Distributed Aggregation Sub-layer & Logical component per DRNI

Mandar Joshi
Fujitsu Network Communications

Summary

- This slide deck assumes
  - Assignment of Services to Gateway (a node in the DRNI portal) is configuration driven.
  - Service connectivity from Gateway to DRNI, Link Selection and DRNI failure handling is entirely enforced at the Distributed Aggregation Sub-layer
  - Virtualization/'Logicalization' of intra-DAS link
  - Logical entity per DRNI (presented by Steve H in earlier slide – these slides take the idea to expand on the thoughts in this proposal)
- This slide deck contains thoughts/ideas on
  - Service Identification, Gateway mapping and Link Selection
  - Single DRNI and Hair-pinning
  - Multiple DRNI
  - Single DRNI with un-protected TESI in the connected network
  - Single DRNI with protected TESI in the connected network
- There are several scenarios that these slides does not cover. The attempt is to convey some high-level ideas

Distributed Aggregation Sub-layer (1)

This proposal assumes that **Service Connectivity from Gateway to DRNI, Link Selection and DRNI failure handling** is entirely handled at the D-Agg. Sub-layer

Some advantages and disadvantages are introduced below.

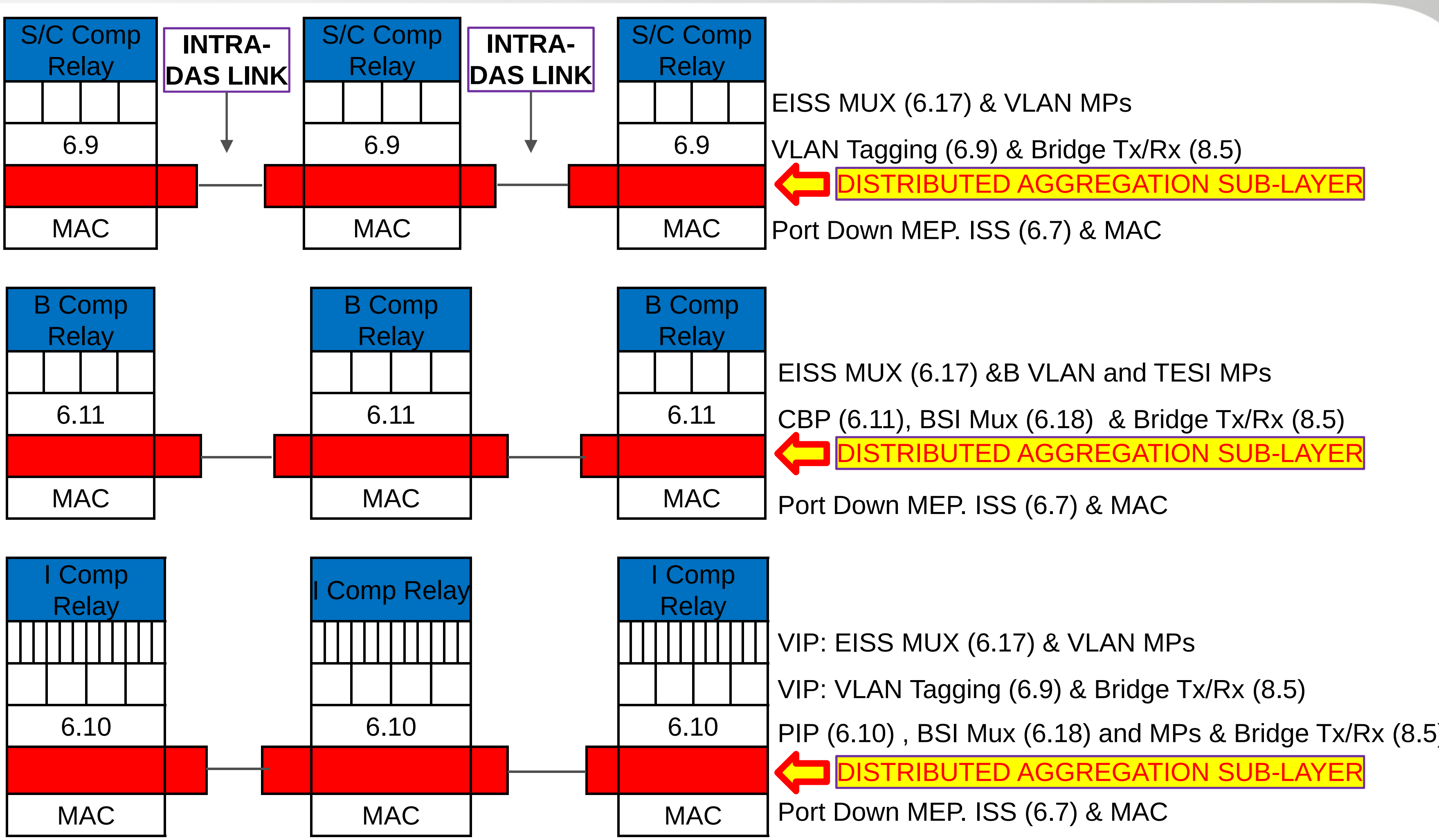
ADVANTAGES

- Preserves upper layer logic – no changes required to the upper layers (For instance RCSI or B-Com relay logic should not have to change because it runs over DRNI)
- A central decision making entity for all DRNI related decisions makes co-ordination between different layers unnecessary (for instance co-ordination of Gateway selection at Relay layer and Link Selection at lower layer)
- The merging should make configuration much easier and less error prone

DISADVANTAGES

- There is layering violation. But layering is already violated in legacy LAG – it would not make it any worse.
- (More likely to come out of these discussions)

Distributed Aggregation Sub-layer (2)
Position in stack



The change to show Distributed Agg-sublayer from Steve H's slide 23

(Ref: http://ieee802.org/1/files/public/docs2011/axbq-haddock-multiple-drni-support-1011-v01.pdf)

Service Identification and Gateway mapping (1)

- Gateway mapping – A mapping of services to Gateways exists both peering networks connected via a DRNI.
- Gateway mapping is likely to be <u>driven by configuration</u> <u>based on policies/criteria</u>
- A frame is classified to belong to a specific service based on header fields AND/OR the configured criteria
- Based on the service to Gateway mapping, the frame is identified as belonging to a specific Gateway
- So far, based on my understanding, the Service Identifier of a frame is SVID or BVID or ISID
- The assumption here is that services will be load-balanced based on these service identifiers
- Although this might be ok when load-balancing based on ISIDs, it seems somewhat restrictive for SVIDs and BVIDs
- Further, it seems to exclude other ways of mapping services to Gateways – for example mapping specific TESIs to specific Gateways (This maybe possible in the models described so far, but it is not clear to me how this can be done)

Service Identification and Gateway mapping (2)

- It might help for a Service to be identified as a combination of different header fields and associated criteria. Some examples:
  - <SVID> - Criteria could be Gateway A for all Odd and Gateway B for all Even
  - Each ESP within a TESI, identified by <ESP-DA, ESP-SA, ESP-VID>, may want to be assigned to different gateway
  - <BVID, ISID> - Different ISIDs (services) within the same BVID may be assigned to different gateways
  - <DST-MAC, SVID> - Frames to a certain DST-MAC for an SVID may be assigned to different gateways
  - Etc. etc.
- This leaves door open for a flexible and extensible Gateway selection algorithm based on the needs of the peering networks – <u>HOWEVER THIS PRESENTATION DOES NOT PROPOSE A SOLUTION – IT MERELY IDENTIFIES, WHAT MIGHT BE AN IMPORTANT CONSIDERATION!</u>
- **For backward compatibility purposes the Gateway mapping functionality could be optional** (Link Selection functionality subsumes Gateway selection functionality in such cases. Described in next slide)

Link Selection

- Link Selection – Link Selection is the stage after the frame has been accepted to be forwarded by this gateway
- Link selection could be driven by a configured criteria
- This criteria could be (as in legacy LAG implementation)
  - Some hashing mechanism (a standard 5 tuple hash, hash that includes MPLS labels etc.)
  - Static configuration mapping specific flows to ports
  - Etc.
- Note that at the link selection stage, this proposal requires that flows be mapped to ports that <u>are local to this node in the DRNI portal</u>
- **Backward compatibility** is required when a node that does not support DRNI connects to different nodes in the peering network that supports DRNI. The following options exist:
  - A coordinated hash algorithm between all the nodes in a DRNI portal ➔ implies that flows could traverse the intra-DAS in normal operating conditions
  - Completely independent hashing schemes in the two networks. Depending on the technology used in the peering network (Bridging vs. P2P), there will more traffic on the intra-DAS links than usual

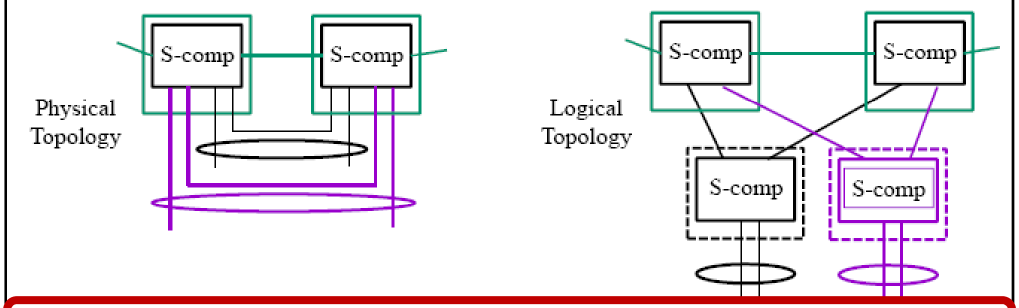
Service ID semantics in the DRNI

- Service ID semantics in the peering networks will likely differ
  - Peering networks may use different SVIDs, BVIDs or ISIDs for the same service
  - Peering networks may also use different technologies in each of their networks. For instance one network may use Bridging and the peering network may use a point-to-point technology (such as PBB-TE or G.8031)
- Propose **Service ID Normalization in the DRNI network**
  - Before forwarding a frame TO the DRNI, the connected network is responsible to translate the Service ID to the semantics of the service in the DRNI network
  - The service semantics in the DRNI are known to each of the peering networks
  - A major advantage of this is that the peering networks do not need to know the semantics of the service in each of their networks. In fact, this would be a major requirement for all carriers
- This proposal assumes the same encapsulation in the DRNI network as in the attached network after normalization

Logical Component per DRNI

- These slides make use of Steve H's Option 1 of creating a logical entity per DRNI
- Argument is that this is a better model to
  - Describe, visualize and understand DRNI functionality
  - Makes it easy to understand the traffic flow
- This model combined with the distributed aggregation layer as the gateway connectivity AND link selection entity, provides for an extensible and cleaner model (Remains to be seen! ☺

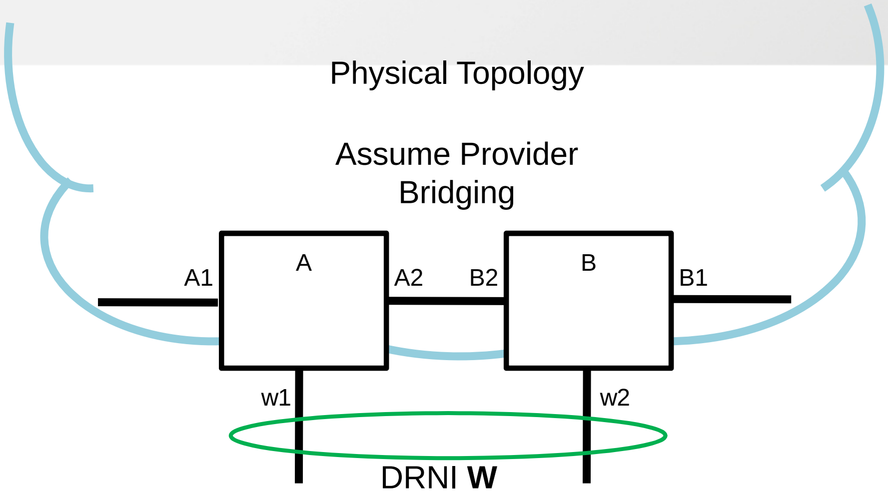
Possible Models of Multiple DRNIs

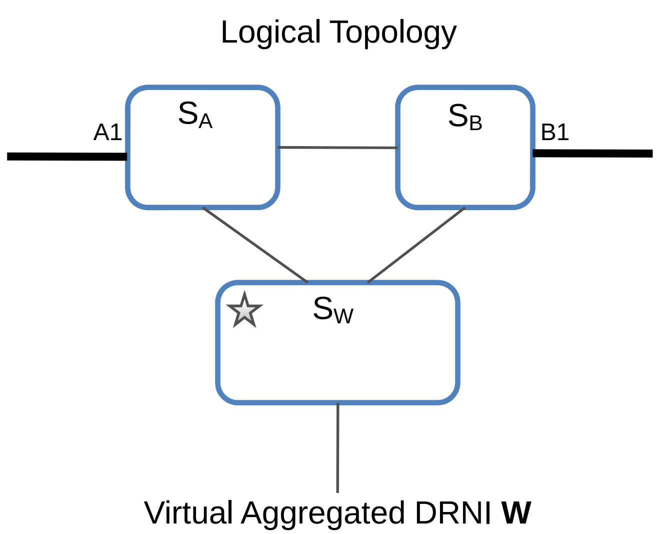


- Option 1: Create logical component per DRNI
  - Advantages:
    - Keeps all DRNI aggregation layers separate.
  - Disadvantages:
    - Requires Intra-DAS Link per DRNI.
      - Separate physical dedicated Intra-DAS Links, or use encapsulation to create separate virtual Intra-DAS Links on a single physical link.
    - Requires separate instances of 'Distributed Higher Layers' per DRNI.

Ref: http://ieee802.org/1/files/public/docs2011/axbq-haddock-multiple-drni-support-1011-v01.pdf

Single DRNI – Simple Example (1)



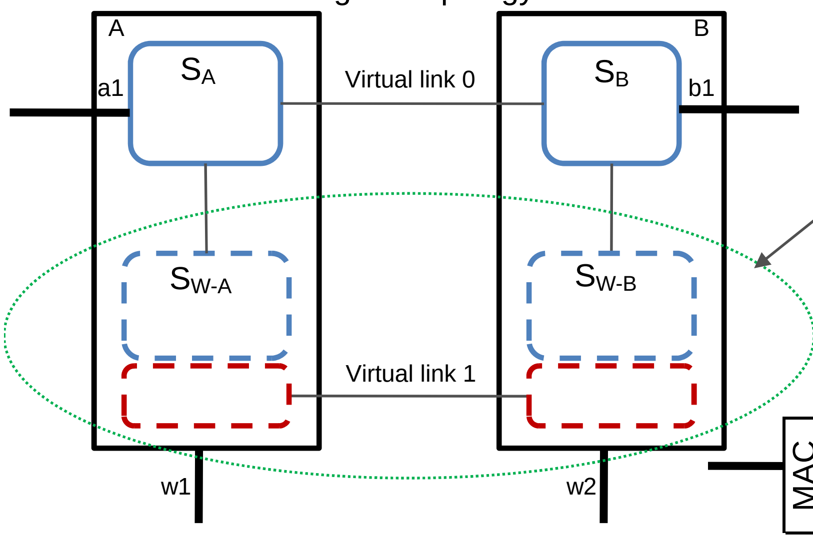


- Assume S-VLAN based service
- Assume physical link between A & B has a dual function – it is part of the provider network and also carries intra-DAS link traffic
- Logical topology is shown below. S-components S_A and S_B are un-aware of the underlying distribution
- The data link between S_A and S_B is a logical connection over the physical link between A and B (link A2-B2).
- S_A and S_B connect through an internal logical interface to the Distributed S-component for the DRNI – S_W
- The Intra-DAS link (NOT SHOWN) logically connects the distributed components of S_W
- This requires a special encapsulation for each such logical connection that uniquely identifies frames
- Note that the DRNI appears as a single link to S_W

Single DRNI – Simple Example (2)

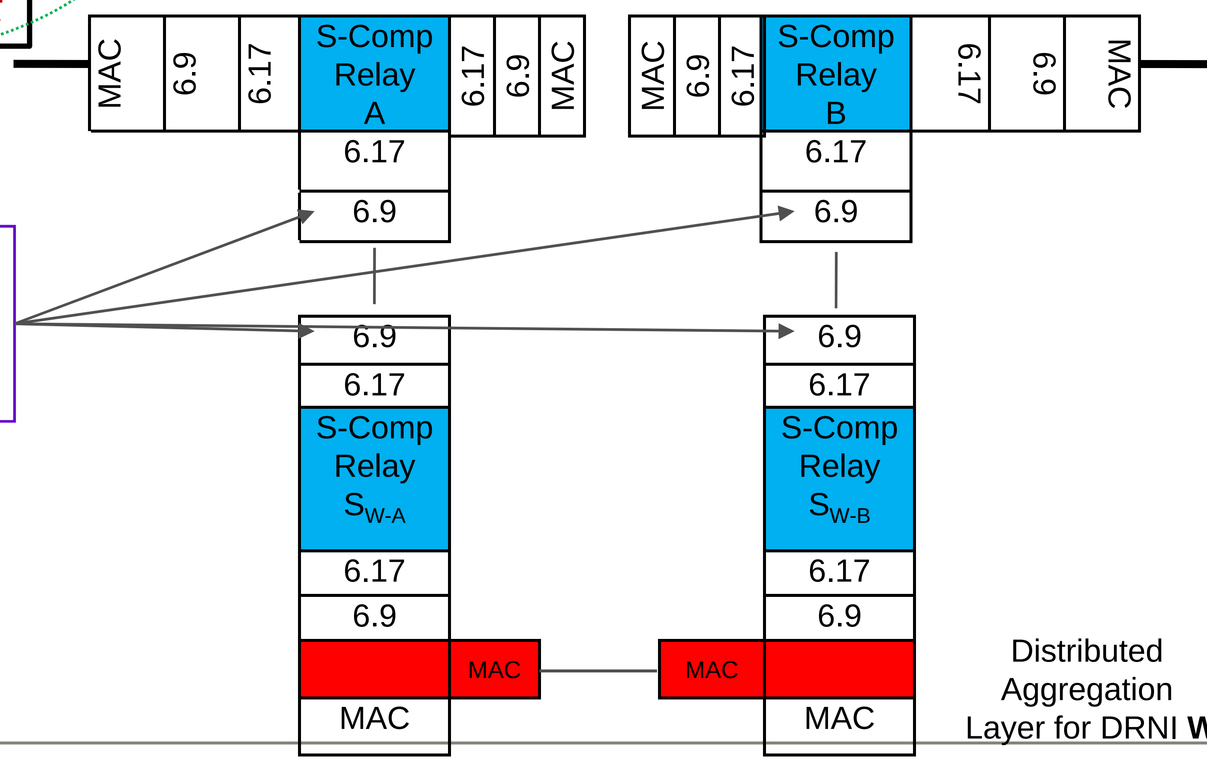
FUJITSU

Logical Topology



- Distributed S-Component is achieved via Distributed Aggregation Layer for DRNI **W**

Bridge model with the 802.1Q layering

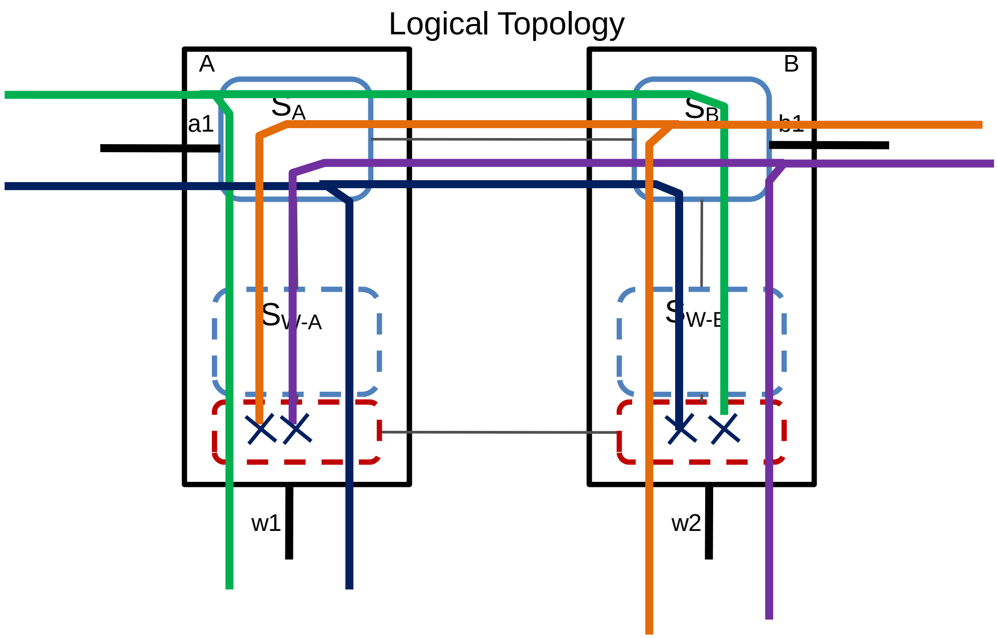


Service ID Normalization

VID Translation to/from DRNI SVID semantics

Distributed Aggregation Layer for DRNI **W**

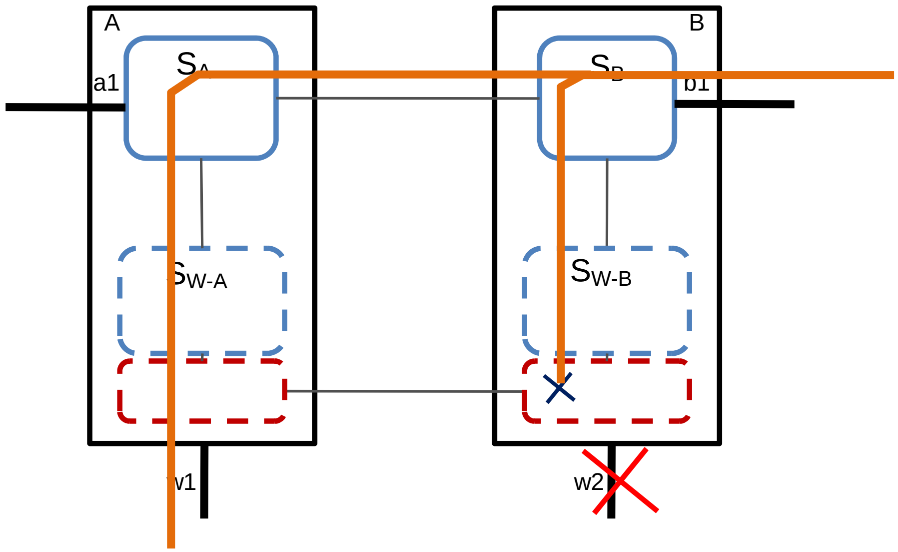
Single DRNI – Simple Example (3)



- Consider the following example:
- Green VLAN: Gateway A
- Blue VLAN: Gateway A
- Orange VLAN: Gateway B
- Purple VLAN: Gateway B

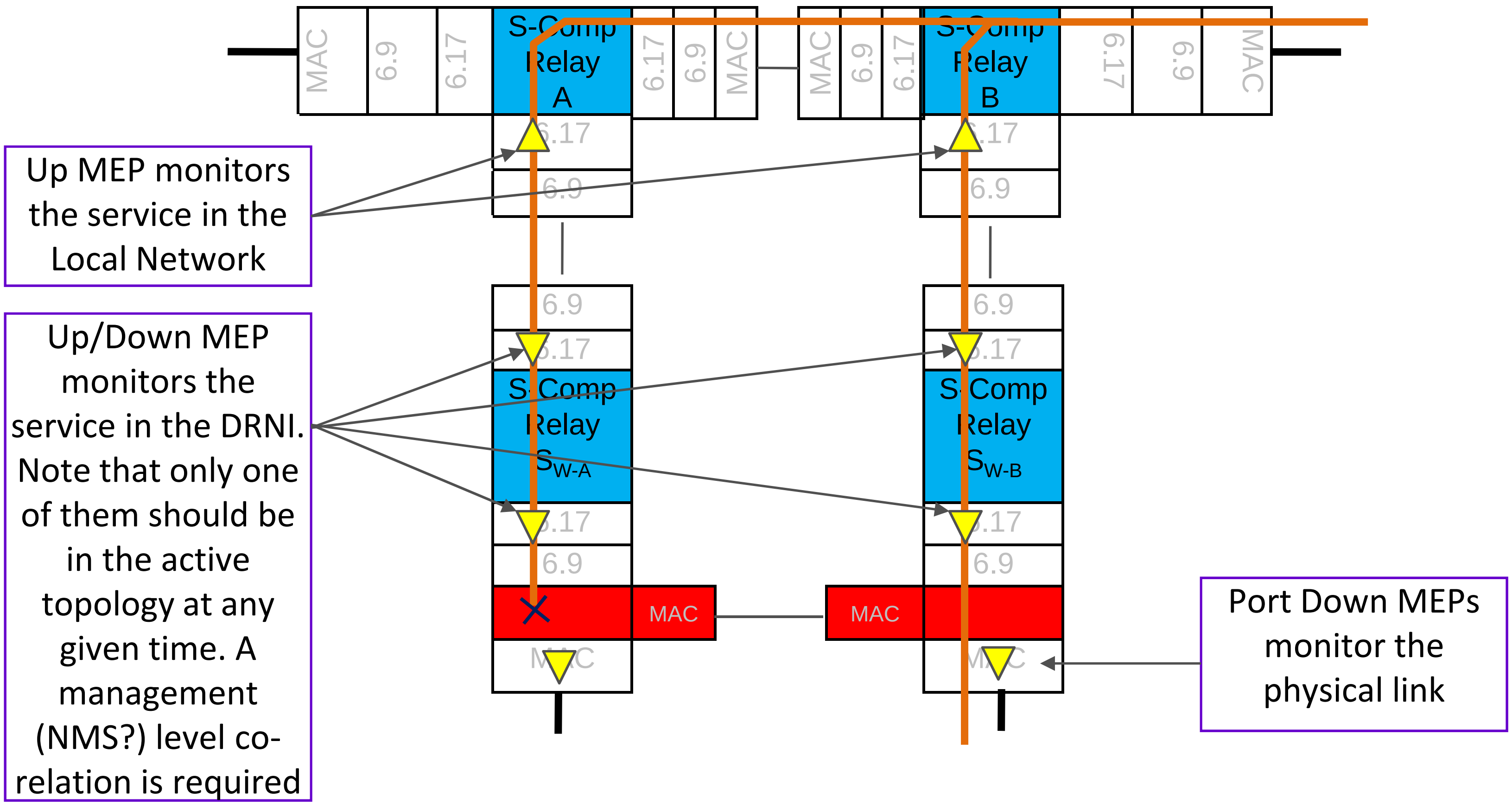
Single DRNI – Simple Example (4)
Link Failure scenario

- Assume that port w2 on B goes down
- The Orange service could now be forwarded via gateway A



Single DRNI – Simple Example (5)
MEP Placement

Bridge Model with the 802.1Q layering



Single DRNI – Simple Example (6)
Link Failure scenario

- Consider a **<u>different scenario</u>** where the **<u>link between A and B is a dedicated intra-DAS link,</u>**
- The logic may dictate that the traffic be forwarded to A over the intra-DAS and out link w1 as shown in Fig. 2.

Fig1. Possibility 1

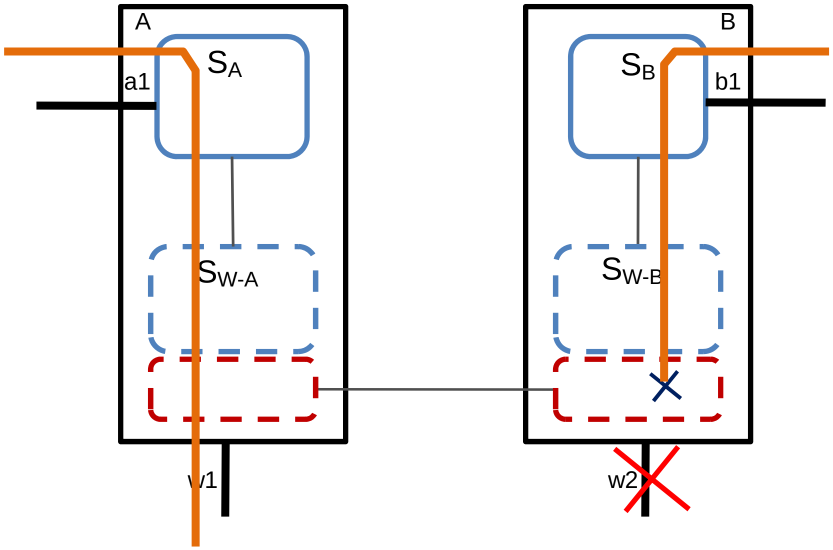
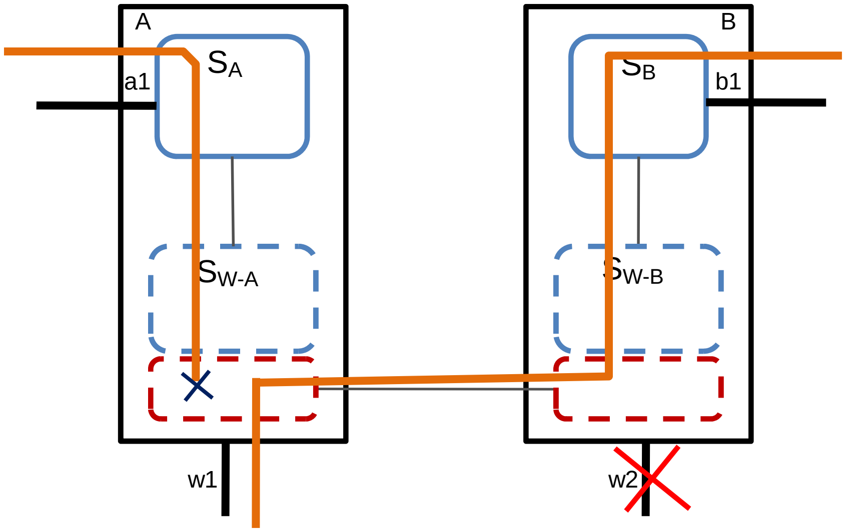
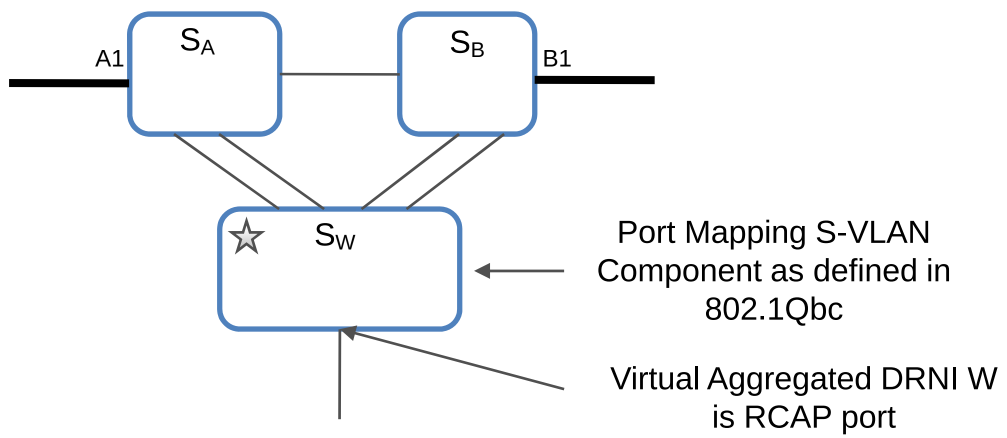


Fig2. Possibility 2

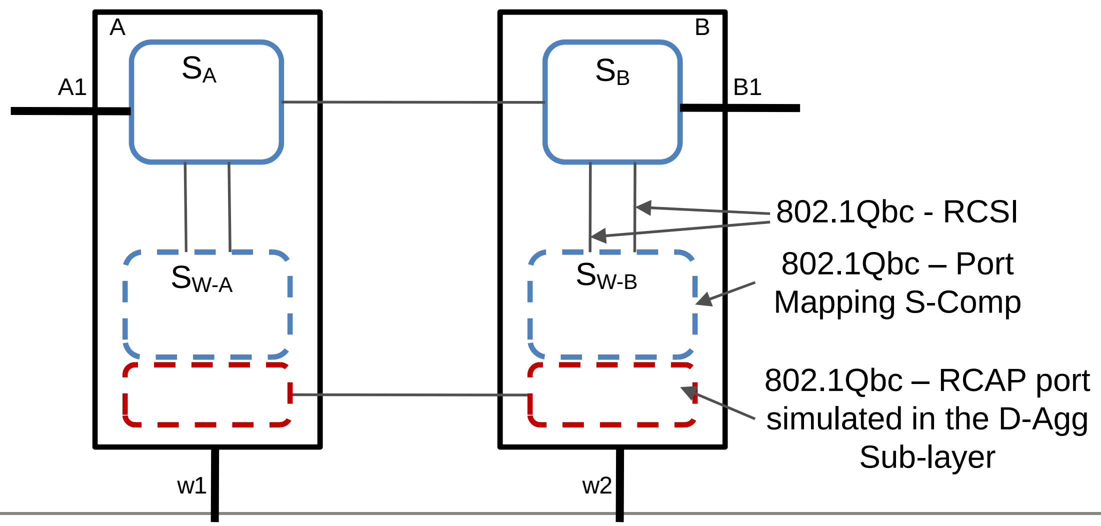


Single DRNI – Hair-pinning (1)

Logical Topology



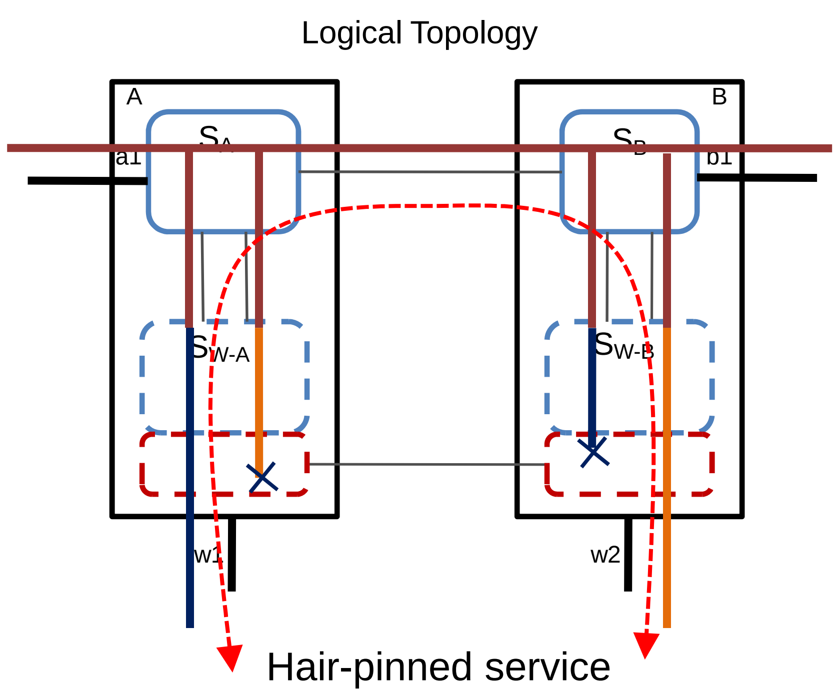
Logical Topology

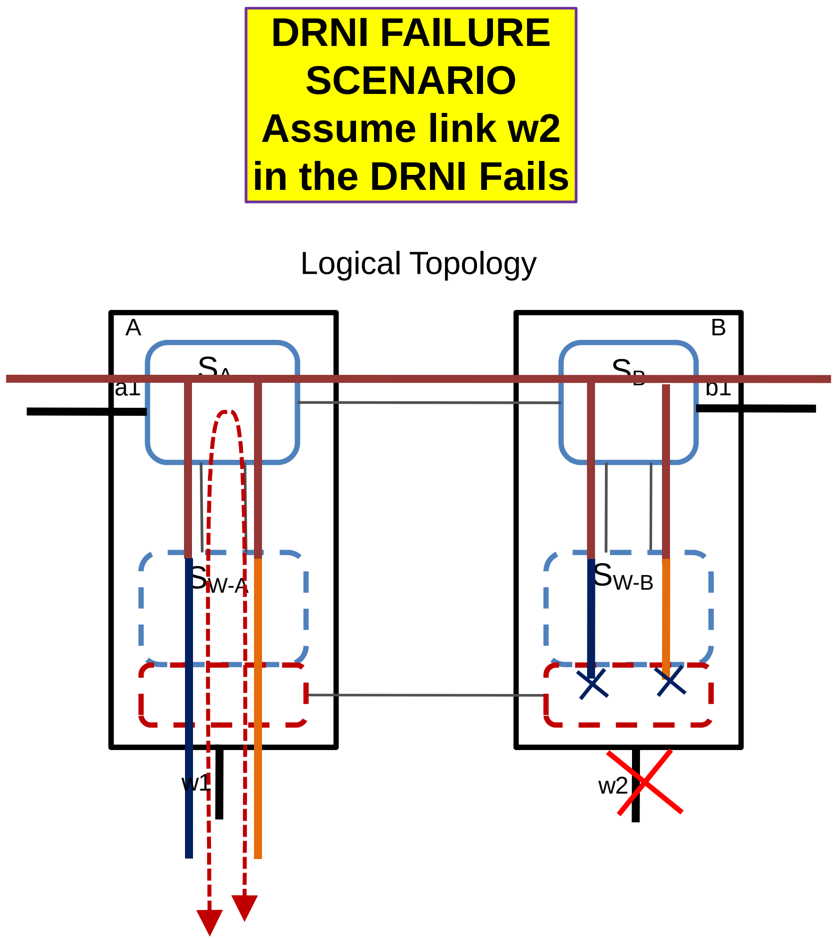


- Same physical topology as in the earlier example
- Assume hair-pinning is required between two services on the DRNI towards the other carrier
- The logical topology is shown here. (along the lines described in 802.1Qbc)
- Only requires that the same configuration be replicated on all nodes in the DRNI portal. Each S-comp on the two physical devices has to have two RCSIs to the distributed S-component as shown.
- **(Note: For simplicity this topology assumes only port based service)**

Single DRNI – Hair-pinning (2)

- The Blue and Orange services require to be hair-pinned. Brown service in the local network.

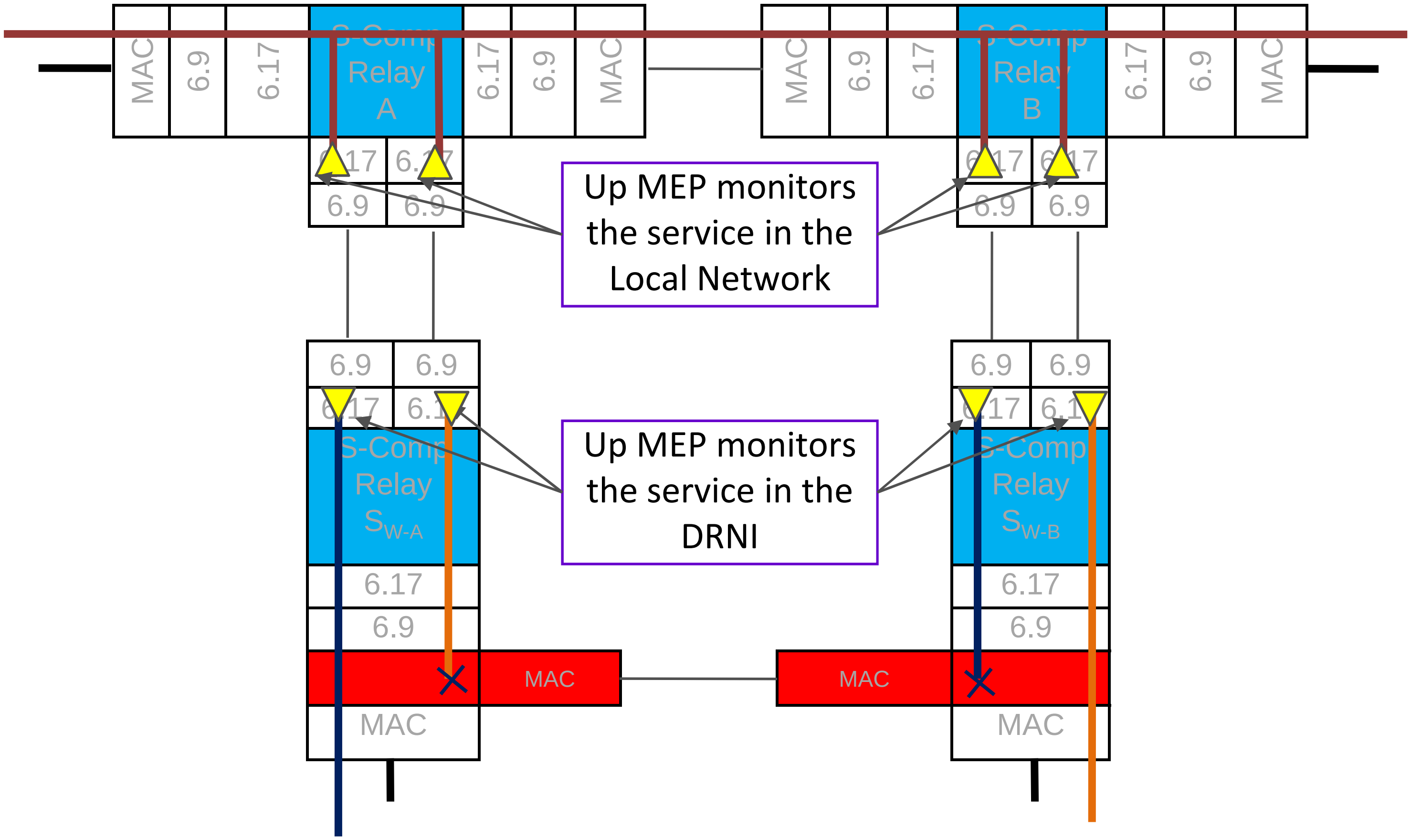




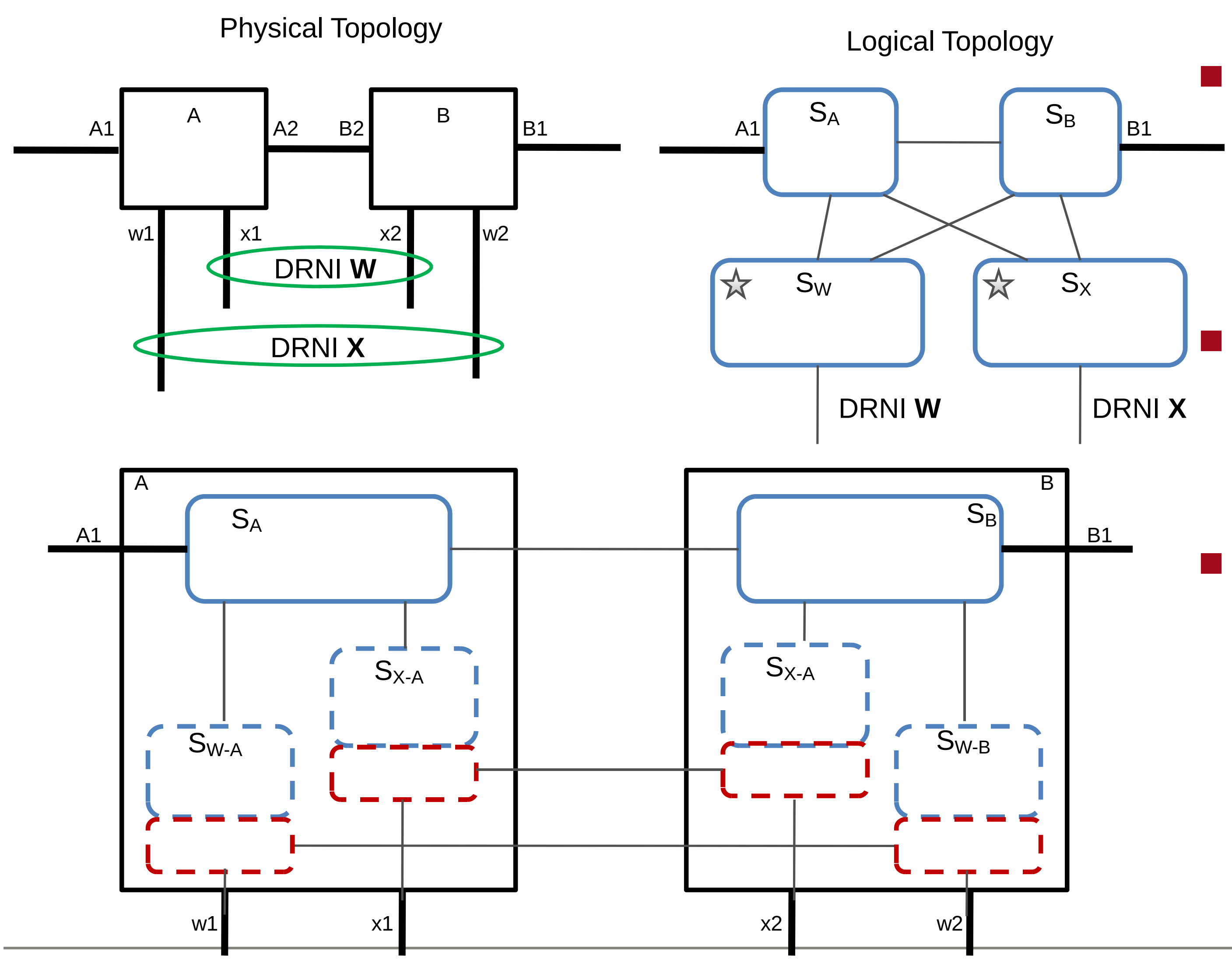
Single DRNI – Hair-pinning (3)
MEP Placement

FUJITSU

- TBD



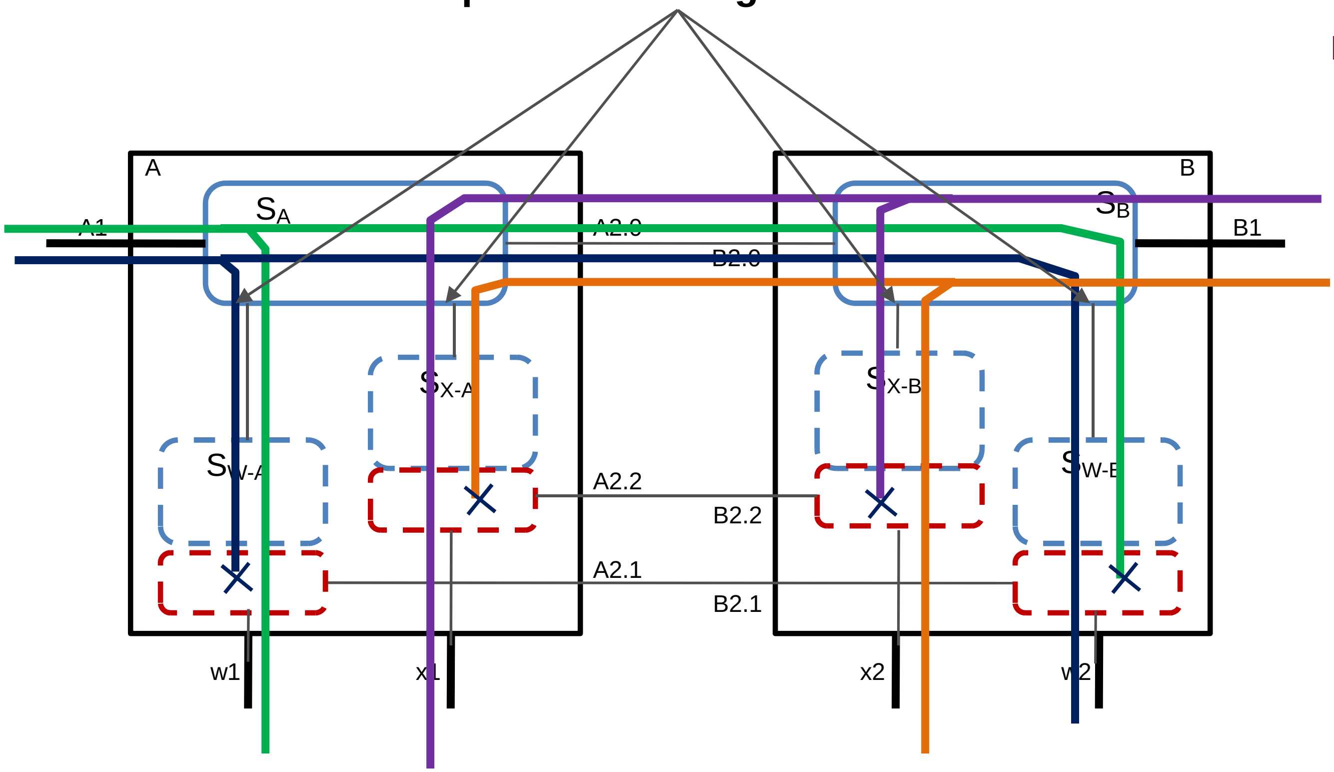
Multiple DRNI – Simple Example (1)



- Assume S-VLAN based service.
- In this example the physical link between A & B has a dual function – it is part of the provider network and also carrier intra-DAS link traffic
- There are two DRNIs – DRNI W and DRNI X. Assuming that each DRNI is connected to a different Carrier network.
- Some services are mapped to DRNI W and other services are mapped to DRNI X.

Multiple DRNI – Simple Example (2)

Service to DRNI Mapping mappings on these ports via configuration



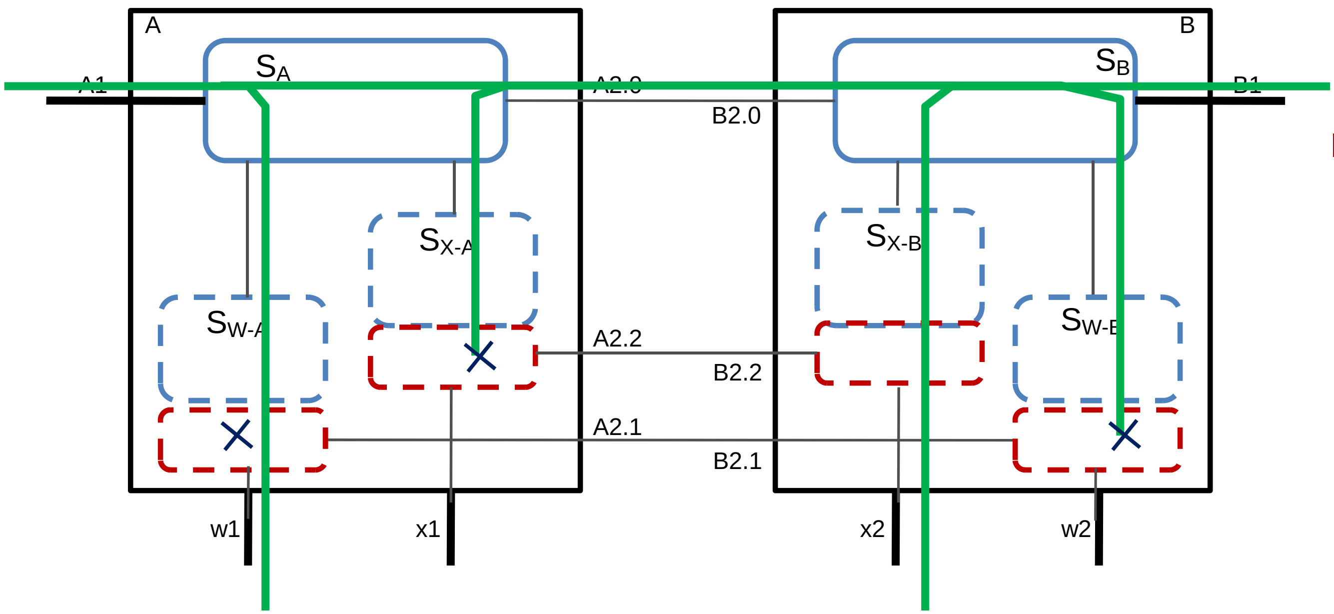
- Assume S-VLAN based service.
- Assume 4 services
  - Services Green and Blue Mapped to DRNI W via configuration.

    Green → Gateway A.
    Blue → Gateway B.
  - Orange and Purple Mapped to DRNI X via configuration.

    Orange → Gateway B.

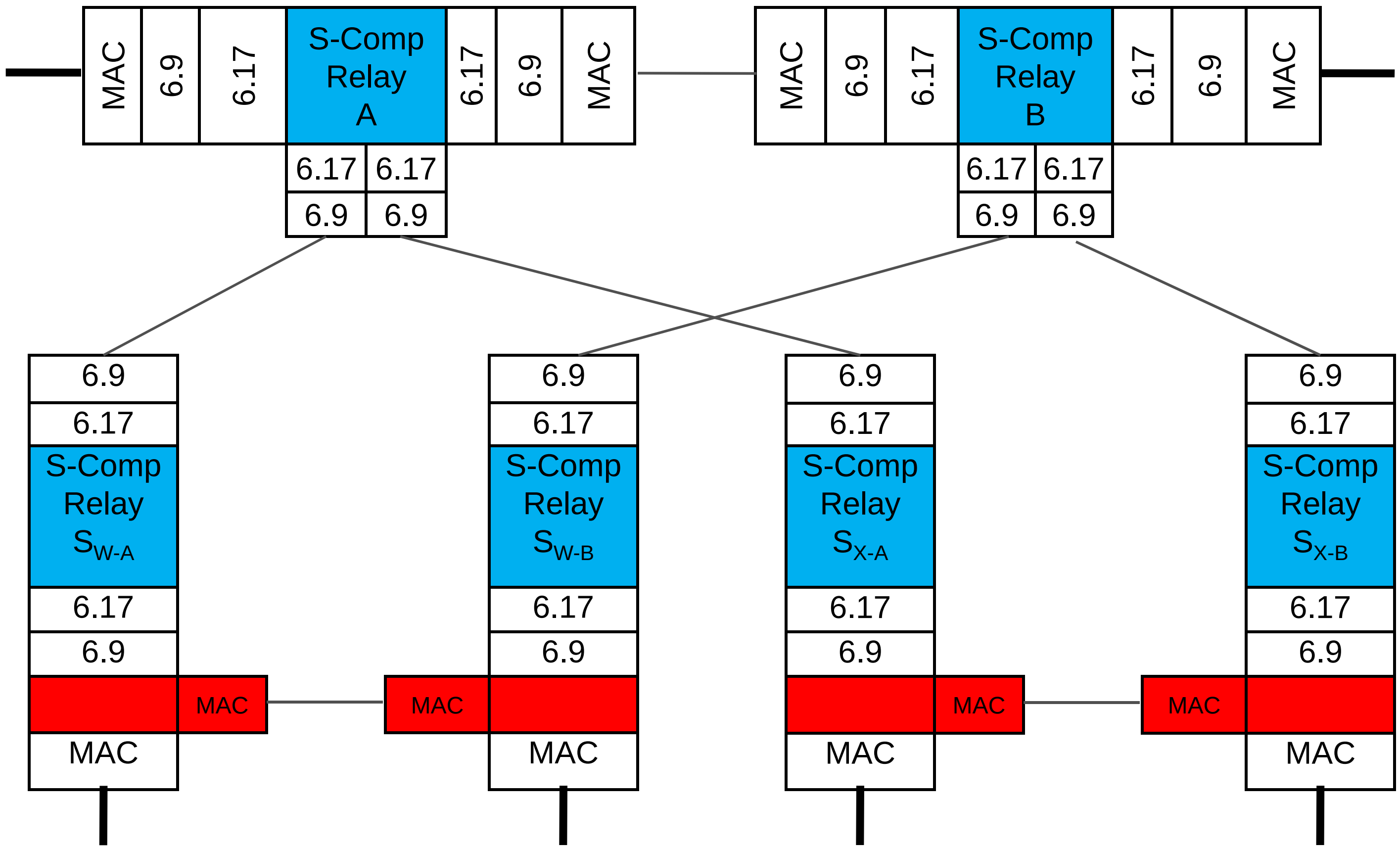
    Purple → Gateway A.

Multiple DRNI - Simple Example (3)
Service spanning across DRNIs



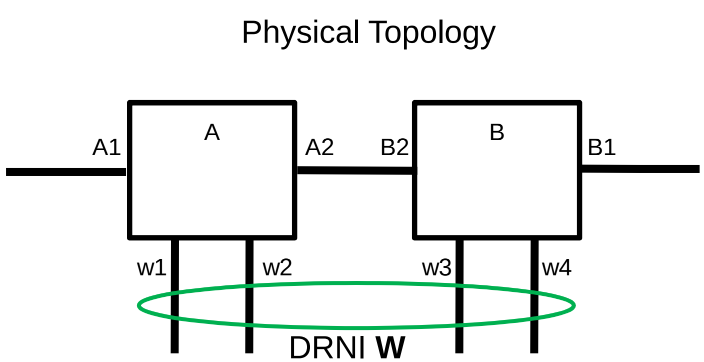
- In a scenario where a service needs to span across multiple DRNIs, the service could be simply created on the logical port connected to the logical DRNI component.
- The figure shows Green Service being forwarded between the two DRNIs

Multiple DRNI – Simple Example (4)
Bridge Model

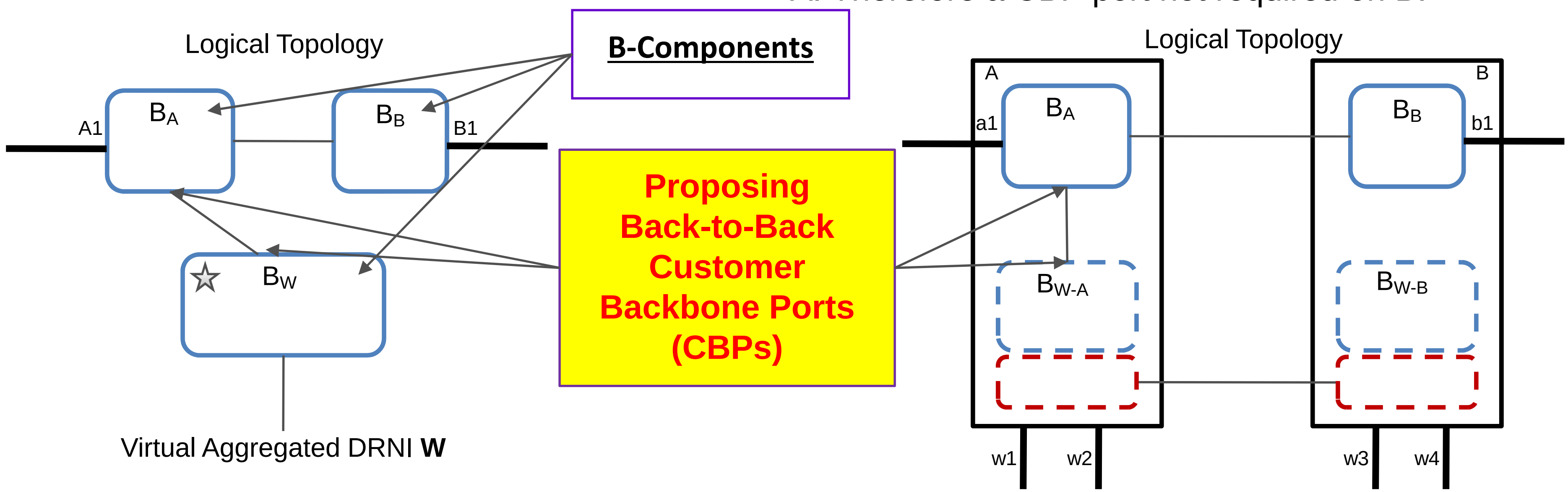


Single DRNI –PBB-TE: No Protection (1)
Un-Protected TESI

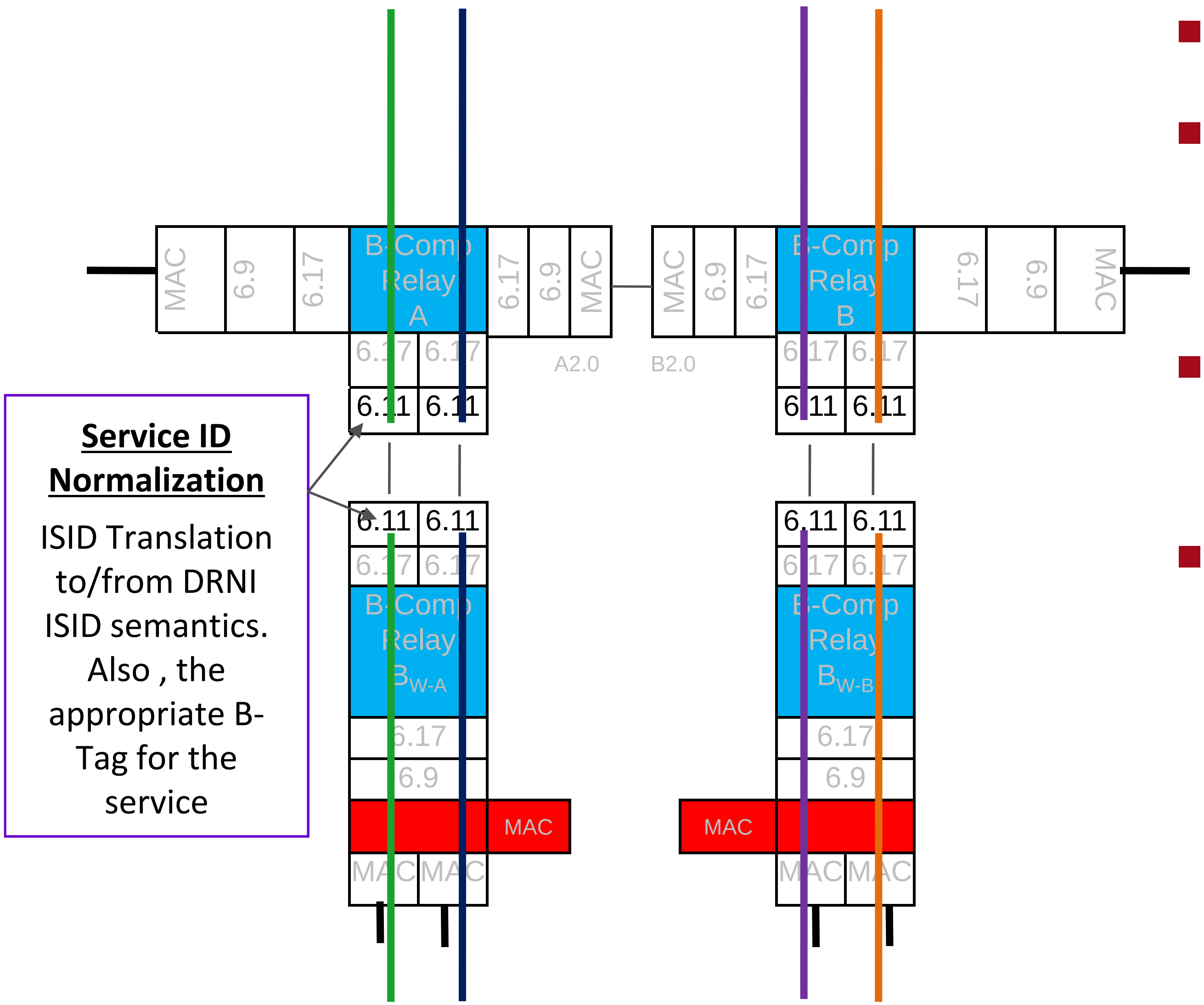
FUJITSU



- Assume a <u>single un-protected</u> point-to-point TESI in the attached network
- (Multiple Service example on next slide)
- **In the connected network an un-protected TESI is configured to terminate on Gateway A or Gateway B**
- **However, the TESI maybe forwarded to the peering network by Gateway A or Gateway B**
- Note: Assume this service terminates on Gateway A. Therefore a CBP port not required on B.



Single DRNI – PBB-TE: No Protection (2)
Un-protected TESI



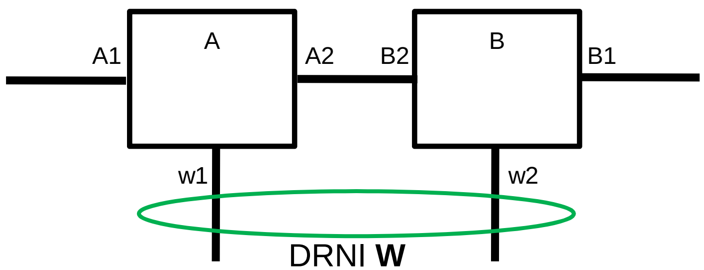
- Assume 4 un-protected services
- Green and Blue services are configured to terminate on Gateway A. Orange and purple terminate on Gateway B.
- Each of these services can be monitored in the connected as well as over the DRNI (MEPs not shown)
- In a failure scenario in the DRNI, the services can be forwarded to the other Gateway. Note that in the connected network these TESIs are unprotected.

Single DRNI – Protected PBB-TE (1)
Protected TESI - Non-Distributed implementation

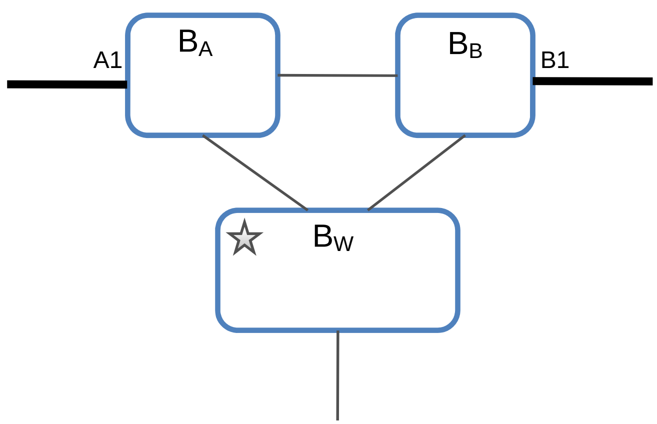
FUJITSU

- Assume two **1:1 protected point-to-point** TESIs in the attached network
- Assume Non-distributed PBB-TE implementation – i.e. both the Working and Protection TESIs either terminate on Gateway A or Gateway B.
- In the attached network, the Green protected TESI is configured to terminate on Gateway A and the Orange protected TESI is configured to terminate on Gateway B.

Physical Topology

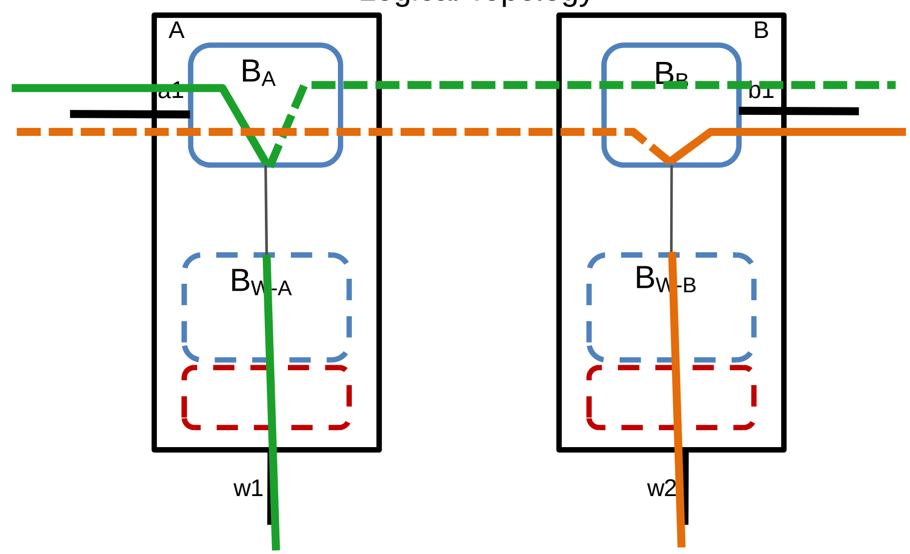


Logical Topology

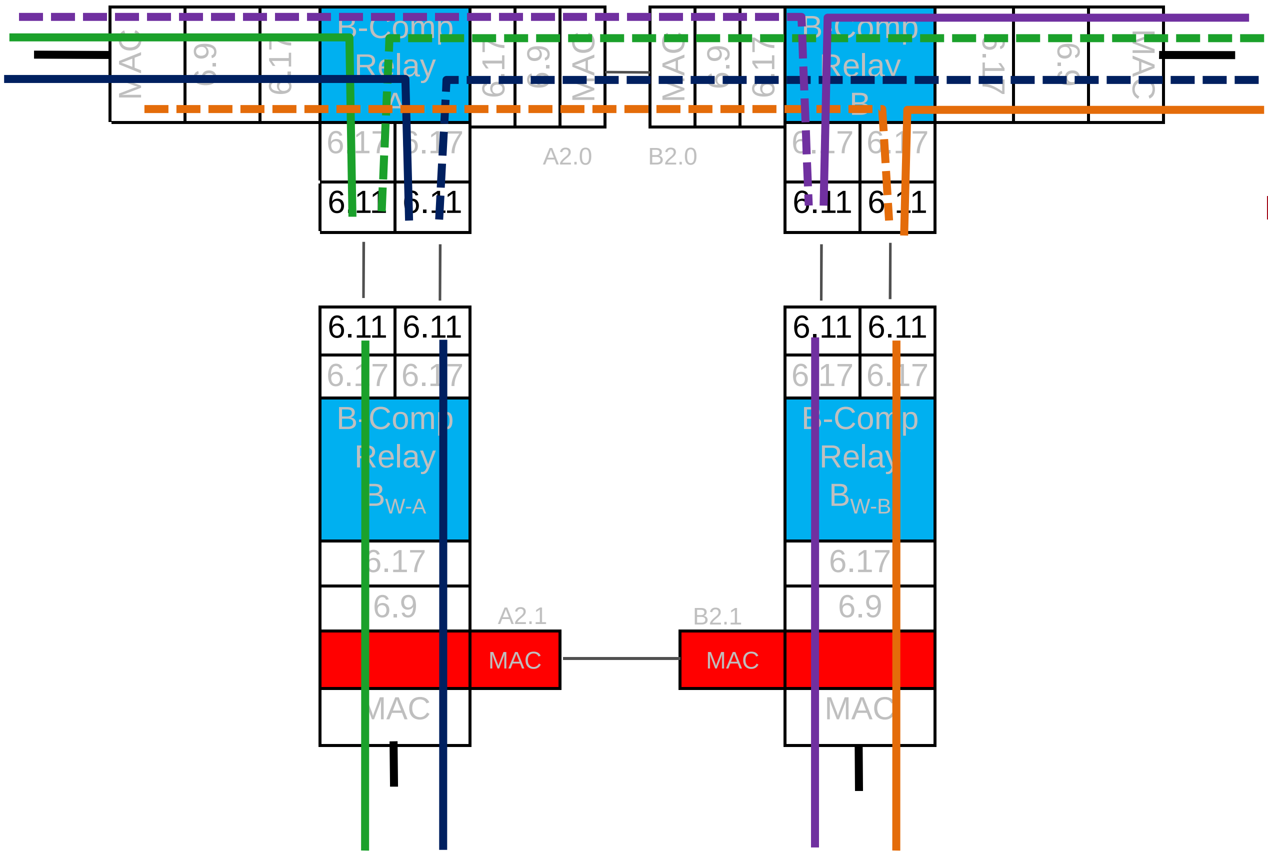


Virtual Aggregated DRNI **W**

Logical Topology



Single DRNI – Protected PBB-TE (2)
Protected TESI - Non-Distributed implementation

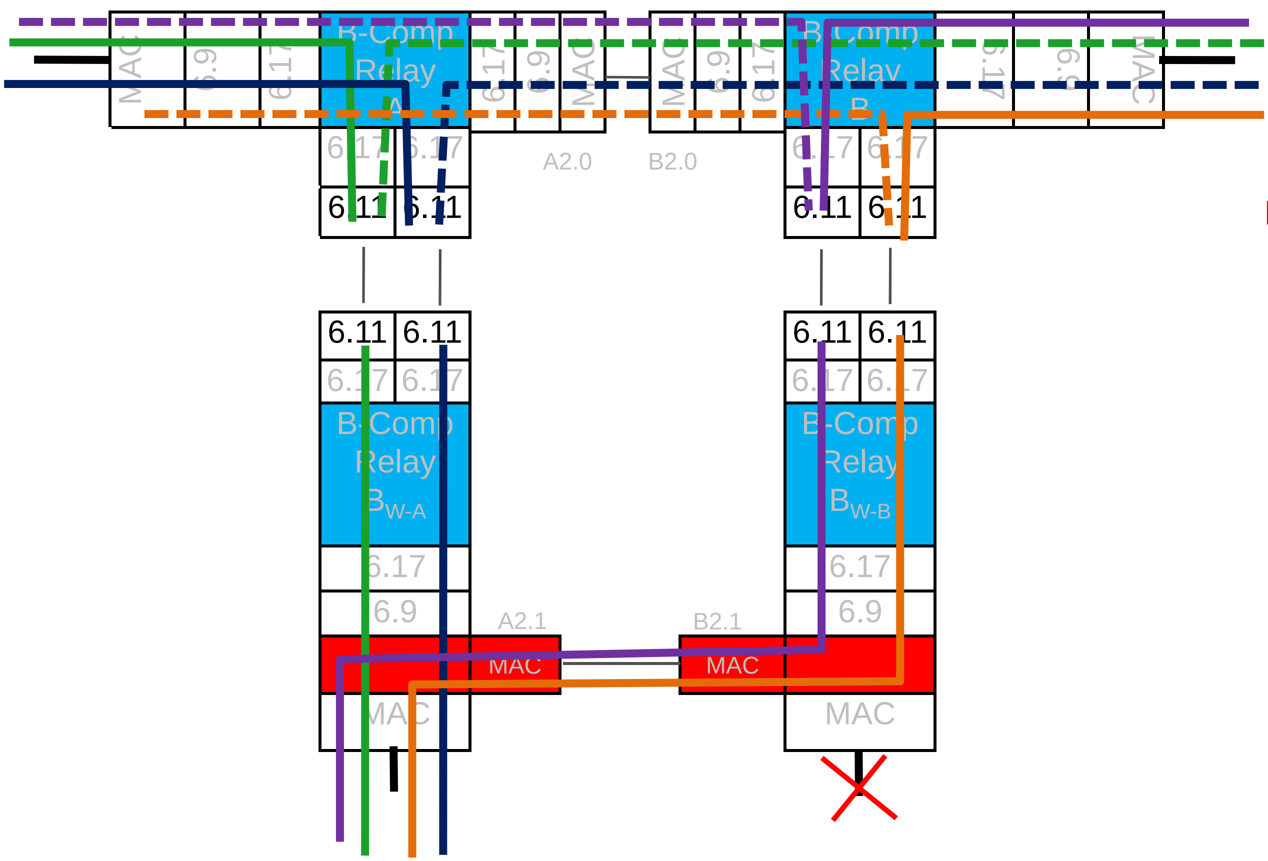


- Assume 4 **1:1** protected services
- Protected Green and Blue services are configured to terminate on Gateway A. Protected Orange and purple terminate on Gateway B.
- Each of these services can be monitored in the connected as well as over the DRNI (MEPs not shown)

Single DRNI – Protected PBB-TE (3)
Protected TESI - Non-Distributed implementation

DRNI FAILURE SCENARIO Assume link w2 in the DRNI Fails

Case 1: Entities DO NOT “move” Gateways

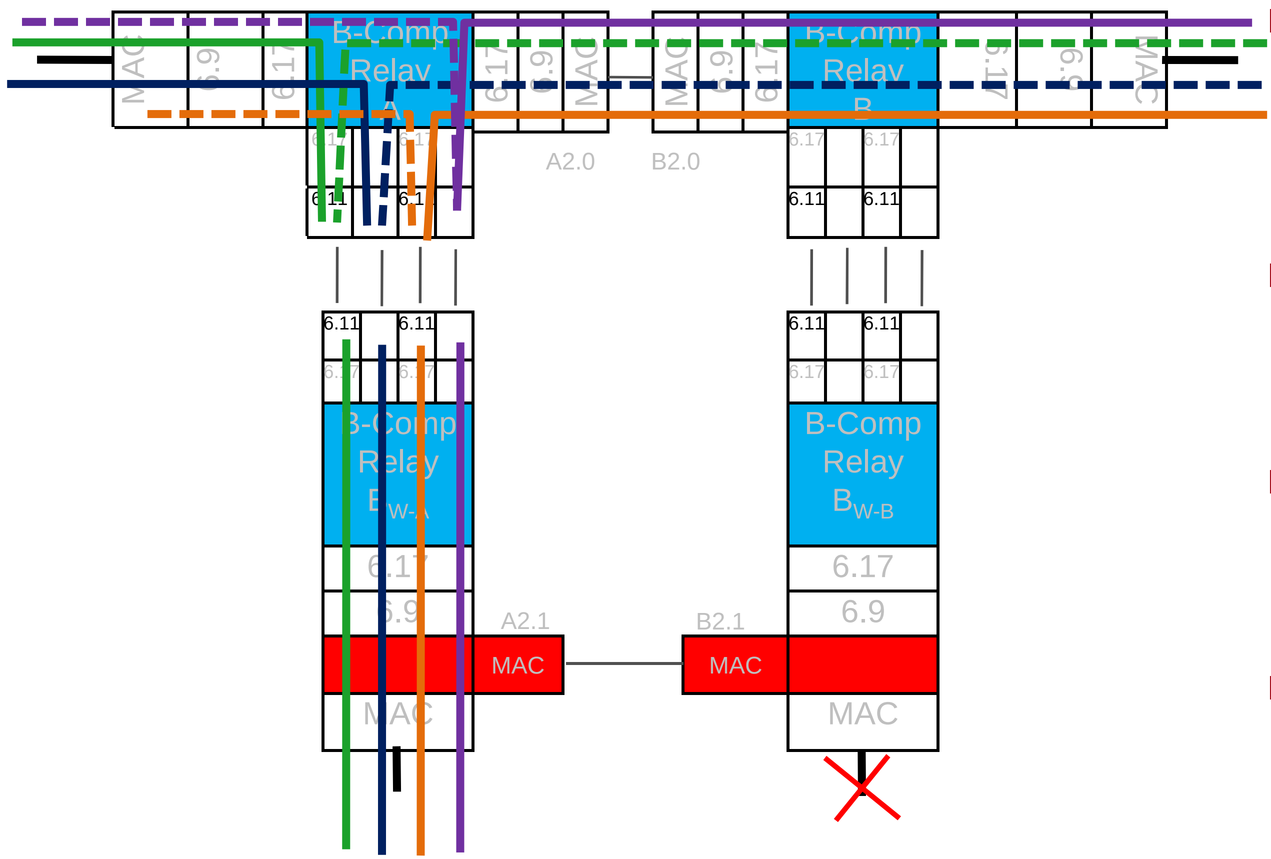


- Assume that link w2 fails
- Green and Blue services continue to be forwarded by Gateway A. The D-Agg Sub-layer now forwards the Orange and Purple services over the Intra-DAS link to be forwarded by Gateway B
- Note that this slide describes the case where the protected Orange and purple services DO NOT “move” to Gateway B as a result of the failure. However, a management entity can decide to move the services to Gateway A (requires virtual-MAC for CBPs and a common configuration on B-Comp A and B-Comp B). Shown on next slide

Single DRNI – Protected PBB-TE (4)
Protected TESI - Non-Distributed implementation

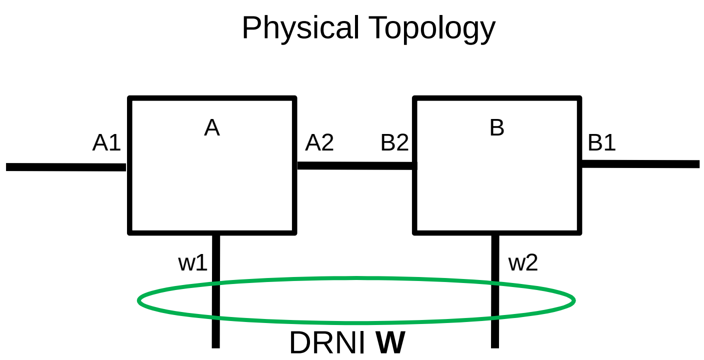
DRNI FAILURE SCENARIO
Assume link w2 in the DRNI Fails

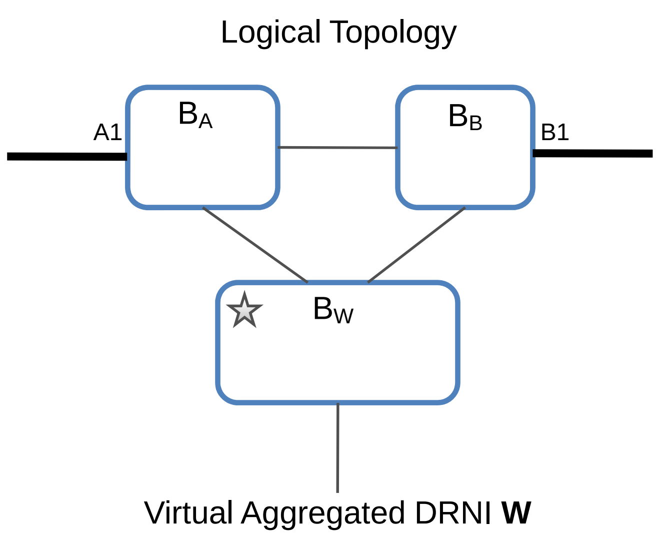
Case 2: Entities "move" Gateways



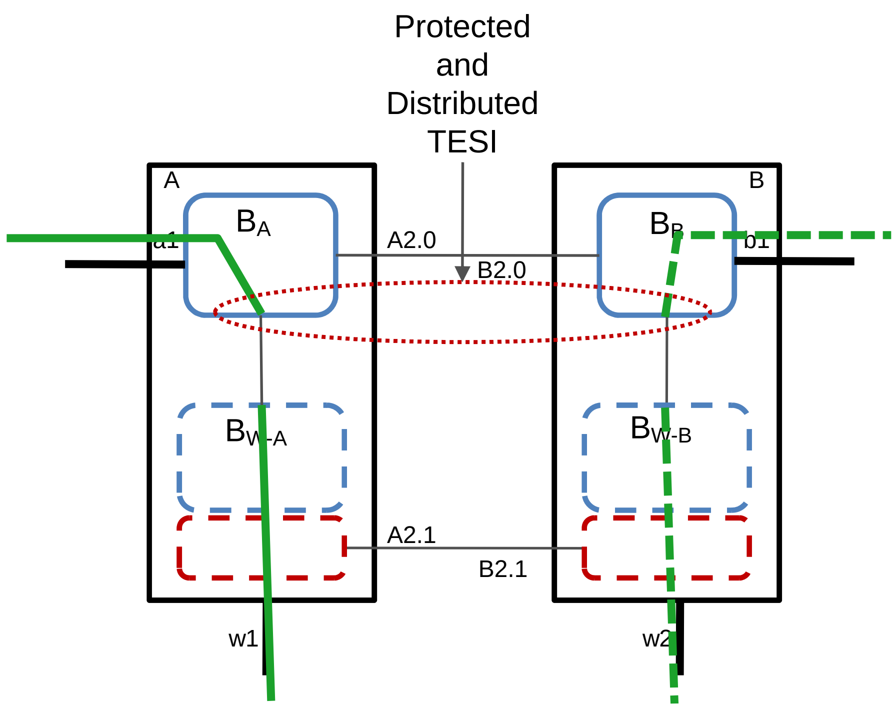
- Assume that link w2 fails
- Green and Blue services continue to be forwarded by Gateway A.
- On failure of link w2 **Management entity decides to switch the gateways for service Orange and purple**
- **How exactly this is achieved should be OUT-OF-SCOPE from the perspective of .1bq**
- In any case, the D-Agg Sub-layer AND the Management entity coordinate the switching of gateways
- Virtual MACs maybe required for CBPs and the forwarding plane needs to be configured appropriately

Single DRNI – Protected PBB-TE (1)
Protected TESI - Distributed implementation

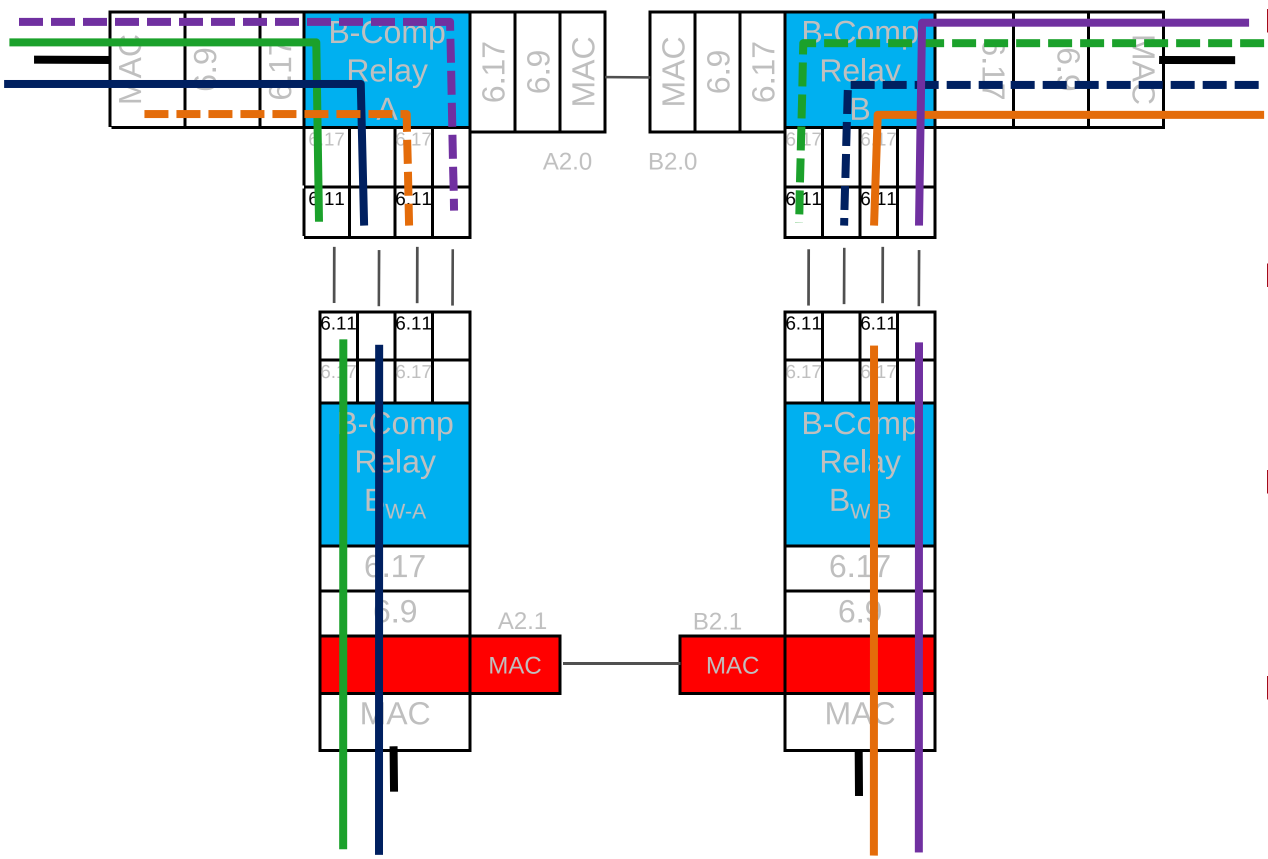




- Assume a single 1:1 protected point-to-point TESIs in the attached network
- Assume a Distributed PBB-TE implementation – i.e. the Working TESI terminates on Gateway A and Protect TESI terminates on Gateway B.
- How this is exactly achieved should be **beyond the scope of .1bq**



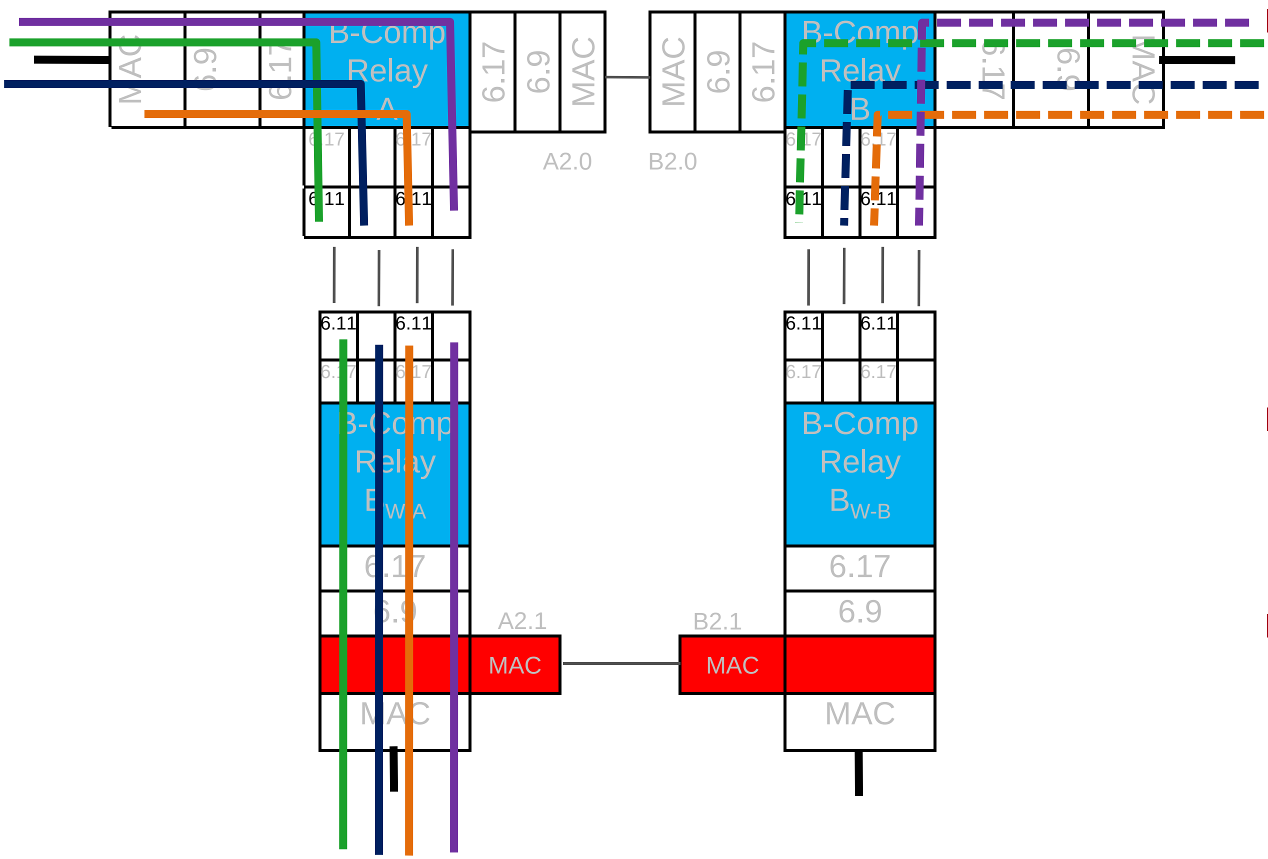
Single DRNI – Protected PBB-TE (2)
Protected TESI – Distributed implementation



- Assume that link w2 fails
- Green and Blue services continue to be forwarded by Gateway A.
- On failure of link w2 **Management entity decides to switch the gateways for service Orange and purple**
- **How exactly this is achieved should be OUT-OF-SCOPE from the perspective of .1bq**
- In any case, the D-Agg Sub-layer AND the Management entity coordinate the switching of gateways
- Virtual MACs maybe required for CBPs and the forwarding plane needs to be configured appropriately

Single DRNI – Protected PBB-TE (3)
Protected TESI – Distributed implementation

DRNI FAILURE SCENARIO
Assume link w2 in the DRNI Fails



- Assume that link w2 fails
- Green and Blue services continue to be forwarded by Gateway A.
- On failure of link w2 **the Distributed PBB-TE implementation could decide to perform a distributed protection switch. So protect TESI is now active. Shown in the picture.**
- **How exactly this is achieved should be OUT-OF-SCOPE from the perspective of .1bq**
- In any case, the D-Agg Sub-layer AND the Distributed PBB-TE entity will have to coordinate

FUJITSU

shaping tomorrow with you