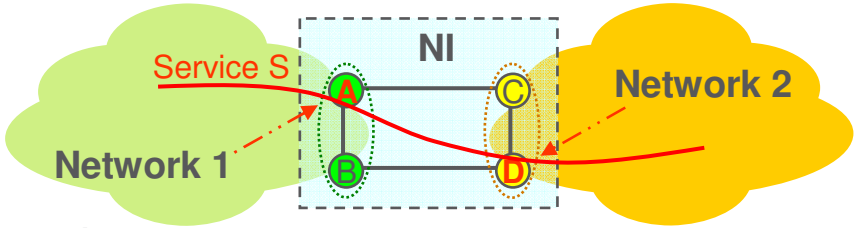
Resilient Network Interconnect Functionalities

János Farkas

Requirements

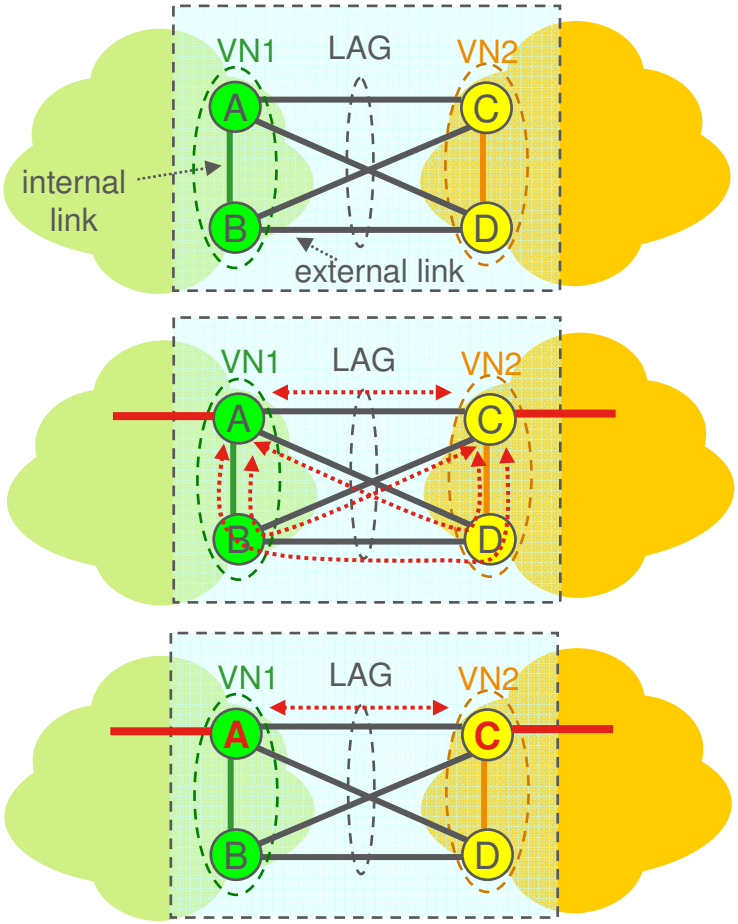
- NI resiliency is independent of the attached networks



- R1 – Independent service assignments
  - A provider may select an NI node for a service independently of the peering provider's selection
- R2 – NI failure isolation
  - NI failure should not cause state change in the provider networks' control protocols
- R3 – Failover time
  - NI should provide sub 50 msec failover time for link failures
- R4 – Connectivity
  - NI should provide loop and duplicate free connectivity
- R5 – Congruency
  - Congruency should be supported
- R6 – NI topology
  - NI topology should be at least two-connected

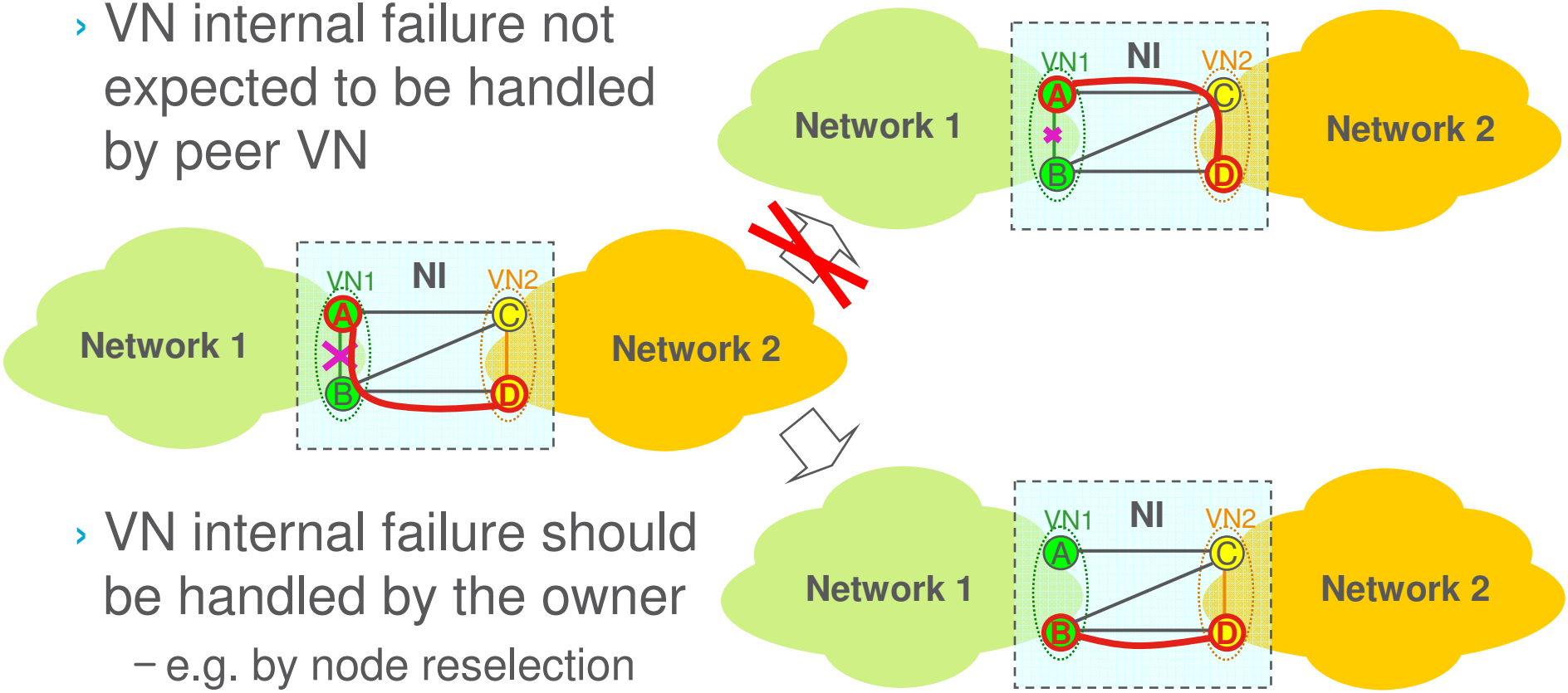
LAG with Virtual Nodes for the NI

- The model
  - NI nodes of the same network comprise a Virtual Node (VN)
  - External links connecting VNs are grouped by LAG
- LAG may cause undesirable traffic load within VN
- Per service differentiation avoids the problem
- If LAG does not distribute traffic, then LAG can be used for
  - Managing (grouping) external links
  - Monitoring the availability of external links



Failure isolation

- VN internal failure not expected to be handled by peer VN

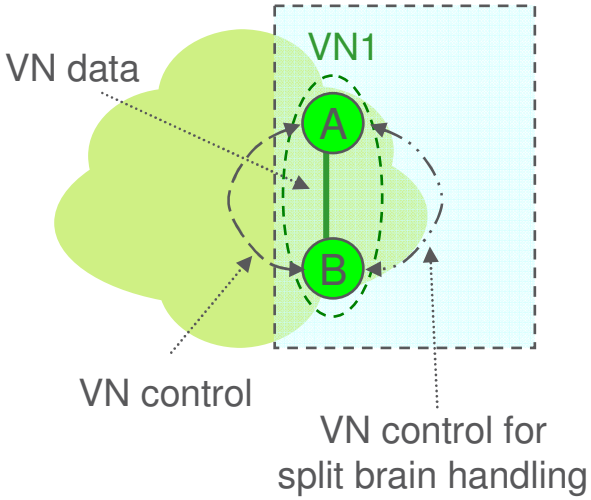


- VN internal failure should be handled by the owner
  - e.g. by node reselection

- External link failure should be completely isolated by the NI

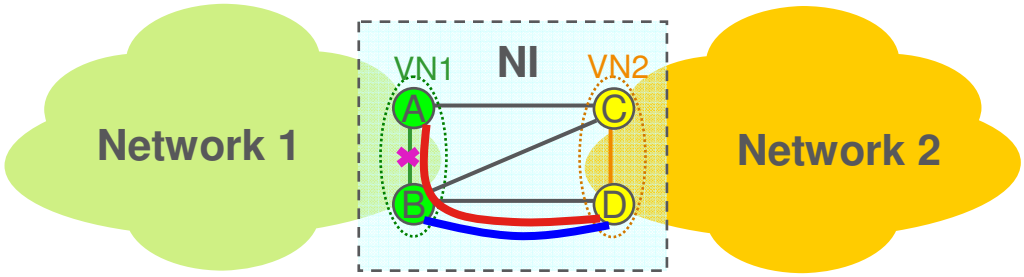
VN internal connections

- Separate data and control connections within a VN
  - Easier to distinguish link and node failures
- Data link break down can be handled by reselection
- Control link break down = split brain
  - Keep a single node
  - The others “suicide”



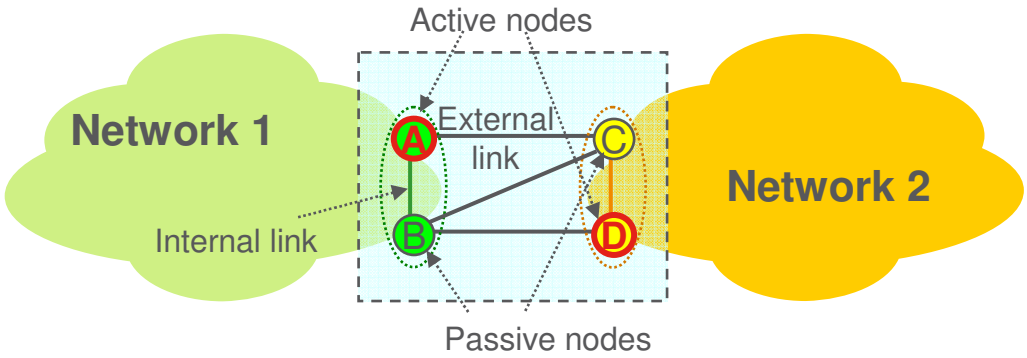
Bundling

- Control plane bundling
  - reduce the amount of configuration, i.e. link priority list is configured per bundle instead of per service
  - reduce the number of events to be managed after a fault is detected, i.e. fault management is per bundle instead of per service
  - reduce the amount of signaling, i.e. outbound link is signaled per bundle instead of per service
  - e.g. bundling VIDs to MSTIs
- Data plane bundling
  - reduce the number of forwarding entries, i.e. forwarding is per bundle instead of per service
  - e.g. bundling I-SIDs into B-VIDs
- *Bundling should be applied on services carried over the same external link and destined to the same NI node*
  - Red and blue should be different bundles as they require different treatment



Picking among nodes comprising a VN

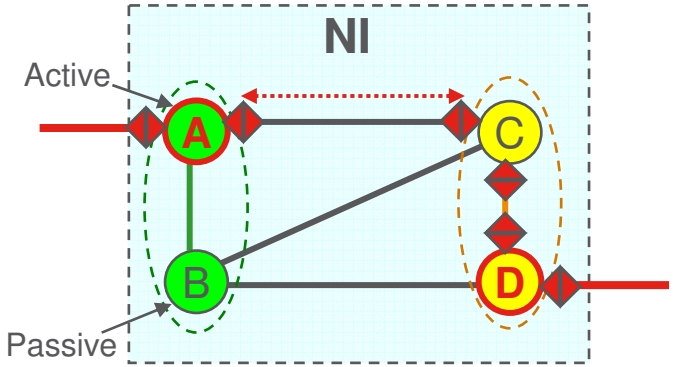
- To ensure loop and duplicate free forwarding Picking protocol selects:
  - a single Active (gateway) node within a VN for a service all other nodes are Passive
  - a single outbound link for a service (= inbound if congruent)



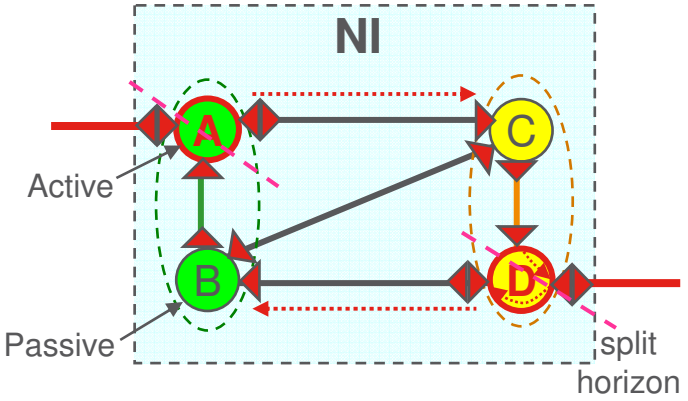
- Active node: forwards frames between the NI and the attached network
- Passive node: Not allowed to forward between the NI and the attached network
- Each network runs its own picking protocol independently of the peer
- Picking is based on priority values
  - Configured by the operator (or set by its routing protocol)
  - Provided by auto-provisioning

Picking results and forwarding

- Congruent
  - Inbound = Outbound
  - Coordinated external link priority
  - Picking
    - Highest priority node
    - Highest priority external link

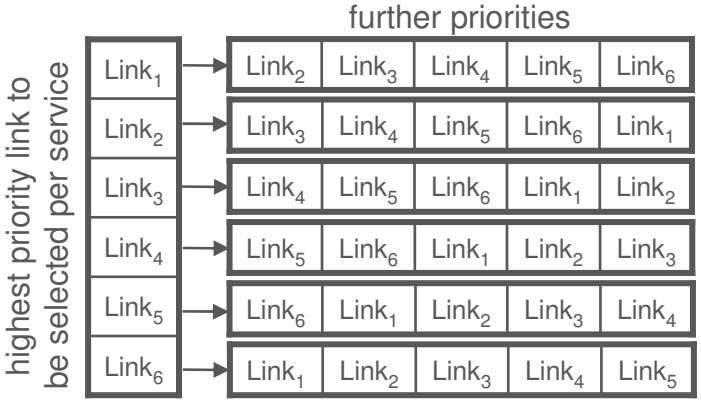


- Non-congruent
  - Inbound link can be any → traffic is directed
  - Active node splits horizon
  - Picking
    - Highest priority node
    - If the node has multiple links: highest priority external link



Auto-provisioning

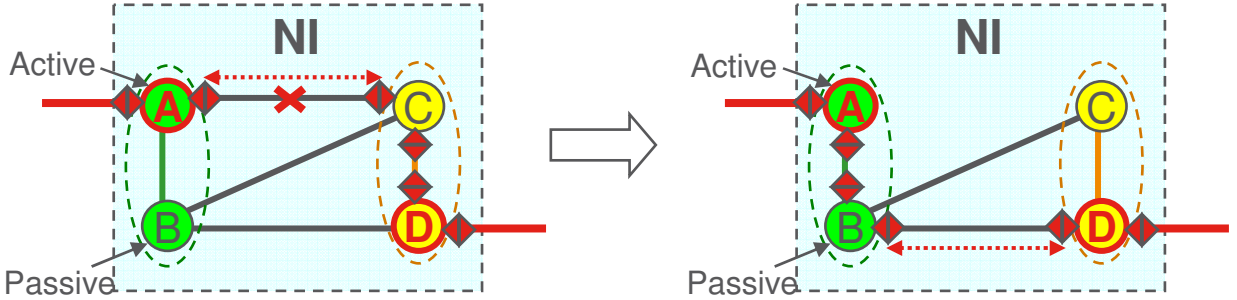
- Use IDs available for both networks
  - Service ID
  - External link ID (LAG: Link ID=$Port_1$,$Port_2$ – has to be unique!)
- Congruent
  - External link priority list has to be the same for a particular service
  - Highest priority link selection should provide load distribution, e.g.: Service ID modulo #external links in the sorted list of link IDs



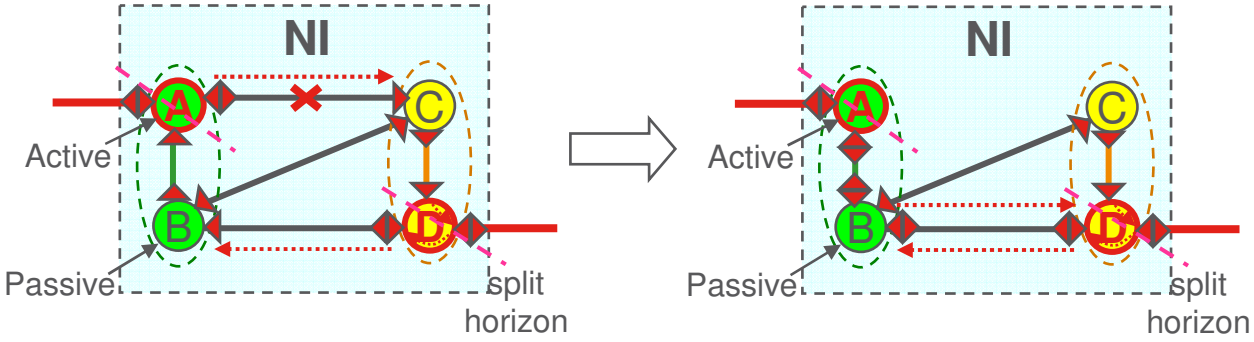
  - Further priorities are determined by highest priority link in order to support bundling (should provide even link utilization)
  - Bundle ID is determined by highest priority external link and local active node
- Non-congruent
  - Priorities are local, any algorithm can be used, e.g. Service ID mod #nodes
  - Bundles have to be coordinated, e.g. Service ID mod #Bundle IDs-i i={0…#Bundle IDs-1} until an unused Bundle ID found

External link failure

- Has to be hidden from the attached networks
- Congruent

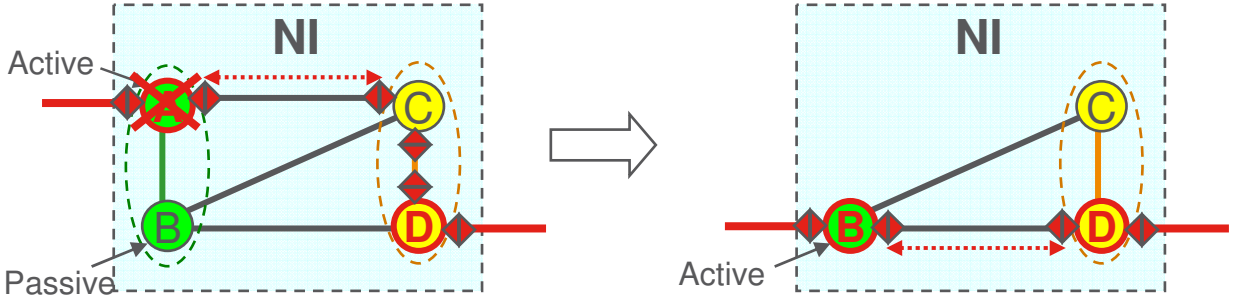


- Non-congruent

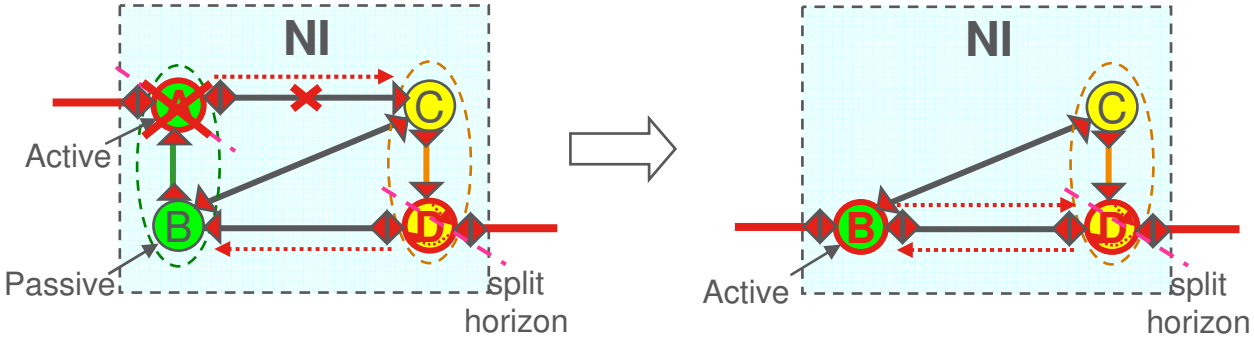


NI node failure

- Perceived as external link failure by the peering network
- Congruent

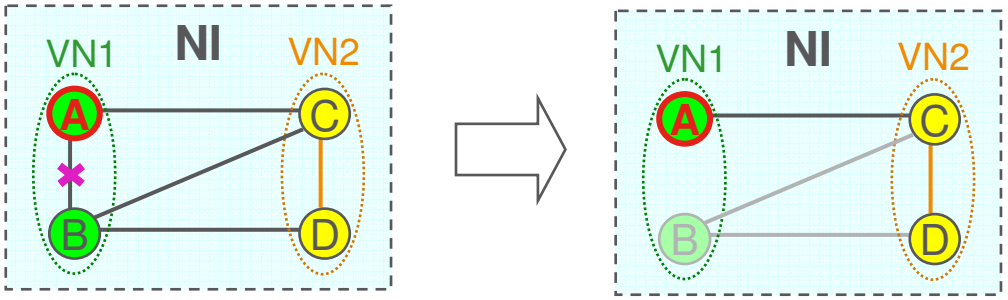


- Non-congruent

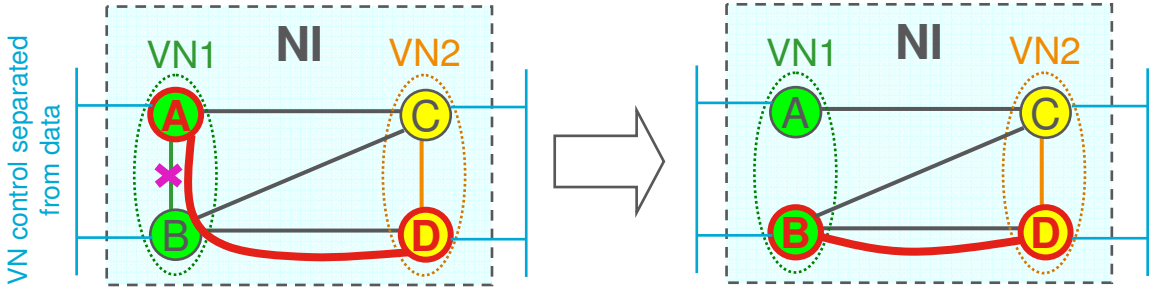


Internal link failure

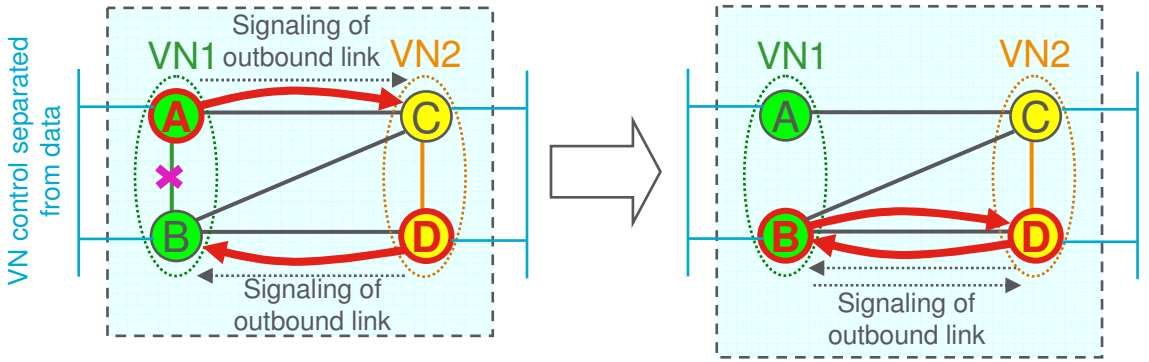
- Common
  - Same node for all services
  - Shut down all other external links
  - No coordination
  - No signaling



- Congruent
  - Re-picking
  - Coordinated external link priorities
  - No signaling



- Non-congruent
  - Re-picking
  - Signaling of outbound link
  - Coordinated bundles



Examples

- S-tagged
  - Service ID = SVID (VID translation if differs from the attached network)
    - Scope of forwarding decisions: SVID
  - Bundling
    - MSTP type of bundling can be used, per port per MSTI forward variable can be reused
    - Congruent: Control and data plane bundling can be applied
    - Non-congruent: Control plane only bundling makes more sense
  - Split horizon for non-congruent can be implemented by SVID translation
- I-tagged
  - Service ID = I-SID
  - Bundling
    - BVID
      - Scope of forwarding decisions: BVID
    - MSTP type of bundling may also be used
  - Split horizon for non-congruent can be implemented by BVID translation