Troubleshooting with Layer 2 Control Protocols

June 15, 2016

Werner Fischer
Principal Networking Consultant | avodaq AG

SharkFest ‘16 • Computer History Museum • June 13-16, 2016
• Background
• Gotchas and Challenges with Layer 2 Control Protocols (L2CP)
• Layer 2
  • LACP
  • UDLD
  • Configuration Test Protocol (loopback)
  • Ethernet Flow-Control
• Between the lines
• Wrap-up
Background
About me

• From Germany (sorry the accent)
• More than 10 year Dual-CCIE (R/S, Security)
• Sniffer Certified Master
• Wireshark Certified Network Analyst
• VMware Certified Professional
• IPv6 Forum Certified Engineer (Gold)
• More than 18 years in the networking area
My first data network analyzer

• Wandel & Goltermann DA-30C – still working 😊
Gotchas & Challenges with L2CP

Capture Files:
https://app.box.com/v/sharkfest2016-layer2
Interference

- Physical Layer (1)
- Data Link Layer (2)
- Network Layer (3)
- Transport Layer (4)
BTW - what is a Link with Ethernet?

- Speed
- Duplex
- MTU
- Auto-Negotiation
- Flow-Control
- MDI/MDI-X
- Remote-Fault / Local-Fault / FEFI
- Carrier-Delay
- Debounce Timer
- EEE
Challenge 1 - Different kind of links

- **Copper**
  - 10/100/1000/10000 traffic

- **Fiber**
  - 10BASE-FL
  - 100BASE-FX
  - 1000-BASE-X
  - 10G/40G/100G
  - …
Challenge 1.1 - Copper links

• Copper – 10/100/1000/10000 traffic
• Taps – Gotchas
  • The Tap negotiates separately with each side of the full-duplex link
  • One Link before – with Tap two Links segments
• SPAN – Gotchas

```
SW_2520(eth-25)# monitor
25: Cannot monitor a dynamic LACP trunk.
SW_2520(eth-25)#
```

• Link Loss Carry Forward or Link Failure Propagation
• PoE (e.g for 802.11ac Wave 1 and Wave2 APs)
Challenge 1.2 - Fiber Links

- Duplex/Simplex
- Single strand BiDi
- Power Level / Split Ratio
- Multi-Wavelength Tap (CWDM/DWDM)
- Taps – the best for single data stream
  - Passive Optical Fiber TAPs
- QSFP+
  - BiDi Transceiver
- CFP, CFP2, CFP4, CXP
- SPAN – Gotchas
Challenge 1.3 – DAC & AOC

- Direct-Attach Cable
  - also known as a twinax cable

- Active Optical Cable (AOC)
Challenge 2 – Display Environment

- Know your MAC-Addresses and write it down
- Use aliases and well-known names
- Use Profiles
Challenge 3 – General

• Location
  • Local versus different Data centers
  • LAN versus WAN / MAN
  • Layer 2 VPNs
  • Virtualization

• Time stamping / correlation

• Cluster Systems
  • Multi-Chassis
  • Fabrics
Time for Questions
Link Aggregation Control Protocol (LACP)
**Link Aggregation Control Protocol (LACP)**

- **What is LACP?**
  - A Layer 2 protocol to logically bundle multiple physical ethernet links into one
- **Why LACP?**
  - for increasing bandwidth and build-in redundancy
- **Who need it?**
  - Everyone from the networking field
- **Benefits?**
  - Failover, load-sharing, acting as one
- **Requirements**
  - “… all interfaces in the channel group must be the same type and speed”

SharkFest ‘16 • Computer History Museum • June 13-16, 2016
LACP Standards

  - 802.3ad
- IEEE Std 802.1AX™-2008
  - 802.1AX not 802.3ax
- IEEE Std 802.1AX™-2014
  - Revision of IEEE Std 802.1AX-2008
Terms

• Link Aggregation
• Link Aggregation Group (LAG)
• Link Aggregation Control Protocol (LACP)
• Link Aggregation interface
• Member interface (member link)
• Active, inactive and standby interfaces
• Aggregator port
• Actor / Partner
• Active / Passive
• Upper / Lower threshold for the number of active interfaces
LACP Requirements

• “… all interfaces in the channel group must be the same type and speed”
• “… as either Layer 2 or Layer 3 interfaces”
• the interfaces that participate in a Port-Channel can include both the copper and fiber-optic ports
• interface attributes

• Really – nothing forgotten?
  • Please remember the Slide "what is a Link with Ethernet"
LACP Notes

• Link Aggregation Control and Marker Protocols are encoded with Ethertype
  • 0x8809
  • Destination Multicast MAC Address: 01-80-C2-00-00-02
• multiple physical links to provide a single logical link between exactly two entities
• in LACP there is no explicit confirmation from a neighbor that he had received LACPDUs
• LACP selects a port for each frame
LACP Load-Balancing

- IPv4 packets
- IPv6 packets
- MPLS packets
- Layer 2 Frames except IPv4, IPv6 and MPLS packets
  - TRILL packets
  - FCoE packets

→ The Load-Balancing code is platform dependent and most use a hashing algorithm by the LAG
→ LACP isn't "additive", it's a LB mechanism!
Load-Balancing - Values in the frame/packet header

- Source MAC address
- Destination MAC address
- Source IP address
- Destination IP address
- Source port
- Destination port
- IPv6 Flow label
- MPLS label(s)
LACP – Marker Protocol

- Marker Generator
- Marker Responder
- Wireshark can dissect it
- The 802.3ad standard also provides two methods to ensure that packets are not disordered when moving conversations. They are time-outs and the Marker Generator
- Never captured by me – and I capture very often
LACP – Marker Protocol

• IEEE Standard versus Dissection

Adjusted by Packet Editor - no real frame
Capturing LACP

Keep all your capture points in time sync

System 1

System 2

SharkFest '16 • Computer History Museum • June 13-16, 2016
Capturing LACP - Reality

- Lab Environment
- The Hardware Ethernet Analyzers provides different methods to capture packets inline and full-duplex.
- Copper or Fiber – up to Gigabit
- Wireshark used for further analysis
LACP - Flow Graph

after reload one component
### LACP – Capture and Display Filter

**Compiled Filter Output**

- **Capture filter for selected interfaces:** `ether proto 0x8809`

**LAN-Verbindung**

<table>
<thead>
<tr>
<th>No.</th>
<th>Time</th>
<th>Delta</th>
<th>SRC-MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>12.538944000</td>
<td>0.000000000</td>
<td>00:16:35:b3:71:9f</td>
</tr>
<tr>
<td>4</td>
<td>12.581696000</td>
<td>0.042752000</td>
<td>b0:b2:dc:6c:c3:86</td>
</tr>
<tr>
<td>6</td>
<td>134.384160000</td>
<td>121.802460000</td>
<td>b0:b2:dc:6c:c3:86</td>
</tr>
</tbody>
</table>

- Frame 3: 124 bytes on wire (992 bits), 124 bytes captured
- Slow Protocols
  - Slow Protocols subtype: LACP (0x01)
  - Link Aggregation Control Protocol

**slow**

- eth.src[0;3] == b0:b2:dc

**lacp**

- Frame 3: 124 bytes on wire (992 bits), 124 bytes captured
- Slow Protocols
  - Slow Protocols subtype: LACP (0x01)
  - Link Aggregation Control Protocol

- Packets: 36 • Displayed: 20 (55.6%) • Load time: 0:0.2

---

*SharkFest ‘16 • Computer History Museum • June 13-16, 2016*
## LACP in Wireshark 2.0

### Display Filter Reference: LACP

**Protocol field name:** lACP

**Versions:** 2.0.0 to 2.0.2

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
<th>Type</th>
<th>Versions</th>
</tr>
</thead>
<tbody>
<tr>
<td>lACP.actorInfo</td>
<td>Actor Information</td>
<td>Unsigned integer, 1 byte</td>
<td>2.0.0 to 2.0.2</td>
</tr>
<tr>
<td>lACP.actorInfoLen</td>
<td>Actor Information Length</td>
<td>Unsigned integer, 1 byte</td>
<td>2.0.0 to 2.0.2</td>
</tr>
</tbody>
</table>

---

SharkFest ‘16 • Computer History Museum • June 13-16, 2016
LACP – in Detail

**Link Aggregation Control Protocol**

**LACP Version Number:** 0x01

- **Actor Information:** 0x01
  - **Actor Information Length:** 0x14
  - **Actor System Priority:** 32768
  - **Actor System:** SWITCH-A-SYSTEM (00:13:aa:d9:c7:00)
  - **Actor Key:** 10
  - **Actor Port Priority:** 32768
  - **Actor Port:** 274
  - **Actor State:** 0x7d, LACP Activity, Aggregation, Synchronization, Collecting, Distributing, Defaulted

- **Partner Information:** 0x02
  - **Partner Information Length:** 0x14
  - **Partner System Priority:** 0
  - **Partner System:** 00:00:00:00:00:00 (00:00:00:00:00:00)
  - **Partner Key:** 0
  - **Partner Port Priority:** 0
  - **Partner Port:** 0
  - **Partner State:** 0x00

- **Collector Information:** 0x03
  - **Collector Information Length:** 0x10
  - **Collector Max Delay:** 32768

- **Terminator Information:** 0x00
  - **Terminator Length:** 0x00

**ACTOR**

**PARTNER**

**COLLECTOR**

**TERMINATOR**
### LACP – Flags (Actor)

<table>
<thead>
<tr>
<th>Flag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>0x3d</code></td>
<td>Actor State: LACP Activity, Aggregation, Synchronization, Collecting, Distributing</td>
</tr>
<tr>
<td><code>1</code></td>
<td>LACP Activity: Yes</td>
</tr>
<tr>
<td><code>0</code></td>
<td>LACP Timeout: No</td>
</tr>
<tr>
<td><code>1</code></td>
<td>Aggregation: Yes</td>
</tr>
<tr>
<td><code>1</code></td>
<td>Synchronization: Yes</td>
</tr>
<tr>
<td><code>1</code></td>
<td>Collecting: Yes</td>
</tr>
<tr>
<td><code>1</code></td>
<td>Distributing: Yes</td>
</tr>
<tr>
<td><code>0</code></td>
<td>Defaulted: No</td>
</tr>
<tr>
<td><code>0</code></td>
<td>Expired: No</td>
</tr>
</tbody>
</table>

- **Activity control value for this link.** Active = 1, Passive = 0 (lacp.actorState.activity), 1 Byte
- **Timeout control value for this link.** Short Timeout = 1, Long Timeout = 0 (lacp.actorState.timeout), 1 Byte
- **Aggregatable = 1, Individual = 0** (lacp.actorState.aggregation), 1 Byte
- **In Sync = 1, Out of Sync = 0** (lacp.actorState.synchronization), 1 Byte
- **Collection of incoming frames is:** Enabled = 1, Disabled = 0 (lacp.partnerState.collecting), 1 Byte
- **Distribution of outgoing frames is:** Enabled = 1, Disabled = 0 (lacp.partnerState.distributing), 1 Byte
- **1 = Actor Rx machine is using DEFAULT Partner info, 0 = using info in Rx'd LACPDU** (lacp.partnerState.defaulted), 1 Byte
- **1 = Actor Rx machine is EXPIRED, 0 = NOT EXPIRED** (lacp.partnerState.expired), 1 Byte
LACP – Lua script for the flags

- Lua dissectors are a great way to enhance Wireshark

<table>
<thead>
<tr>
<th>No.</th>
<th>Time</th>
<th>Source</th>
<th>Destination</th>
<th>Protocol</th>
<th>Length</th>
<th>LACP Flags</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>19.000995712</td>
<td>HewlettP_b3:71:9b</td>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>128</td>
<td>**DCSG<em>A,**DCSG</em>A</td>
</tr>
<tr>
<td>8</td>
<td>121.845225892</td>
<td>ZyxelCom_6c:c3:86</td>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>128</td>
<td>EF<em><strong><em>G</em>A,*FDC</strong>S</em></td>
</tr>
<tr>
<td>9</td>
<td>121.848485247</td>
<td>HewlettP_b3:71:9f</td>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>128</td>
<td><strong>DCSG<em>A,<em>F</em></em></strong>G*A</td>
</tr>
<tr>
<td>10</td>
<td>122.037726486</td>
<td>ZyxelCom_6c:c3:87</td>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>128</td>
<td>EF<em><strong><em>G</em>A,*FDC</strong>S</em></td>
</tr>
<tr>
<td>11</td>
<td>122.060262887</td>
<td>HewlettP_b3:71:9c</td>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>128</td>
<td><strong>DCSG<em>A,<em>F</em></em></strong>G*A</td>
</tr>
<tr>
<td>12</td>
<td>122.198438937</td>
<td>ZyxelCom_6c:c3:89</td>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>128</td>
<td>EF<em><strong><em>G</em>A,*FDC</strong>S</em></td>
</tr>
<tr>
<td>13</td>
<td>122.221484901</td>
<td>HewlettP_b3:71:9a</td>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>128</td>
<td><strong>DCSG<em>A,<em>F</em></em></strong>G*A</td>
</tr>
</tbody>
</table>

- Frame 7: 128 bytes on wire (1024 bits), 128 bytes captured (1024 bits)
- Ethernet II, Src: HewlettP_b3:71:9b (00:16:35:b3:71:9b), Dst: Slow-Protocols (01:80:c2:00:00:02)
- Slow Protocols
- Link Aggregation Control Protocol
- LACP Actor Flags
  - LACP Flags: **DCSG*A
- LACP Partner Flags
  - LACP Flags: **DCSG*A

See Lua Script “TCP Flags for Wireshark” by Didier Stevens - modified for LACP
LACP – System-ID

System-ID = System Priority plus System MAC address

- The endpoint with the lower System-ID makes the decision about which ports are actively participating in the port-channel at any given time.
- **The lower the value** becomes the Actor and determines the links between the LACP partner switches that are in active and standby states for each LACP port channel.
- When the System Priority is same, the device with lower System MAC will have higher system-priority.
Actor Key

- Value assigned to aggregator ports and physical ports that are candidates for joining a LAG.
- Only ports with matching keys are allowed to aggregate.
Port-ID = Port Priority plus Port Number

- The lower the range of the Port-ID, the more likely that the interface will be used for LACP transmission
- Port Priority decides which ports should be put in standby mode when there is a limitation that prevents all compatible ports from aggregating and which ports should be put into active mode.
LACP – Actor Election

Which ports are part of the aggregation?
LACP interaction with LLDP

Link Layer Discovery Protocol
- Chassis Subtype = MAC address, Id: 00:1e:58:b4:0f:c3
- Port Subtype = Locally assigned, Id: 1/1
- Time To Live = 120 sec
- Port Description = RMON Port 1 on Unit 1
- System Name =
- System Description = Gigabit Ethernet Switch
- Capabilities
- IEEE 802.3 - MACPHY Configuration/Status
  - IEEE 802.3 - Link Aggregation
    - 111 111. .... .... = TLV Type: Organization Specific (127)
    - .... ....0 0000 1001 = TLV Length: 9
    - Organization Unique Code: IEEE 802.3 (0x00120f)
    - IEEE 802.3 Subtype: Link Aggregation (0x03)
    - Aggregation Status: 0x01
      - .... ...1 = Aggregation Capability: Yes
      - .... ...0 = Aggregation Status: Disabled
    - Aggregated Port Id: 0
  - IEEE 802.3 - Maximum Frame Size
  - End of LLDPDU

SharkFest ‘16 • Computer History Museum • June 13-16, 2016
MC-LAG with LACP

- MC-LAG - LAG terminate on separate chassis
- MC-LAG is not covered under IEEE standard
- Multi-homing for redundancy
- Active-active to utilize all links which otherwise may get blocked by Spanning-Tree
- no modification of LAG partner
- Temporary loops or duplicates not acceptable
- Split brain handling
- One the way for multi-vendor implementation
MC-LAG – different vendors – different names

• Cisco:
  • StackWise
  • Virtual Switching System (VSS)
  • Virtual Port Channel (vPC)

• Juniper
  • Virtual Chassis (VC)

• HP
  • Intelligent Resilient Framework (IRF)

• Extreme Networks
  • Inter-Switch-Connection

• Force10
  • Virtual Link Trunking

• Avaya (Nortel)
  • Split multi-link trunking

• Cumulus Networking
  • Multi-Chassis Link Aggregation

• Arista Networks
  • MLAG

• … and many others
Figure 1. A basic MC-LAG Example

MC-LAG peering information is exchanged and adjacency formed.
MLAG with LACP

A.  

B. Peer-Link

C.  

D. Peer-Link

E.  

F.  

2 tier (2 layer, full mesh)
LACP in Fabric/Cluster/vPC/Grouping/Stacking

SYSTEM 1

SYSTEM 2

SYSTEM 3

SYSTEM 4

SharkFest ‘16 • Computer History Museum • June 13-16, 2016
LACP Challenges from the field
<table>
<thead>
<tr>
<th>Destination</th>
<th>Protocol</th>
<th>Length</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>124</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>126</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>124</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>124</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>126</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>124</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>124</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>126</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>124</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7</td>
</tr>
<tr>
<td>Slow-Protocols</td>
<td>LACP</td>
<td>124</td>
<td>Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7</td>
</tr>
</tbody>
</table>
Timer different – not short or long

<table>
<thead>
<tr>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
<th>Link Aggregation Co</th>
</tr>
</thead>
<tbody>
<tr>
<td>203.88121.6800</td>
<td>204.88121.6800</td>
<td>205.53494.4600</td>
<td>205.88121.6800</td>
<td>206.88124.6000</td>
<td>207.53497.6000</td>
<td>207.88124.6000</td>
<td>208.88128.0000</td>
<td>209.53494.4600</td>
</tr>
</tbody>
</table>

Every two seconds from the link partner
Time for Questions
Unidirectional Link Detection (UDLD)
UDLD Basics

- Cisco UDLD feature is documented in RFC 5171

<table>
<thead>
<tr>
<th>Network Working Group</th>
<th>Cisco Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request for Comments: 5171</td>
<td>April 2006</td>
</tr>
<tr>
<td>Category: Informational</td>
<td>M. Foschino</td>
</tr>
</tbody>
</table>

- Different names and implementations
  - Device Link Detection Protocol (DLDP)
  - D-Link Unidirectional Link Detection (DULD)

- Three Paket Formats
  - Probe
  - Echo
  - Flush
UDLD Basics

• Many vendors have their own proprietary solution
  • LACP protocol in a single member LAG
  • Own Ethertype
• Layer 1 "fault" indication is the "loss of light"
• Why it is needed – we use Auto-Negotiation with Remote-Fault?
  • Different wavelengths of optical signaling (10/100/1000)
  • EoSDH
• Used for miswiring detection
Cisco UDLD Notes

- Cisco UDLD are encoded with LLC, standard Subnetwork Access Protocol (SNAP) format and Protocol ID 0x111
- Destination Multicast MAC Address: 01:00:0C:CC:CC:CC:CC
- Fast Hello enhancement available
UDLD Modes

• Aggressive Mode:
  • UDLD will declare link as unidirectional and will disable interface, if no reply has been received for subsequent 8 PDU message transmitted at an interval of 1 sec.

• Normal Mode:
  • Link will be disabled immediately if PDU reply has not been received within predefine timeout interval.
Field occurrence is very useful in such case. Every udld.data field as a custom column.
UDLD Challenges from the field
UDLD non-Cisco

There is a workaround 😊

Use Wireshark Legacy with Packet Edit
UDLD non-Cisco

Works for this vendor

Maybe a bug – will discuss that with the wireshark developers 😊 afterwards
Time for Questions
Configuration Test Protocol (loopback)
Loop Detection Protocol

- Loop detection protocol
  - Pro Port
  - Pro VLAN (Trunk)
- Ethertype 0x9000
- Different Destination MAC Addresses
  - CF-00-00-00-00-00
  - 01-0F-E2-00-00-07
  - 00-00-F4-27-71-01
  - 01-A0-C5-AA-AA-AB
  - ...

SharkFest ‘16 • Computer History Museum • June 13-16, 2016
Loop Detection Protocols

• CTP from the archive

For more information see http://www.mit.edu/~jhawk/ctp.pdf
Loopback from the field
Loop Detection Protocols

Type: Loopback (0x9000)

Configuration Test Protocol (loopback)
- skipCount: 0
- Relevant function: Unknown (256)
- Function: Unknown (256)

Data (42 bytes)
- Data: 000100000001e58b40fc30fc300000000000000000000...
- [Length: 42]

Packet comments
- [2016-03-26 - Werner Fischer]
  - "config loopdetect mode vlan-based"

Frame 4: 60 bytes on wire (480 bits), 60 bytes captured (480 bits) on interface 0
- Destination: Ethernet-Configuration-Test-protocol-(Loopback) (cf:0:00:00:00:00:00:00:00:00:00:00:00:00)
- Source: D-LinkCo_b4:0f:c3 (00:1:e:58:b:4:0f:c3)
- Type: Loopback (0x9000)

Configuration Test Protocol (loopback)
- skipCount: 0
- Relevant function: Unknown (256)
- Function: Unknown (256)

Data (42 bytes)
- Data: 0001000001e58b40fc30fc300000000000000000000...
- [Length: 42]
Loop Detection Protocols

• Every vendor has its own solution
  • TLV coded

• You should read the HEX-code – also in 2016
Loop Detection Protocols – Port Testing

• Every vendor has its own solution
Time for Questions
Ethernet Flow-Control
Ethernet Flow-Control

• Hard to catch
  • Depends on your capture equipment
• Ethertype 0x8808
• Different Modes
  • No PAUSE
  • Symmetric PAUSE
  • Asymmetric PAUSE
  • Symmetric PAUSE and Asymmetric PAUSE
• With Auto-Negotiation or without it
### Table 37-4—Pause priority resolution

<table>
<thead>
<tr>
<th>Local Device</th>
<th>Link Partner</th>
<th>Local Resolution</th>
<th>Link Partner Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAUSE</td>
<td>ASM_DIR PAUSE</td>
<td>Disable PAUSE Transmit and Receive</td>
<td>Disable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>PAUSE</td>
<td>PAUSE</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0 PAUSE</td>
<td>Disable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1 0 PAUSE</td>
<td>Disable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1 1 PAUSE</td>
<td>Enable PAUSE transmit, Disable PAUSE receive, Disable PAUSE transmit</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1 1 1 PAUSE</td>
<td>Enable PAUSE receive, Disable PAUSE transmit</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0 0 PAUSE</td>
<td>Disable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1 0 PAUSE</td>
<td>Enable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1 1 PAUSE</td>
<td>Enable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0 0 0 PAUSE</td>
<td>Disable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0 1 PAUSE</td>
<td>Enable PAUSE receive, Disable PAUSE transmit</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1 0 PAUSE</td>
<td>Enable PAUSE Transmit and Receive</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1 1 PAUSE</td>
<td>Enable PAUSE Transmit and Receive</td>
</tr>
</tbody>
</table>
Ethernet Flow-Control

- Depending on the bandwidth of the link, the PAUSE frames are sent at a specific interval of time.
- The PAUSE time is measured in units of PAUSE "quanta" and is defined to be 512 bit times
  - Fast Ethernet 5.12µs, 0.512µs for Gigabit Ethernet, 0.0512µs for 10-Gigabit Ethernet, 0.0128µs for 40-Gigabit Ethernet and 0.00512µs for 100-Gigabit Ethernet (e.g. 512Bits/1.000.000.000Bit/sec for GE)
- \( 65535 \times 512 / 1.000.000.000 = 0.03355392 \text{ seconds} = 33.55\text{ms}. \)
Ethernet Flow-Control - Settings

```
# ethtool ens817
Settings for ens817:
  Supported ports: [ FIBRE ]
  Supported link modes: 1000BaseKX/Full
                        10000BaseX/Full
                        10000BaseKR/Full
                        40000BaseCR4/Full
                        40000BaseSR4/Full

Supported pause frame use: Symmetric Receive-only
Supports auto-negotiation: yes
Advertised link modes: 1000BaseKX/Full
                        10000BaseX/Full
                        10000BaseKR/Full
                        40000BaseCR4/Full
                        40000BaseSR4/Full

Advertised pause frame use: Symmetric
Advertised auto-negotiation: yes
Link partner advertised link modes: 40000BaseCR4/Full
Link partner advertised pause frame use: No
Link partner advertised auto-negotiation: Yes
Speed: 40000Mb/s
Duplex: Full
Port: Direct Attach Copper
PHY: 0
Transceiver: internal
Auto-negotiation: on
Supports Wake-on: d
Wake-on: d
Current message level: 0x00000014 (2)
  link isdown
Link detected: yes
```

Routerfish controller FastEthernet 8/9 | b PHY registers
PHY registers:
  Register 0x00: 1000 782D 0000 51E4 01E1 0000 0000 0001
  Register 0x08: 0000 0000 0000 0000 0000 0000 0000 0000
  Register 0x10: 0000 0000 0000 0000 0000 0000 0000 0000
  Register 0x18: 0038 851F 9F00 008A 082B 0000 0000 0000
Bytes_recv 2768829297 Bytes_sent 636868627 Frames_recv 54499753 Frames_sent 7456645
Total_bytes_RX 2768829297 Total_frames_RX 54499753 Bcast_frames_recv 18797931
Mcast_frames_RX 21756491 CRC_err 0 Ovr_size frames 0
Fragments 0 Jabber 0 collision 0
Late_collision 0 64B frame 26323978; 55 127B_frames 1799379
128-255B frames 14967201 256-511B_frames 2319326 512-1023B_frames 723842
1024_maxd_frames 500252 Rx_error 0 Dropped_frames 0
Mcast_frames_tx 1488314 Bcast_frames_tx 2839987 Smll_frame_recv 0

SharkFest ‘16 • Computer History Museum • June 13-16, 2016
100 Mbit/s, 1 Gbit/s, 10 Gbit/s, 40 Gbit/s or 100 Gbit/s?
Flow-Control information with LLDP

- IEEE 802.3 - MAC/PHY Configuration/status
  - 1111 1111 ... ... - TLV Type: Organization Specific (127)
    - 0000 1001 - TLV Length: 9
    - Organization Unique Code: IEEE 802.3 (0x001006)
  - IEEE 802.3 Subtype: MAC/PHY Configuration/Status (0x21)
  - Auto-Negotiation Support/Status: 0x03
    - ... 1 = Auto-Negotiation: Supported
    - ... 0 = Auto-Negotiation: Enabled
  - PMD Auto-Negotiation Advertised Capability: 0x001006
    - ... ... ... 1 = 1000BASE-T (full duplex mode): Not Supported
    - ... ... ... 0 = 1000BASE-T (half duplex mode): Not Supported
    - ... ... ... 0 = 1000BASE-X (LX, SX, CX full duplex mode): Not Supported
    - ... ... ... 0 = 1000BASE-X (LX, SX, CX half duplex mode): Not Supported
    - ... 1 = Asymmetric and Symmetric PAUSE (for full-duplex links): Not Supported
    - ... 0 = Symmetric PAUSE (for full-duplex links): Not Supported
    - ... 0 = Asymmetric PAUSE (for full-duplex links): Not Supported
    - ... 0 = PAUSE (for full-duplex links): Not Supported
    - ... 0 = IEEE 802.3T (full duplex mode): Not Supported
    - ... 0 = IEEE 802.3T (half duplex mode): Not Supported
    - ... 1 = 100BASE-TX (full duplex mode): Not Supported
    - ... 1 = 100BASE-TX (half duplex mode): Supported
    - ... 0 = 100BASE-T4: Not Supported
    - ... 1 = 100BASE-T (full duplex mode): Supported
    - ... 1 = 100BASE-T (half duplex mode): Supported
    - 0 = Other or unknown: Not Supported

- Some information (wrong) bitorder...
  - Operational NDU Type: 1000baseTFO - Four-pair Category 5 UTP, full duplex mode (0x001e)

- IEEE 802.3 - Link Aggregation
PFC – Priority Based Flow-Control

- Frame: 64 bytes on wire (512 bits), 64 bytes captured (512 bits)
- Ethernet II, Src: 0e:fc:00:d5:3d:00 (0e:fc:00:d5:3d:00), Dst: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
  - Type: MAC Control (60000)

- MAC Control
  - Opcode: Class Based Flow Control [CBFC] Pause (0x0101)
    - CBFC Class Enable Vector: 0x0008, C3
      - . . . . . . 0 = C0: False
      - . . . . . . 1 = C1: False
      - . . . . 0 = C2: False
      - . . . . 1 = C3: True
      - . . . . 0 = C4: False
      - . . . . 0 = C5: False
      - . . . . 0 = C6: False
      - . . . . 0 = C7: False

- CBFC Class Pause Times
  - C0: 0
  - C1: 0
  - C2: 0
  - C3: 65535
  - C4: 0
  - C5: 0
  - C6: 0
  - C7: 0
Flow-Control with Copper Taps

• Remember the Taps – Gotchas
  • two Links segments for the Network Ports
  • also two Links for the Monitoring Ports
Flow-Control Challenges from the field
Strange Flow-Control Implementation

Source: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)

[Expert Info (Warn/Protocol): Source MAC must not be a group address: IEEE 802.3-2002, Section 3.2.3(b)]

Address: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
   ...0: InfiniBand: Ignored
   ...1: Globally unique address (factory default)
   ...2: Group address (multicast/broadcast)

Type: MAC Control (0x8000)

Opcode: Pause (0x0001)
pause_time: 65535
Time for Questions
Session Summary

• Wireshark’s capabilities of dissection, filtering and others will help your analysis in a Layer 2 environment
• Pay attention to the capture points and any data that could be used as a “signature” to correlate traces with Layer 2 events
• Time sync of all capture points is a must
• Read standards from IETF / IEEE / MEF – reflex and ask yourself what’s going on the wire here
Please provide Session Feedback

20 - Troubleshooting with Layer 2 Control Protocols

2:15pm - 3:30pm
Grand Hall Classroom

Instructor

Content

Comments (Optional)
Thank You for attending the session and enjoy the rest of the conference / party 😊