INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

IEC 62439-3 Clause 5:

HSR – High-availability Seamless Redundancy

Zero-delay recovery fault-tolerance in Industrial Ethernet in a ring or meshed topology

Prof. Dr. Hubert Kirrmann, Solutil, Switzerland

2017 February 8

Abstract

HSR (High-availability Seamless Redundancy

is a Layer 2 Ethernet (IEEE 802.3) redundancy protocol

- provides zero switchover time in case of failure
- allows to chain nodes for cost effective networking
- allows complex topologies such as rings and rings of rings
- is easily implemented in hardware
- supports deterministic transmission of highest priority frames
- is standardized as IEC 62439-3 Clause 5

Objectives

This standardization addresses the dependability and real-time requirements of demanding applications such as substation automation and motion control.

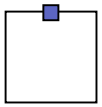
The technical solutions have been developed in

IEC SC65C WG15
(highly available automation networks), resulting in IEC 62439-3, and in

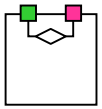
IEC TC57 WG10
(substation automation), for introduction into IEC 61850.

- Cost effective redundancy with no single point of failure and zero recovery time
- Fulfill the dependability and real-time requirements of the most demanding applications such as substation automation and motion control
- Protocol-independent, applicable to most industrial Ethernet
- Applicable to a variety of topologies, principally rings and rings of rings
- Do not require switches

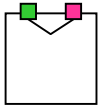
HSR Topologies : conventions



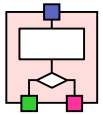
SAN singly attached node (not HSR)



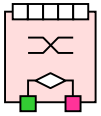
DANH node with 2 HSR ports



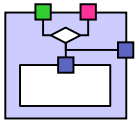
DANP node with 2 PRP ports



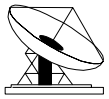
redbox with one single port



redbox switch (RSTP) to HSR

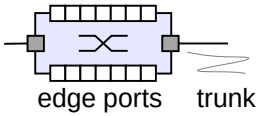


HSR node with auxiliary port



GPS time server

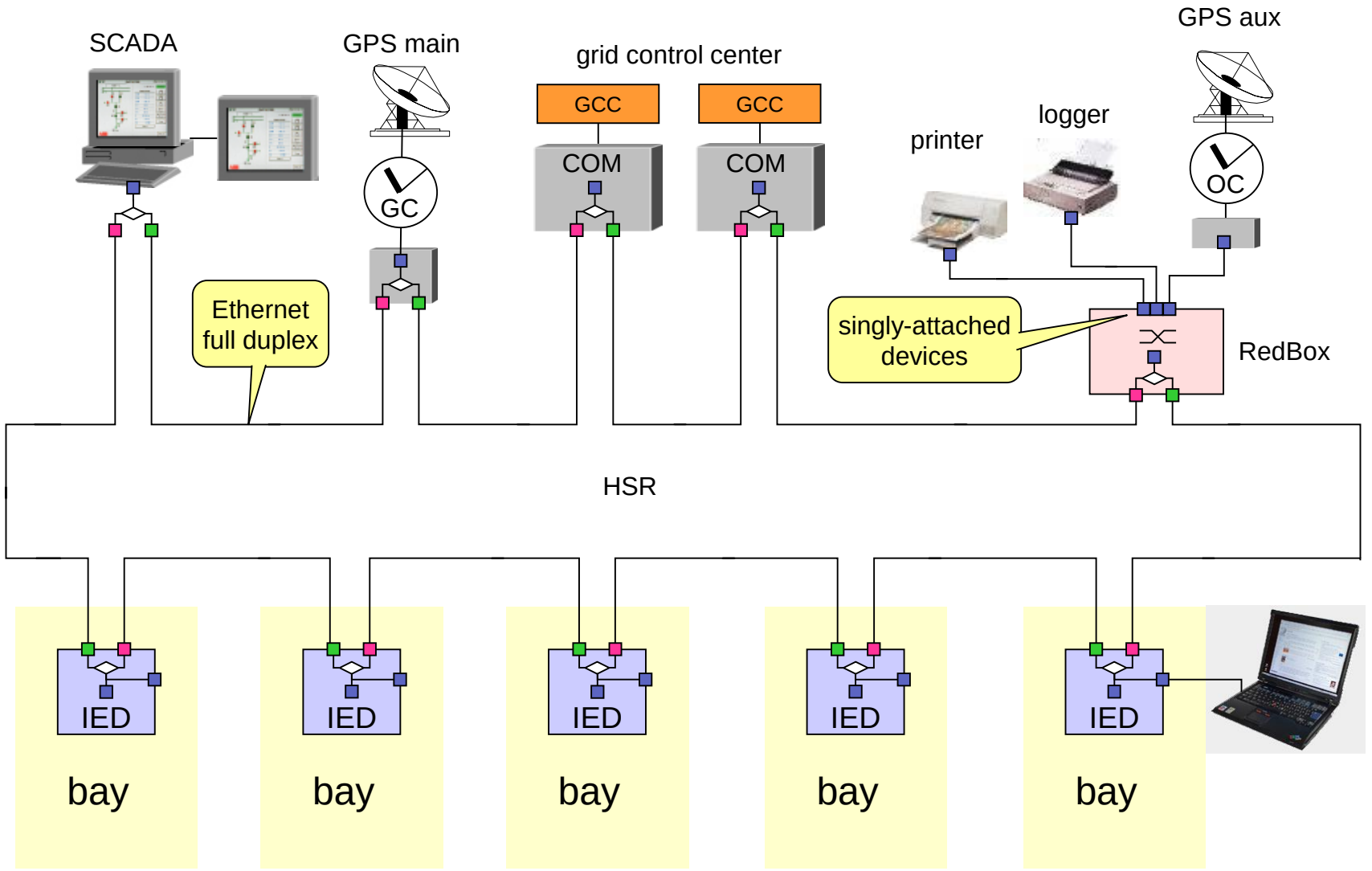
clock
GC = grandmaster clock
TC = transparent clock
BC = boundary clock
OC = ordinary clock
NC = network clock



layer 2 bridge

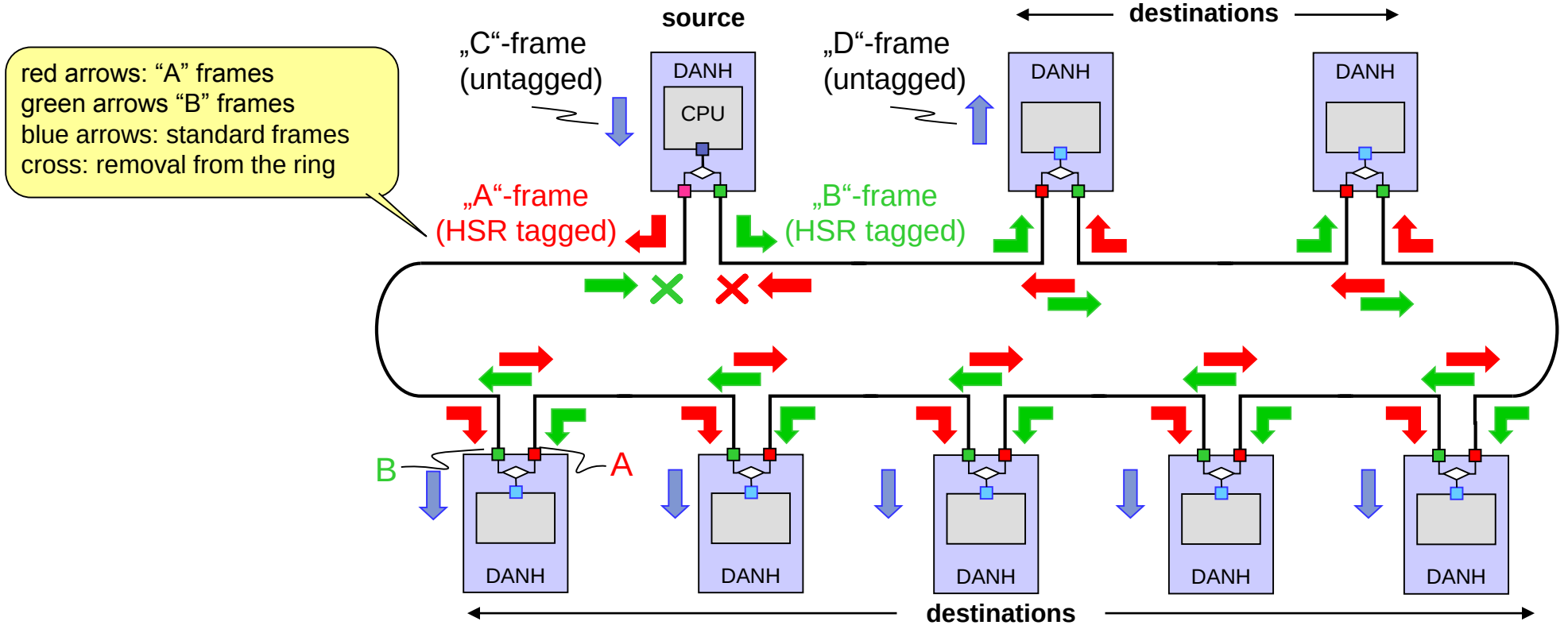
ports
- □ 100 Mbit/s Tx
- ○ 100 Mbit/s Fx
- ⊠ 1Gbit/s Tx
- ⊗ 1 Gbit/s Fx

(Substation) Automation network ring (fibre or copper full duplex) with “switching nodes”



Cost-effective: all nodes are “switching nodes”, there are no dedicated switches in the ring
Non-ring nodes are attached through a “RedBox”.

HSR principle (Multicast)



Nodes are arranged as a ring, each node has two identical interfaces, port A and port B.

For each frame to send (“C”-frame), the source node sends two copies over port A and B.

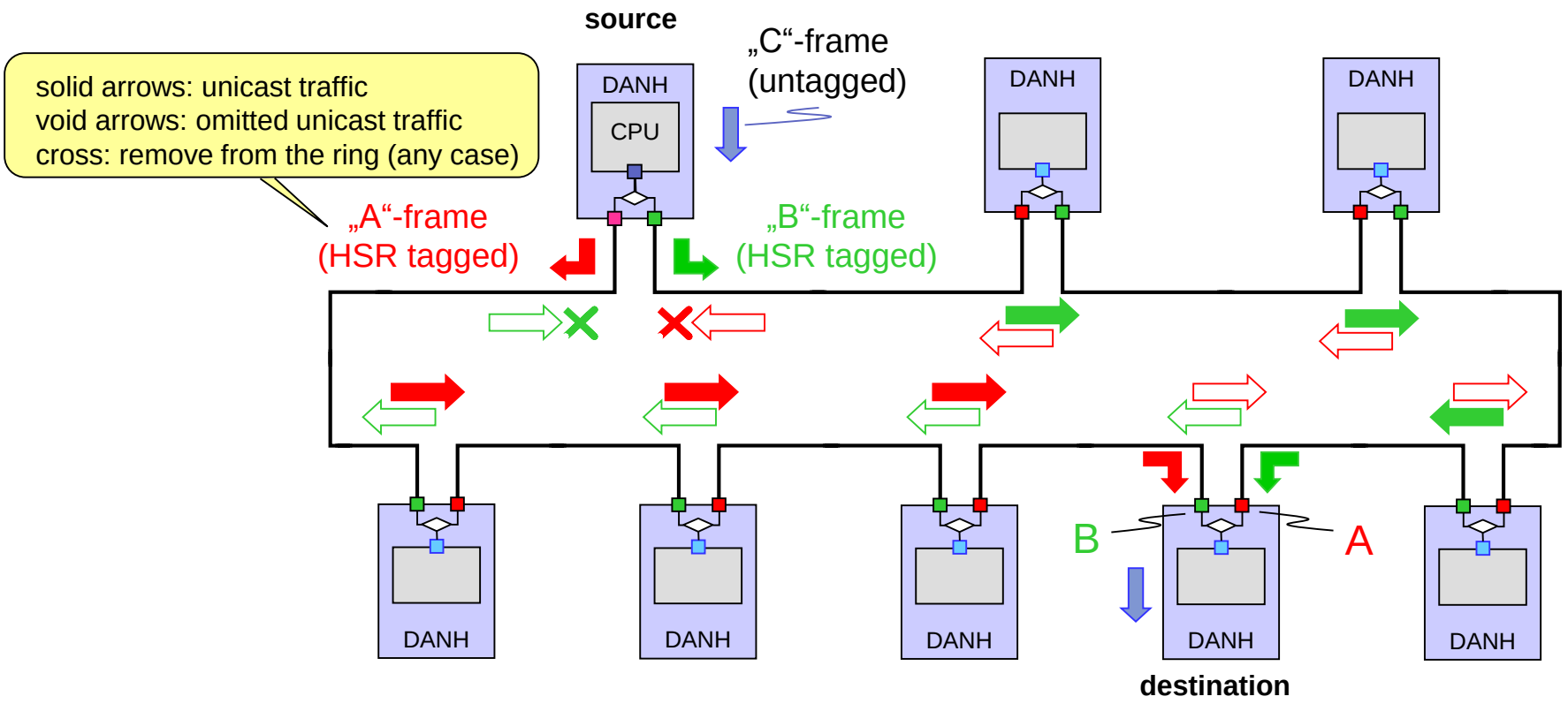
The source node removes the frames it injected into the ring.

Each node relays a frame it receives from port A to port B and vice-versa, except if already forwarded.

The destination nodes consumes the first frame of a pair (“D-frame”) and discards the duplicate.

If the ring is broken, frames still arrive over the intact path, with no impact on the application.
Loss of a path is easily detected since duplicates cease to come.

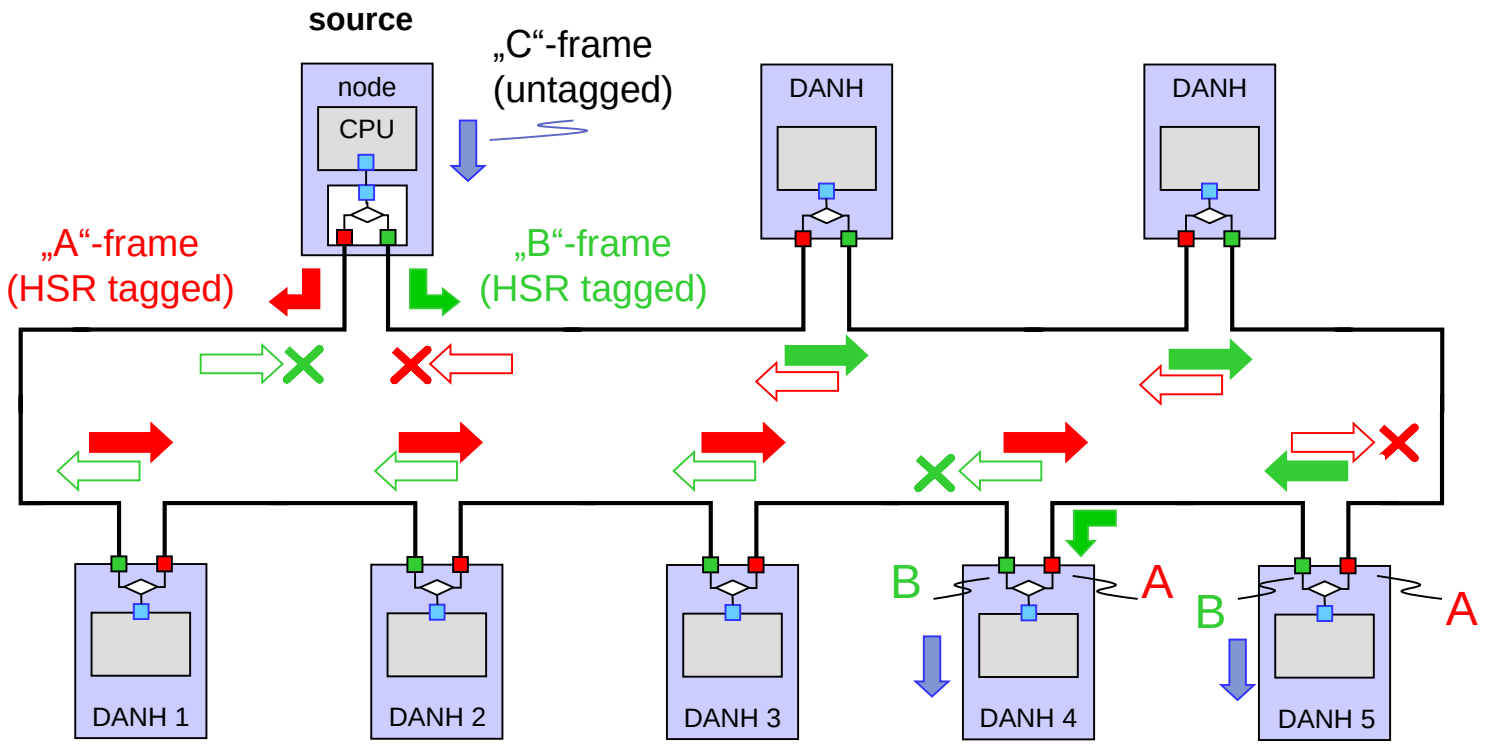
HSR principle (mode U, unicast frames removal)



To reduce unicast traffic, a node refrains from sending a frame addressed uniquely to itself (unicast). Should this rejection fail, the frame would be discarded by the other nodes.

Mode U may be disabled for testing purpose (network monitoring) or redundant nodes set-up.

HSR principle (mode X, multicast frame removal)



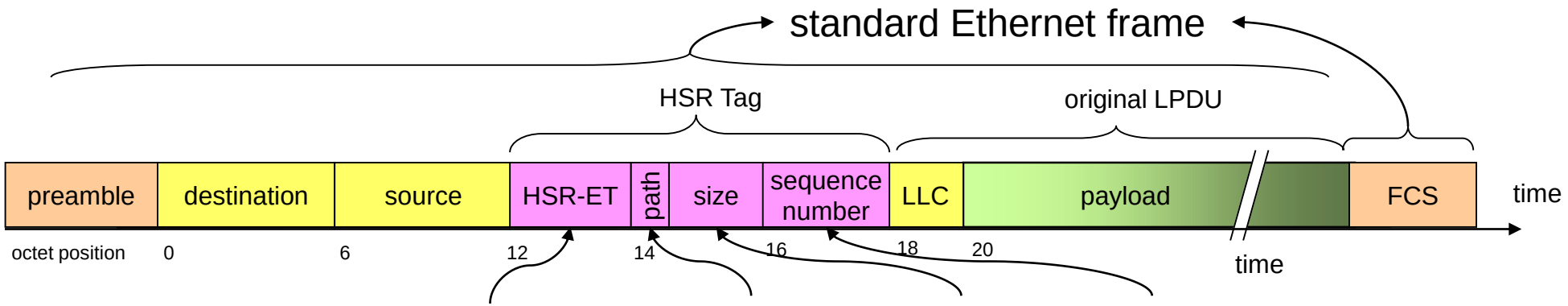
To reduce multicast traffic, a port refrains from sending a frame if it already received a duplicate of that frame from the opposite direction.

E.g. Port B in DANH 4 does not forward the “B”-frame since it received previously the “A”-frame.
Port A in DANH 5 does not forward the “A”-frame since it received previously the “B”-frame.

Would this mechanism fails, the frames would still be removed by the other nodes or by the node that originally injected the frame.

Mode X is not applicable to PTP messages and to supervision messages.
It may be disabled for testing purpose (network monitoring) or for redundant node operation.

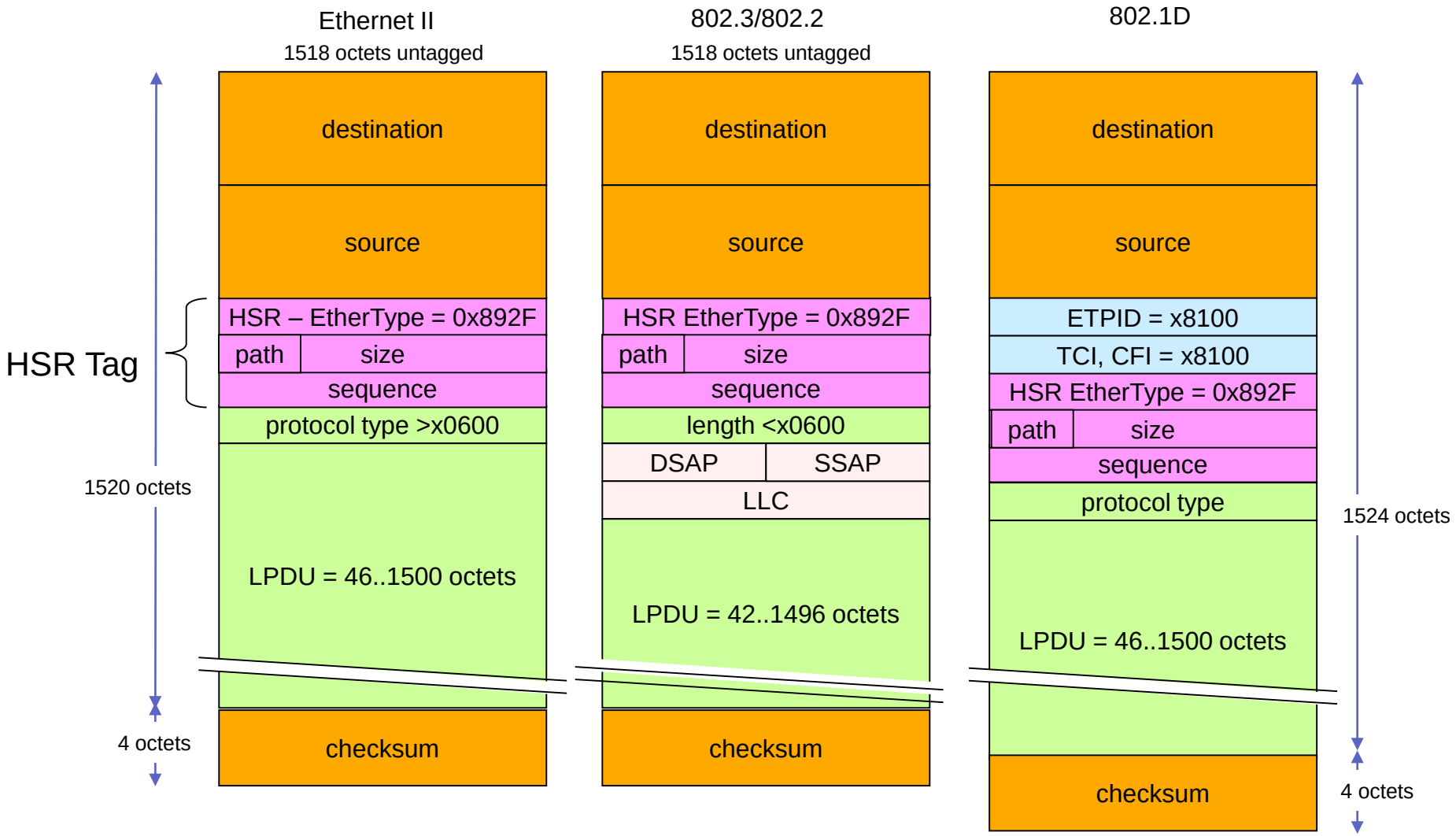
HSR Frame identification for duplicate rejection



- each frame has an HSR Ethertype, a path indicator, a size field and a sequence number, inserted as an HSR tag in the same way a VLAN tag is inserted.
- the sender inserts the same sequence number in both frames of a pair, and increments the sequence counter by one for each sending from this node.
- the receiver keeps track of the sequence counter for each source MAC address it receives frames from. Frames with the same source and sequence number value coming from different lines are discarded.

  to supervise the network, a node may keep a table of all other nodes in the network from which it receives frames. This allows to detect nodes absence and bus errors at the same time.
- a node recognize the frame it sent through its source address and sequence number

HSR Frames types: tag position



The additional six bytes of the HSR tag could generate oversize frames of more than 1522 octets. However, this is private ring traffic and does not affect Ethernet controllers.

Duplicate recognition

Each node increments the sequence number field monotonically for each frame sent.

A duplicate frame is recognized in a receiver or forwarding node by its:

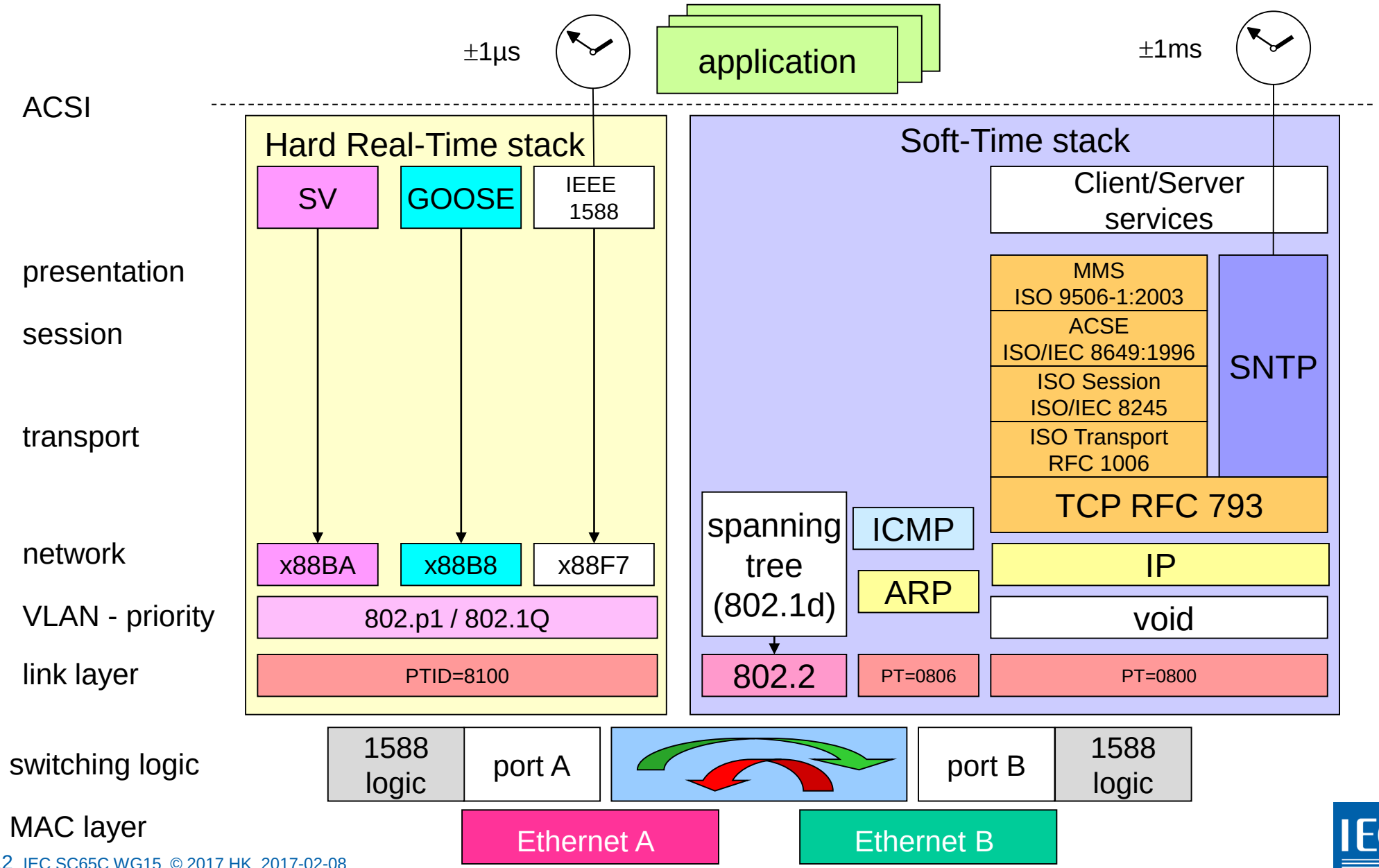
- source address
- sequence number in the HSR tag.

HSR nodes shall never reject a frame that they did not receive before and shall detect nearly all duplicates, but infrequent duplicates do not disturb.

The duplicate detection algorithm is not specified. Hash tables, queues and tracking of sequence numbers are possible methods.

PRP (IEC 62439-3) only considered discard of duplicates on a “best effort” basis. HSR has an improved coverage.

Layering in IEC 61850: HSR is independent from stack



Addresses and layer 2 redundancy

Each node has the same MAC address on both ports.

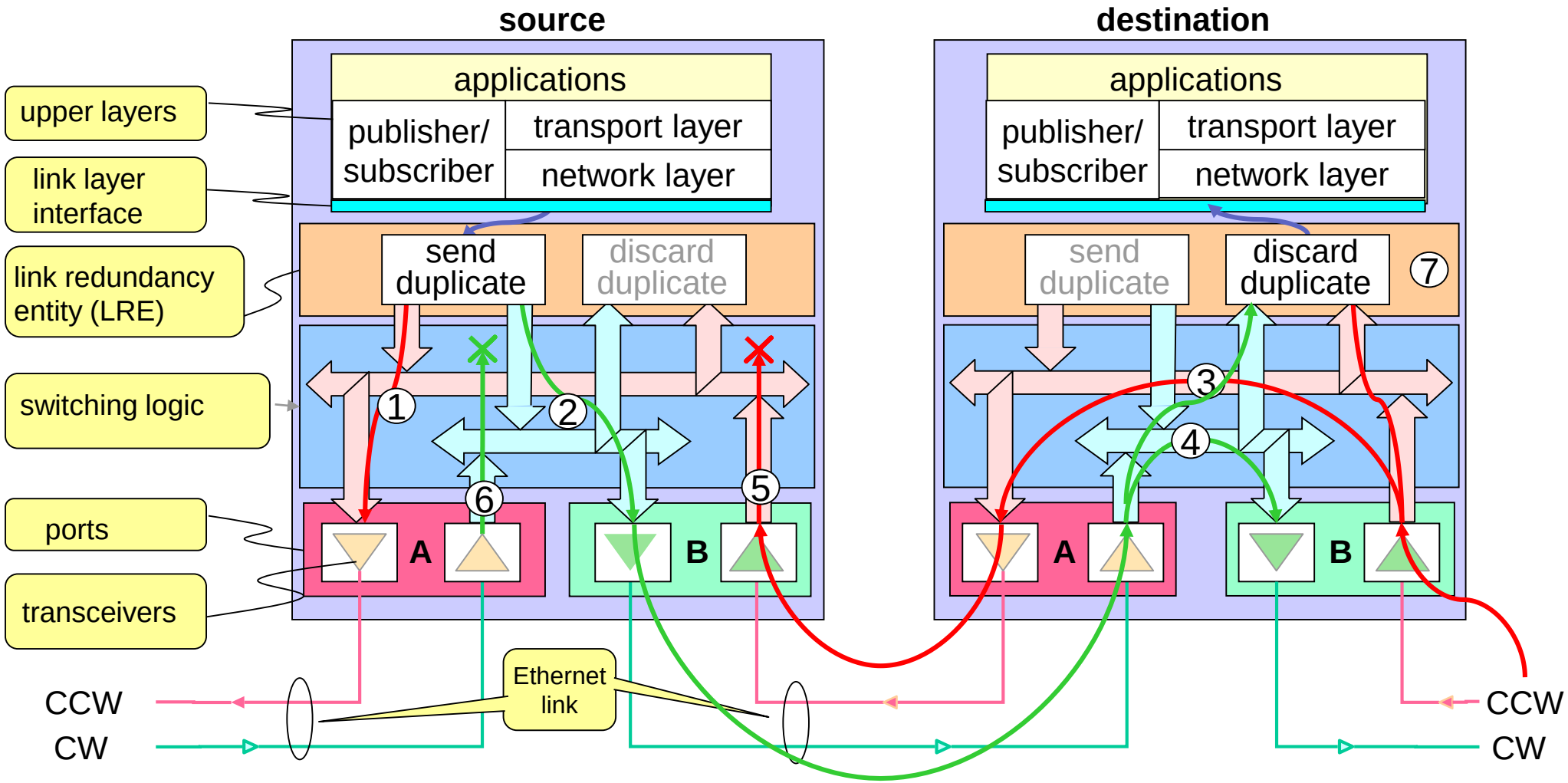
Each node operates with the same IP address(es)* for both ports.

Therefore, management protocols such as ARP operate as usual and assign that MAC address to the IP address(es) of that node.

TCP/IP traffic is not aware of the Layer 2 redundancy, it is required to treat duplicates.

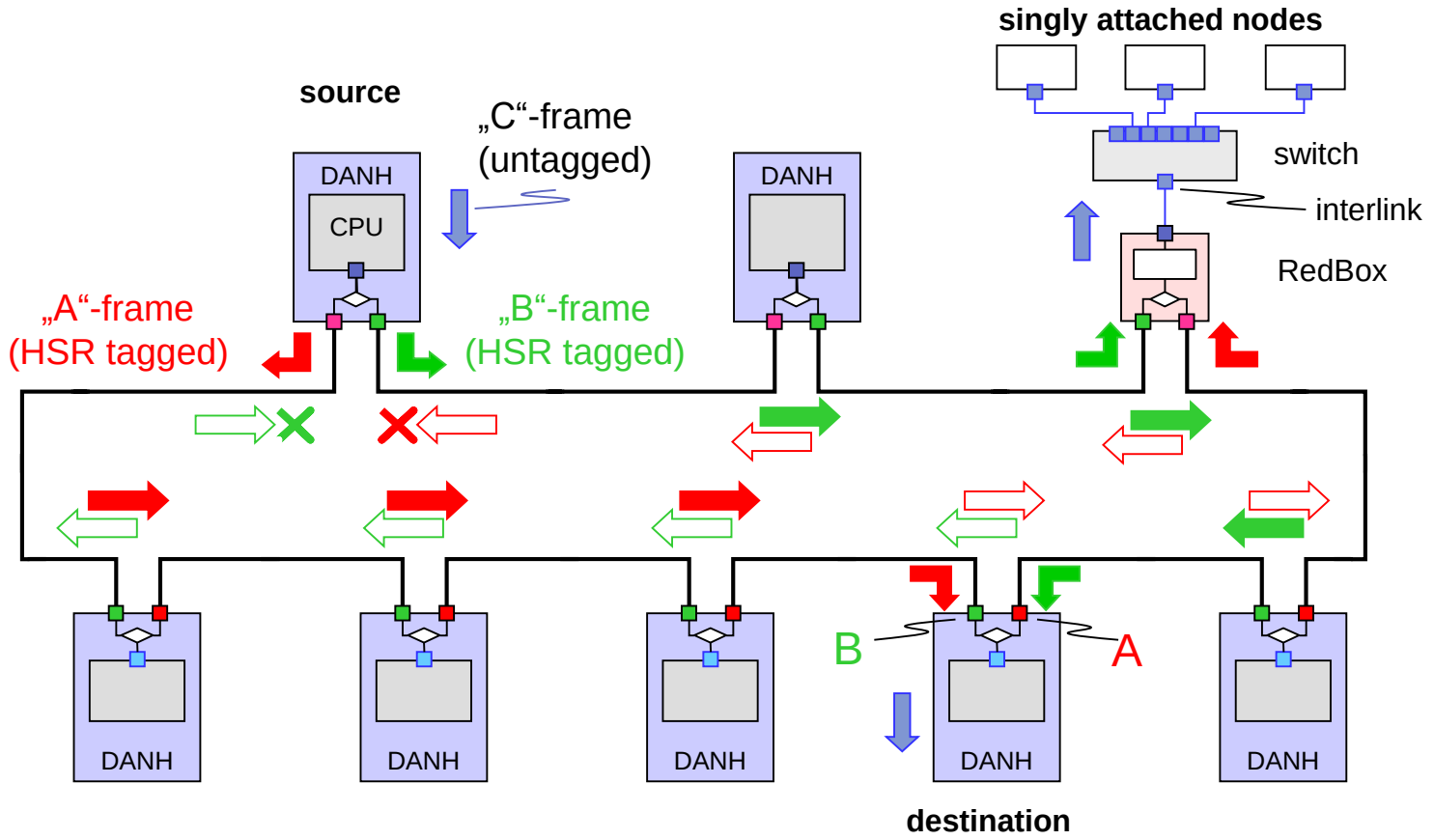
* a node may respond to several IP addresses

HSR Node Operation



send: the LRE sends each frame to send simultaneously over port A and port B (1), (2).
forward: the switching logic resend frames from one port over the other port (3),(4) except own frames (5),(6)
receive: the LRE receives both frames, keeps the first frame and discards the duplicate (7).

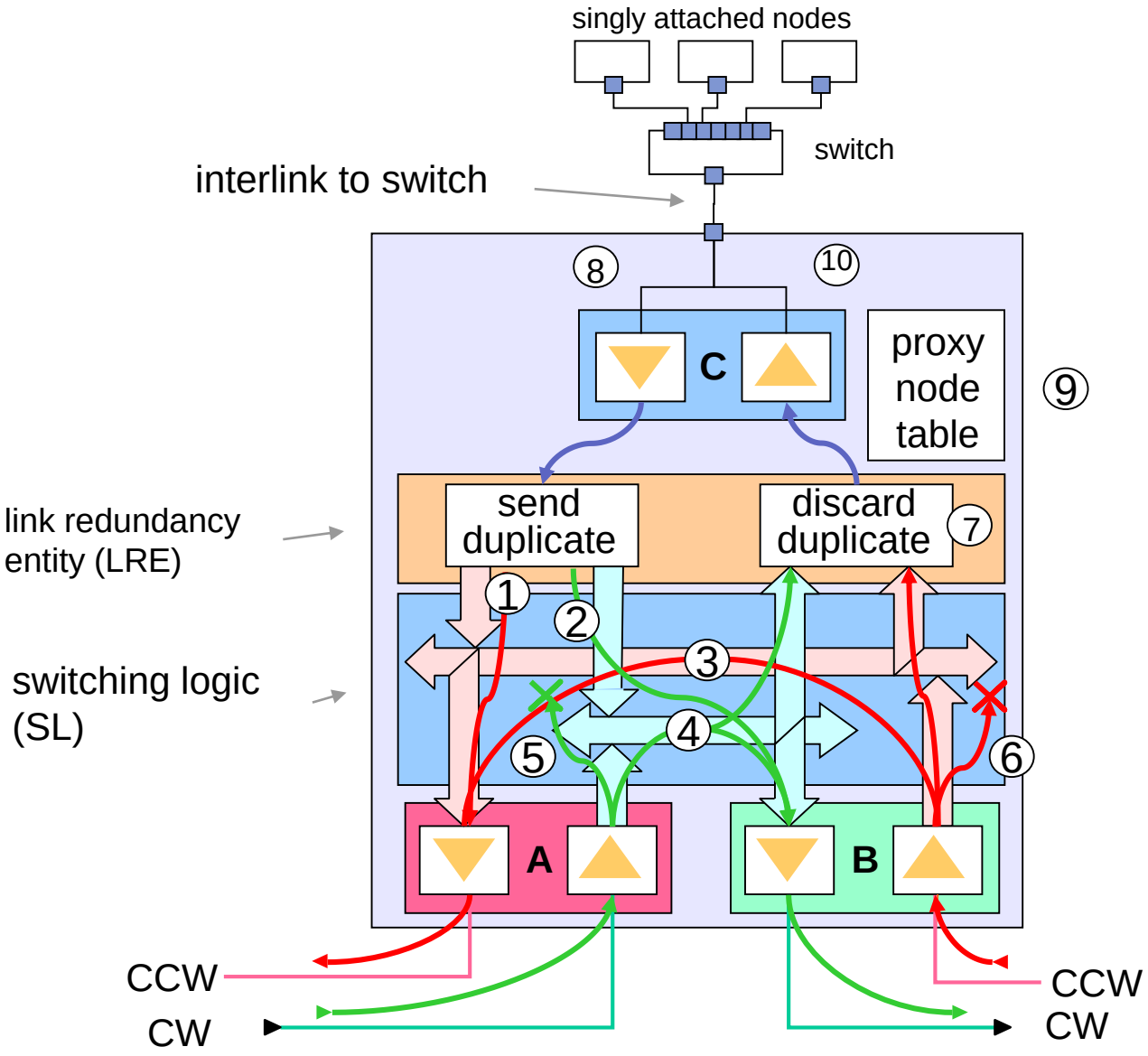
Attachment of legacy devices: RedBox



Legacy nodes such as laptops or printers do not recognize the HSR tag and must be attached through a RedBox (Redundancy Box) which acts as their proxy.

The RedBox generates the same management frames as if its represented nodes would be inserted directly in the ring, and removes the frames it injected into the ring when they come back

Redundancy Box operation (RedBox H)

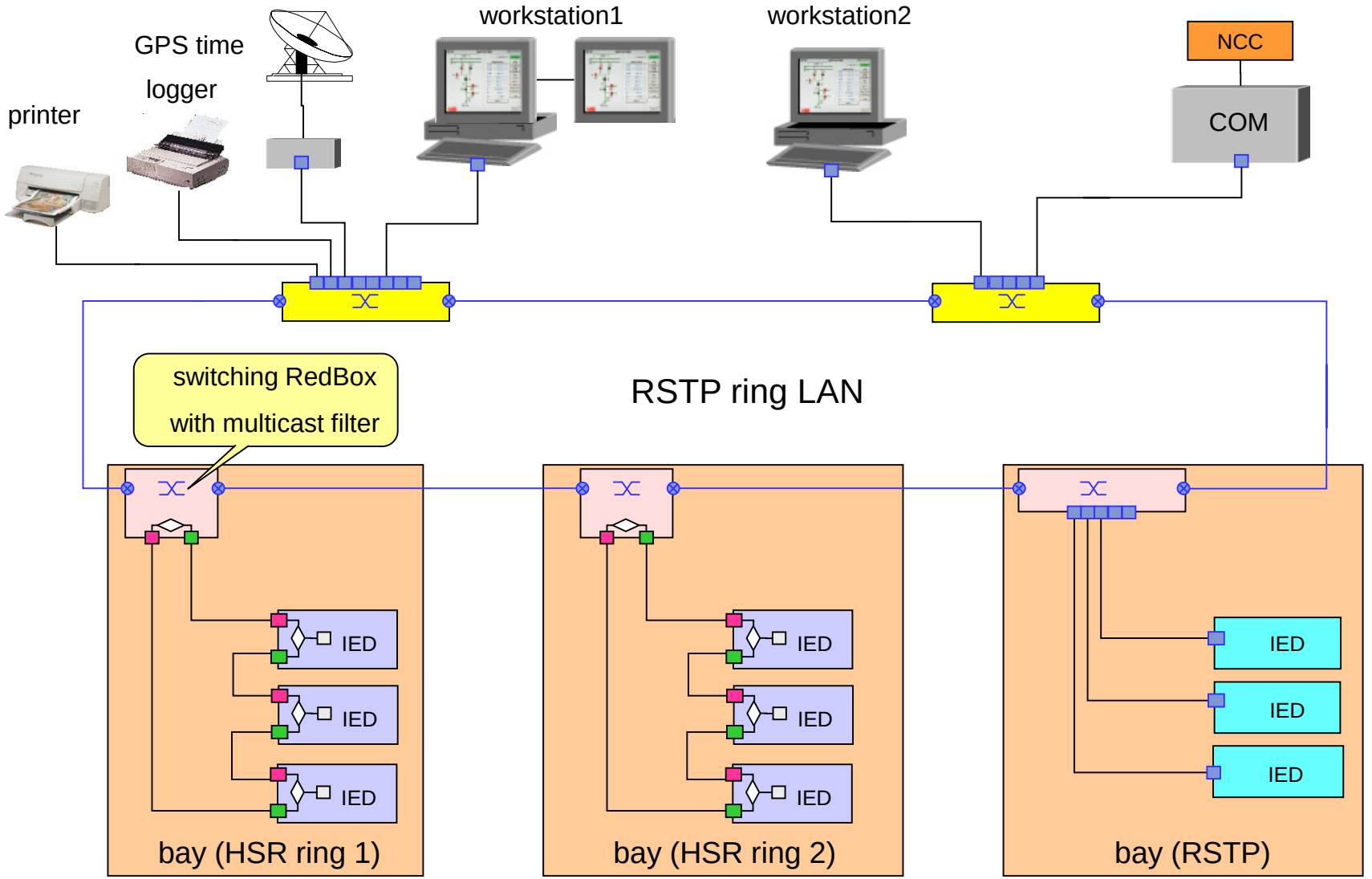


The RedBox H operates as a proxy for a number of singly attached nodes.

To remove the frames it send from the ring, the RedBox keeps a table of nodes for which it is the proxy, e.g. by listening to the received frames (8).
It can ping the SANs to clean up the list of removed or inoperative nodes, or remove the entries after a time-out (e.g. 1 minute).

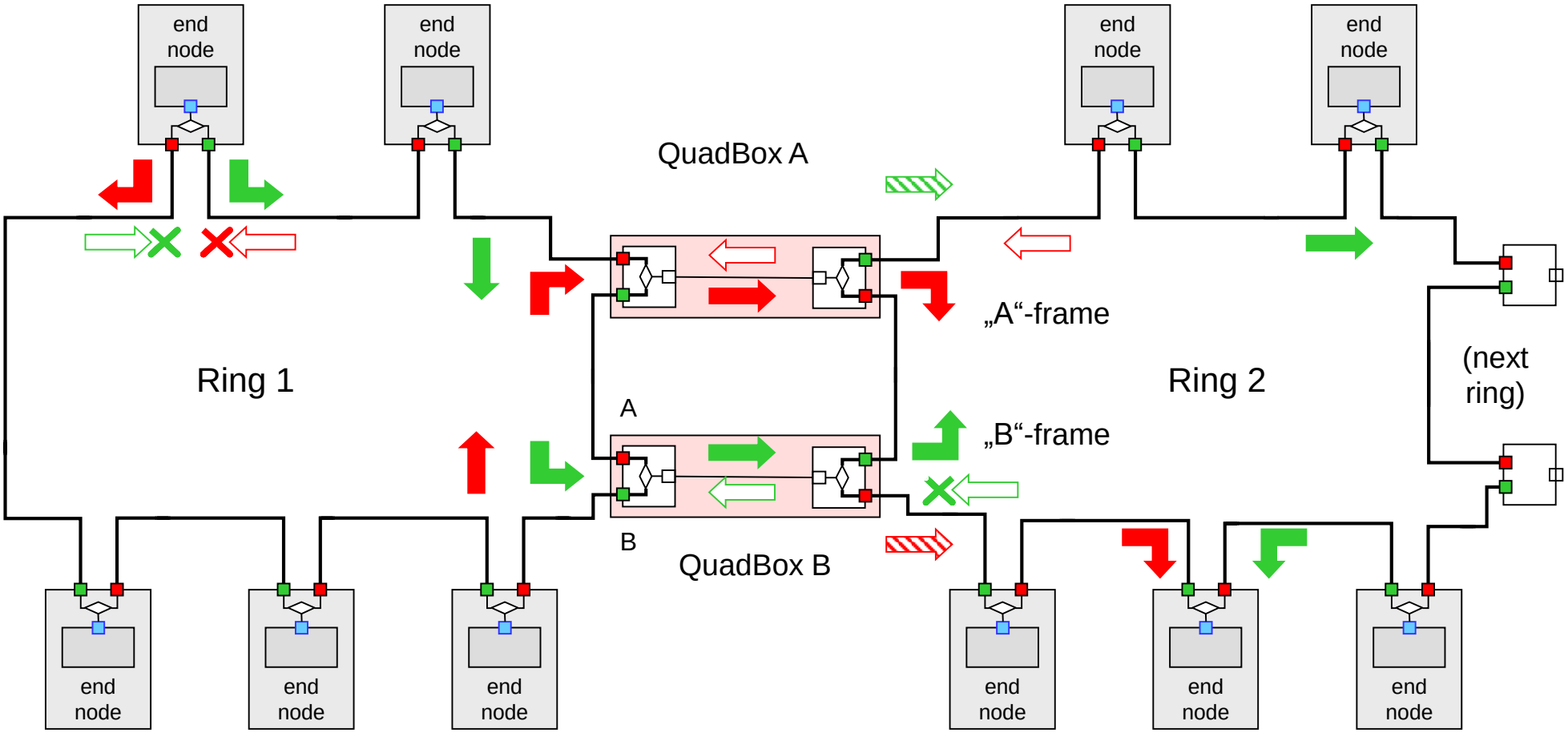
The RedBox behaves as a bridge for non-HSR traffic, the protocol is defined in the PICS.

Non-redundant topology: 2-level (RSTP – HSR) hierarchy



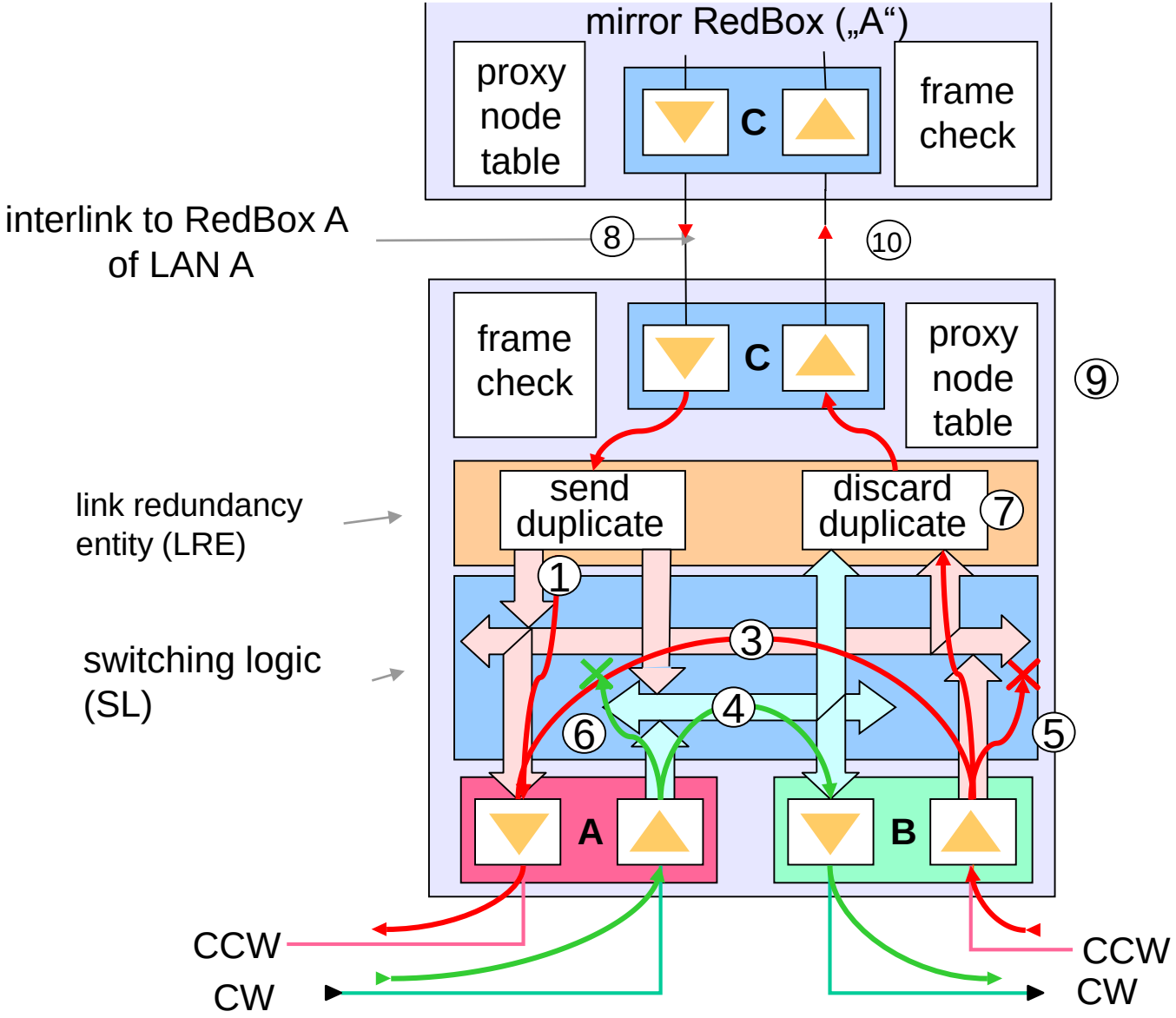
Mixing non-redundant ring and HSR rings (partial redundancy)

Coupling two HSR rings with a QuadBox

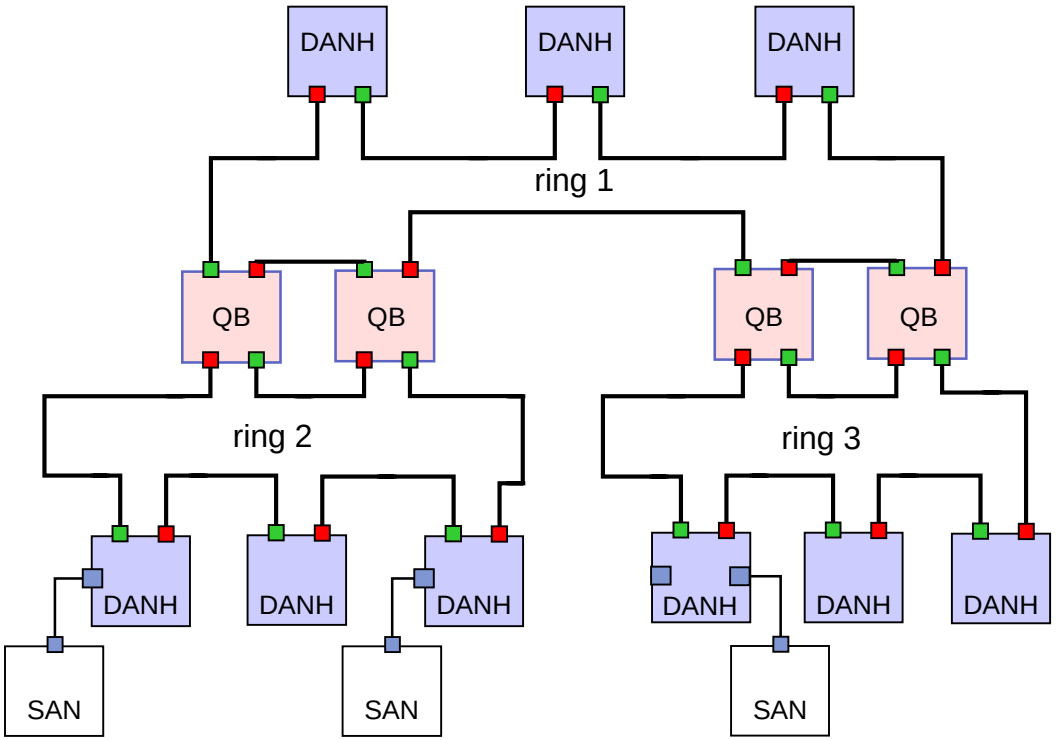


Two quadboxes are needed to avoid a single point of failure

Quadbox = 2 x RedBox (in principle)



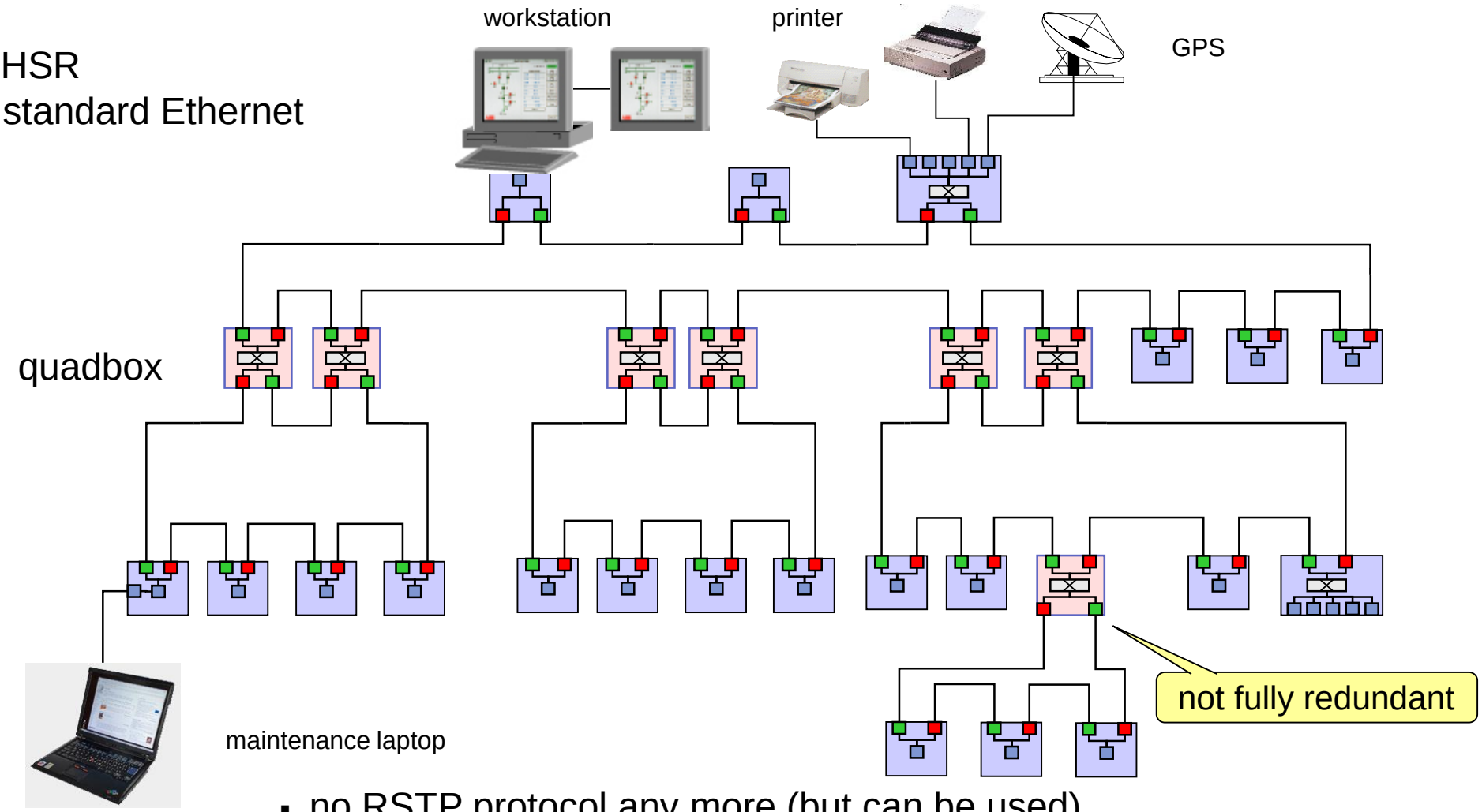
Topology with full coverage: ring of rings



- Needs two quadboxes for failure-independence
- Makes only sense if VLAN or Multicast filtering is used

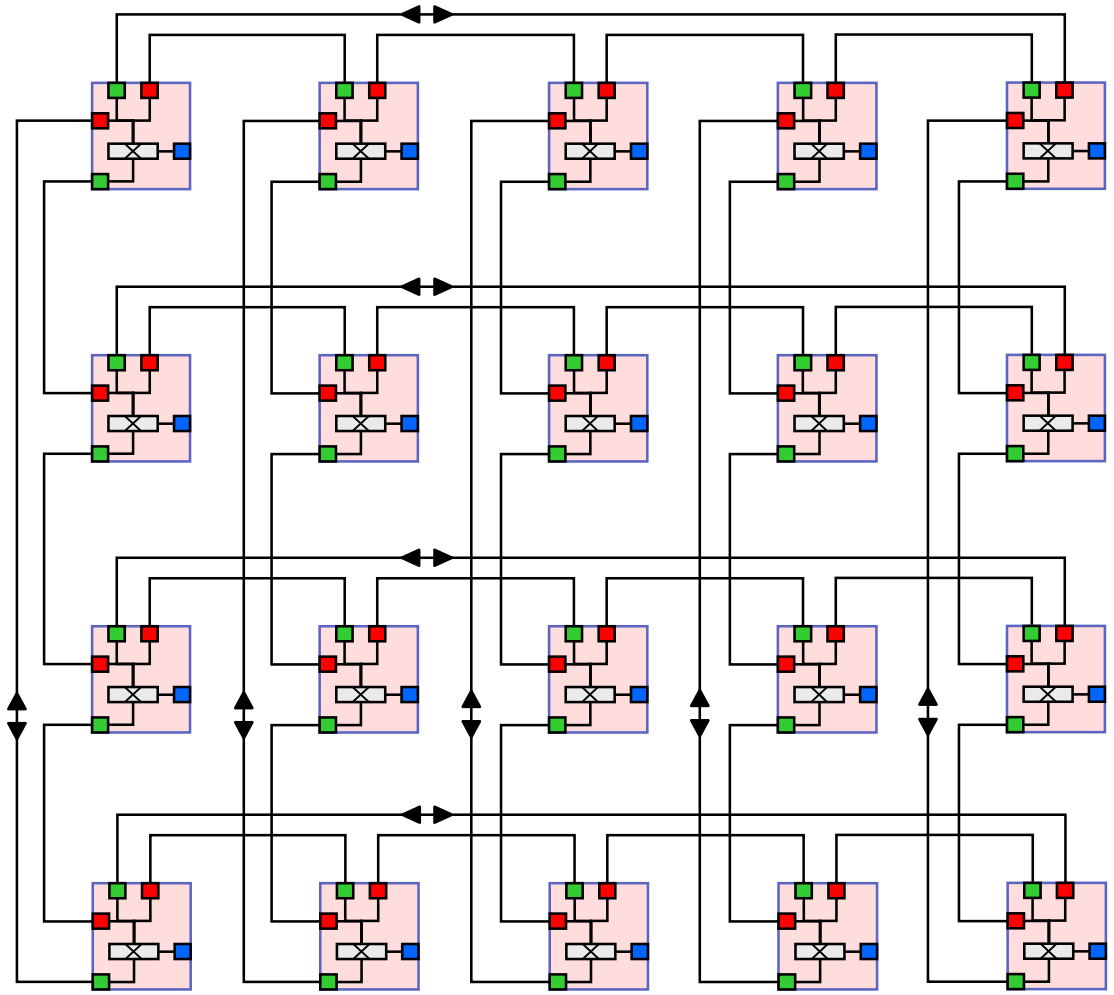
Generalizing the topology: three levels

- HSR
- standard Ethernet



- no RSTP protocol any more (but can be used)
- note that level 3 is singly attached (only one quadbox)

Meshed topology “transputer”



- any meshing allowed

HSR COUPLING TO OTHER NETWORKS

Compatibility PRP - HSR

PRP is a redundancy protocol operating on the same principles as HSR, but without requiring special hardware.

It is standardized as IEC 62439-3 Clause 4

PRP compatibility

A node can operate in HSR mode or PRP mode with the same hardware.

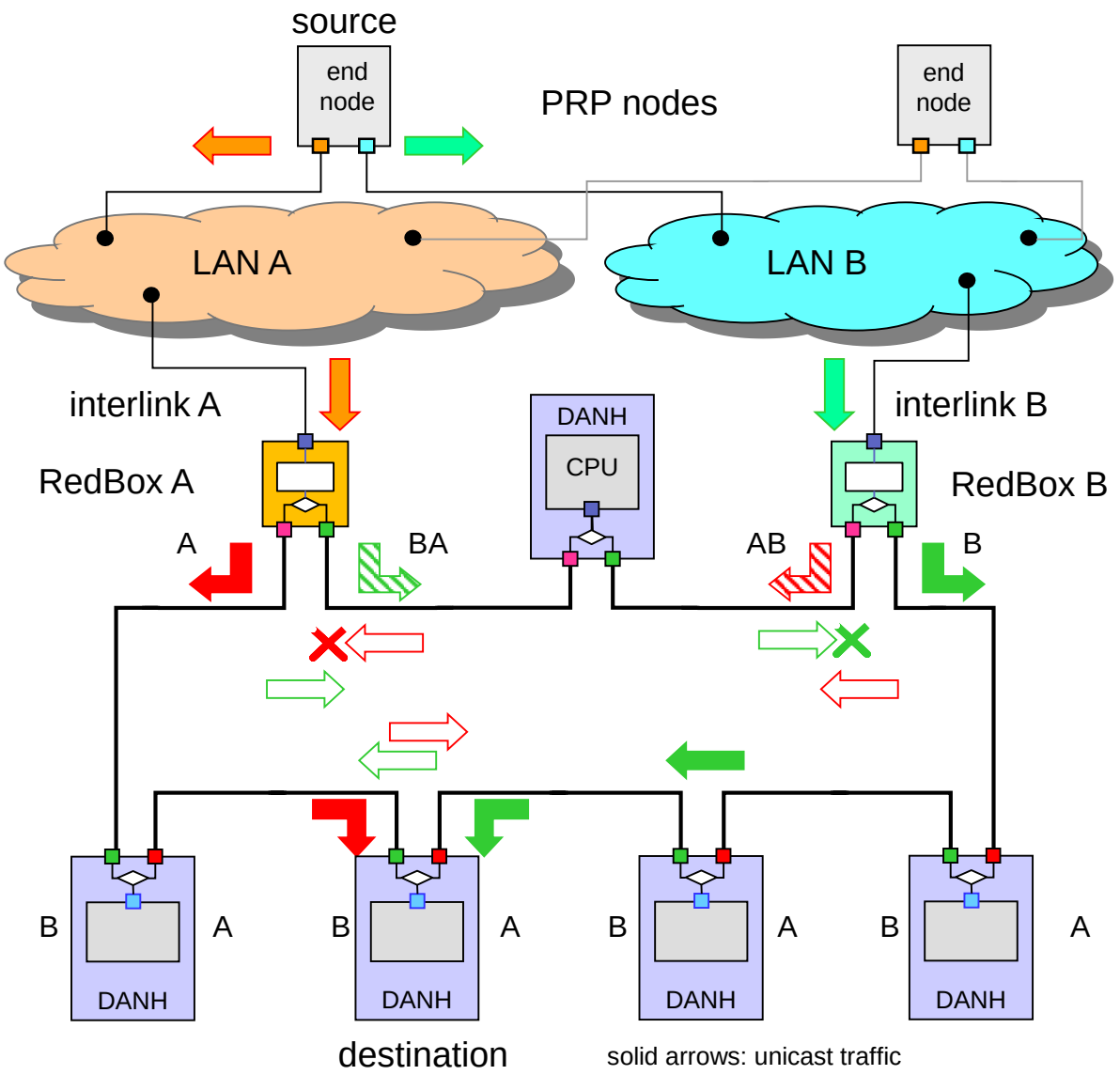
Therefore, the basic infrastructure of PRP can be used.

But forwarding frames requires hardware that is currently not needed in PRP.

The frame format is different.

Since HSR frames have the same size as PRP frames, segmentation is avoided (the HSR Tag remains in the ring and does not arrive to the Ethernet controller).

Coupling HSR and two PRP LANs (sender in PRP)



solid arrows: unicast traffic
void arrows: multicast or not received unicast traffic
patterned arrows: duplicate from other RedBox

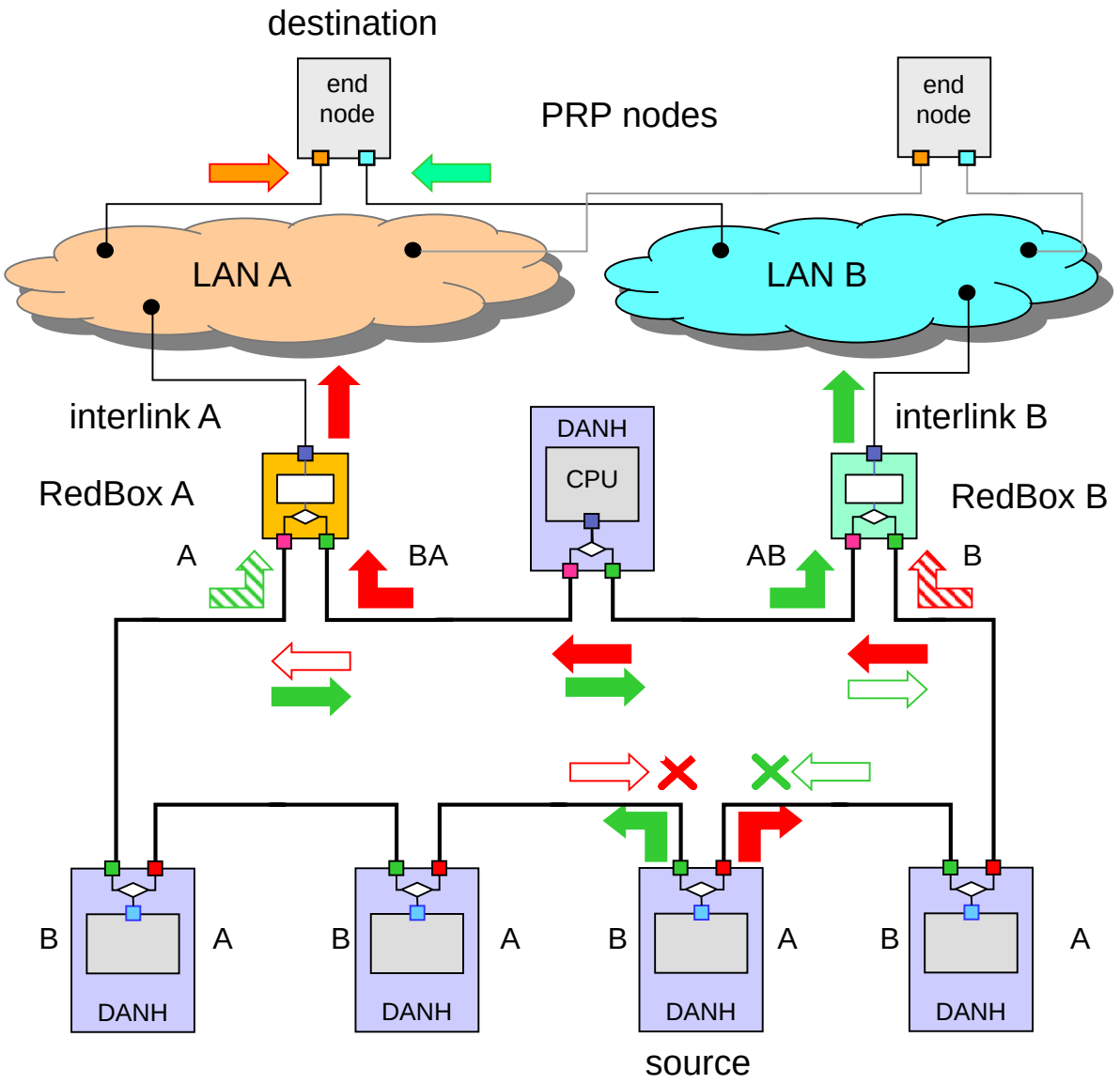
The Red Box receives frames from their interlink and store their source address in the Proxy Node Table.

The RedBox sends such frames in both directions on the ring, tagged as “A” and “B”, except if it already forwarded the same frame in that direction (since there are two red boxes, this depends on the order of sending)

A RedBox forwards frames received by one port to the other, except if it already sent it.
To raise throughput, a node may not forward a unicast frame directed to it.

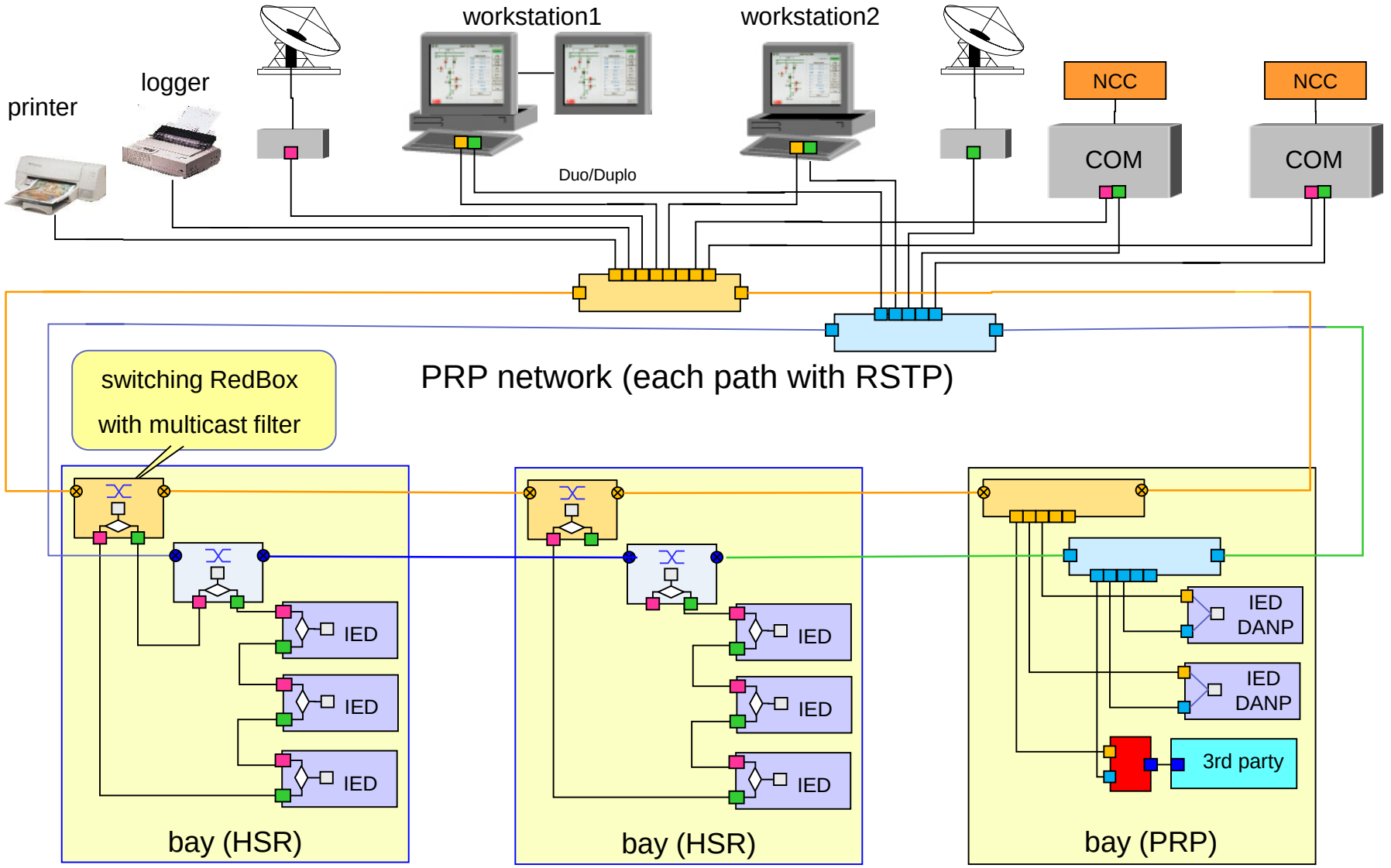
The RedBox forwards to the interlink any frame received from the ring that does not have its source registered in the Uplink Node Table and that has the correct LAN identifier (A or B).

Coupling HSR and two PRP LANs (sender in ring)



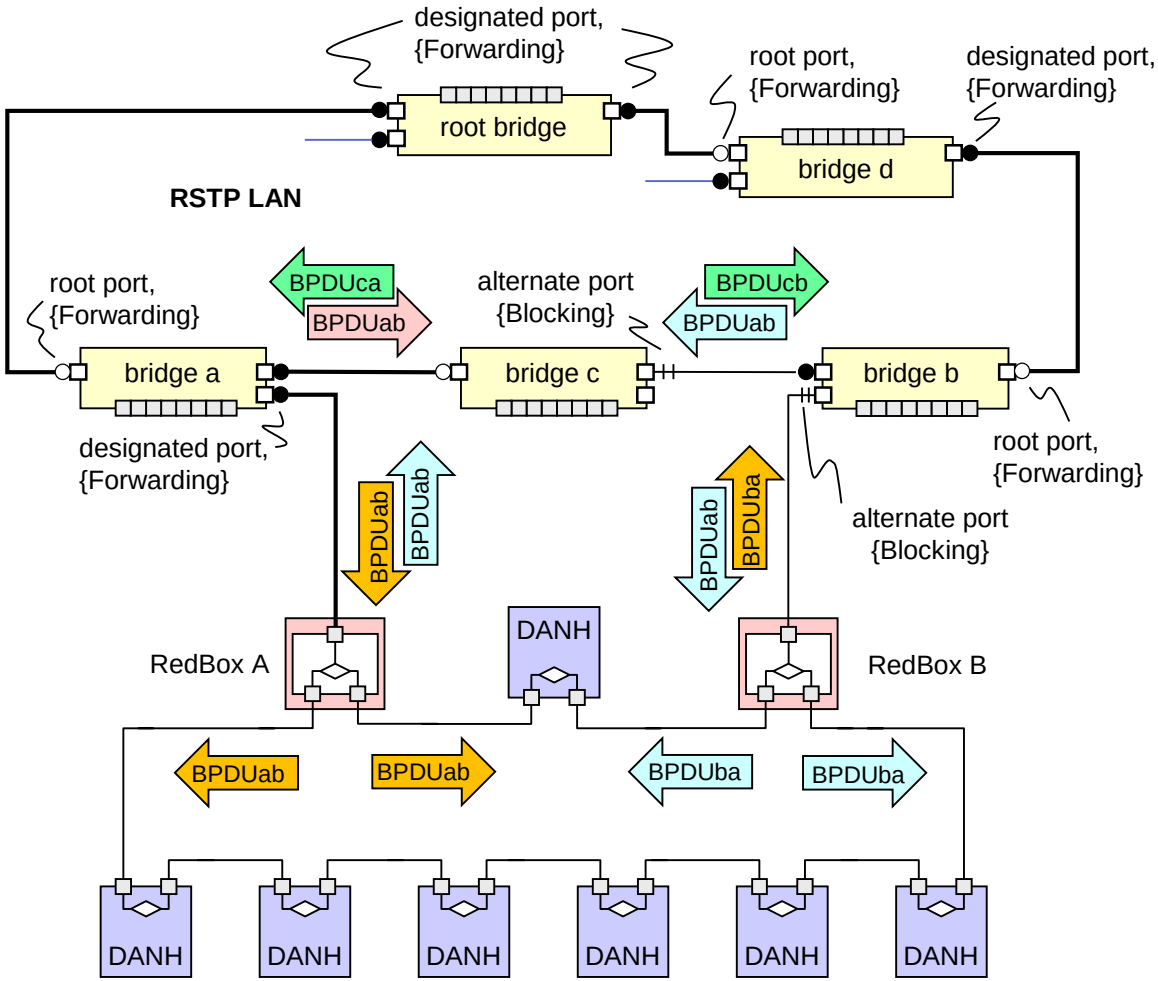
if RedBox A fails, connectivity would be lost between PRP and ring.
Therefore, a RedBox forwards whichever frame
A or B comes first and tags it with its color.
(the shaded frames are used for that purpose)

Example of full-redundant PRP/HSR network hierarchy



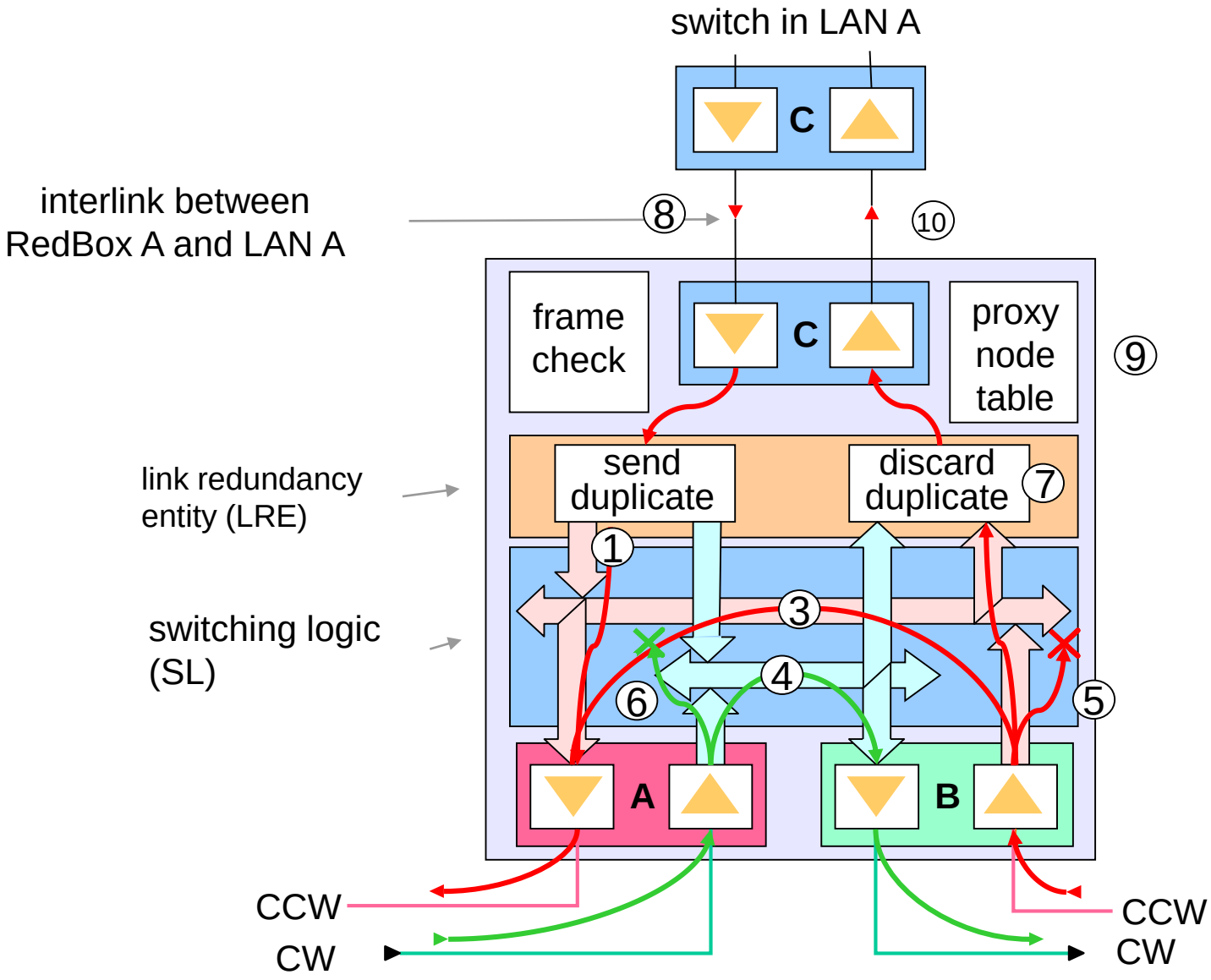
Mixing redundant, non-redundant, HSR and PRP

Coupling to RSTP networks (experimental)



An HSR ring can be connected by two RedBoxes to an RSTP LAN, provided the RedBoxes support the exchange of BPDUs (experimental extension)

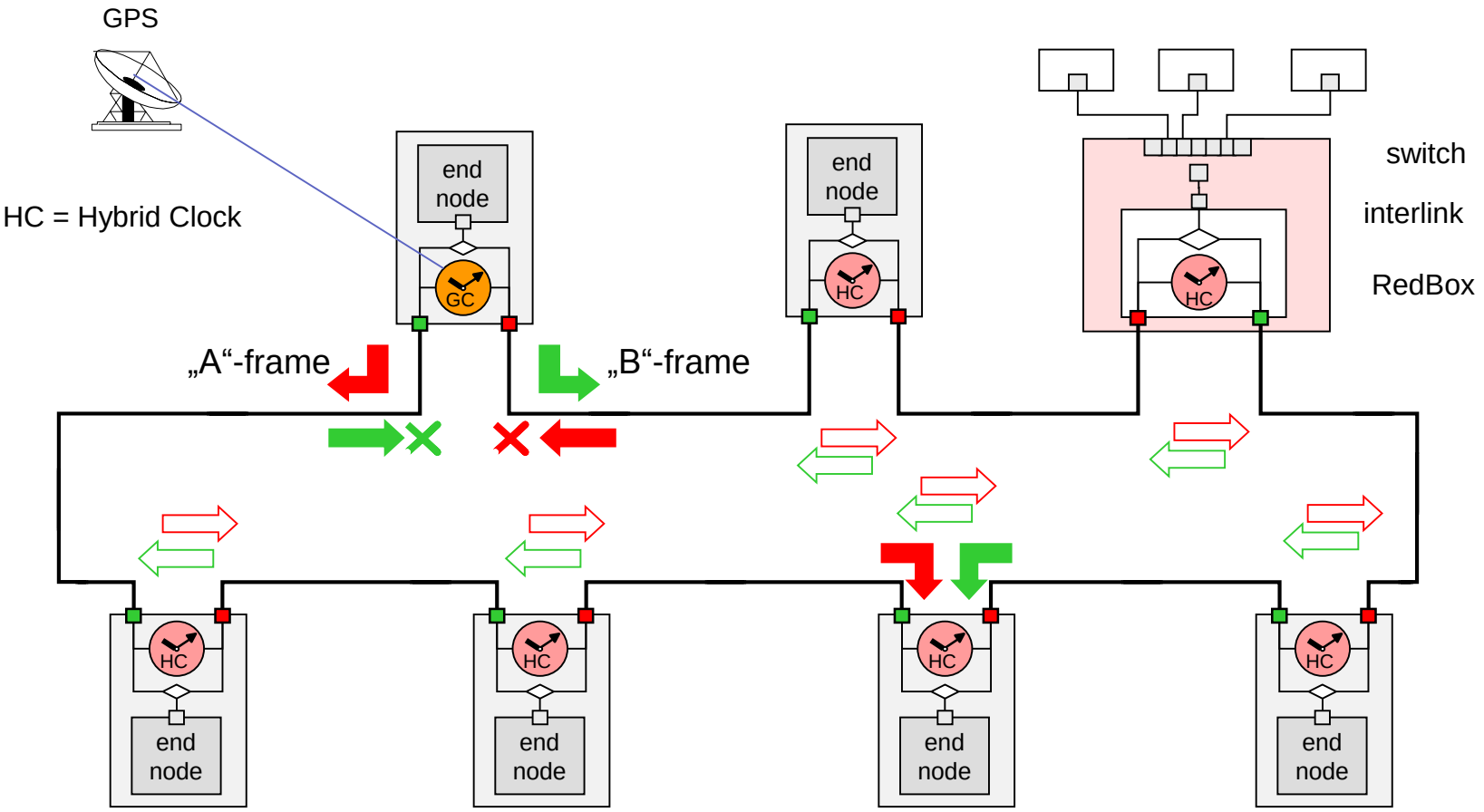
RedBox A coupling a ring to a PRP or RSTP network



CLOCK SYNCHRONIZATION

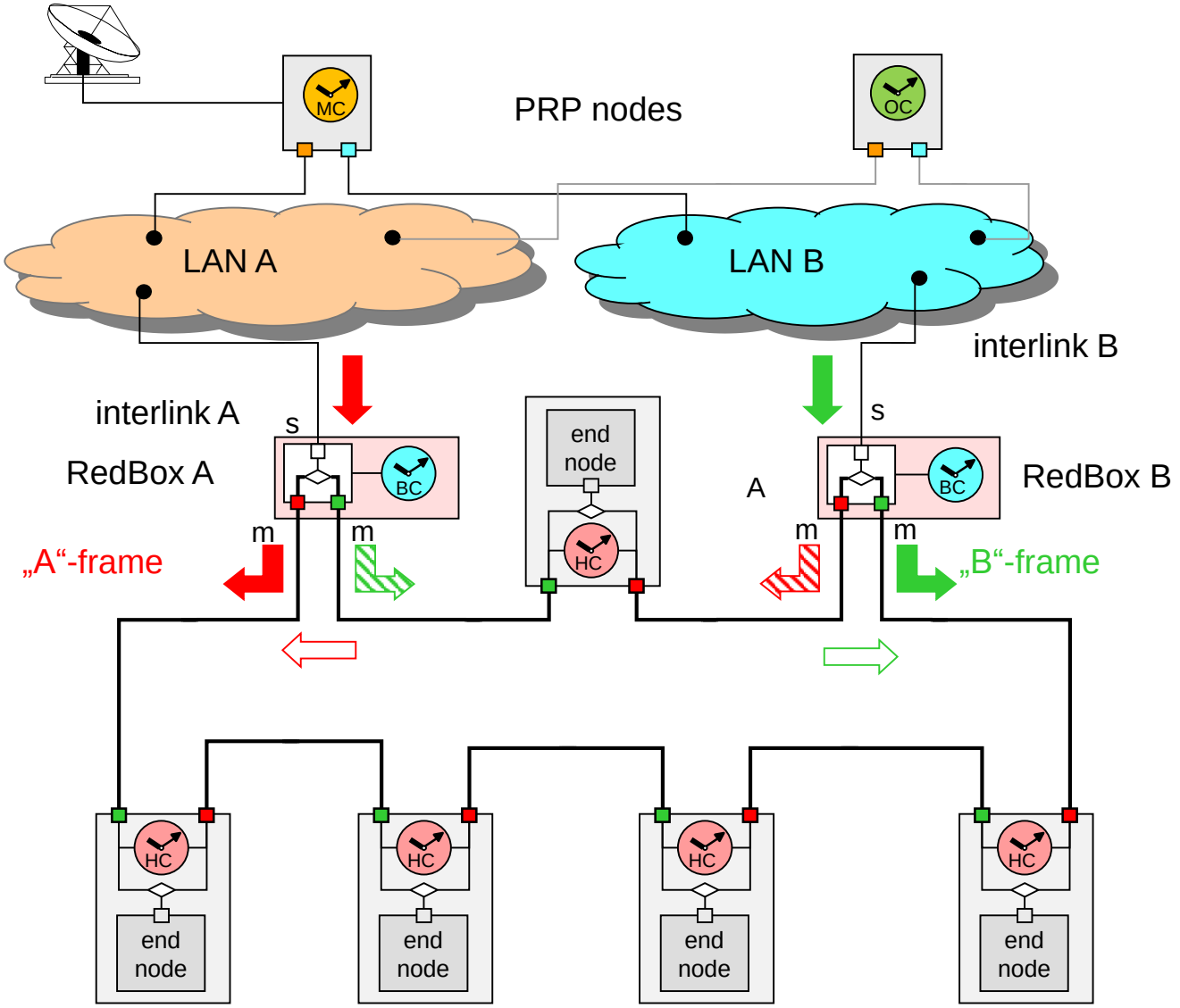
IEEE 1588v2 Layer 2 Peer-to-peer is the only clock protocol considered for HSR (see IEC 62439-3 Annexes)

Clocks in HSR



The transparent clocks (hybrid clocks) operate in both directions
The ordinary clock of the hybrid clock takes the time from the SYNC messages, from whichever direction

Clocks: coupling PRP and HSR



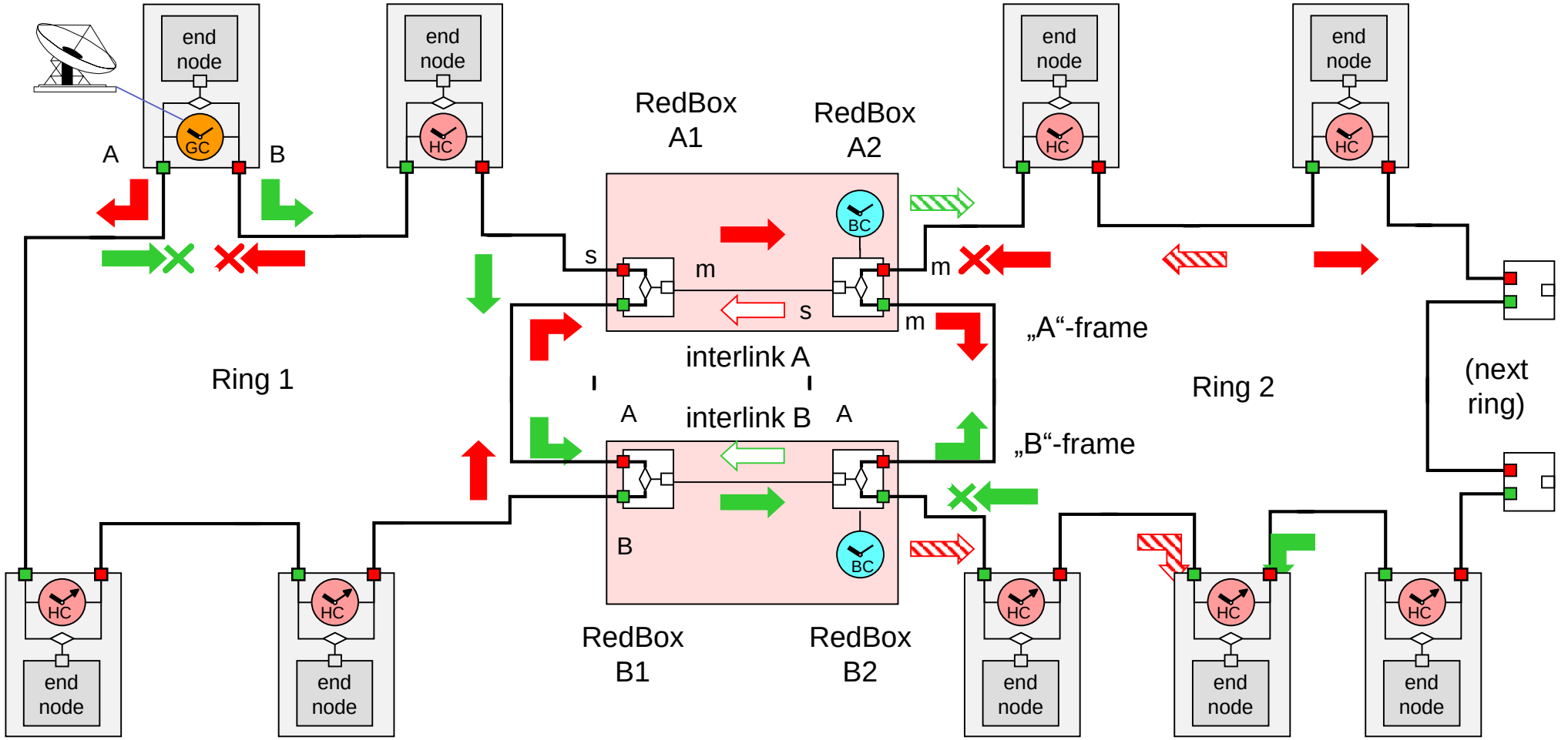
distinguish four cases:

- A received from A,
- A received from B
- B received from A
- B received from B

Clocks in two HSR rings coupled by QuadBoxes

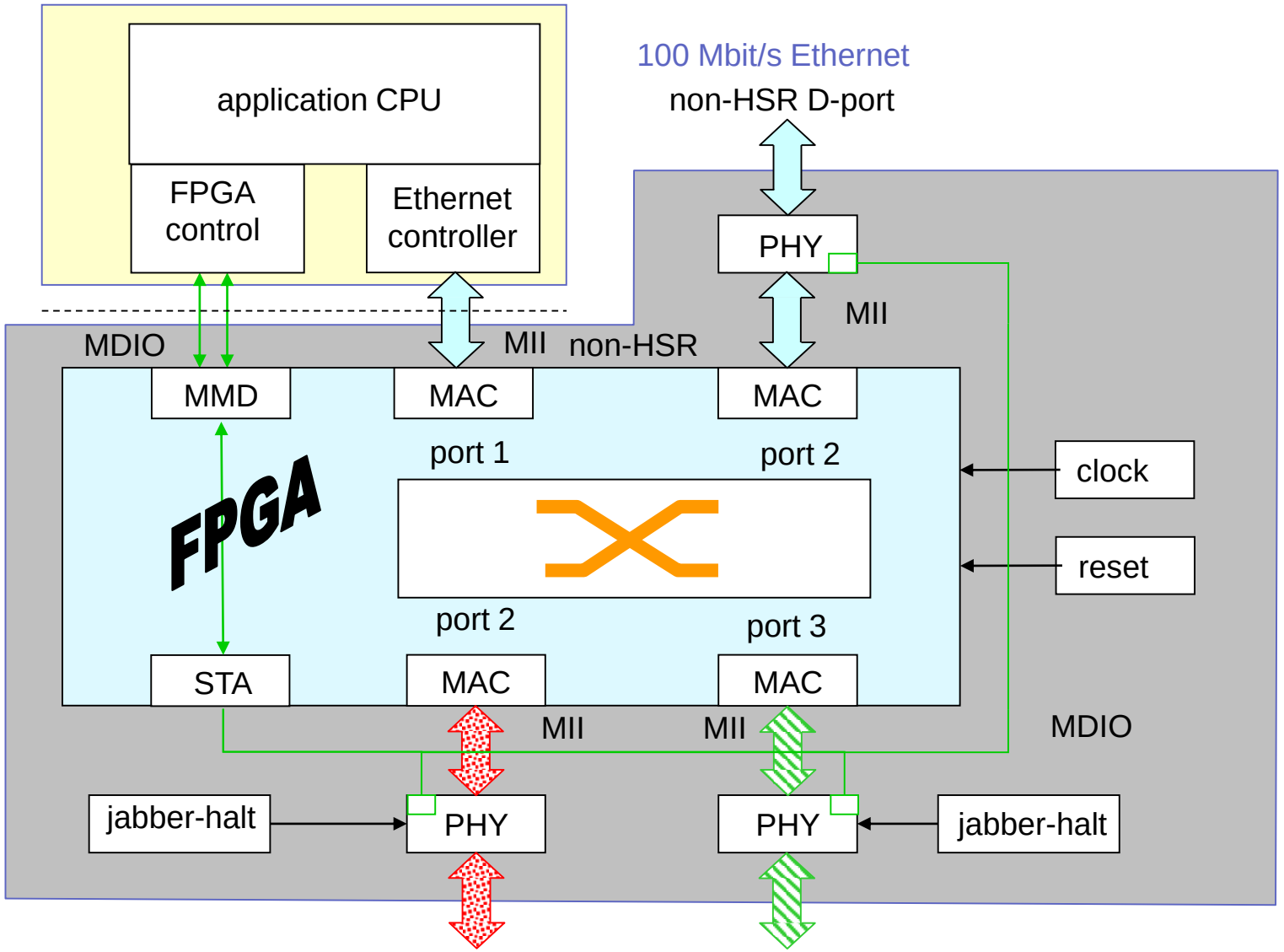
BC = boundary clock

HC = hybrid clock
(transparent clock + ordinary clock)

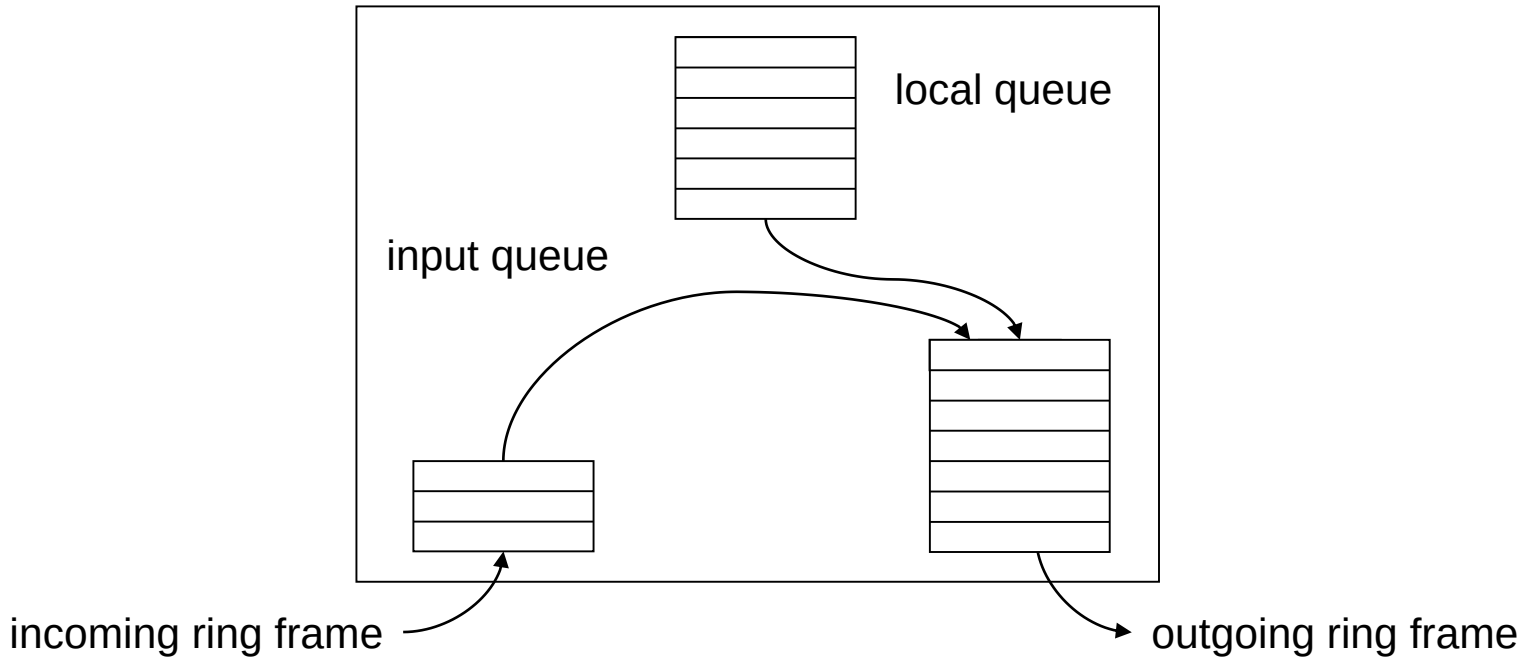


HSR IMPLEMENTATION

Implementation example



Cut-through



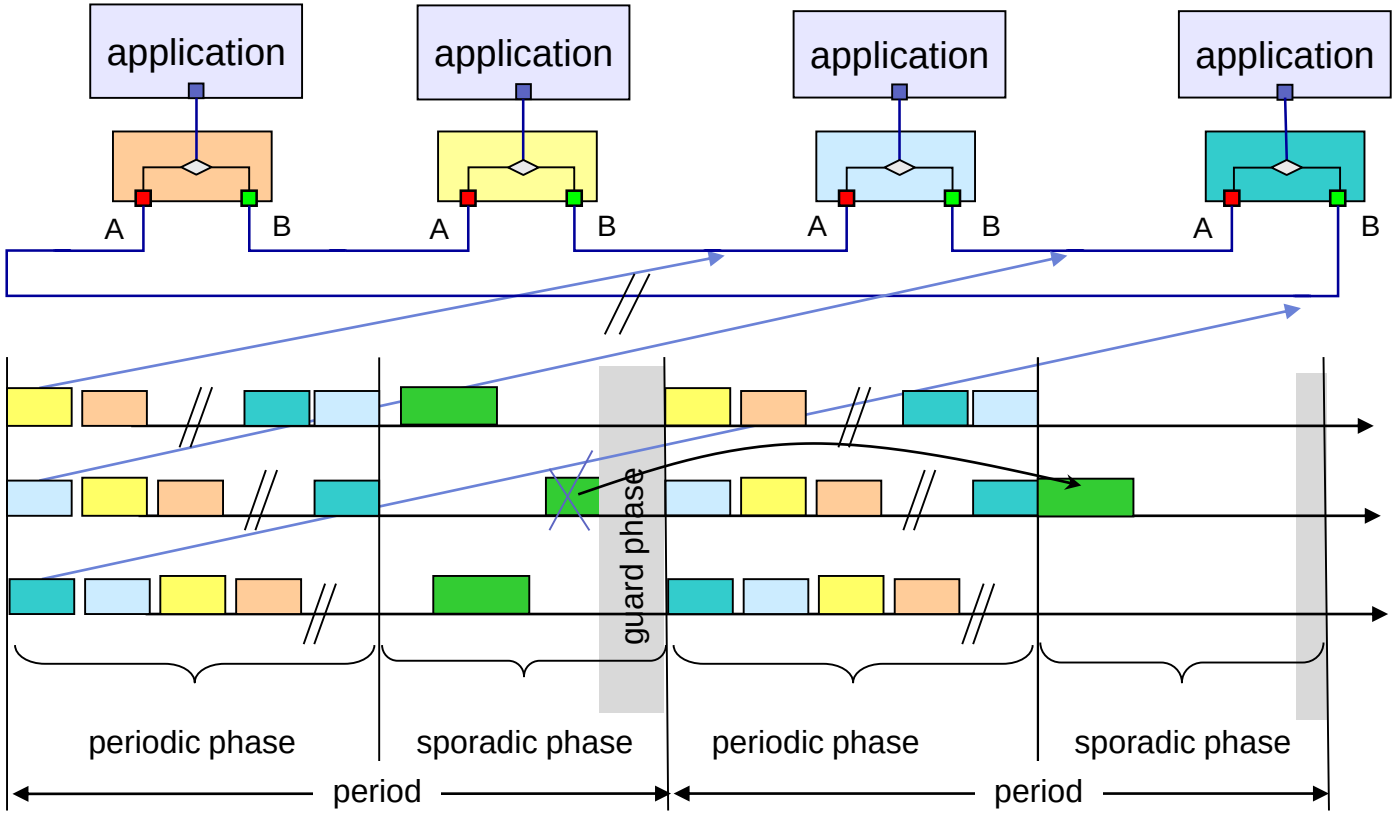
Cut-through (forwarding a frame as soon as its MAC header is received) improves the average delays, but the worst case delay occurs when a node just started sending an own frame of maximum length (1536 octets = 123 µs @ 100 Mbit/s) when a ring frame arrives. For this it has to buffer the ring frames up to a size of 1536 octets.

The node recognizes a frame it sent itself based on the MAC source address, but to remove damaged or ownerless frames from the ring, a node must store-and-forward frames coming from a source that once sent a damaged frame until a sufficient number of good frames came from that node.

HSR PRIORITIES AND DETERMINISM

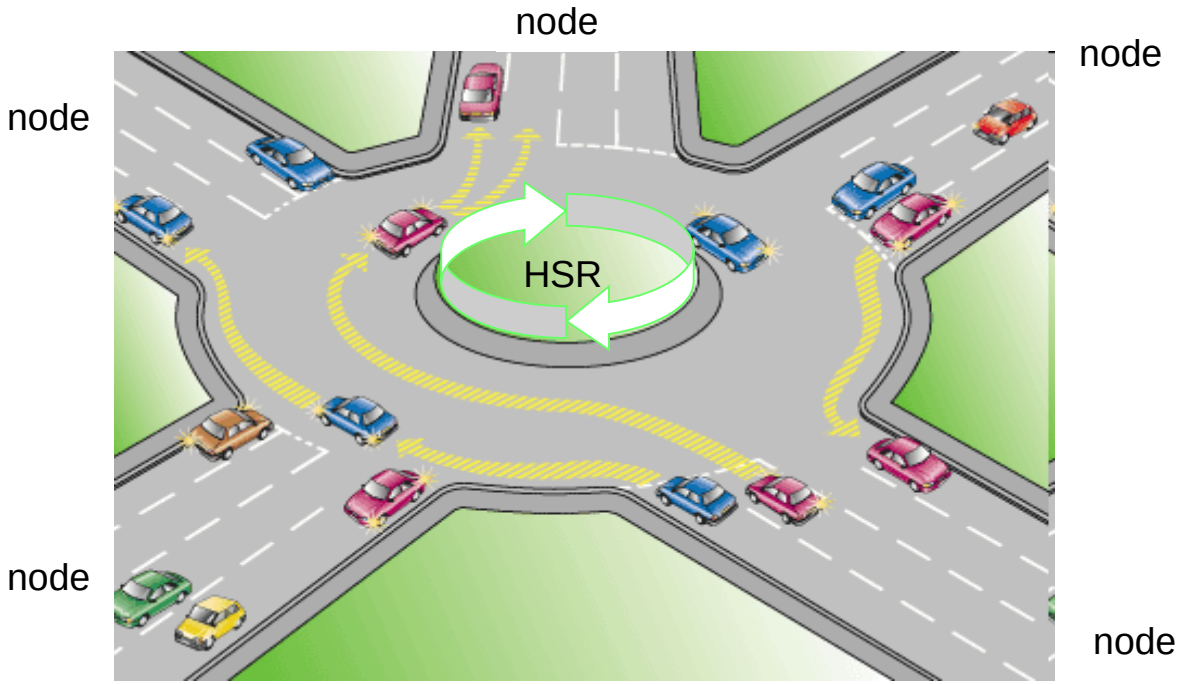
Deterministic, real-time scheduling

Relying on the precision clock given by IEEE 1588, all nodes transmit their (buffered) time-critical data (cyclic Sampled Measurement Values in IEC 61850) at the same time.
This queues the time-critical traffic and leaves a continuous slot for the aperiodic messages.
Sub-cycles with a power of 2 multiple of the base period are possible.
Nodes delay sending aperiodic messages if they would overlap the start of the next period (guard).



HSR priorities

HSR behaves like a roundabout: frames in the ring have a higher priority than inserted frames.
Cut-through allows wire-speed transmission from node to node, but is ineffective when another frame is already being transmitted in the next node (e.g. when a long truck is entering the roundabout)



HSR NETWORK MANAGEMENT

Network supervision and detection of lurking faults

Each node continuously checks all paths.

In order not to rely on application cyclic data for this, each node sends periodically a supervision frame (beacon) (over both ports) that indicates its state.
This frame is received by all nodes, including the sender, who can check the continuity of the network.

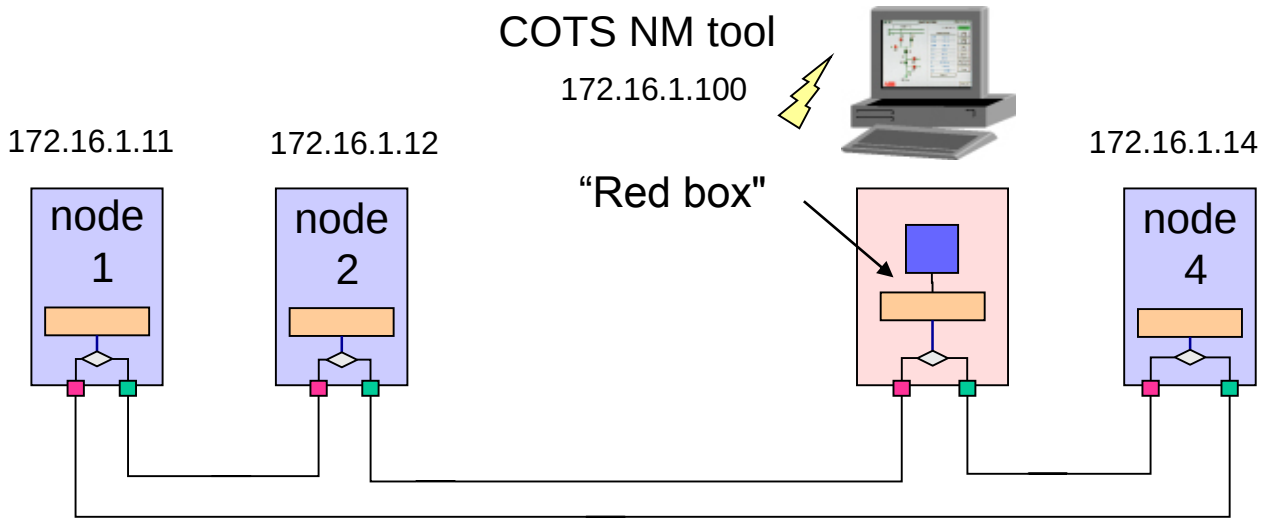
The beacon period is relatively long (some seconds) since the supervision frame is not needed for failover, but only to check redundancy.

The “duplicate discard” mode allows to keep track of all nodes in the network.

All nodes keep a node table of all detected partners and registers the last time a node was seen as well as missing duplicates and out-of-sequence frames.

Changes to the topology are communicated over SNMP or to the Link Management Entity, which can communicate them using the application protocol.

COTS attachment and network management



Each node has an SNMP agent to keep track of redundancy.

COTS devices are attached through a “RedBox” that hides the HSR traffic from the device.

HSR: Pros & Cons

+ seamless failover in case of failure of a node or reinsertion of a repaired node
+ needs no switches or bridges, linear topology possible
+ needs no duplication of the network, economical redundancy
+ supervises constantly the redundancy
+ monitors actual topography (over network management / SNMP)
+ international standard (IEC 62439-3 Clause 5)
+ interoperable with PRP (IEC 62439-3 Clause 4)
+ supports clock synchronization (IEEE 1588) with a transparent clock in every node
+ can be used for any Industrial Ethernet
+ application-protocol independent
+ no intellectual property: open specification and free licence

- uses four fibres (100 Fx) or 2 cables (100Tx) per node
- non-HSR devices (COTS) can only be inserted over a “RedBox or a “Quadbox”
- limited to a layer 2 broadcast domain
- requires a hardware implementation (ASIC or FPGA) to meet the real-time constrains,
  + which can also be used for clock synchronization.

Application to IEC 61850

- supports the layer 2 communication of GOOSE (IEC 61850-8-1) and SMV (IEC 61850-9-2)

- offers the seamless switchover as defined in 61850-5 § 14

- offers the same redundancy scheme and hardware for the station bus and the process bus

- can expose the link layer redundancy objects through the management interface directly as IEC 61850 objects or also using SNMP.

- can use the same SCD files as the non-redundant structure since the IP addresses are not affected and the MAC addresses are the same. In the communication section, the redundant switches appear as additional devices with their own IP address.

Implementation

Several companies implemented the protocol only relying on the specifications.

Intellectual property is available under fair and non-discriminatory conditions.

An interoperability test allowed to check the implementations.

A first implementation was done in software, which precludes cut-through.
It did not meet the real-time requirements, but served as proof of concept
and is available for PCs free of charge.

A switch fabric is highly recommended to increase performance.
At least four FPGA implementations exist (August 2010).

Experienced switch manufacturers will provide implementations that can be used under license by any company.

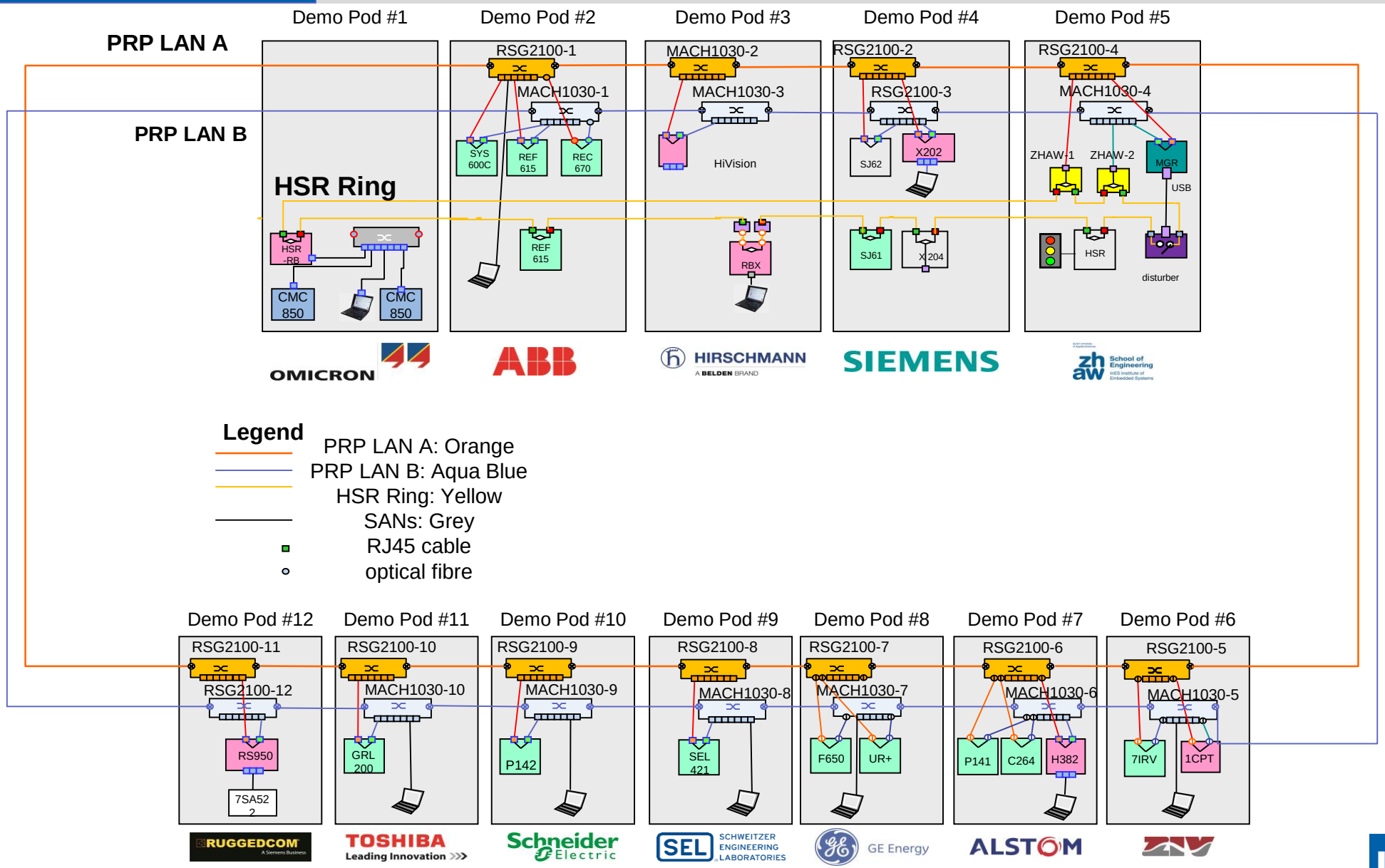
HSR TESTING AND EXPERIENCE

CIGRE 2010 demo

- Hirschmann, Siemens, ABB, ZHAW and Flexibilis presented an HSR interoperability demo at CIGRE 2010 in Paris

CIGRE 2012 Demo

CIGRE 2012 Demo network



HSR conclusion

- IEC standard 62439-3 since February 2010
- specified as the redundancy solution in IEC 61850 Ed. 2
- clock profile specified in IEC 62439-3 Annex A,B,C,D,E
- fulfills the most critical redundancy and real-time requirements
- has the potential of displacing all other layer 2 protocols in industry
- demonstrated by ABB, Siemens, Hirschmann, ZHAW, RuggedCom, Flexibilis, SoCE
- synchronized by an IEEE 1588 one-step clock, allowing deterministic operation
- simulated for large networks
- complements and compatible with PRP - can be operated in mixed topologies
- can be implemented with FPGAs of reasonable size and price
  (Altera Cyclone III, Xylinx Spartan 6)