SharkFest '16 Europe

Troubleshooting with Layer 2 Control Protocols

2016-10-19

Werner Fischer
Principal Networking Consultant | avodaq AG

#sf16eu

Agenda

- Background
- Gotchas and Challenges with Layer 2 Control Protocols (L2CP)
- Layer 2
  - LACP
  - UDLD
  - Ethernet Flow-Control
  - Loopback
- Between the lines
- Wrap-up

Background

About you / Little Poll

- Who is interested in three-dimensional networking?
- Who visited SharkFest US this year?

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

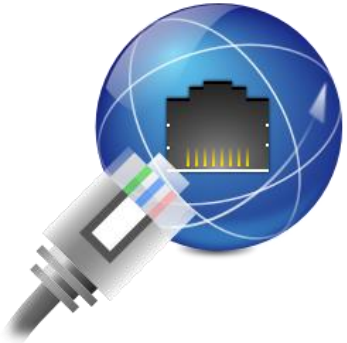
Capture Files:
https://app.box.com/v/sharkfest2016-layer2

Some Topics

- 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
  - …

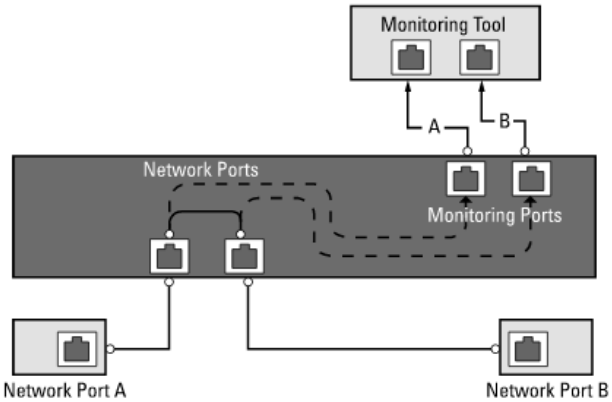
Challenges with copper links

- Speed - 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)#
```

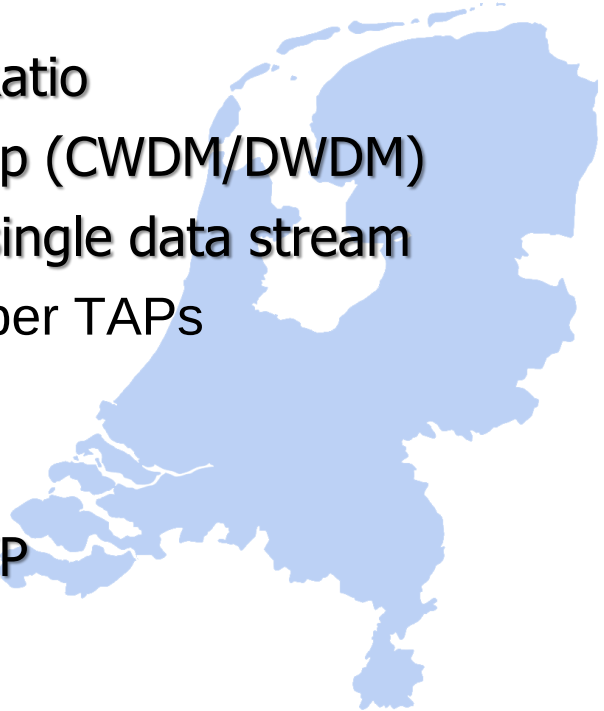
- Power-over-Ethernet (802.3af, 802.3at)
- Link Loss Carry Forward or

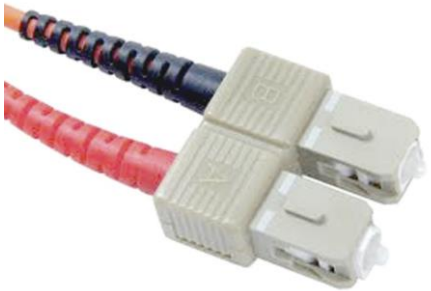
Link Failure Propagation

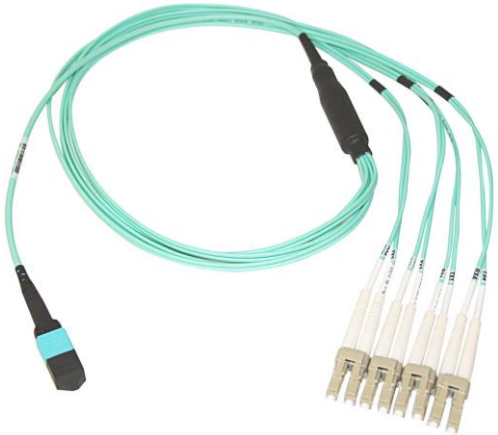


Challenges with fiber links

- 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







Challenges with DAC and AOC

- Direct Attach Cable (DAC)
  - also known as a twinax cable
  - Cost effective solution over optical transceiver and cables for short reach applications
- Active Optical Cable (AOC)
  - Alternative for optical tranceivers

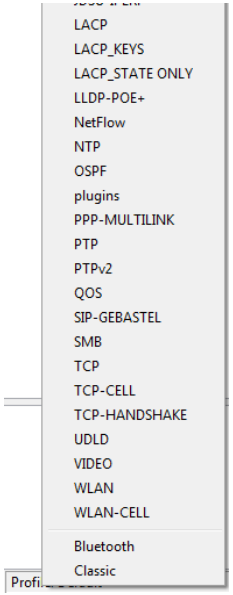
Challenge 2 – Display Environment

- Know your MAC-Addresses and write it down
- Use aliases and well-known names

ethers - Editor

Datei Bearbeiten Format Ansicht ?

```
00:26:b9:bc:9c:87       EIGENE-MAC-WF
00:0e:83:16:f5:10       SWITCH-A--PORT25
00:13:c4:12:0f:0d       SWITCH-B--PORT22
00:19:aa:d9:e1:80       SWITCH-B-SYSTEM
00:19:aa:d9:c7:00       SWITCH-A-SYSTEM
00:80:C8:37:A1:1B       USB-101
00:80:C8:3B:53:CC       USB-104
00:1D:45:7F:63:04       SW1
00:1C:B0:83:A2:84       SW2
```



- 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

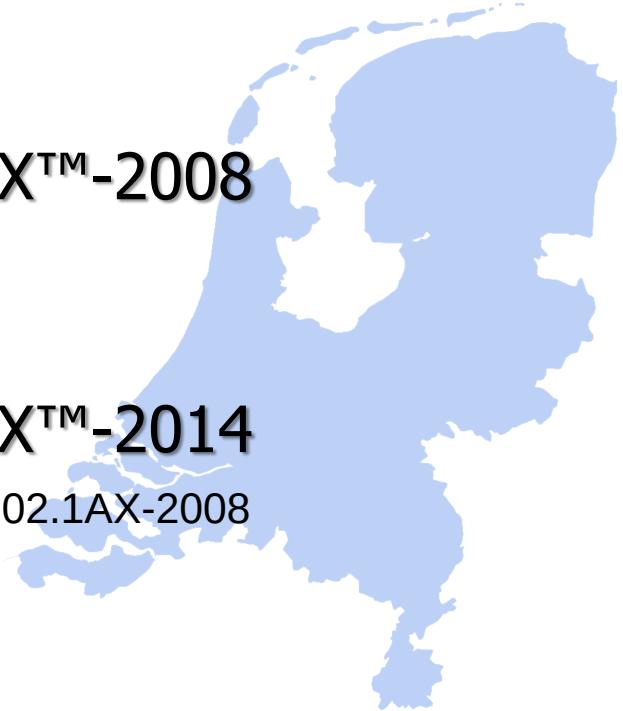
Link Aggregation Control Protocol (LACP)

Link Aggregation Control Protocol (LACP)

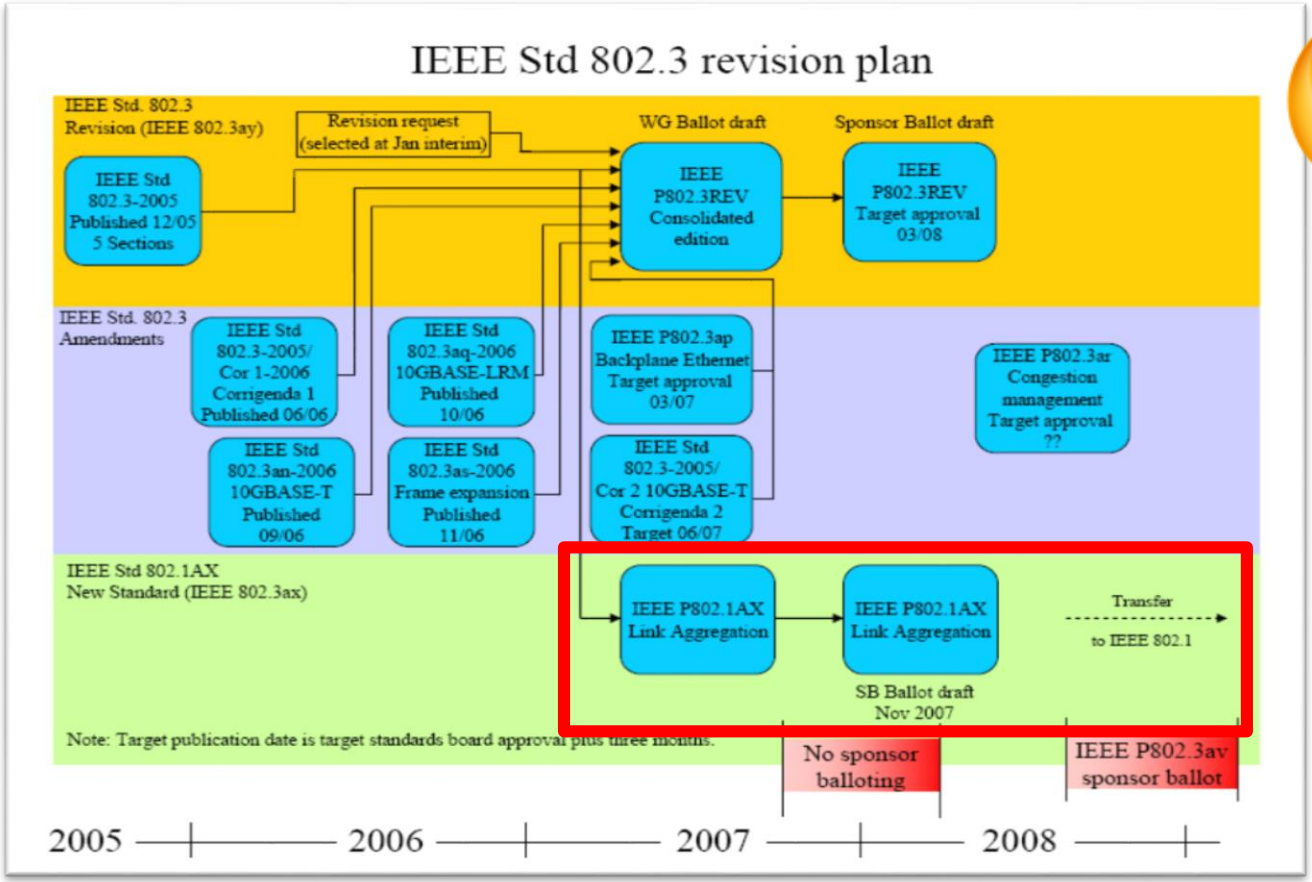
- What is LACP?
  - A Layer 2 protocol to logically bundle multiple physical ethernet links into one
- Why LACP?
  - for increasing bandwidth and build-in redundancy
- Who need it?
  - Everyone from the networking field
- Benefits?
  - Failover, load-sharing, acting as one
- Requirements
  - "… all interfaces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

LACP Standards

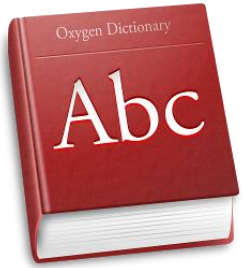
- IEEE Std 802.3, 2000 Edition - Clause 43
  - 802.3ad
- IEEE Std 802.1AX™-2008
  - 802.1AX not 802.3ax
- IEEE Std 802.1AX™-2014
  - Revision of IEEE Std 802.1AX-2008



LACP - 802.1AX versus 802.3ad



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 LACPDU
- 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 headers

- 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

```
Frame 192: 124 bytes on wire (992 bits), 124 bytes captured (992 bits)
Ethernet II, Src: CiscoInc_7f:63:02 (00:1d:45:7f:63:02), Dst: Slow-Protocols (01:80:c2:00:0
Slow Protocols
    Slow Protocols subtype: Marker Protocol (0x02)
Marker Protocol
    Version Number: 0x01
    TLV Type: Marker Response Information (0x02)
    TLV Length: 0x14
    Requester Port: 32768
    Requester System: CiscoInc_7f:63:00 (00:1d:45:7f:63:00)
    Requester Transaction ID: 98304
    TLV Type: Unknown (0x3c)
    TLV Length: 0x00
    Requester Port: 0
    Requester System: MS-NLB-PhysServer-20_80:00:00:1c (02:14:80:00:00:1c)
    Requester Transaction ID: 2961416832
    TLV Type: Unknown (0x80)
    TLV Length: 0x00
    Requester Port:
    Requester Syste
    Requester Trans
```

Adjusted by Packet Editor - no real frame

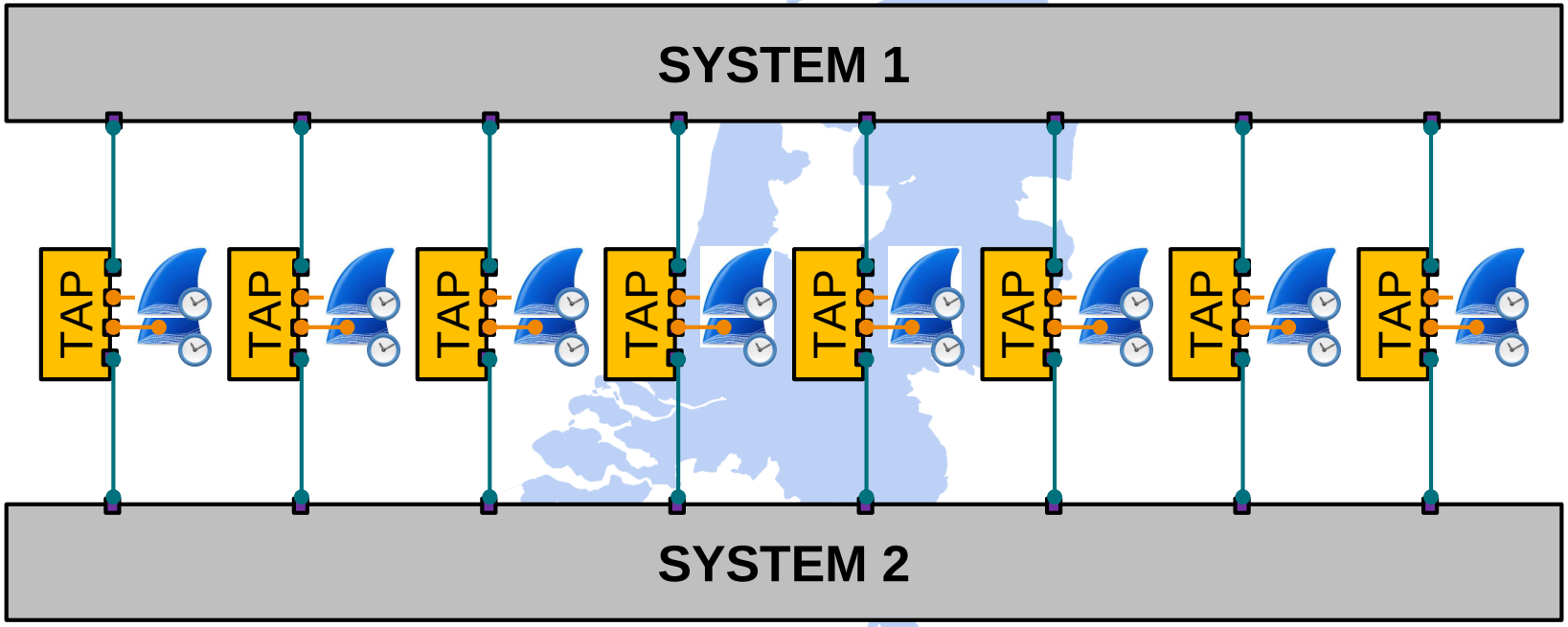
IEEE Std 802.1AX-2014
IEEE STANDARD FOR LOCAL AND METROPOLITAN AREA NETWORKS—LINK AGGREGATION

Marker PDU	Octets	Marker Response PDU
Subtype = Marker	1	Subtype = Marker
Version Number	1	Version Number
TLV_type = Marker Information	1	TLV_type = Marker Response Information
Marker_Information_Length= 16	1	Marker_Response_Information_Length = 16
Requester_Port	2	Requester_Port
Requester_System	6	Requester_System
Requester_Transaction_ID	4	Requester_Transaction_ID
Pad = 0	2	Pad = 0
TLV_type = Terminator	1	TLV_type = Terminator
Terminator_Length = 0	1	Terminator_Length = 0
Reserved	90	Reserved
FCS	4	FCS

Figure 6-27—Marker PDU and Marker Response PDU structure

Capturing LACP

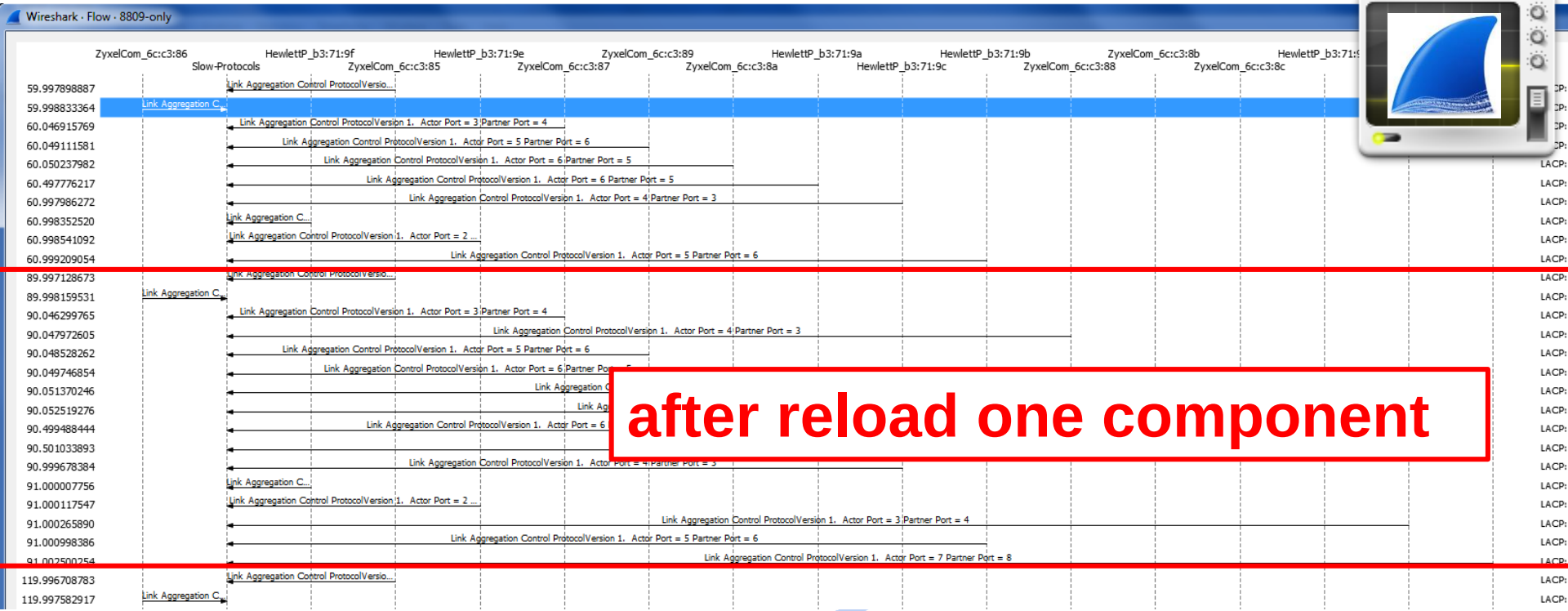
Keep all your capture points in time sync



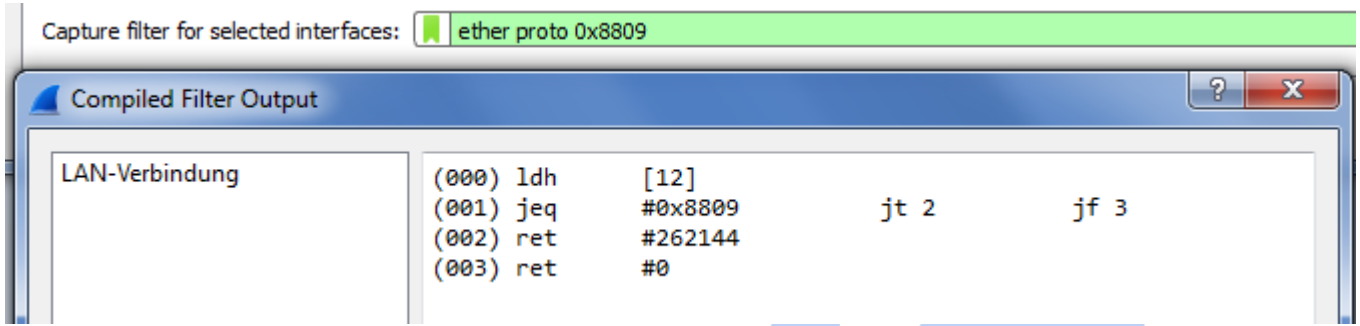
Capturing LACP – from Reality

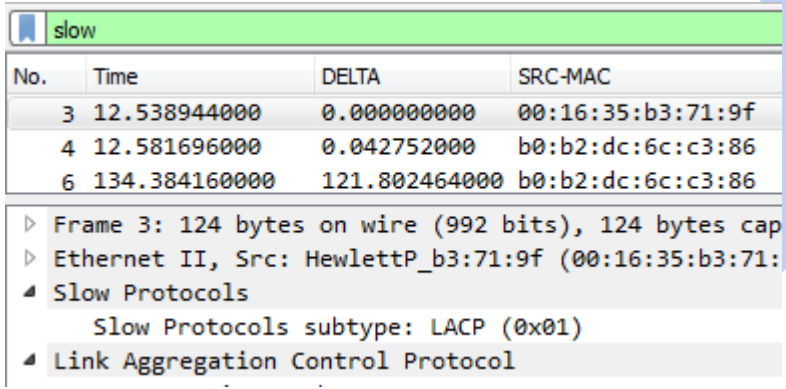
- Lab Environment
- The Hardware Ethernet Analyzers provides different methods to capture packets inline and full-duplex.
- Copper or Fiber – up to Gigabit
- Wireshark used for further analysis

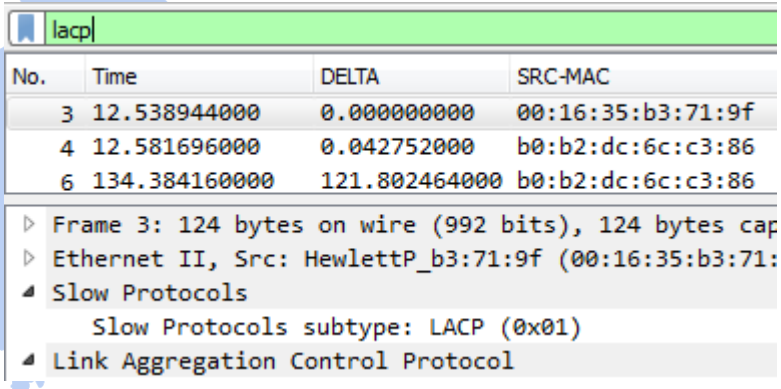
LACP - Flow Graph

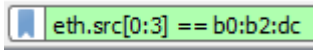


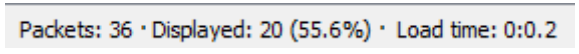
LACP – Capture and Display Filter



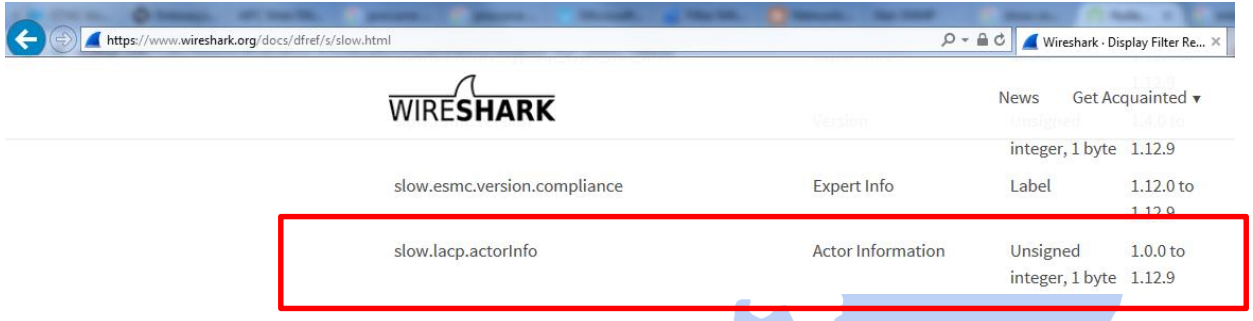


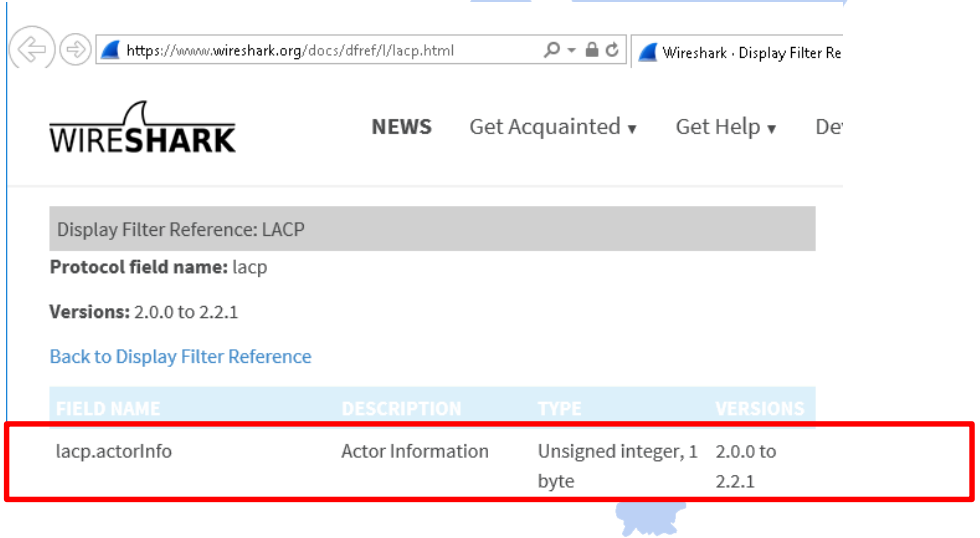






LACP in Wireshark 2.2





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 (00:19:aa:d9:c7:00)
    Actor Key: 10
    Actor Port Priority: 32768
    Actor Port: 274
  ▷ Actor State: 0x7d, LACP Activity, Aggregation, Synchronization, Collecting, Distributing, Defaulted
    Reserved: 000000
    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
    Reserved: 000000
    Collector Information: 0x03
    Collector Information Length: 0x10
    Collector Max Delay: 32768
    Reserved: 00000000000000000000000000
    Terminator Information: 0x00
    Terminator Length: 0x00
    Reserved: 000000000000000000000000000000000000000000000000...
```

ACTOR

PARTNER

COLLECTOR

TERMINATOR

LACP – Flags (Actor)

```
▲ Actor State: 0x3d, LACP Activity, Aggregation, Synchronization, Collecting, Distributing
      .... ...1 = LACP Activity: Yes
      .... ..0. = LACP Timeout: No
      .... .1.. = Aggregation: Yes
      .... 1... = Synchronization: Yes
      ...1 .... = Collecting: Yes
      ..1. .... = Distributing: Yes
      .0.. .... = Defaulted: No
      0... .... = Expired: No
```

- 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 = is NOT EXPIRED (lacp.partnerState.expired), 1 Byte

LACP – Flags (as an enhancement)

```
> Link Aggregation Control Protocol
v LACP Actor Flags
    LACP Flags: *FDC*G*A
v LACP Partner Flags
    LACP Flags: E****GS*
```

After SharkFest US 2016 I opened an enhancement request via https://bugs.wireshark.org – and now it is part of every Wireshark installation.

→Thank you Wireshark developers / Note: Everyone can do it ☺

LACP – System-ID

```
Link Aggregation Control Protocol
    LACP Version Number: 0x01
    Actor Information: 0x01
    Actor Information Length: 0x14
    Actor System Priority: 32768
    Actor System: SWITCH-A-SYSTEM (00:19:aa:d9:c7:00)
    Actor Key: 10
    Actor Port Priority: 32768
    Actor Port: 274
  ▷ Actor State: 0x7d, LACP Activity, Aggregation, Synchronization, Collecting, Distributing, Defaulted
    Reserved: 000000
```

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.

LACP – Actor Key

```
Link Aggregation Control Protocol
    LACP Version Number: 0x01
    Actor Information: 0x01
    Actor Information Length: 0x14
    Actor System Priority: 32768
    Actor System: SWITCH-A-SYSTEM (00:19:aa:d9:c7:00)
    Actor Key: 10
    Actor Port Priority: 32768
    Actor Port: 274
  ▷ Actor State: 0x7d, LACP Activity, Aggregation, Synchronization, Collecting, Distributing, Defaulted
    Reserved: 000000
```

KEY

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.

LACP – Port-ID

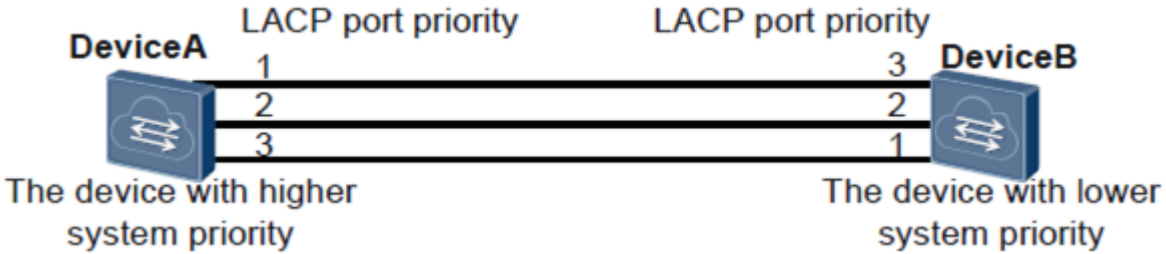
```
Link Aggregation Control Protocol
    LACP Version Number: 0x01
    Actor Information: 0x01
    Actor Information Length: 0x14
    Actor System Priority: 32768
    Actor System: SWITCH-A-SYSTEM (00:19:aa:d9:c7:00)
    Actor Key: 10
    Actor Port Priority: 32768
    Actor Port: 274
  ▷ Actor State: 0x7d, LACP Activity, Aggregation, Synchronization, Collecting, Distributing, Defaulted
    Reserved: 000000
```

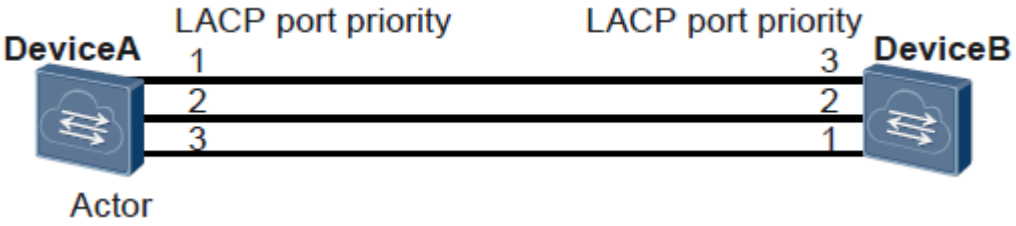
Port-ID

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





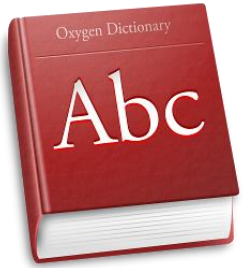
Decision Maker → Which ports are part of the aggregation

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 - MAC/PHY Configuration/Status
    IEEE 802.3 - Link Aggregation
        1111 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
```

```
IEEE 802.3 - Link Aggregation
    1111 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: 0x03
        .... ...1 = Aggregation Capability: Yes
        .... ..1. = Aggregation Status: Enabled
    Aggregated Port Id: 6
```

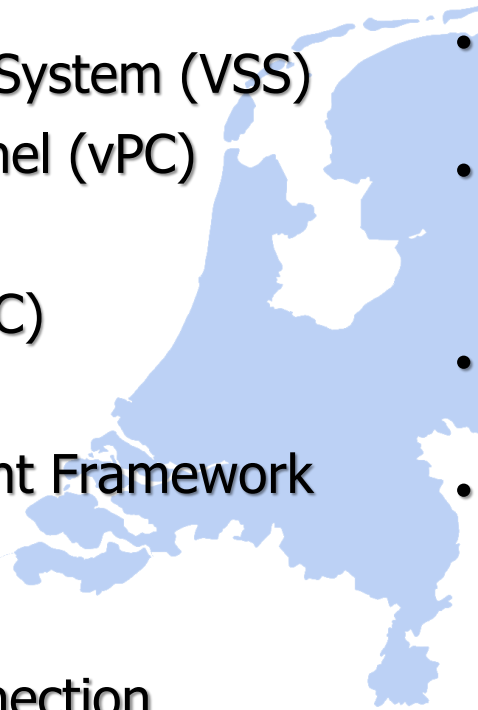
MC-LAG with LACP



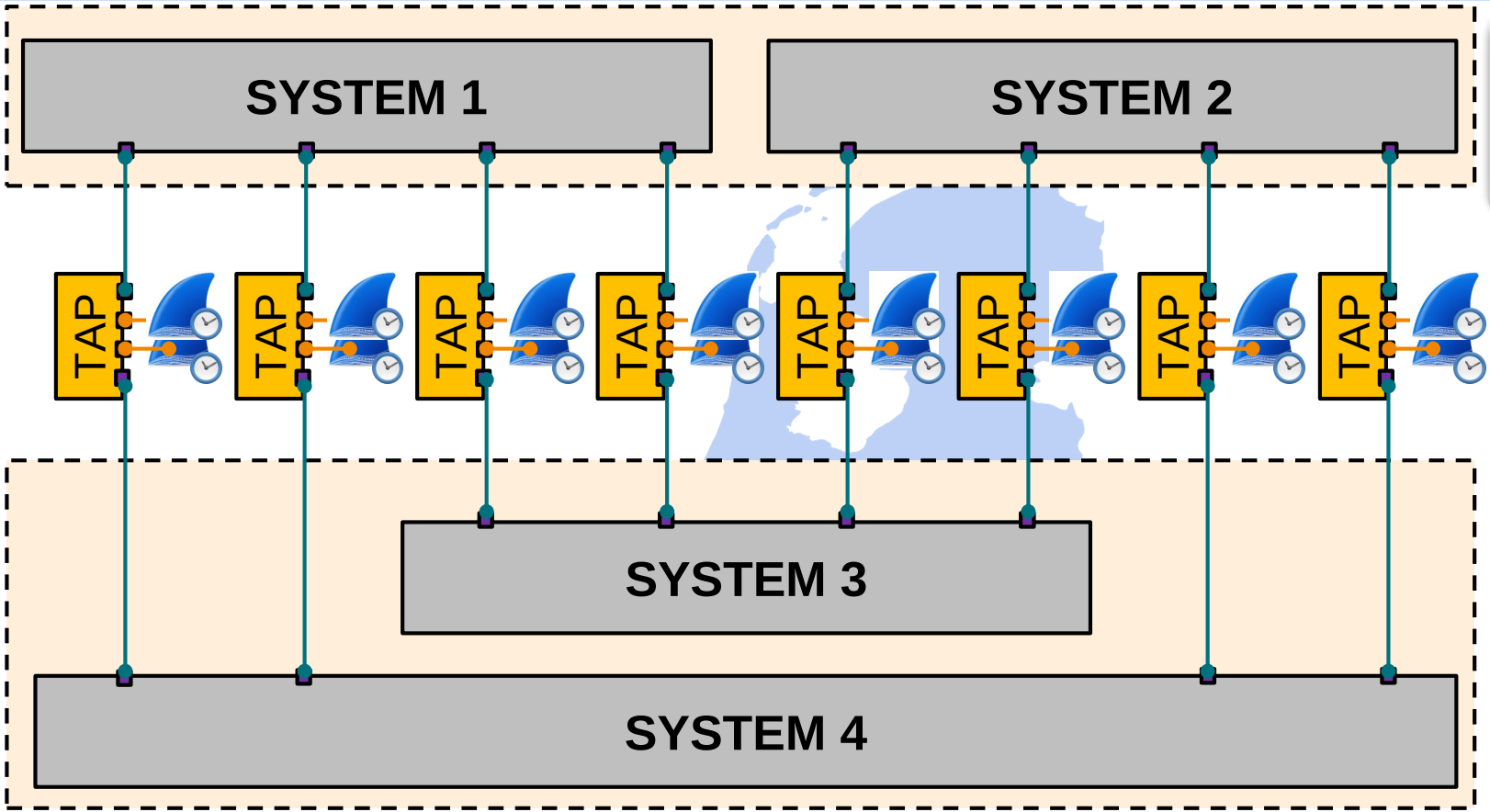
- MC-LAG - LAG terminate on separate chassis
- MC-LAG is not covered under IEEE standard
- Multi-homing for redundancy
- Active-active to utilize all links which otherwise may get blocked by Spanning-Tree
- no modification of LAG partner
- Temporary loops or duplicates not acceptable
- Split brain handling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



LACP capture in a distributed environment



MC-LAG with LACP

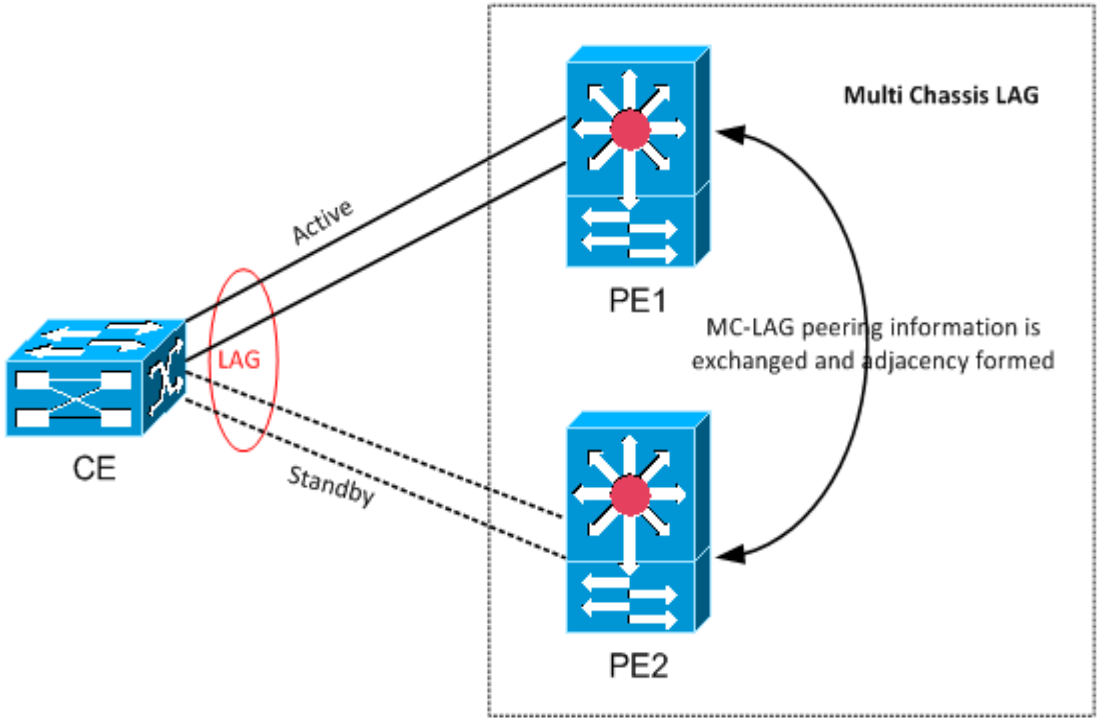
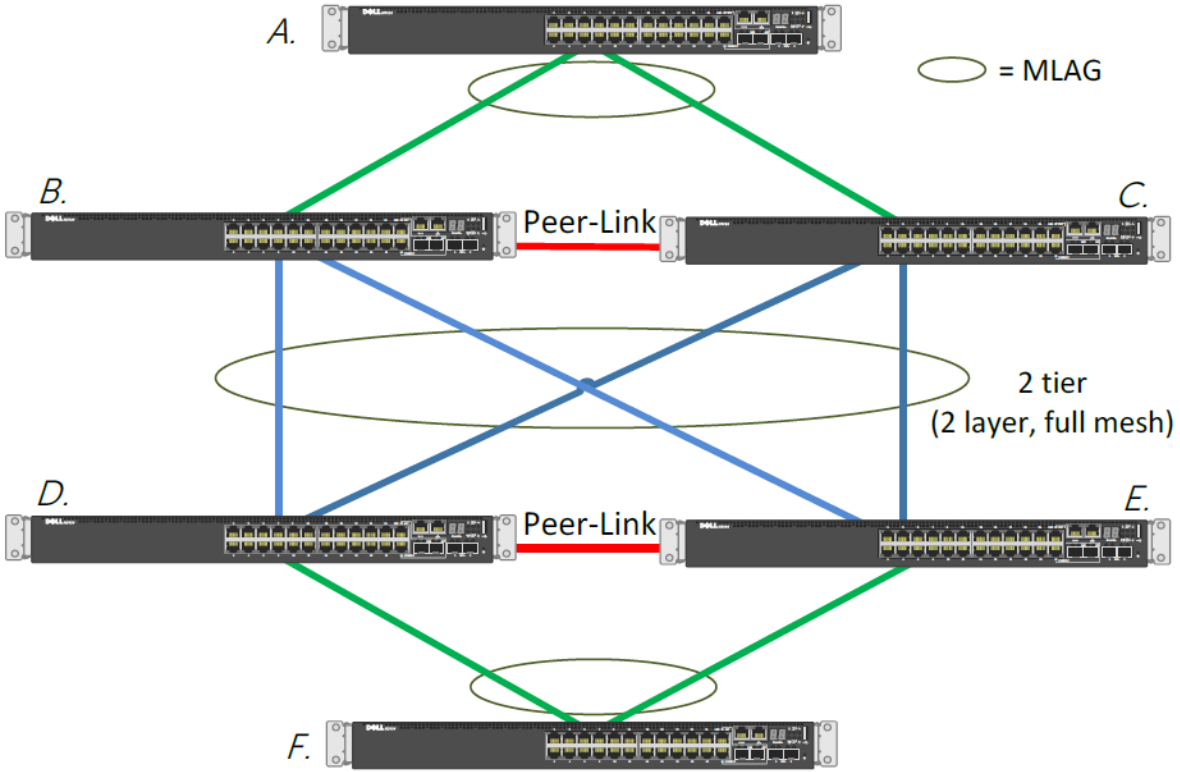


Figure 1. A basic MC-LAG Example

MLAG with LACP



LACP Challenges from the field

LACP Frame Size different

Destination	Protocol	Length	Info
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	126	Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	126	Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	126	Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	126	Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
Slow-Protocols	LACP	124	Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7

LACP Timer different – not short or long

203.881216000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
204.881248000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
205.534944000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
205.881216000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
206.881248000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
207.534976000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
207.881248000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
208.881280000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
209.534944000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
209.881280000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
210.881312000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
211.534944000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
211.881312000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
212.881312000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
213.534944000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
213.881312000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
214.881344000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
215.534944000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
215.881344000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
216.881376000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
217.534944000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
217.881376000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
218.881376000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7
219.534944000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 7 Partner Port = 1
219.881376000	Link Aggregation Co	LACP: Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 7

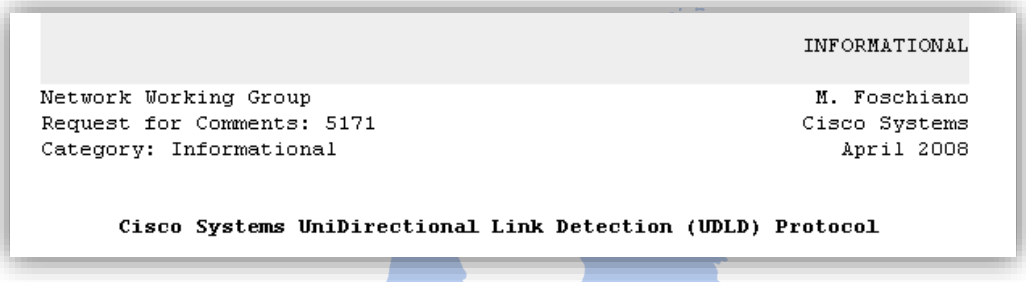
Every two seconds from the link partner

Time for questions

Unidirectional Link Detection (UDLD)

UDLD Basics

- Cisco UDLD feature is documented in RFC 5171



INFORMATIONAL

Network Working Group — M. Foschiano
Request for Comments: 5171 — Cisco Systems
Category: Informational — April 2008

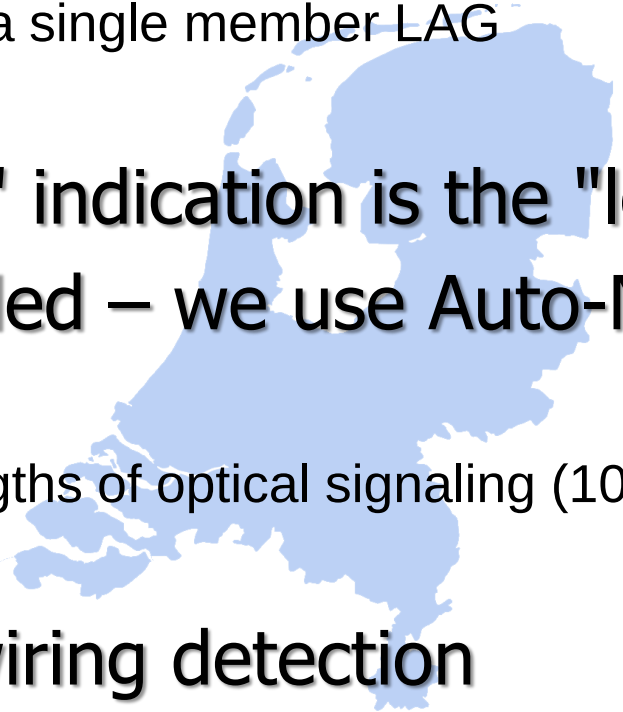
Cisco Systems UniDirectional Link Detection (UDLD) Protocol

Different names and implementations

- Device Link Detection Protocol (DLDP)
- D-Link Unidirectional Link Detection (DULD)

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



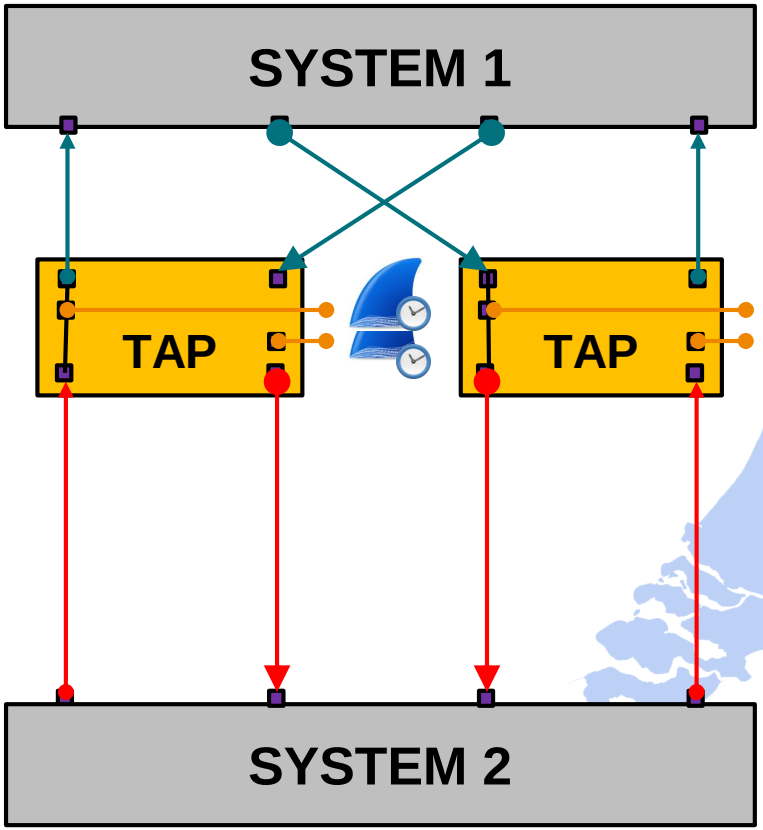
UDLD Notes

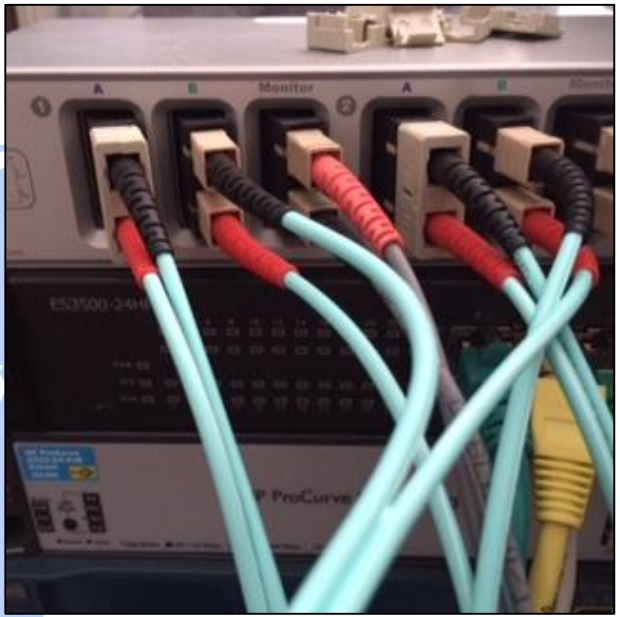
- Cisco UDLD are encoded with LLC, standard Subnetwork Access Protocol (SNAP) format and Protocol ID 0x111
- Destination Multicast MAC Address: 01:00:0C:CC:CC:CC
- Fast Hello enhancement available

UDLD Modes

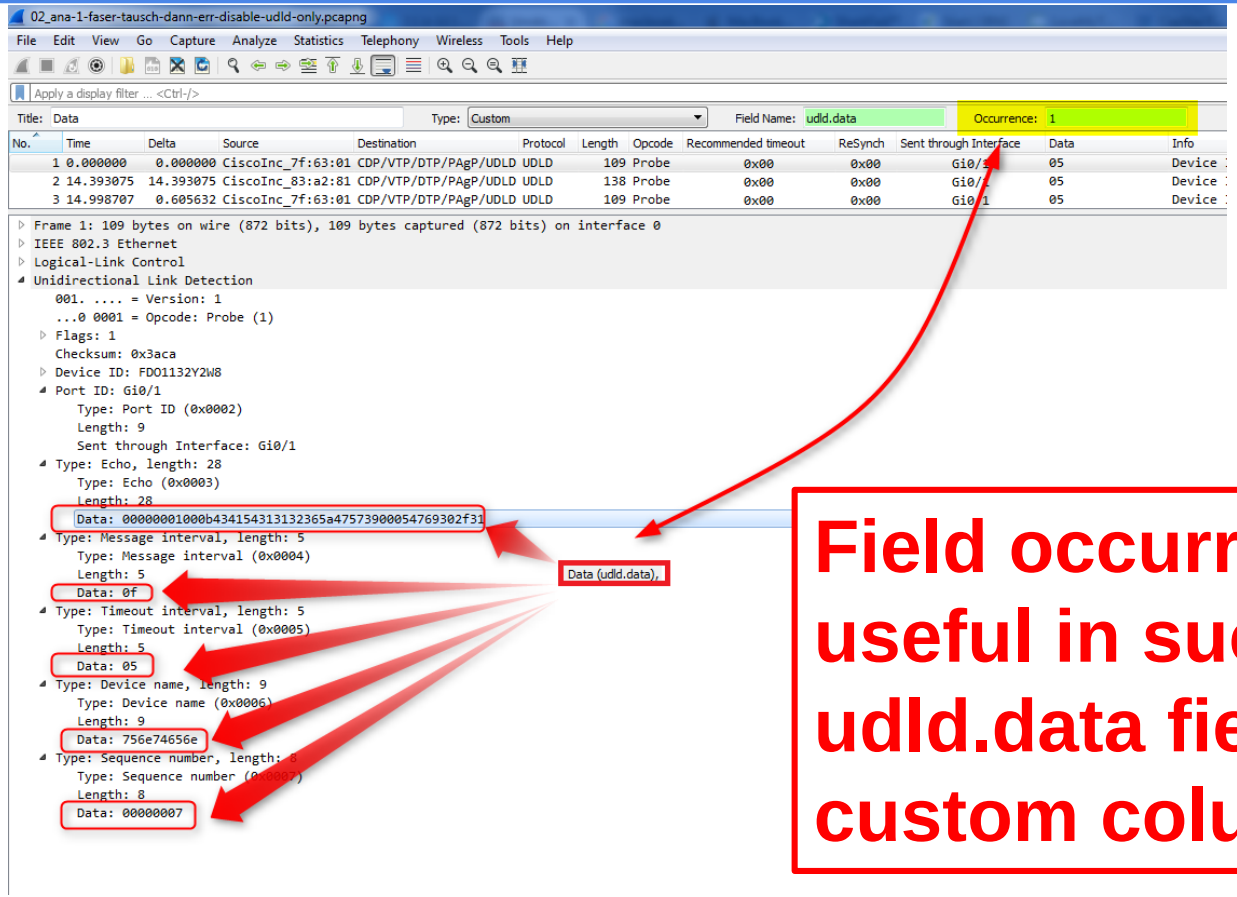
- Aggressive Mode:
  - UDLD will declare link as unidirectional and will disable interface, if no reply has been received for subsequent 8 PDU message transmitted at an interval of 1 sec.
- Normal Mode:
  - Link will be disabled immediately if PDU reply has not been received within predefine timeout interval.

UDLD Capture



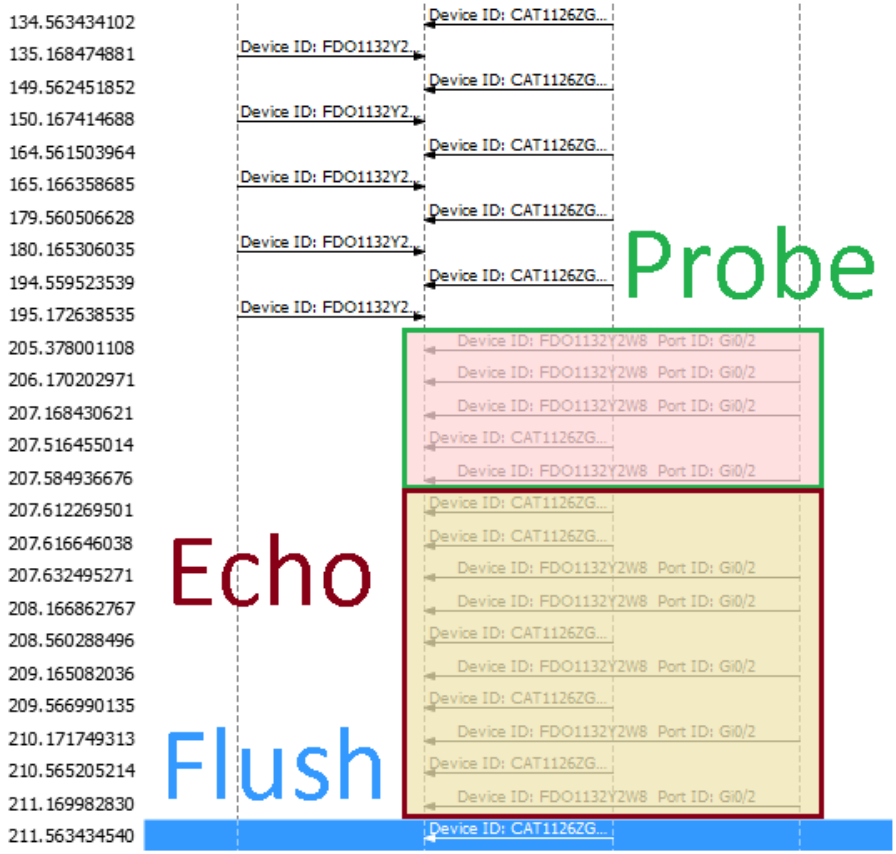


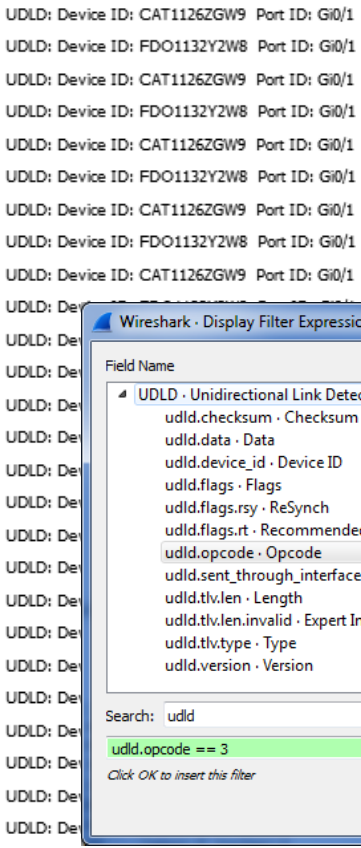
UDLD in Wireshark – Custom Columns



Field occurrence is very useful in such case. Every udld.data field as a custom column

UDLD in Wireshark – Flow Graph

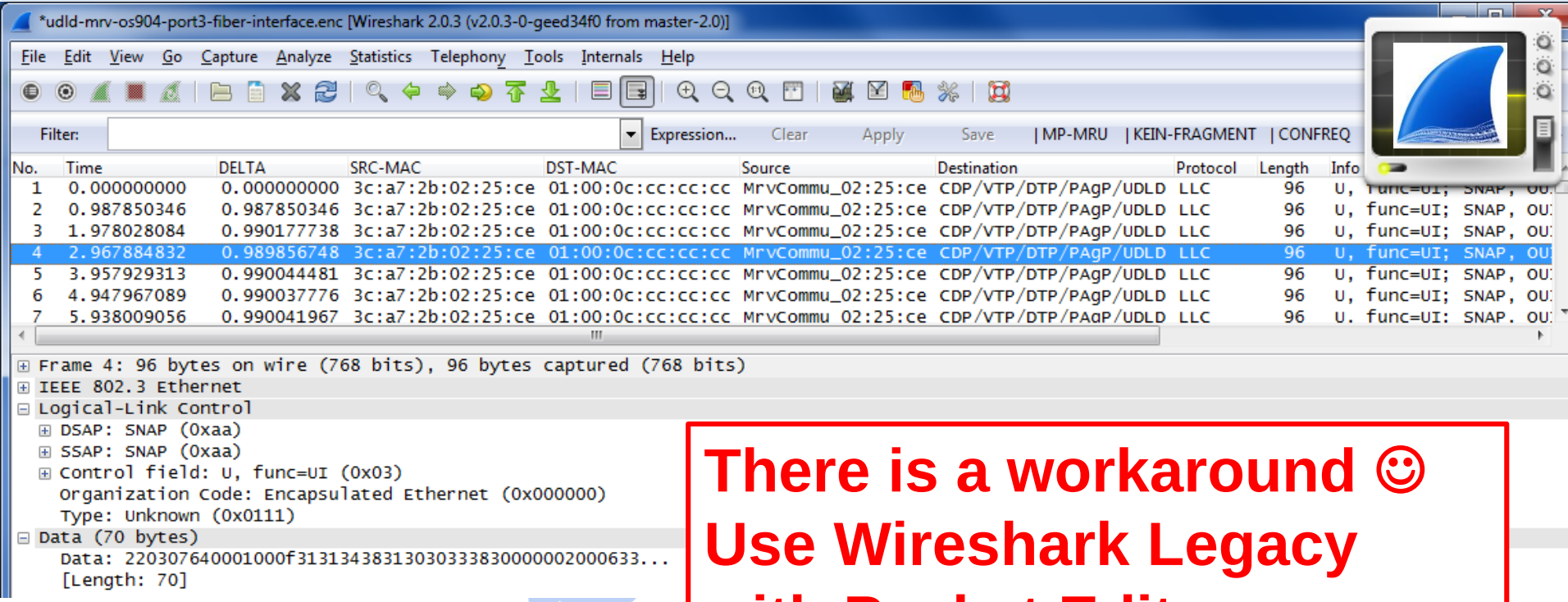




Challenges with Unidirectional Link Detection (UDLD) from the field

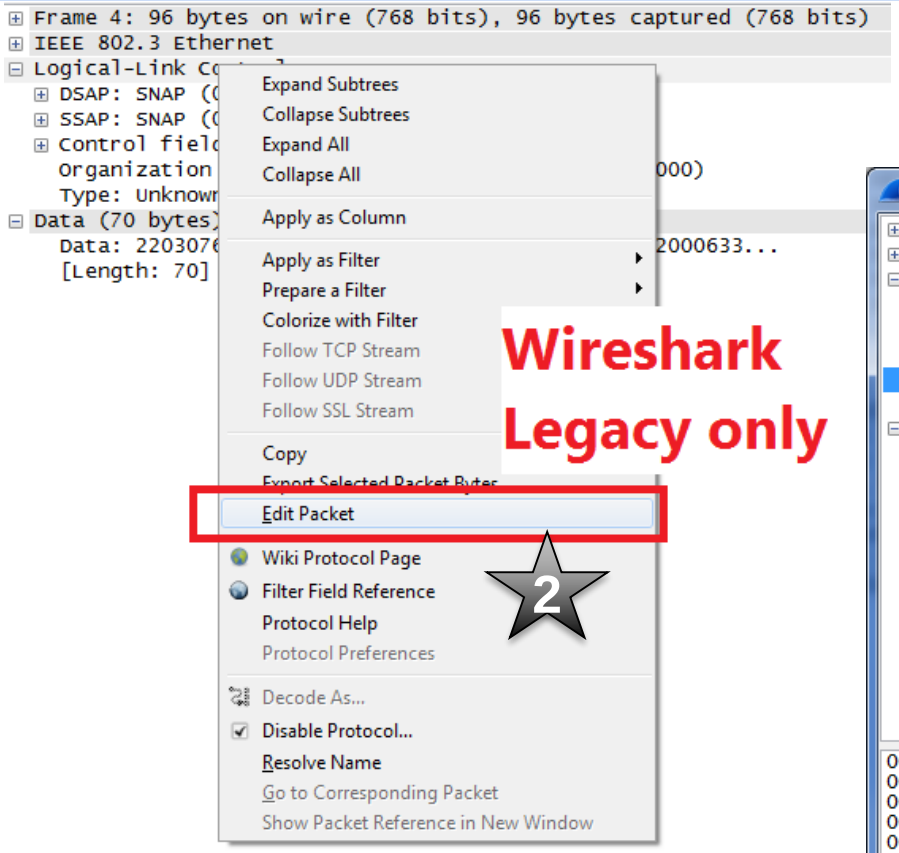
SharkFest '16 Europe • Arnhem, Netherlands • October 17-19, 2016 • #sf16eu

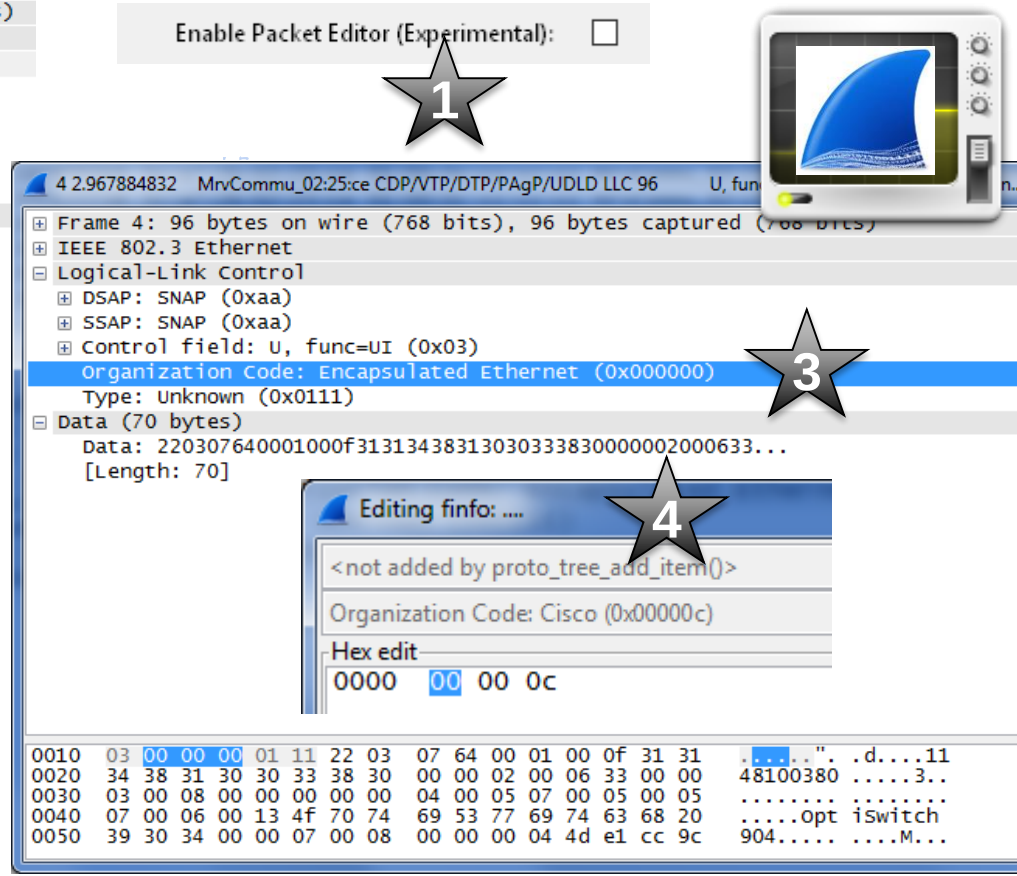
UDLD non-Cisco



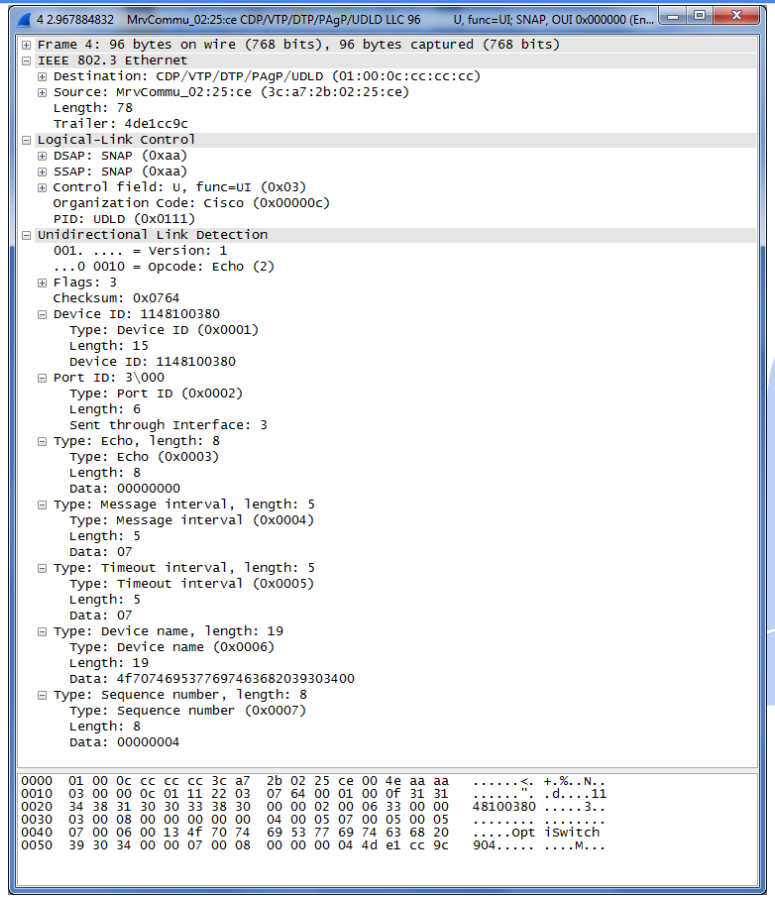
There is a workaround ☺
Use Wireshark Legacy with Packet Edit

UDLD non-Cisco





UDLD non-Cisco



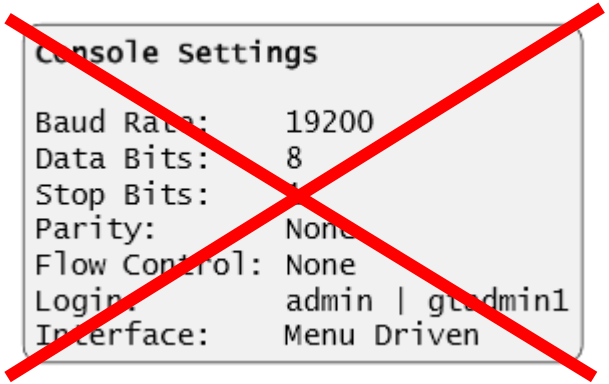
Works for this vendor
Maybe a bug – will discuss that with the wireshark developers ☺ afterwards

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



Ethernet Flow-Control Priority Resultion

Table 37–4—Pause priority resolution

Local Device		Link Partner		Local Resolution	Link Partner Resolution
PAUSE	ASM_DIR	PAUSE	ASM_DIR		
0	0	—	—	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
0	1	0	—	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
0	1	1	0	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
0	1	1	1	Enable PAUSE transmit, Disable PAUSE receive	Enable PAUSE receive, Disable PAUSE transmit
1	0	0	—	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
1	0	1	—	Enable PAUSE Transmit and Receive	Enable PAUSE Transmit and Receive
1	1	0	0	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
1	1	0	1	Enable PAUSE receive, Disable PAUSE transmit	Enable PAUSE transmit, Disable PAUSE receive
1	1	1	—	Enable PAUSE Transmit and Receive	Enable PAUSE Transmit and Receive

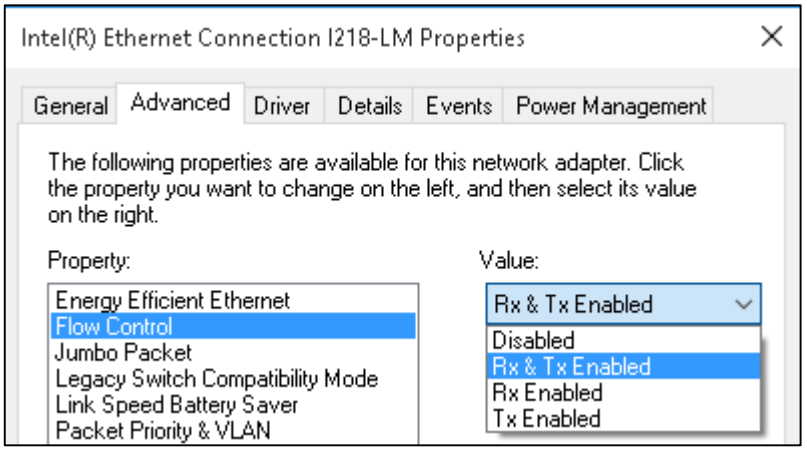
Ethernet Flow-Control

- 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*512/1.000.000.000 = 0.03355392 seconds = 33.55ms.

Ethernet Flow-Control Priority Resultion




```
Router#sh controller FastEthernet 0/0 | b PHY registers
 PHY registers:
   Register 0x00:   1000  782D  0040  61E4  01E1  C1E1  000D  2001
   Register 0x08:   0000  0000  0000  0000  0000  0000  0000  0000
   Register 0x10:   1000  0300  0000  0000  0200  007B  0300  0000
   Register 0x18:   003B  851F  9F00  008A  082B  0000  80A0
Bytes_recvd 2706029297 Bytes_sent 636605627 Frames_recvd 54499753 Frames_sent 7456645
total_bytes_RX 2706029297 Total_frames_RX 54499753 Bcast_frames_recvd 18797931
Mcast_frames_RX 21756491 CRC_err 0 Ovr_sized_frames 0
Fragments 0 Jabber 0 collision 0
Late_collision 0 64B frame 26323978; 65_127B_frames 17093799
128_255B_frames 14987201 256_511B_frames 2319326 512_1023B_frames 723842
1023_maxB_frames 508252 Rx_error 0 Dropped_frames 0
Mcast_frames_tx 1488314 Bcast_frames_tx 2839987 Sml_frame_recvd 0
```

```
# ethtool ens817
   Settings for ens817:
     Supported ports: [ FIBRE ]
     Supported link modes:   1000baseKX/Full
                             10000baseKX4/Full
                             10000baseKR/Full
                             40000baseCR4/Full
                             40000baseSR4/Full
     Supported pause frame use: Symmetric Receive-only
     Supports auto-negotiation: Yes
     Advertised link modes:  1000baseKX/Full
                             10000baseKX4/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
     PHYAD: 0
     Transceiver: internal
     Auto-negotiation: on
     Supports Wake-on: d
     Wake-on: d
     Current message level: 0x00000014 (20)
            link ifdown
     Link detected: yes
```

MAC PAUSE Frames

```
Frame 453: 64 bytes on wire (512 bits), 64 bytes captured (512 bits)
Ethernet II, Src: EIGENE-MAC-WF (00:26:b9:bc:9c:87), Dst: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
    Destination: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
    Source: EIGENE-MAC-WF (00:26:b9:bc:9c:87)
    Type: MAC Control (0x8808)
MAC Control
    Opcode: Pause (0x0001)
    pause_time: 1664
```

macc

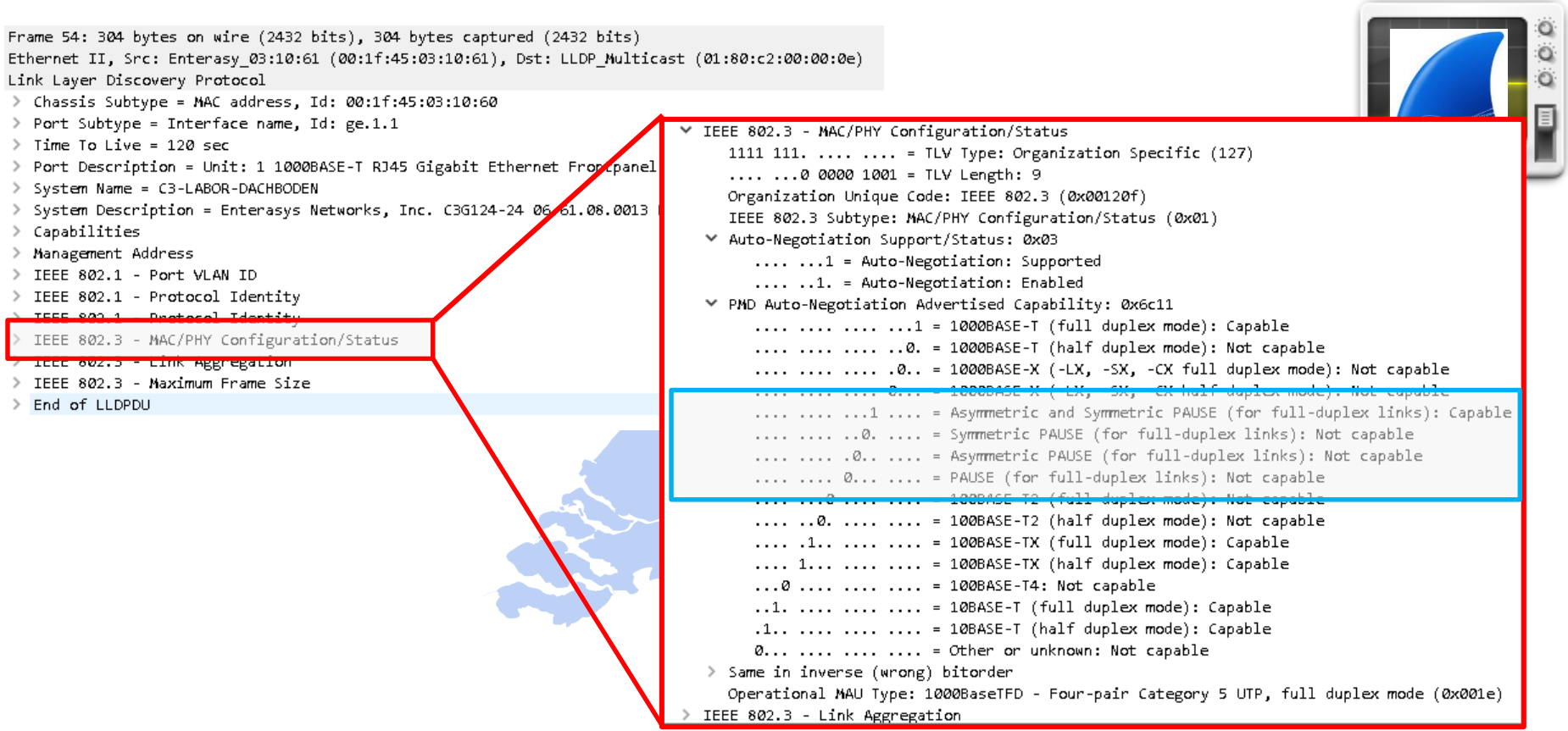
```
0000  01 80 c2 00 00 01 00 26  b9 bc 9c 87 88 08 00 01
0010  06 80 00 00 00 00 00 00  00 00 00 00 00 00 00 00
0020  00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00
0030  00 00 00 00 00 00 00 00  00 00 00 00 ac c3 7a 28
```

```
Frame 455: 64 bytes on wire (512 bits), 64 bytes captured (512 bits)
Ethernet II, Src: EIGENE-MAC-WF (00:26:b9:bc:9c:87), Dst: Spanning-tree-(for-bridges)
    Destination: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
    Source: EIGENE-MAC-WF (00:26:b9:bc:9c:87)
    Type: MAC Control (0x8808)
MAC Control
    Opcode: Pause (0x0001)
    pause_time: 0
```

```
0000  01 80 c2 00 00 01 00 26  b9 bc 9c 87 88 08 00 01   .......& ........
0010  00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
0020  00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
0030  00 00 00 00 00 00 00 00  00 00 00 00 97 90 16 84   ........ ........
```

100 Mbit/s,
1 Gbit/s,
10 Gbit/s,
40 Gbit/s
or 100 Gbit/s?

Flow-Control in interaction with LLDP



PFC – Priority Based Flow-Control

```
▷ Frame 1: 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)
   ▷ Destination: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
   ▷ Source: 0e:fc:00:d5:3d:00 (0e:fc:00:d5:3d:00)
     Type: MAC Control (0x8808)
◢ MAC Control
     Opcode: Class Based Flow Control [CBFC] Pause (0x0101)
   ◢ CBFC Class Enable Vector: 0x0008, C3
        .... .... .... ...0 = C0: False
        .... .... .... ..0. = 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 Gotchas with the Taps
  - two Links segments for the Network Ports
  - also two Links for the Monitoring Ports

Flow-Control ???

PORTABLE TAP KEY FEATURES

- Small portable form factor: 3.942"x1.20"x3.942" (W x H x D)
- Four (4) Copper RJ-45 Ports
- Plug-n-Play design with zero configuration
- Link Speed Synchronization
- Captures Full Duplex Traffic up to 2 Gigabits without dropping any packets
- Supports Jumbo Frames
- Passes physical layer errors
- Rack Shelf supports up to four (4) TAPs in a 1U space

Flow-Control Challenges from the field

Strange Flow-Control Implementation

AT900-FLOWCONTROLTEST.enc

File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help

Apply a display filter ... <Ctrl-/>

No.	Time	DELTA	SRC-MAC	DST-MAC	Source	Destination	Protocol	Length	Info
1	0.000000000	0.000000000	01:80:c2:00:00:01	01:80:c2:00:00:01	Spanning-tree-(for-bridges)_01	Spanning-tree-(for-bridges)_01	MAC CTRL	64	Pause: pause_time: 65535 quanta
2	0.000862401	0.000862401	01:80:c2:00:00:01	01:80:c2:00:00:01	Spanning-tree-(for-bridges)_01	Spanning-tree-(for-bridges)_01	MAC CTRL	64	Pause: pause_time: 0 quanta
3	0.008910637	0.008048236	01:80:c2:00:00:01	01:80:c2:00:00:01	Spanning-tree-(for-bridges)_01	Spanning-tree-(for-bridges)_01	MAC CTRL	64	Pause: pause_time: 65535 quanta
4	0.009773037	0.000862400	01:80:c2:00:00:01	01:80:c2:00:00:01	Spanning-tree-(for-bridges)_01	Spanning-tree-(for-bridges)_01	MAC CTRL	64	Pause: pause_time: 0 quanta
5	0.017822111	0.008049074	01:80:c2:00:00:01	01:80:c2:00:00:01	Spanning-tree-(for-bridges)_01	Spanning-tree-(for-bridges)_01	MAC CTRL	64	Pause: pause_time: 65535 quanta
6	0.018683674	0.000861563	01:80:c2:00:00:01	01:80:c2:00:00:01	Spanning-tree-(for-bridges)_01	Spanning-tree-(for-bridges)_01	MAC CTRL	64	Pause: pause_time: 0 quanta
7	0.026732748	0.008049074	01:80:c2:00:00:01	01:80:c2:00:00:01	Spanning-tree-(for-bridges)_01	Spanning-tree-(for-bridges)_01	MAC CTRL	64	Pause: pause_time: 65535 quanta

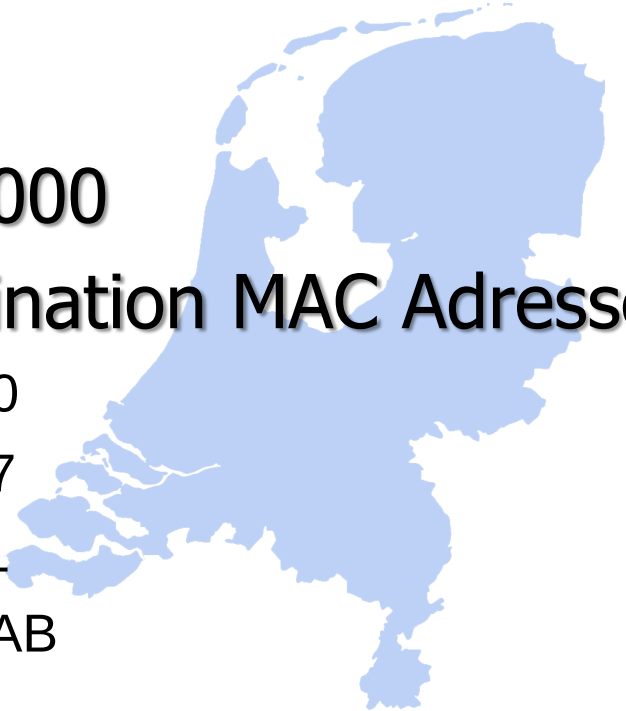
```
▷ Frame 1: 64 bytes on wire (512 bits), 64 bytes captured (512 bits)
◢ Ethernet II, Src: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01), Dst: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
  ▷ Destination: Spanning-tree-(for-bridges)_01 (01:80:c2:00:00:01)
  ◢ Source: Spanning-tree-(for-bridges)_01 (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. .... .... .... .... = LG bit: Globally unique address (factory default)
      .... ...1 .... .... .... .... = IG bit: Group address (multicast/broadcast)
    Type: MAC Control (0x8808)
◢ MAC Control
    Opcode: Pause (0x0001)
    pause_time: 65535
```

Time for questions

Configuration Test Protocol (loopback)

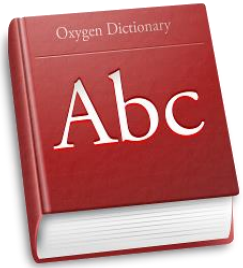
Loop Detection Protocol

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



Loop Detection Protocol

- CTP from the archive



ETHERNET SPECIFICATION: Configuration Testing Protocol

8. ETHERNET CONFIGURATION TESTING PROTOCOL

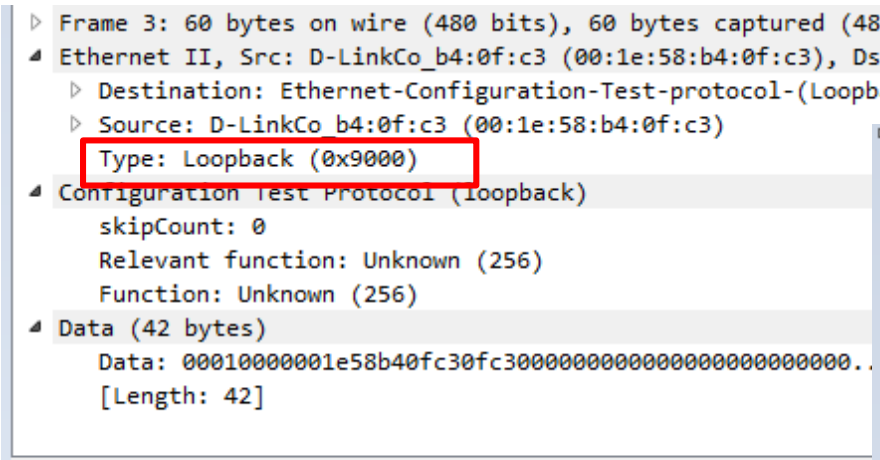
The *Ethernet Configuration Testing Protocol* provides a minimum testing capability of communication between stations on an Ethernet. It is the only Client Layer protocol specified in this document and has the only assigned Ethernet type field value in this document. All Ethernet stations must support the configuration testing functions.

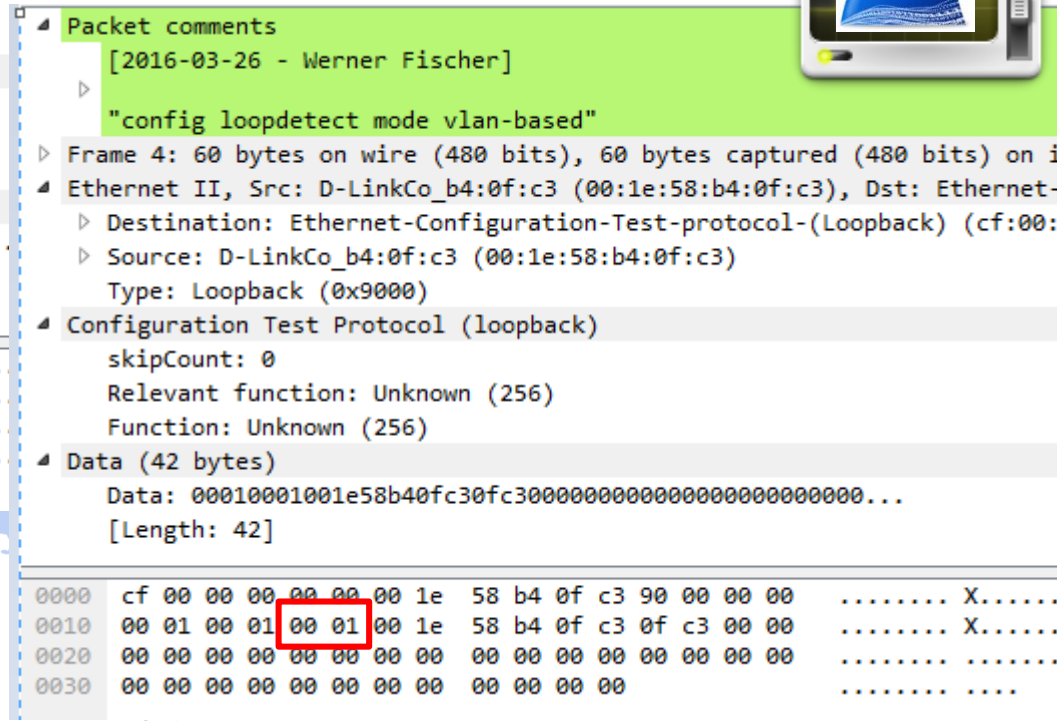
8.1 Goals

For more information see
http://www.mit.edu/~jhawk/ctp.pdf

Loopback from the field

Loop Detection Protocol





Loop Detection Protocol – Vendor XYZ

- Every vendor has it own solution
  - TLV coded

```
0000   01 a0 c5 aa aa ab fc f5   28 4d 6d 47 90 00 01 00   ........ (MmG....
0010   27 75 4b 01 06 fc f5 28   4d 6d 47 02 02 00 06 03   'uK....( MmG.....
0020   06 45 53 33 35 30 30 04   04 00 08 c1 40 05 06 45   .ES3500. ....@..E
0030   53 33 35 30 30 00 00 00   00 00 00 00 3e 43 dd 3c   S3500... ....>C.<
```

  - You should read the HEX-code – also in 2016

Loop Detection Protocols – Port Testing

- Every vendor has it own solution

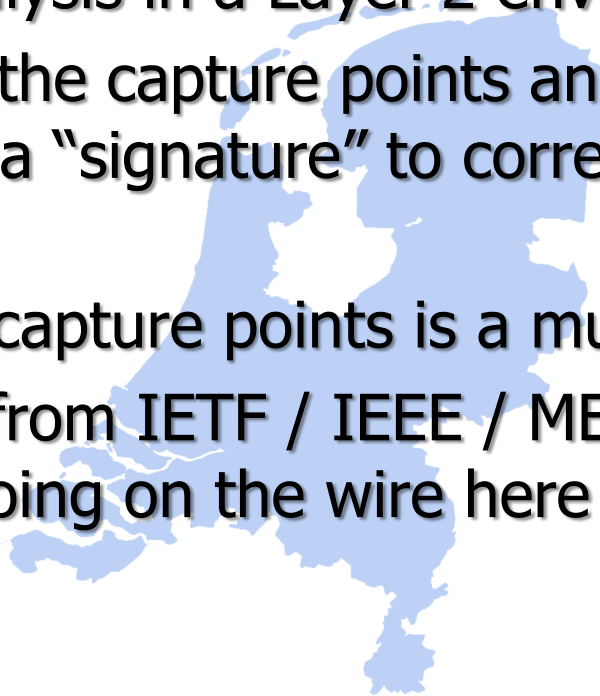
```
▷ Frame 1: 64 bytes on wire (512 bits), 64 bytes captured (512 bits)
▷ Ethernet II, Src: AlliedTe_00:00:00 (00:00:cd:00:00:00), Dst: AlliedTe_00:00:01 (00:00:cd:00:00:01)
◢ Configuration Test Protocol (loopback)
    skipCount: 0
    Relevant function: Unknown (3840)
    Function: Unknown (3840)
◢ Data (46 bytes)
    Data: 3c54657374696e6720706f7274303e00000000000000000...
    [Length: 46]
```

```
0000  00 00 cd 00 00 01 00 00  cd 00 00 00 90 00 00 00   ........ ........
0010  00 0f 3c 54 65 73 74 69  6e 67 20 70 6f 72 74 30   ..<Testi ng port0
0020  3e 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   >....... ........
0030  00 00 00 00 00 00 00 00  00 00 00 00 82 61 e5 ca   ........ .....a..
```

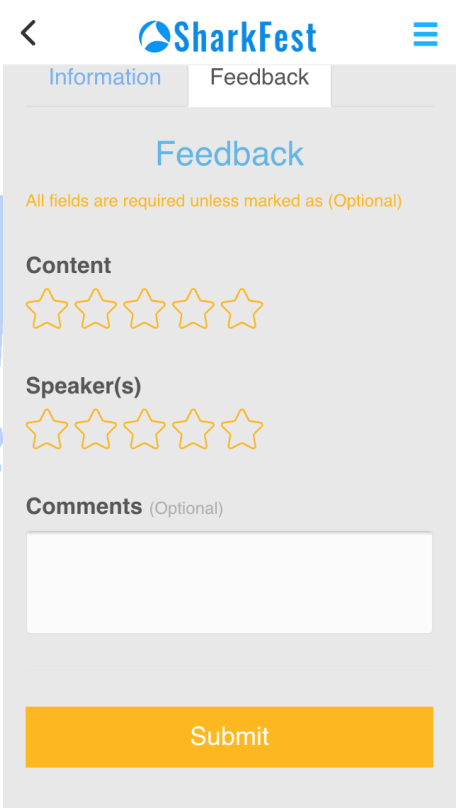
Time for questions

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



THANK YOU