Link Aggregation - IEEE 802.3ad

Mario Baldi
Politecnico di Torino
mario.baldi[at]polito.it
staff.polito.it/mario.baldi

Pietro Nicoletti
Studio Reti
piero[at]studioreti.it
www.studioreti.it

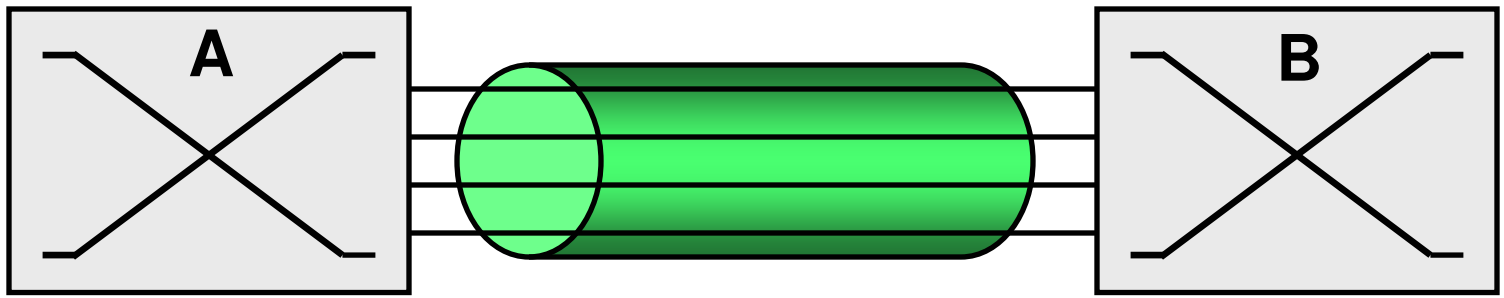
Based on chapter 8 of:
M. Baldi, P. Nicoletti, "Switched LAN", McGraw-Hill, 2002, ISBN 88-386-3426-2

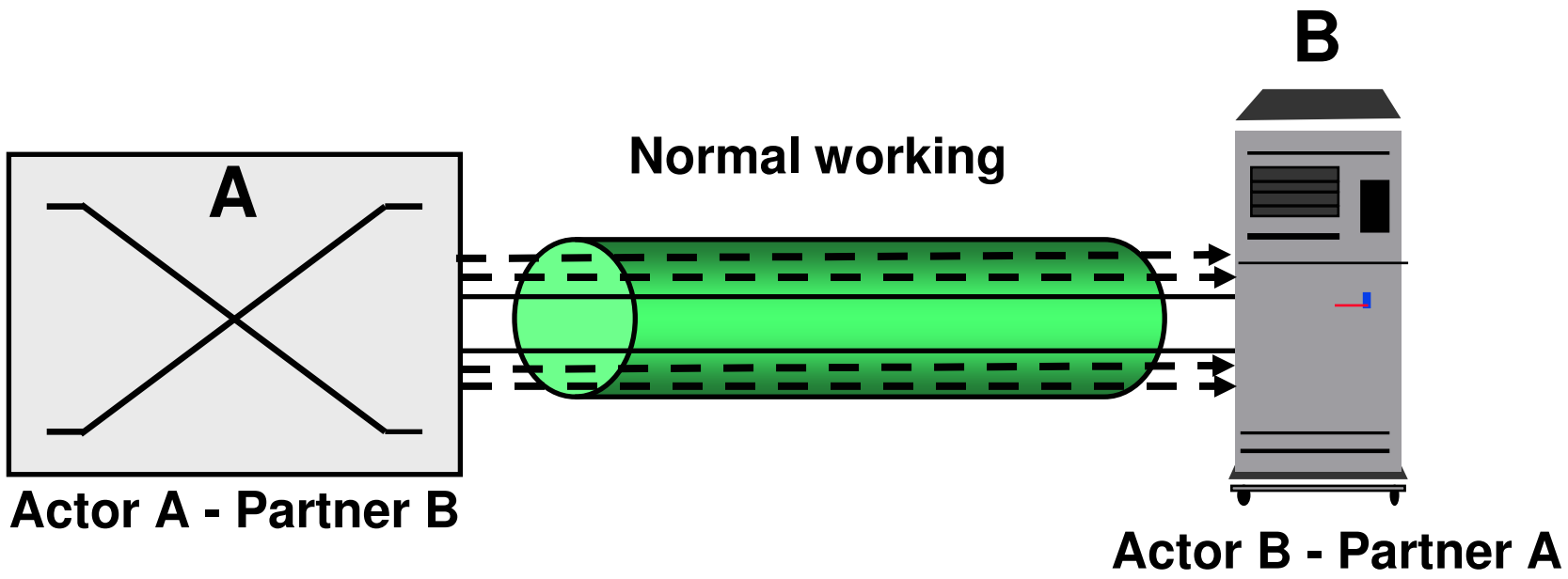
Copyright notice

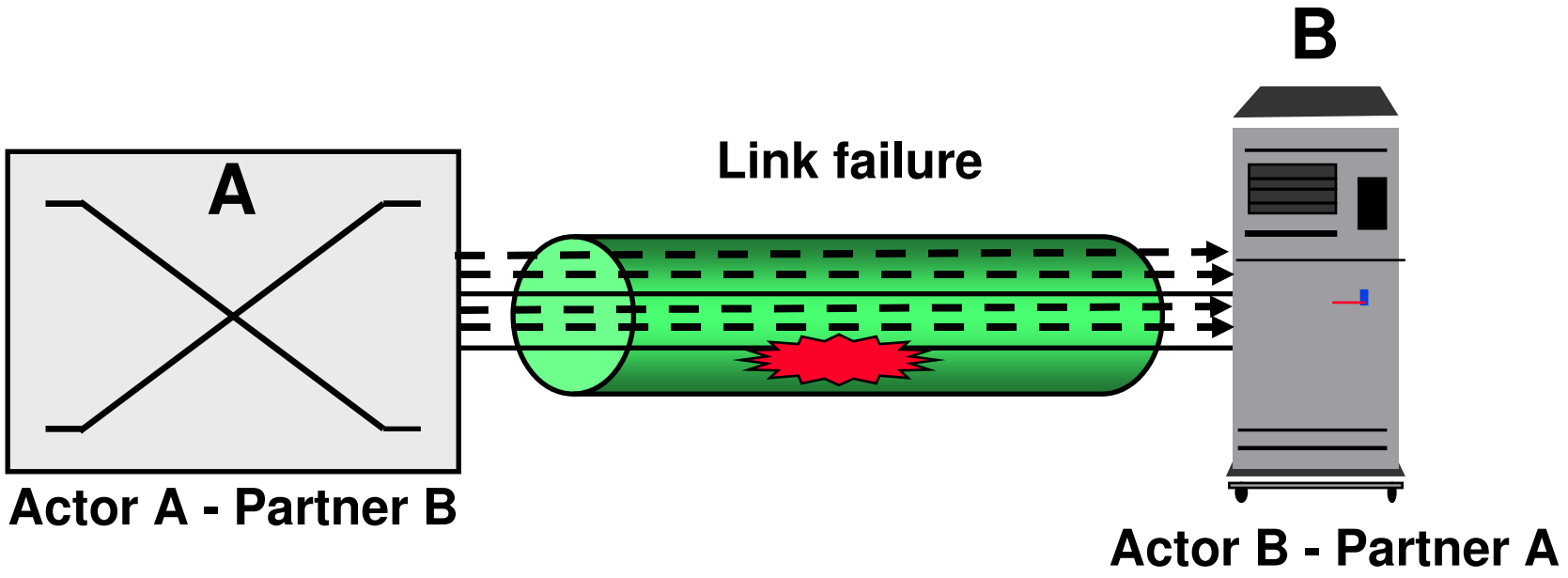
- This set of transparencies, hereinafter referred to as slides, is protected by copyright laws and provisions of International Treaties. The title and copyright regarding the slides (including, but not limited to, each and every image, photography, animation, video, audio, music and text) are property of the authors specified on page 1.
- The slides may be reproduced and used freely by research institutes, schools and Universities for non-profit, institutional purposes. In such cases, no authorization is requested.
- Any total or partial use or reproduction (including, but not limited to, reproduction on magnetic media, computer networks, and printed reproduction) is forbidden, unless explicitly authorized by the authors by means of written license.
- Information included in these slides is deemed as accurate at the date of publication. Such information is supplied for merely educational purposes and may not be used in designing systems, products, networks, etc. In any case, these slides are subject to changes without any previous notice. The authors do not assume any responsibility for the contents of these slides (including, but not limited to, accuracy, completeness, enforceability, updated-ness of information hereinafter provided).
- In any case, accordance with information hereinafter included must not be declared.
- In any case, this copyright notice must never be removed and must be reported even in partial uses.

IEEE 802.3ad

- Standard which replaces proprietary solution to group/aggregate more ports
  - useful to increase bandwidth
  - It's able to manage efficiently redundacy link between two network devices
  - normally used on link between switches, seldom between switch and computer







802.3ad details

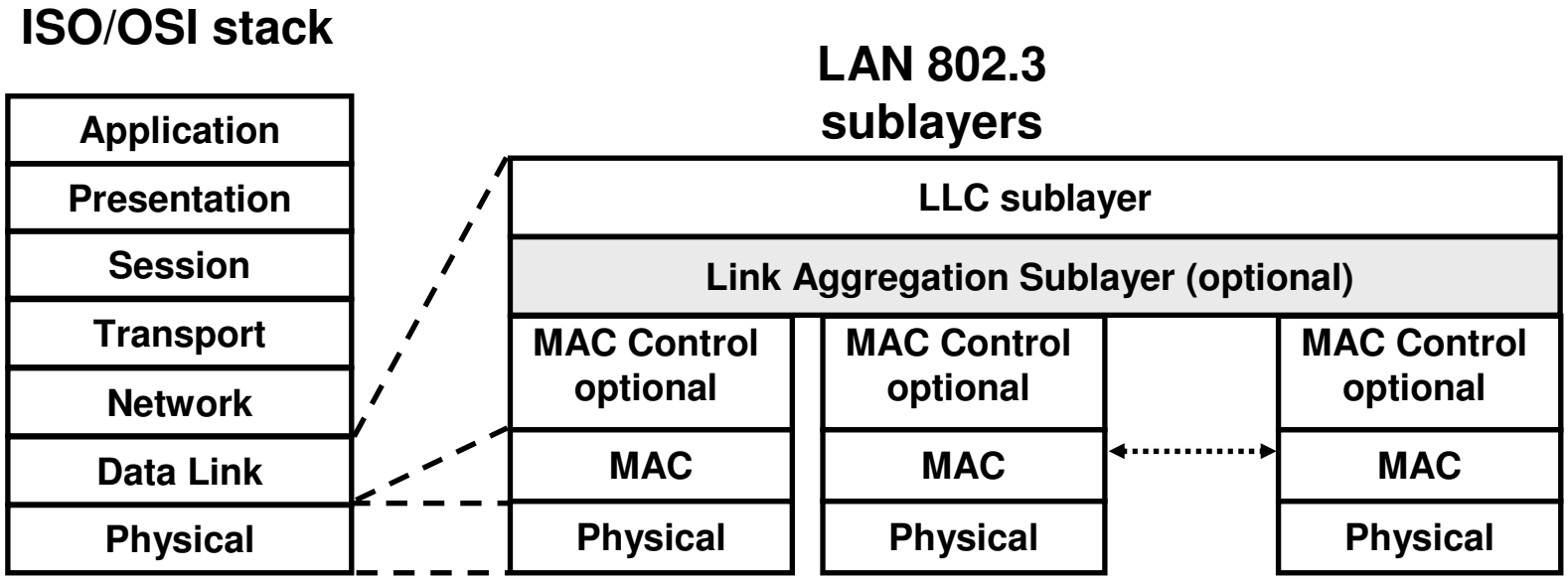
- Link aggregation is possible only on full duplex links
- Band increasing
  - Multiple link are grouped in a logical link
  - Bandwidth grow up is incremental
- Load sharing
  - Client's traffic can be shared out on multiple link
  - Increased reliability
  - link failure in an aggregation set doesn't affect comunication between partners at the edge

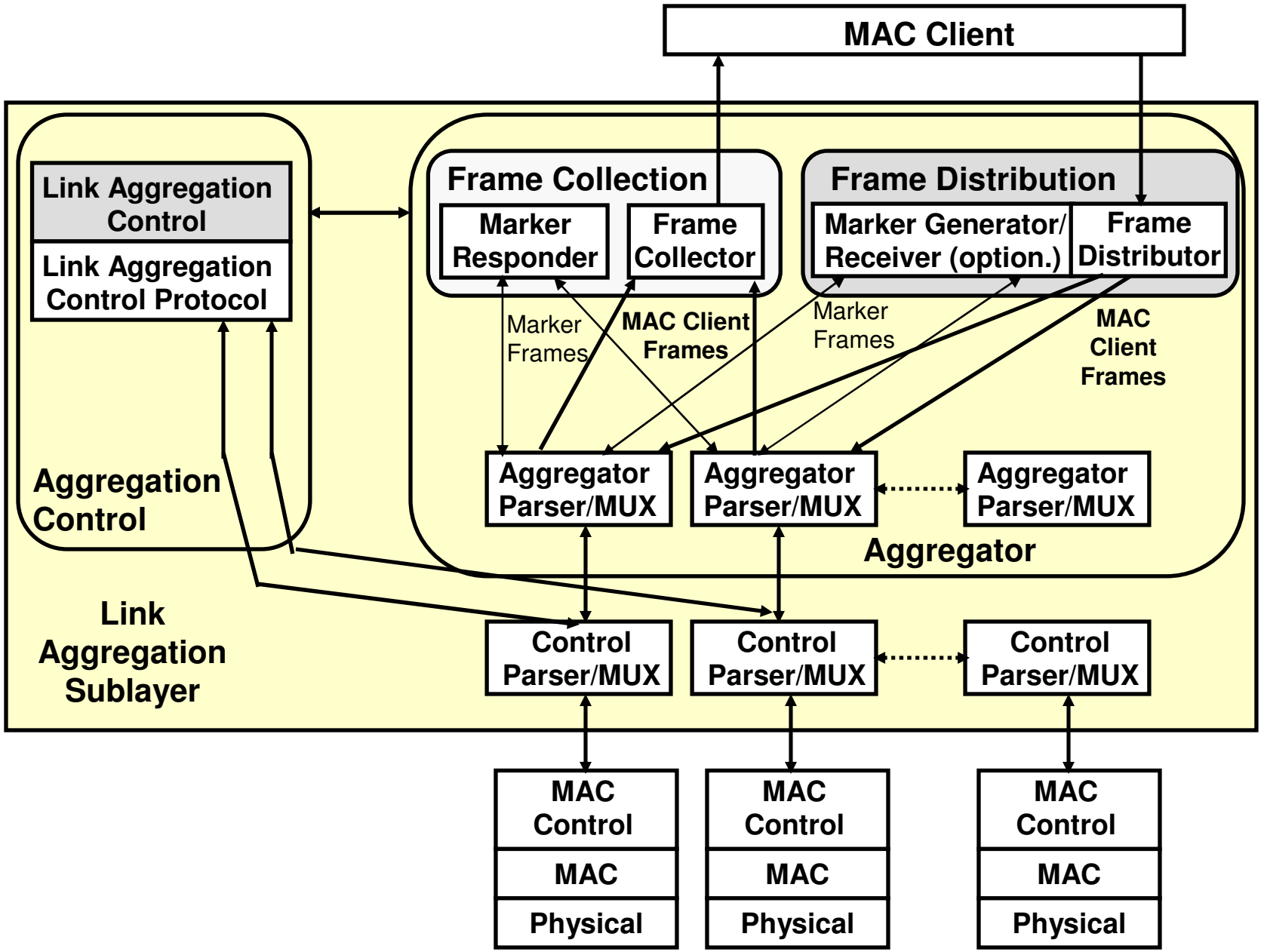
802.3ad details

- Fast convergence
  - aggregation set can converge in a new configuration in less than 1 second
- Every physical link within an aggregation group must have same speed
- Automatic aggregation configuration through LACP (Link Aggregation Control Protocol)
  - Multicast transmission of LACPDU

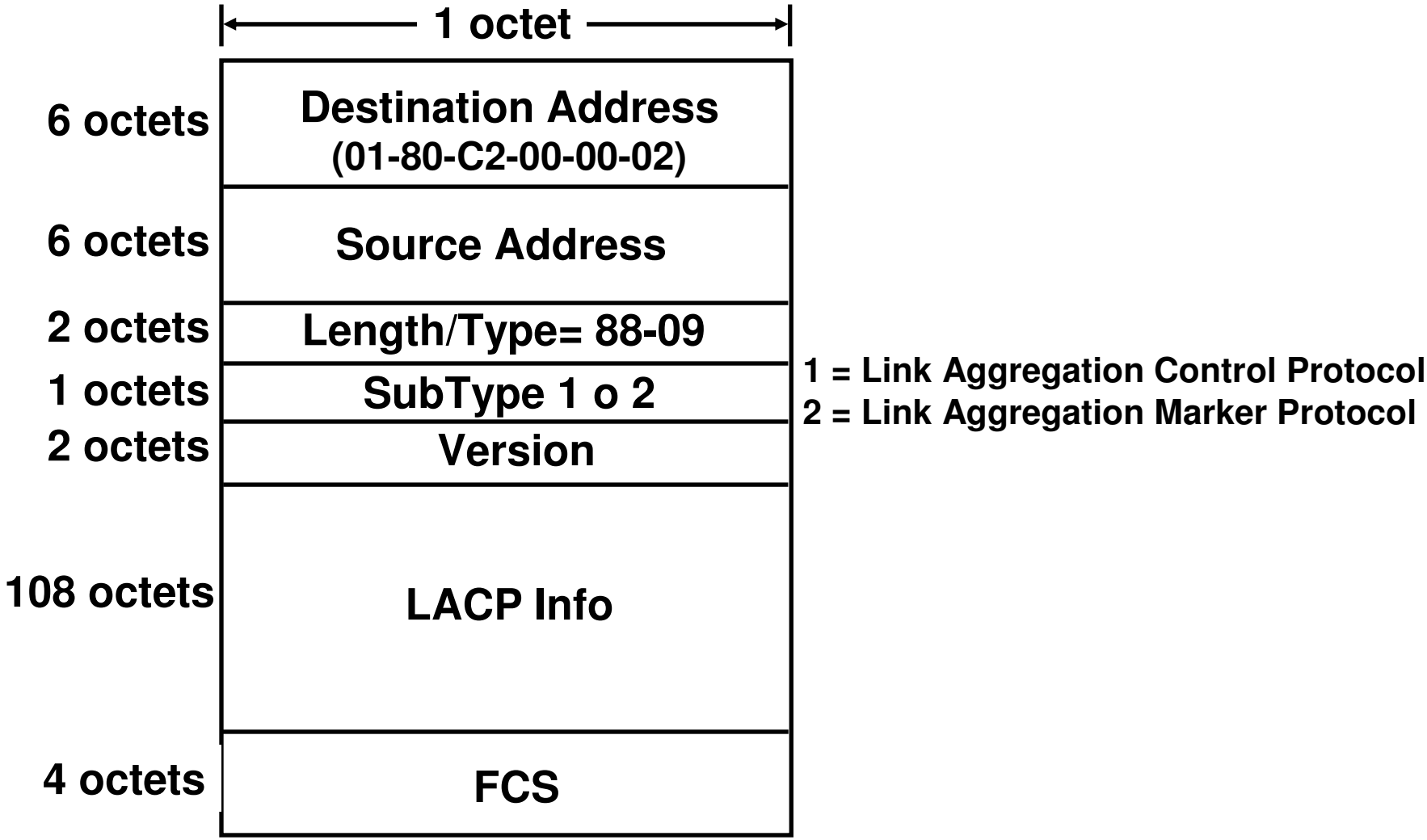
802.3ad and OSI layer

- Link aggregation sublayer is inserted between LLC level and MAC layers of the single ports to aggregate





LACPDU



Link Aggregation Group Identifier (LAG ID)

- Used to automatically verify if two ports of a link share the same membership group
  - Complete LAG ID is made up of a local and remote LAG ID
    - Actor LAG ID is the local one
    - Partner LAG ID is the remote one

Link Aggregation Group Identifier (LAG ID)

- LAG ID has the following parameters:
  - System Identifier (System Priority + MAC Address)
  - Operational Key assegned to ports in the LAG
  - Port Identifier (Port Priority + port number)
    - Unnecessary parameter in some cases, if so it is set to zero

Link Aggregation Group Identifier (LAG ID)

- To establish the membership at the same aggregation group between two ports of a link, local (actor) and remote (Partner) LAG ID are written in LACP packet
  - S e T variable for local and remote System ID
  - K e L variable for local and remote Operational Key value
  - P e Q variable for local and remote Port ID

Example of Partner parameter to build the complete LAG ID of a link

	Partner SKP	Partner TLQ
System Parameters (S, T)	System Priority = 0x8000 (see 43.4.2.2) System Identifier = AC-DE-48-03-67-80	System Priority = 0x8000 (see 43.4.2.2) System Identifier = AC-DE-48-03-FF-FF
Key Parameter (K, L)	Key = 0x0001	Key = 0x00AA
Port Parameters (P, Q)	Port Priority = 0x80 (see 43.4.2.2) Port Number = 0x0002	Port Priority = 0x80 (see 43.4.2.2) Port Number = 0x0002

The complete LAG ID derived from this information is represented as follows, for an Individual link:

[(SKP), (TLQ)] = [(8000,AC-DE-48-03-67-80,0001,80,0002), (8000,AC-DE-48-03-FF-FF,00AA,80,0002)]

Packets distribution on aggregate ports

- Standard doesn't define an algorithm to distribute packets to the port
  - No packets segmentation and reassembling
  - More conversations over a port
  - A conversation can be moved to an another port because of load balancing or *link failure*
  - Aggregation can have one or more link
  - Standard lists possible packets distribution criteria over ports

Packets distribution on aggregate ports

- Standard doesn't define an algorithm to distribute packets to the port
  - Two switches of different vendors can use different packets distribution algorithms
    - it can not be working
    - distribution could be not optimal
  - It will be better if switches at the edge of an aggregation link belong to the same vendor

Packets distribution on aggregate ports

- Conversation are assigned on:
  - Source MAC Address
  - Destination MAC Address
  - Receiving port
  - Destination kind (singlecast, multicast, broadcast)
  - Length/Type value
  - Higher Layer Protocol (for example 4 Layer ports)
  - Mix of previous criteria

Cisco distribution criteria

MAC address pairs **Source and Destination**	**Last** **2 bit**	**X-OR** **result**	**Chosen** **link**
Source MAC Address 00-00-00-00-00-01 Destination MAC Addr. 00-00-00-00-00-04	**01** 00	**01**	**Link 2**
Source MAC Address 00-00-00-00-00-02 Destination MAC Addr. 00-00-00-00-00-05	**10** 01	**11**	**Link 4**
Source MAC Address 00-00-00-00-00-03 Destination MAC Addr. 00-00-00-00-00-07	**11** 11	**00**	**Link 1**
Source MAC Address 00-00-00-00-00-06 Destination MAC Addr. 00-00-00-00-00-08	**10** 00	**10**	**Link 3**

Standby link

- Enables link to act as backup of other links
- Links belonging to two aggregation groups
  - Link with higher priority became active
  - Link with lesser priority became standby
- When the aggregable link number is *lesser* of the number of link between two switches (which share the same assigned administrative key):
  - link with higher priority are activated while the other are put in standby mode

Standby link and dynamic keys assignment

- Administrative key assignment to active link
  - Different keys for standby link and active link
  - standby link uses a new key but it keep trace of the active link's group key
  - in case of active link failure, a new aggregation group must be created. This group will exclude the broken link and will include the stand-by one
    - Key reassegnment for the new group

Dynamic keys assignament

- The two equipments use the same key on the port aggregation
- The key is assigned by the equipment that has the higher priority
  - priority criteria is the spanning tree one's
    - system priority (bridge priority if it is a switch) + MAC address
    - lower the value, higher the priority
    - Example: 8000-08-00-2B-50-20-00 has higher priority than 8000-08-00-2B-C4-E6-AA

Standby and equal level cost load sharing

- Assigned an unique administrative key for all ports; in practice ports are grouped together, although it is possible to group a smaller number of ports
  - Port can be active for a group and standby for another one
  - The higher priority aggregate become immediataly active using the administrative key
    - Priority = System-ID - Port ID
  - After a few seconds the lesser priority aggregate become active using another key

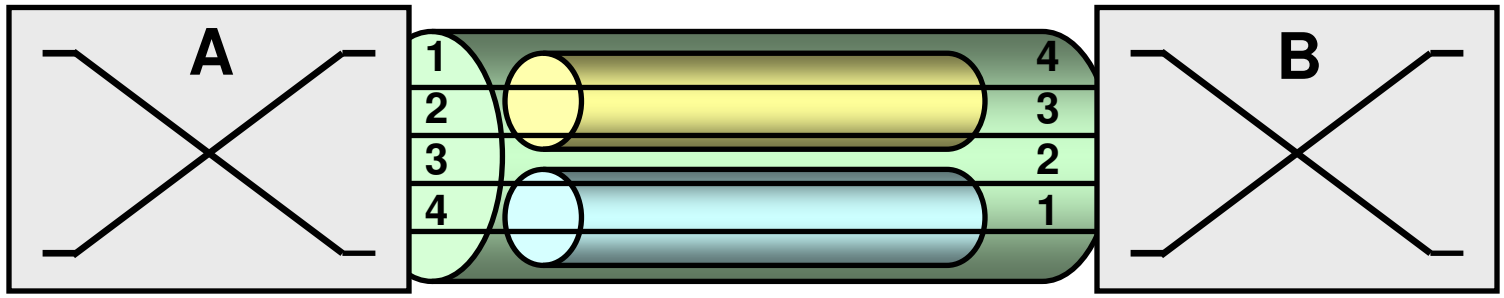
Example 1: standby and equal level cost load sharing

- Example 1: 4 parallel link, only 2 can be aggregate
  - Aggregation key = 1
- Switch A has higher priority
- Aggregation and backup rules on Switch A:
  - port 1 with port 2 or 4
  - port 2 with port 3 or 1
  - port 3 with port 4 or 2
  - port 4 with port 1 or 3

Example 1: standby and equal level cost load sharing

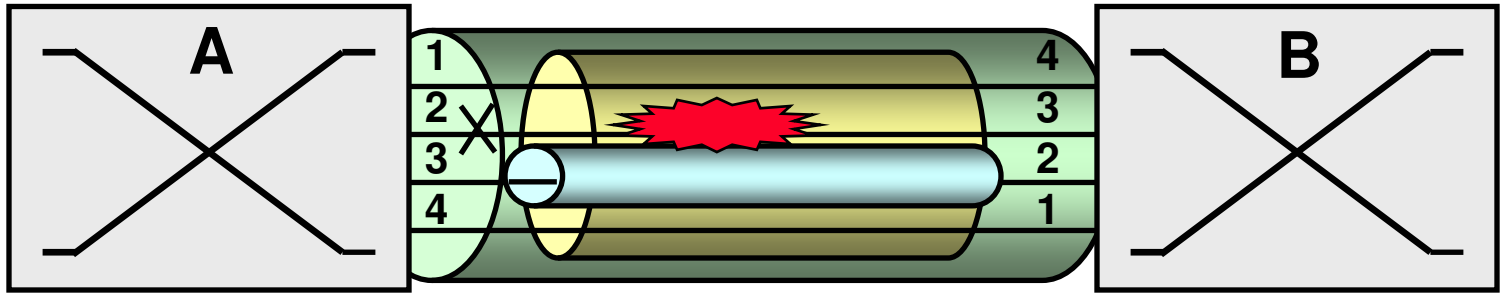
- The alghoritm:
  - select the 2 higher priority link as active
    - assign key = 1 to link A1-B4 and A2-B3
    - consider other link as standby
    - keep trace of key 1 on link A4-B1 and change the key on link A3-B2 and A4-B1
  - After few second it active link A3-B2 and A4-B1 as second port aggregate
  - load sharing on the two aggregate

Example 1: normal and fault conditions



Higher priority switch

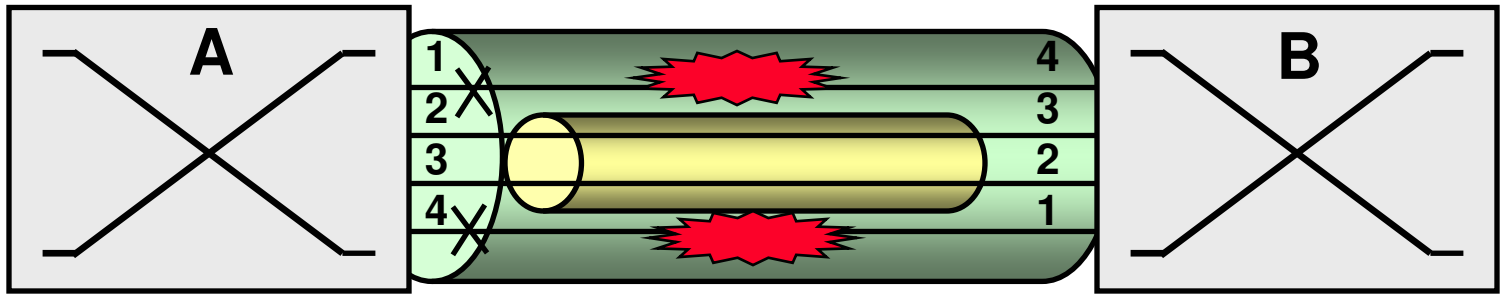
Normal working



Higher priority switch

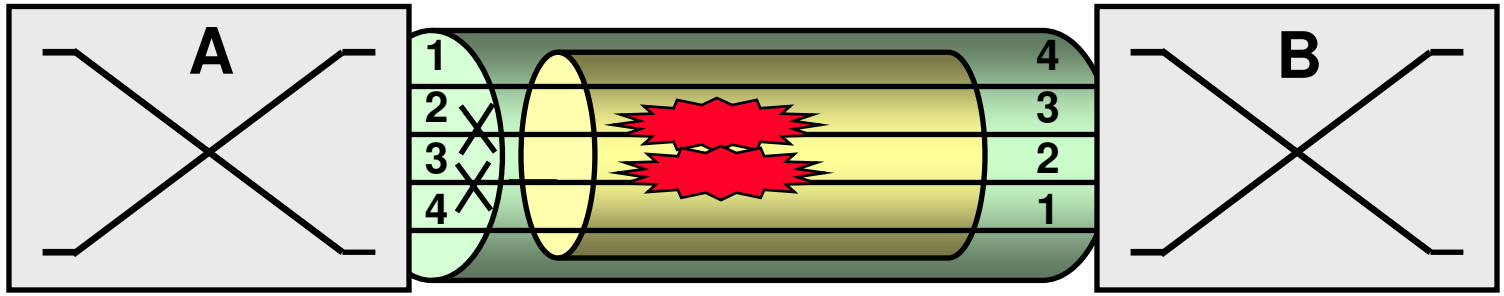
Failure on link A2-B3

Example 1: fault conditions



Higher priority switch

Failures on link A1-B4 and A4-B1



Higher priority switch

Failures on link A2-B3 and A3-B2

Link Aggregation and STP/RSTP

- It doesn't affect STP or RSTP
- We have to set Path Cost value to ports in order to reflect aggregate bandwith value
  - disable automatic path cost on ports
  - It will be better to use RSTP because has a wider range of Path Cost values

Link Aggregation and STP/RSTP