Network Protocols – ARP, LAGs & LLDP

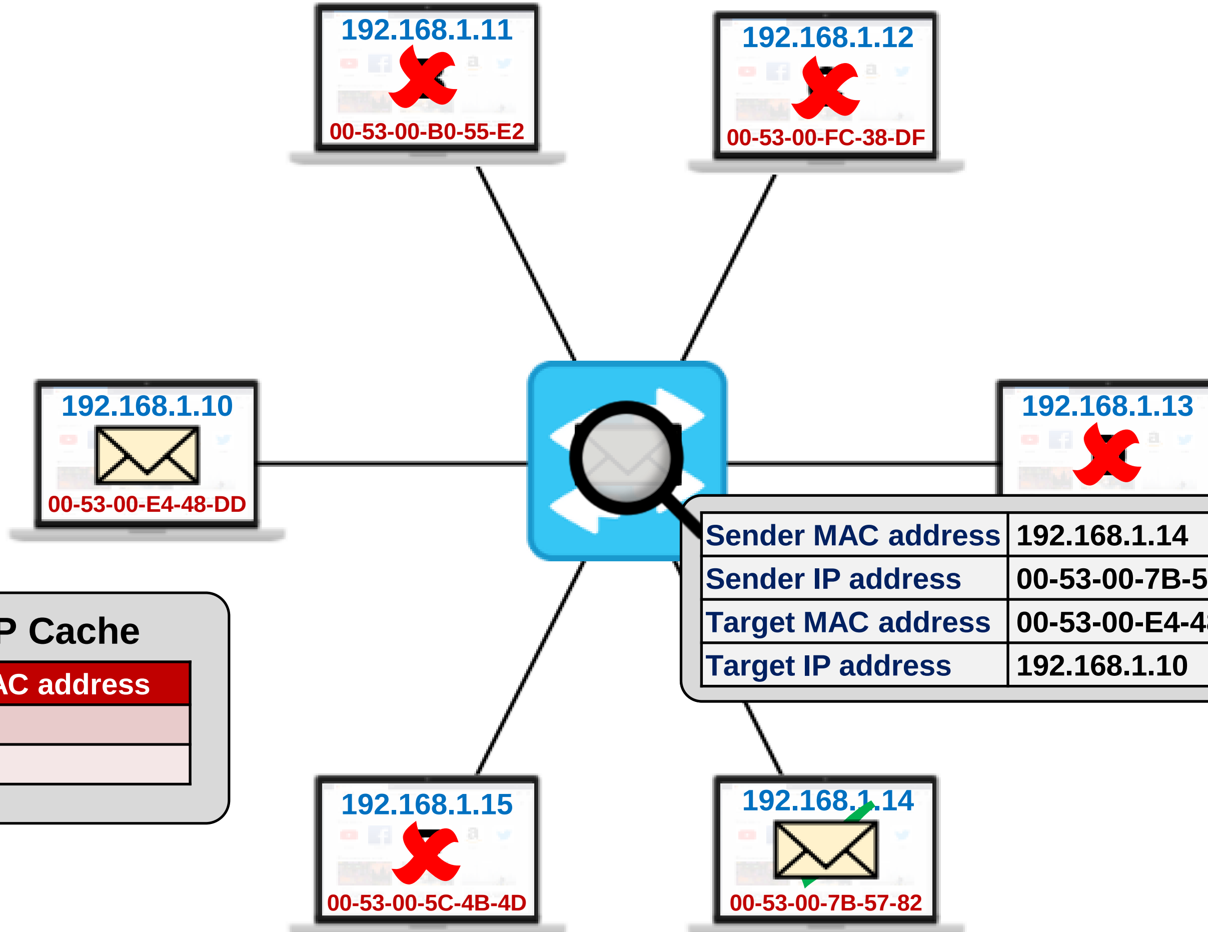
INFO-6078 – Managing Enterprise Networks

FANSHAWE

Address Resolution Protocol (ARP)

- Address resolution protocol (ARP) is a communication protocol used to resolve network layer addresses (IP addresses) into the corresponding data link layer address (MAC address)
- ARP is defined in RFC 826
- ARP operates in layer 2 of the OSI model and is therefore restricted to operate within the local subnet
- Devices that use ARP maintain a cache of mapped addresses, and maintain the cache for a set period
  - On Windows hosts, ARP cache is retained for 2 minutes

ARP Operation



Sender MAC address	192.168.1.14
Sender IP address	00-53-00-7B-57-82
Target MAC address	00-53-00-E4-48-DD
Target IP address	192.168.1.10

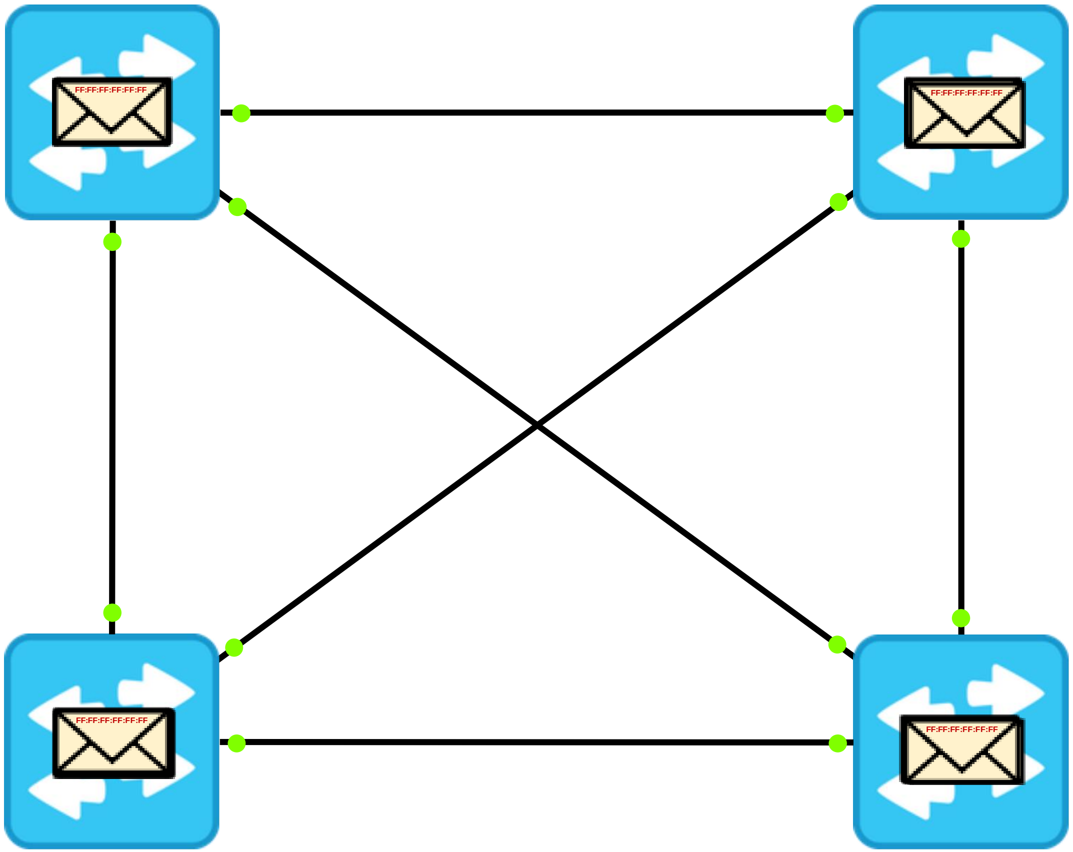
Host A – ARP Cache

IP address	MAC address
192.168.1.14	

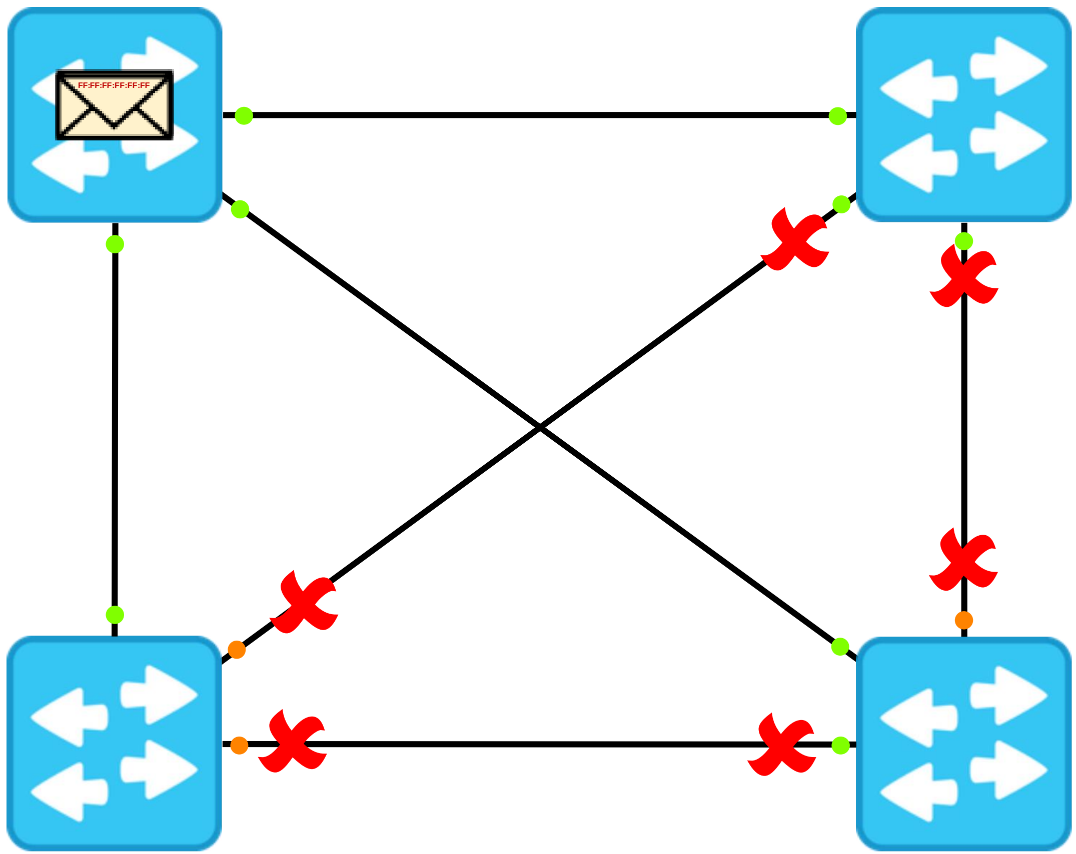
Spanning Tree Protocol (STP)

- Spanning tree protocol (STP) is a communication protocol that is used to detect and remove loops in layer 2 networks
- Spanning tree is described in IEEE 802.1D
- Spanning tree allows a network to have redundant links without creating broadcast storms resulting from broadcast radiation
- When spanning tree is active, redundant links are turned to standby mode to prevent a loop forming

Network Without STP – Broadcast Radiation



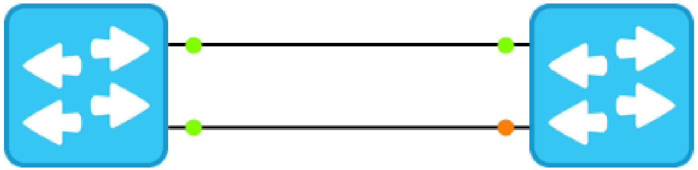
Network With STP



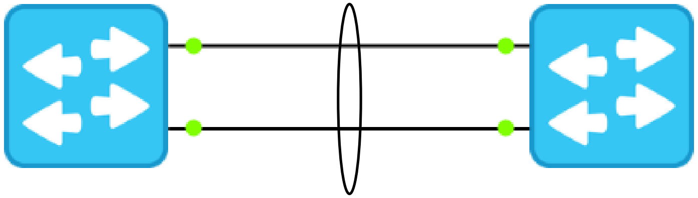
Link Aggregation

- Spanning Tree provides a means to keep the network running in the event of a link failure but does so at a cost
- If a device has redundant links, spanning tree will block one link rendering the bandwidth of the blocked link unusable
- Link aggregation allows the device to combine the links and present them as one physical link, thus allowing both ports to forward traffic

Spanning Tree with Redundant Links



Spanning Tree with LAG



Link Aggregation

- Link aggregation refers to the process of combining multiple physical connections or interfaces into a virtual connection called a Link Aggregation Group (LAG) such that a MAC client can treat it as a single link
- A LAG can provide increased throughput, and/or redundancy should a single link fail
- Link aggregation can be an effective and budget friendly way to incrementally increase bandwidth of the network backbone while also increasing overall reliability
- Most backbone installations includes additional unused cabling

Link Aggregation

- Link Aggregation Control Protocol (LACP) is the open standard for establishing aggregation defined in IEEE's 802.1AX standard (previously 802.3ad)
- Proprietary link aggregation protocols include Cisco's EtherChannel, Port Aggregation Protocol (PAgP) and Virtual Port Channeling (vPC) or Juniper's Aggregated Ethernet
- Link aggregation may be know by many different names such as NIC/Port bonding (Linux), NIC teaming (Microsoft) or port bundling
- On Cisco devices a LAG is known as a port-channel

LAG Bandwidth

- IEEE standards require that all interfaces participating in a LAG operate in full-duplex mode and provide the same bandwidth (Fast Ethernet, Gigabit Ethernet, etc.)
- It is not important that interfaces share a common media
- In most implementations, LAGs will share a common meda

Link Aggregation Load Balancing

- Devices running different platforms or from different vendors can take any number of variables into consideration when determining how to place traffic on LAG members
- Some of the variables include:
  - Source and destination Interface
  - Source and destination MAC address
  - Source and destination IP address
  - Source and destination port numbers
  - VLAN Tag

FANSHAWE

Link Aggregation Load Balancing

- These variables are combined with the aide of a hashing function to identify a flow
- The frame would then be placed on a specific LAG member for the duration of the flow
- Using a single link for each flow also means that the hosts communicating cannot exceed the signaling rate of a single link (e.g. 1 Gbit/s)
- This also means that transmitted frames are less likely to arrive out-of-order and reduces overall overhead on the receiving host

Link Aggregation Load Balancing

- Implementations that send traffic based on Layer 2 or Layer3 hashes often favor a limited number of channels
- This results in only some of the links that make up the LAG being fully utilized, while others go under-utilized unused
- To achieve a more balanced distribution of traffic, it is recommended to use Layer 4 hashing that includes TCP/UDP port numbers

Link Aggregation and Spanning Tree

- When a LAG participates in a spanning tree enables network, the logical LAG interface is the only interface presented to spanning tree
- The individual ports do not appear in the spanning tree topology
- If multiple redundant LAGs are configured between devices, spanning tree will disable one of the links
- Normally, spanning tree adjusts the cost associated with the LAG to reflect the increased bandwidth

Link Aggregation – Static vs Dynamic

- LAGs may be either configured statically or negotiated dynamically using a protocol
- When configured statically, a virtual interface is created and each member of the LAG is added to the group
- Static LAG configuration can introduce a loop into the network if improperly created or modified at a later date
- Using a combination of static and dynamic configurations will fail to create a LAG

Link Aggregation Control Protocol

- LACP is the protocol defined by the IEEE in the 802.1AX specification
- LACP provides a method to automatically negotiate a LAG between devices from a range of manufacturers
- If a physical link in the LAG goes down, the group only loses the bandwidth of the missing link
- When the physical link is restored, it is dynamically re-added to the LAG

LACP – Interface States

- In LACP, ports can be configured in either an active or passive state
  - **Active interfaces** – Actively participate in LAG negotiation by sending LACP PDUs at a rate of one PDU per second
  - **Passive interfaces** – Will only send LACP PDUs in response to a PDU received from another device

Port State Combinations

	Port 1 Active	Port 1 Passive	Port 1 On	Port 1 No Config
Port 2 Active	✓	✓	✘	✘
Port 2 Passive	✓	✘	✘	✘
Port 2 On	✘	✘	✓	✘
Port 2 No Config	✘	✘	✘	✘

LACP – LAG Formation and Maintenance

- When actively synchronizing, LACP sends a PDU out every interface participating in the group every second until the LAG is formed
- The device on the other side of the link will also send PDUs out every interface in the group enabling both devices to identify all participating interfaces and combine them into one logical link
- LACP PDUs are sent to the multicast MAC address 01:80:c2:00:00:02

LACP – LAG Formation and Maintenance

- When synchronization is complete, the LAG interfaces exchange PDUs every thirty seconds to keep the LAG alive
- If the interface does not receive a PDU from its neighbor in within ninety seconds the link is considered failed and enters a down state
- The LACP timers can also be adjusted so that maintenance PDUs are sent every second (fast mode)

Linux Bonding

- The Linux Bonding driver allows for the creation of LAGs using multiple interfaces called slaves into a single interface called a bonded interface
- While the terminology may differ from other devices, the process for defining and creating a LAG is very similar

Linux Bonding Modes

- **Mode 0 (balance-rr) Round-robin**
  - Transmits in sequential order from the first slave to the last
- **Mode 1 (active-backup)**
  - A single active link with backup links ready to take over should the active link fail
  - Can be used in a multi switch setup
- **Mode 2 (balance-xor) XOR**
  - Utilizes a Layer 2 hash to determine interface preference
- **Mode 3 (broadcast)**
  - All traffic is transmitted out all slave interfaces
  - Special use cases only

Linux Bonding Modes

- **Mode 4 (802.3ad) LACP**
  - Dynamic link aggregation compliant with 802.3ad
- **Mode 5 (balance-tlb) Adaptive transmit load balancing**
  - Outgoing traffic is distributed across slaves
  - Incoming traffic is received only by the active slave
- **Mode 6 (balance-alb) Adaptive load balancing**
  - Includes balance-tlb plus receive load balancing (rlb)
  - Receive load balancing is achieved by ARP intercept

FANSHAWE

Microsoft NIC Teaming

- Starting with Windows Server 2012, Microsoft has included NIC Teaming with its server operating systems
- Any network interface that has passed the Windows Hardware Qualification and Logo test (WHQL tests) can be used as a team member in a team interface

NIC Teaming Modes

- **Static Teaming (switch dependent)**
  - A generic mode that presents all the members of the Nic team to the switch as a single interface
- **Switch Independent**
  - Can operate in active-active mode, or active-passive mode
  - Outbound traffic is distributed based on address hash
  - Only one interface receives incoming traffic
  - Alternatively, load balancing can occur based on Hyper-V port for both incoming and outgoing traffic
- **LACP**
  - Compliant with 802.3ax

Link Layer Discovery Protocol (LLDP)

- Network devices use the Link Layer Discovery Protocol to identify neighbor devices on the LAN and to advertise their own presence to neighbors
- LLDP is specified in IEEE 802.1AB and is also known as Station and Media Access Control Connectivity Discovery
- LLDP is an open protocol that was designed to provide similar functionality to that of proprietary protocols such as Cisco Discovery Protocol, Foundry Discovery Protocol, Nortel Discovery Protocol, etc.

Link Layer Discovery Protocol (LLDP)

- Information gathered by LLDP is stored in a Management Information Database (MIB) and is available for remote access through SNMP.
- LLDP can share information related to:
  - System Name/Description
  - Port Names/Descriptions
  - VLAN Names
  - Management Addresses
  - System Capabilities
  - MAC Information
  - Power over Ethernet (PoE)
  - Link Aggregation

LLDP Frames

- LLDP information is sent out of all LLDP enabled ports every 30 seconds by default in the form of an LLDP Data Unit (LLDPDU)
- Each frame is a sequence of type-length-value (TLV) structures

6	6	2						4
Destination MAC 01:80:c2:00:00:0e	**Source MAC**	**Type 0X88 cc**	**Chassis ID TLV**	**Port ID TLV**	**TTL TLV**	**Optional TLVs**	**End of LLDPDU TLV**	**FCS**

LLDP Frames

- Normally the destination of LLDPDUs is to the multicast MAC address 01:80:c2:00:00:0e, but other multicast or unicast addresses are allowed
- Each LLDPDU must contain the mandatory TLVs Chassis ID, Port ID and Time-to-Live
- Each TLV field contains a 7-bit Type value, followed by a 9-bit Length value, followed by a value of variable length from 0 to 511 bytes
- Type 127 identifies custom TLVs know as Organizationally Specific TLVs
- Custom TLVs are formatted as a 24-bit organizationally unique identifier, followed by a 1-byte subtype, followed by the value data.

LLDP TLV Types

#	Description
0	End of LLDPDU
1	Chassis ID
2	Port ID
3	Time-to-Live
4	Port Description
5	System Name
6	System Description
7	System Capabilities
8	Management Address
9-126	Reserved
127	Custom

Mandatory Optional

References

- https://1.ieee802.org/tsn/802-1ax-rev/
- https://thenetworkway.wordpress.com/2015/05/01/an-overview-of-link-aggregation-and-lacp/
- https://www.alliedtelesis.com/sites/default/files/documents/feature-guides/linkag_feature_overview_guide.pdf
- https://docs.microsoft.com/en-us/windows-server/networking/technologies/nic-teaming/nic-teaming-settings