An Engineering guide to IEEE 802.1Q and IEEE 802.1p

Silvano Gai

Agenda

- VLAN
- IEEE 802 committees
- IEEE 802.1p
- IEEE 802.1Q
- The Cisco solution

Compass

- **VLAN**
  - **assigning frames to VLANs**
  - **tagging and baby giant frames**
  - **spanning tree(s)**
  - **access or independent VLANs**
  - **single/multiple filtering data base(s)**
  - **internetworking between VLANs**
- **IEEE 802 committees**
- **IEEE 802.1p**
- **IEEE 802.1Q**
- **The Cisco solution**

Assigning frames to VLANs

- **A station may be member of one or more VLANs**
- **Membership may be:**
  - **static**
    - **per port**
  - **dynamic**
    - **per MAC address**
    - **per protocol**
    - **per layer 3 address**
    - **per multicast address**
    - **“policy-based” (per application, per user, etc.)**

Frame tagging

- The tag contains the VLAN membership information
- Implicit tagging
  - no tag is added to the frame
  - easy in connection-oriented approaches
  - difficult for multicast/broadcast frames
- Explicit tagging
  - a tag is added to each frame
  - the tag carries the VLAN membership information
  - the tag may carry additional information

Baby Giants

- **The addition of extra bytes for the tag makes frames "Baby Giants"**
- **How to accommodate the extra bytes for the tag in the frame?**
  - **802.1 is persuading 802.3 to increase the maximum frame size from 1518 to 1522 (4 extra bytes)**

Explicit tagging

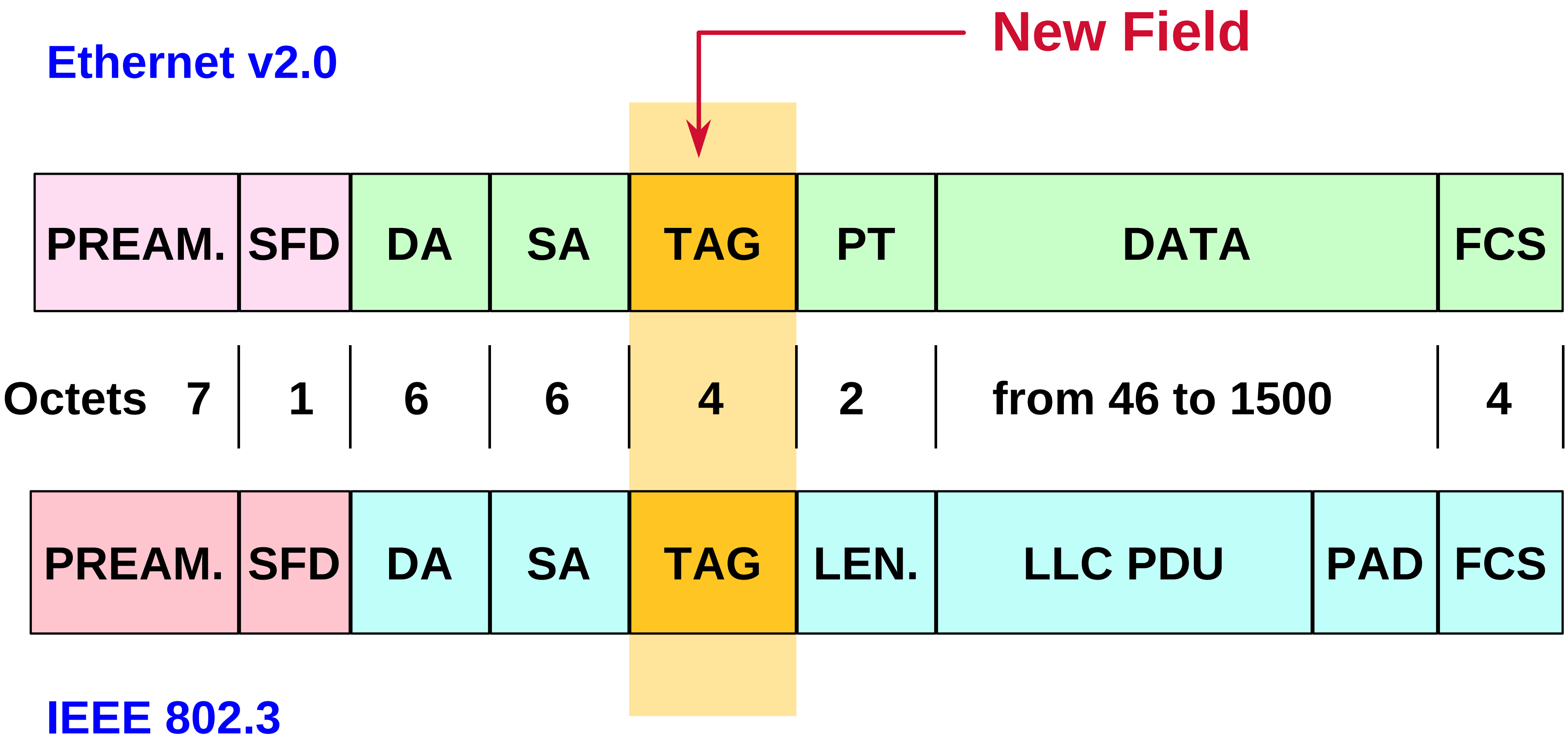
- **Where to position the tag in the frame?**
- **Two possibilities:**
  - **One level tagging**
    - **also called "Internal tagging"**
  - **Two level tagging**
    - **also called "External tagging"**
- **Both require to be implemented in ASIC for wire speed performance**

One level tagging

- The original frame is modified with the addition of the tag inside the frame
- The tagged frame has a valid format also for the “VLAN unaware”devices
  - MAC SA and DA are unchanged
  - an exception: it may be a baby giant

Example of one level tagging

- Tagging Ethernet - IEEE 802.3



Two-level tagging

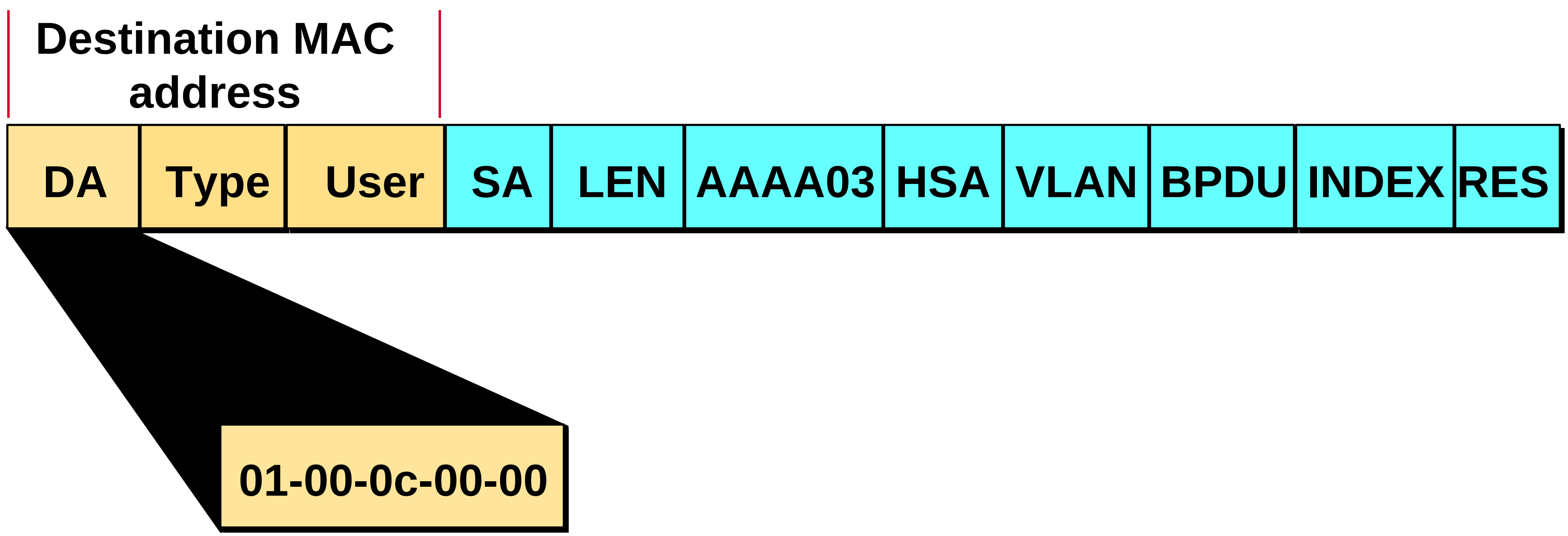
- The original frame is left unchanged
- A new external header is added in front of the original frame
  - New SA, DA, (RIF), Ethertype, and VLAN-ID
  - It is possible to support giant frames
- The RIF works better:
  - two-level tagging is a tunnelling mechanism
  - it is unclear how source routing works in 1Q
- A Tricky FCS fix-up in the new header would allow original frame FCS to be retained

Inter-Switch Link (ISL)

- **Original frame is encapsulated with ISL header and FCS, i.e. two level tagging**
- **Support up to 1,024 VLANs**
- **Implemented in ASICs provides wire speed performance**

ISL Header 26 bytes	Encapsulated frame 1 ... 24.5 KBytes	FCS 4 bytes

ISL Header Format



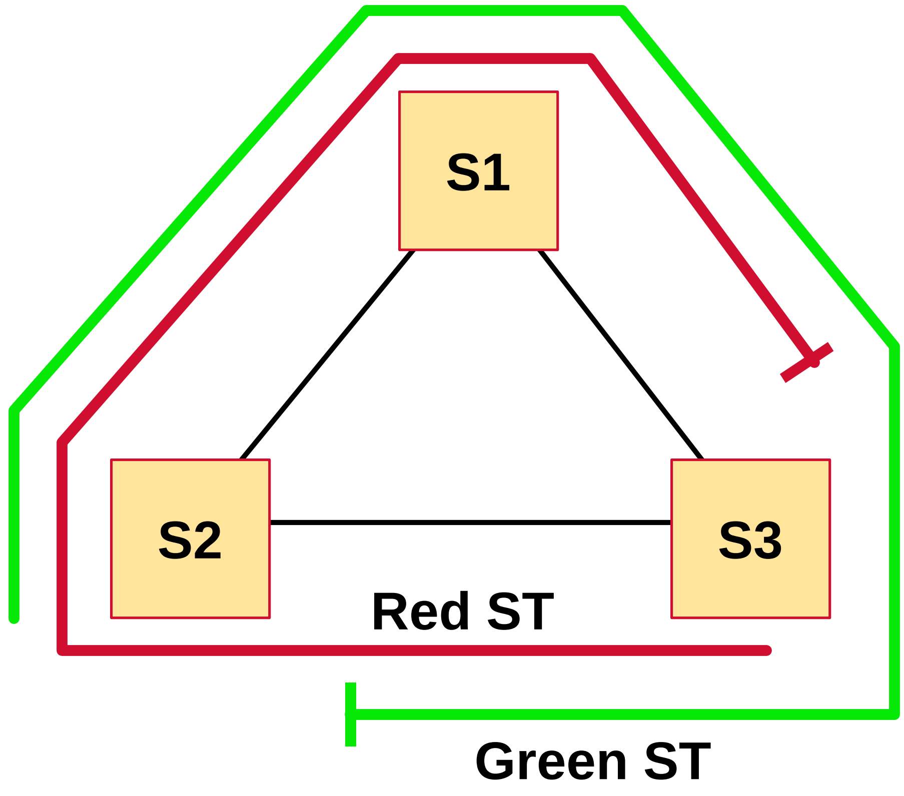
- The higher 40 bit - multicast destination address
- Lowest 8 bits used by type and user field

Spanning Tree

- Three different possibilities:
  - Single spanning tree
  - Per VLAN spanning tree
  - Shared spanning tree
- Single spanning tree does not allow:
  - multiple active topologies
  - load balancing
- Cisco implements one spanning tree per VLAN, at present

Multiple Spanning Trees

- **All links in the network are simultaneously used**



What is a VLAN?

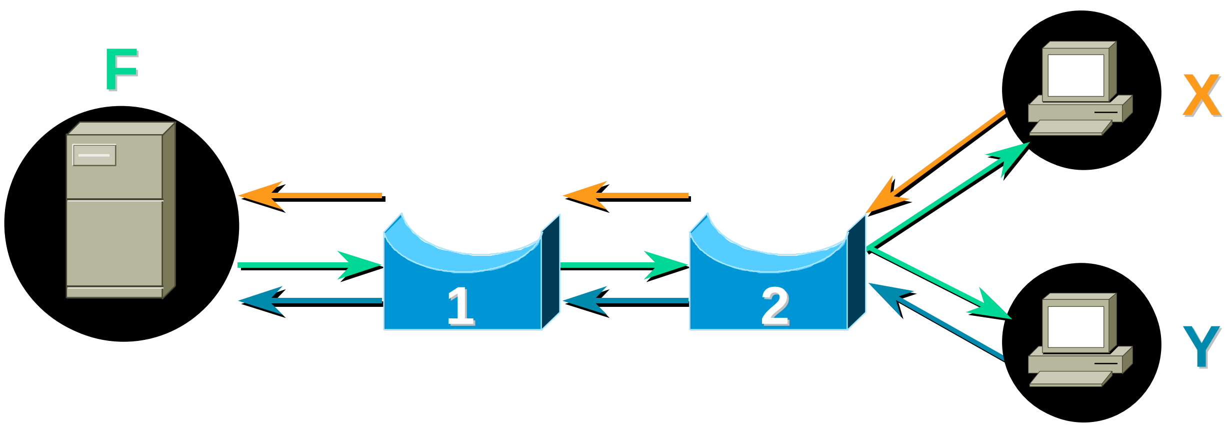
- **Two possible models**
  - **Access VLANs**
    - **VLANs are a clever way to specify filters to limit endstation-to-endstation connectivity on a single, bridged LAN**
  - **Independent VLANs**
    - **VLANs are a clever way to utilize one physical plant to carry multiple, independent bridged LANs**

Access VLANs

- It is a single bridged LAN, with filters
  - filtering helps in scaling somewhat larger
- Access VLANs require a single spanning tree for the whole network, because they have one filtering database for all VLANs in each bridge

Access VLANs

- One-way VLANs
  - Half-duplex conversations between different VLANs
  - Bridge 1 never sees F's source on yellow or blue, nor X's or Y's sources on green
  - Filtering database *must* ignore "color"

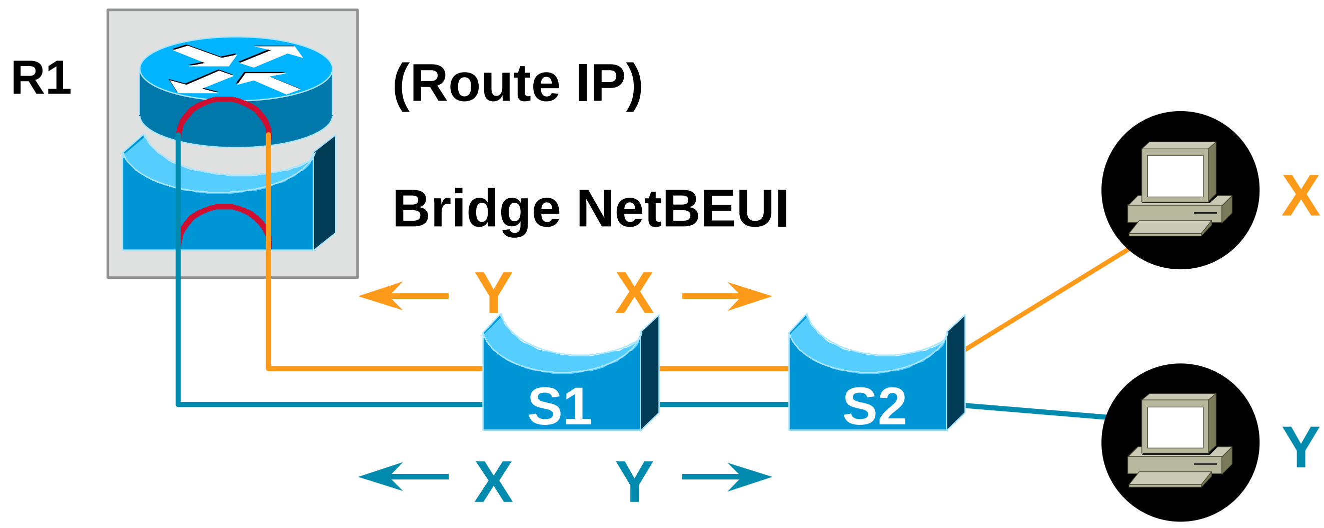


Independent VLANs

- It is possible to build large networks
  - if the scope of each VLAN is not global
  - routers plus bridged LANs are known to scale well
- Per VLAN filtering database
- They work with:
  - a single spanning tree
  - one spanning tree per VLAN
  - multiple VLANs in each of several spanning trees

Independent VLANs

- They support duplicate MAC addresses
  - DECNet phase IV routers and two-Ethernet Sun workstations
  - to route some protocols and bridge others



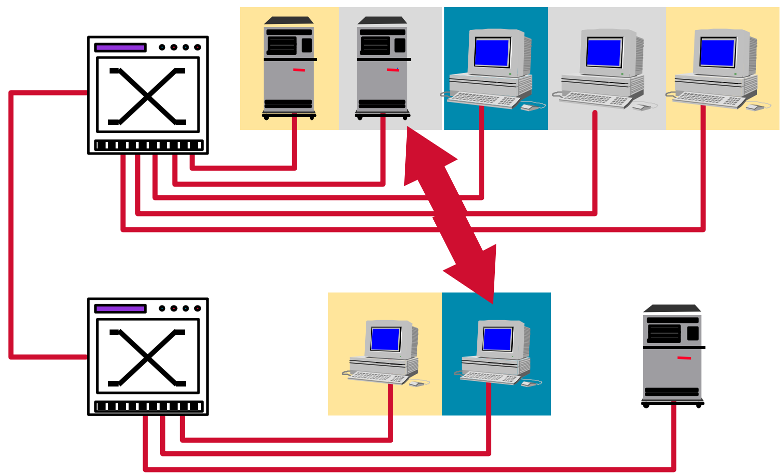
When R1 bridges some protocol between X and Y on different VLANs, S1 and S2 see duplicate MAC addresses for Y and X.

Number of “filtering databases”

- **MFD/SE**
  - **Multiple Filtering Database - Single Entry**
  - **Natural solution for independent VLANs**
  - **Compatible with multiple spanning trees**
- **SFD/ME**
  - **Single Filtering Database - Multiple Entry**
  - **Solution adopted in Access VLAN to try to support duplicated MAC addresses**
  - **Requires a single spanning tree**
- **Duplicate MAC addresses are common!!!**

Internetworking between VLANs

- Using routers
  - classical approach
  - scale well
- Layer 2 shortcuts
  - switches create shortcuts between VLANs
  - limited scalability



Compass

- VLAN
- **IEEE 802 committees**
  - **IEEE 802.1**
  - **IEEE 802.3ac**
  - **standard tagging scheme**
- IEEE 802.1p
- IEEE 802.1Q
- The Cisco solution

IEEE 802 LMSC

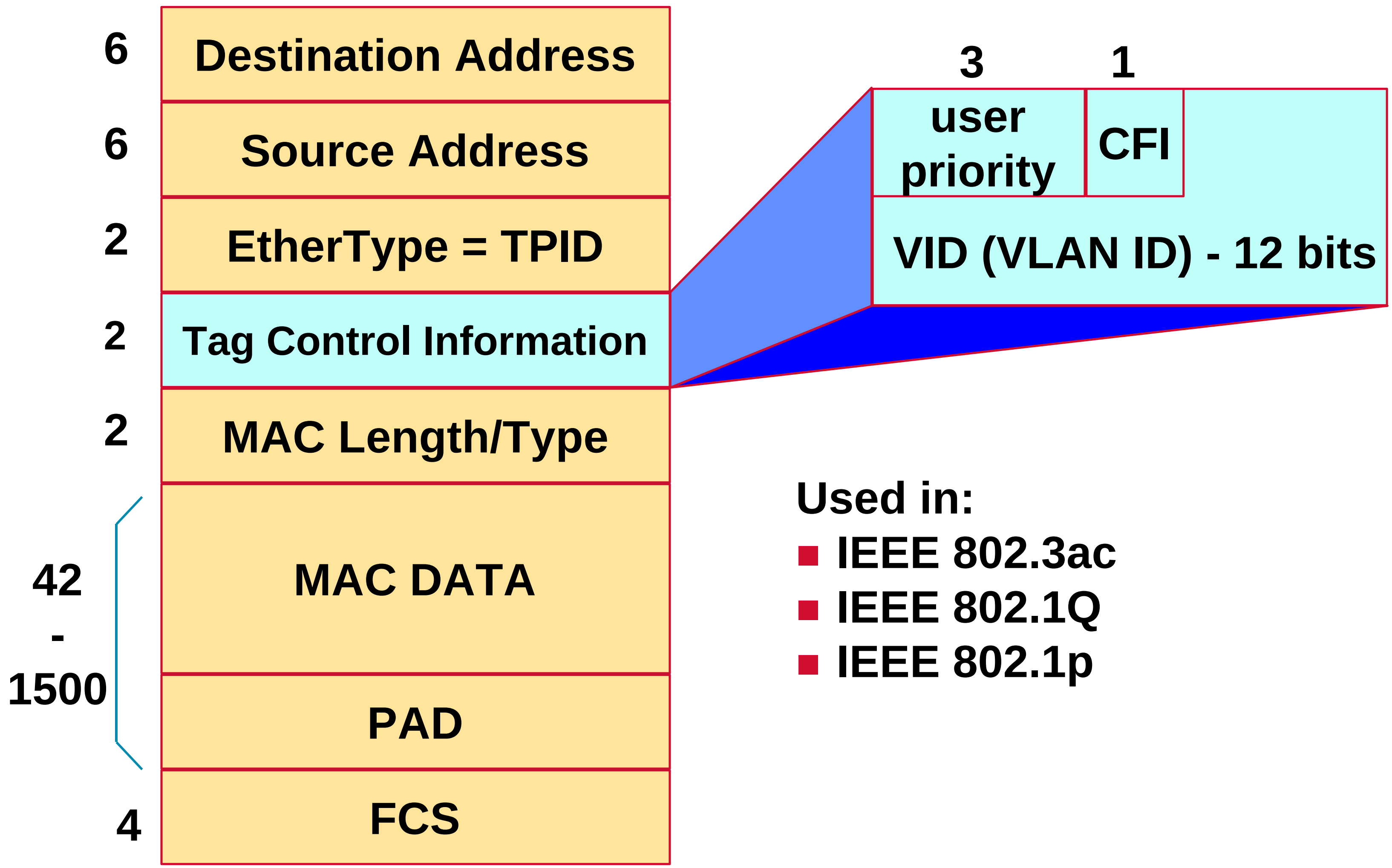
- 802 LAN/MAN Standards Committee
  - 802.1: Higher Layer Interfaces (*)
    - 802.1D (transparent bridging)
    - 802.1G (metro transparent bridging)
    - 802.1H (translation bridging)
    - **802.1D Reaffirmation**
    - **802.1p Priorities/GARP/GMRP**
    - **802.1Q VLANs/GVRP**
  - 802.3: CSMA/CD (Ethernet)
    - **802.3ac**
  - 802.5: Token Ring
  - Others

(*) IEEE 802.1 started working on VLANs in late 1995 and it has still not finished

IEEE 802.3ac

- **IEEE Standards for Local and Metropolitan Area Networks:**
  - **Supplement to Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method & Physical Layer Specification**
    - **Frame Extension for Virtual Bridged Local Area Networks (VLAN) Tagging on 802.3 Networks.**
- **Draft 1**
- **Main topic:**
  - **Extend Maximum Frame size from 1518 to 1522 octets**

Tagging scheme



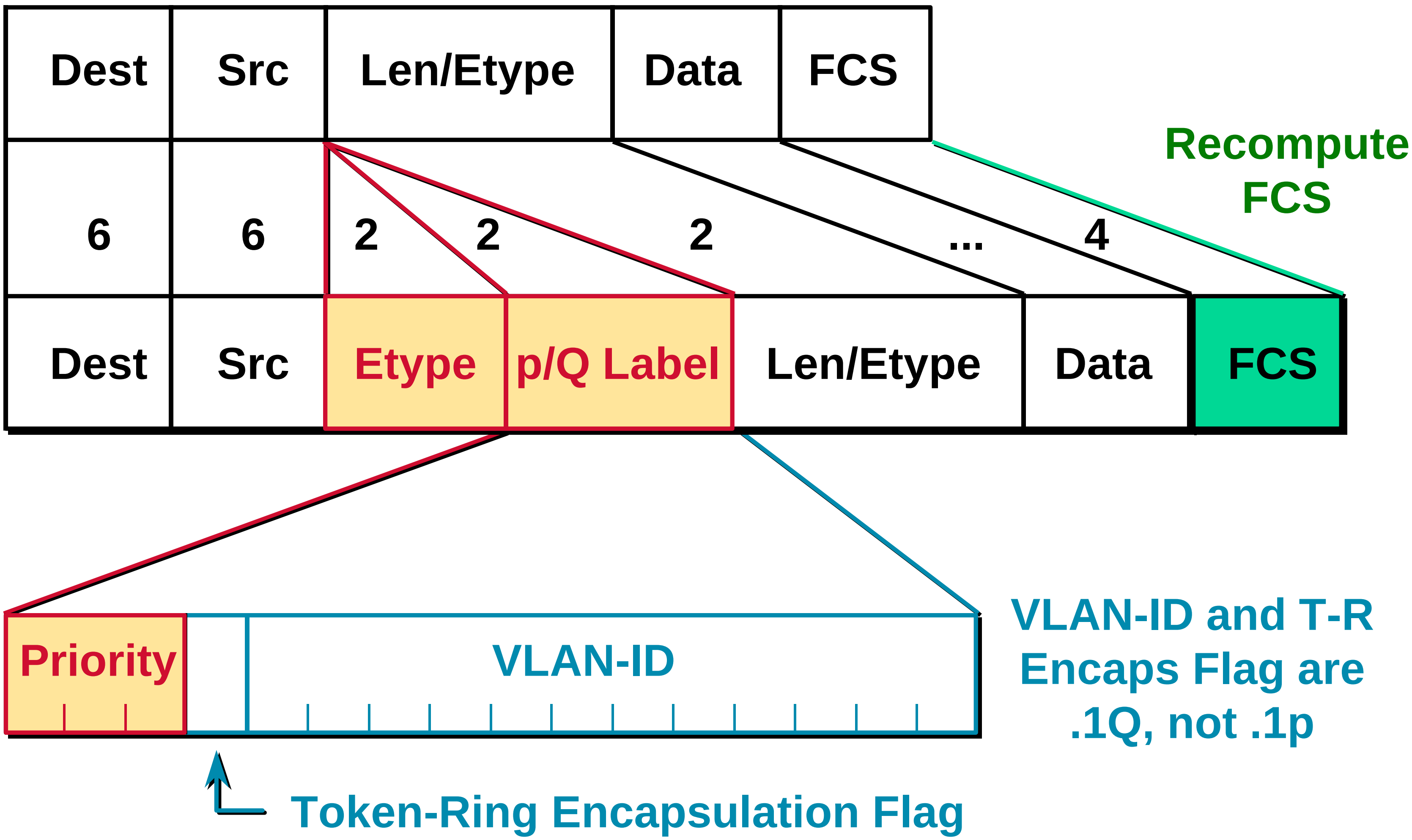
Used in:

- IEEE 802.3ac
- IEEE 802.1Q
- IEEE 802.1p

One-level tagging

- **Insert Ethertype and VLAN-ID after MAC source (or RIF), but before original Ethertype/Length (or LLC)**
- **Includes T-R Encapsulation bit so that T-R frames can be carried across Ethernet backbones without 802.1H translation of data contents**
- **802.1p and 802.1Q share the same tag**

802.1p/Q tags



Dest
Src
Len/Etype
Data
FCS
6
6
2
2
2
...
4
Recompute FCS
Dest
Src
Etype
p/Q Label
Len/Etype
Data
FCS
Priority
VLAN-ID
VLAN-ID and T-R Encaps Flag are .1Q, not .1p
Token-Ring Encapsulation Flag

Compass

- VLAN
- IEEE 802 committees
- **IEEE 802.1p**
  - **Expedited traffic capabilities**
  - **Bridge architecture**
  - **GARP, GMRP**
- IEEE 802.1Q
- The Cisco solution

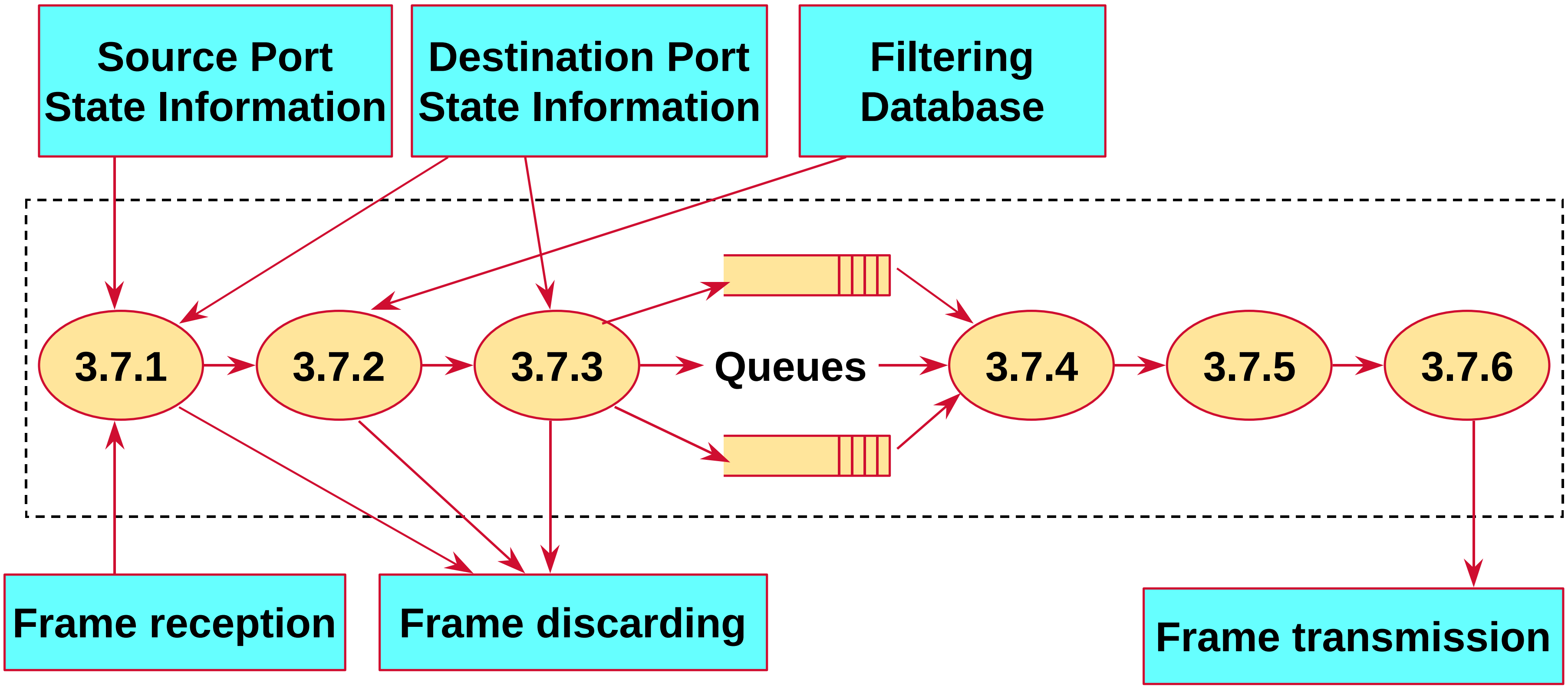
IEEE 802.1p

- **IEEE Standards for Local and Metropolitan Area Networks:**
  - **Supplement to Media Access Control (MAC) bridges: Traffic Class Expediting and Dynamic Multicast Filtering**
- **Draft 8**
- **Two main topics:**
  - **Expedited traffic capabilities**
  - **Filtering services to support the dynamic use of Group MAC addresses**

Expedited traffic capabilities

- **Priority labeling**
  - **MAC-layer priority in the add-on tag**
  - **Priority *not* derived from MAC address**
- **Multiple output queues per output port**
  - **output queue selection based on 802.1p tag**
  - **maintains ordering only between frames at same priority**
- **802.1 is cooperating with IETF's ISSLL (Integrated Services over Specific Lower Layers)**
  - **mapping L3 RSVP requests to 802.1p priorities via a subnet bandwidth manager**

Bridge architecture



3.7.1 Enforcing topology restriction
3.7.2 Filtering Frames
3.7.3 Queueing Frames

3.7.4 Selecting frames for transmission
3.7.5 Mapping priority
3.7.6 Recalculating FCS

GARP

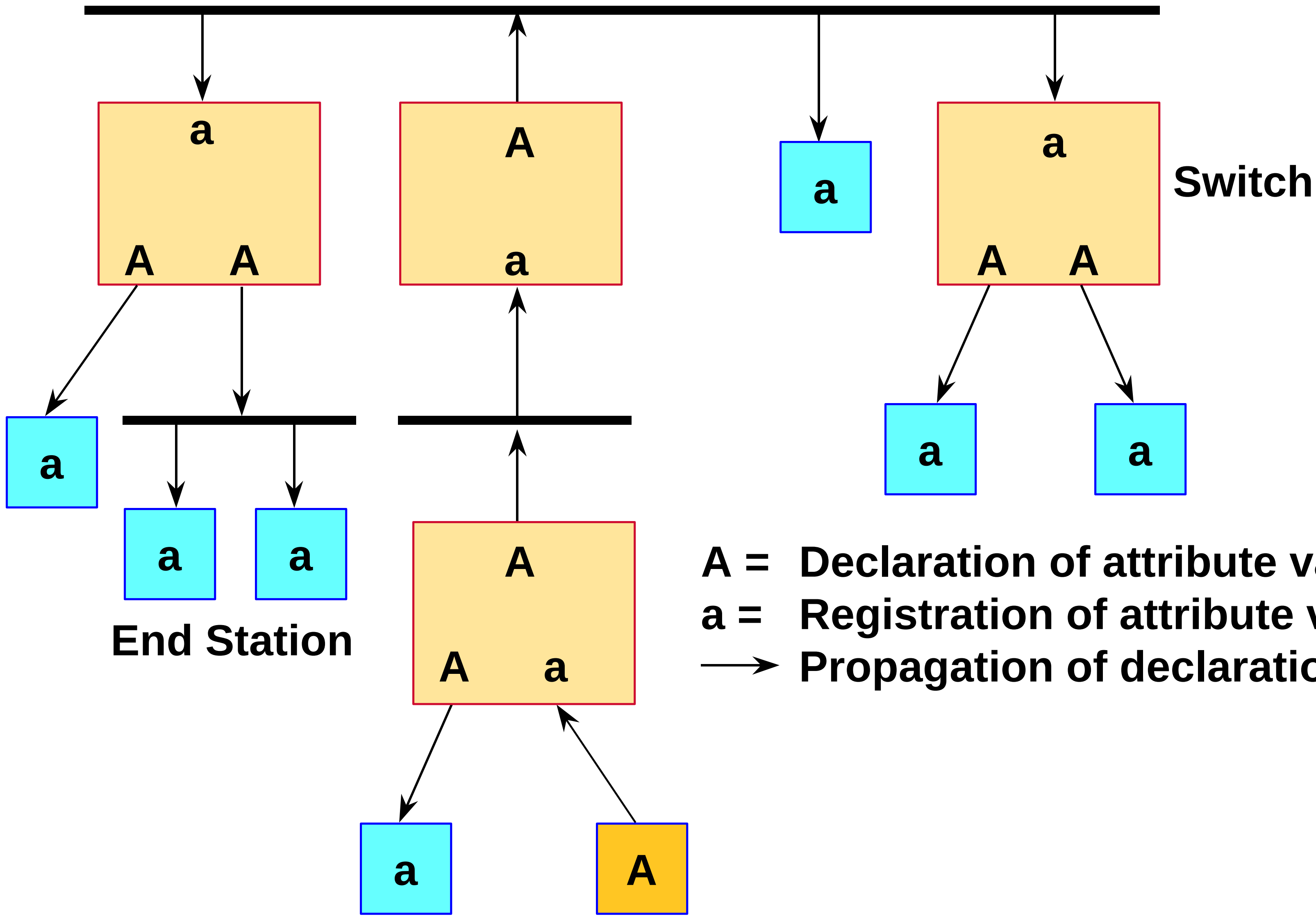
- **Generic Attribute Registration Protocol**
  - **Generic attribute dissemination capability**
  - **Used by participants in "GARP Applications (*)" (GARP Participants) to Register and de-register attribute values with other GARP participants within a bridged LAN**
  - **Attribute types and attribute values are specific of each GARP application**
  - **Designed to register anything**

(*) there are two GARP Applications already defined: GMRP and GVRP

GARP operation

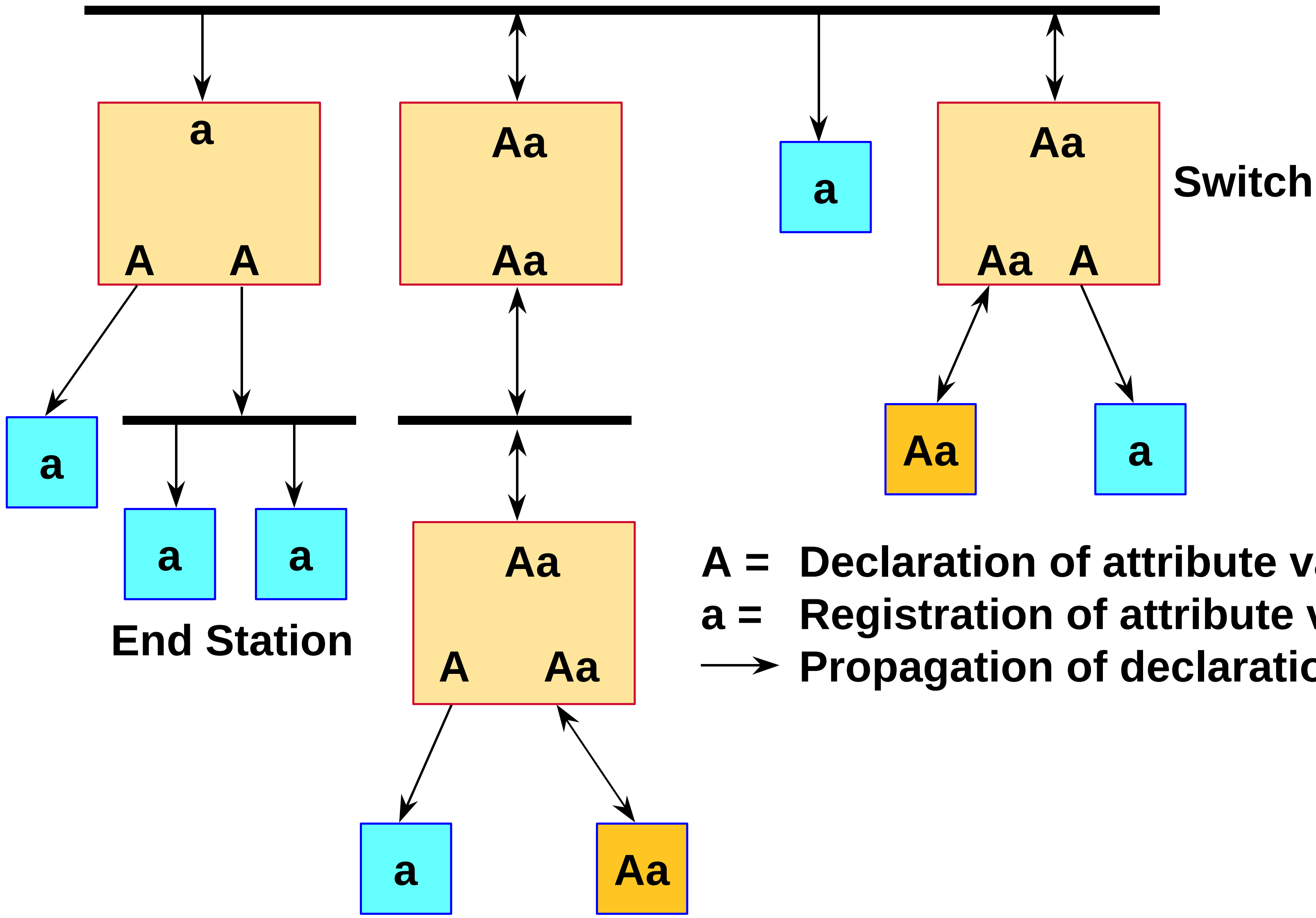
- GARP Applications
  - make/withdraw declarations relative to attribute values
  - this results in registration/deregistration of attribute values in other GARP participants
  - registration/deregistration is recorded in a state variable in the “Registar state machine”
    - only on the port that receives the GARP PDU containing the declaration
    - even on ports that are not ST forwarding
  - attribute values registered on ports belonging to the active topology are propagated to all the other bridge ports belonging to the active topology by the “Applicant state machine”

Declaration and Registration



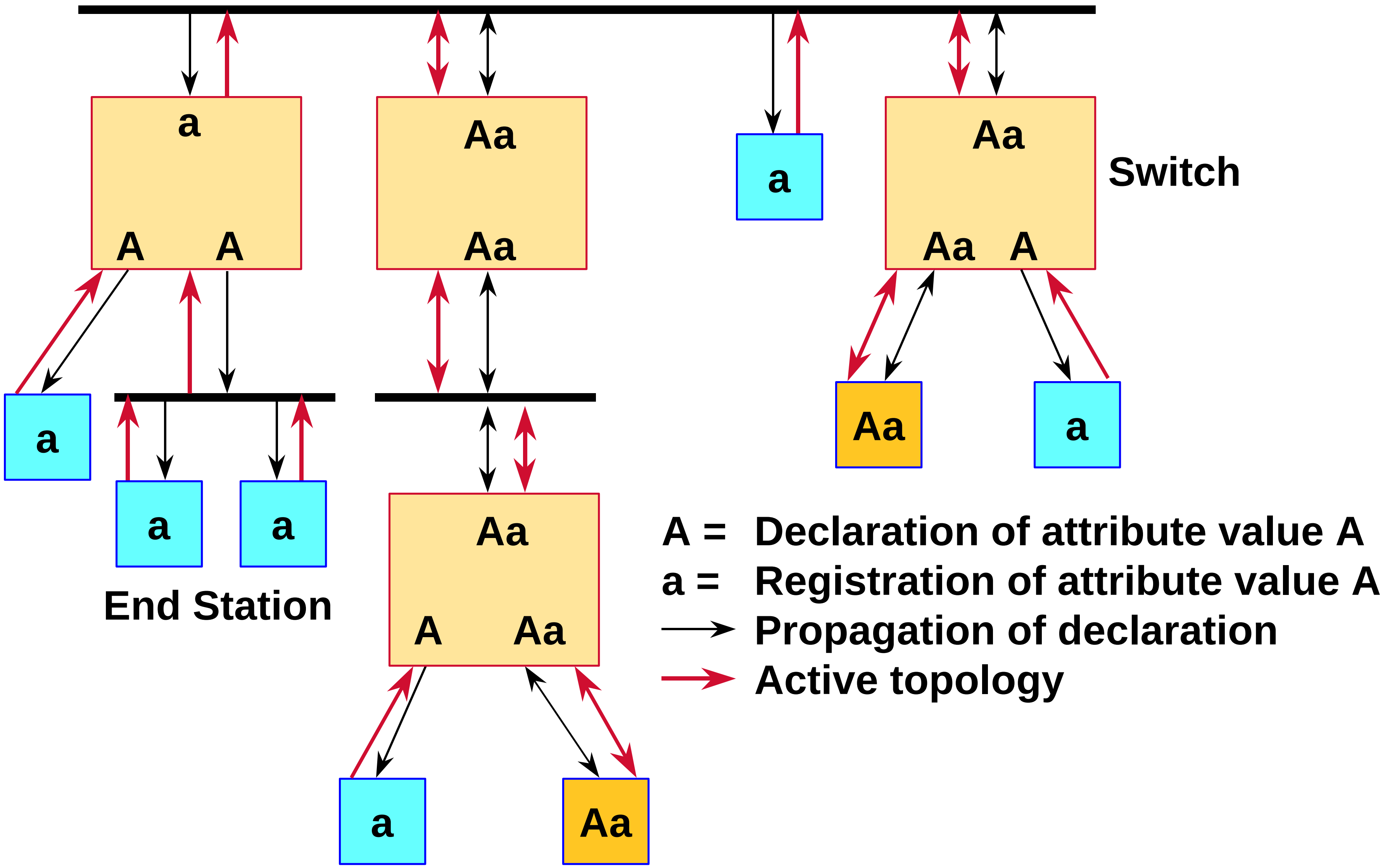
A = Declaration of attribute value A
a = Registration of attribute value A
→ Propagation of declaration

From two end stations



A = Declaration of attribute value A
a = Registration of attribute value A
→ Propagation of declaration

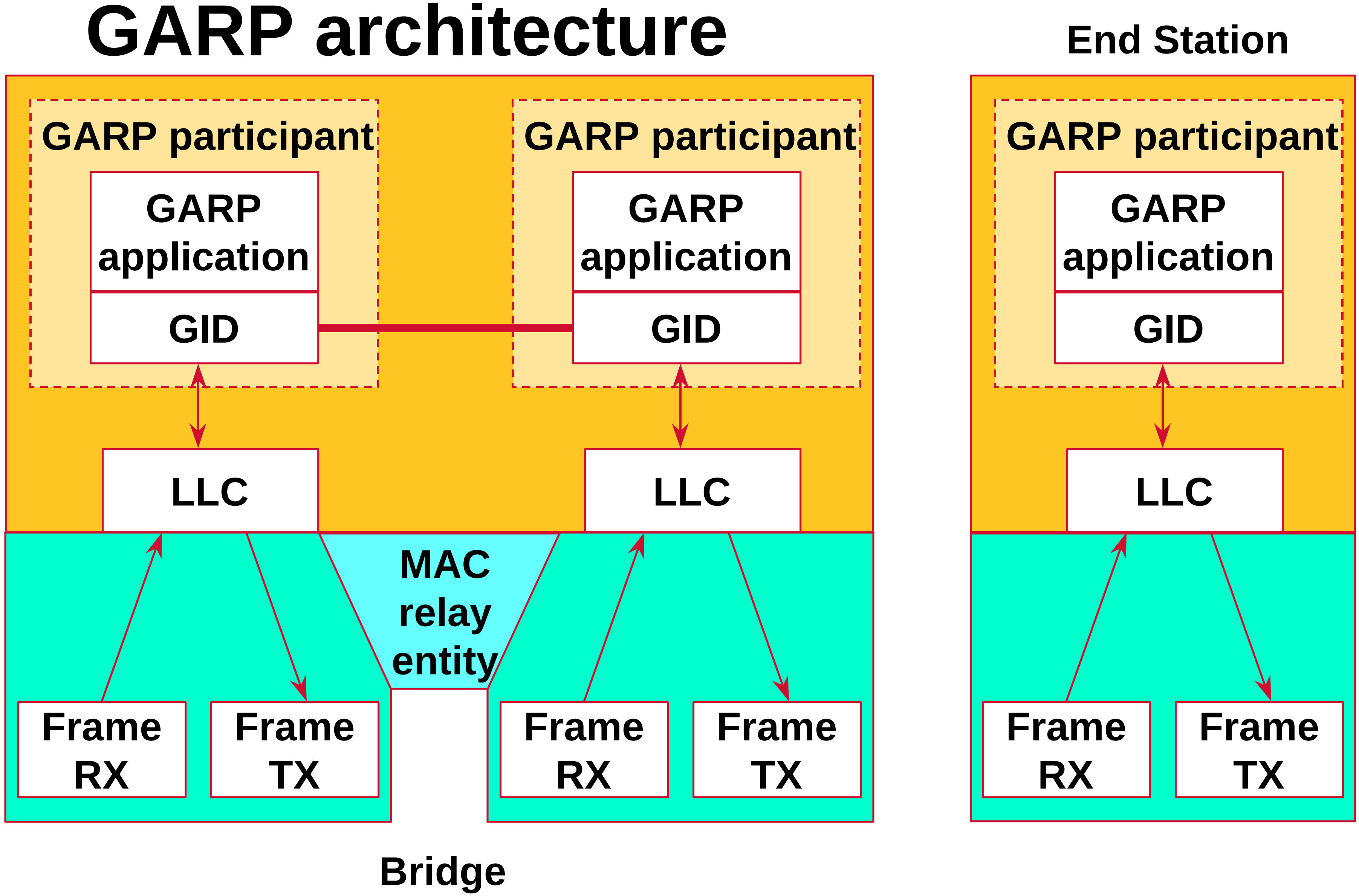
Active topology



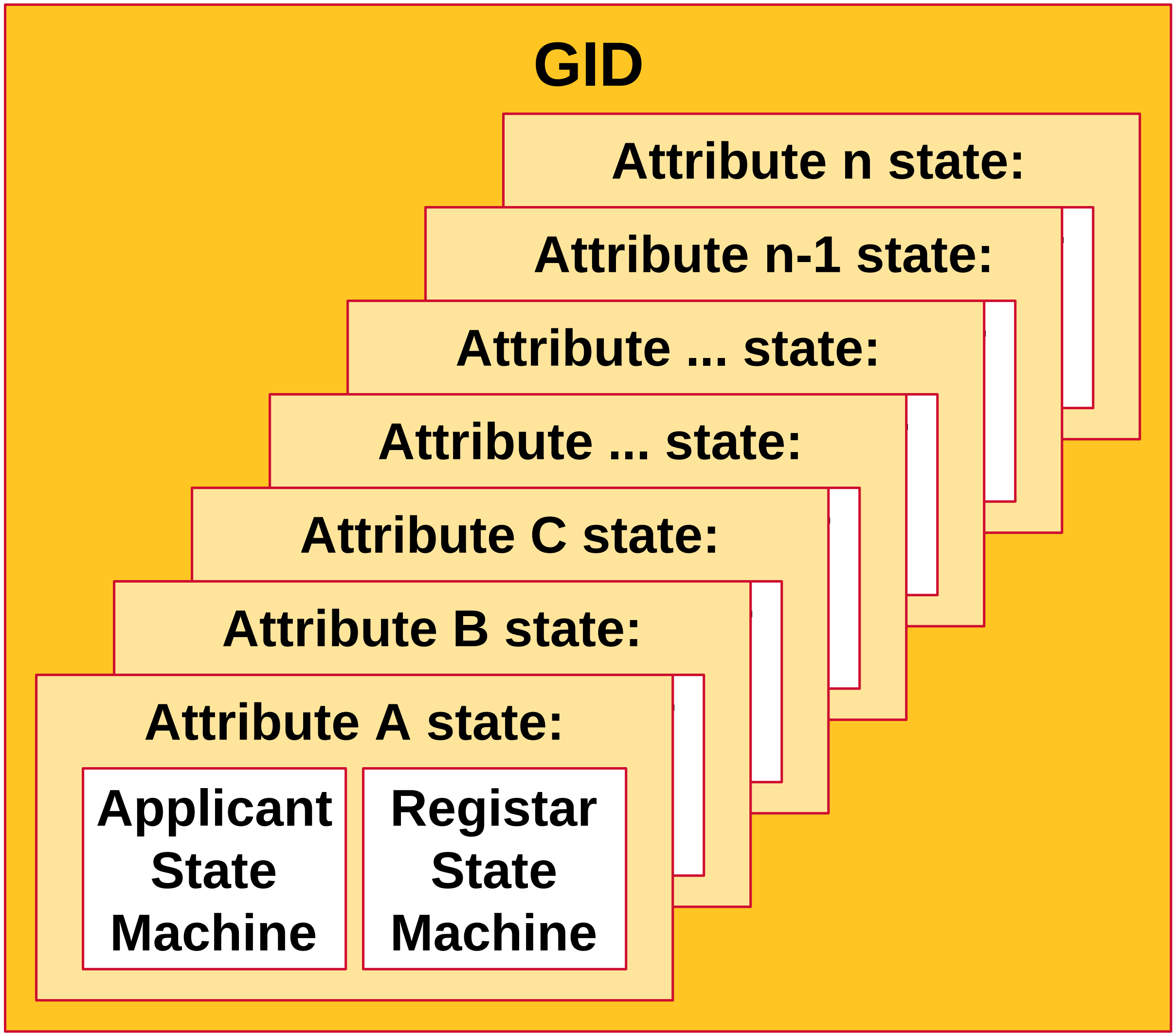
GARP Participant

- In each bridge it consists of a GARP application and a GID per each port
- GID (GARP Information Distribution)
  - a set of state machines that defines the current registration and declaration state of all attribute values
- GIP (GARP Information Propagation)
  - propagation of information between GARP participants
  - within a bridge
  - between bridges
    - based on LLC Type 1 service

GARP architecture



GID architecture



GMRP

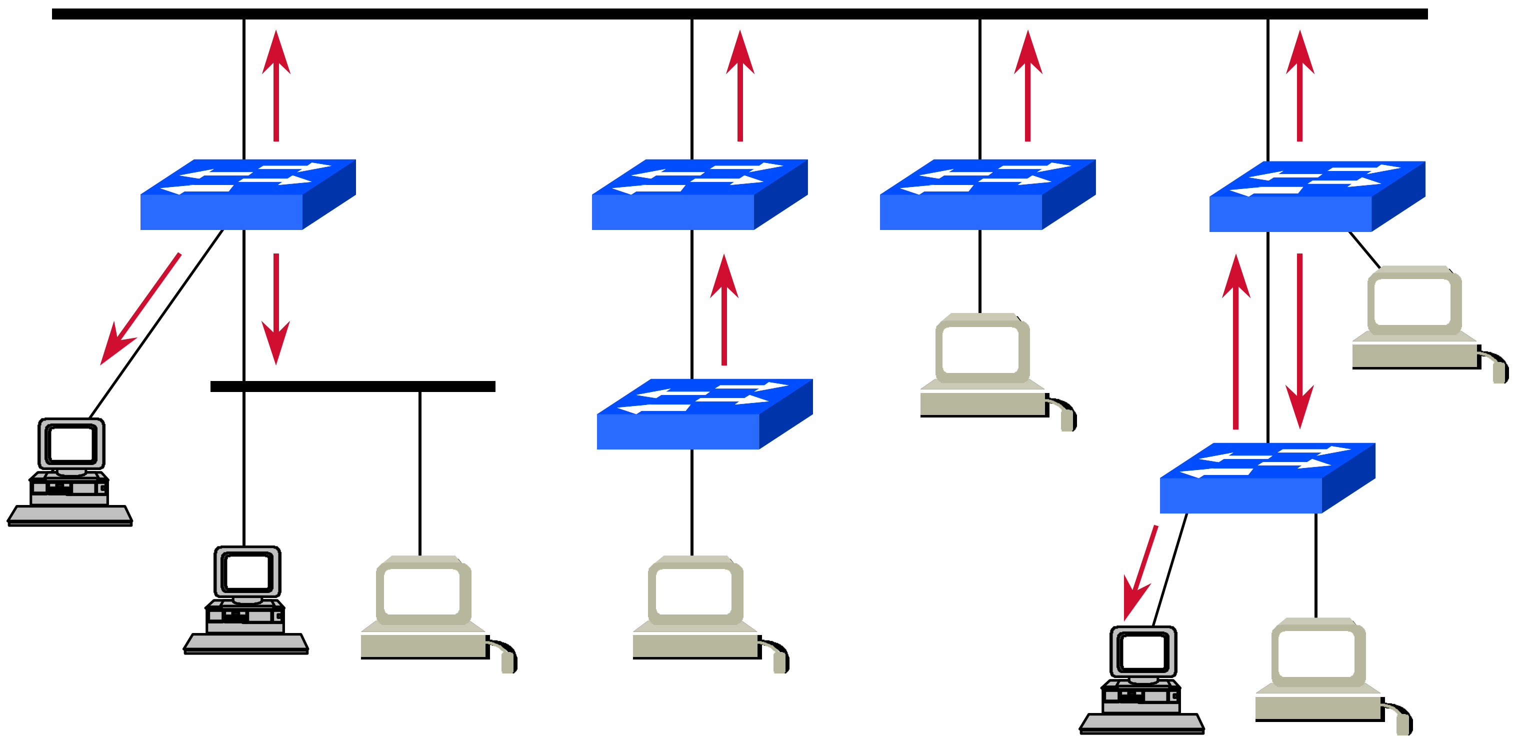
(GARP Multicast Registration Protocol)

- **Multicast group membership at MAC layer (MAC layer version of IGMP)**
  - **Allows endstations to register for the MAC multicasts that they want**
  - **Tracks which ports request each multicast address**
  - **Allows frame switches to send multicasts only where they're needed**
  - **Allows endstations to register for all MAC multicasts**

GMRP

- **GMRP attributes**
  - **Group Membership Information**
    - **Composition of the group**
    - **The attribute type is the 48-bit multicast MAC address**
    - **Updating Filtering database to indicate the ports on which members of the groups have been registered**
  - **Default Group behavior:**
    - **Filter Unregistered Groups (default)**
    - **Forward All Group**
    - **Forward Unregistered Group**

GMRP



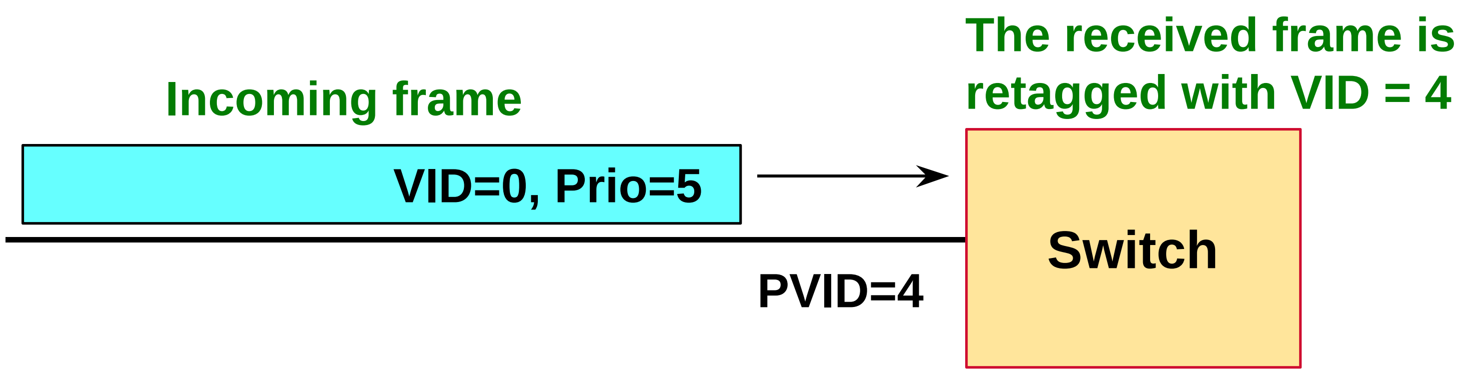
Application station belonging to Group M

Application station NOT belonging to Group M

Group Registration entry for Group M

Priority tagged frame

- The 1Q/1p tag can be used also in absence of VLANs
- Useful to carry priorities
- It is sufficient to set the VID = 0
  - The receiving port will retag the frame using the PVID (Port VLAN ID)



Compass

- VLAN
- IEEE 802 committees
- IEEE 802.1p
- IEEE 802.1Q
  - architectural model
  - GVRP
  - relaying function
  - port-based VLANs - native VLAN
  - spanning tree issues
  - interaction between 1Q and 1p
- The Cisco solution

IEEE 802.1Q

- **IEEE Standards for Local and Metropolitan Area Networks:**
  - **Virtual Bridged Local Area Network**
- **Draft 7**
- **Two main topics:**
  - **Bridged/switched networks**
  - **VLANs (Virtual LANs)**

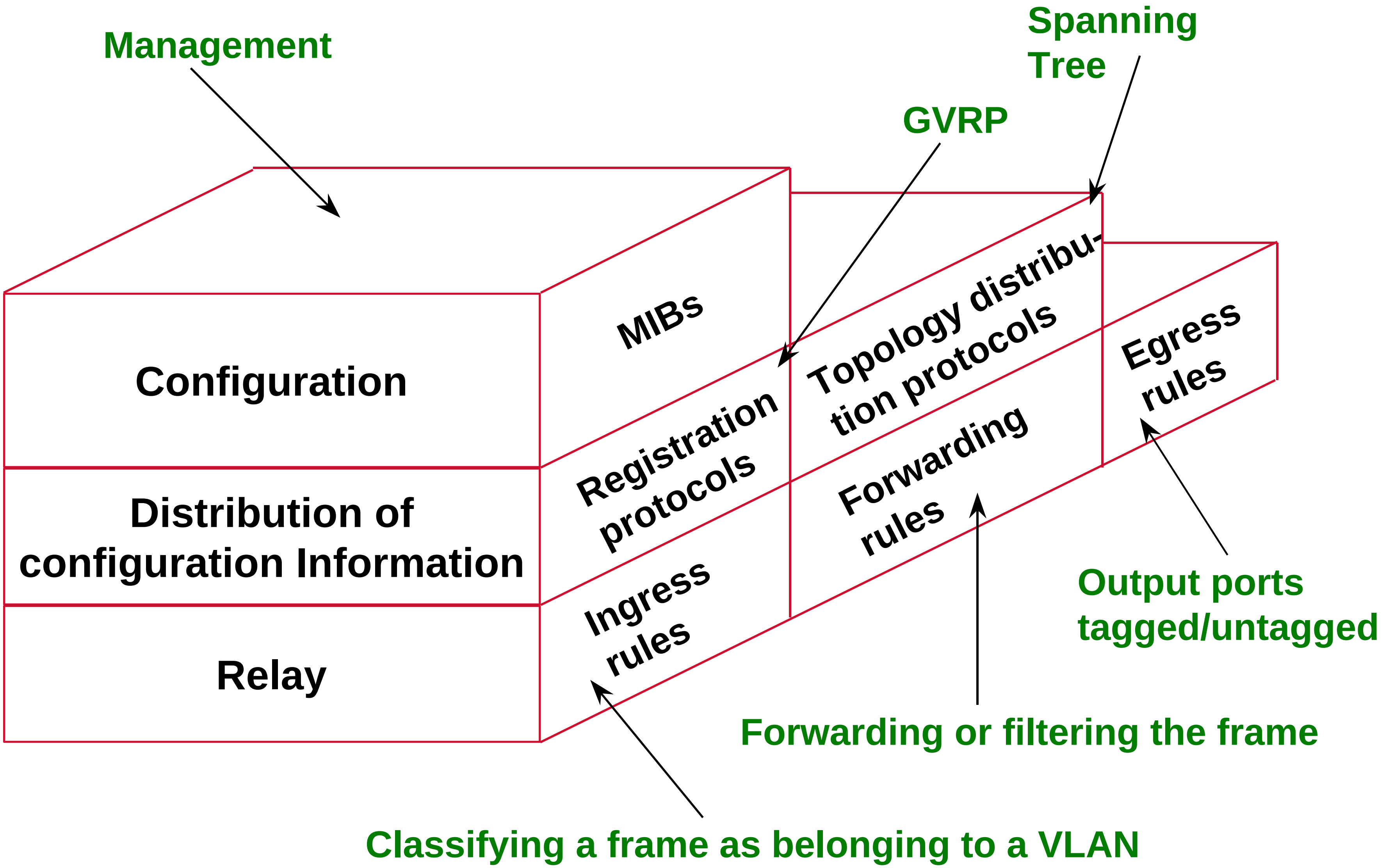
IEEE 802.1Q

- Defines the capabilities of a “VLAN-aware” bridge
- Adds to IEEE 802.1D the VLAN support
  - it interoperates with “VLAN-unaware” bridges
- Compared to proprietary solutions:
  - it is late
  - it has a limited set of features

Architectural choices

- **Per-port VLANs only**
  - **Assigning frames to VLANs by filtering, e.g. using L3 information, is allowed but not specified in the standard**
- **Single spanning tree**
- **Explicit tagging**
  - **one level tagging**
- **Supports both SFD/ME and MFD/SE**

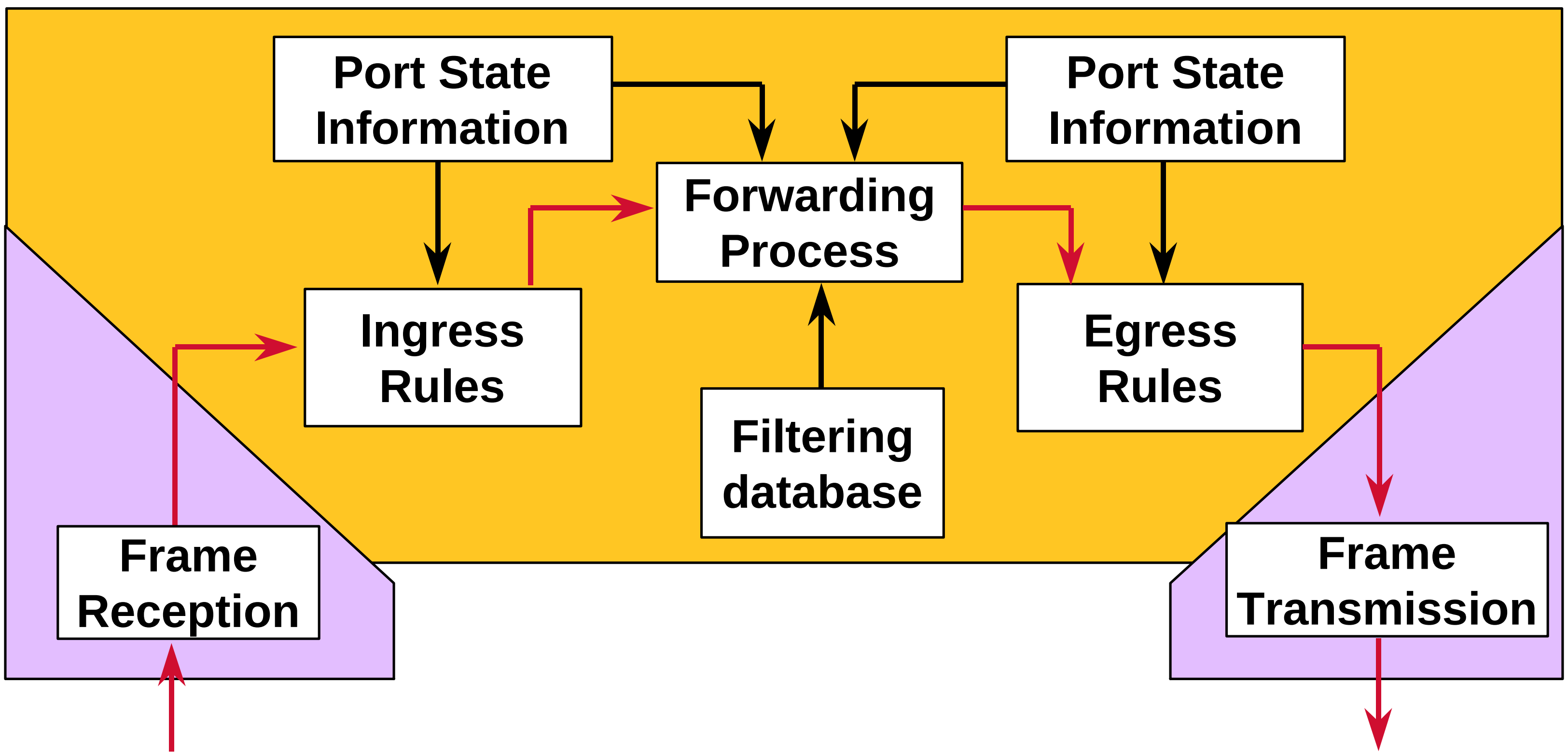
Architectural model



802.1Q VLANs

- **VLANs based on three-layer approach**
  - **Configuration: netadmin sets parameters**
  - **Distribution: switches agree on working details (to minimize configuration requirements)**
  - **Relay: frames are assigned to VLANs and distributed**
- **The distribution protocol is GVRP**
  - **it tells which switches want which VLANs**

Relaying function



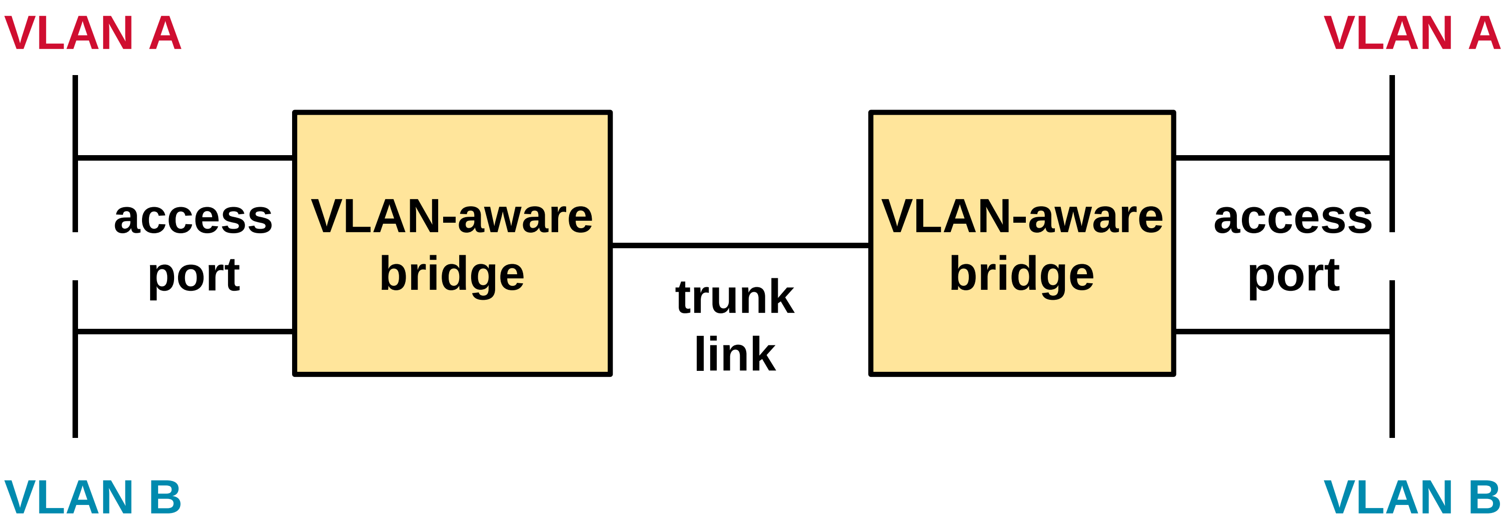
GVRP

(GARP VLAN Registration Protocol)

- **VLAN membership**
  - **End stations and bridges may issue or revoke declarations relating to the membership of VLANs**
  - **The attribute type is the 12 bits VID (VLAN ID)**
- **Service primitives**
  - **ES_REGISTER_VLAN_MEMBER(VID)**
  - **ES_DEREGISTER_VLAN_MEMBER(VID)**

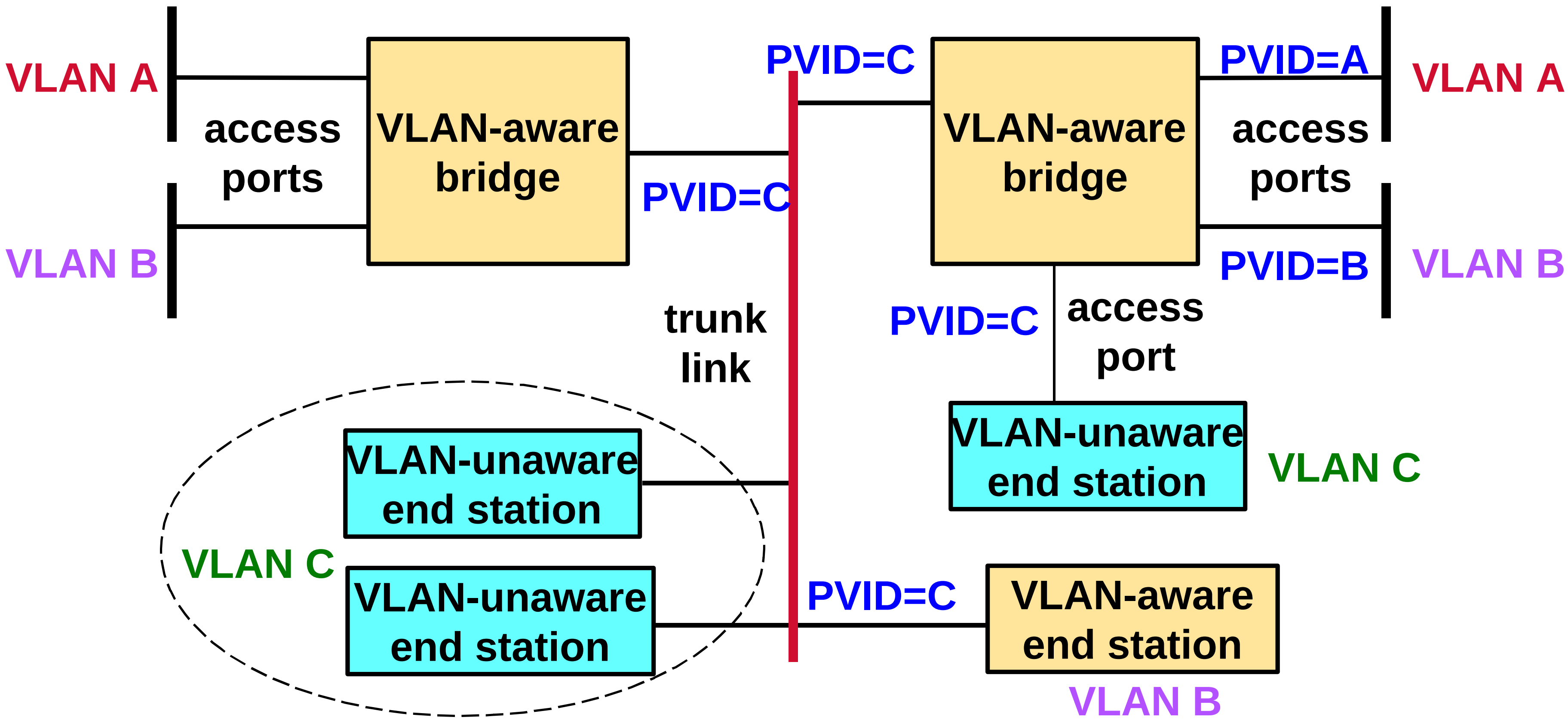
Port-based VLANs

- **Does not support user mobility nor decision based on higher level information**



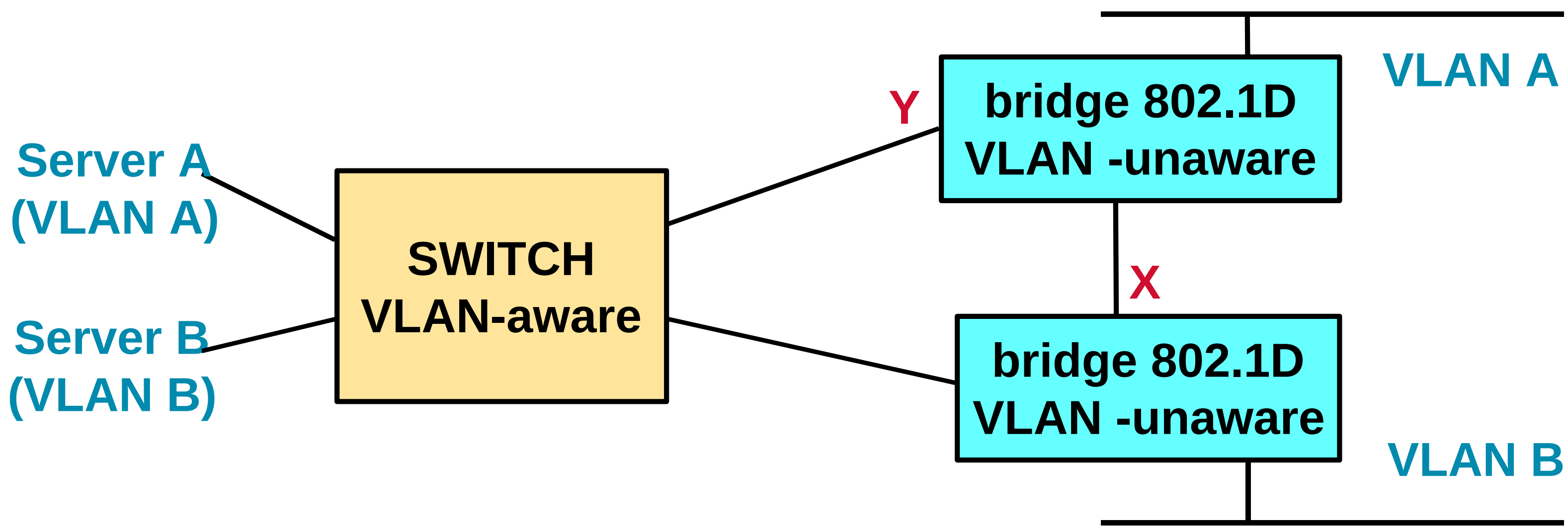
Native VLAN

- **Each physical port has a PVID (Port VLAN-ID) to which all untagged frames are assigned**



A spanning tree problem

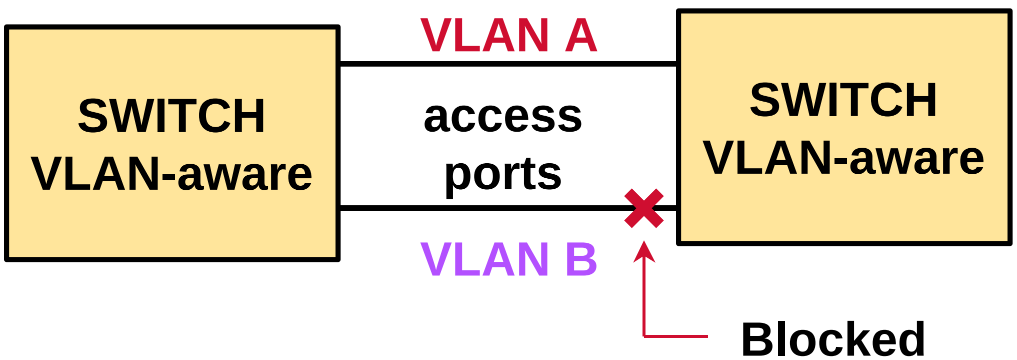
- **Mixing VLAN-aware and VLAN-unaware switches**



- **If the port X is blocked the network works fine**
- **If the port Y blocks VLAN A is partitioned (the server A is unreachable from VLAN A)**

A worse spanning tree problem

- The single spanning tree will block one of the two VLANs
- This situation is normal and unavoidable in the case of two VLANs implemented with ATM/LANE



Interaction between 1Q and 1p

- **1p introduces GARP**
  - **1p specifies how to forward frames in a multicast environment using GMRP**
- **1Q introduces GVRP and specifies how to forward frames in a VLAN environment**
  - **on which ports**
  - **on which VLANs**
  - **tagged or native format**
- **1p specifies how to encode the frame priority**