Leon Bruckman

Dynamic Host Configuration Protocol for IPv6 - DHCPv6

- DHCPv6 enables DHCP servers to pass configuration parameters such as IPv6 network addresses to IPv6 nodes.
  - It offers the capability of automatic allocation of reusable network addresses and additional configuration flexibility.
- This protocol is a stateful counterpart to Stateless Address Autoconfiguration
  - It can be used separately or concurrently with the latter to obtain configuration parameters.
- Clients and servers exchange DHCP messages using UDP.
- The client uses a link-local address or addresses determined through other mechanisms for transmitting and receiving DHCP messages.
- DHCP servers receive messages from clients using a reserved, link-scoped multicast address.
  - A DHCP client transmits most messages to this reserved multicast address, so that the client need not be configured with the address or addresses of DHCP servers.
- To allow a DHCP client to send a message to a DHCP server that is not attached to the same link, a DHCP relay agent on the client's link will relay messages between the client and server.
  - The operation of the relay agent is transparent to the client.

Leon Bruckman

How is DHCPv6 different from DHCP in IPv4?

- No baggage.
  - DHCP is based on an earlier protocol called BOOTP. This packet layout is wasteful in a lot of cases.
  - A lot of the options turn out to be not useful, or not as useful as they can be, but it is hard to change a protocol with such a large installed base.
  - There are a lot of "tweaks" that implementations need in order to be compatible with the buggy clients.
  - DHCPv6 leaves all this behind.
- Two features of IPv6 greatly improve DHCPv6:
  - IPv6 hosts have "link local addresses". IPv6 hosts can use this to send requests for "real" addresses. IPv4 hosts have to use system specific hacks to work before they have an address.
  - All IPv6 systems support multicasting. All DHCPv6 servers register that they want to receive DHCPv6 multicast packets. This means the network knows where to send them. In IPv4, clients broadcast their requests, and networks do not know how far to send them.
- One exchange configures all interfaces.
  - A single DHCPv6 request may includes all interfaces on a client.
- DHCPv6 allows normal address allocation, as well as temporary address allocation.

Leon Bruckman

Client-server Exchanges Involving Two Messages

- When a DHCP client **does not need** to have a DHCP server assign it IP addresses, the client can obtain configuration information such as a list of available DNS servers or NTP servers through a single message and reply exchanged with a DHCP server.
  - To obtain configuration information the client first sends an **Information-Request** message to the All_DHCP_Relay_Agents_and_Servers multicast address (FF02::1:2).
  - Servers respond with a **Reply** message containing the configuration information for the client.
- When a server has **IPv6 addresses** and other configuration information **committed to a client:**
  - The client sends a **Solicit** message to the All_DHCP_Relay_Agents_and_Servers requesting the assignment of addresses and other configuration information.
    - This message includes an indication that the client is willing to accept an immediate Reply message from the server.
  - The server that is willing to commit the assignment of addresses to the client immediately responds with a **Reply** message.
    - The configuration information and the addresses in the Reply message are then immediately available for use by the client.

135

Leon Bruckman

Client-server Exchanges Involving Four Messages

- To request the assignment of one or more IPv6 addresses, a client first locates a DHCP server and then requests the assignment of addresses and other configuration information from the server.
- The client sends a **Solicit** message to the All_DHCP_Relay_Agents_and_Servers address to find available DHCP servers.
- Any server that can meet the client's requirements responds with an **Advertise** message.
- The client then chooses one of the servers and sends a **Request** message to the server asking for confirmed assignment of addresses and other configuration information.
- The server responds with a **Reply** message that contains the confirmed addresses and configuration.

136

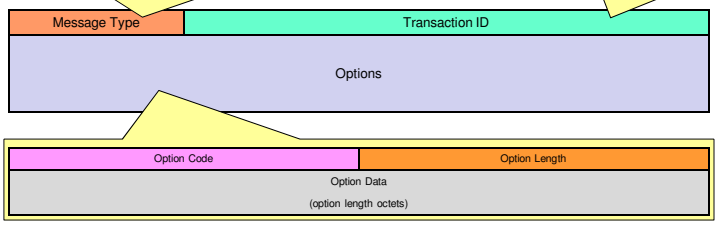
Leon Bruckman

Client/Server Message Formats

- Clients listen for DHCP messages on UDP port 546. Servers and relay agents listen for DHCP messages on UDP port 547.

SOLICIT	A client sends a Solicit message to locate a Server
ADVERTISE	A server sends an Advertise message to indicate that it is available for DHCP service, in response to a Solicit message received from a client.
REQUEST	A client sends a Request message to request configuration parameters, including IP addresses, from a specific server.
CONFIRM	A client sends a Confirm message to any available server to determine whether the addresses it was assigned are still appropriate to the link to which the client is connected.
RENEW	A client sends a Renew message to the server that originally provided the client's addresses and configuration parameters to extend the lifetimes on the addresses assigned to the client and to update other configuration parameters.
REBIND	A client sends a Rebind message to any available server to extend the lifetimes on the addresses assigned to the client and to update other configuration parameters; this message is sent after a client receives no response to a Renew message.
REPLAY	A server sends a Reply message containing assigned addresses and configuration parameters in response to a Solicit, Request, Renew, Rebind message received from a client. A server sends a Reply message containing configuration parameters in response to an Information-request message. A server sends a Reply message in response to a Confirm message confirming or denying that the addresses assigned to the client are appropriate to the link to which the client is connected. A server sends a Reply message to acknowledge receipt of a Release or Decline message.
RELEASE	A client sends a Release message to the server that assigned addresses to the client to indicate that the client will no longer use one or more of the assigned addresses.
DECLINE	A client sends a Decline message to a server to indicate that the client has determined that one or more addresses assigned by the server are already in use on the link to which the client is connected.
RECONFIGURE	A server sends a Reconfigure message to a client to inform the client that the server has new or updated configuration parameters, and that the client is to initiate a Renew/Reply or Information-request/Reply transaction with the server in order to receive the updated information.
INFORMATION-REQUEST	A client sends an Information-request message to a server to request configuration parameters without the assignment of any IP addresses to the client.

The transaction ID for this message exchange.



137

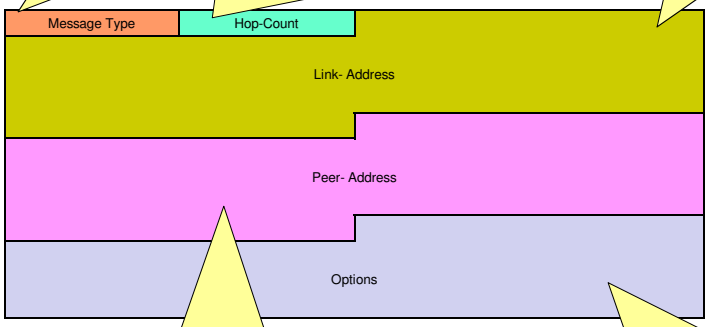
Leon Bruckman

Relay Agent/Server Message Formats

RELAY-FORW	A relay agent sends a Relay-forward message to relay messages to servers, either directly or through another relay agent. The received message, either a client message or a Relay-forward message from another relay agent, is encapsulated in an option in the Relay-forward message.
RELAY-REPL	A server sends a Relay-reply message to a relay agent containing a message that the relay agent delivers to a client. The Relay-reply message may be relayed by other relay agents for delivery to the destination relay agent. The server encapsulates the client message as an option in the Relay-reply message, which the relay agent extracts and relays to the client.

RELAY-FORW: A global or site-local address that will be used by the server to identify the link on which the client is located.
RELAY-REPL: Copied from the Relay-forward message

RELAY-FORW: Number of relay agents that have relayed this message.
RELAY-REPL: Copied from the Relay-forward message



RELAY-FORW: The address of the client or relay agent from which the message to be relayed was received.
RELAY-REPL: Copied from the Relay-forward message

Must include the Relay Message option that includes the message to be relayed

138

Leon Bruckman

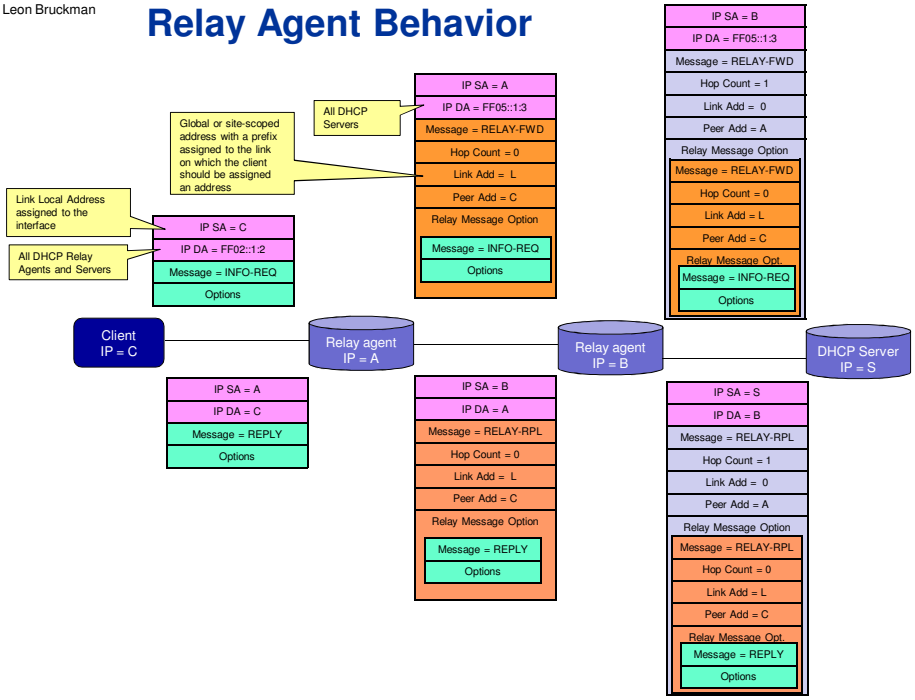
DHCP Unique Identifier (DUID)

- Each DHCP client and server has a DUID.
  - **DHCP servers** use DUIDs to identify clients for the selection of configuration parameters and in the association of IAs with clients.
    - An "identity-association" (IA) is a construct through which a server and a client can identify, group, and manage a set of related IPv6 addresses.
    - Each IA consists of an IAID and associated configuration information.
  - **DHCP clients** use DUIDs to identify a server in messages where a server needs to be identified.
- A DUID consists of a two-octet type code represented in network byte order, followed by a variable number of octets that make up the actual identifier.
  - A DUID can be no more than 128 octets long (not including the type code).
  - The following types are currently defined:
    - Link-layer address plus time
    - Vendor-assigned unique ID based on Enterprise Number
    - Link-layer address
  - The motivation for having more than one type of DUID is that the DUID must be globally unique, and must also be easy to generate.
    - The sort of globally-unique identifier that is easy to generate for any given device can differ quite widely.

139

Leon Bruckman

Relay Agent Behavior



140

Leon Bruckman

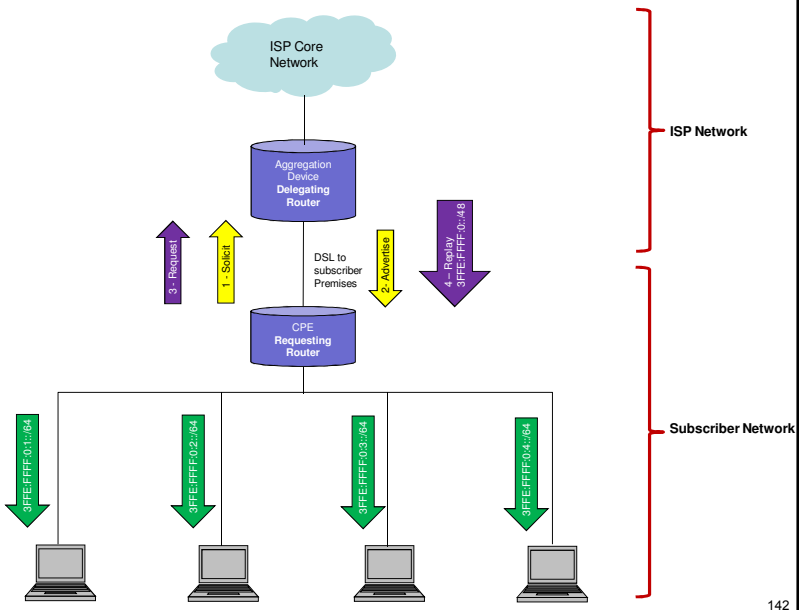
Prefix Delegation option

- The prefix delegation mechanism is intended for simple delegation of prefixes from a delegating router to requesting routers.
- It is appropriate for situations in which the delegating router does not have knowledge about the topology of the networks to which the requesting router is attached, and the delegating router does not require other information aside from the identity of the requesting router to choose a prefix for delegation.
  - For example, these options would be used by a service provider to assign a prefix to a Customer Premise Equipment (CPE) device acting as a router between the subscriber's internal network and the service provider's core network.
- **Requesting router:** The router that acts as a DHCP client and is requesting prefix(es) to be assigned.
- **Delegating router:** The router that acts as a DHCP server, and is responding to the prefix request.

141

Leon Bruckman

DHCPv6 Prefix Delegation example



142

Leon Bruckman

DNS Extensions to Support IPv6

- Current support for the storage of Internet addresses in the Domain Name System (DNS) cannot easily be extended to support IPv6 addresses since applications assume that address queries return 32-bit IPv4 addresses only.
- To support the storage of IPv6 addresses in the DNS the following extensions are defined:
  - A resource record type to map a domain name to an IPv6 address.
  - A domain to support lookups based on address.
  - Existing queries that perform additional section processing to locate IPv4 addresses are redefined to perform additional section processing on both IPv4 and IPv6 addresses.
- The IP protocol version used for querying resource records is independent of the protocol version of the resource records; e.g., IPv4 transport can be used to query IPv6 records and vice versa.

Leon Bruckman

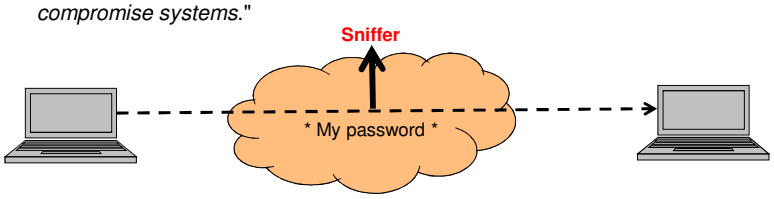
Multicast Listener Discovery (MLD) for IPv6

- MLD is derived from version 2 of IPv4's Internet Group Management Protocol, IGMPv2.
- MLDv2 is a translation of the IGMPv3 protocol for IPv6 semantics.
- The MLDv2 protocol, when compared to MLDv1, adds support for "source filtering", i.e., the ability for a node to report interest in listening to packets *only* from specific source addresses, as required to support Source-Specific Multicast, or from *all but* specific source addresses, sent to a particular multicast address.
  - MLDv2 is designed to be interoperable with MLDv1.
- One important difference to note is that MLD uses ICMPv6 (IP Protocol 58) message types, rather than IGMP (IP Protocol 2) message types.
- **Please refer to RFC 3810 for details**

Leon Bruckman

Why Do We Need IPSec?

- The Internet provides amazing opportunities, but not without some risk. Without the proper controls, your data is subject to several types of attacks.
- **Loss of Privacy**
  - A perpetrator may observe confidential data as it traverses the Internet. This ability is probably the largest inhibitor of business-to-business communications today. Without encryption, every message sent may be read by an unauthorized party.
  - The Computer Emergency Response Team Coordination Center (CERT CC) in its 1996 annual report: "*Intruders continued to install packet sniffers on root-compromised systems. These sniffers, used to collect account names and passwords, were frequently installed as part of a widely available kit that also replaced common system files with Trojan horse programs. These kits provided `cookbook' directions that even novice, unskilled intruders could use to compromise systems.*"

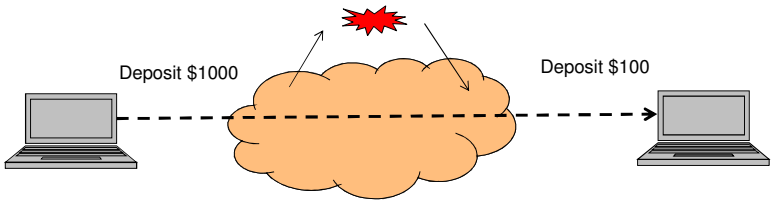


145

Leon Bruckman

Loss of Data Integrity

- Even for data that is not confidential, one must still take measures to ensure data integrity.
- For example, you may not care if anyone sees your routine business transaction, but you would certainly care if the transaction were modified.
- If you were able to securely identify yourself to your bank using digital certificates, you would still want to ensure that the transaction itself is not modified in some way, such as changing the amount of the deposit

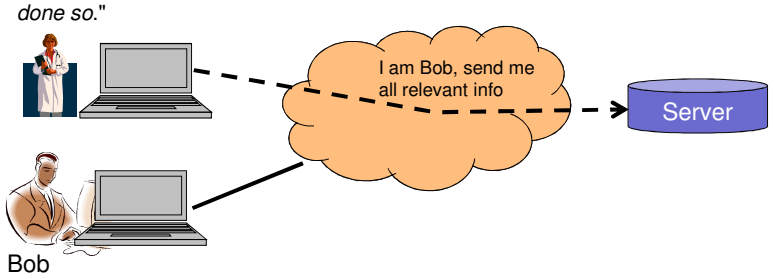


146

Leon Bruckman

Identity Spoofing

- Moving beyond the protection of data itself, you must also be careful to protect your identity on the Internet.
  - A crafty intruder may be able to impersonate you and have access to confidential information.
  - Many security systems today rely on IP addresses to uniquely identify users. Unfortunately this system is quite easy to fool.
  - This was another vulnerability that CERT-CC cited in its 1996 annual report, saying: "*We continued to receive several reports each week of IP spoofing attacks. Intruders attacked by using automated tools that are becoming widespread on the Internet. Some sites incorrectly believed that they were blocking such spoofed packets, and others planned to block them but had not yet done so.*"

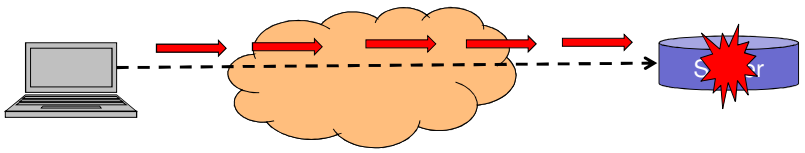


147

Leon Bruckman

Denial-of-service

- As organizations take advantage of the Internet, they must take measures to ensure that their systems are available. Over the last several years attackers have found deficiencies in the TCP/IP protocol suite that allows them to arbitrarily cause computer systems to crash.
- The CERT CC reported that: "*Instructions for executing denial-of-service attacks and programs for implementing such attacks were widely distributed this year. After this information was published, we noticed a significant and rapid increase in the number of denial-of-service attacks executed against sites.*"



148

Leon Bruckman

Public-key cryptography

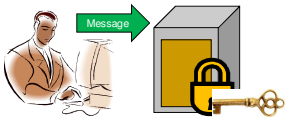
- Public-key cryptography is a cryptographic approach which involves the use of asymmetric key algorithms instead of or in addition to symmetric key algorithms.
  - Unlike symmetric key algorithms, it does not require a secure initial exchange of one or more secret keys to both sender and receiver.
- The asymmetric key algorithms are used to create a mathematically related key pair: a secret private key and a published public key.
  - Use of these keys allows protection of the authenticity of a message by creating a digital signature of a message using the private key, which can be verified using the public key.
    - An analogy for digital signatures is the sealing of an envelope with a personal wax seal. The message can be opened by anyone, but the presence of the seal authenticates the sender.
  - It also allows protection of the confidentiality and integrity of a message, by public key encryption, encrypting the message using the public key, which can only be decrypted using the private key.
    - An analogy to public-key encryption is that of a locked mailbox with a mail slot. The mail slot is exposed and accessible to the public. Anyone knowing the street address can go to the door and drop a written message through the slot; however, only the person who possesses the key can open the mailbox and read the message.

149

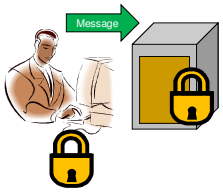
Leon Bruckman

A postal analogy

Symmetric Key System



Asymmetric Key System



150

Leon Bruckman

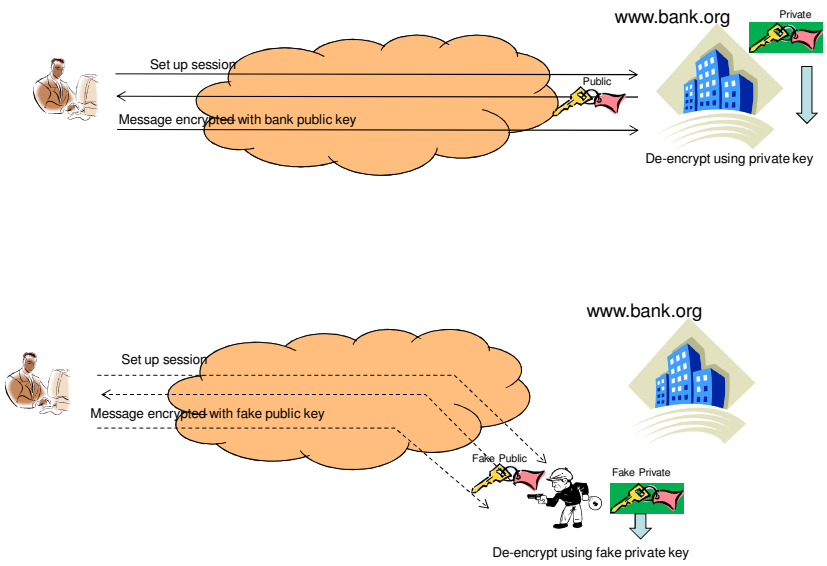
Certificate authority - CA

- A certificate authority or certification authority (CA) is an entity that issues digital certificates.
- The digital certificate certifies the ownership of a public key by the named subject of the certificate.
  - This allows others (relying parties) to rely upon signatures or assertions made by the private key that corresponds to the public key that is certified.
  - In this model of trust relationships, a CA is a trusted third party that is trusted by both the subject (owner) of the certificate and the party relying upon the certificate.
- CA's obligation is to verify an applicant's credentials, so that users and relying parties can trust the information in the CA's certificates.
  - In essence, the Certificate Authority is responsible for saying "yes, this person is who they say they are, and we, the CA, verify that".
- Commercial CAs often use a combination of authentication techniques including leveraging government bureaus, the payment infrastructure, third parties' databases and services, and custom heuristics.
  - In some enterprise systems, local forms of authentication such as **Kerberos** can be used to obtain a certificate.

151

Leon Bruckman

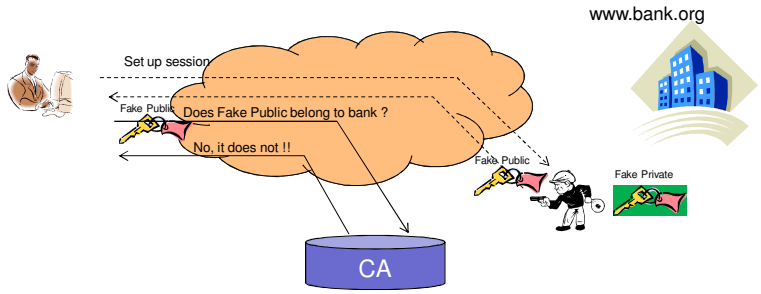
Communication hi-jacking



152

Leon Bruckman

Identity verification



- Since www.bank.org uses a public key that the certification authority certifies therefore a fake www.bank.org can only use the same public key.
- Since the fake www.bank.org does not know the corresponding private key, it cannot decrypt the user's answer.

153

Leon Bruckman

Diffie–Hellman key exchange

- Diffie–Hellman key exchange is a cryptographic protocol that allows two parties that have no prior knowledge of each other to jointly establish a shared secret key over an insecure communications channel.
  - This key can then be used to encrypt subsequent communications using a symmetric key cipher.
- Example with non-secret values in **green**, and secret values in **red**:
  - Alice and Bob agree to use a prime number **p=23** and base **g=5**.
  - Alice chooses a secret integer **a=6**, then sends Bob **A** = g^a mod **p**
    - **A** = 5^6 mod **23** = **8**.
  - Bob chooses a secret integer **b=15**, then sends Alice **B** = g^b mod **p**
    - **B** = 5^{15} mod **23** = **19**.
  - Alice computes **s** = B^a mod **p**
    - 19^6 mod **23** = **2**.
  - Bob computes s = A^b mod **p**
    - 8^{15} mod **23** = **2**.
- Much larger values of a, b, and p would be needed to make this example secure, since it is easy to try all the possible values of g^{ab} mod 23.

154

Leon Bruckman

IPsec general

- Internet Protocol Security (IPsec) is a protocol suite for securing Internet Protocol (IP) communications by authenticating and encrypting each IP packet of a communication session.
  - IPsec also includes protocols for establishing mutual authentication between agents at the beginning of the session and negotiation of cryptographic keys to be used during the session.
- IPsec is an end-to-end security scheme operating in the Internet Layer of the Internet Protocol Suite.
  - It can be used in protecting data flows between a pair of hosts, a pair of security gateways or between a security gateway and a host.
- Some other Internet security systems in widespread use, such as Secure Sockets Layer (**SSL**), Transport Layer Security (**TLS**) and Secure Shell (**SSH**), operate in the upper layers of the TCP/IP model.
  - Hence, IPsec protects any application traffic across an IP network. Applications do not need to be specifically designed to use IPsec.
  - The use of TLS/SSL, on the other hand, must be designed into an application to protect the application protocol.

155

Leon Bruckman

IPsec architecture

IPsec uses the following protocols to perform various functions:

- Internet Key Exchange (**IKE** and **IKEv2**) or Kerberized Internet Negotiation of Keys (**KINK**) sets up a security association (**SA**) by handling the negotiation of protocols and algorithms and by generating the encryption and authentication keys to be used by IPsec.
- Authentication Header (**AH**) to provide connectionless integrity and data origin authentication for IP datagrams and to provide protection against replay attacks.
- Encapsulating Security Payload (**ESP**) to provide confidentiality, data origin authentication, connectionless integrity, an anti-replay service (a form of partial sequence integrity), and limited traffic flow confidentiality.

156

Leon Bruckman

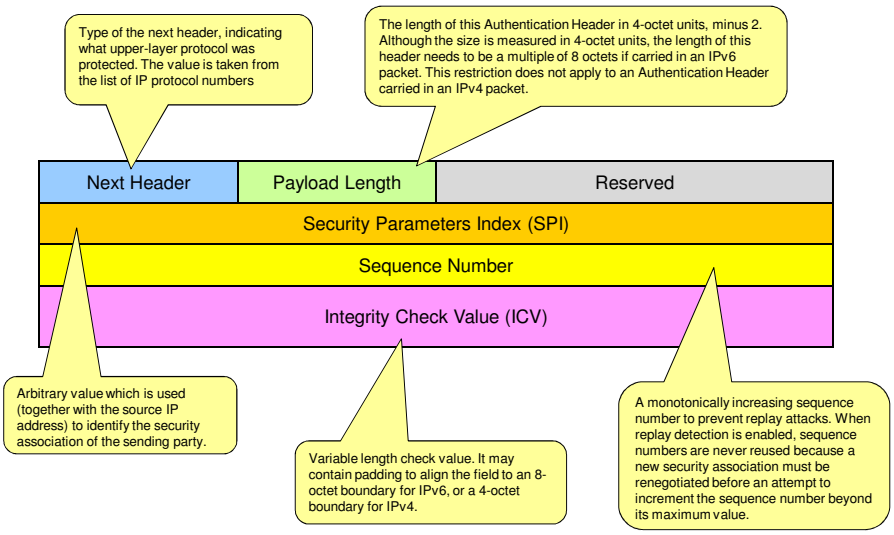
Authentication Header (AH)

- AH guarantees connectionless integrity and data origin authentication of IP packets.
  - Further, it can optionally protect against replay attacks by using the sliding window technique and discarding old packets.
- In IPv4, the AH protects the IP payload and all header fields of an IP datagram except for mutable fields (i.e. those that might be altered in transit).
  - Mutable (and therefore unauthenticated) IP header fields are DSCP/TOS, ECN, Flags, Fragment Offset, TTL and Header Checksum.
- In IPv6, the AH protects the AH itself, the Destination Options extension header after the AH, and the IP payload.
  - It also protects the fixed IPv6 header and all extension headers before the AH, except for the mutable fields: DSCP, ECN, Flow Label, and Hop Limit.
- AH operates directly on top of IP, using IP protocol number 51.

157

Leon Bruckman

Authentication Header format



158

Leon Bruckman

Modes of operation

Transport mode

- In transport mode, only the payload (the data you transfer) of the IP packet is encrypted and/or authenticated.
- The routing is intact, since the IP header is neither modified nor encrypted; however, when the authentication header is used, the IP addresses cannot be translated, as this will invalidate the hash value.
  - The transport and application layers are always secured by hash, so they cannot be modified in any way (for example by translating the port numbers). A means to encapsulate IPsec messages for NAT traversal has been defined.
  - Transport mode is used for host-to-host communications.

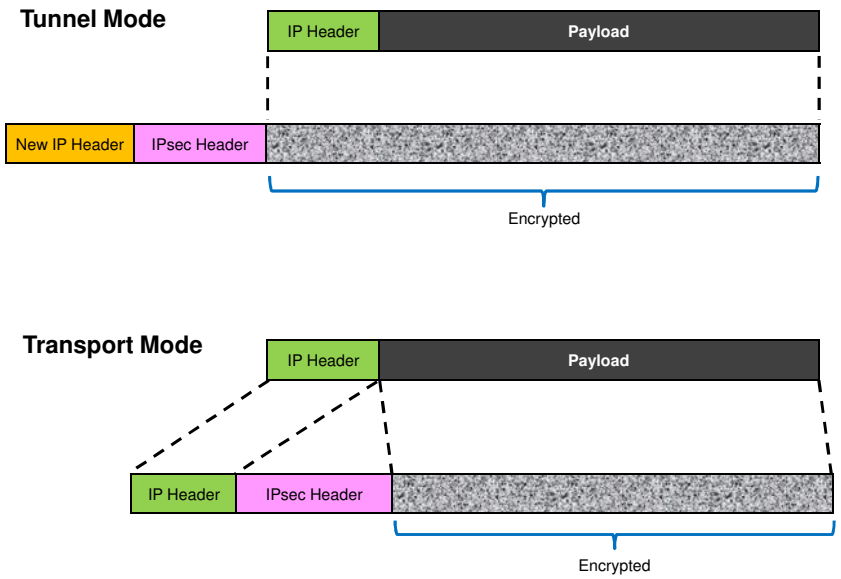
Tunnel mode

- In tunnel mode, the entire IP packet is encrypted and/or authenticated. It is then encapsulated into a new IP packet with a new IP header.
  - Tunnel mode is used to create Virtual Private Networks.
  - Tunnel mode supports NAT traversal.

159

Leon Bruckman

Tunnel and transport modes

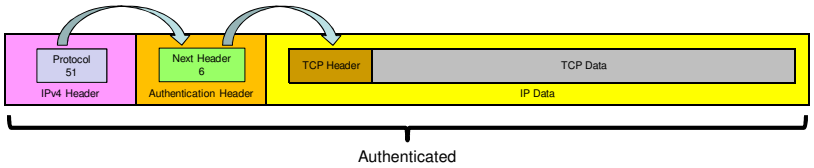


160

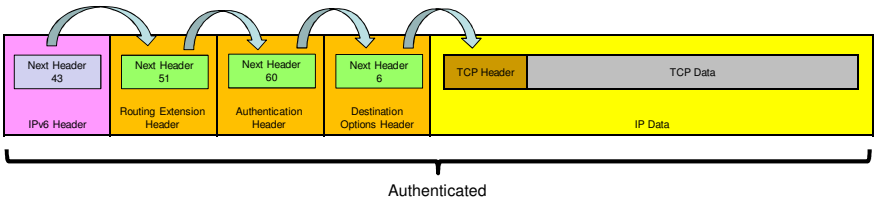
Leon Bruckman

AH Datagrams – Transport mode

IPv4



IPv6

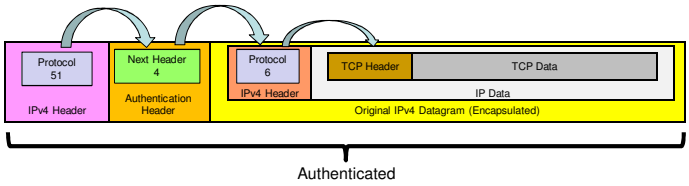


161

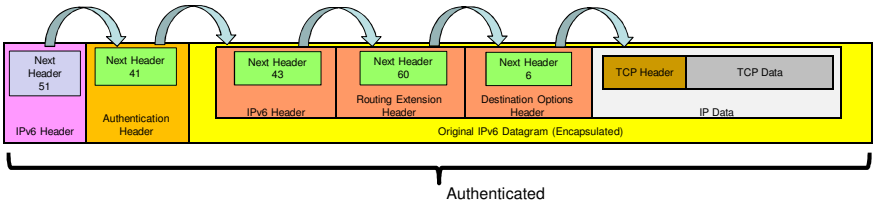
Leon Bruckman

AH Datagrams – Tunnel Mode

IPv4



IPv6



162

Leon Bruckman

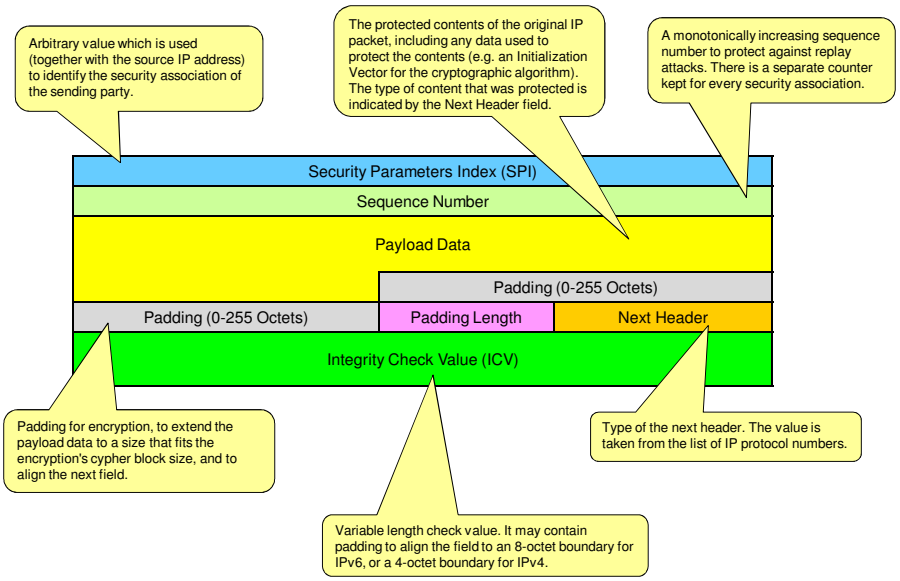
Encapsulating Security Payload

- Encapsulating Security Payload (ESP) in IPsec it provides origin authenticity, integrity, and confidentiality protection of packets.
- ESP also supports encryption-only and authentication-only configurations.
  - Using encryption without authentication is strongly discouraged because it is insecure.
- Unlike Authentication Header (AH), ESP does not protect the IP packet header.
  - However, in Tunnel Mode, where the entire original IP packet is encapsulated with a new packet header added, ESP protection is afforded to the whole inner IP packet (including the inner header) while the outer header remains unprotected.
- ESP operates directly on top of IP, using IP protocol number 50.

163

Leon Bruckman

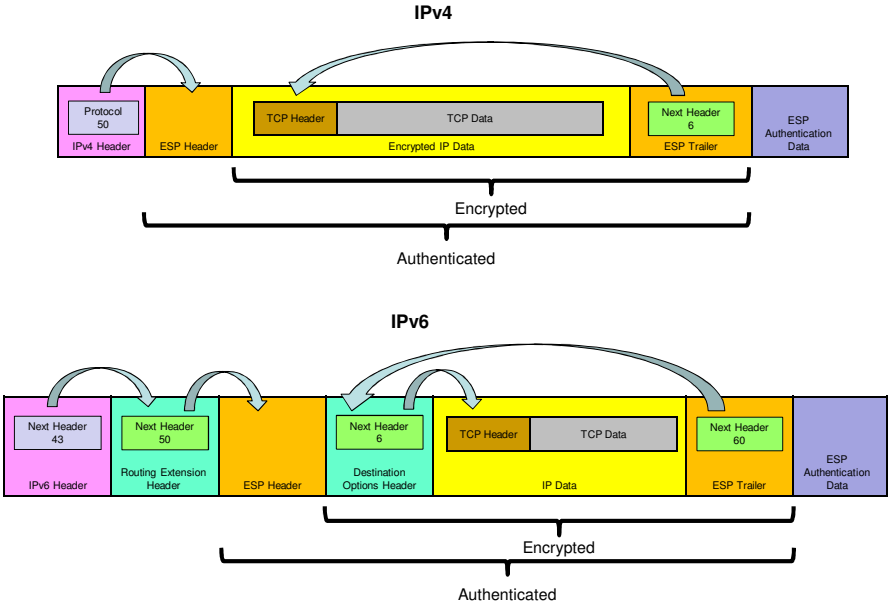
Encapsulating Security Payload format



164

Leon Bruckman

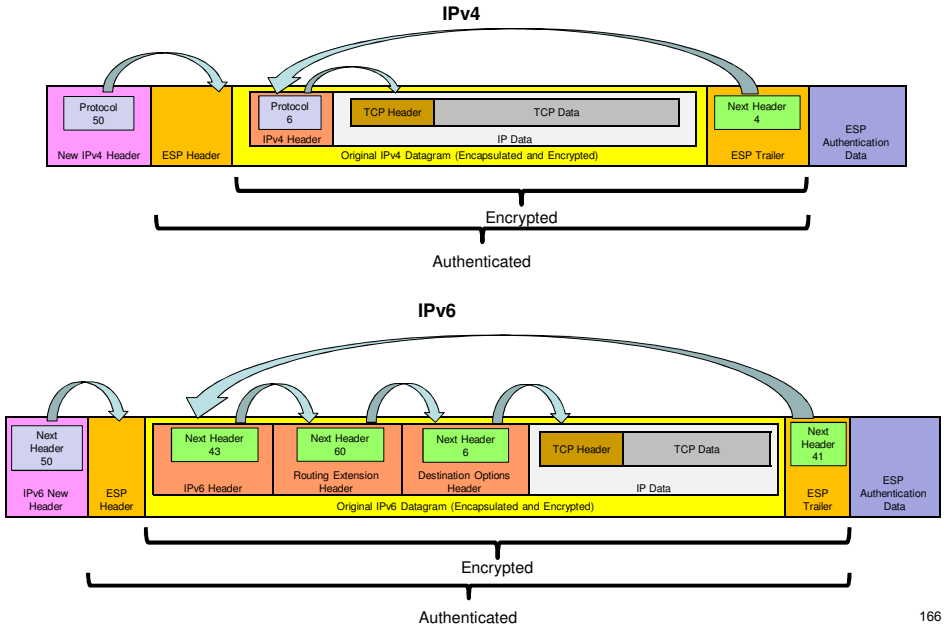
ESP Datagrams – Transport Mode



165

Leon Bruckman

ESP Datagrams – Tunnel Mode



166

Leon Bruckman

Security Association

- The IP security architecture uses the concept of a security association as the basis for building security functions into IP.
- A security association is simply the bundle of algorithms and parameters (such as keys) that is being used to encrypt and authenticate a particular flow in one direction.
  - Therefore, in normal bi-directional traffic, the flows are secured by a pair of security associations.
  - The actual choice of encryption and authentication algorithms (from a defined list) is left to the IPsec administrators.
- In order to decide what protection is to be provided for an outgoing packet, IPsec uses the Security Parameter Index (SPI), an index to the security association database (SADB), along with the destination address in a packet header, which together uniquely identify a security association for that packet.
  - A similar procedure is performed for an incoming packet, where IPsec gathers decryption and verification keys from the security association database.
- If different classes of traffic are sent on the same SA, and if the receiver is employing the optional anti-replay feature available in both AH and ESP, this could result in inappropriate discarding of lower priority packets due to the windowing mechanism used by this feature.
  - Therefore, a sender should put traffic of different classes on different SAs to support Quality of Service (QoS) appropriately.

167

Leon Bruckman

Internet Key Exchange (IKE)

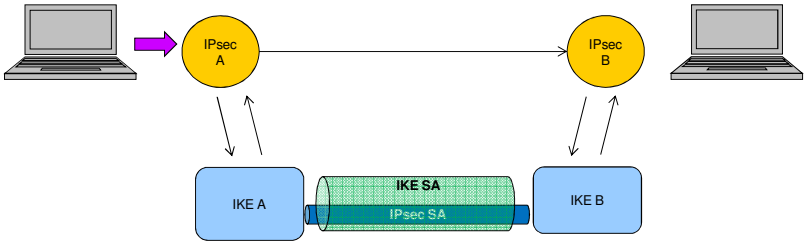
- IKE is the protocol used to set up a SA in the IPsec protocol suite.
- IKE uses a Diffie–Hellman key exchange to set up a shared session secret, from which cryptographic keys are derived.
  - Public key techniques (e.g. RSA) or, alternatively, a pre-shared key, are used to mutually authenticate the communicating parties.
- IPsec implementations consist of an IKE daemon that runs in user space and an IPsec stack in the kernel that processes the actual IP packets.
  - User-space daemons have easy access to mass storage containing information, such as the IPsec endpoint addresses, keys and certificates.
  - Kernel modules can process packets efficiently and with minimum overhead.
- The IKE protocol uses UDP packets, usually on port 500.
- The negotiated key material is then given to the IPsec stack.
  - This could be an Advanced Encryption Standard (AES) key, information identifying the IP endpoints and ports that are to be protected, as well as what type of IPsec tunnel has been created.
  - The IPsec stack, in turn, intercepts the relevant IP packets if and where appropriate and performs encryption/decryption as required.

168

Leon Bruckman

IPsec and IKE

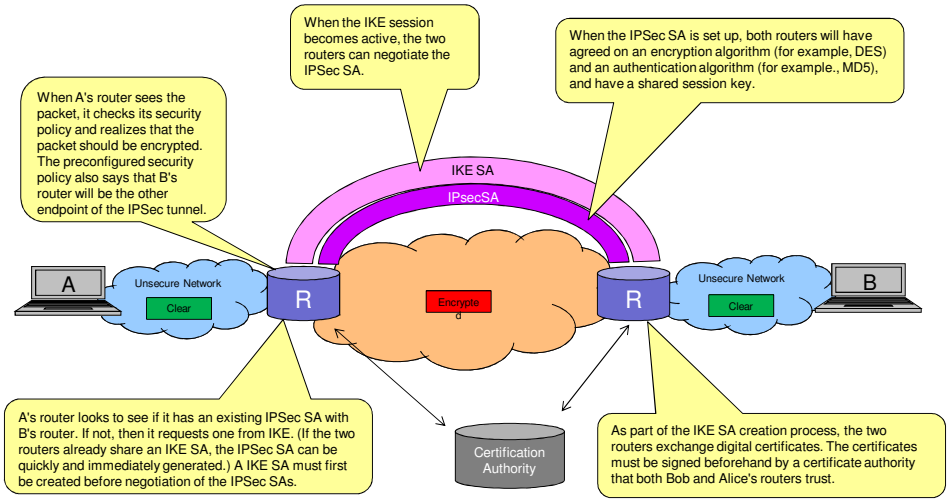
- **A**'s first packet to **B** that should be encrypted triggers the IKE process.
- The IKE process builds a secure tunnel between **B** and **A**.
- The IPSec SA is negotiated over this tunnel.
- **A** can then use this SA to send secure data to **B**.



169

Leon Bruckman

IPSec and IKE in Practice



170

Leon Bruckman

Secure Socket Layer (SSL)

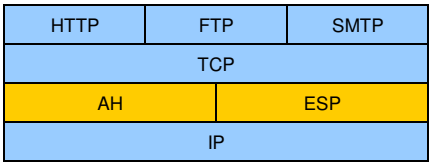
- Transport Layer Security (TLS) and its predecessor, Secure Socket Layer (SSL), are cryptographic protocols that provide security for communications over networks such as the Internet.
- TLS and SSL encrypt the segments of network connections at the Application Layer to ensure secure end-to-end transit at the Transport Layer.
- Several versions of the protocols are in widespread use in applications like web browsing, electronic mail, Internet faxing, instant messaging and voice-over-IP (VoIP).
- The TLS protocol allows client/server applications to communicate across a network in a way designed to prevent eavesdropping and tampering. TLS provides endpoint authentication and communications confidentiality over the Internet using cryptography.
- In typical end-user/browser usage, TLS authentication is unilateral: only the server is authenticated
  - TLS also supports the more secure bilateral connection mode

171

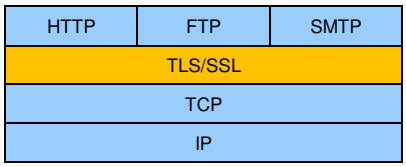
Leon Bruckman

TLS/SSL versus IPsec

Network-layer IPsec



Application-layer SSL



- Network-layer VPNs provide users with the same full, continuous access to the network as if they were physically connected.
- This is ideal for facilitating regular communications and resource sharing among users at geographically separate offices to improve productivity enterprise-wide.

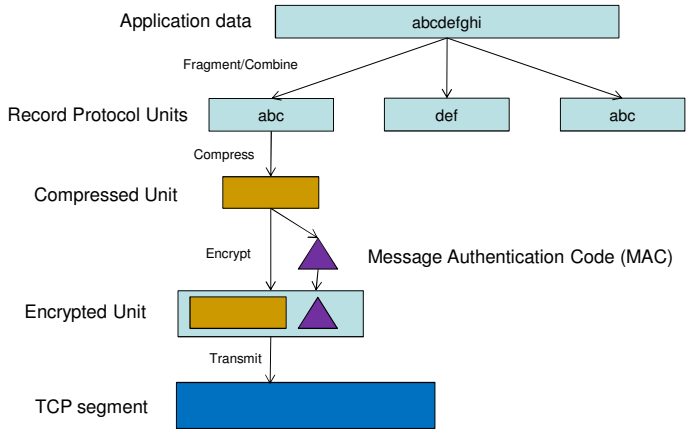
- With an SSL VPN, the connection between the mobile user and the internal resource happens via a Web connection at the application layer.
- The use of SSL is ideal for the mobile user because:
  - SSL VPN does not require client software to be preinstalled and maintained on the device being used to access corporate resources.
  - SSL VPN does not need to be configured on the endpoint machine by a user or administrator
  - SSL VPN is available from any standard Web browser, so users don't need a company laptop

172

Leon Bruckman

TLS architecture

- TLS defines Record Protocol to transfer application and TLS information
- A session is established using a Handshake Protocol



173

Leon Bruckman

TLS handshake protocol

The TLS Handshake Protocol provides connection security that has three basic properties:

- The peer's identity can be authenticated using asymmetric, or public key, cryptography.
  - This authentication can be made optional, but is generally required for at least one of the peers.
- The negotiation of a shared secret is secure: the negotiated secret is unavailable to eavesdroppers, and for any authenticated connection the secret cannot be obtained, even by an attacker who can place himself in the middle of the connection.
- The negotiation is reliable: no attacker can modify the negotiation communication without being detected by the parties to the communication.

- TLS is application protocol independent.
  - Higher-level protocols can layer on top of the TLS protocol transparently.
  - The TLS standard, however, does not specify how protocols add security with TLS; the decisions on how to initiate TLS handshaking and how to interpret the authentication certificates exchanged are left to the judgment of the designers and implementers of protocols that run on top of TLS.

174

Leon Bruckman

IPsec or SSL VPN ?

- Administrators who need to achieve high-performance, redundant site-to-site connectivity will be well served by IPsec VPN offerings.
  - IPsec VPNs were created to meet the challenge of securely providing employees around the world with the always-on connectivity.
  - IPsec VPNs provide users at geographically distributed locations an experience akin to that of logging in at the corporate headquarters, allowing them to easily access all the network resources.
  - The management resources required for deployment and maintenance are fairly limited in the site-to-site use case: the number of sites is limited; the device, which usually serves also as a firewall, is managed; and the session is fixed.
- Administrators who need to allow mobile employees, contractors, offshore employees, business partners, and/or customers access to certain corporate resources will be better served by SSL VPNs.
  - SSL VPNs are designed to address the needs of diverse audiences that need secure access to administrator-specified corporate resources from anywhere. They allow the administrator to change both access methods and the resources that can be accessed as the users' circumstances change.
  - SSL VPNs offer users the convenience of being able to access corporate resources using any Web-enabled device from anywhere with no preinstalled client software needed on the endpoint device.

175

Leon Bruckman

Scenarios For Using SSL VPN vs. IPsec VPN

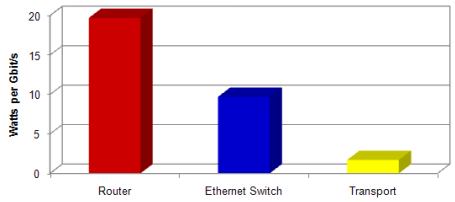
TYPE OF APPLICATION	TYPE OF ENDPOINT DEVICE	REMOTE NETWORK SECURITY	TYPE OF CONNECTION	TYPE OF VPN
Remote office/branch office	Corporate	Managed, trusted	Fixed site-to-site	IPsec
Mobile employee	Corporate or Non-corporate	Unmanaged, untrusted	Mobile	SSL VPN
Partner/customer extranet	Non-corporate	Unmanaged, untrusted	Mobile	SSL VPN
Employee remote access	Corporate or Non-corporate	Managed, trusted	Mobile	SSL VPN

176

Leon Bruckman

Optical networks trends

- Global IP traffic continues to grow rapidly driven by both consumer and business applications.
- As much as 60 percent of traffic passing through core routers is transit traffic
- Core router ports are expensive:
  - CAPEX
  - OPEX
  - Physical rack space
  - Power

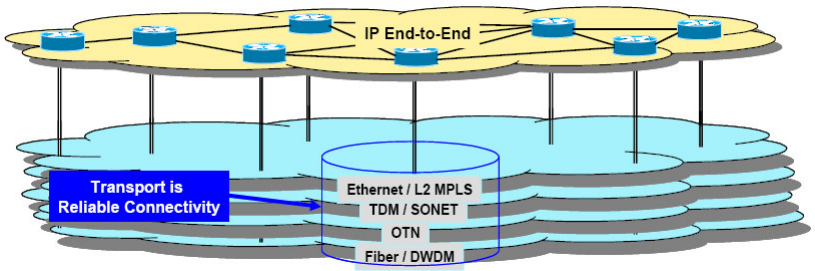


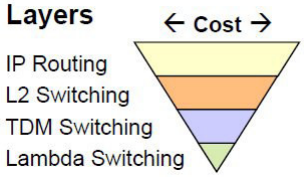
Source: Verizon 2009

177

Leon Bruckman

Switch when you can route when you must





G. Wellbrock, Verizon

178

Leon Bruckman

Ethernet Frame Format - Addresses

6	Destination MAC address
6	Source MAC address
2	EtherType (0x8100)
2	VLAN
2	EtherType
	Payload
	PAD
4	FCS

Destination MAC Address

– 48 bit IEEE Destination Node address.

Source MAC Address

– 48 bit IEEE Source Node address

OUI – Organizationally Unique Identifier

AC-DE-48-00-00-80

C A E D 8 4 0 0 0 0 0 8

lsb msb

0011 0101-0111 1011-0001 0010-0000 0000-0000 0000-0000 0001

I/G U/L

I/G = 0 – Individual Address
I/G = 1 – Group Address

U/L = 0 – Globally Administered Address
U/L = 1 – Locally Administered Address

179

Leon Bruckman

Ethernet Frame Format – VLAN tag

6	Destination MAC address
6	Source MAC address
2	EtherType (0x8100)
2	VLAN tag
2	Length/Type
	Payload
	PAD
4	FCS

PCP	CFI	VID
VID		

PCP 3 bit Priority Code Point

CFI 1 bit Canonical Format Indicator. Usually not used.

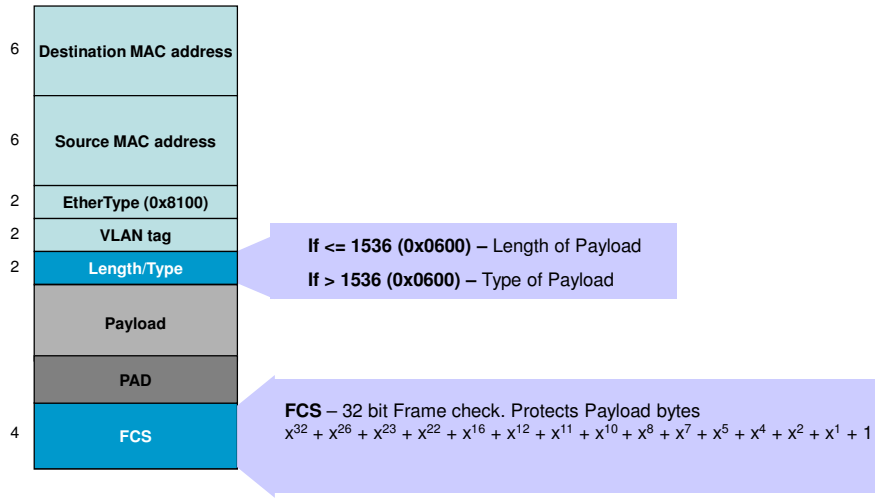
VID 12 bit VLAN Identifier.

VID	Use
0	Tag includes only priority (PCP)
1	Default asigned to untagged frames
FFF	Reserved

180

Leon Bruckman

Ethernet Frame Format – Length/Type



181

Leon Bruckman

Ethernet frame transmission

Preamble: Used to allow the physical layer circuitry to reach its steady-state synchronization with the received frame's timing – (101010…).

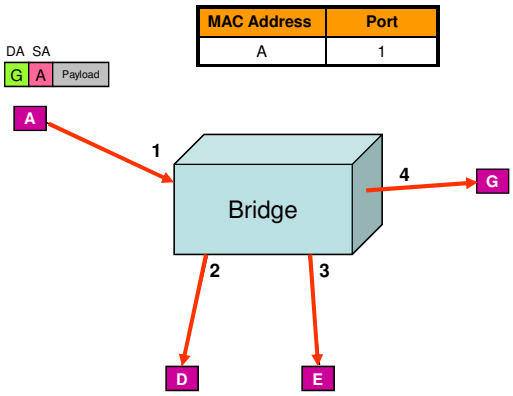
SFD: Start Frame Delimiter – 10101011

IFG: Inter-frame Gap – 12 bytes minimum (96 nsec at 1Gbps)

182

Leon Bruckman

Learning bridge – Unknown flood

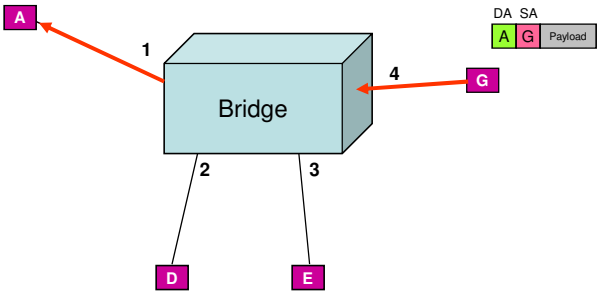


MAC Address	Port
A	1

183

Leon Bruckman

Learning bridge - Unicast



MAC Address	Port
A	1
G	4

184

Leon Bruckman

MAC learning

- In a VLAN network there are two options of MAC learning:
  - Shared VLAN Learning (SVL) – Not VLAN aware
    - Easier to implement, but unable to support duplicate MACs in separate VLANs
  - Independent VLAN Learning (IVL)– VLAN aware
    - More complicated but supports duplicated MACs in separated VLANs
- MAC addresses learned are removed if:
  - Not refreshed (aged)
  - Learned in another port (Device moved)

185

Leon Bruckman

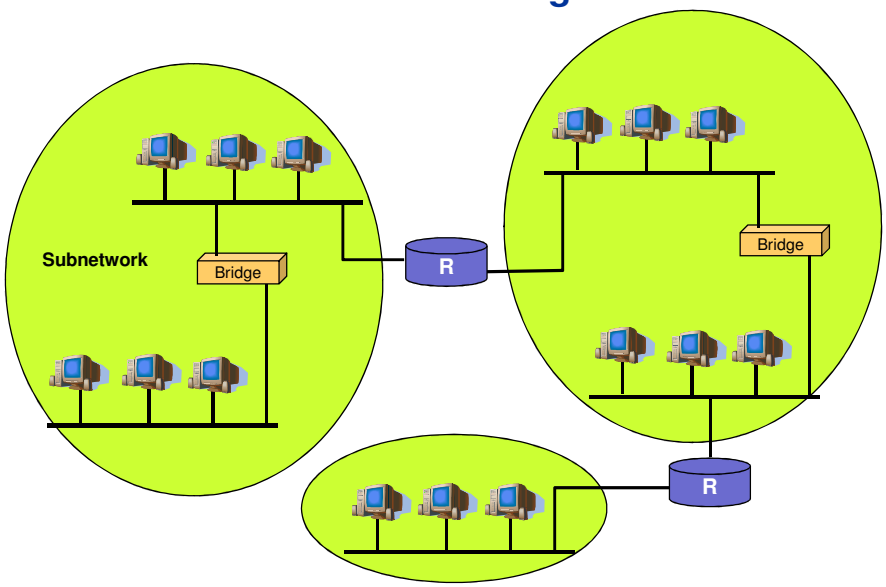
Layer 2 networking limitations vs Layer 3

- Flat addressing scheme limits the size of the network
  - Layer 3: Networks divided into subnetworks
- Flooding over the entire network create congestion and security hazards
  - Flooding limited to layer 2 network
- Spanning tree lack of flexibility and load sharing
  - Open Shortest Path First (OSPF) routing protocols and MPLS

186

Leon Bruckman

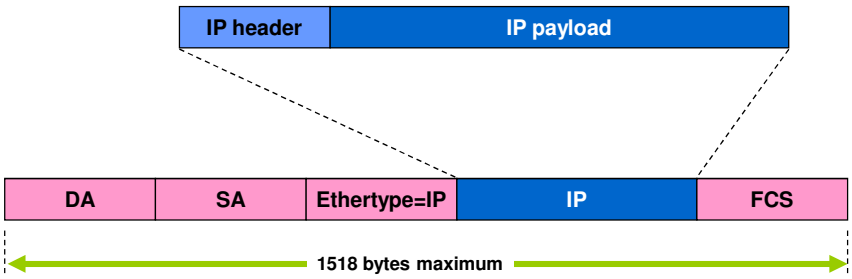
IP networking



187

Leon Bruckman

IP over Ethernet encapsulation



188

Leon Bruckman

What is Layer 2 missing to support WAN ?

- Scalability
- Traffic Engineering
- Topologies
- Protection and restoration
- Quality of Service
- Legacy services support
- Operation and management

189

Leon Bruckman

Scalability

- VLAN
  - The number of VLANs is limited to 4k
    - Not enough for a MAN or WAN
  - Solution: Provider Bridges (QinQ) IEEE 802.1ad
- MAC address
  - MAC address is large enough, but…
  - Bridges must learn all MAC addresses in the network
    - Very large number in MAN or WAN
  - Solution: Provider Backbone Bridges (MAC in MAC) IEEE 802.1ah

190

Leon Bruckman

Traffic engineering

- Connectionless protocols are hard to traffic engineer
  - Very limited control over flow path
- No load balancing in spanning tree
- No fairness protocols
  - Very elementary flow control tools (pause)
- No standard reservation protocols

Leon Bruckman

Topologies

- Ethernet networks are usually built using:
  - Hub and spoke
  - Full mesh
  - Partial mesh
- Fiber networks are usually layed-out as rings
  - Ethernet protocols are not efficient handling rings
    - Do not take advantage of the topology

Leon Bruckman

Protection and restoration

- Slow protection scheme
  - 10s of seconds
  - Even RSTP is not fast enough to meet 50msec requirement
- No protection maintenance commands
- No differential protection schemes
  - All flows are protected the same way

193

Leon Bruckman

Quality of Service

- Usually 8 Strict priority queues only
  - No sophisticated QoS options
- Lower priority classes starvation
- Topology dependent advantages
- No standard definitions and mechanisms for QoS parameters
  - E.g. delay, delay variation

194

Leon Bruckman

Legacy service transport

- No synchronization between nodes
  - Complicates and limits circuit emulation over Ethernet
- No QoS guarantees for high priority services

195

Leon Bruckman

Operation and Management

- Rudimentary failure detection schemes
  - Some support at the physical level when using auto-discovery
- No maintenance tools
  - E.g. Loops

196

Leon Bruckman

Metro Ethernet Forum - MEF

- The MEF, as the defining body for Carrier Ethernet is a global industry alliance comprising more than 150 organizations including telecommunications service providers, cable MSOs, network equipment/software manufacturers, semiconductors vendors and testing organizations.
- The MEF's mission is to accelerate the worldwide adoption of Carrier-class Ethernet networks and services.
- The MEF develops Carrier Ethernet technical specifications and implementation agreements to promote interoperability and deployment of Carrier Ethernet worldwide.

197

Leon Bruckman

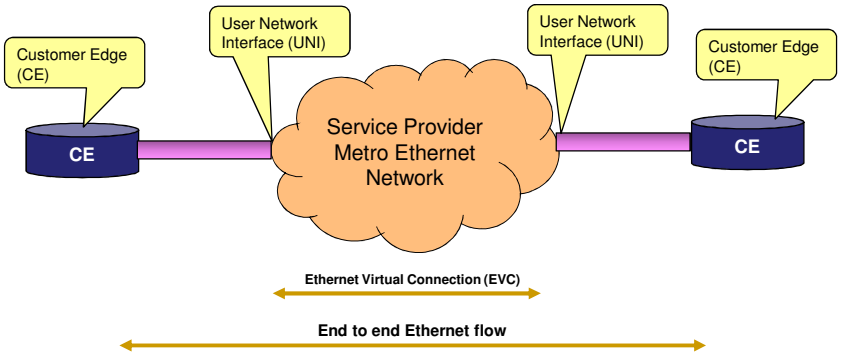
Objectives

- **Build consensus** and unite service providers, equipment vendors and end customers on Ethernet service definition, technical specifications and interoperability.
- **Facilitate implementation** of existing and new standards, Ethernet service definition, test procedures and technical specifications of the MEF to allow delivery of Ethernet services and make Carrier Ethernet-based core, metro and access networks truly carrier class.
- **Enhance** worldwide **awareness** of the **benefits** of **Ethernet services**, enabled applications and Ethernet based networks.

198

Leon Bruckman

Ethernet Service Model



- Connectivity between UNIs is specified by the Ethernet Virtual Connection (**EVC**).
  - There are a number of types of EVC and a number of service attributes that an EVC can have.
- There are a number of different service attributes for a **UNI**.

199

Leon Bruckman

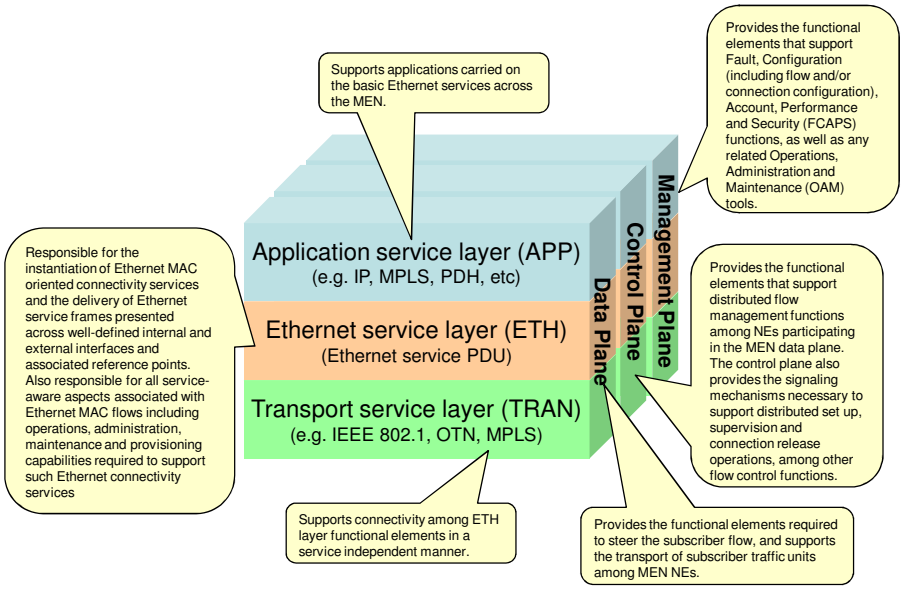
Ethernet flows and connections

- An **Ethernet flow** represents a particular unidirectional stream of Ethernet frames that share a common treatment for the purpose of transfer across the MEN.
  - In particular, an end-to-end Ethernet flow refers to the flow of Ethernet frames between the communicating terminal equipment (TE) that creates and terminates the Ethernet frames.
- The **Ethernet Virtual Connection** (EVC) is the architecture construct that supports the association of UNI reference points for the purpose of delivering an Ethernet flow between subscriber sites across the MEN.
  - There may be one or more subscriber flows mapped to a particular EVC (e.g., there may be more subscriber flows identified by the flow classification rules at the ingress point to a network than EVCs).

200

Leon Bruckman

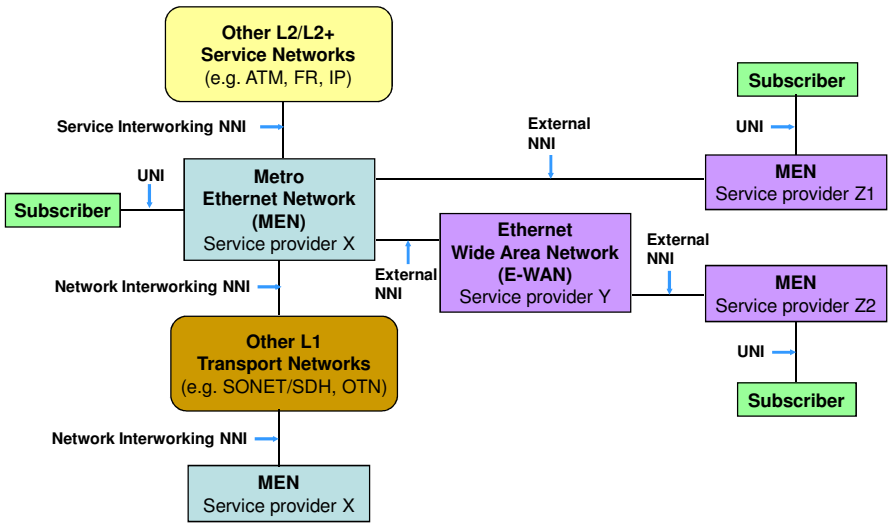
MEN layer network model



201

Leon Bruckman

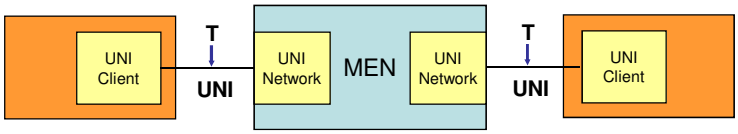
MEN External Interfaces and associated reference points



202

Leon Bruckman

UNI – User Network Interface

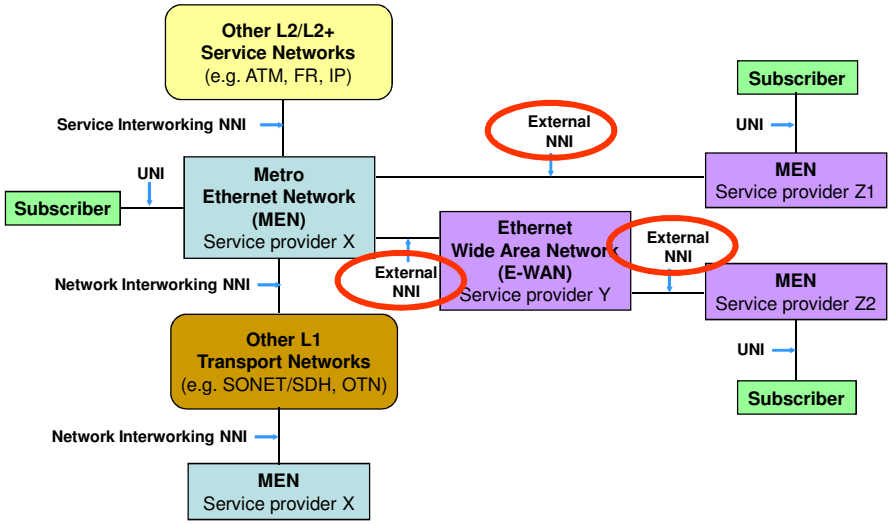


- UNI client (UNI-C)
  - The UNI-C is a compound architectural component of a MEN that represents all of the functions required to connect a subscriber to a MEN.
    - From the perspective of the MEN, the UNI-C supports the set of functions required to exchange data, control and management plane information with the MEN subscriber.
- UNI network (UNI-N)
  - The UNI-N is a compound architectural component of a MEN that represents all of the functions required to connect a MEN to a MEN subscriber.
    - From the perspective of the subscriber, the UNI-N supports the set of functions required to exchange data, control and management plane information with the MEN.

203

Leon Bruckman

MEN External Interfaces and associated reference points



204

Leon Bruckman

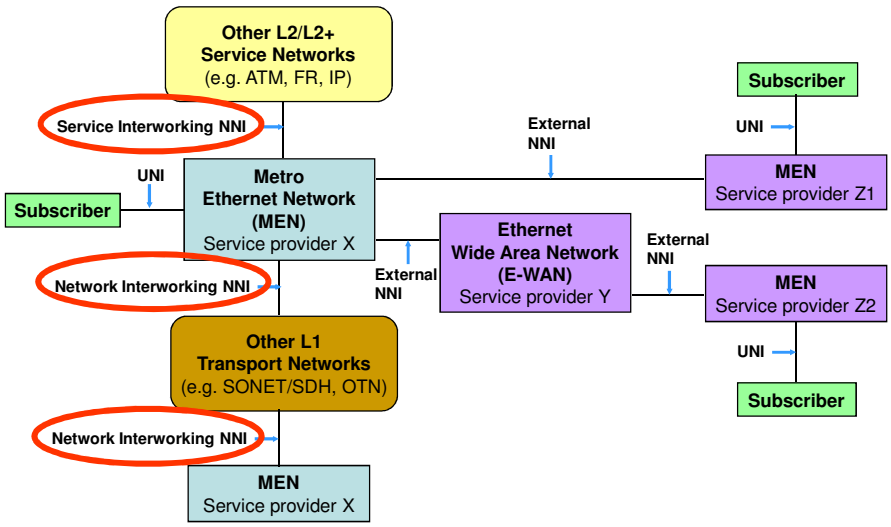
NNIs – Network to Network Interfaces

- External NNI (E-NNI)
  - The E-NNI is an open interface used to interconnect two MEN service providers.
    - It provides a reference point for network equipment (NE) and Ethernet service demarcation between the two directly attached MENs.
    - It also provides a reference point for NEs and Ethernet service demarcation between a MEN and an Ethernet service aware Wide Area Network (E-WAN).
- Internal NNI (I-NNI)
  - The I-NNI is an open interface used to interconnect NEs from a given MEN service providers.
    - The I-NNI provides a reference point for Ethernet service demarcation between the two directly attached NEs.

205

Leon Bruckman

MEN External Interfaces and associated reference points



206

Leon Bruckman

Interworking Interfaces

- Network Interworking NNI (NI-NNI)
  - The NI-NNI is an open interface that supports the extension of transport facilities used to support Ethernet services, and associated EVCs, over an external transport network(s) not directly involved in the end-to-end Ethernet service.
    - The NI-NNI is intended to preserve the characteristic information of a subscriber's flow.
    - The NI-NNI also provides a reference point for demarcation between the two MEN service provider interfaces attached via public transport networks.
- Service Interworking NNI (SI-NNI)
  - The SI-NNI is an interface that supports the interworking of an MEF service with services provided via other service enabling technologies (e.g., Frame Relay, ATM, IP, etc.).
    - The SI-NNI provides a reference point for demarcation between a MEN and another public service network.

207

Leon Bruckman

MEN Physical Components

- Customer Edge (CE) Equipment
  - The CE equipment is a physical element of the MEN architecture that contains all of the functional elements in the customer network required to request services from a MEN.
  - Individual functional elements in a CE may be either entirely in the subscriber domain, or may be entirely in the service provider domain (and managed by the service provider/network operator).
  - From the perspective of the MEN, the CE must support at the minimum the set of functions associated with the UNI-C.
  - The CE may be implemented by an (Ethernet) Switch, (IP/MPLS) Router or a Host System. As such, the CE includes functional elements at the ETH, TRAN layers, and (optionally) APP layers.
- Provider Edge (PE) Equipment
  - The PE equipment is a physical element of the MEN architecture that contains the functional elements in the MEN required to support either a subscriber access link or an access link to another external network in the ETH layer.
  - There may be TRAN layer physical components between a PE and its associated CE(s).
  - From the perspective of the Subscriber, the PE supports at a minimum the set of functions associated with UNI-N.
  - The subscriber itself has no other perspective of the physical/logical implementation of the PE. A subscriber sees the functions provided via the UNI

208

Leon Bruckman

MEN Physical Components – (Cont'd)

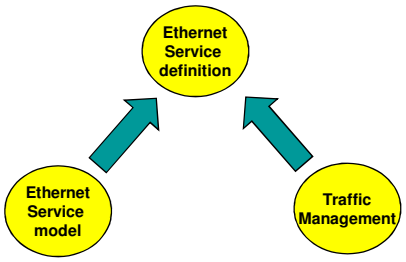
- Provider (P) Core Equipment
  - The P core equipment is a physical element of the MEN architecture that represents any other provider equipment participating in ETH layer network that does not support any functional elements associated with a PE.
  - There may be TRAN layer physical components between a P device and other P/PE devices in the service provider network.
  - From the perspective of the Provider, the P devices are those that do not participate in UNI-N/E-NNI functions.
- Network Termination (NT) Equipment
  - The NT equipment is a physical device that supports the functional elements associated with the end point of the provider TRAN layer network and the beginning of the Subscriber TRAN layer network.
  - The NT's responsibilities include physical medium performance monitoring, timing, line encoding/conversion, among others.
- Transport Edge (TE) Equipment
  - The TE equipment is a physical device that enables TDM/packet multiplexing of multiple customer flows into single physical link.
  - The TE equipment does not need to be able to inspect Ethernet frames generated by a subscriber.
  - A TE may also be a converter from one TRAN type to another.

209

Leon Bruckman

Ethernet services

- Ethernet has its origins in providing ***network connectivity*** and was not originally used to provide ***services***.
- With the introduction of Metro Ethernet services, Service Providers started using this Ethernet "***connectivity***" technology to provide Ethernet "***services***".
- The services are from a Subscriber perspective and are defined based on the service attributes that might appear in a **Service Level Agreement** (SLA) or **Service Level Specification** (SLS).



210

Leon Bruckman

Ethernet services definitions

- The technical definition of a service is in terms of what is seen by each CE.
  - This includes the UNI, which is the physical demarcation point between the responsibility of the Service Provider and the responsibility of the Subscriber.
  - A UNI must be dedicated to a single Subscriber.
- The CE and MEN exchange Service Frames across the UNI.
  - The Service Frame consists of the first bit of the Destination MAC Address through the last bit of the Frame Check Sequence.
  - The protocol as seen by the CE operating at the UNI must be standard Ethernet with the exception that may have a length greater than that specified in IEEE.

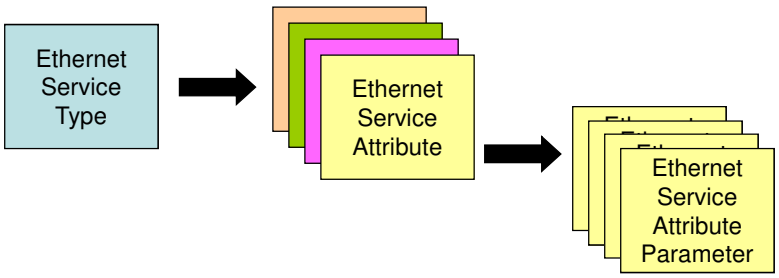
DA	SA	EtherType	VLAN	EtherType	Payload	CRC

- There are no assumptions about the details of the Metro Ethernet Network.
  - It could consist of a single switch or an agglomeration of networks based on many different technologies.

211

Leon Bruckman

Ethernet service definition framework

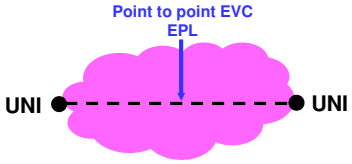


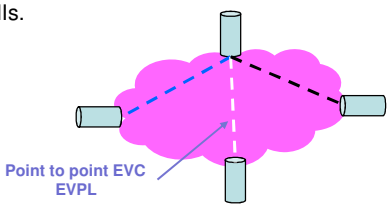
212

Leon Bruckman

Ethernet Line (E-Line) – point to point EVC service

- Any Ethernet service that is based on a Point-to-Point Ethernet Virtual Connection is designated as an Ethernet Line (E-Line) Service type.
  - EPL – Ethernet Private Line: Port based
  - EVPL - Ethernet Virtual Private Line: VLAN based
- An E-Line Service type can provide:
  - Symmetrical or asymmetrical bandwidth for data sent in either direction.
  - With performance assurances or without performance assurances
- For EVPL Service Multiplexing may occur at the UNIs.
  - For example, more than one E-Line Service may be offered on the same physical port at one or both of the UNIs.



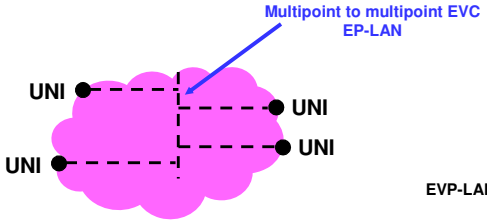


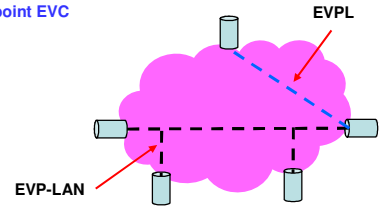
213

Leon Bruckman

Ethernet LAN (E-LAN) – multipoint to multipoint EVC service

- Any Ethernet Service that is based upon a Multipoint-to-Multipoint Ethernet Virtual Connection is designated as an Ethernet LAN (E-LAN) Service type.
  - EP-LAN – Ethernet Private LAN: Port based
  - EVP – LAN – Ethernet Virtual Private LAN: VLAN based
- An E-LAN Service type can provide:
  - Best effort service with no performance assurances.
  - Some performance assurances (some are not defined e.g. delay, loss)
- For EVP-LAN Service Multiplexing may occur at the UNIs.
  - For example, an E-LAN Service type and an E-Line Service type may be service multiplexed at the same UNI .



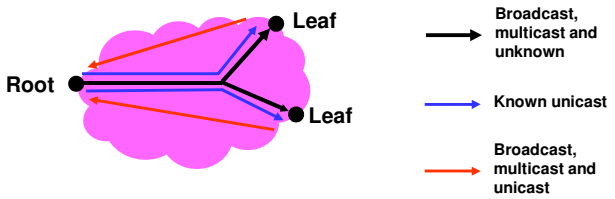


214

Leon Bruckman

Rooted multipoint service

- In a Rooted-Multipoint service one or more of the UNIs is designated as a Root and each of the other UNIs are designated as a Leaf.
  - An ingress Service Frame mapped to the EVC at a Root UNI may be delivered to one or more of the other UNIs in the EVC.
  - An ingress Service Frame mapped to the EVC at a Leaf UNI does not result in an egress Service Frame at another Leaf UNI but may result in an egress Service Frame at some or all of the Root UNIs.

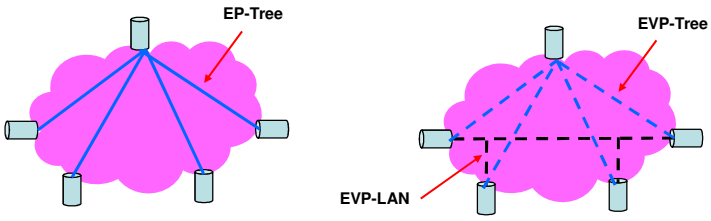


215

Leon Bruckman

E-Tree services

- Ethernet Private Tree service – EP-Tree
  - Provides services distributed from a centralized site (or few such sites) where the distribution site is designated as a Root and all the remaining sites are designated as leaves.
- Ethernet Virtual Private Tree service – EVP-Tree
  - Provides access to certain applications or content services from well-defined access points.
  - One or more of the Subscriber's UNIs may also support other services, e.g., EVPL or EVP-LAN.



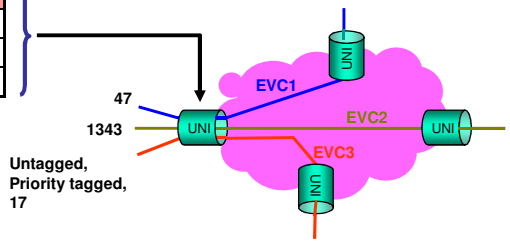
216

Leon Bruckman

Customer Edge-VLAN ID/EVC Map Service Attribute

- At each UNI there must be a mapping of each CE-VLAN ID to at most one EVC.
- The mapping of one or more CE-VLAN IDs to an EVC is an attribute associated with the EVC at the UNI.
- The collection of all of these mappings is called the CE-VLAN ID/EVC Map.
  - Note that a given CE-VLAN ID may not be mapped to any EVC.

CE-VLAN ID	EVC
47	EVC1
1343	EVC2
17*	EVC3



217

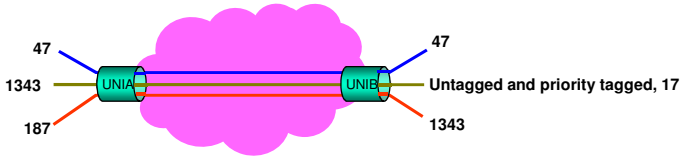
Leon Bruckman

CE-VLAN significance

- CE-VLAN ID values may only be significant at a given UNI.
  - The CE-VLAN ID/EVC mapping for a given EVC at a UNI may be different from the mapping at another UNI in the EVC.
  - Note that when the CE-VLAN ID Preservation attribute applies to an EVC, the mappings for the EVC are identical as is the case for EVC1

UNI-A	
CE-VLAN ID	EVC
47	EVC1
1343	EVC2
187	EVC3

UNI-B	
CE-VLAN ID	EVC
47	EVC1
17*	EVC2
1343	EVC3



218

Leon Bruckman

CE-VLAN ID Preservation Service Attribute

- A Service Frame is defined to have its CE-VLAN ID Preserved when the relationship between the ingress Service Frame and its corresponding egress Service Frame(s) is as follows:

Ingress Service Frame	Egress Service Frame
No IEEE 802.1Q Customer VLAN Tag	No IEEE 802.1Q Customer VLAN Tag
Contains IEEE 802.1Q Customer VLAN Tag	Contains IEEE 802.1Q Customer VLAN Tag with VLAN ID equal to the VLAN ID of the Tag on the ingress Service Frame

- An EVC with the CE-VLAN ID Preservation Service Attribute must preserve the CE-VLAN ID for Service Frames as described in the following table:

CE-VLAN ID/EVC Map Characteristic	Service Frames with CE-VLAN ID Preserved
All to One Bundling at all UNIs	All Data Service Frames
All other cases	All tagged Data Service Frames with VLAN ID in the range 1 – 4094

219

Leon Bruckman

Untagged and priority tagged Service Frames

- Untagged and priority tagged Service Frames must have the same CE-VLAN ID and that value must be configurable to any value in the range 1, 2, …, 4094.
- When the **CE-VLAN ID Preservation Service Attribute** is **not in force** for an EVC to which the CE-VLAN ID for untagged and priority tagged Service Frames is mapped, egress Service Frames for this EVC at the given UNI must be untagged.
- When **CE-VLAN ID Preservation Service Attribute** is **in force** for an EVC to which the CE-VLAN ID for untagged and priority tagged Service Frames is mapped, ingress untagged and priority tagged Service Frames at this UNI are not mandated to have their CE-VLAN ID preserved except in the case of All to One Bundling.

220

Leon Bruckman

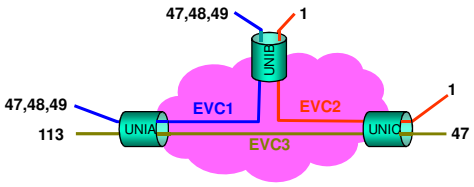
Bundling Service Attribute

- When a UNI has the Bundling attribute, it must be configurable so that more than one CE-VLAN ID can map to a particular EVC at the UNI.
- An EVC with more than one CE-VLAN ID mapping to it must have the CE-VLAN ID Preservation Service Attribute and the list of CE-VLAN IDs mapped to the EVC must be the same at each UNI in the EVC

UNI-A	
CE-VLAN ID	EVC
47,48,49	EVC1
113	EVC3

UNI-B	
CE-VLAN ID	EVC
47,48,49	EVC1
1	EVC2

UNI-C	
CE-VLAN ID	EVC
1	EVC2
47	EVC3



221

Leon Bruckman

All to One Bundling Service Attribute

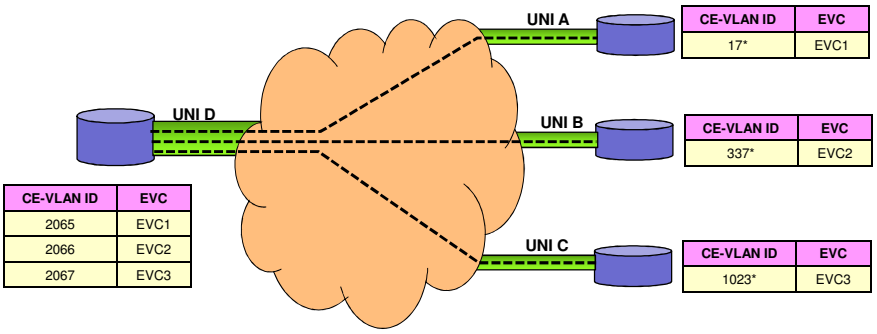
- When a UNI has the All to One Bundling attribute set, **all** CE-VLAN IDs must map to a **single** EVC at the UNI.
- The EVC at the UNI must have the CE-VLAN ID Preservation Service Attribute and the list of CE-VLAN IDs mapped to the EVC must include all CE-VLAN IDs and be the same at each UNI in the EVC. This means that:
  - Such a UNI cannot have Service Multiplexing
  - All UNIs in the EVC must have the All to One Bundling Service Attribute
- All to One Bundling is a special case of Bundling but it is sufficiently important to be called out as a separate attribute.
- EPL, EP-LAN and EP-Tree must use All in one Bundling

222

Leon Bruckman

Example 1 of the Use of the CE-VLAN ID/EVC Map

- Untagged UNIs
  - In connecting branch enterprise locations to a hub enterprise location, it is desirable to make the configuration of the branch CEs simple.
  - A similar objective applies to providing access to higher layer services, e.g., Internet Access, where the configuration of the CE at the sites accessing the service should be kept simple.

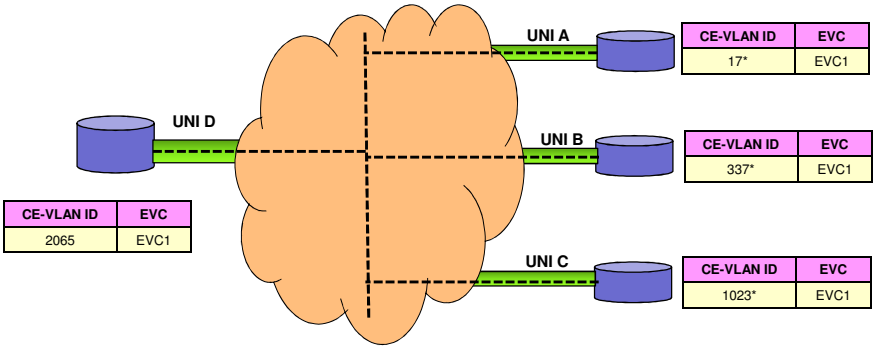


223

Leon Bruckman

Example 2 of the Use of the CE-VLAN ID/EVC Map

- Use of Rooted-Multipoint EVC
  - A higher layer service is being provided to three different customers.
  - By using a Rooted-Multipoint EVC, all three customers can be reached by the higher layer service provider at UNI D using a single EVC. This can save a large number of Point-to-Point EVCs when there are a large number of customers.
  - Each customer's CE can only send to the higher layer service CE thus keeping each customer from seeing other customers' traffic.

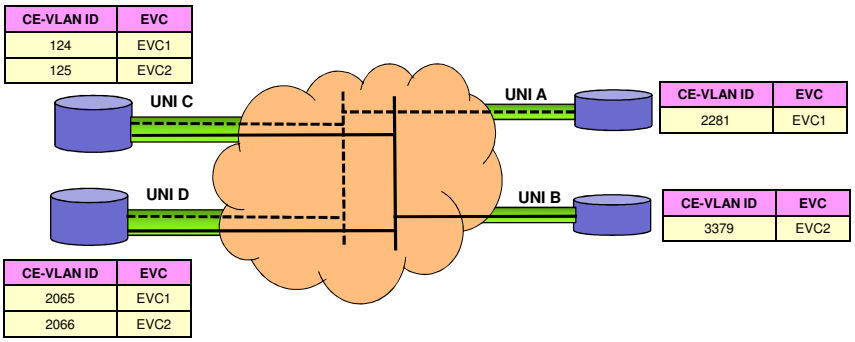


224

Leon Bruckman

Example 3 of the Use of the CE-VLAN ID/EVC Map

- Redundant Higher Layer Service Access
  - A Multipoint-to-Multipoint EVC is used for each customer of the higher layer service. Higher layer service routers are attached to two UNIs in each such EVC.
  - Routing protocols running among the two higher layer service routers and the customer router allow the customer to access the higher layer service in a redundant fashion.



225

Leon Bruckman

Layer 2 Control Protocol Processing Requirements

- The requirements are intended to provide guidance for actual deployments of the Ethernet services defined, while at the same time allowing for flexibility among the Service Provider offerings.
- A Layer 2 Control Protocol is identified by one of the following MAC Destination Addresses:
  - 01-80-C2-00-00-00 through 01-80-C2-00-00-0F - Bridge Block of protocols
  - 01-80-C2-00-00-20 through 01-80-C2-00-00-2F - GARP Block of protocols
- For each service, protocols are configured to:
  - Tunnel
    - Frames are transparently passed to a given EVC for transport across the MEN to the destination UNI(s).
  - Peer
    - The MEN will actively participate with the protocol if the DA is as specified.
  - Discard
    - The MEN will discard ingress L2CP frames of a given <protocol, DA> pair and will not generate that <protocol, DA> pair on egress from the MEN.

226

Leon Bruckman

CoS Identifier Based on EVC

- All ingress Data Service Frames mapped to the EVC shall have the same CoS Identifier.

EVC	Service Frame Type	CoS
EVC1	Data	Gold
	Tunneled Layer 2 Control Protocol	Gold
EVC2	Data	Silver
	Tunneled Layer 2 Control Protocol	Platinum

227

Leon Bruckman

CoS Identifier Based on Priority Code Point Field

- The CoS Identifier for an ingress Data Service Frame shall be determined by the EVC and non-overlapping sets of values of the CE-VLAN CoS.
- If the ingress Data Service Frame is untagged, it shall have the same CoS Identifier as an ingress Data Service Frame with Priority Code Point field = 0.
- The union of the sets of CE-VLAN CoS values must contain all of the possible CE-VLAN CoS values.

EVC	PCP	CoS
EVC1	4,5,6,7	Gold
	0,3	Silver
	1,2	Discard
	Untagged	Silver
EVC2	7	Platinum
	0,1,2,3,4,5,6	Gold

228

Leon Bruckman

Performance monitoring details

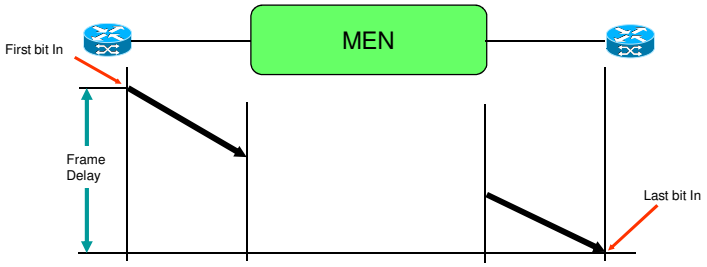
- Performance Attributes must not apply to Service Frames with the level of conformance determined to be Yellow or Red.
  - Typically, the Frame Loss Ratio Performance will be degraded for Service Frames determined to be Yellow.
- For a given EVC and CoS instance, Performance Objectives may be specified over any given subset of the Ordered Pairs of UNIs (describing transmission direction) on the EVC.
  - Once a subset of UNI pairs is defined, then all attributes in this section shall have performance objectives applying to that subset.
- Values of the Service Frame delay, delay variation, and loss performance during periods of unavailable time must not be used to determine Service Frame delivery compliance.
  - A process must be established to exclude all performance during unavailable periods from comparison with Service Frame performance objectives.
- The assessment of all performance attributes should account for unexpected arrival phenomena, such as frame duplication, or frames arriving in a different order from that observed on ingress
  - The presence of these phenomena alone do not necessarily exclude a Service Frame from the set of **Qualified Service Frames**.

229

Leon Bruckman

Frame delay performance for point to point EVCs

- The Frame Delay for a Service Frame is defined as the time elapsed from reception at the ingress UNI of the first bit of the ingress Service Frame until the transmission of the last bit of the Service Frame at the egress UNI.



- **Formally**: Define the objective (D), the measurement interval (T) and the percentile (P). The performance is met if the delay of P % of the measurements during T is lower or equal to D.

230

Leon Bruckman

Frame delay performance for multipoint EVCs

- Calculate the point to point delay performance for any subset of UNI pairs in the service
  - For a Rooted-Multipoint EVC, the subset must be such that all ordered pairs contain at least one UNI that is designated as a Root.

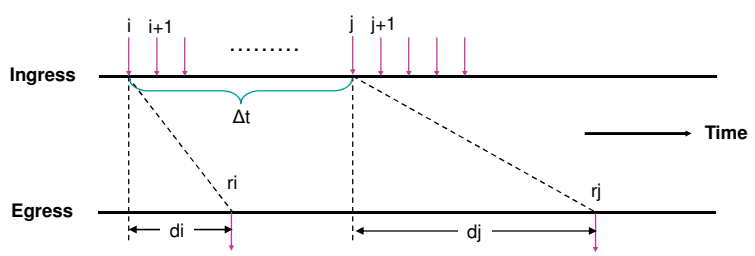
- **Formally**: Define the objective (D), the measurement interval (T), the subset (S) and the percentile (P).
  - The performance is met if the delay of P % of the measurements during T for S is lower or equal to D.

231

Leon Bruckman

Inter-frame delay variation performance for point to point EVCs

- Inter Frame Delay Variation (IFDV) is the difference between the one-way delays of a pair of selected Service Frames.
  - FDV Performance is applicable to successfully delivered Service Frames



- **Formally**: Define the objective (IFDV), the measurement interval (T), the separation between frame pairs (Δt) and the percentile (P). The performance is met if the measured delay variation of P % of the measurements during T is lower or equal to FDV.

232

Leon Bruckman

Frame loss performance for point to point EVCs

- For a given T:

$$FLR = \begin{cases} \left(\frac{I_t - E_t}{I_t}\right) \times 100 \text{ if } I_t \geq 1 \\ 0 \text{ otherwise} \end{cases}$$

- Where
  - **FLR** = Frame Loss Ratio
  - I_t = the number of Service Frames that arrive at an ingress UNI during the interval T that should be delivered to the egress UNI
  - E_t = the number of such Service Frames that are delivered

233

Leon Bruckman

Availability Performance

- Availability Performance is the percentage of time within a specified time interval during which the Frame Loss Ratio Performance is small.
  - As an example, a service provider can define the availability performance to be measured over a month and the value for the Availability Performance objective to be 99.9%.
  - In a month with 30 days and no scheduled downtime this parameter will allow the service to be unavailable for approximately 43 minutes out of the whole month.
- The Availability for a particular Class of Service instance on a Point-to-Point EVC for a time interval T is based on the following four parameters:
  - **Δt** - a time interval much smaller than T,
  - **Cu** - a loss ratio threshold which if equaled or exceeded suggests unavailability
  - **Ca** - a loss ratio threshold which if not exceeded suggests availability with Ca≤Cu
  - **n** - number of consecutive small time intervals Δt over which to assess availability.

234

Leon Bruckman

Availability calculation (n=4)

Available				Unavailable			
				FLR ≥ Cu	FLR ≥ Cu	FLR ≥ Cu	FLR ≥ Cu
Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt

Available				Available			
				FLR ≥ Cu	FLR < Cu	FLR ≥ Cu	FLR ≥ Cu
Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt

Unavailable				Available			
				FLR ≤ Ca	FLR ≤ Ca	FLR ≤ Ca	FLR ≤ Ca
Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt

Unavailable				Unavailable			
				FLR ≤ Ca	FLR > Ca	FLR ≤ Ca	FLR ≤ Ca
Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt

235

Leon Bruckman

Bandwidth profile service attributes

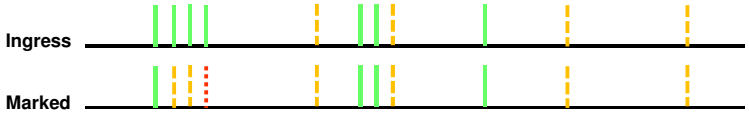
- Committed Information Rate (CIR) expressed as bits per second. CIR must be ≥ 0.
- Committed Burst Size (CBS) expressed as bytes. When CIR > 0, CBS must be greater than or equal to the largest Maximum Transmission Unit size among all of the EVCs that the Bandwidth Profile applies to.
- Excess Information Rate (EIR) expressed as bits per second. EIR must be ≥ 0.
- Excess Burst Size (EBS) expressed as bytes. When EIR > 0, EBS must be greater than or equal to the largest Maximum Transmission Unit size among all of the EVCs that the Bandwidth Profile applies to.
- Coupling Flag (CF) must have only one of two possible values, 0 or 1.
- Color Mode (CM) must have only one of two possible values, “color-blind” and “color-aware.”

236

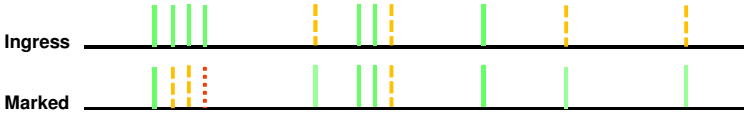
Leon Bruckman

Color mode

- The Bandwidth Profile algorithm is said to be in **color aware** mode when each Service Frame already has a level of compliance (i.e., a color) associated with it and that color is taken into account in determining the level of compliance by the Bandwidth Profile algorithm.



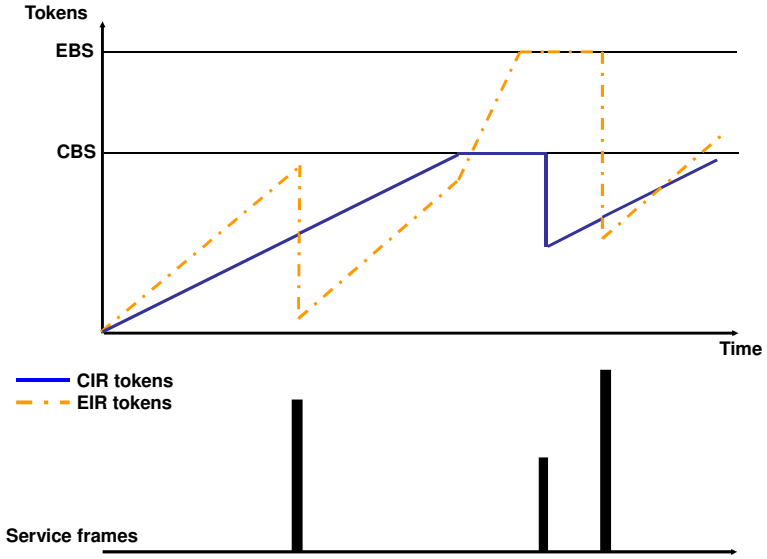
- The Bandwidth Profile algorithm is said to be in **color blind** mode when the color (if any) already associated with each Service Frame is ignored by the Bandwidth Profile Algorithm.



237

Leon Bruckman

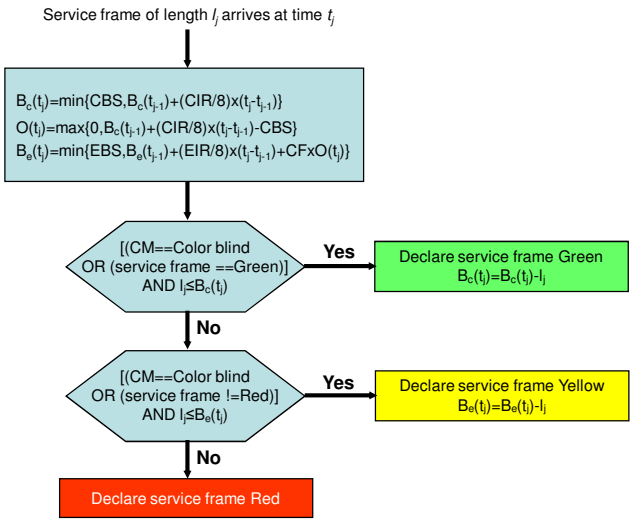
Token accumulation - Coupling flag =1



238

Leon Bruckman

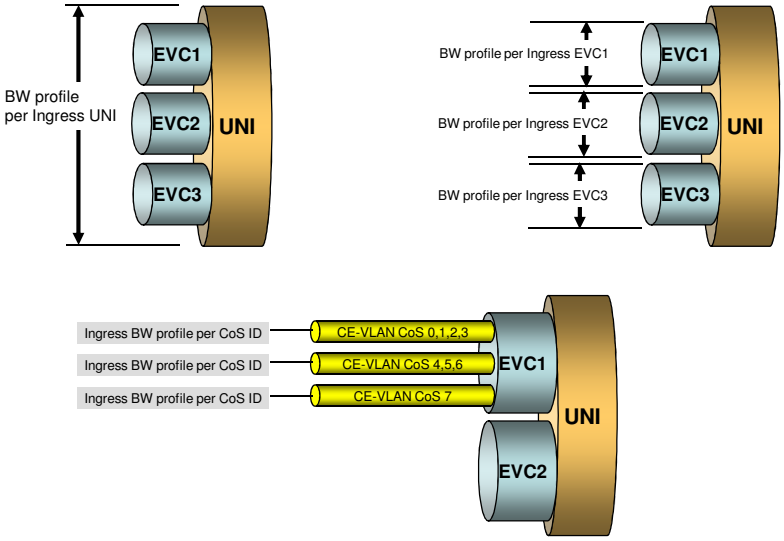
The Bandwidth Profile Algorithm



239

Leon Bruckman

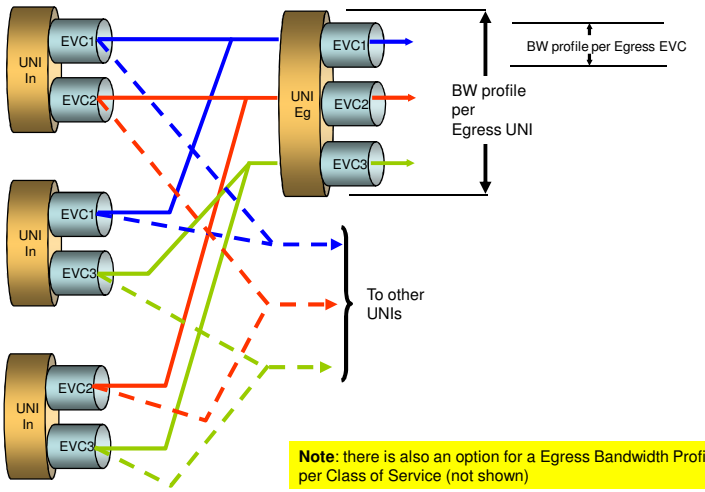
Ingress bandwidth profiles



240

Leon Bruckman

Egress Bandwidth profiles



Note: there is also an option for a Egress Bandwidth Profile per Class of Service (not shown)

241

Leon Bruckman

UNI and EVC per UNI Service Attributes

Attribute	Type of Parameter Value
UNI Identifier	Any string
Physical Medium	A Standard Ethernet PHY
Speed	10 Mbps, 100 Mbps, 10/100 Mbps Auto-Negotiation, 1 Gbps, or 10 Gbps
Mode	Full Duplex
MAC Layer	IEEE 802.3 – 2005
UNI Maximum Transmission Unit Size	Integer ≥ 1522.
Service Multiplexing	Yes or No
UNI EVC ID	A string formed by the concatenation of the UNI ID and the EVC ID
CE-VLAN ID for untagged and priority tagged Service Frames	A number in 1, 2, …, 4094.
CE-VLAN ID/EVC Map	Map
Maximum Number of EVCs	Integer ≥ 1
Bundling	Yes or No
All to One Bundling	Yes or No
Ingress Bandwidth Profile Per Ingress UNI	No or parameters
Ingress Bandwidth Profile Per EVC	No or parameters for each EVC
Ingress Bandwidth Profile Per Class of Service Identifier	No or parameters for each Class of Service Identifier
Egress Bandwidth Profile Per Egress UNI	No or parameters
Egress Bandwidth Profile Per EVC	No or parameters for each EVC
Egress Bandwidth Profile Per Class of Service Identifier	No or parameters for each Class of Service Identifier
Layer 2 Control Protocols Processing	A list of Layer 2 Control Protocols with each being labeled with one of Discard, Peer, Pass to EVC, Peer and Pass to EVC

242

Leon Bruckman

EVC Service Attributes

Attribute	Type of Parameter Value
EVC Type	Point-to-Point, Multipoint-to-Multipoint, or Rooted-Multipoint
EVC ID	An arbitrary string, unique across the MEN, for the EVC supporting the service instance
UNI List	A list of <UNI Identifier, UNI Type> pairs
Maximum Number of UNIs	Integer. **must** be 2 if EVC Type is Point-to-Point. **must** be greater than or equal to 2 otherwise.
EVC Maximum Transmission Unit Size	Integer ≥ 1522.
CE-VLAN ID Preservation	Yes or No
CE-VLAN CoS Preservation	Yes or No
Unicast Service Frame Delivery	Discard, Deliver Unconditionally, or Deliver Conditionally. If Deliver Conditionally is used, then the conditions **must** be specified.
Multicast Service Frame Delivery	Discard, Deliver Unconditionally, or Deliver Conditionally. If Deliver Conditionally is used, then the conditions **must** be specified.
Broadcast Service Frame Delivery	Discard, Deliver Unconditionally, or Deliver Conditionally. If Deliver Conditionally is used, then the conditions **must** be specified.
Layer 2 Control Protocols Processing	A list of Layer 2 Control Protocols labeled Tunnel or Discard.
EVC Performance	Performance objectives for Frame Delay Performance, Frame Delay Variation Performance, Frame Loss Ratio Performance, and Availability Performance and associated Class of Service Identifier(s)

243

Leon Bruckman

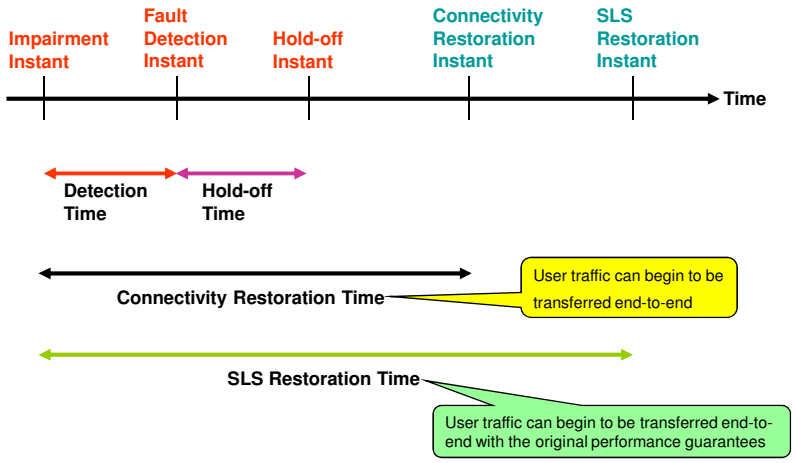
Protection principles

- Protection in Metro Ethernet Networks (MEN) is a self-healing property of the network that allows it to continue to function with minimal or no impact to the network users upon disruption, outages or degradation of facilities or equipment in the MEN.
- Network protection can be viewed in two ways:
  - From the viewpoint of the **user** of the MEN services the actual methods and mechanisms are of minor concern and it is the availability and quality of the services that are of interest. These can be described in a Service Level Specification (SLS).
  - From the viewpoint of the **network provider**. The provider is tasked with translating the SLSs of all the customers (and future customers) into requirements on the network design and function.

244

Leon Bruckman

Protection event timing

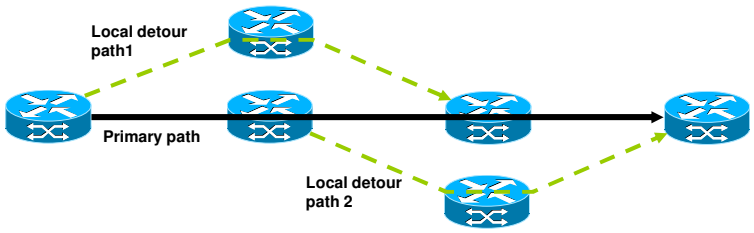


245

Leon Bruckman

Aggregated Line and Node Protection (ALNP)

- ALNP provides protection against local link and nodal failure by using local path detour mechanisms.
- The detour path may provide 1:n protection or 1:1 protection of the primary paths in the network.

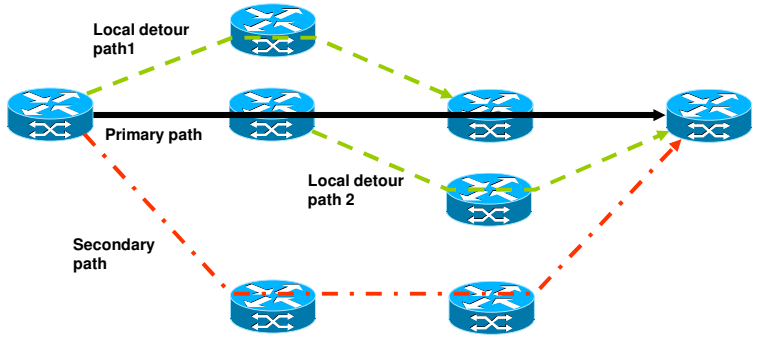


246

Leon Bruckman

End-to-End Path Protection (EEPP)

- End-to-end path protection (EEPP) is the ability to provide a redundant end-to-end path for the primary path. This mechanism can be used to augment ALNP.
- A variation of this method can be used to protect partial segments of the end-to-end path within the same layer



247

Leon Bruckman

External NNI – Phase 1

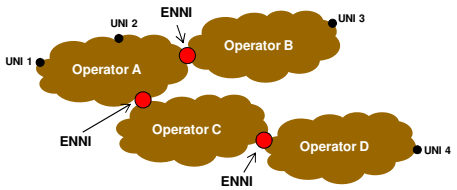
- Support for Point-to-Point and Multipoint-to-Multipoint EVCs spanning an arbi-trary number of operator MENs and ENNIs.
  - Support for Rooted Multipoint EVCs is out of scope.
- Ethernet frames at the ENNI with formats according to the Provider Bridges
- Gigabit Ethernet or 10-Gigabit physical links
- Color aware Bandwidth Profiles at the ENNI.
- Hairpin switching, where ENNI Frames associated with an EVC may be sent back across an ENNI from which they were received by the Operator.
- Link protection based on Link Aggregation
- Link layer OAM

248

Leon Bruckman

ENNI model

- For a given EVC, the Subscriber contracts with a Service Provider to be responsible for delivering Ethernet Services among the UNIs in the EVC.
- The Service Provider, in turn, selects and contracts with various MEN Operators to deliver the UNI-to-UNI services.
  - It is the responsibility of the Service Provider to ensure that the appropriate service and interface attribute values from each Operator are such that the UNI to UNI service features purchased by the Subscriber can be delivered.

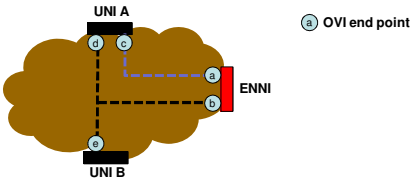


249

Leon Bruckman

Operator Virtual Connection - OVC

- An Operator Virtual Connection (OVC) is the building block for constructing an EVC spanning multiple Operator MENs.
- An OVC is an association of OVC End Points.
  - In turn each OVC End Point is associated with either a UNI or an ENNI.
  - At least one OVC End Point associated by an OVC is at an ENNI.

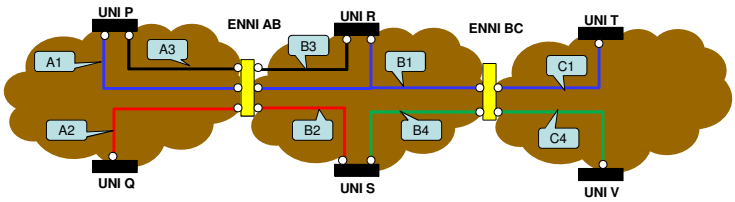


  - An EVC is an association of two or more UNIs.
  - When an EVC associates UNIs attached to more than one Operator MEN, the EVC is realized by concatenating OVCs.

250

Leon Bruckman

Relationship of OVCs to EVCs

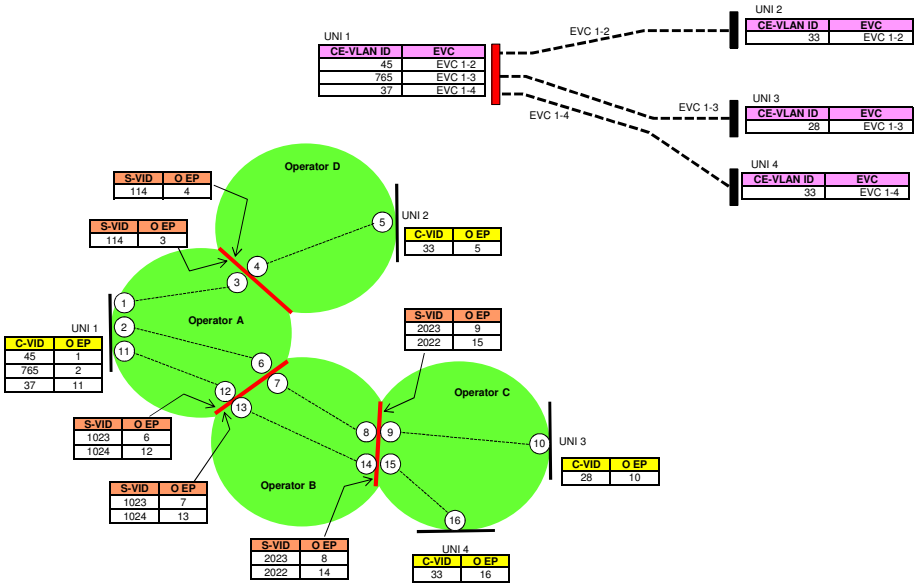


EVC	UNIs	OVCc
Blue	P, R, T	A1, B1, C1
Red	Q, S	A2, B2
Black	P, R	A3, B3
Green	S, V	B4, C4

251

Leon Bruckman

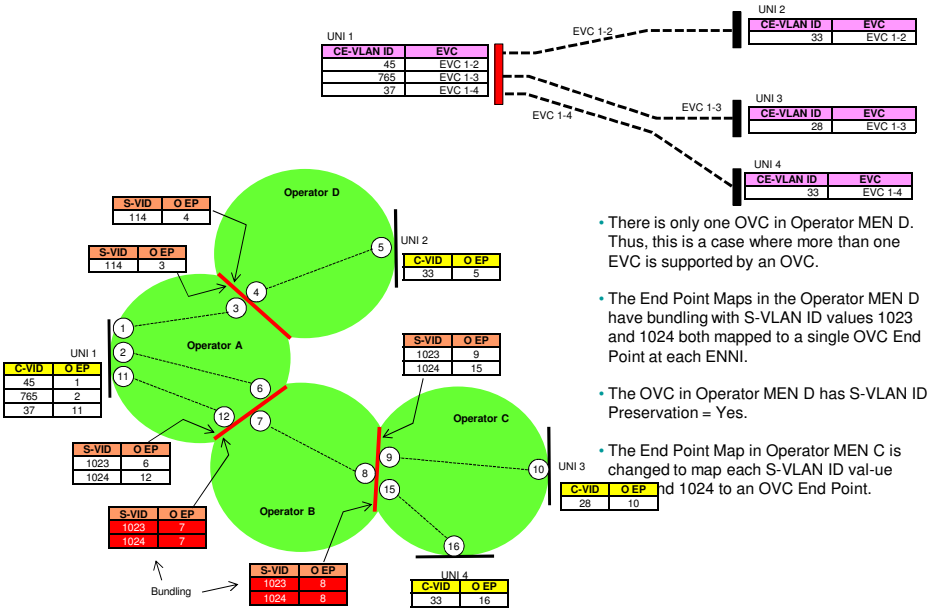
Ethernet Virtual Private Lines to a Hub Location



252

Leon Bruckman

Ethernet Virtual Private Lines to a Hub Location - Bundling

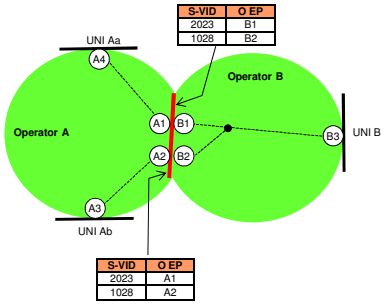


- There is only one OVC in Operator MEN D. Thus, this is a case where more than one EVC is supported by an OVC.
- The End Point Maps in the Operator MEN D have bundling with S-VLAN ID values 1023 and 1024 both mapped to a single OVC End Point at each ENNI.
- The OVC in Operator MEN D has S-VLAN ID Preservation = Yes.
- The End Point Map in Operator MEN C is changed to map each S-VLAN ID val-ue and 1024 to an OVC End Point.

253

Leon Bruckman

Hair pinning

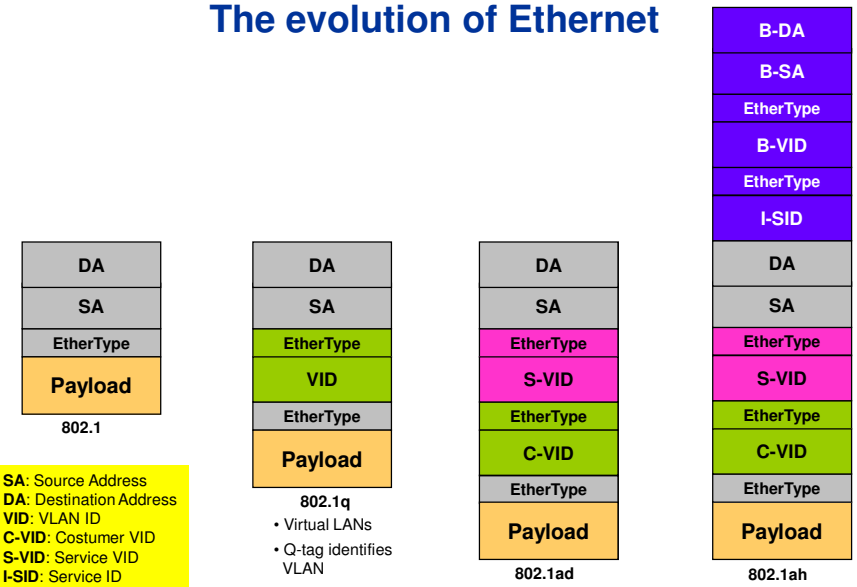


- With this configuration, a Service Frame sent from UNI Aa to UNI Ab will pass through the Operator B MEN where it will be hairpin switched from B1 to B2.
  - A similar path will be followed by a Service Frame sent from UNI Ab to UNI Aa.
- Note that in this example, two OVCs are used in Operator MEN A to implement the single EVC.
- The improper use of Hairpin Switching can lead to data loops at an ENNI.

254

Leon Bruckman

The evolution of Ethernet

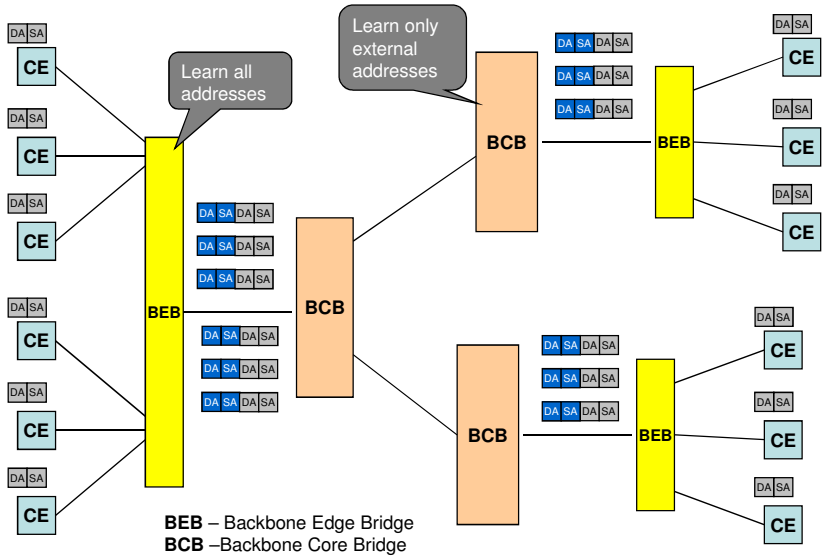


SA: Source Address
DA: Destination Address
VID: VLAN ID
C-VID: Costumer VID
S-VID: Service VID
I-SID: Service ID
B-VID: Backbone VID
B-SA: Backbone SA
B-DA: Backbone DA

255

Leon Bruckman

MACinMAC application – Scalability of Core

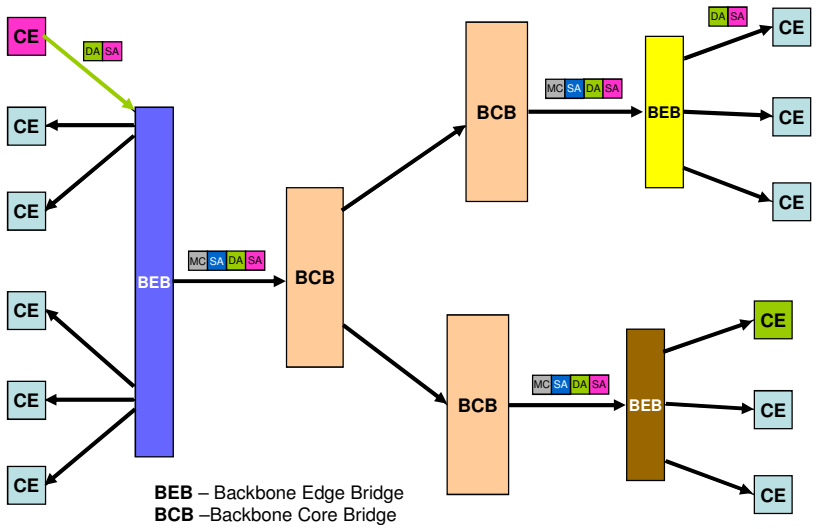


BEB – Backbone Edge Bridge
BCB –Backbone Core Bridge

256

Leon Bruckman

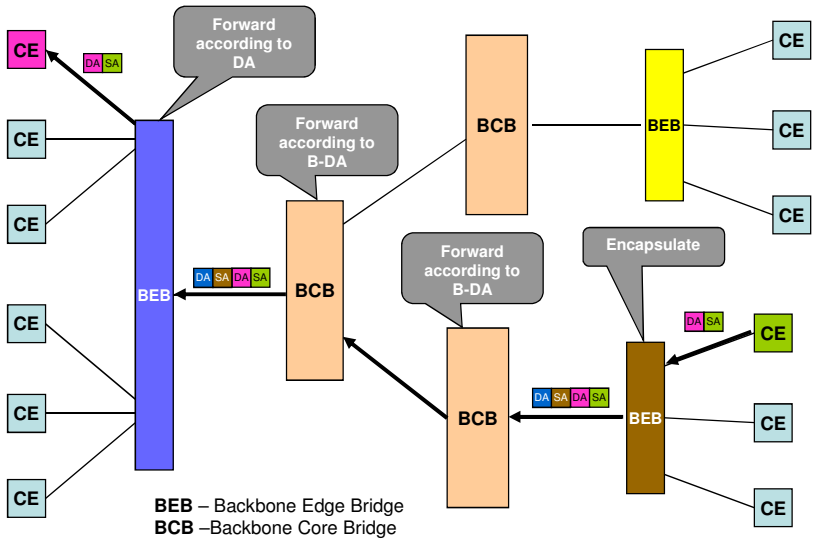
MACinMAC learning



257

Leon Bruckman

MACinMAC unicast



258

Leon Bruckman

PBB-TE – 802.1Qay

- PBB-TE is a framework to provide **carrier-grade point-to-point** Ethernet connectivity over a bridged Ethernet network
- The data plane is based on IEEE 802.1ah (PBB – Provider Backbone Bridge, or MAC-in-MAC).
  - Forwarding is based on <B-DA, B-VID>, where B-DA defines a destination BEB (Backbone Edge Bridge) and B-VID defines a path to that BEB
- **NMS** is used for Ethernet Switched Path (ESP) **provisioning**.
  - Control plane may be introduced in the future
- Ethernet OAM mechanisms based on IEEE 802.1ag
- End-to-end Ethernet Switched Path (ESP) protection between edge bridges

259

Leon Bruckman

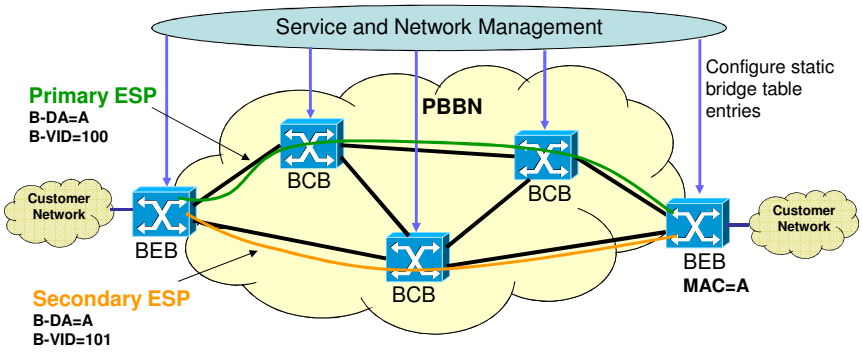
PBB-TE details

- Give control of the forwarding tables to the management plane and hence control forwarding behavior
  - If VLAN-ID exists and DA exists forward, else drop
- Turn off specific bridge functions
  - No broadcast and multicast
  - No flooding of unknows
  - No MAC learning
  - No spanning tree protocol

260

Leon Bruckman

Configuring Ethernet trunks with PBB-TE

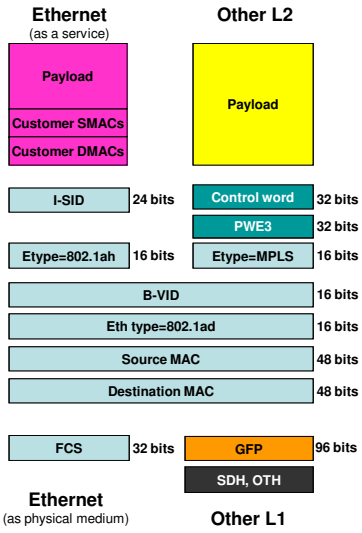


PBBN – Provider Backbone Bridged Network
BEB – Backbone Edge Bridge
BCB – Backbone Core Bridge
ESP – Ethernet Switched Path

261

Leon Bruckman

Multiservice over PBB-TE



262

Leon Bruckman

Operating PBB together with PBB-TE

- PBB is connectionless while PBB-TE is connection oriented
- B-VID space should be divided between PBB and PBB-TE
- Potential problems:
  - Migration from PBB to mixed PBB and PBB-TE
    - B-VIDs should be assigned "intelligently" from the beginning
  - QoS provisioning
    - PBB can not be considered by PBB-TE CAC
    - PBB-TE services may not meet their SLA
    - **Solution**: Strict priority of PBB-TE over PBB

263

Leon Bruckman

The problem – From IEEE 802.1ag

- The most common use of bridged networks has been in an environment where a single administration operates the network and where physical access to all bridges is available.
- A service instance, on the other hand, can be provided to a customer by more than one interconnected Provider Bridged Network.
- Furthermore, each network can be under different and independent administrative control, each with restricted management access to each other's equipment.
  - Physical access to Provider Bridges and other network equipment can be difficult and expensive; equipment can be many kilometers from the service personnel and can be positioned underground, atop a tower, or on the customer's premises.
- As a result, the methods commonly used to ensure the availability of the services offered by IEEE 802.1D and IEEE 802.1Q bridged networks, which assume a single administration and easy access to equipment, are inadequate to manage interconnected Provider Bridged Networks.

264

Leon Bruckman

Principles of Connectivity Fault Management operation

- Connectivity Fault Management (CFM) comprises capabilities for detecting, verifying, and isolating connectivity failures in Virtual Bridged Local Area Networks.
  - These capabilities can be used in networks operated by multiple independent organizations, each with restricted management access to each other's equipment.
- CFM is designed to be transparent to the customer data transported by a network and to be capable of providing maximum fault coverage.
  - Customer data is forwarded transparently by CFM Entities while CFM PDUs are generated and processed

265

Leon Bruckman

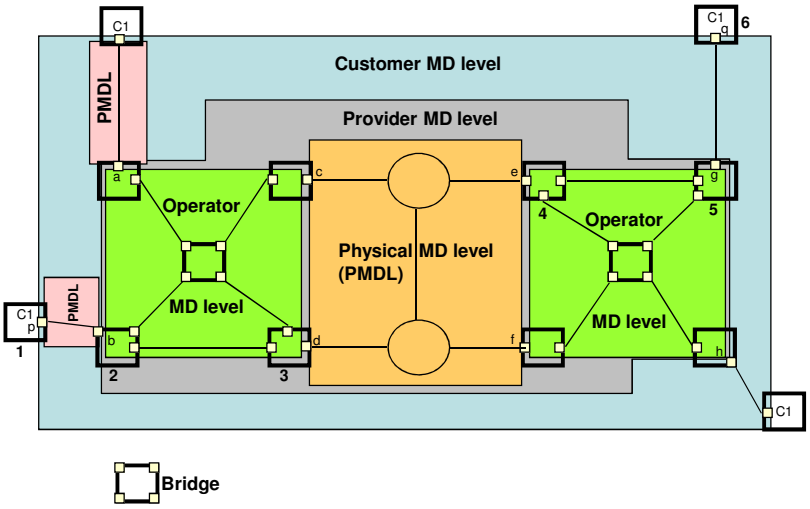
CFM concepts

- **Maintenance Entity** (ME) represents an entity that requires management and is a relationship between two **maintenance entity group** (MEG) end points.
- **ME group** (MEG also known as Maintenance Association MA) includes different MEs that satisfy the following conditions:
  - MEs in a MEG exist in the same administrative boundary
  - MEs in a MEG have the same MEG level
  - MEs in a MEG belong to the same point-to-point connection or multipoint connectivity.
    - For a point-to-point ETH connection, a MEG contains a single ME.
    - For a multipoint ETH connectivity containing n end-points, a MEG contains n*(n–1)/2 MEs.
- In case MEGs are nested, the OAM flow of each MEG has to be clearly identifiable and separable from the OAM flows of the other MEGs. The **MEG level** in the OAM frame distinguishes between the OAM flows of nested MEGs.
- **MEG end point** (MEP) marks the end point of an ETH MEG which is capable of initiating and terminating OAM frames for fault management and performance monitoring.
- **MEG intermediate point** (MIP) is an intermediate point which is capable of reacting to some OAM frames. A MIP does not initiate OAM frames.

266

Leon Bruckman

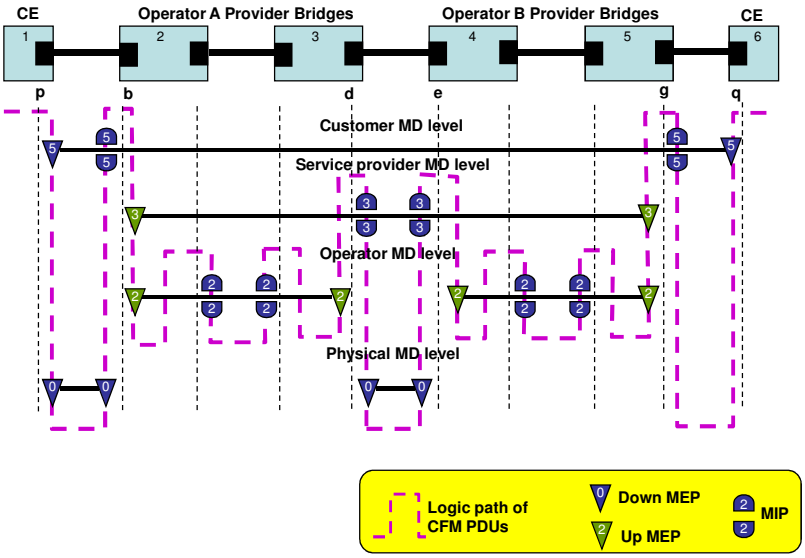
One service instance in a provider network



267

Leon Bruckman

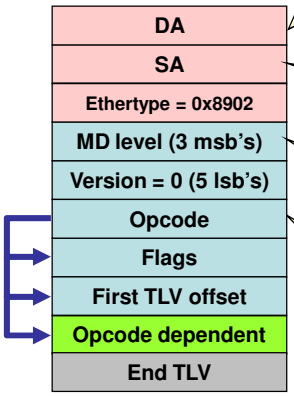
MEPs, MIPs and MD levels



268

Leon Bruckman

CFM PDU



There are three classes of CFM PDUs in terms of the CFM entities to which they are addressed:
a) Those addressed to all MEPs in a service instance (CCM, LBM);
b) Those addressed to the set of MPs immediately adjacent to the transmitting MP, at a certain MD Level, in a service instance (LTM)
c) Those addressed to a single, specific MP (LBM, LBR, LTR).

Group DA = 01-80-C2-00-00-3y , y=axxx where xxx is the MD level and a=1 indicates that the group includes all MPs (LTM), and a=0 indicates that only the MEPs are included in the group (CCM)

The Individual MAC address of the MP transmitting the PDU. An MP's MAC address is not necessarily unique; MPs configured on the same Bridge Port can share the same MAC address.

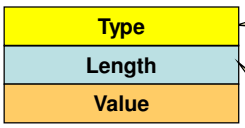
Integer identifying the MD Level of the packet. Higher numbers correspond to higher MEGs, with the highest values for customers' CFM packets. Lower numbers correspond to lower MEG's, with the lowest values for single Bridges or physical links.

CFM PDU or organization	OpCode range
Reserved for IEEE 802.1	0
Continuity Check Message (CCM)	1
Loopback Reply (LBR)	2
Loopback Message (LBM)	3
Linktrace Reply (LTR)	4
Linktrace Message (LTM)	5
Reserved for IEEE 802.1	6 - 31
Defined by ITU-T Y.1731	32 - 63
Reserved for IEEE 802.1.	64 - 255

269

Leon Bruckman

Type, Length, Value (TLV) format



TLV or organization	Type field
End TLV	0
Sender ID TLV	1
Port Status TLV	2
Data TLV	3
Interface Status TLV	4
Reply Ingress TLV	5
Reply Egress TLV	6
LTM Egress Identifier TLV	7
LTR Egress Identifier TLV	8
Reserved for IEEE 802.1	9 - 30
Organization-Specific TLV	31
Defined by ITU-T Y.1731	32 - 63
Reserved for IEEE 802.1	64 - 255

Required if the Type field is not 0. Not present if the Type field is 0. The 16 bits of the Length field indicate the size, in octets, of the Value field. 0 in the Length field indicates that there is no Value field.

270

Leon Bruckman

Continuity Check protocol

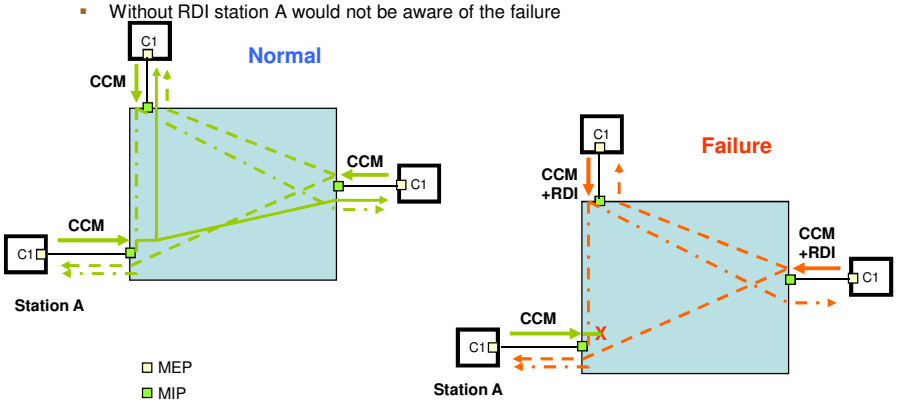
- The Continuity Check Message provides a means to detect connectivity failures in a MEG.
- Each MEP can be configured to periodically transmit a CCM.
  - Period between 3+1/3 msec to 10 minutes
- A connectivity failure is then defined as either:
  - The inability of a MEP to receive three consecutive CCMs from any one of the other MEPs in its MEG, indicating either a MEP failure or a network failure
  - The reception by a MEP of a CCM with an incorrect transmission interval, indicating a configuration error
  - The reception by a MEP of a CCM with an incorrect MEP_ID or MEG_ID, indicating a configuration error or a cross connect error
  - The reception by a MEP of a CCM with an MD Level lower than that of the MEP, indicating a configuration error or a cross connect error
  - The reception by a MEP of a CCM containing a Port Status TLV or Interface Status TLV indicating a failed Bridge Port or aggregated port

271

Leon Bruckman

CCM and Remote Defect Indication (RDI)

- The Remote Defect Indication (RDI), a single bit, is carried by the CCM.
- The absence of RDI in a CCM indicates that the transmitting MEP is receiving CCMs from all configured MEPs
  - Without RDI station A would not be aware of the failure

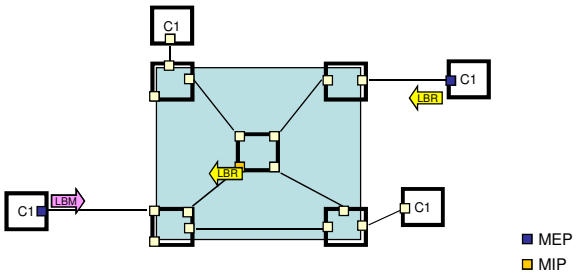


272

Leon Bruckman

Loopback protocol - Unicast

- A unicast Loopback Message (LBM) is used for Fault verification and isolation.
  - A MEP can be instructed by a system administrator to issue one or more LBMs.
- The LBM is initiated by a MEP with specified DA, priority, and drop eligible parameters, the DA being the Individual MAC address of another MP (MEP or MIP) within the same MEG as the transmitting MEP.
- The receiving MP (MEP or MIP) responds to the LBM with a unicast Loopback Reply (LBR).



273

Leon Bruckman

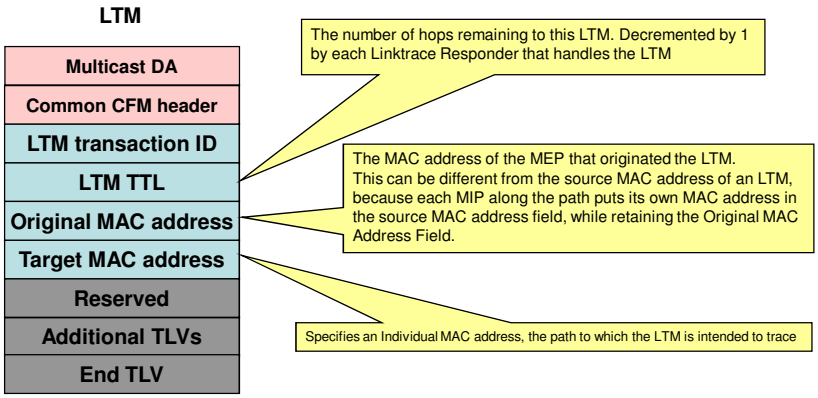
Link Trace protocol

- Link trace function (LT) is an on-demand OAM function that can be used for the following two purposes:
  - **Adjacent relation Retrieval** –LT function can be used to retrieve adjacency relationship between a MEP and a remote MEP or MIP. The result of running the LT function is a sequence of MIPs from the source MEP until the target MIP or MEP. Each MIP and/or MEP is identified by its MAC address.
  - **Fault localization** – LT function can be used for fault localization. When a fault (e.g., a link and/or a device failure) or a forwarding plane loop occurs, the sequence of MIPs and/or MEP will likely be different from the expected one. Difference in the sequences provides information about the fault location.
- After transmitting a LT Message (LTM) frame, the MEP expects to receive frames with LT reply (LTR) information within a specified period of time.
- A network element containing MIP or MEP responds with a LTR frame upon receiving a valid LTM frame (including TTL>0) only if:
  - The network element where the MIP or MEP resides is aware of the Target MAC address in the LTM and associates it to a single egress port, where the egress port is not the same as the port on which the LTM frame was received; or
  - The Target MAC address is the same as the MIP's or MEP's own MAC address.
- In addition, if the MP through which the LTM was received was a MIP, the LT Responder forwards (unless it is the Target MAC) an altered version of the LTM out of a single Bridge Port in the direction of the target MAC address.
- As each network element containing the MIPs or MEP needs to be aware of the Target MAC address in the received LTM frame and associates it to a single egress port, in order for the MIP or MEP to reply, a Unicast LB to the Target MAC address could be performed by a MEP before transmitting the LTM frame.
  - This would ensure that the network elements along the path to the Target MAC address would have information about the route to the Target MAC address if the Target MAC address is reachable in the same MEG.

274

Leon Bruckman

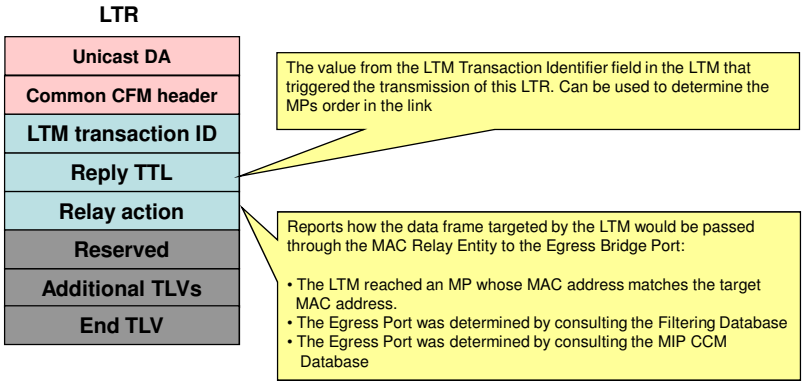
Link Trace message format



275

Leon Bruckman

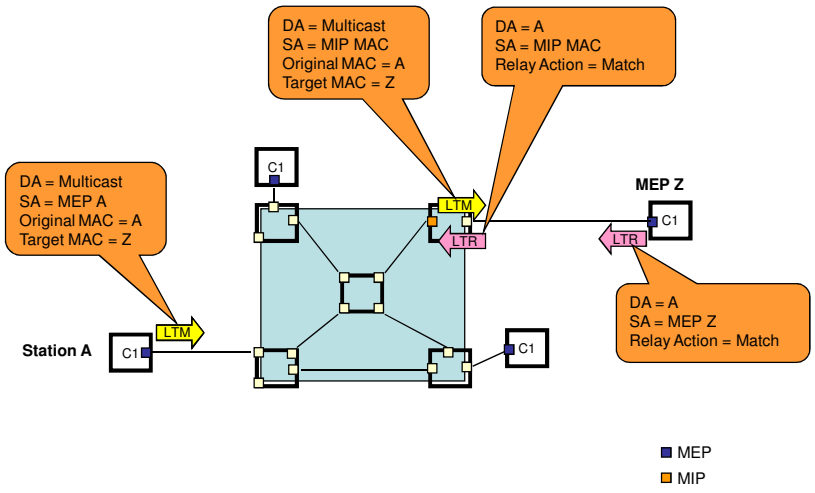
Link Trace reply format



276

Leon Bruckman

Linktrace operation



277