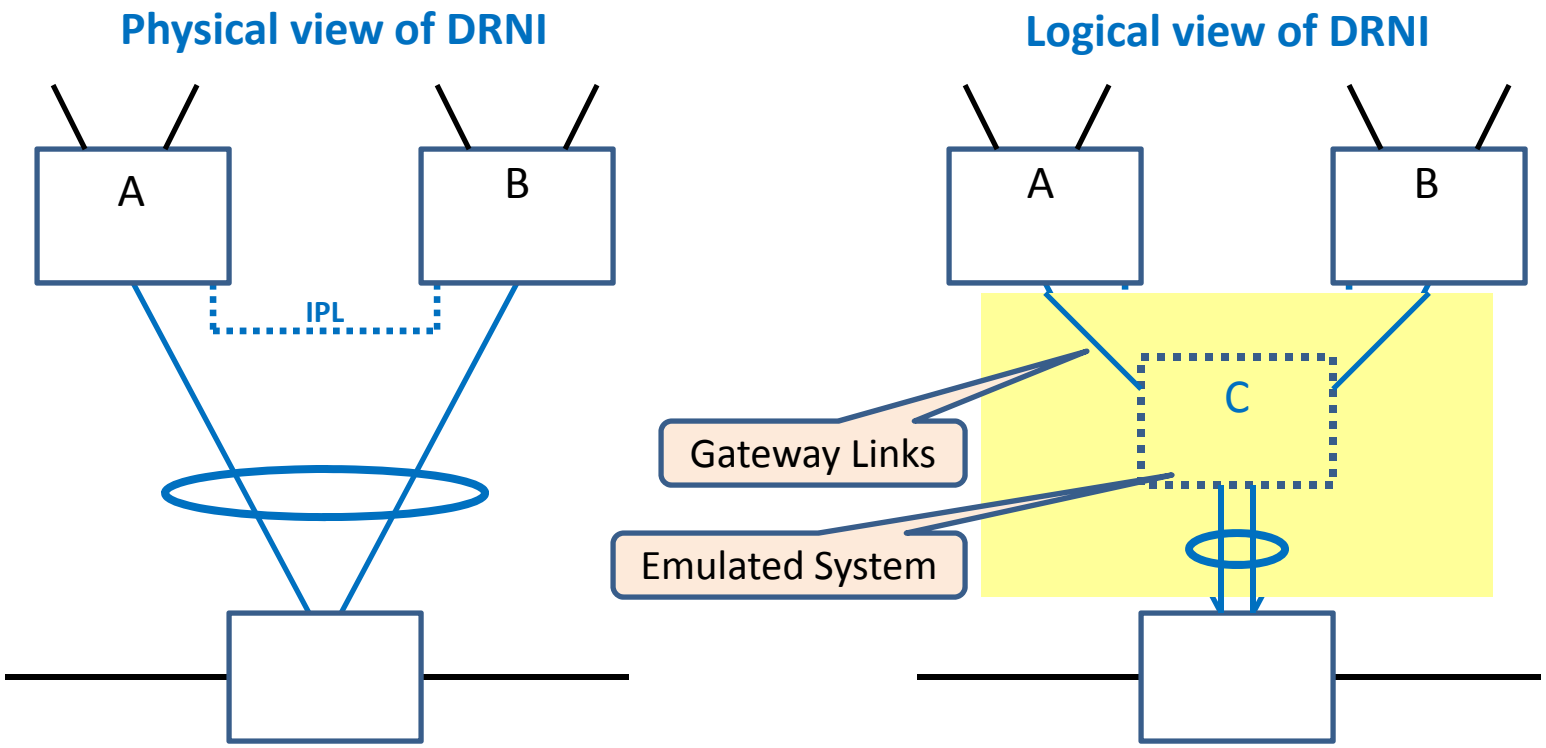
RSTP and DRNI

Version 1

Stephen Haddock

January 17, 2014

What does a DRNI look like to RSTP?



The “right” way to run RSTP across a DRNI is to have an instance of RSTP that runs in the Emulated System and exchanges BPDUs across the Gateway Links with the Portal Systems and across the LAG with the LAG Partner.

DRNI and RSTP

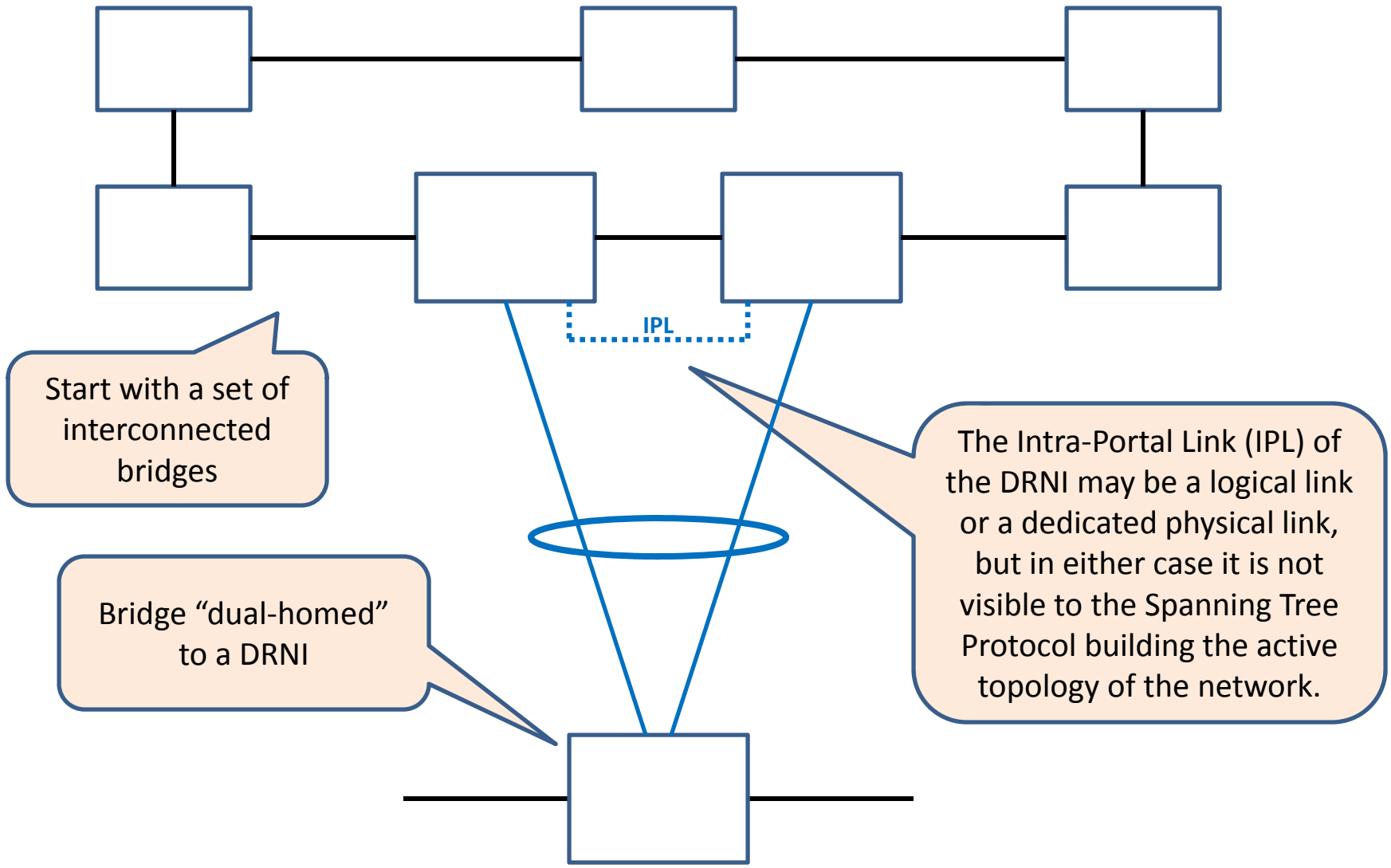
- In practice, running an instance of RSTP for the Emulated System is not an ideal solution.
  - The implementation of the Emulated System would need to be distributed between the Portal Systems, which is not trivial.
  - The Gateway Links are logical links within a single Portal System. Exchanging BPDUs across them, and maintaining the RSTP state machines for the logical ports at each end of the logical link, is exceptionally inefficient.
- It is possible to create a more optimal solution that provides the same behavior.
  - To get there we need to understand the behavior we are looking for by understanding how RSTP would work if we did run an instance for the Emulated System.

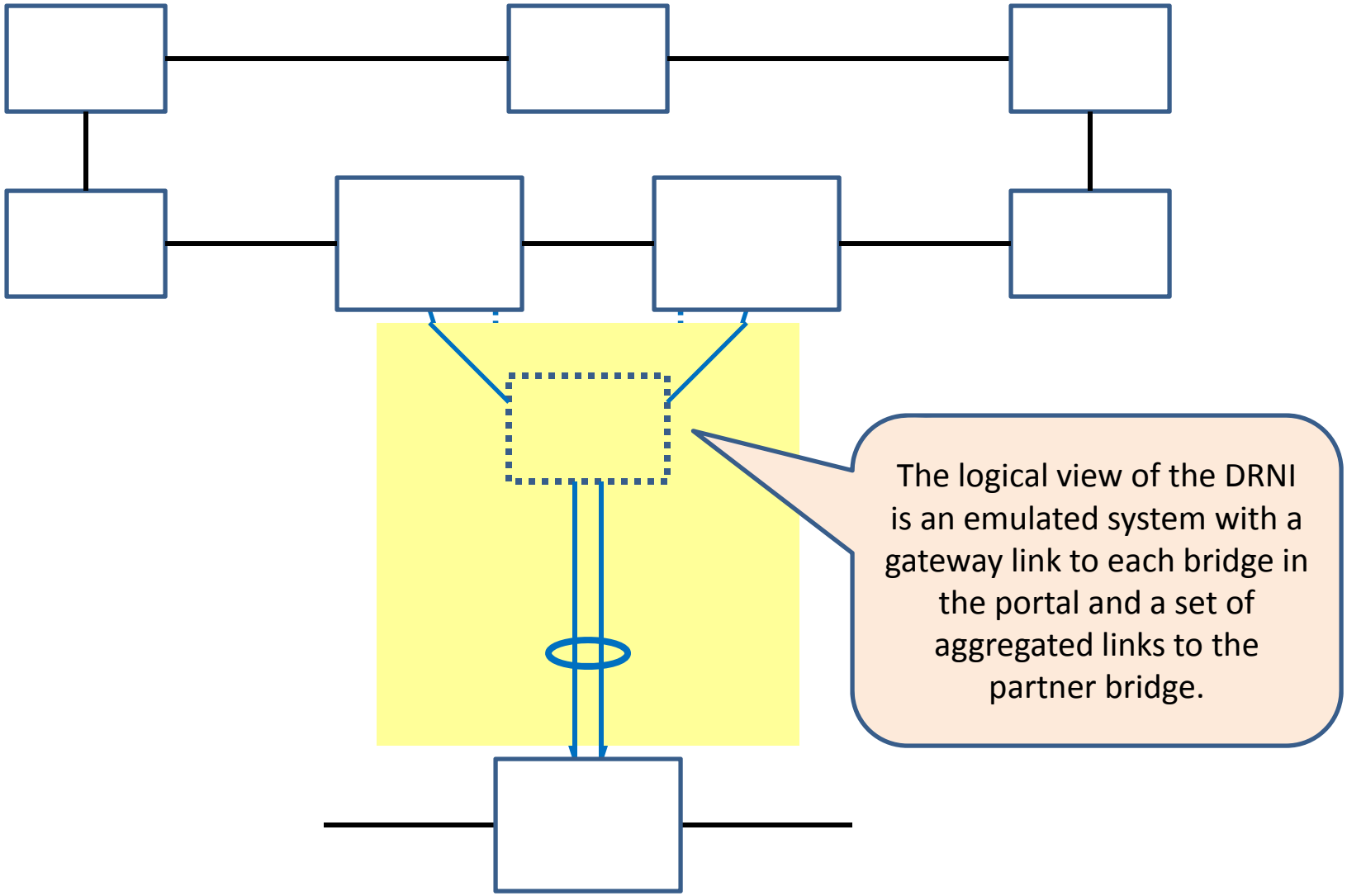
DRNI and RSTP

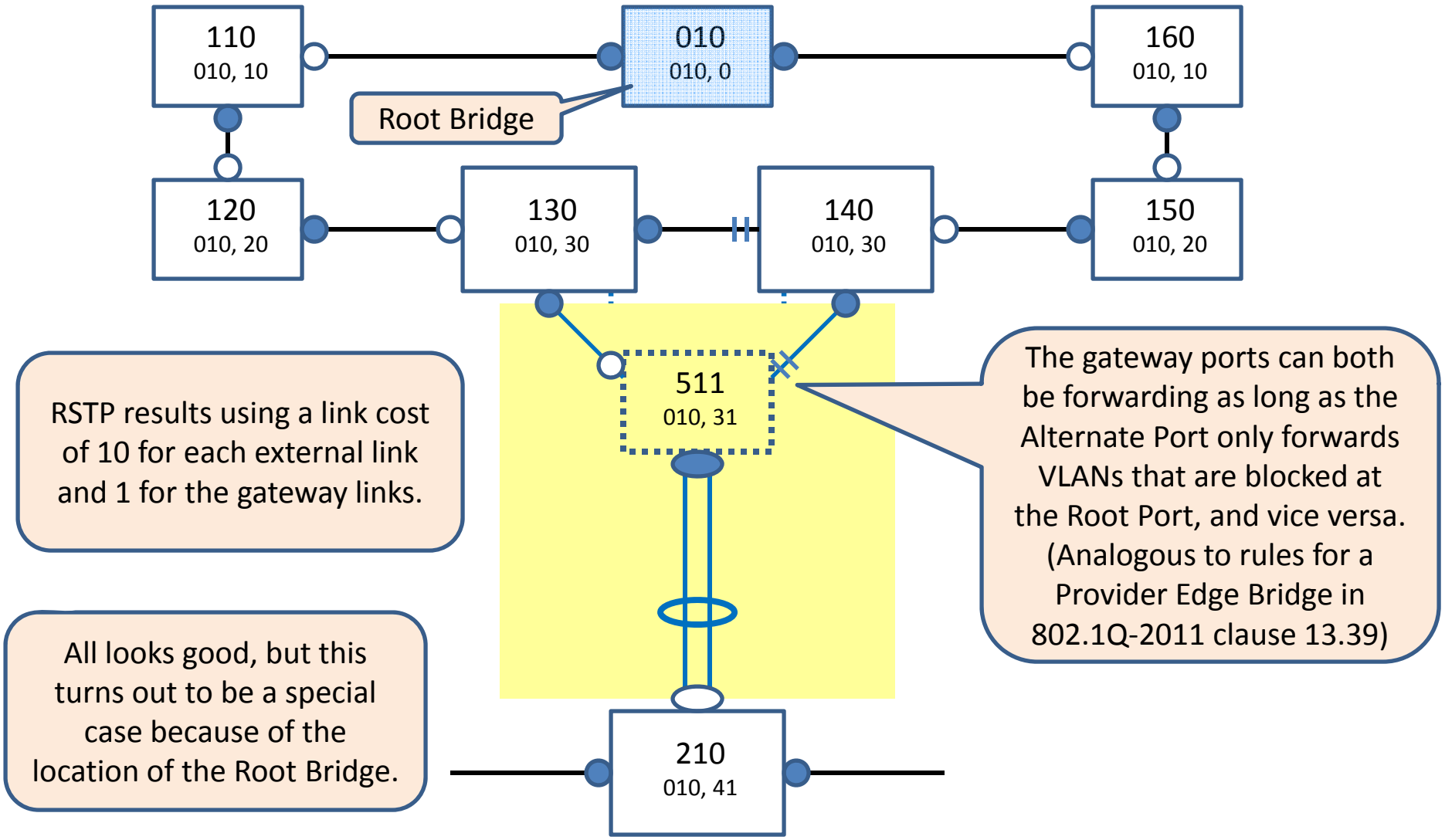
- If we were to run an instance of RSTP on the Emulated System, we immediately encounter a problem:
  - How do you prevent the Gateway Links from being incorporated into the active topology, thus forming a “transit path” in the network?
- The solution to this is to use the Gateway Links as “alternate paths”.

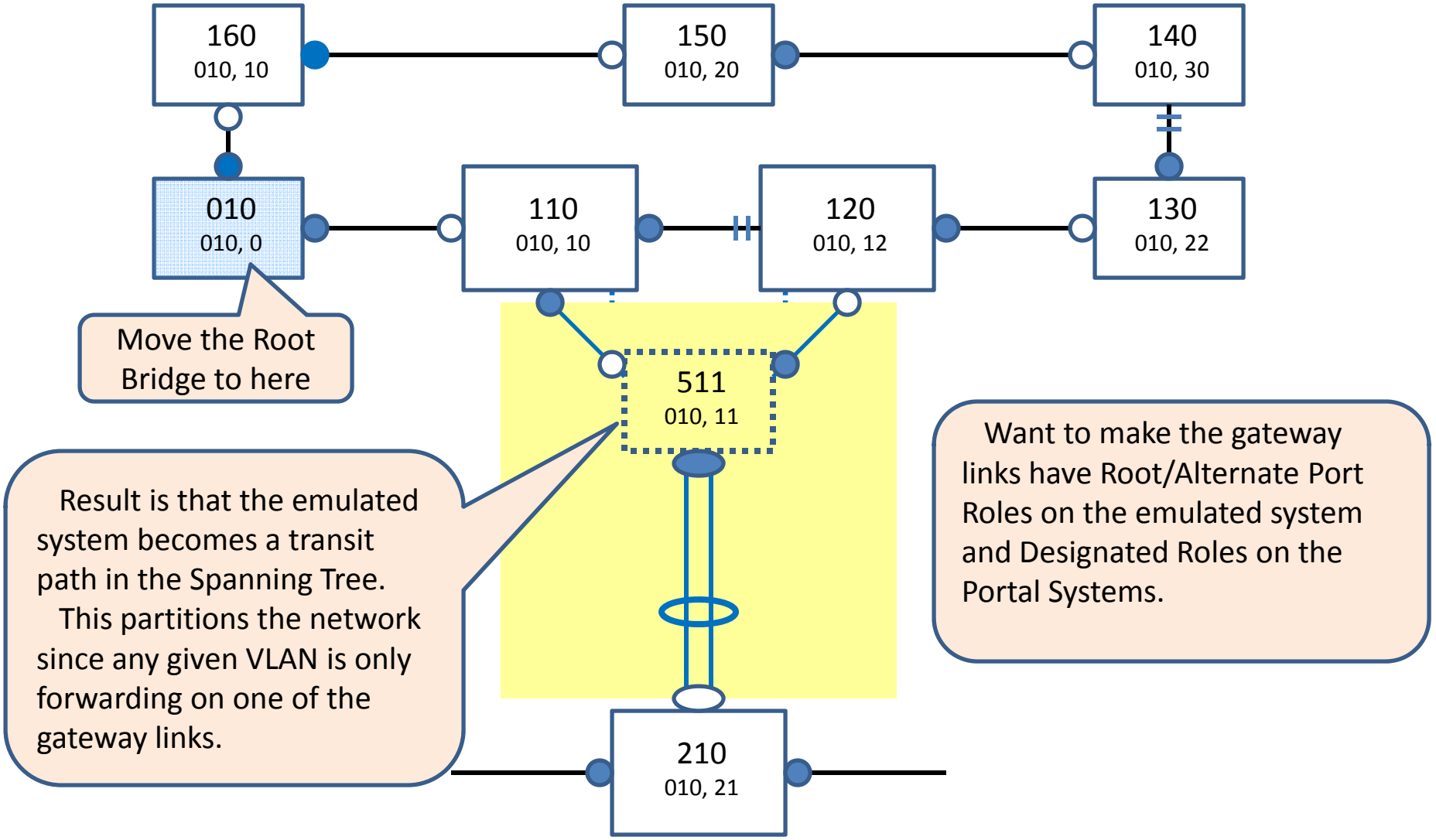
  Step 1: Force one of the Gateway Links to have a port with a role of Root, and force the other Gateway Link(s) to have a port with a role of Alternate.

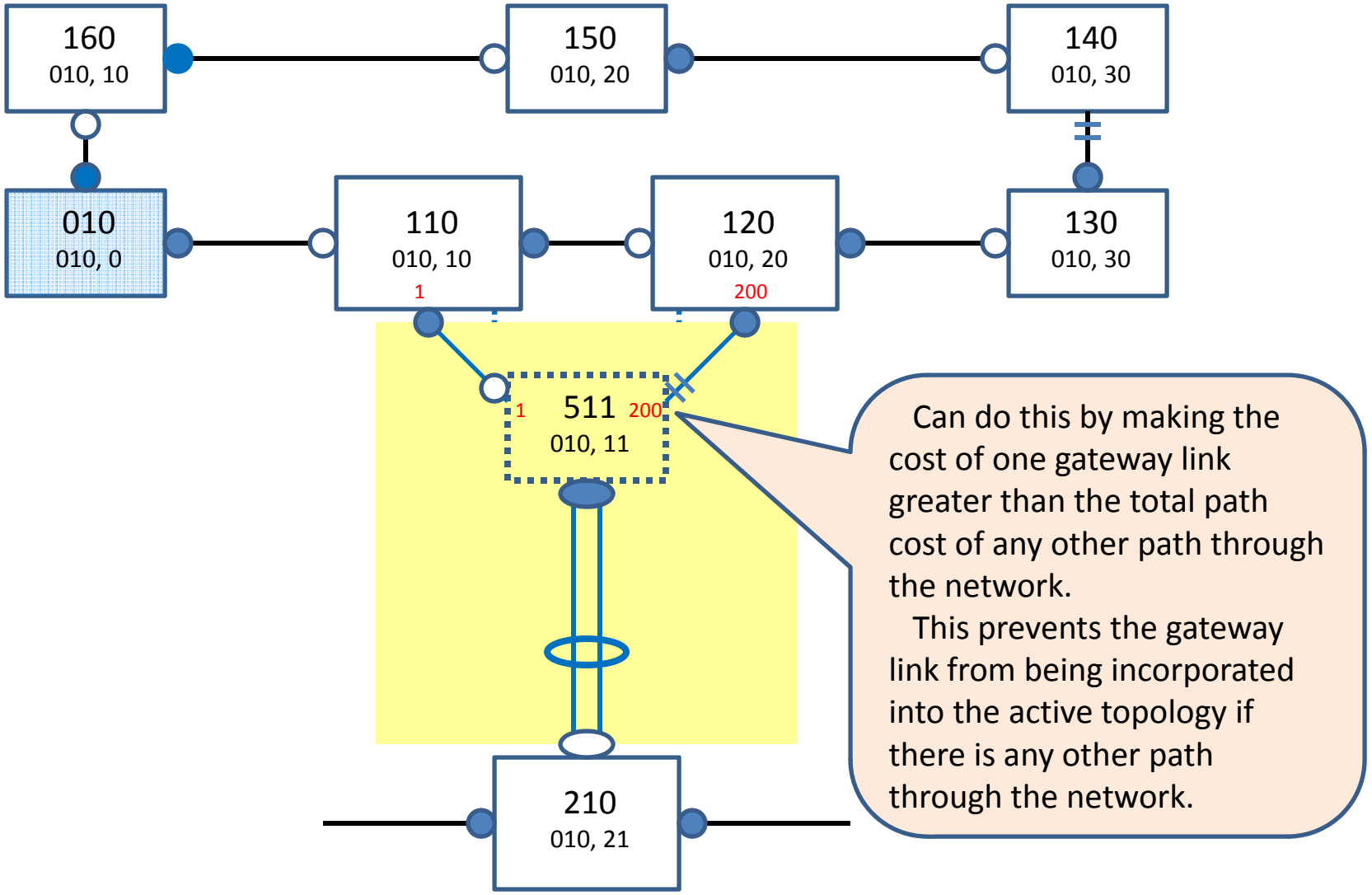
  Step 2: Allow the Gateway Link with the Alternate Port to be Forwarding for VIDs that are not Forwarding on the Gateway Link with the Root Port.
    - This is analogous to, but an extension of, the concept that allows Alternate Ports to be Forwarding for selected VLANs in a Provider Edge Bridge (802.1Q-2011 subclause 13.39).

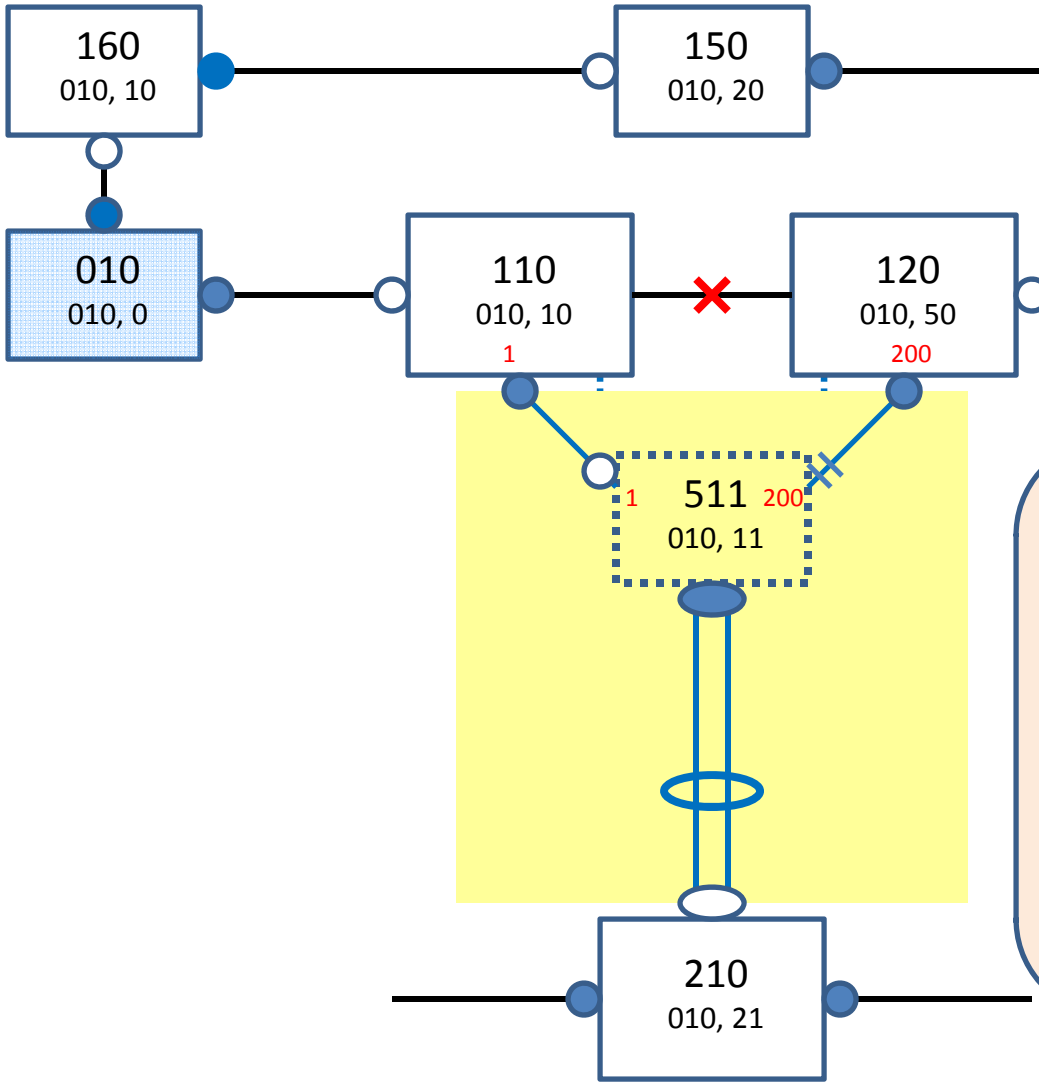

IPL
Start with a set of interconnected bridges
Bridge “dual-homed” to a DRNI
The Intra-Portal Link (IPL) of the DRNI may be a logical link or a dedicated physical link, but in either case it is not visible to the Spanning Tree Protocol building the active topology of the network.


The logical view of the DRNI is an emulated system with a gateway link to each bridge in the portal and a set of aggregated links to the partner bridge.

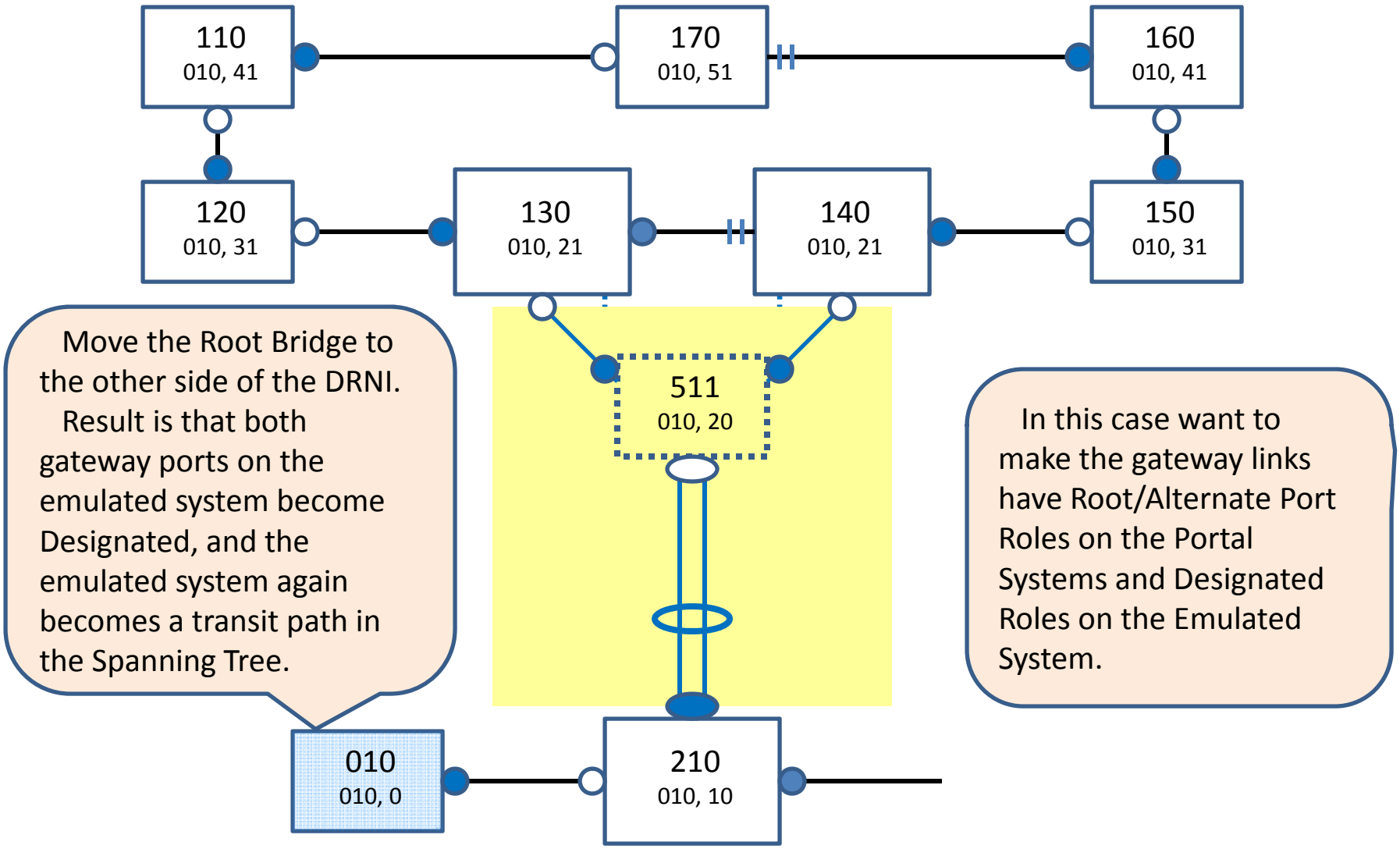

110
010, 10
010
010, 0
160
010, 10
Root Bridge
120
010, 20
130
010, 30
140
010, 30
150
010, 20
511
010, 31
210
010, 41
RSTP results using a link cost of 10 for each external link and 1 for the gateway links.
The gateway ports can both be forwarding as long as the Alternate Port only forwards VLANs that are blocked at the Root Port, and vice versa. (Analogous to rules for a Provider Edge Bridge in 802.1Q-2011 clause 13.39)
All looks good, but this turns out to be a special case because of the location of the Root Bridge.

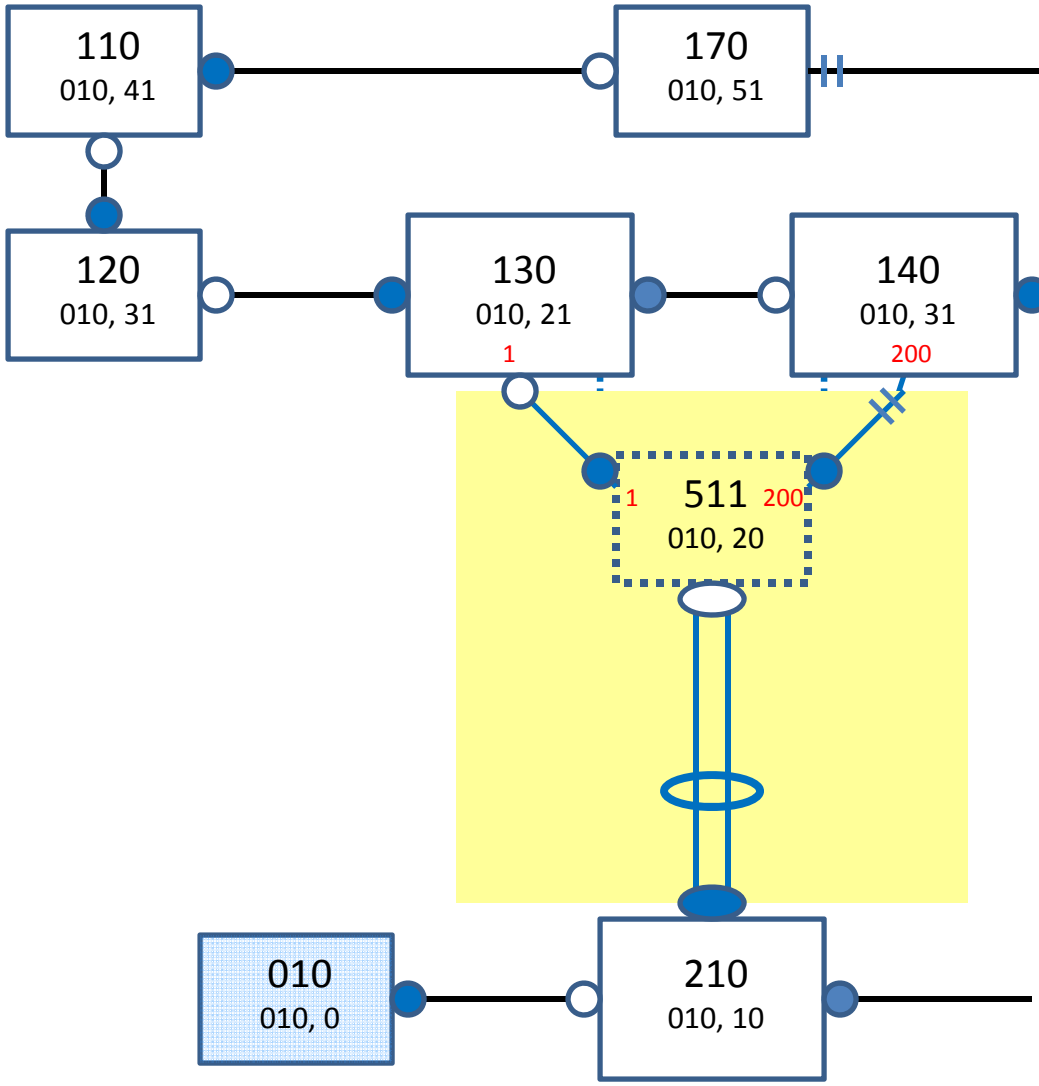

160
010, 10
150
010, 20
140
010, 30
010
010, 0
110
010, 10
120
010, 12
130
010, 22
Move the Root Bridge to here
511
010, 11
Result is that the emulated system becomes a transit path in the Spanning Tree.
This partitions the network since any given VLAN is only forwarding on one of the gateway links.
Want to make the gateway links have Root/Alternate Port Roles on the emulated system and Designated Roles on the Portal Systems.
210
010, 21


160
010, 10
150
010, 20
140
010, 30
010
010, 0
110
010, 10
1
120
010, 20
200
130
010, 30
1
511
200
010, 11
210
010, 21
Can do this by making the cost of one gateway link greater than the total path cost of any other path through the network.
This prevents the gateway link from being incorporated into the active topology if there is any other path through the network.



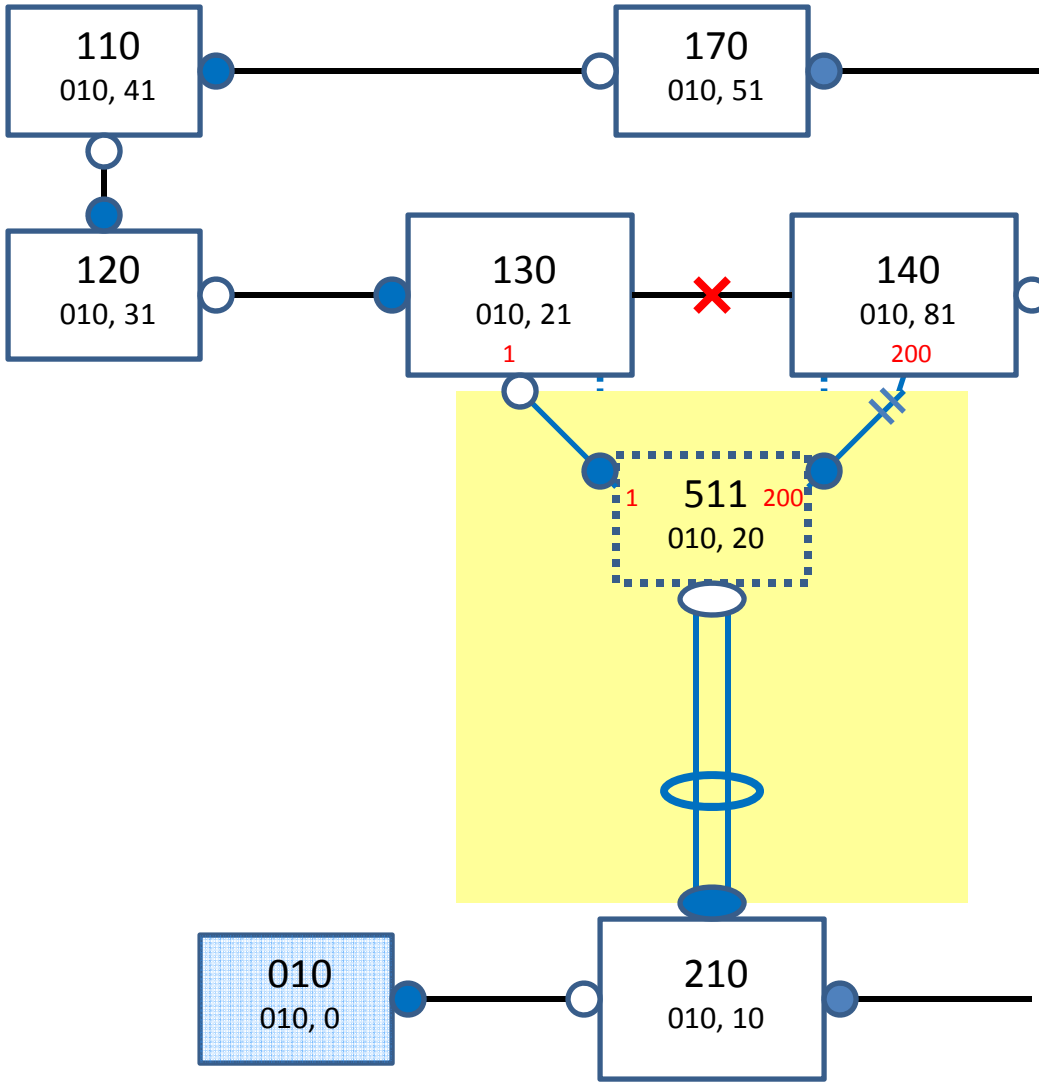
If the active path between the Portal Systems breaks, the network will reconfigure to a new active topology still avoiding the gateway link. (There may be a transient that incorporates the gateway link, but there will be incomplete connectivity in the network during the transient anyway.)


110
010, 41
170
010, 51
160
010, 41
120
010, 31
130
010, 21
140
010, 21
150
010, 31
511
010, 20
010
010, 0
210
010, 10
Move the Root Bridge to the other side of the DRNI.
Result is that both gateway ports on the emulated system become Designated, and the emulated system again becomes a transit path in the Spanning Tree.
In this case want to make the gateway links have Root/Alternate Port Roles on the Portal Systems and Designated Roles on the Emulated System.



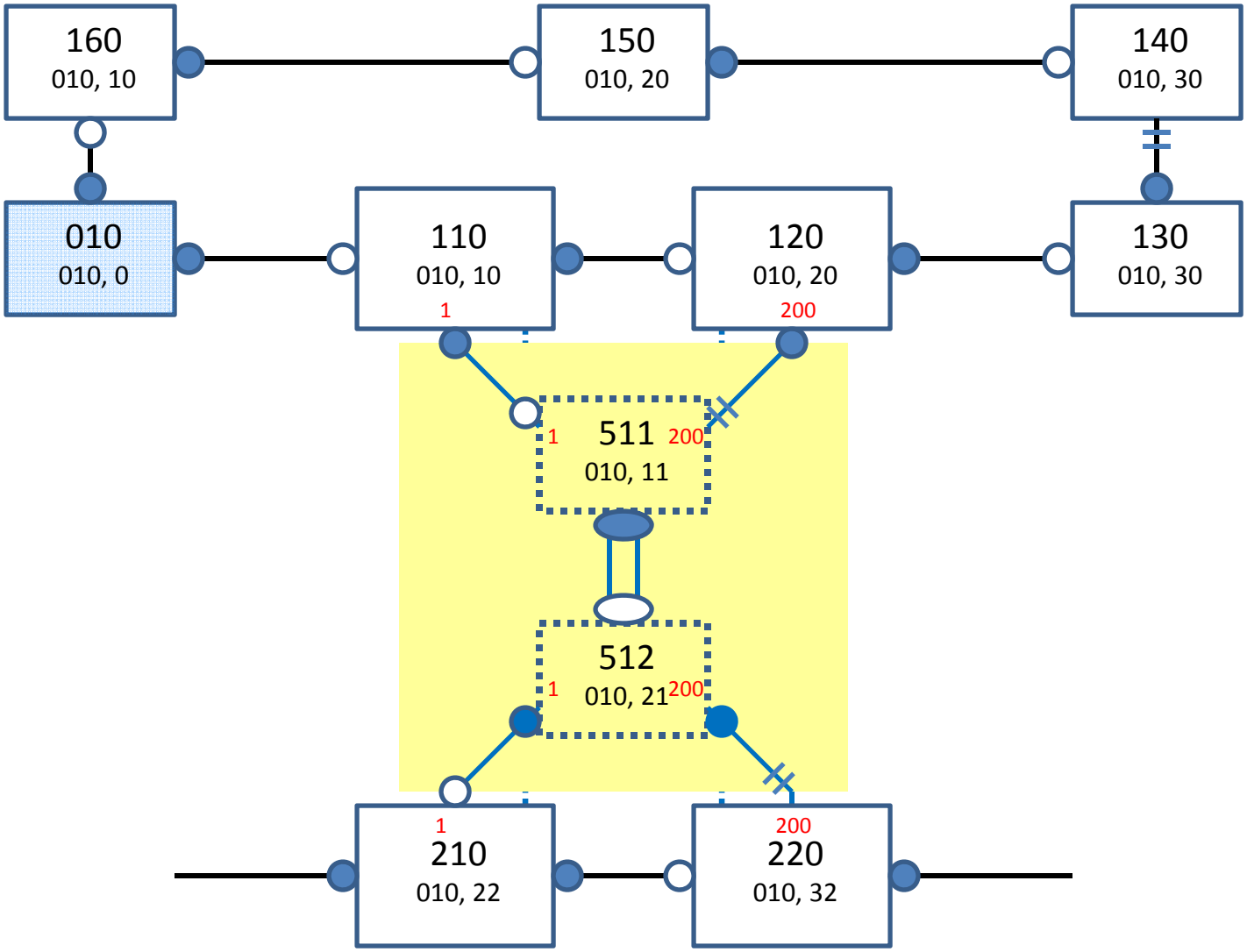
The same assignment of gateway link costs resolve this situation.

Once again, both gateway links can be forwarding as long as any given VLAN is forwarded only on the Alternate and blocked on the Root, or vice versa.

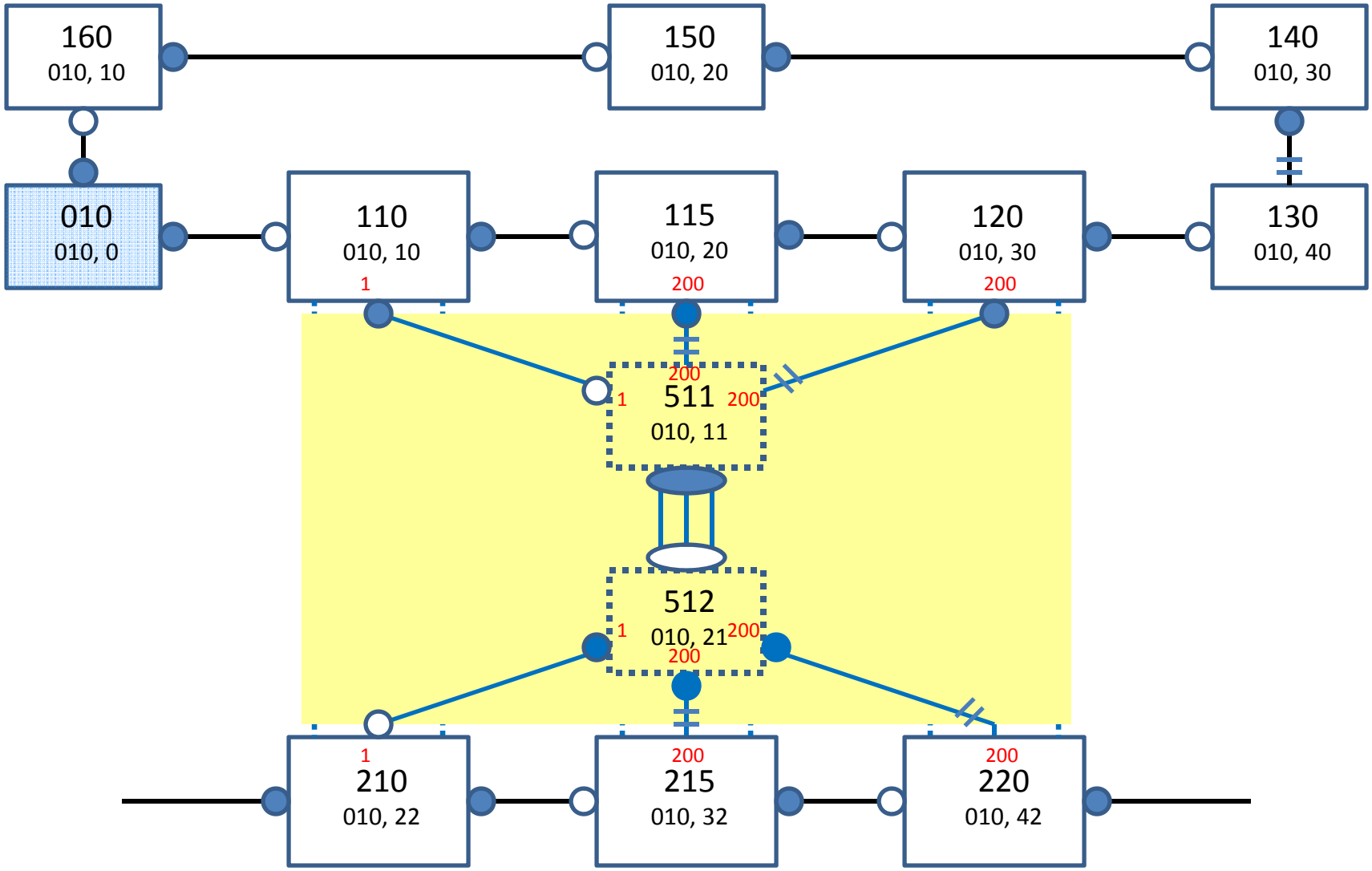


If the path between the Portal Systems breaks then the gateway port on 140 temporarily becomes Root, but ultimately this port becomes Alternate again. There is incomplete connectivity during this transient.

Example with DRNI on both sides of the interface.



Example with three Portal Systems per Portal



Optimizing DRNI RSTP Processing

1. Select one Gateway Link to be visible to RSTP
2. Hide all other Gateway Links
   - The examples above use Path Cost to effectively prevent some Gateway Links from being integrated into the active topology.
   - The same result can be achieved by not running RSTP over these links. (RSTP does not “see” the bridge ports supported by these links.)
3. Make the Emulated System transparent to RSTP
   - When all Gateway Links except one are hidden, the Emulated System looks like a two port device.
   - A two port device that does not do learning can be transparent to RSTP (does not run RSTP and simply relays received BPDUs).
   - Then the BPDUs generated/processed over the one visible Gateway Link are the same as the BPDUs transmitted/received across the DRNI.
4. When the visible Gateway Link is forwarding, the other Gateway Links can be used as alternate paths.

Result

- RSTP sees the DRNI as a single Bridge Port in one of the Portal Systems in each Portal
  - This Bridge Port (call it the Root Gateway) operates RSTP state machines generates/processes BPDUs exchanged over the DRNI.
    - These BPDUs may be transmitted/received over an Aggregation Link attached to this Portal System, or carried over the IPL to an Aggregation Link attached to another Portal System.
  - The other Portal Systems in each Portal do not do any RSTP processing on behalf of the DRNI.
    - These Portal Systems provide the Alternate Gateways.
    - When the Root Gateway RSTP state is Forwarding, the Alternate Gateways are Forwarding.
    - The VIDs that are transmitted/received at the Root vs Alternate Gateways are controlled by the Gateway Conversation ID to Gateway mapping . This is independent of RSTP operation.

Issues

- Selecting which Portal System provides the Root Gateway.
- Communicating when the Root Gateway RSTP state is Forwarding.
  - When the Root Gateway transitions out of Forwarding, do we need acknowledgement that the Alternate Gateways are no longer Forwarding?
- How do the Portal Systems providing Alternate Gateways know when to flush learned addresses?
- Does this solution extend to MSTP? SPB?
  - I think the answer is yes, (at least if all Portal Systems of a given Portal are in the same Region and maybe even if not).

Selecting the Root Gateway

Option 1: The Portal System with the gateway that handles untagged frames is the Portal System for the Root Gateway.

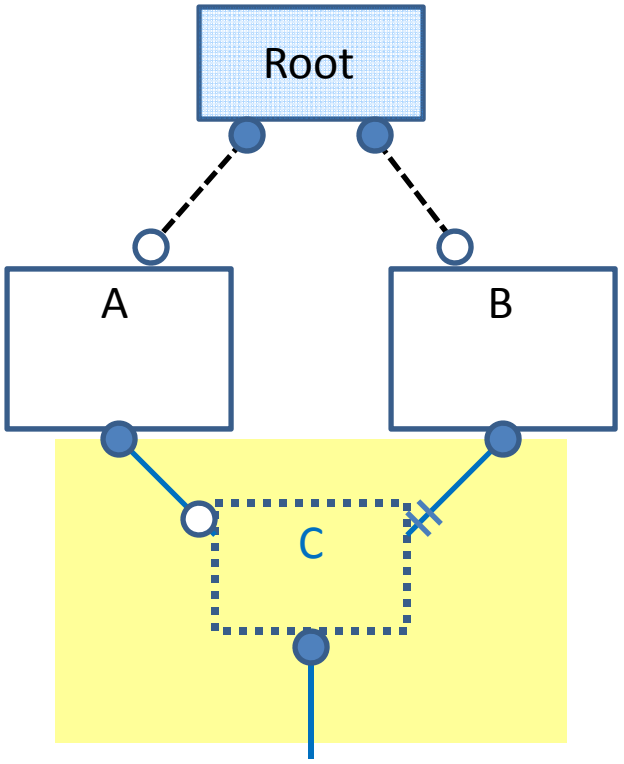
- **Pro:** No additional configuration required.
- **Pro:** The Distributed Relay treats BPDUs just like any other untagged frame.
- **Con:** Remapping the Gateway Conversation ID for untagged frames to a different Gateway causes a RSTP topology change.

Option 2: Each Portal System has a unique number 1,2,or 3. Decree that the Root Gateway is in Portal System 1.

- **Pro:** No additional configuration required.
- **Con:** The Distributed Relay may need to send BPDUs to a different gateway than other untagged frames.
- **Pro:** Remapping the Gateway Conversation ID for untagged frames to a different Gateway does not cause a RSTP topology change.

Backup Slides

RSTP and Loop-Free Alternate Paths



- The Alternate Port on System C forms a Loop-Free Alternate Path.
- If the link with the Root Port on System C were to break, RSTP would re-assign the Alternate Port to Root and immediately make it Forwarding.
  - This is a local event, meaning it does not require any BPDU exchanges.
- If System C “breaks” the Root Port path for some VLANs, those VLANs can be forwarded on the Alternate Port path without any risk of forming loops.
  - Again this is a local event. No other system needs to be involved in the decision or even know System C has made this decision.