Control of the DRNI

János Farkas
Balázs Péter Gerő
Panagiotis Saltsidis
Minimalist control for DRNI

› Minimal coordination
  - Service ID is the absolute minimum
  - Congruency can be ensured by coordinated external link priorities

› Minimal communication
  - Providers do not like to share their network internal information
  - LACP (or CCMs) already provide the status of the external links to the peering parties

› Keep LACP as simple as possible
  - It is possible to provide control for DRNI with a simple LACP if
  - Portals run a standardized control protocol, thus
  - Portal controls can be run independently
  - Therefore, provider preferences are hidden from the peer
Picking selects Gateway and external link

- Picking protocol selects:
  - Single Active Gateway node within a Portal for a service, (all other nodes are Passive)
  - Single outbound link for a service, the inbound is the same for congruent services

- Each Portal 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)
  - If not set, then provided by auto-provisioning
Auto-provisioning

› Non-congruent services
  – Auto-provisioning may be independent as no coordination is required

› Congruent services
  – Auto-provisioning has to provide coordination
  – Standardized algorithm can ensure that peering parties determine the same external link priorities for a particular service
  – No need for message based coordination

› Auto-provisioning should distribute the load for normal operation
  – (Providing connectivity is the main goal after failures, not load distribution)

› A simple example algorithm for congruent services
  – Input parameters
    › S = Service ID
    › N = number of external links
    › L = sorted list of external link IDs (determined by LAG)
  – Highest priority link: H
    › P = S modulo N
    › H = P-th element of L
  – Further priorities are relative to the highest one in a pre-defined manner, e.g. as shown in the figure
Implications of Picking on forwarding

- Data plane properties (new-farkas-RNI-data-plane-0111-v02)
  - Each external port is prepared to receive data frames:
  - Transmission is only on the outbound port (able to receive too):
  - Active Gateway node splits horizon
- Picking and link status determine forwarding
  - Congruent service
    - Inbound link = Outbound link
    - Coordinated external link priorities
    - Picking selects
      - Highest priority node
      - Highest priority external link
  - Non-congruent service
    - Either link can be the inbound link
    - Outbound link is connected to the Gateway
    - Picking selects
      - Highest priority node
      - Highest priority external link of the highest priority node
Fault Management: External link failure

- External link failure has to be hidden from the attached networks
- New outbound link is selected instead of the broken one

- Congruent

- Non-congruent
Fault Management: Node failure

- DRNI node failure has to be hidden from the peering network
  - New Gateway is selected in the affected Portal
    - In principle, the network selects the Gateway, re-selection may be affected by the convergence of the control protocol(s) of the network,
    - The gateway is probably not reachable from a remote network node until the end of network convergence
    - 50 ms should not be required for Portal node failure
  - New outbound link is selected in both networks

- Congruent

- Non-congruent
Portal Internal link: physical link and/or overlay tunnel

- Direct, protected physical link is preferred
- Sometimes, it is not possible to have a direct link
  - Geographically dislocated Portal nodes
- A solution covering overlay tunnel covers direct link too
- Protection for portal internal link is not a DRNI task
  - Protected by other means in case of protected physical link
  - Protected by the control protocol of the network in case of an overlay tunnel, thus
  - Overlay tunnel breakdown = split network
Fault Management: Internal link/tunnel failure

- Internal link/tunnel failure is perceived as node failure
  - It is caused by multiple failures as the internal link/tunnel is protected
- It is not a node failure \(\rightarrow\) Split Brain handling

Due to Split Brain treatment
  - The disconnected Portal node is excluded from the DRNI
  - Split network (=overlay tunnel breakdown) is not aimed to be connected by the peering partner

Portal control if overlay tunnel breakdown is perceived
  - If the other portal node is up, then Split Brain handling
  - If the other portal node is down, then node failure handling (page 7)
  - Portal control should wait the end of network convergence before declaring node breakdown
  - 50 ms should not be required for handling of multiple failures causing split network

- Having monitoring through the peering network too could help to distinguish Portal node and Portal internal link failures
Bundling for an S-tagged interface

What is worth to bundle?

- Data plane bundling
  - The forwarding decision is based on S-VIDs, see e.g. new-farkas-RNI-data-plane-0111-v02
- Control plane bundling
  - S-VIDs may be bundled to reduce control plane complexity

Control plane bundling may be useful

Independent control allows for independent bundles

- Bundles are network internal
- No need for coordination of bundles
Summary

› DRNI control should be Portal internal
  – Portals run the same standardized control independently of each other
    › Standard auto-provisioning provides the coordination necessary for congruent services
  – Simple LACP

› Portal internal control allows for network internal bundles

› Fault management can be based on the status information and on the results of Picking