Resilient Network Interconnect:
Requirement for a Single Portal Address

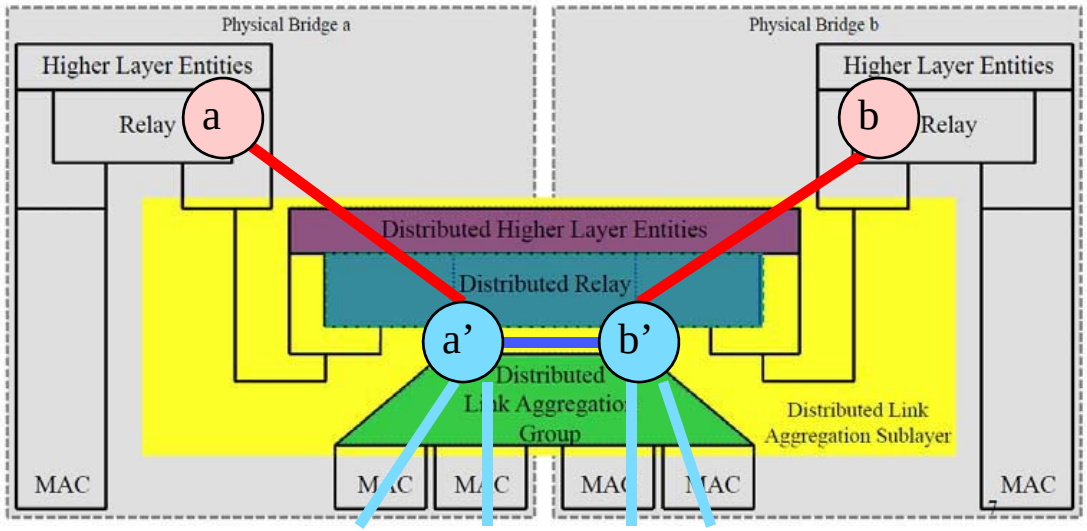
Version 1

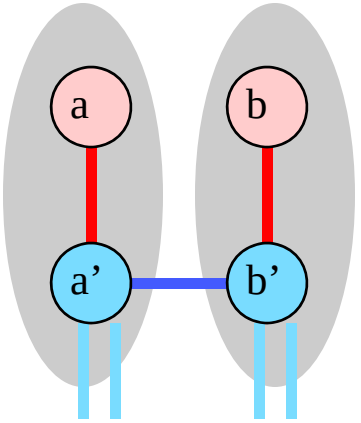
Stephen Haddock
Extreme Networks
November 22, 2010

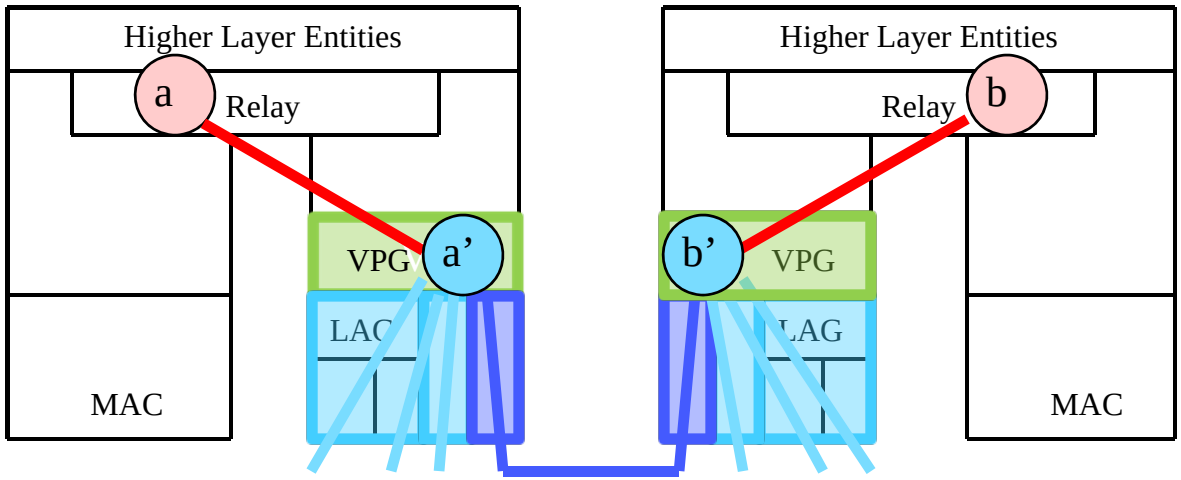
Introduction

- Three models for Resilient Network Interconnect :
  - Virtual Protection Group (VPG) model
  - Buffer Network model (a.k.a. “heavy” model)
  - Distributed Link Aggregation model
- All three models are basically the same in the data plane:
  - All separate the “Terminal Node” into a portion considered part of the “Area Network” and portion considered part of the “Buffer (or Interconnect) Network”.
  - All have a “Gateway” concept connecting the two portions of the terminal node.
  - All allow the service-to-Gateway assignments to be made independently from the service-to-interconnect-link assignments.
- This presentation explores a common requirement in the control plane.
  - The need for a single addressable entity at the “Portal” to the Resilient Network-Network Interconnect (RNNI).
  - This need is independent of the model and independent of the control protocol used in the Buffer Network

Commonality in the RNNI Data Plane Model







- Overlaying Norm's Buffer Network Virtual Node model on Steve's Distributed LAG model and Zehavit's Virtual Protection Group model.
- The data plane functionality is the same in each model.

Difference in the RNNI Control Plane Model

- Make a distinction between:
  1. The Control Plane Protocol

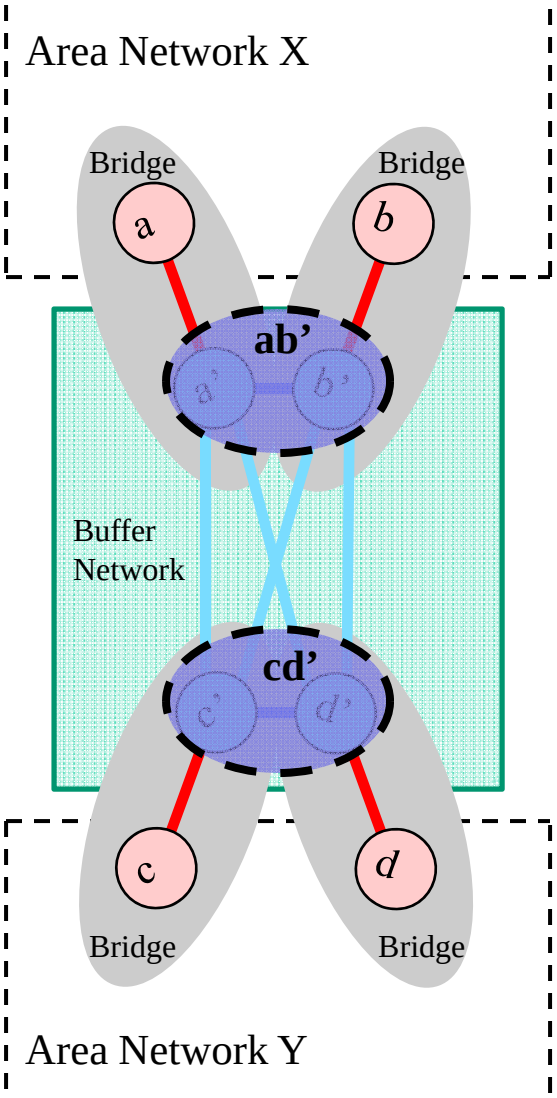
     Exchanges state information, and controls link and gateway selection, among the nodes in the Buffer Network (or Interconnect Network).
  2. The Control Plane Model

     How the control plane elements of the RNNI Bridges appear to the rest of the Area Network.
- The Distributed LAG model creates a single logical control plane element in each Area Network for the RNNI:
  - Addressable as a single entity.
  - Viewed from the Buffer Network as a single entity
    - So can create a single Link Aggregation Group distributed between the Bridges.
  - Viewed from the Area Network as a single entity
    - So could run a single instance of RSTP/MSTP/MVRP distributed between the Bridges (more significant for a general Distributed LAG solution than for RNNI).

Having a single logical control element at the RNNI

- Even if LACP is not chosen as the RNNI protocol, there are other situations where there is value in having a single addressable control element at the RNNI.
- These situations are independent of:
  - The Data Plane Model
    - i.e. the need is common to Norm's, Steve's, and Zehavit's models.
  - The Control Plane Protocol used in the Buffer/Interconnect Network
- These situations include:
  1. Protection Switching in the Area Network
  2. Backbone Area Network with an S-tagged RNNI
  3. Backbone Area Network with an I-tagged RNNI
  4. E-Line services and point-to-point OVCs
  5. Service OAM (CFM and Y.1731)

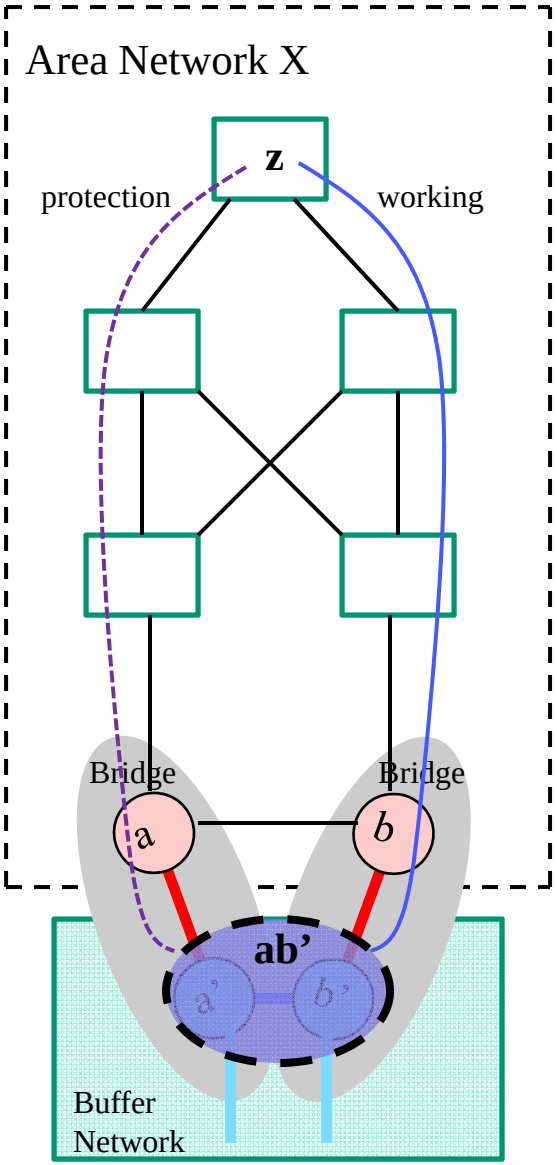
RNNI control element model



- Each Bridge is independently addressable
  - a, b, c, d
- There is an additional addressable entity at each RNNI
  - ab', cd'
  - Visible to control protocols in each Area Network.
  - The internal structure of ab' and cd' is not visible to the Area Network control protocols

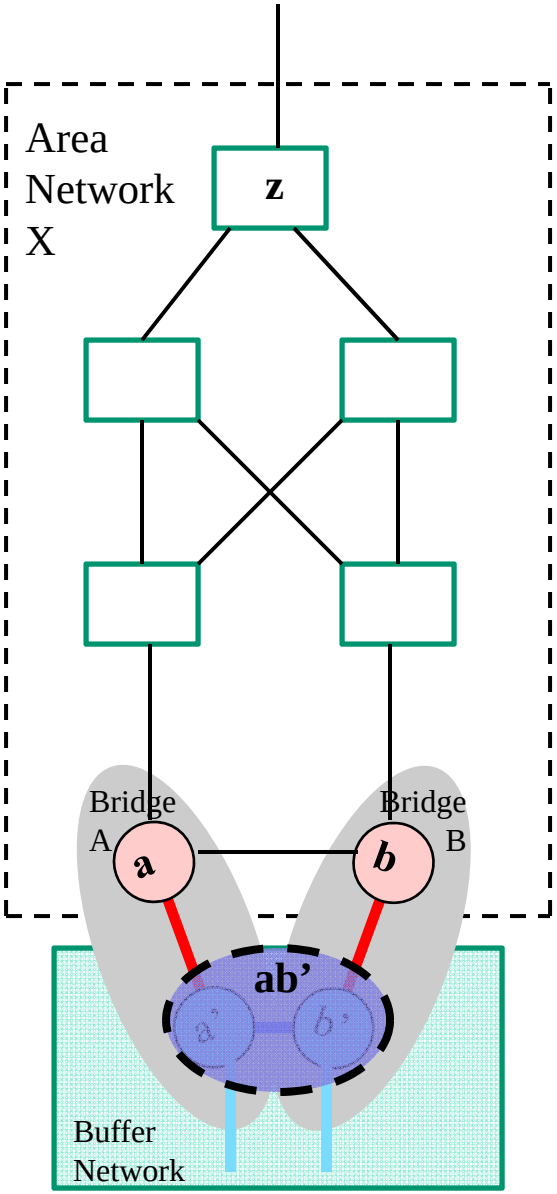
- There may need to be addresses for the portion of each bridge in the Buffer Network
  - a', b', c', d'
  - Used for RNNI control protocol
  - Can probably share addresses (with a, b, c, d) or use nearest bridge group address.
  - Does not need to be decided for the issues discussed in this presentation.

1. Protection Switching in the Area Network



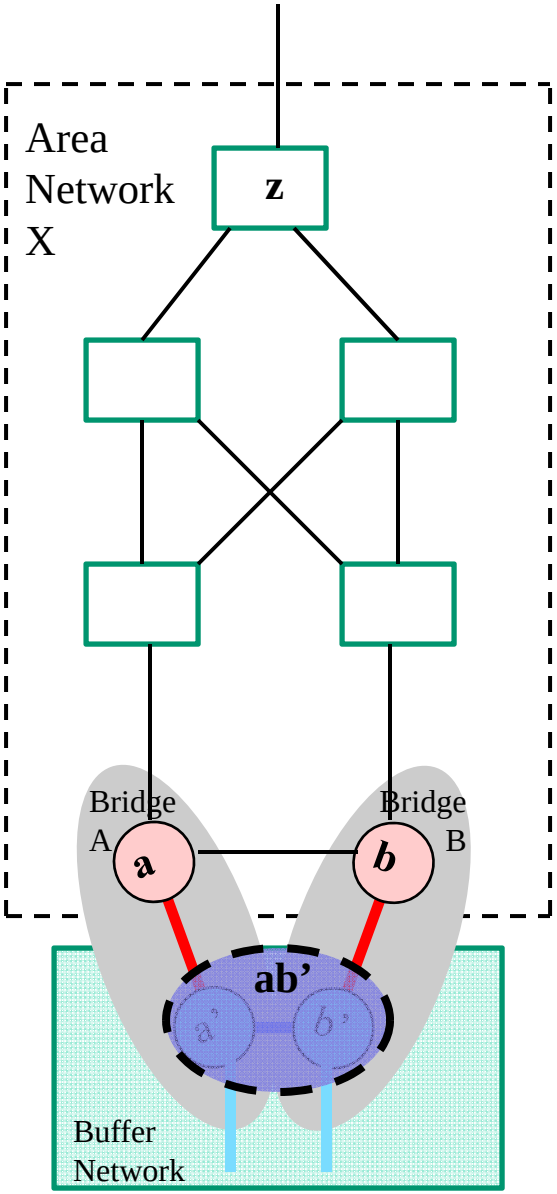
- Addressable entity **ab'** allows provisioning of Working and Protection paths between common endpoints (**z ← → ab'**)
  - Provisioning disjoint paths assures one path goes through **a** while the other goes through **b**.
- Alternative is provisioning to different endpoints (W: **z ← → b**, P: **z ← → a**).
  - Likely not how provisioning tool works.
  - Doesn't work if endpoint address carried in frame (e.g. PBB-TE in Area Network).
- The single addressable entity approach is consistent with the theory Norm developed in new-nfinn-nni-framework-0110-v01:
  - **ab'** is the address of the Portal
  - Area responsibility is to deliver frames to the Portal, not a specific node in the Portal.

2. Backbone Edge Bridge at an S-tagged RNNI



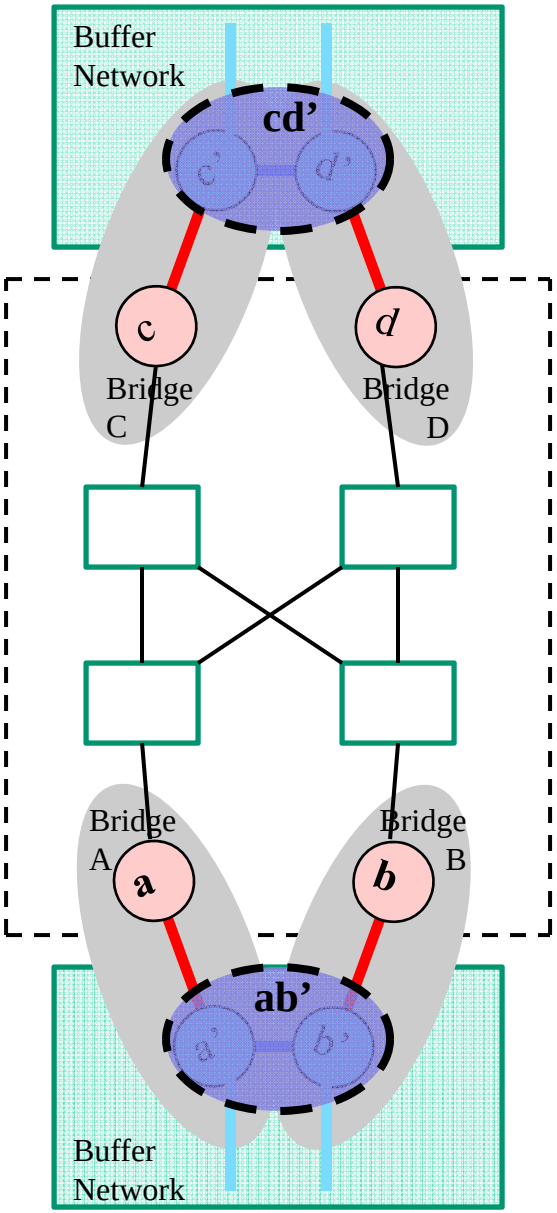
- Bridges A, B, and Z are all IB-BEBs.
- The I-component in Z learns C-MAC address to B-MAC address associations from frames encapsulated at the RNNI.
  - If A and B use independent PIP addresses (**a** and **b**), then Z has to flush and re-learn the C-MAC/B-MAC associations following any gateway changes at the RNNI.
  - If A and B use the **ab'** address for encapsulation (effectively modeling the PIP as part of the common entity), then gateway changes at the RNNI are transparent to Z.
- There are other reasons to model the PIP as part of the common entity
  - In particular, gateway selection is based on B-VID (to avoid learning issues in Area X), so the S-VID → I-SID → B-VID assignments occur prior to gateway selection on ingress.

3. Point-to-Point Backbone Service Instances



- Consider a P2P BSI between an S-tagged NNI at Z and an S-tagged RNNI at A+B.
  - The PIP at Z learns the individual address of the gateway at the RNNI as the Default Backbone Destination address for all broadcast, multicast, and unknown C-MAC addresses.
  - If A and B use independent PIP addresses (**a** and **b**), then Z has to re-learn the Default Backbone Destination address following any gateway changes at the RNNI.
  - If A and B use the **ab'** address for encapsulation, then gateway changes at the RNNI are transparent to Z.
- Consider a P2P BSI between an I-tagged NNI at Z and an I-tagged RNNI at A+B.
  - The CBP at Z can be configured with the individual address of the gateway at the RNNI as the Default Backbone Destination address for any frames using the B-DA derived from the I-SID.
  - As a configured value, this only makes sense if A and B use the **ab'** address for encapsulation, or else gateway changes would require re-configuration.

4. E-Line Services and Point-to-Point OVCs



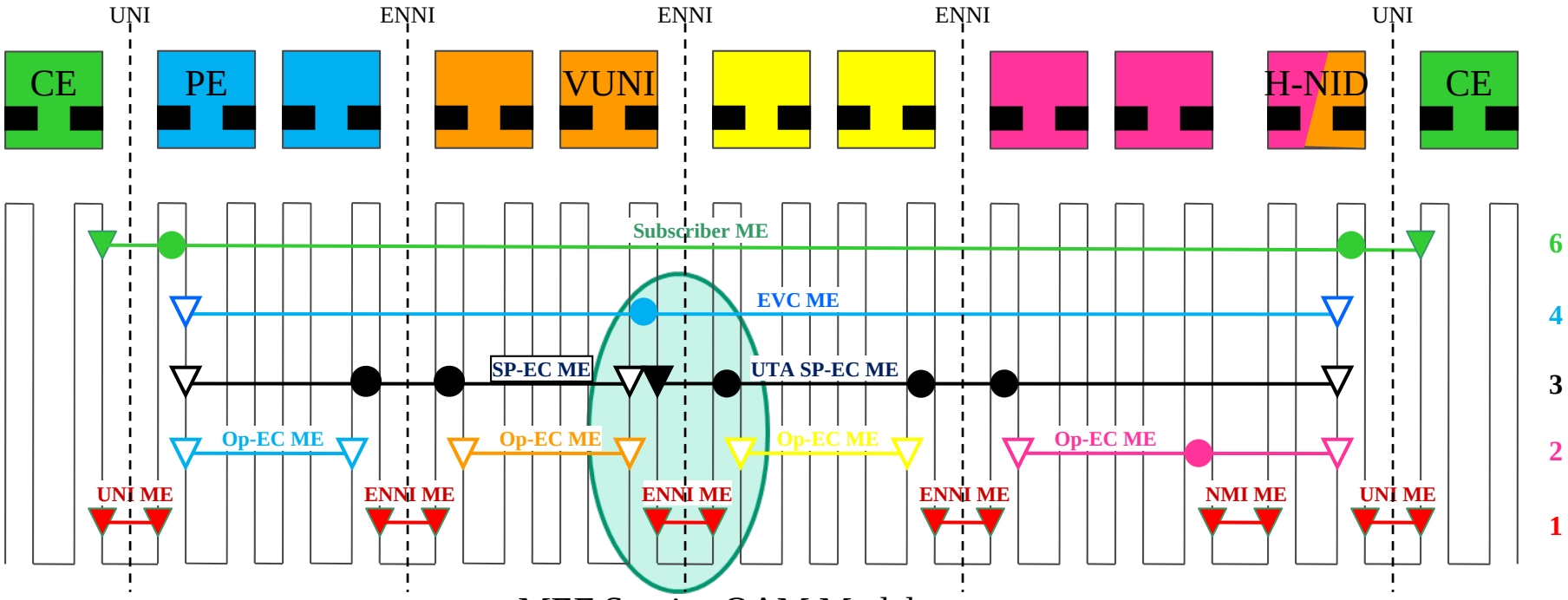
- Consider an E-Line service traversing an operator network between two RNNIs.

  How is the connection between the RNNIs modeled?

  1. A multipoint connection between **a**, **b**, **c**, and **d**?
  2. A point-to-point connection between the current gateways (**a - c**, **a - d**, **b - c**, or **b - d**) where the connection endpoints change when the gateway changes?
  3. A point-to-point connection between **ab'** and **cd'**, where only the path through the network changes when the gateway changes?

- The MEF model expects this to be configured as a point-to-point operator virtual connection (OVC).
  - The E-Line service itself is by definition point-to-point, and is expected to be consistently point-to-point.
    - In some cases attributes of the service depend upon it (e.g. learning constraints, CESoE, UTA/VUNI,…)
  - An ENNI with redundant connections is still considered a single interface.
- The OVC can only be configured as P2P using **ab'** and **cd'** as the addressable endpoints.

5. Service OAM (CFM and Y.1731)



MEF Service OAM Model

- Focus on the MEPs and MIPs at an ENNI. Do they change at an RNNI?
  - Related to previous issue: How many MEPs are at the RNNI on a given level?
- At some level want the RNNI to look like a single interface.
  - If don't have single addressable entity at the RNNI, the individual gateways are visible to the highest level of service provider monitoring (the MIP on the EVC-ME).
  - If have a single addressable entity, the RNNI can look like a single interface down to the ENNI ME level (and still have transport level 0 for individual links).

Conclusion

There are several reasons to have a single addressable entity at an Resilient Network-Network Interface (RNNI).

All of these are independent of the protocol used to at the RNNI.

But there is a dark side …

Single Addressable Entity – the Dark Side

- Each of these issues requires that the RNNI maintain the same address through the failure (and recovery) of any node or link in the portal.
  - One of the nodes, but never more than one node, assumes the address in the presence of a failure.
  - Means each node must be able to reliably detect the difference between a link failure in the portal and a failure of another node in the portal.
  - This is the most challenging of the “split-brain” issues to solve.