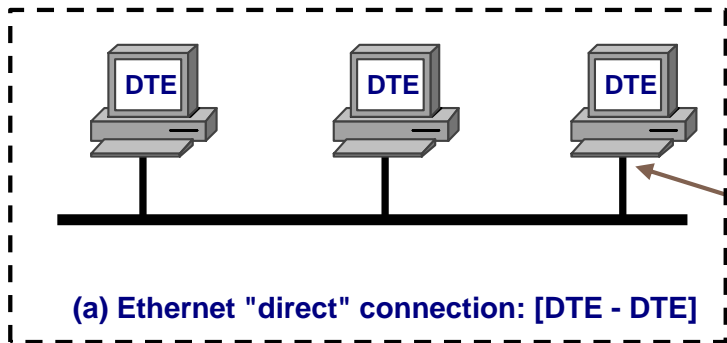
Ethernet Fabric Routing (EFR) with LLC/802.2: Advantages

- **802.3 Standard:** preserve the MAC/PLS service interfaces
- **Scalability:** no forwarding tables, no bridging/routing protocols
- **Inherent security:** address fixed by the node, not modifiable
- **Very high end-to-end throughput:** L1 switching, L2 operation
- **LLC/HDLC:** proven and well known advantages
 - Performance: bit oriented and designed to operate in hardware
 - Layer 2 operation (instead of TCP's layer 4)
 - L2 Congestion Information does not need to be passed on to ULPs
 - Flow Control mechanisms built in
- **Opportunity to apply Ethernet to:**
 - Carrier Grade, Clustering, IPC, SAN/NAS
 - iSCSI over LLC/ETH: FCH performance, much more cost effective

EFR: Addressing and Switching principle

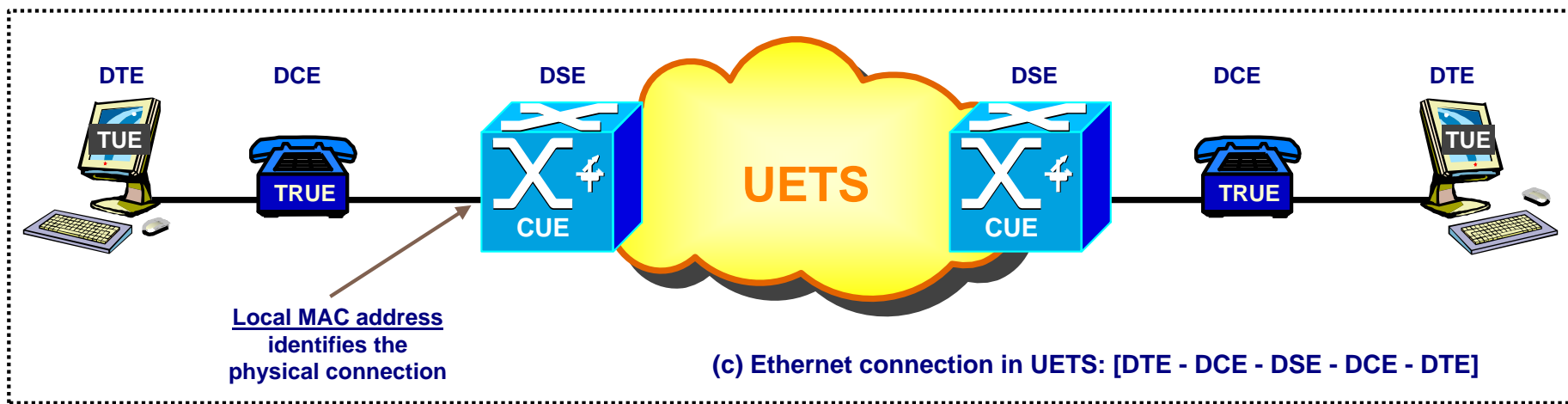


IEEE 802.3 Ethernet: IEEE 802.1 vs UETS Architectures



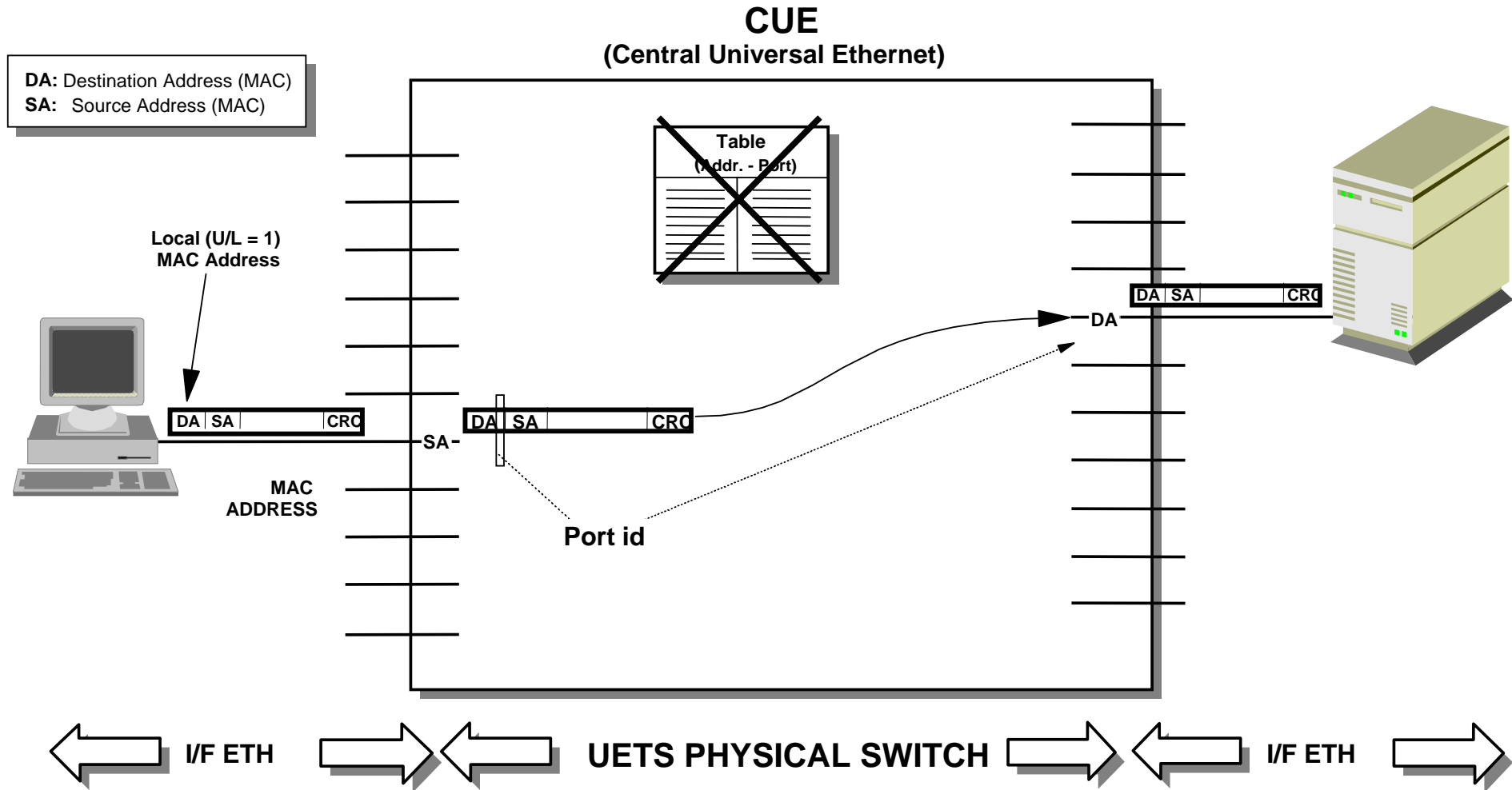
MAC address identifies the DTE

DTE - Data Terminal Equipment
 DCE - Data Communications Equipment
 DSE - Data Switching Equipment

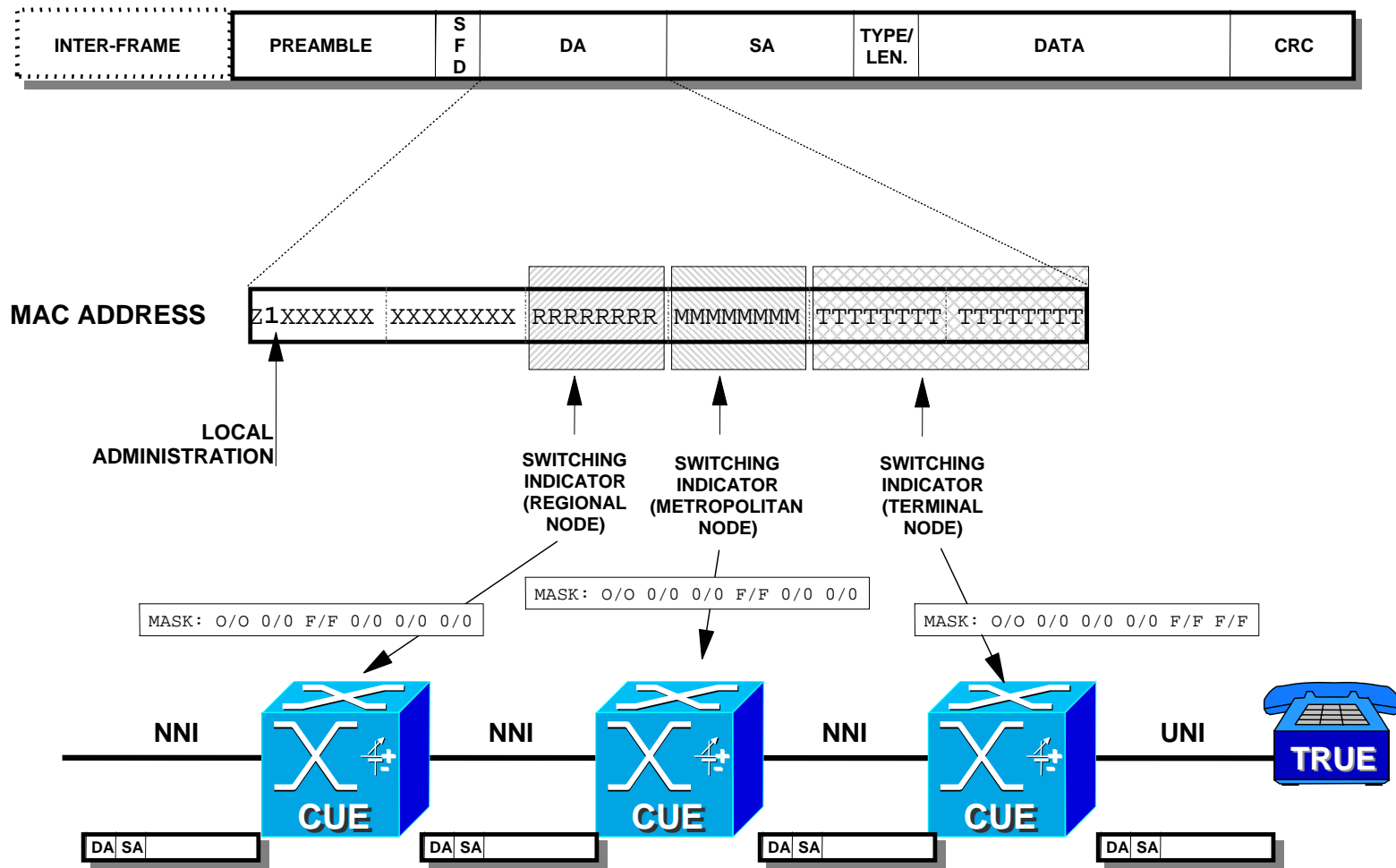


Local MAC address identifies the physical connection

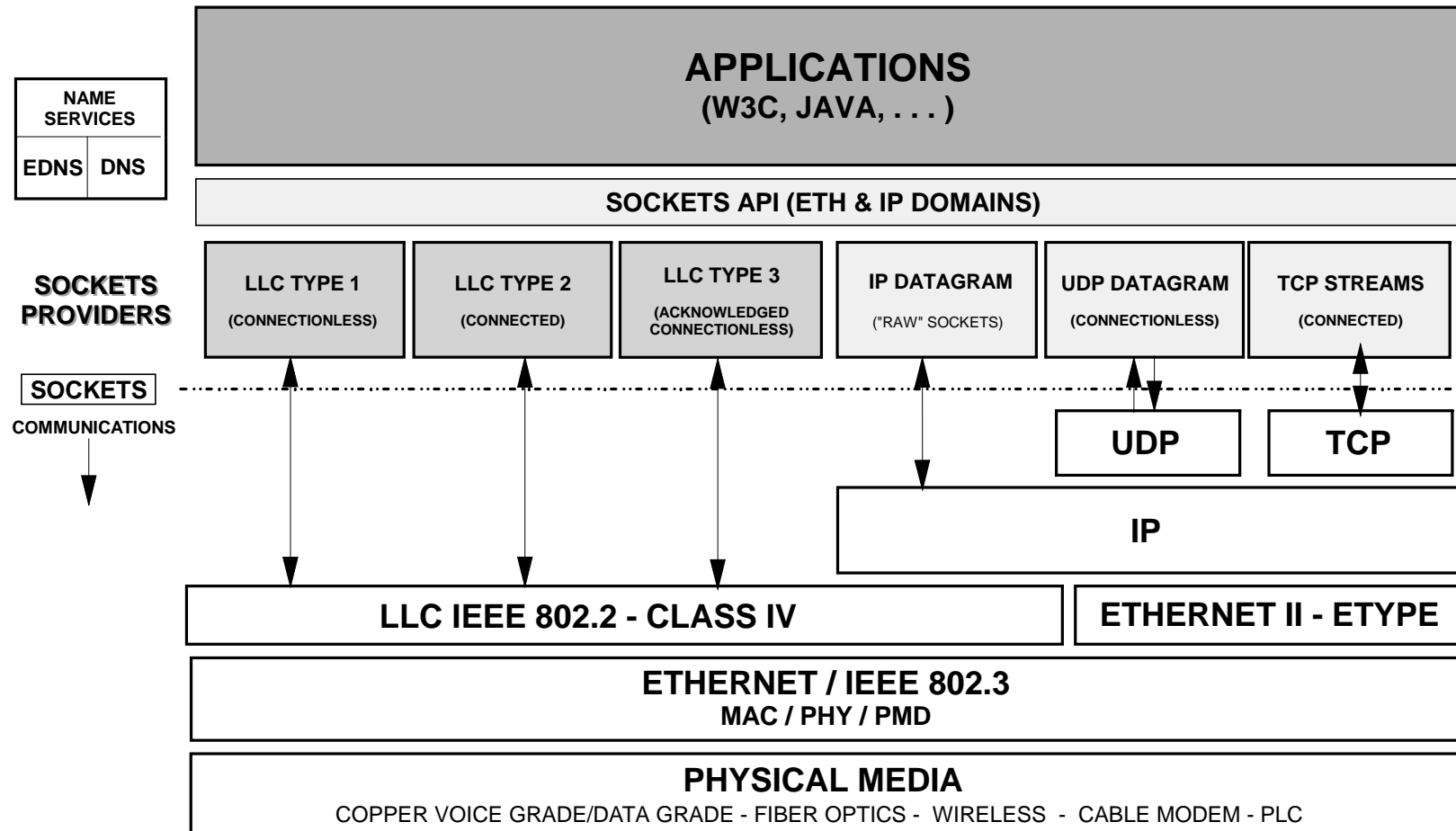
CUE implementation from bridge, L3 switch or IP Router fabric



Network scalability: hierarchical addresses

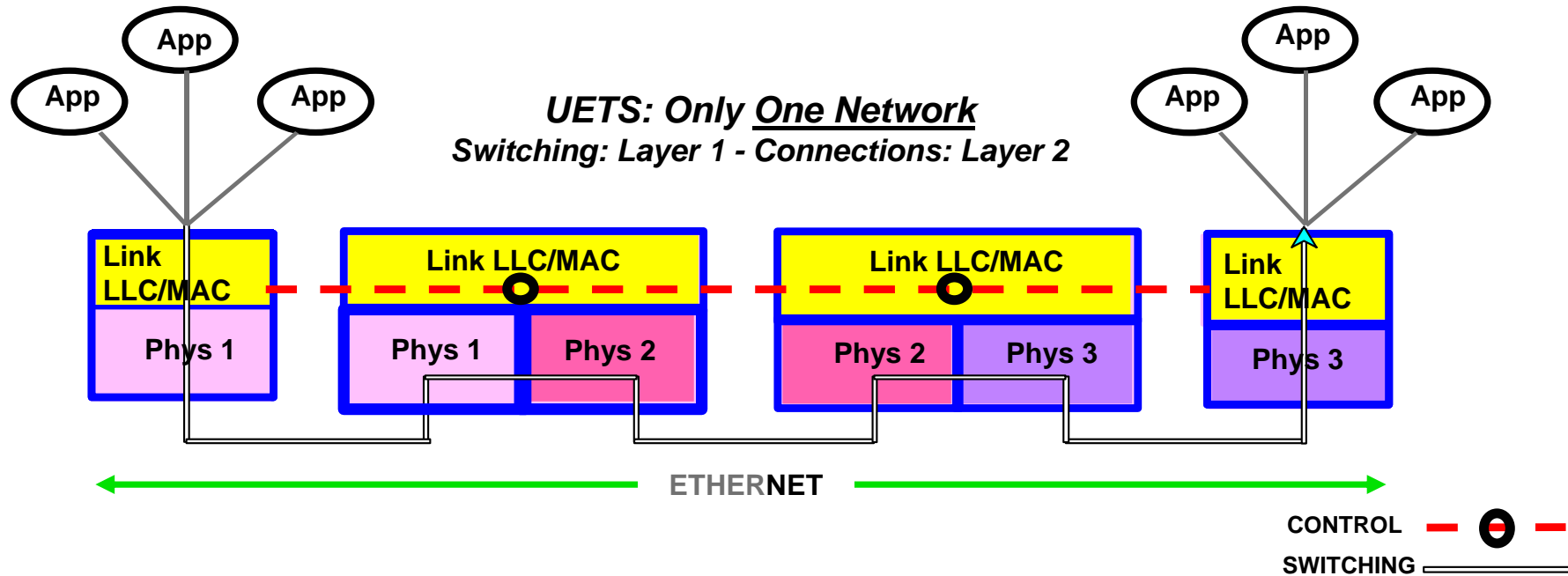


UETS/EFR Reference Model: double stack LLC and TCP/IP



The UETS system reduces the complexity of the Network to an astonishing simplicity. Only two protocols (LLC/ETH) put together in different patterns make, essentially, everything: Ethernet / 802.3 transports the information, and LLC / 802.2 performs the control.

UETS: the "Thinnest Waist" of the Earth's Internet



- **LLC-1:** "PREMIUM" services (rt-CBR / rt-VBR)
- **LLC-2:** "MISSION CRITICAL" services (nrt-VBR)
- **LLC-3:** "Stop and Go" services (HDX, BSC type)
- **ETYPE:** "BEST EFFORT" services (TCP/IP)

CONCLUSIONS

• Features

- Dual Stack: full compatibility and interoperability with TCP/IP
- Scalability from minimum UETS domain sizes to world size
- High Performance: direct hardware switching, minimum latency
- Multipoint to Multipoint Datagrams Network
- Inherent Layer Two Security
- Table-free, no label swapping, no spanning tree limitations
- Power management for energy saving

• Applications

- LAN/MAN/WAN, Enterprise, ISP, Transport Networks
- Computer on Net, HDTV Distribution, Home Networking, Grid
- High Performance Computing, SAN/NAS, Network Of Workstations
- Secure networks, multipoint-to-multipoint L2VPNs

UETS page available on-line at:

`http://www.LMdata.es/uets.htm`