Split Brain Handling

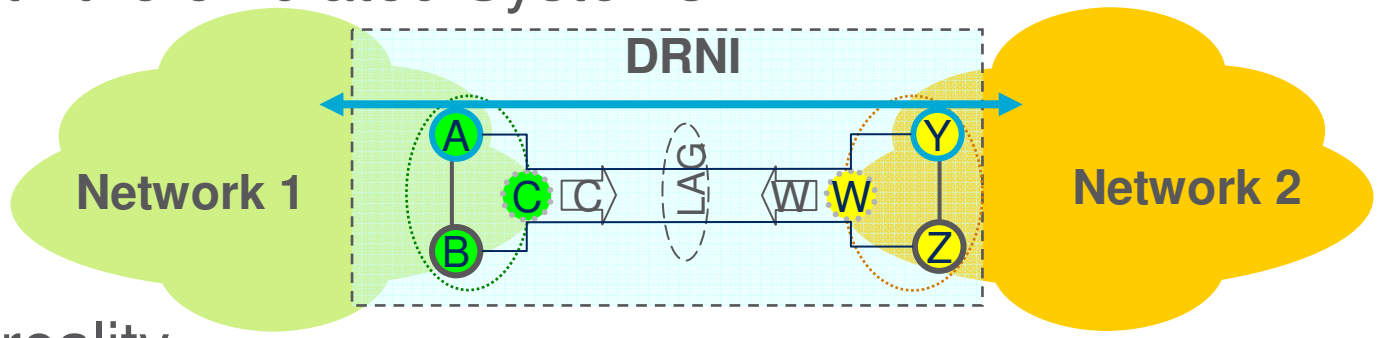
János Farkas, Panagiotis Saltsidis and
Balázs Péter Gerő

Background

- Split Brain related issues have been already discussed, e.g. in
  - http://www.ieee802.org/1/files/public/docs2010/new-nfinn-light-nni-0710-v04.pdf (http://www.ieee802.org/1/files/public/docs2011/axbq-nfinn-graceful-name-change-0511-v1.pdf)
  - http://www.ieee802.org/1/files/public/docs2010/new-haddock-RNNI-split-brain-avoidance-1210-v1.pdf
  - http://www.ieee802.org/1/files/public/docs2011/new-farkas-DRNI-control-0311.pdf
  - http://www.ieee802.org/1/files/public/docs2012/axbq-bragg-split-brain-detection-0912-v00.ppt
- This presentation aims to provide an additional spin from a little bit different angle

Normal operation

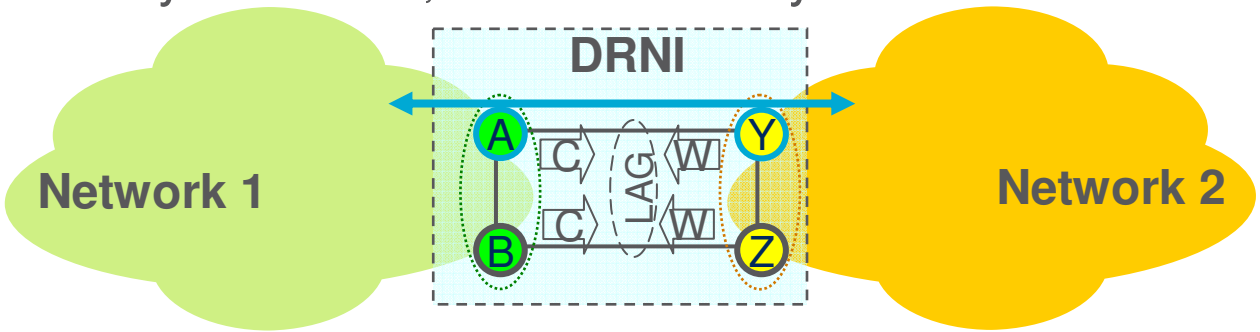
- A and Y are the active gateway for the blue service
- With the emulated Systems



Active gateway

Passive

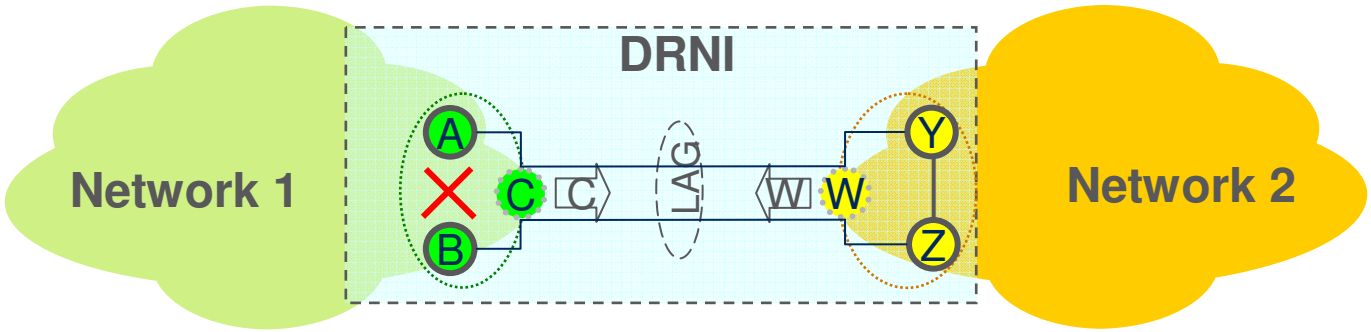
- In reality
  - A and B use System ID C, Y and Z use System ID W in LACPDUs



  - Note that System ID never changes in case of 'perfect' emulation

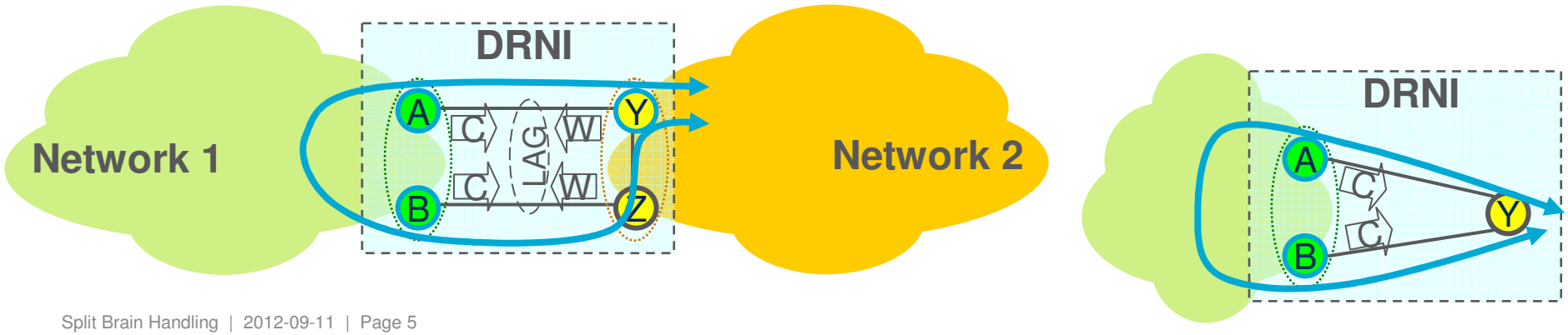
Split brain

- The fellow/mate Systems performing the emulation get disconnected without being aware of the fact that the fellow/mate System is still up and running



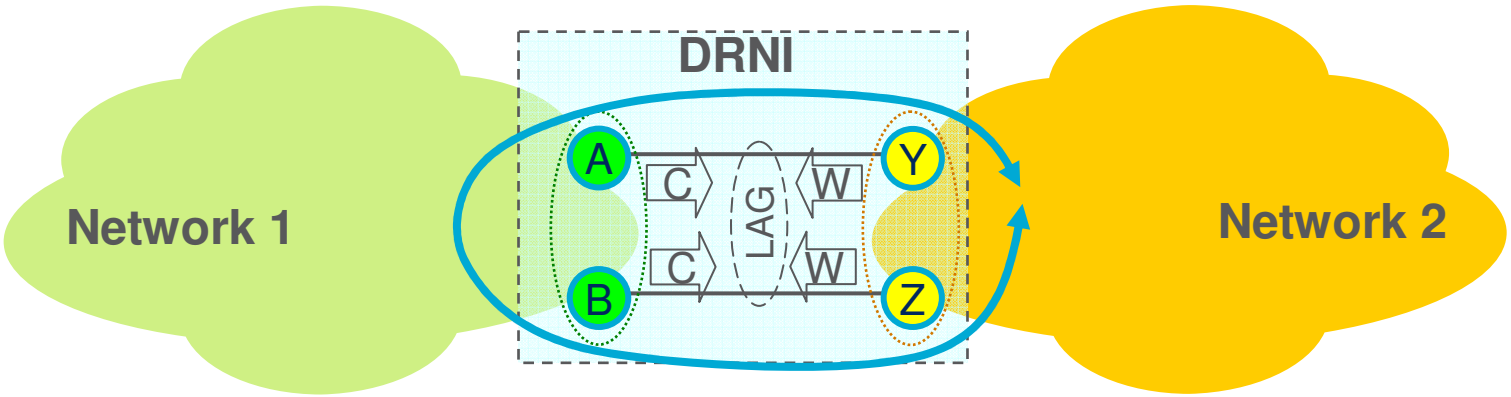
Looping in case of single split brain

- Frames may be looped back due to split brain
- If a System assumes that the fellow/mate System within the same portal is down, then it may become the active gateway for some or for all the services
- Having more than one active gateways may cause looping back of frames if congruency is not enforced



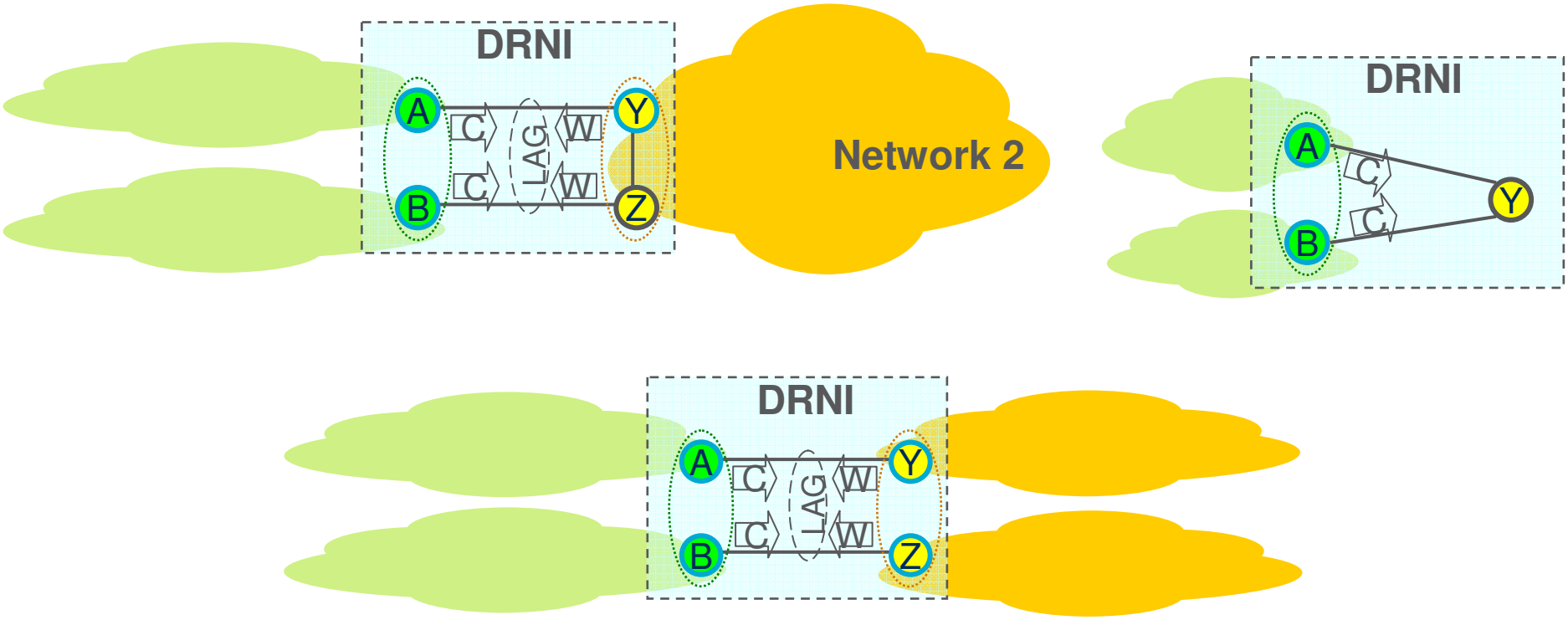
Looping in case of double split brain

- Frames may be looped if both sides of the DRNI are in split brain
- More than one active gateways and more than one active links case frame loop even if congruency is enforced



No looping if split brain = split network

- If it is ensured that split brain only occurs when the network is split then no loop can occur
  - Learning and blackholing issues may appear

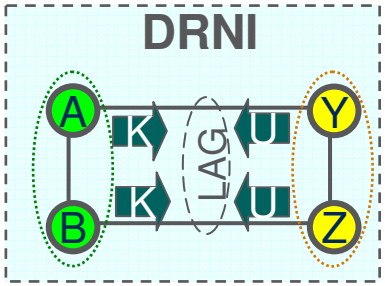


What to do with Split Brain?

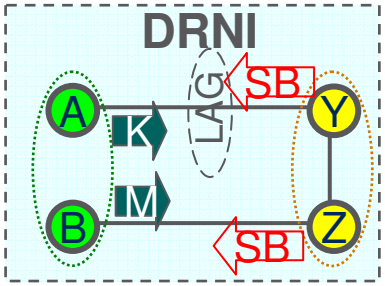
- There were proposals to minimize the probability or even avoid that a split brain situation could occur, e.g.
  - Protected intra portal link
  - Fellow System is reachable as long as network is not split
    - Overlay tunnel can be restored as long as network is not split
    - Monitoring within the own network outside the Portal too
  - Monitoring through the peering party too
- There were also proposals based on changing the System ID if split brain occurs
  - Easy split brain avoidance
  - Hard split brain avoidance
  - Graceful name change
- We might want to handle split brain aiming for ‘perfect’ emulation, i.e. without changing the System ID
- We might also want to do something about the double split brain

Handling of a single split brain

- We need to rely on some other information carried in LACPDUs if we want to preserve the System ID for 'perfect' emulation, e.g. the Actor Key
- Normal operation

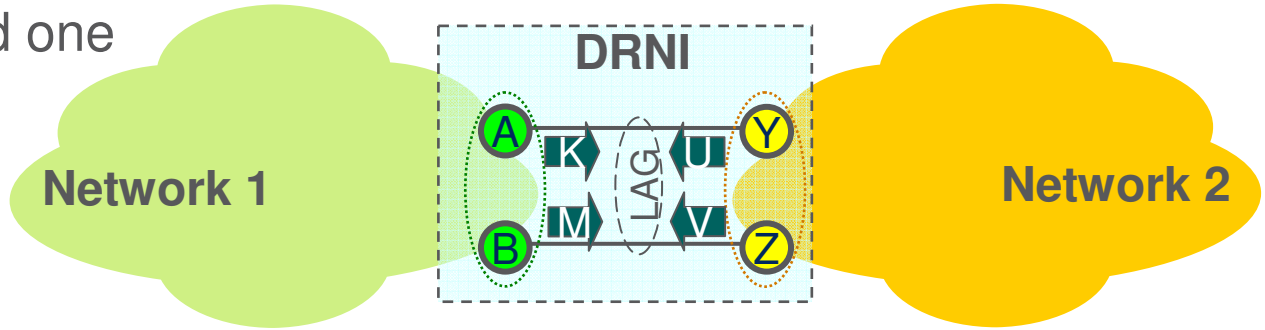


- Failure scenario
  - The highest priority System keeps the common Actor Key (K), all the other Systems change to another Actor Key
  - If the highest priority system went down, then the transition can be graceful (new&old Keys in LACPDUs)
  - The peer receiving multiple Keys declares split brain and only keeps the link belonging to the old Key in LAG

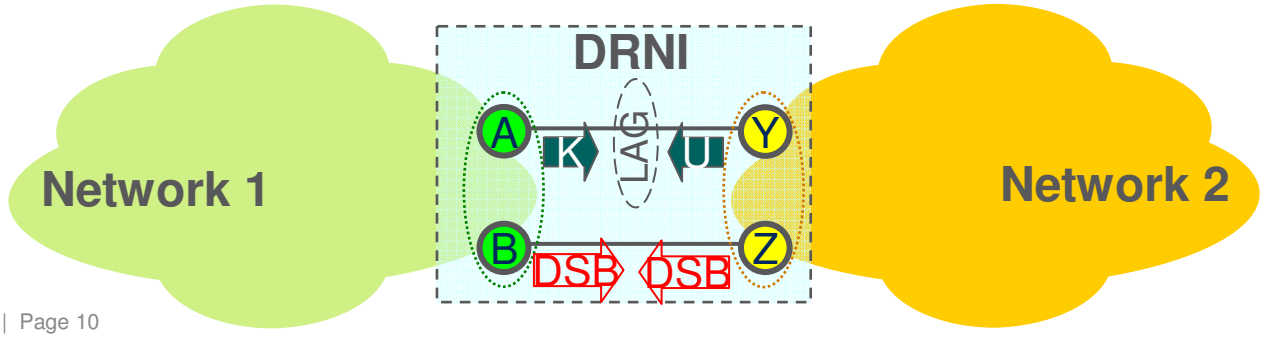


Handling of double split brain

- Double split brain (DSB) scenario can be detected:
  - Fellow System is not reachable within the same portal
  - The information received in the LACPDU from the peer differs from the expected one



- If DSB declared, then the link is excluded from LAG



Summary

- Avoid the occurrence of split brain situation as much as possible
- Handle split brain if occurs
  - Might slightly increase the configuration burden, depending on how safe are we want to be

ERICSSON