Rapid spanning tree protocol

Mario Baldi
Politecnico di Torino
http://staff.polito.it/mario.baldi

Pietro Nicoletti
Studio Reti
http://www.studioreti.it

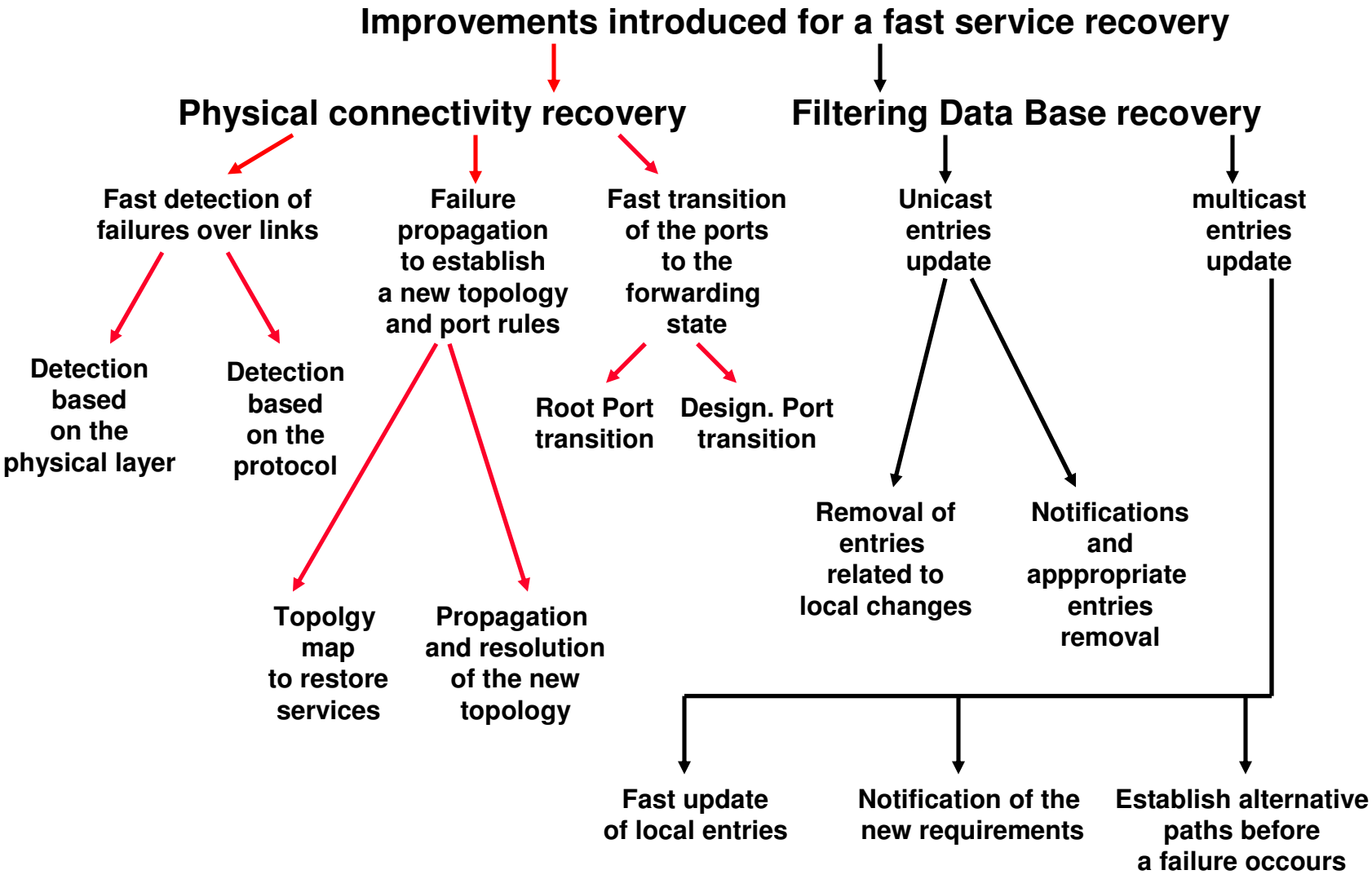
Based on chapter 4 of:
M. Baldi, P. Nicoletti, "Switched LAN", McGraw-Hill, 2002, ISBN 88-386-3426-2

Copyright note

Rapid Spanning Tree Protocol (RSTP)

- Fast convergence Spanning tree:
  - Less than 1 second
- Standard IEEE 802.1w approved in fall 2001
- Interoperable with STP IEEE 802.1D
  - Without fast convergence
- Modern solution for mission critical BLAN
- Operate only on point-to-point links
  - Direct connections (no hubs)
- Replace proprietary solutions for resiliency with fast convergence introduced by some vendors

Improvements over STP



Port roles

- Blocking state replaced by Alternate & Backup port rules
- Alternate
  - Port that offers an alternate path in the direction of the Root Bridge to that provided by the Bridge own Root Port.
  - Alternate port to be used in case of root port failure
- Backup
  - Acts as a backup for the path provided by a Designated Port in the direction of the leaves of the Spanning Tree.
  - Backup Ports exist only where there are two or more connections from a given Bridge to a given LAN.
- Root
- Designated
- Disabled

Alternate port and Backup port

- IEEE 802.1w assumes that there is not a Blocking state
- Ports connected to alternate paths to the root are selected as Alternate
- Different ports of the same Bridge connected to the same LAN:

  A designated port

  The remaining ones are backup ports

  - Point-to-point connection between two ports of a Bridge
  - Two or more ports of a Bridge connected by a repeater
    - The two ports are connected to the same LAN

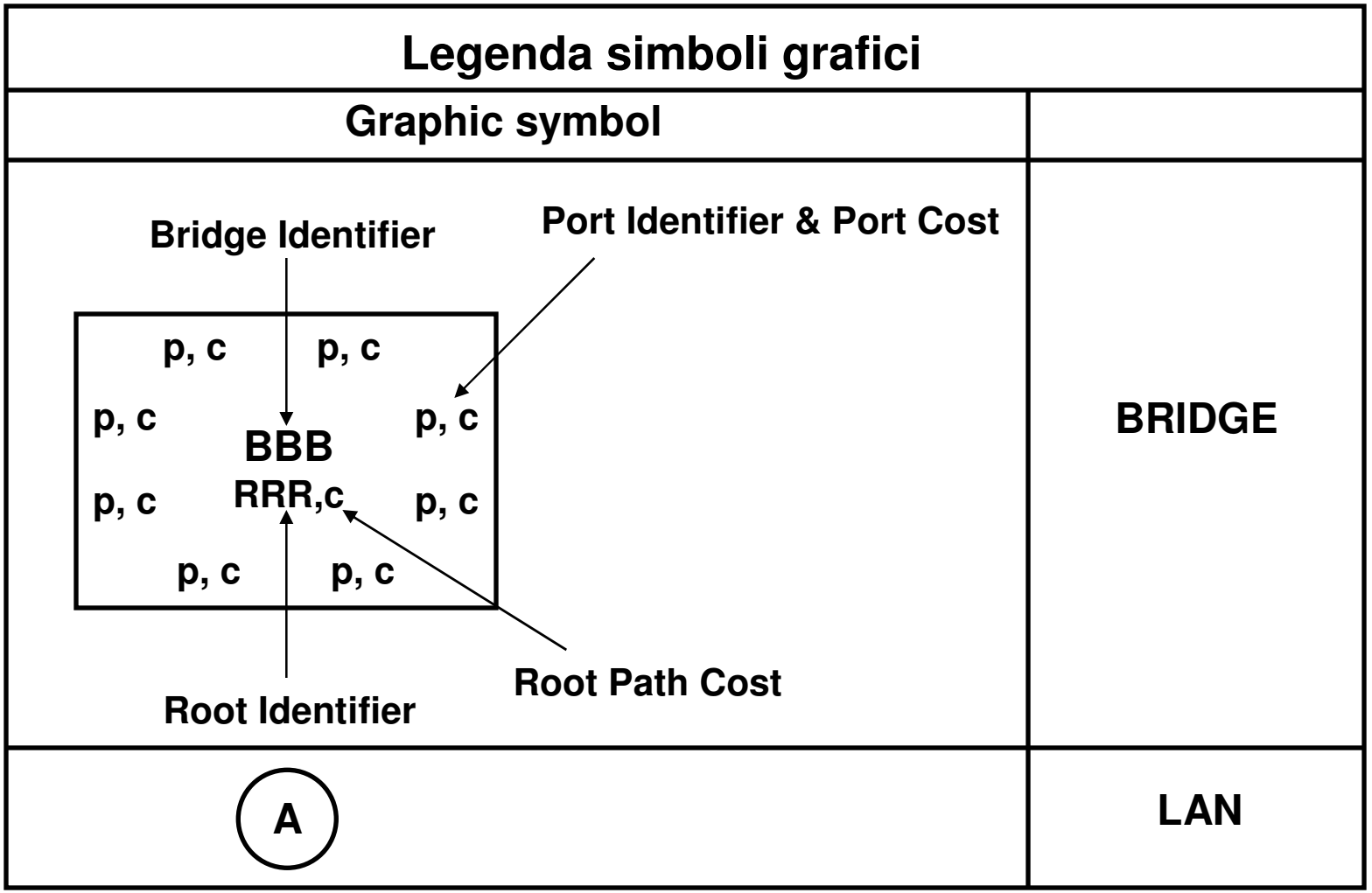
Edge port

- Edge port is a terminating port that connect a end-station
- Edge port does not work as other port:
  - As soon detect a Link Integrity Test go immediately to Forwarding state without Listening e Learning state
  - A port changing state do not cause a TCN BPDU transmission through root port

Port states: similarities between STP and RSTP

STP – port states	Administrative Port states	MAC address operativity	RSTP – port states	Active RSTP topology (port roles)
DISABLED	Disabled	FALSE	Discarding	Excluded (Disabled)
DISABLED	Enabled	FALSE	Discarding	Excluded (Disabled)
BLOCKING	Enabled	TRUE	Discarding	Excluded (Alternate, Backup)
LISTENING	Enabled	TRUE	Discarding	Included (Root, Designated)
LEARNING	Enabled	TRUE	Learning	Included (Root, Designated)
FORWARDING	Enabled	TRUE	Forwarding	Included (Root, Designated)

Diagrammatic conventions



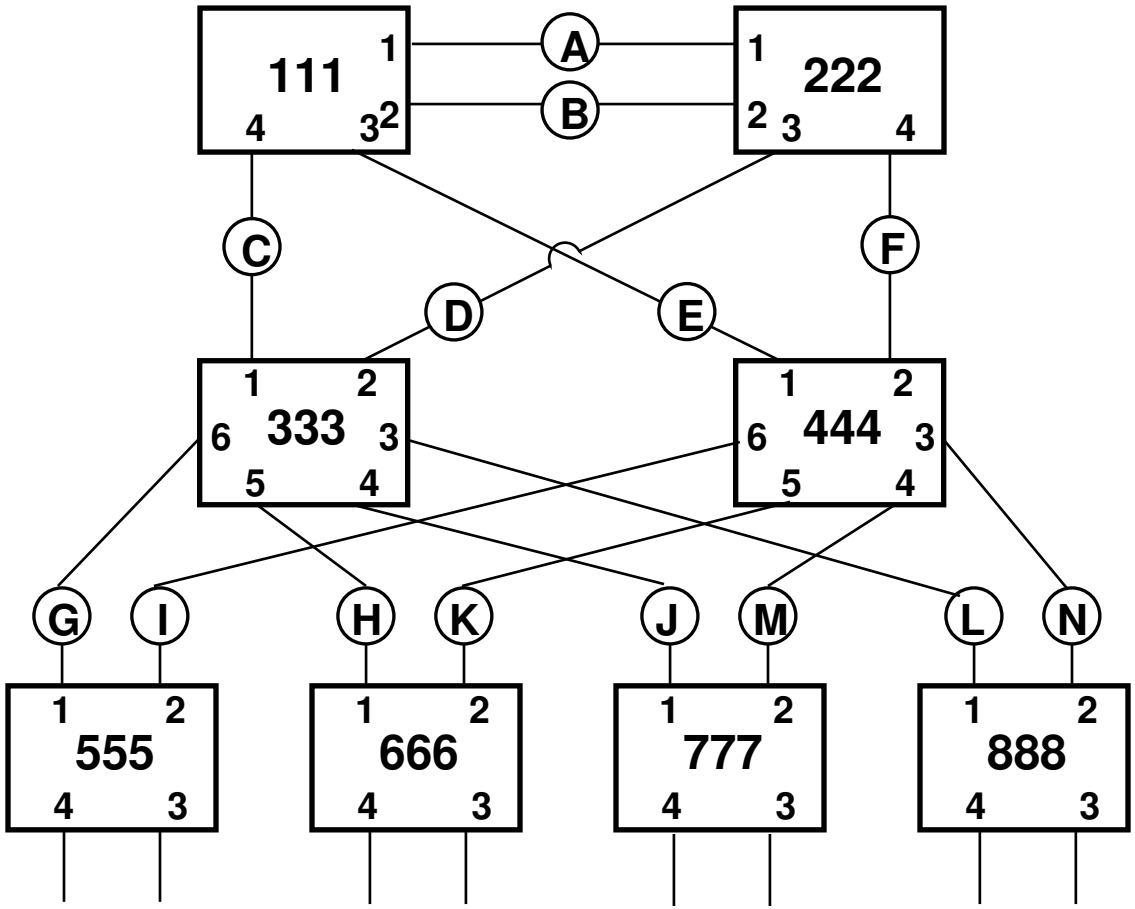
Conventional graphic symbols

Port role	Port state	Legend
Designated	**Discarding** **Learning** **Forwarding**	
edge port	**Forwarding**	
Root Port	**Discarding** **Learning** **Forwarding**	
Alternate	**Discarding**	
Backup	**Discarding**	
Disabled	**-----**	
BPDU sent		
Designated **Designated Proposal** **Root** **Root Proposal**		

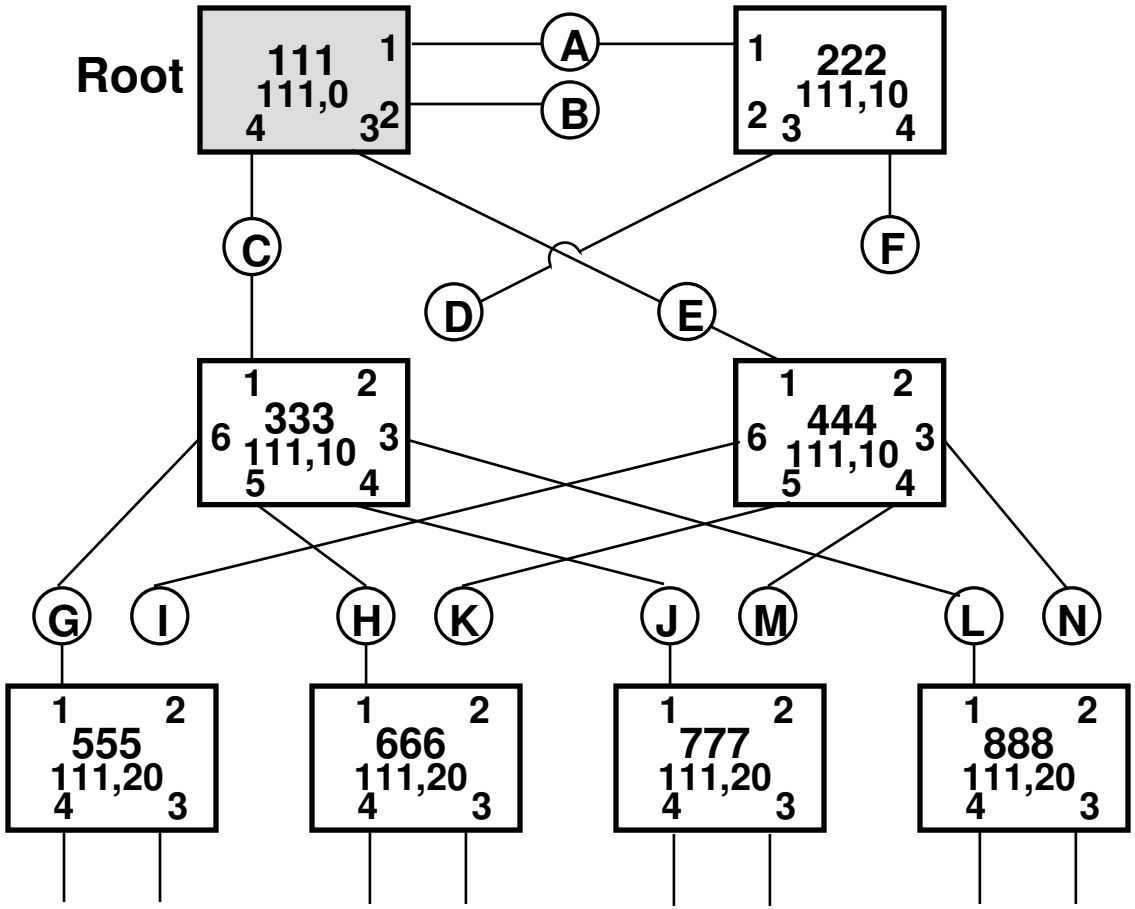
The alghoritm principle

- The tree is created in the same way as RTP
  - Root bridge election
  - Root port definition
  - Designated port definition
- The port not selected as root or designated are selected as:
  - Alternate if connected to a port on different bridge
  - Backup if connected to a different port of the same bridge

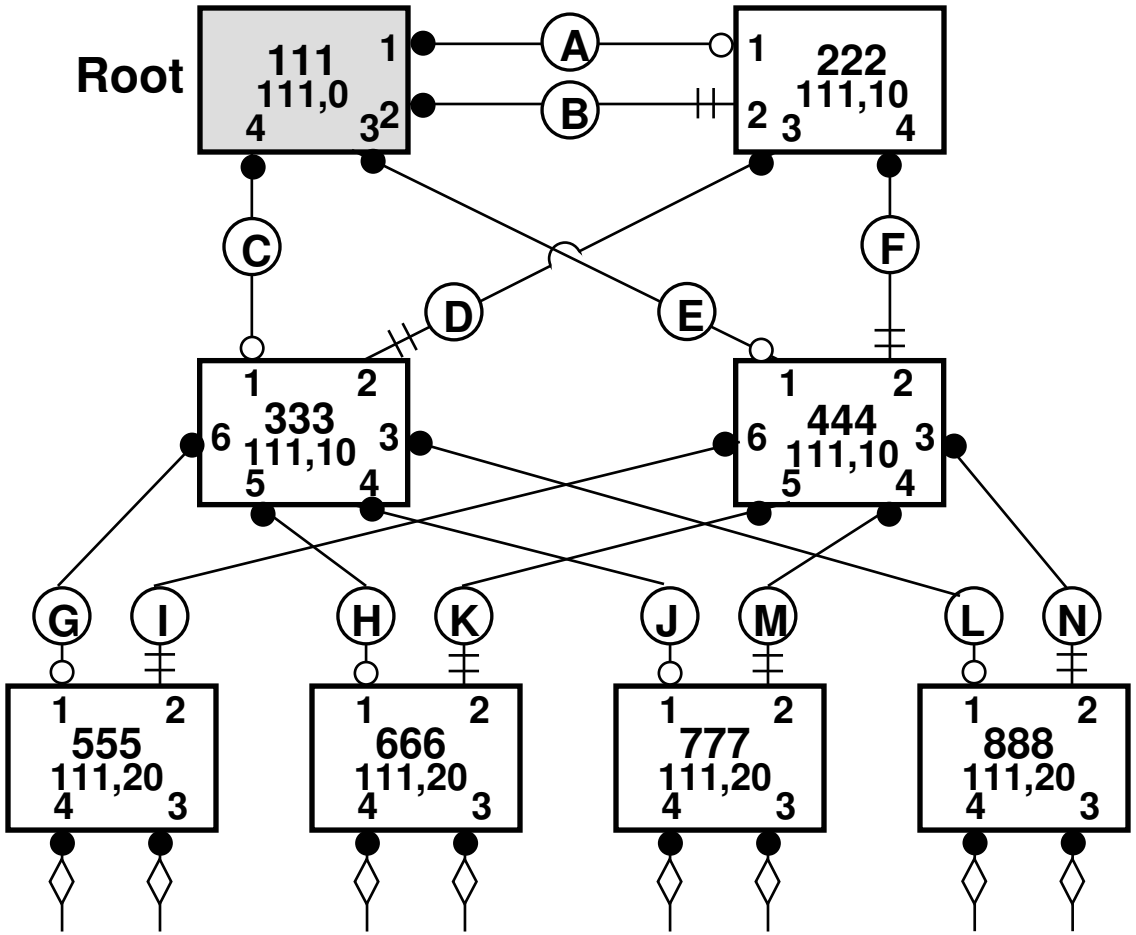
Physical redundant star topology



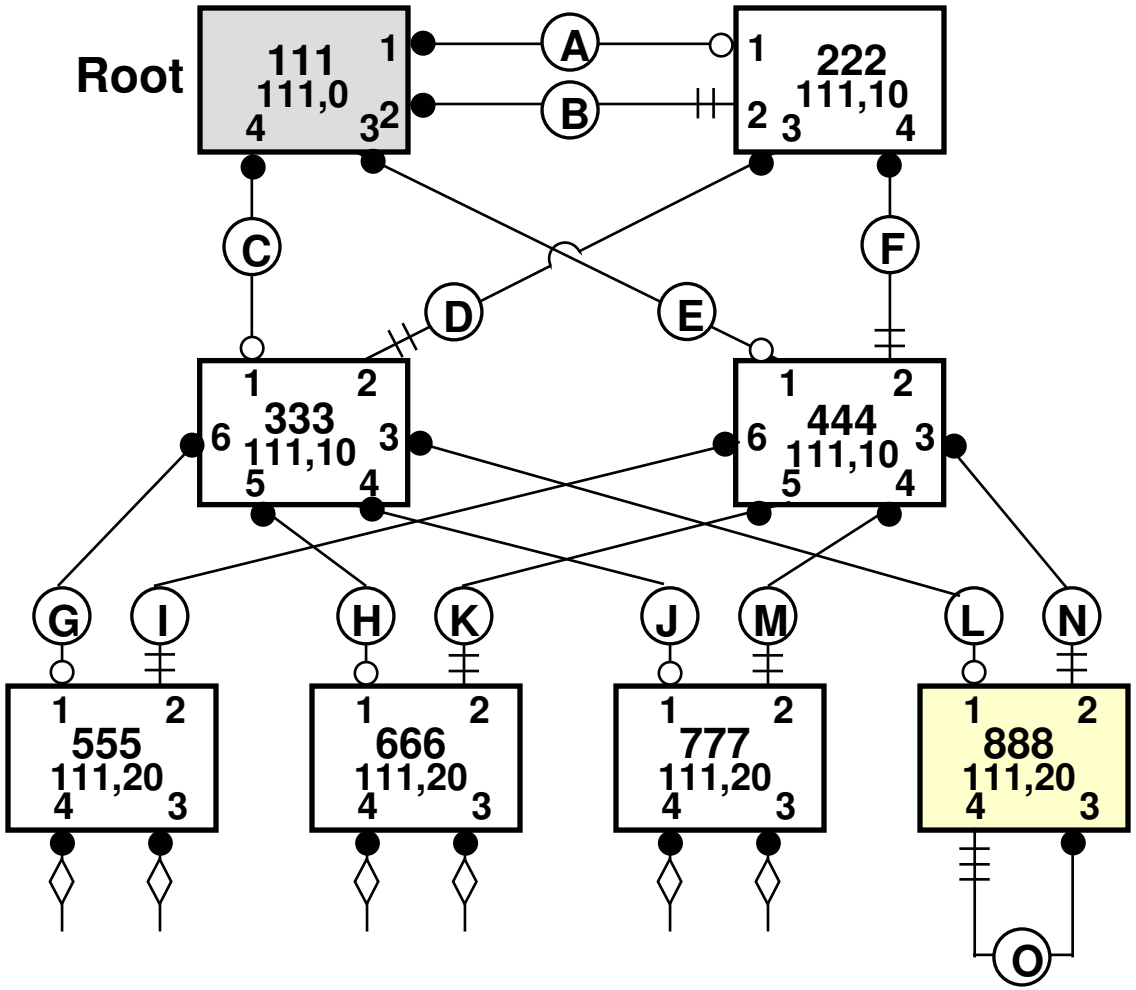
RSTP active topology



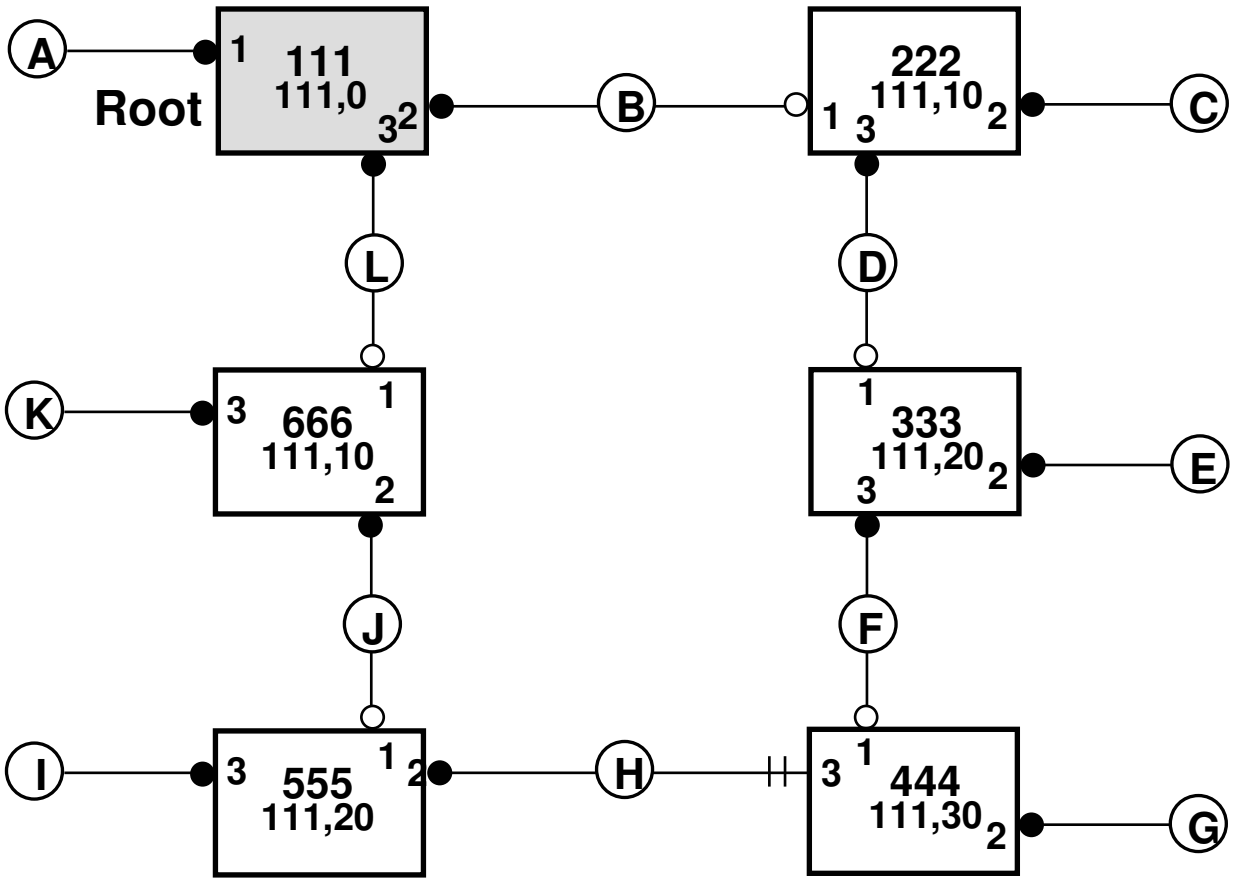
RSTP tree & port state



Backup port example

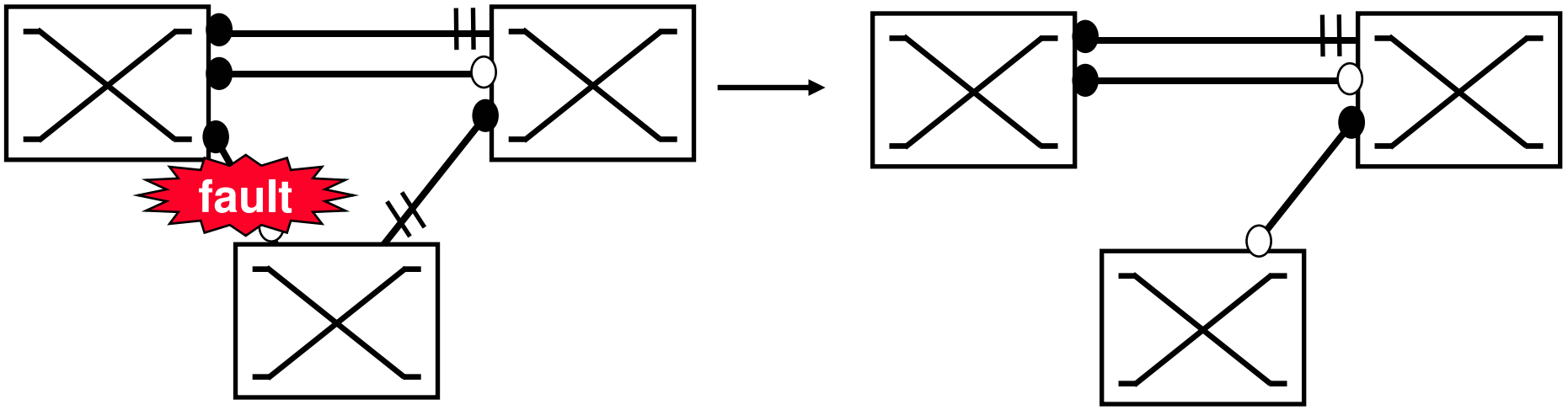


Ring topology



Fault recovery based on physical level fault detection

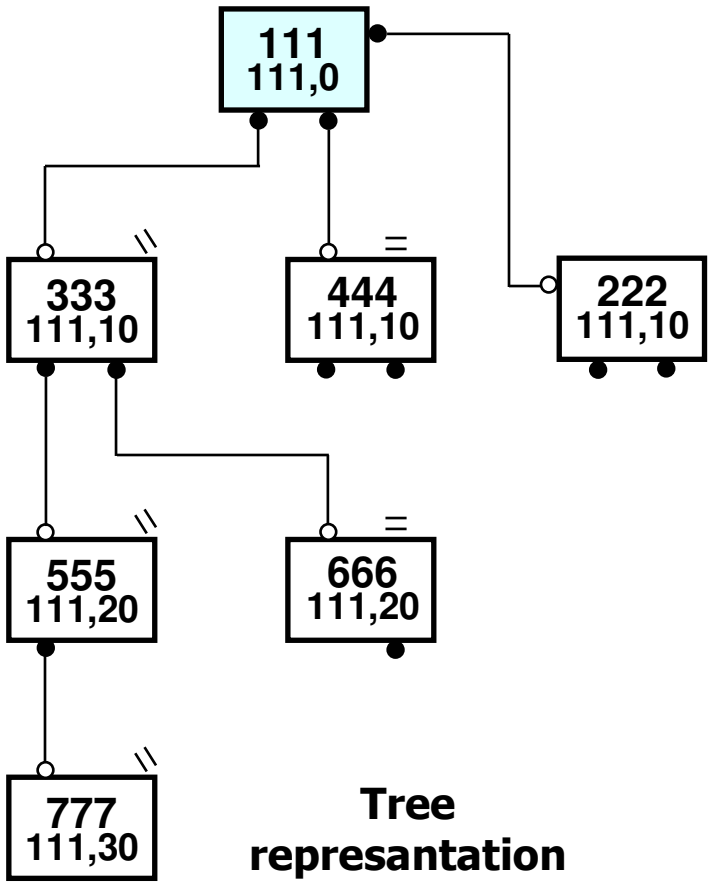
- If the switch looses a root port, the alternate port is activated immediately: that port becomes a new root port
  - Fault detection and recovery based on physical layer



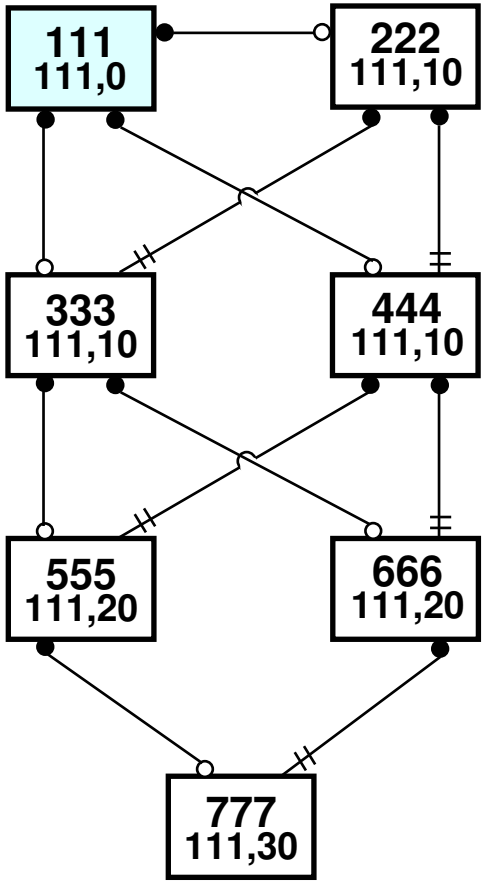
Fault detection and flush entry

- The bridge flushes immediately the entries that have a faulty port as destination port
  - Local flush entries on bridge ports at the end of a faulty link
  - Bridges directly connected to a faulty bridge
- If a root port fails, an alternate port becomes root
- A bridge that turns an Alternate Port into a Root Port sends a ***Topology Change Notification*** (TCN) BPDU through the root port
  - On each port in the forwarding state
- When a TCN is received:
  - Flush the entries on every ports except the port were the TCN as been received
  - Forward TCN on every ports in forwarding state
    - except the port where the TCN has been received

Flush entry example: initial configuration

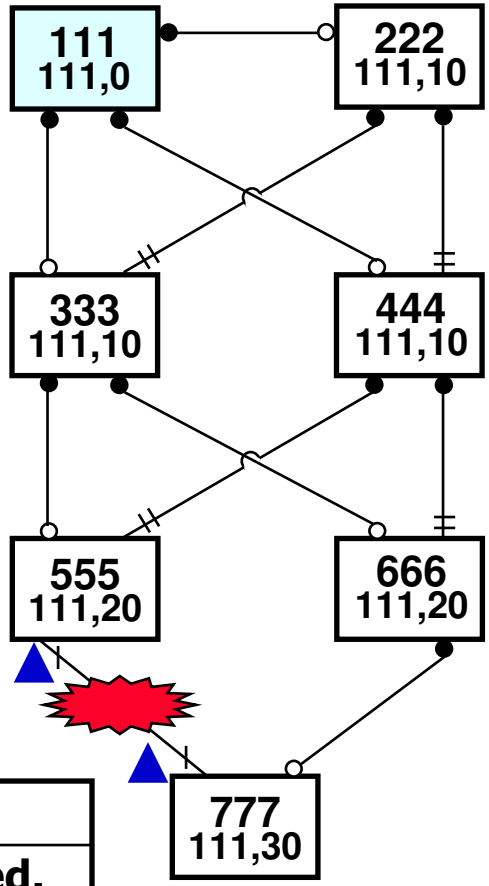


Tree representation



Flush entries: step 1

- The link 555-777 fails
- both 555 and 777 notice it
- It can be seen that all the Ports indicated will need to be flushed.

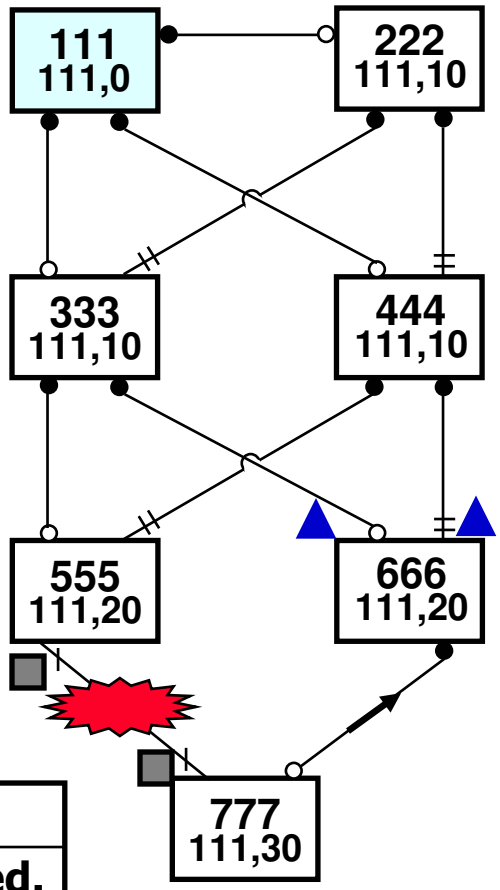


Legend

- **Addresses learnt on these Ports need to be flushed.**
- **Addresses learnt on these Ports have been flushed.**
- **TCN transmitted in the direction of the arrow**

Flush entries: step 2

- 555 and 777 flush addresses for their faulty Ports.
- 777 changes the port state from alternate to root
- 777 sends a TCN message (an RSTP BPDU with TC set).
  - The TCN message is just like a Configuration-BPDU with TC flag set to 1

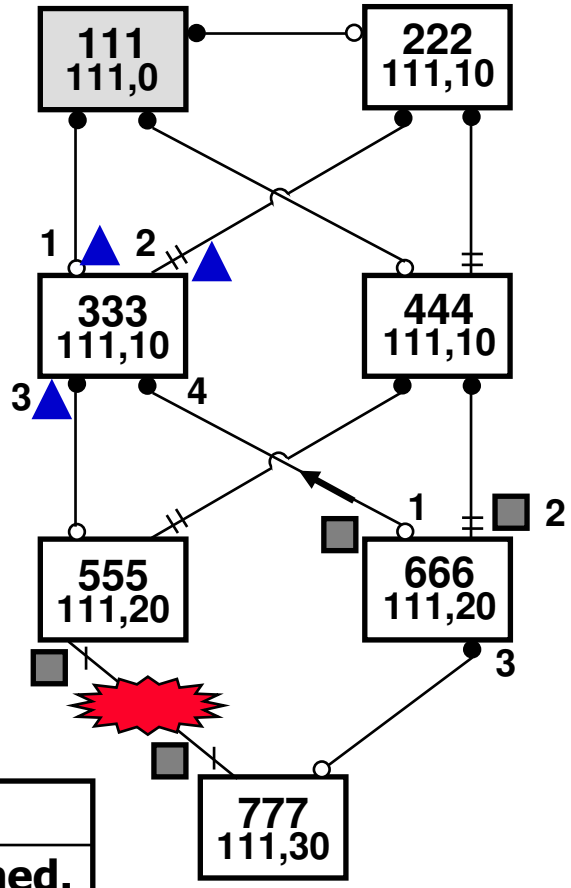


Legend	
▲	**Addresses learnt on these Ports need to be flushed.**
■	**Addresses learnt on these Ports have been flushed.**
→	**TCN packets sent in the direction of the arrow**

Flush entries: step 3

bridge 666

- Receives the TCN message
- Removes the entries related to other ports
  - Port 1 and port 2
- Sends the TCN message from every other port in the forwarding state
  - Port 1 up to bridge 333

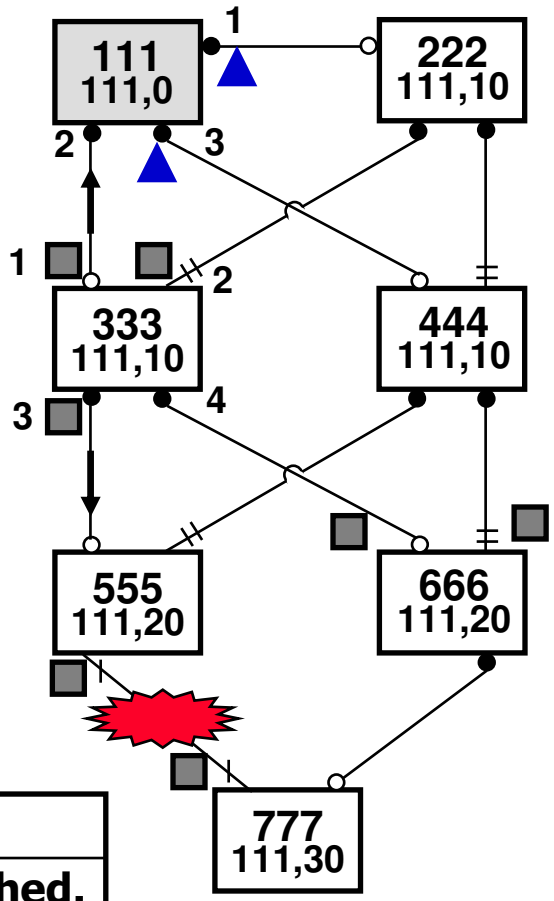


Legend	
▲	**Addresses learnt on these Ports need to be flushed.**
■	**Addresses learnt on these Ports have been flushed.**
→	**TCN packets sent in the direction of the arrow**

Flush entries: fase 4

Bridge 333

- Receives the TCN message from port 4
- Removes the entries related to other ports
  - Port 1, port 2 and port 3
- Sends the TCN message from all its port in the forwarding state
  - Port 1 up to bridge 111
  - Port 3 up to bridge 555

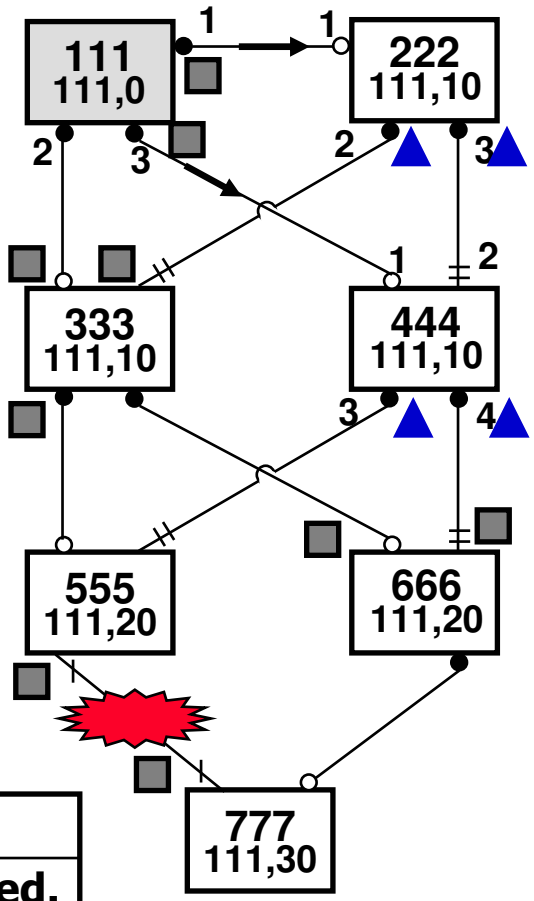


Legend	
▲	**Addresses learnt on these Ports need to be flushed.**
■	**Addresses learnt on these Ports have been flushed.**
→	**TCN packets sent in the direction of the arrow**

Flush entries: step 5

Bridge 111

- Receives the TCN message from port 2
- Removes the entries related to other ports
  - Port 1 and port 3
- Sends the TCN message on the other ports in the forwarding state
  - Port 1 up to bridge 222
  - Port 3 up to bridge 444

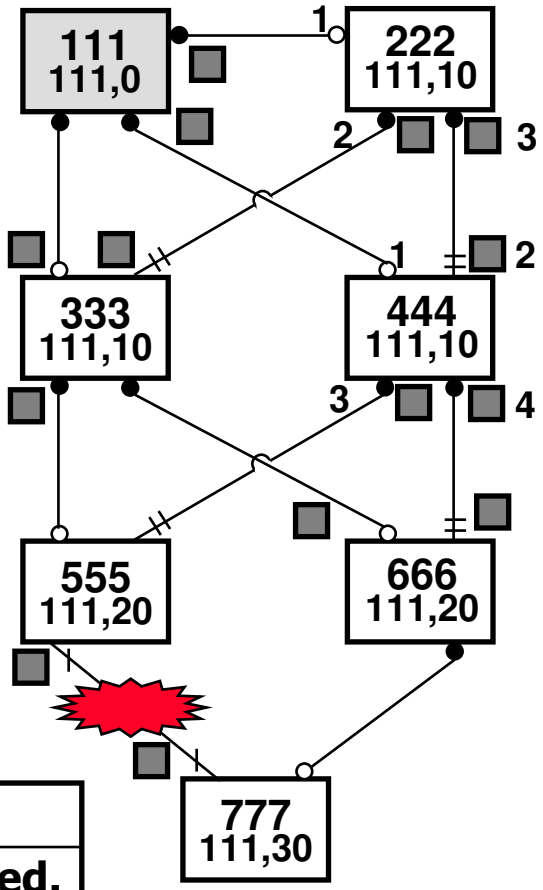


Legend	
▲	**Addresses learnt on these Ports need to be flushed.**
■	**Addresses learnt on these Ports have been flushed.**
→	**TCN packets sent in the direction of the arrow**

Flush entries: step 6

bridge 222 and 444

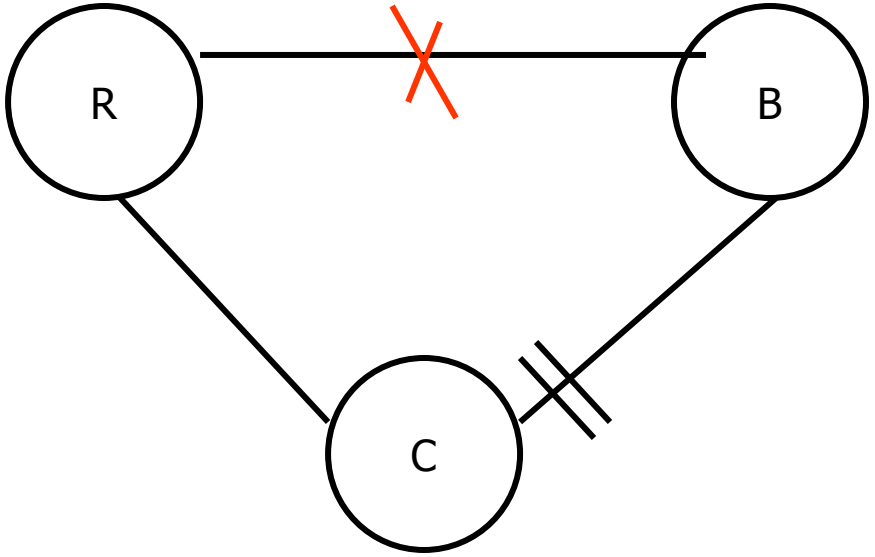
- Receive a TCN message from one of their ports
- Removes the entries related to other ports (as shown in the figure)
- TCN messages sent by them reach only ports in the discarding state ???



Legend	
▲	**Addresses learnt on these Ports need to be flushed.**
■	**Addresses learnt on these Ports have been flushed.**
→	**TCN packets sent in the direction of the arrow**

Proposal/agreement mechanism

- A bridge B near to the root, as soon as the link falls, proposes itself as root.
- The other bridges (i.e. bridge C) inform bridge B that another root is present and bridge B accepts it.



IEEE 802.1w recommended path cost

Port speed	recommended value	recommended range	Acceptable value
<= 100 Kb/s	200.000.000	20.000.000 - 200.000.000	1 - 200.000.000
1 Mb/s	20.000.000	2.000.000 - 200.000.000	1 - 200.000.000
10 Mb/s	2.000.000	200.000 - 20.000.000	1 - 200.000.000
100 Mb/s	200.000	20.000 - 2.000.000	1 - 200.000.000
1 Gb/s	20.000	2.000 - 200.000	1 - 200.000.000
10 Gb/s	2000	200 - 20.000	1 - 200.000.000
100 Gb/s	200	20 - 2000	1 - 200.000.000
1 Tb/s	20	2 - 200	1 - 200.000.000
10 Tb/s	2	1 -20	1 - 200.000.000

DSAP	SSAP	Length	DSAP	SSAP	Control	
Multicast 01-80-C2-00-00-00	**Singlecast** Bridge Address	**XY**	**042H**	**042H**	**XID**	**Configuration message**

00000000 = STP
00000010 = RSTP

Configuration BPDU STP and RSTP

Byte	Configuration message
1÷2	**Protocol Identifier** 00-00
3	**Protocol Ver. Identifier** 00
4	**BPDU Type** 00
5	**TC** | **Flags** P PR PR A | **TCA**
6÷13	**Root Identifier** First 2 byte = Bridge Priority last 6 byte = MAC address of the Root Bridge
14÷17	**Root Path Cost**
18÷25	**Bridge Identifier** first 2 byte = Bridge Priority last 6 byte = MAC address of the Bridge that sent the BPDU
26÷27	**Port Identifier** First byte = Port Priority Second Byte = port number
28÷29	**Message Age**
30÷31	**Max Age**
32÷33	**Hello Time**
34÷35	**Forward Delay**

RSTP and STP compatibility

- RSTP bridges may be configured to operate in STP mode
  - Vital if there is a repeater between bridges
  - The RSTP behavior may lead to instability and to temporary loops
- RSTP bridges (IEEE 802.1w) work automatically in STP mode (IEEE 802.1D) when one or more bridges operate in STP mode
  - BPDUs received with the protocol version identifier field set to 0 are handled in a different way
  - Working contemporary in STP mode and RSTP mode may cause instability in the spanning tree definition
    - Possible packets duplication
    - Possible receipt of out of sequence packets

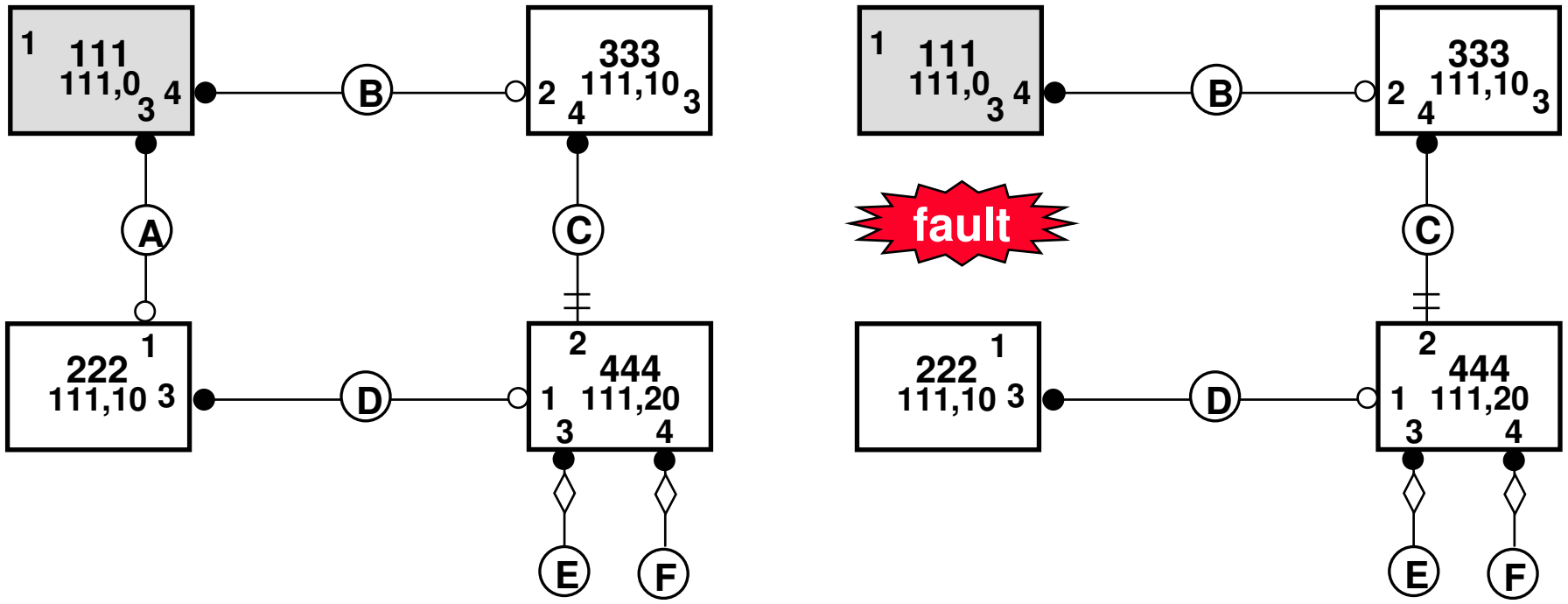
RSTP: fast convergence

- An alternate path may be created in 10 – 20 ms
- Fast convergence relies upon
  - Ability to detect a failure in an reliable way
  - Ability to quickly detect a failure
    - Recovery based physical layer
- For this purpose the transmission systems used are relevant
  - High stability due to reliable hardware parts
    - Intermittent failures may create stability issues
  - Transceiver able to locate a local or remote failure over point-to-point links

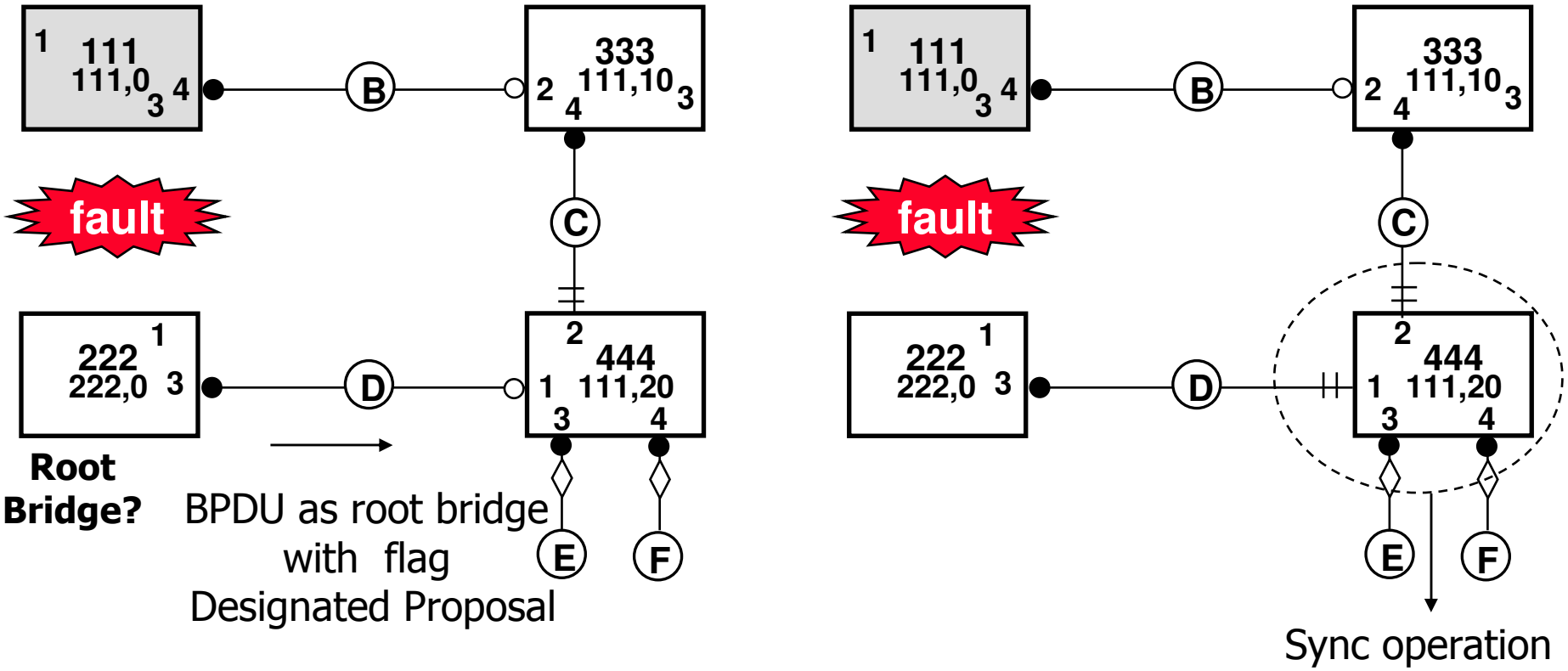
Ring topology and fast recovery

- Fault on the ring
  - If the bridge has an alternate port, its state is changed immediately from alternate to root state
    - Recovery based physical layer
  - If the bridge doesn't have an alternate port proposes itself as Root Bridge
    - Recovery based on new protocol handshaking
      - Fast BPDU Proposal and Agreement type handshaking
- Sync & Agree operations
  - Ports receiving BPDUs with *Designated Proposal flag* execute a *Sync operation*
    - Discard all packet received on non-edge ports
  - Ports discarding packets (Sync) go immediately in Forwarding State when receive a BPDU with *Designated Agreement flag*

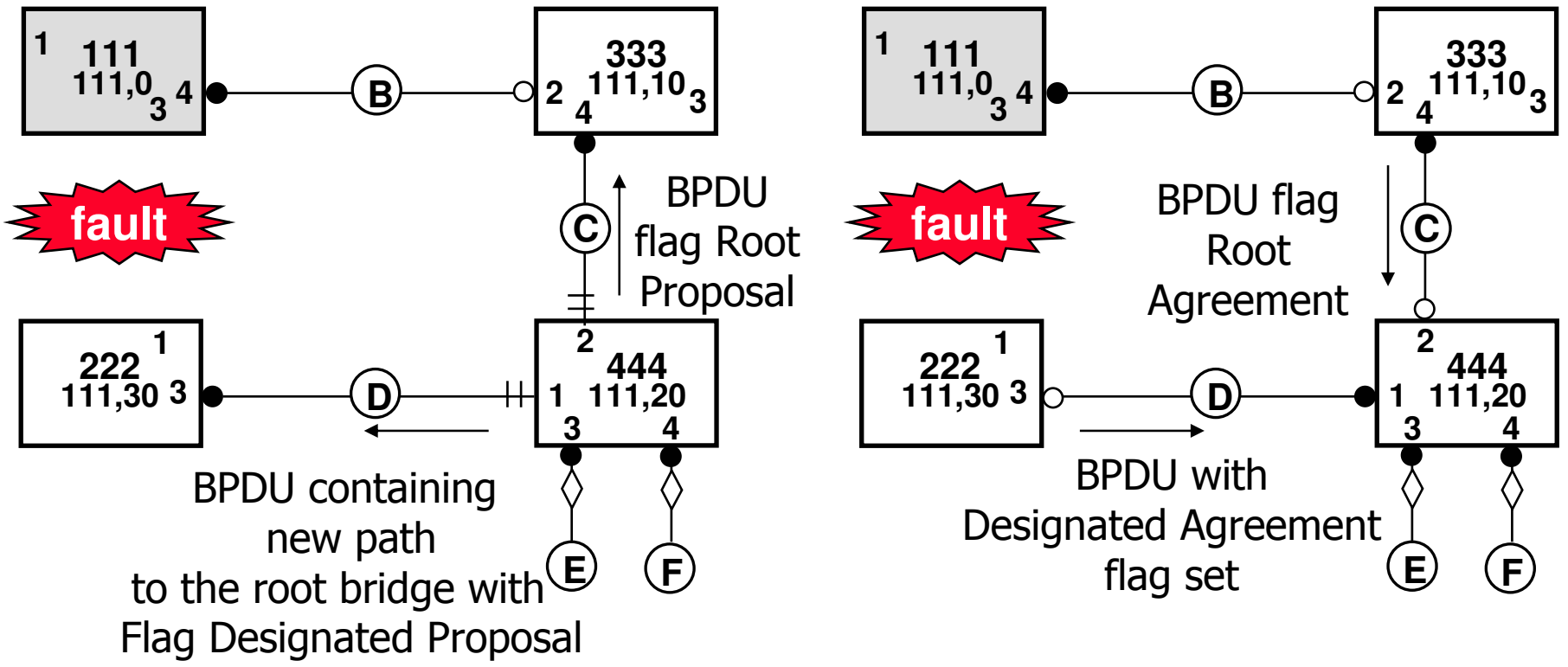
Ring topology fault: step 1



Ring topology fault and new Root proposal



Ring topology fault: Sync Operation



Issues related to fast convergence

- Increased probability of packets duplication
- Increased probability of packets delivered out of order
  - Due to the chance to have
    - Temporary loops
    - Intermittent changes in the spanning tree

Be careful when using protocol relaying on Ethernet!

- LAT, NETBEUI and LLC2
  - They cannot manage out of order and duplicated packets
- It is useful to disable RTSP mode (802.1w)

Compatibilty and interoperability

- Products from different vendors and IEEE 802.w compliant must interoperate without problems
- Be careful with products supporting IEEE 802.1w during and before 2001!!
  - The standard has been approved in late 2001
  - Such products surely support a pre-standard version
    - This may cause interoperability issues with products compliant to the standard