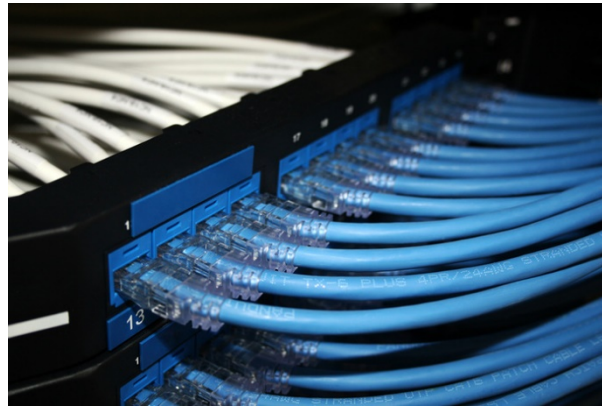# Data Center Ethernet

Raj Jain

Washington University in Saint Louis

Saint Louis, MO 63130

Jain@cse.wustl.edu

These slides and audio/video recordings of this class lecture are at:

http://www.cse.wustl.edu/~jain/cse570-19/

Overview

1. Residential vs. Data Center Ethernet
2. Review of Ethernet Addresses, devices, speeds, algorithms
3. Enhancements to Spanning Tree Protocol
4. Virtual LANs

Quiz: True or False?

Which of the following statements <u>are</u> <u>generally</u> true?

T F

❑❑ Ethernet is a local area network (Local $\leq$ 2km)

❑❑ Token ring, Token Bus, and CSMA/CD are the three most common LAN access methods.

❑❑ Ethernet uses CSMA/CD.

❑❑ Ethernet bridges use spanning tree for packet forwarding.

❑❑ Ethernet frames are 1518 bytes.

❑❑ Ethernet does not provide any delay guarantees.

❑❑ Ethernet has no congestion control.

❑❑ Ethernet has strict priorities.

Residential vs. Data Center Ethernet

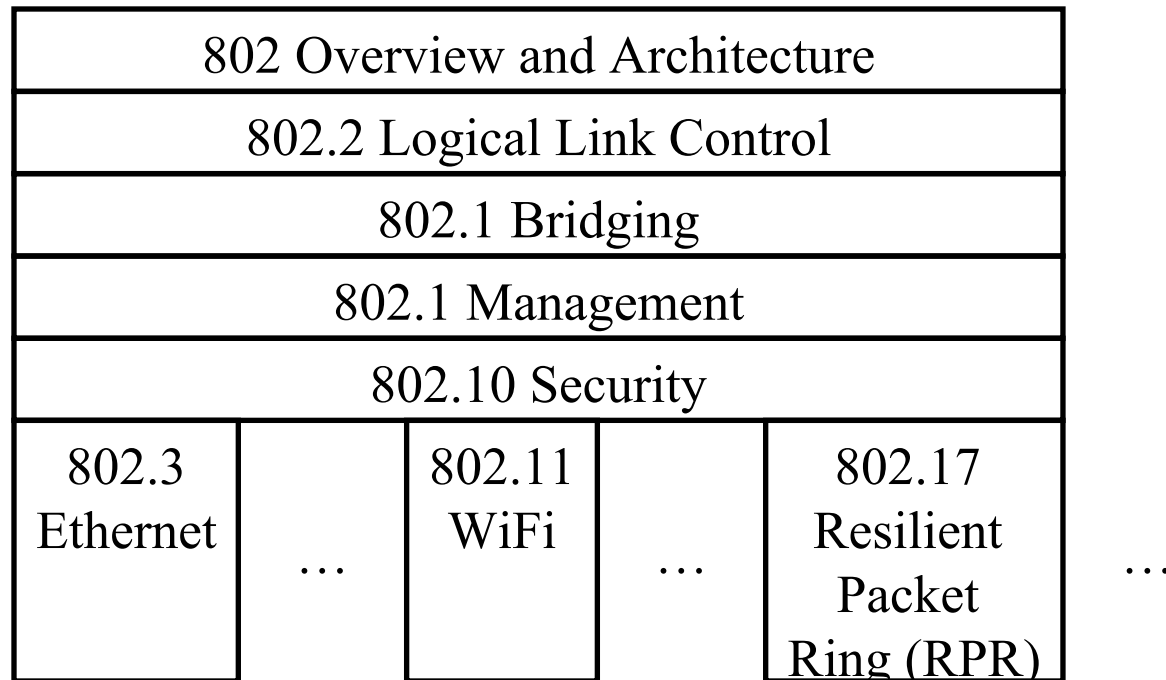| Residential	Data Center
❑ Distance: up to 200m	❑ No limit
❑ Scale: ➢ Few MAC addresses ➢ 4096 VLANs	❑ Millions of MAC Addresses ❑ Millions of VLANs Q-in-Q
❑ Protection: Spanning tree	❑ Rapid spanning tree, … (Gives 1s, need 50ms)
❑ Path determined by spanning tree	❑ Traffic engineered path
❑ Simple service	❑ Service Level Agreement. Rate Control.
❑ Priority ⇒ Aggregate QoS	❑ Need per-flow/per-class QoS
❑ No performance/Error monitoring (OAM)	❑ Need performance/BER

IEEE 802 Address Format

t 48-bit:1000 0000 : 0000 0001 : 0100 0011
: 0000 0000 : 1000 0000 : 0000 1100
= 80:01:43:00:80:0C

Organizationally Unique Identifier (OUI)			24 bits assigned by OUI Owner
Individual/ Group	Universal/ Local		
1	1	22	24

❑ Multicast = "To all bridges on this LAN"

❑ Broadcast = "To all stations" (Note: Local bit is set)
= 111111....111 = FF:FF:FF:FF:FF:FF

IEEE Standards Numbering System

❑ IEEE 802.* and IEEE 802.1* standards (e.g., IEEE 802.1Q-2018) apply to all IEEE 802 technologies:
  ➢ IEEE 802.3 Ethernet
  ➢ IEEE 802.11 WiFi

802 Overview and Architecture					
802.2 Logical Link Control					
802.1 Bridging					
802.1 Management					
802.10 Security					
802.3 Ethernet	…	802.11 WiFi	…	802.17 Resilient Packet Ring (RPR)	…

IEEE Standards Numbering (Cont)

❑ IEEE 802.3* standards apply only to Ethernet,
e.g., IEEE802.3ba-2010

❑ Standards with all upper case letters are base standards
E.g., IEEE 802.1AB-2009

❑ Standards with lower case are additions/extensions/revisions.
Merged with the base standard in its next revision.
e.g., IEEE 802.1w-2001 was merged with IEEE 802.1D-2004

❑ Standards used to be numbered, sequentially, e.g., IEEE
802.1a, …, 802.1z, 802.1aa, 802.1ab, …

❑ Recently they started showing base standards in the additions,
e.g., IEEE 802.1Qau-2010

Names, IDs, Locators
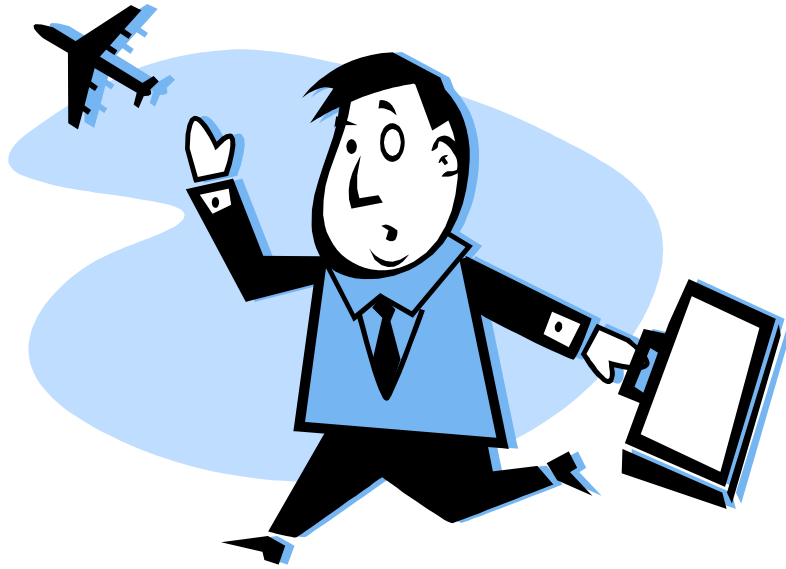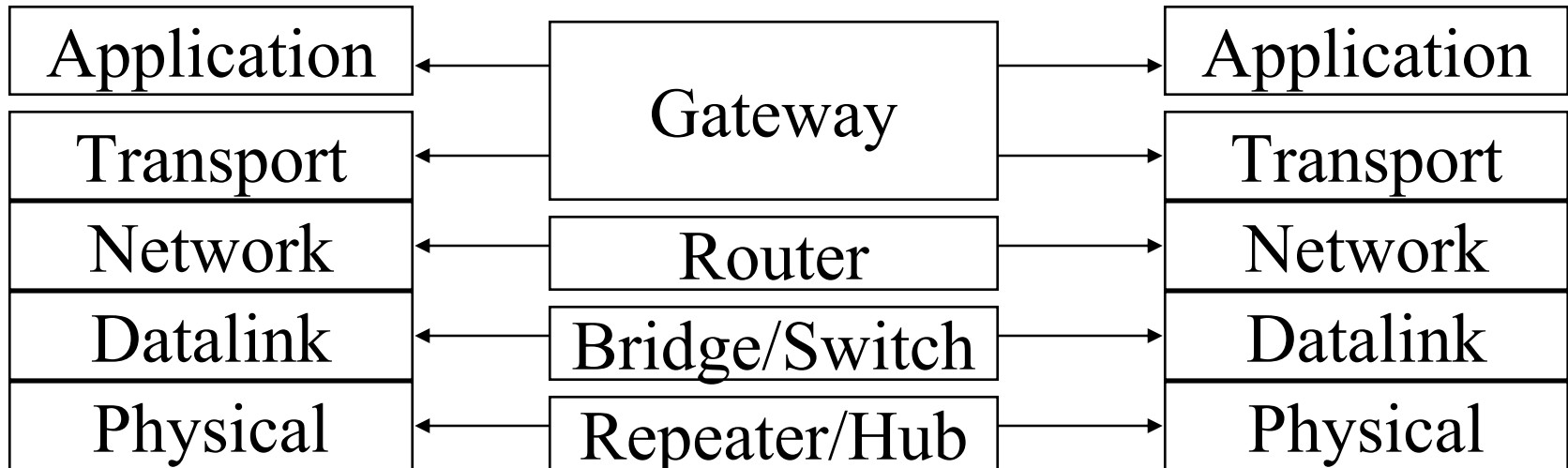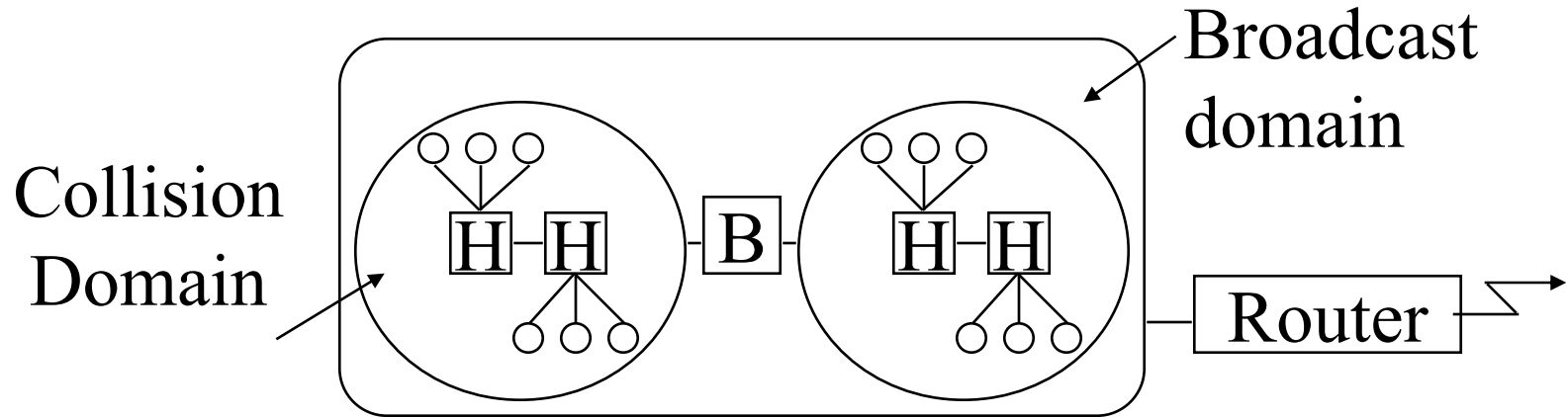
Name: John Smith

ID: 012-34-5678

Locator:
1234 Main Street
Big City, MO 12345
USA

❑ Locator changes as you move, ID and Names remain the same.

❑ **Examples**:

➢ Names: Company names, DNS names (Microsoft.com)

➢ IDs: Cell phone numbers, 800-numbers, Ethernet addresses, Skype ID, VOIP Phone number

➢ Locators: Wired phone numbers, IP addresses

Interconnection Devices

Broadcast domain

Collision Domain

H–H B H–H Router

Application		Application
Transport	Gateway	Transport
Network	Router	Network
Datalink	Bridge/Switch	Datalink
Physical	Repeater/Hub	Physical

Interconnection Devices (Cont)

❑ **Repeater**: PHY device that restores data and collision signals

❑ **Hub**: Multiport repeater + fault detection and recovery

❑ **Bridge**: Datalink layer device connecting two or more collision domains. MAC multicasts are propagated throughout the LAN.

❑ **Router**: Network layer device. IP, IPX, AppleTalk.
Does not propagate MAC multicasts.

❑ **Switch**: Multiport bridge with parallel paths

❑ These are functions. Packaging varies.

Ethernet Speeds

❑ IEEE 802.3cu is working on 400G Ethernet standard. Ethernet Alliance is discussing 800G/1.6T standards

❑ 10Mbps, 100 Mbps, 1 Gbps versions have both CSMA/CD and Full-duplex versions

❑ No CSMA/CD in 10G and up

❑ No CSMA/CD in practice now even at home or at 10 Mbps

❑ 1 Gbps in residential, enterprise offices

❑ 10 Gbps in Data centers, moving to 40 Gbps and 100 Gbps

❑ 100G in some carrier core networks
100G is still more expensive than 10×10G

❑ Note: only decimal **bit** rates are used in networking
No cheating like binary byte values used in storage
1 Gbps = 10^9 b/s, Buy 256 GB Disk = 238.4 GB storage

IS-IS Protocol
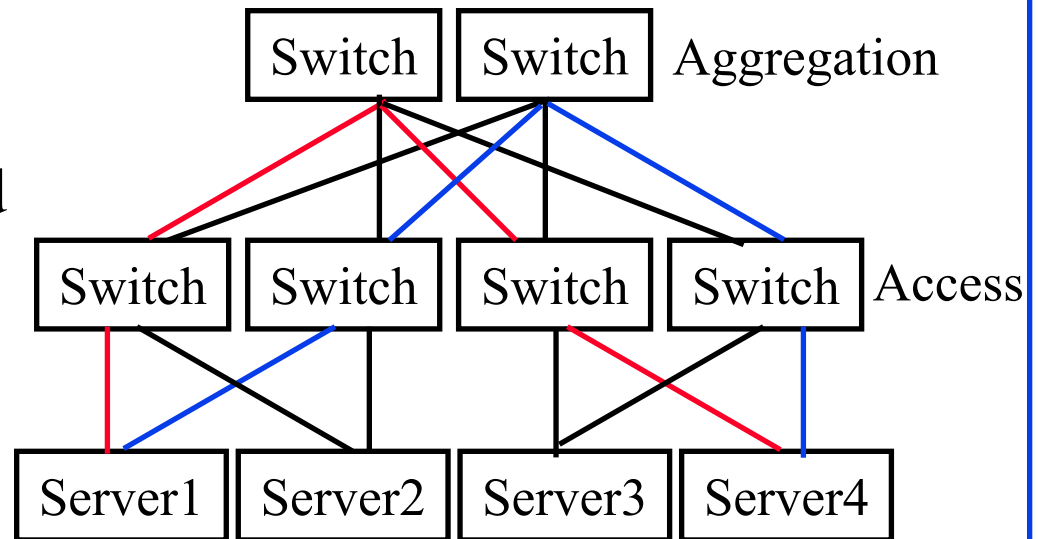
❑ Intermediate System to Intermediate System (IS-IS) is a protocol to build routing tables. Link-State routing protocol => Each nodes sends its connectivity (link state) information to all nodes in the network

❑ Dijkstra's algorithm is then used by each node to build its routing table.

❑ Similar to OSPF (Open Shortest Path First).

❑ OSPF is designed for IPv4 and then extended for IPv6.
IS-IS is general enough to be used with any type of addresses

❑ OSPF is designed to run on the top of IP
IS-IS is general enough to be used on any transport
⇒ Adopted by Ethernet

Washington University in St. Louis   http://www.cse.wustl.edu/~jain/cse570-19/   ©2019 Raj Jain

Shortest Path Bridging

❑ IEEE 802.1aq-2012 (later incorporated in 802.1Q-2014)

❑ Allows all links to be used ⇒ Better CapEx

❑ IS-IS link state protocol (similar to OSPF) is used to build shortest path trees for each node to every other node within the SPB domain
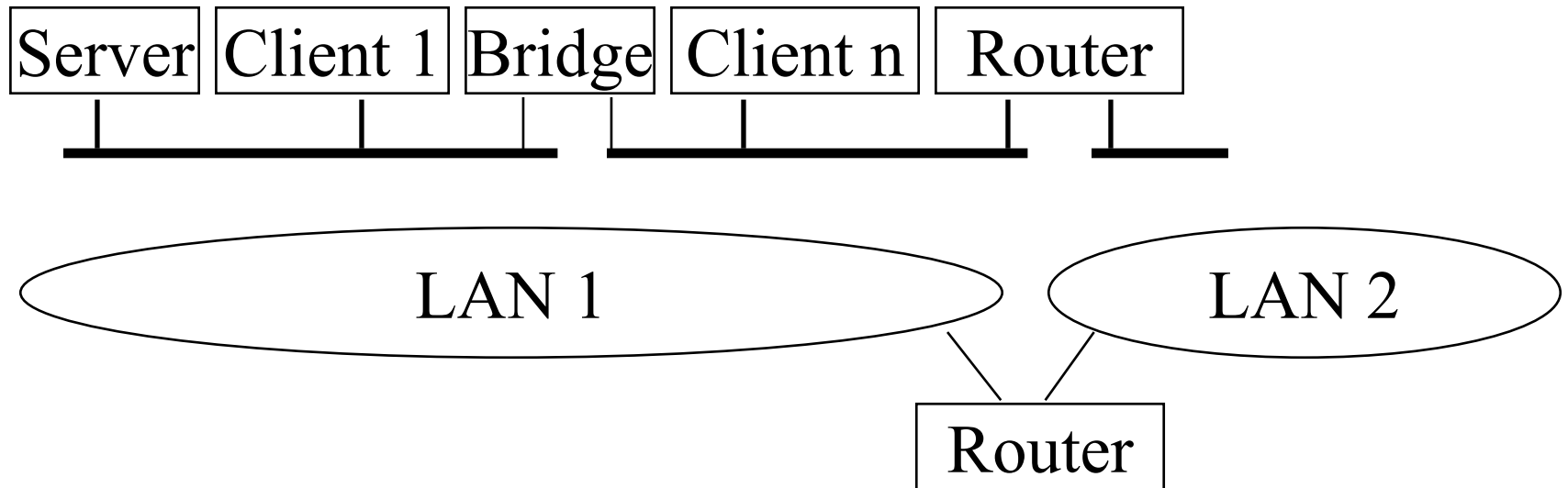
❑ Equal-cost multi-path (ECMP) used to distribute load

| Switch | Switch | Aggregation |

| Switch | Switch | Switch | Switch | Access |

| Server1 | Server2 | Server3 | Server4 |

Ref: http://en.wikipedia.org/wiki/Shortest_Path_Bridging

http://www.cse.wustl.edu/~jain/cse570-19/

©2019 Raj Jain

What is a LAN?

Server	Client 1	Bridge	Client n	Router

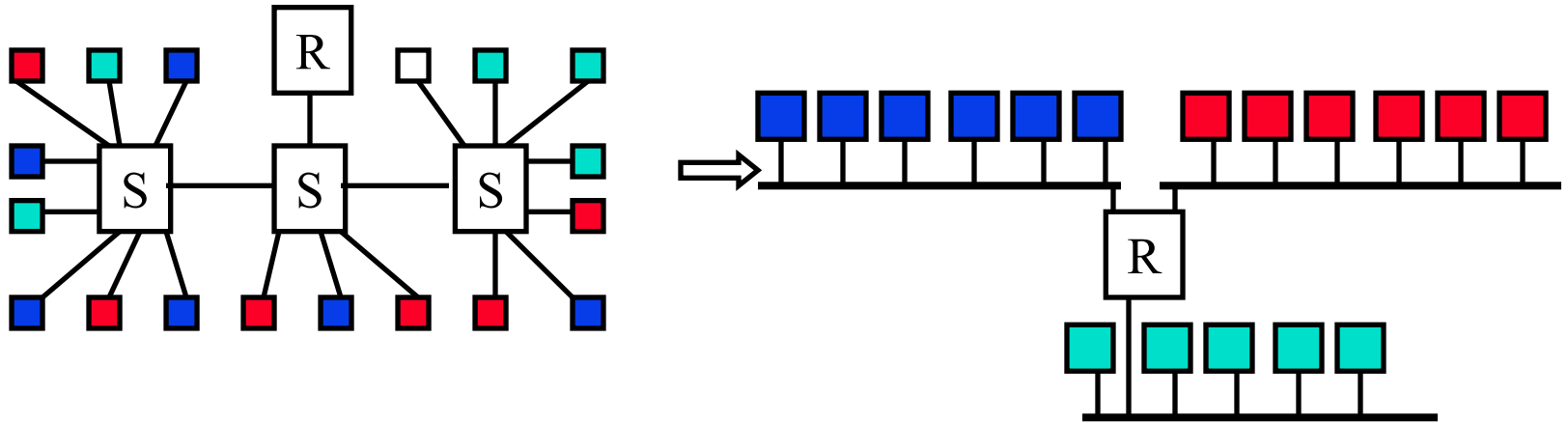LAN 1   LAN 2

Router

❑ LAN = Single broadcast domain = Subnet

❑ No routing between members of a LAN

❑ Routing required between LANs

Virtual LAN

❑ Virtual LAN = Broadcasts and multicast goes only to the nodes in the virtual LAN

❑ LAN membership defined by the network manager ⇒ Virtual

IEEE 802.1Q-2011 Tag

❑ Tag Protocol Identifier (TPI)

❑ Priority Code Point (PCP): 3 bits = 8 priorities 0..7 (High)

❑ Canonical Format Indicator (CFI): 0 ⇒ Standard Ethernet, 1 ⇒ IBM Token Ring format (non-canonical or non-standard)

❑ CFI now replaced by Drop Eligibility Indicator (DEI)

❑ VLAN Identifier (12 bits ⇒ 4095 VLANs)

❑ Switches forward based on MAC address + VLAN ID Unknown addresses are flooded.

Untagged Frame

DA	SA	T/L	Data	CRC

32b IEEE 802.1Q-2011 Header

Tagged Frame

DA	SA	TPI	Priority	CFI/DEI	VLAN ID	T/L	Data	CRC
48b	48b	16b	3b	1b	12b	16b		32b

Ref: Canonical vs. MSB Addresses, http://support.lexmark.com/index?page=content&id=HO1299
Ref: G. Santana, "Data Center Virtualization Fundamentals," Cisco Press, 2014, ISBN:1587143240

Summary

1. Ethernet's use of IDs as addresses makes it very easy to move systems in the data center $\Rightarrow$ Keep traffic on the same Ethernet

2. Spanning tree is wasteful of resources and slow.
   Ethernet now uses shortest path bridging (similar to OSPF)

3. VLANs allow different non-trusting entities to share an Ethernet network

List of Acronyms

- BER   Bit Error Rate
- CapEx   Capital Expenditure
- CD   Collision Detection
- CSMA   Carrier Sense Multiple Access with Collision Detection
- DA   Destination Address
- DEI   Drop Eligibility Indicator
- DNS   Domain Name System
- ECMP   Equal-cost multi-path
- GB   Giga Byte
- ID   Identifier
- IP   Internet Protocol
- IEEE   Institution of Electrical and Electronics Engineers
- IS-IS   Intermediate System to Intermediate System
- LAN   Local Area Network

List of Acronyms (Cont)

- MAC   Media Access Control
- ID   Identifier
- IP   Internet Protocol
- IEEE   Institution of Electrical and Electronics Engineers
- IS-IS   Intermediate System to Intermediate System
- LAN   Local Area Network
- MAC   Media Access Control
- OSPF   Open Shortest Path First
- OUI   Organizationally Unique Identifier
- PCP   Priority Code Point
- PHY   Physical layer
- QoS   Quality of Service
- SPB   Shortest Path Bridging
- TPI   Tag Protocol Identifier
- VLAN   Virtual Local Area Network
- VOIP   Voice over IP
- WiFi   Wireless Fidelity

Reading List

❑ G. Santana, "Data Center Virtualization Fundamentals," Cisco Press, 2014, ISBN:1587143240 (Safari Book) (Chapter 3 up to Figure 3.6).

Wikipedia Links

- http://en.wikipedia.org/wiki/10-gigabit_Ethernet
- http://en.wikipedia.org/wiki/100_Gigabit_Ethernet
- http://en.wikipedia.org/wiki/Data_center
- http://en.wikipedia.org/wiki/Data_center_bridging
- http://en.wikipedia.org/wiki/Data_link_layer
- http://en.wikipedia.org/wiki/Ethernet
- http://en.wikipedia.org/wiki/Ethernet_frame
- http://en.wikipedia.org/wiki/Fast_Ethernet
- http://en.wikipedia.org/wiki/Gigabit_Ethernet
- http://en.wikipedia.org/wiki/IEEE_802.1Q
- http://en.wikipedia.org/wiki/IEEE_802.3
- http://en.wikipedia.org/wiki/IS-IS

Wikipedia Links (Cont)

❑ http://en.wikipedia.org/wiki/Organizationally_unique_identifier
❑ http://en.wikipedia.org/wiki/Shortest_Path_Bridging
❑ http://en.wikipedia.org/wiki/Virtual_LAN

Scan This to Download These Slides

Raj Jain
http://rajjain.com

Related Modules

CSE567M: Computer Systems Analysis (Spring 2013),

https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),

https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcgy5e_10TiDw

Wireless and Mobile Networking (Spring 2016),

https://www.youtube.com/playlist?list=PLjGG94etKypKeb0nzyN9tSs_HCd5c4wXF

CSE571S: Network Security (Fall 2011),

https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u

Video Podcasts of Prof. Raj Jain's Lectures,

https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw