Finn Model for IB-BEB DRNI

Rev. 1

Norman Finn

nfinn@cisco.com

Model parameters

- On what basis are frames assigned to Gateways?
- What stateful functions must be distributed?
- Is each function distributed active-active, or active-standby?
- On what basis are the stateful function distribution choices made?
- How many layers of DRNI Distributed Relay are required?

Assigning frames to Gateways

- It depends on the network protection protocol used.

  For B-VLANs subject to control by SPB or MSTP, gateway choice is **per-B-VLAN**.

  For B-VLANs subject to control by PBB-TE, for each protection N-tuple, one path is assigned to each physical Portal System.

- This is necessary to prevent MAC address learning problems (where learning is employed) and to ensure separation of protected paths (where protection switching is employed).

What stateful functions are distributed?

- Obviously, the LACP protocol must be distributed in order to present a single face to the other end of the DRNI.
- All of the services using a single **protection set**, consisting of a selection state machine and two (or more) MEPs each monitoring a path (whether ESP or Segment), are co-resident in a single physical Portal System and **not distributed**.
- Services belonging to different protection sets can be distributed among different physical Portal Systems.
- If protection switching is not employed, then there are no stateful functions to distribute.

Active-active or active-standby?

- **All** protection sets are **active-standby**.
- Every protection set can have a standby in another Portal System (but is not required to have one).
- When a transfer is made of a protection set from one Portal System to another, basic state (which path is selected, which paths are happy and which are failed) must be carried over. Other state (e.g. frame counters) are not transferred, and a glitch is seen by the other end of the protection set.
- Aside from LACP, which of course, is active-active, no other state machines need by distributed.

How to make stateful distribution choices

- Whole protection sets are distributed, in order to minimize the amount of state that must be passed in a hot-standby transfer.
- Each protection set resides in a single Portal System so that no MEP or protection selection machine needs an active-active emulation protocol defined.

  In particular, distributed CCM connectivity and CCM frame loss measurement counters need not be simulated.
- If protection switching is not employed, then any distribution choice is fine, because the CBP has no stateful entities visible to outsiders except management, which must be aware of reality, anyway.

Distributed Relay Layers

- Two layers of Distributed Relays are required.
- These layers may or may not require some kind of encapsulation, depending on whether protection switching is used, and on how functions are distributed.

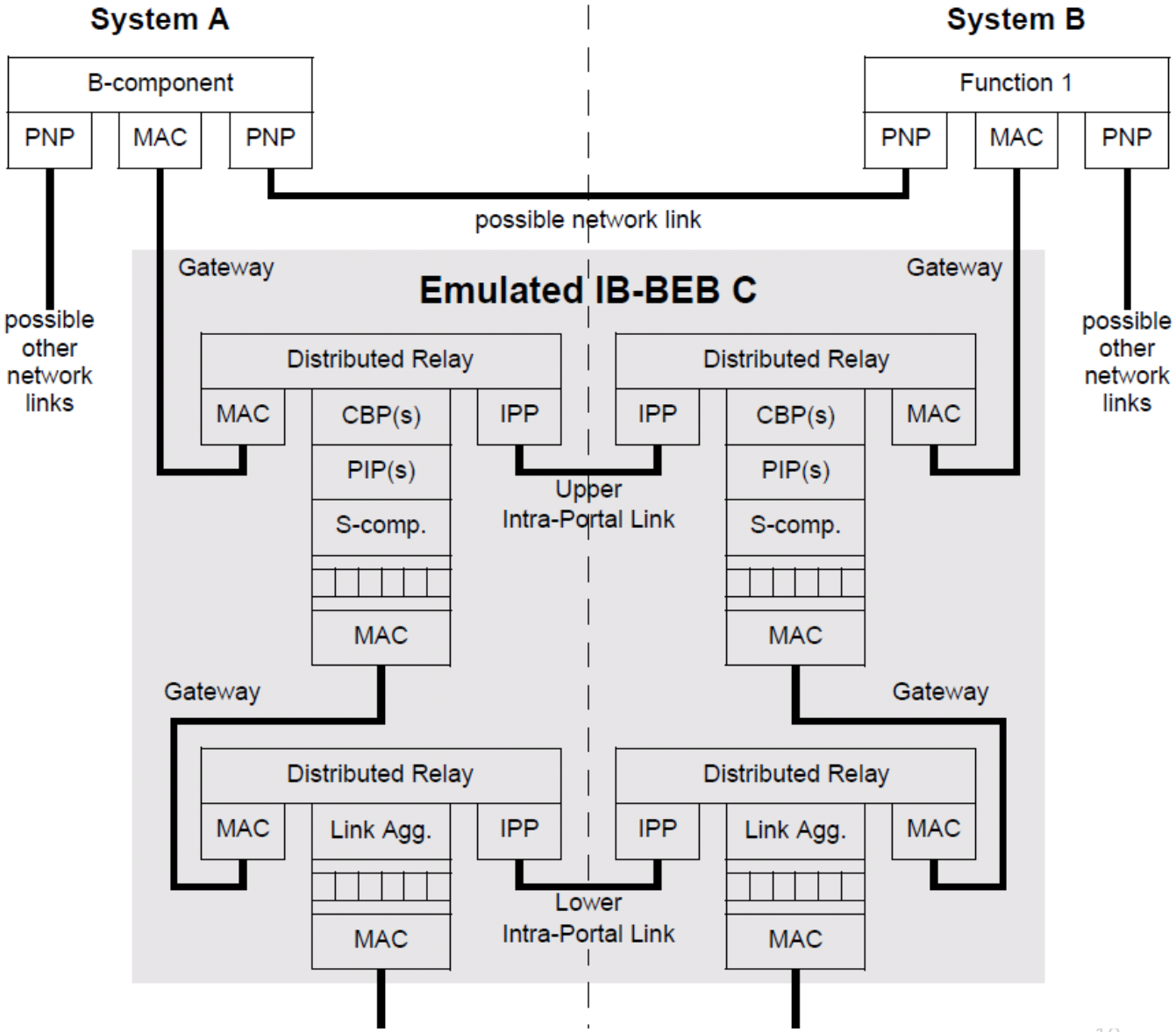
Bandwidth issues

Setup

- In the following examples, we will assume that whole CBPs are distributed, one CBP in one Portal System, and one in the other.
- A CBP can be split, as long as no protection pair is split.

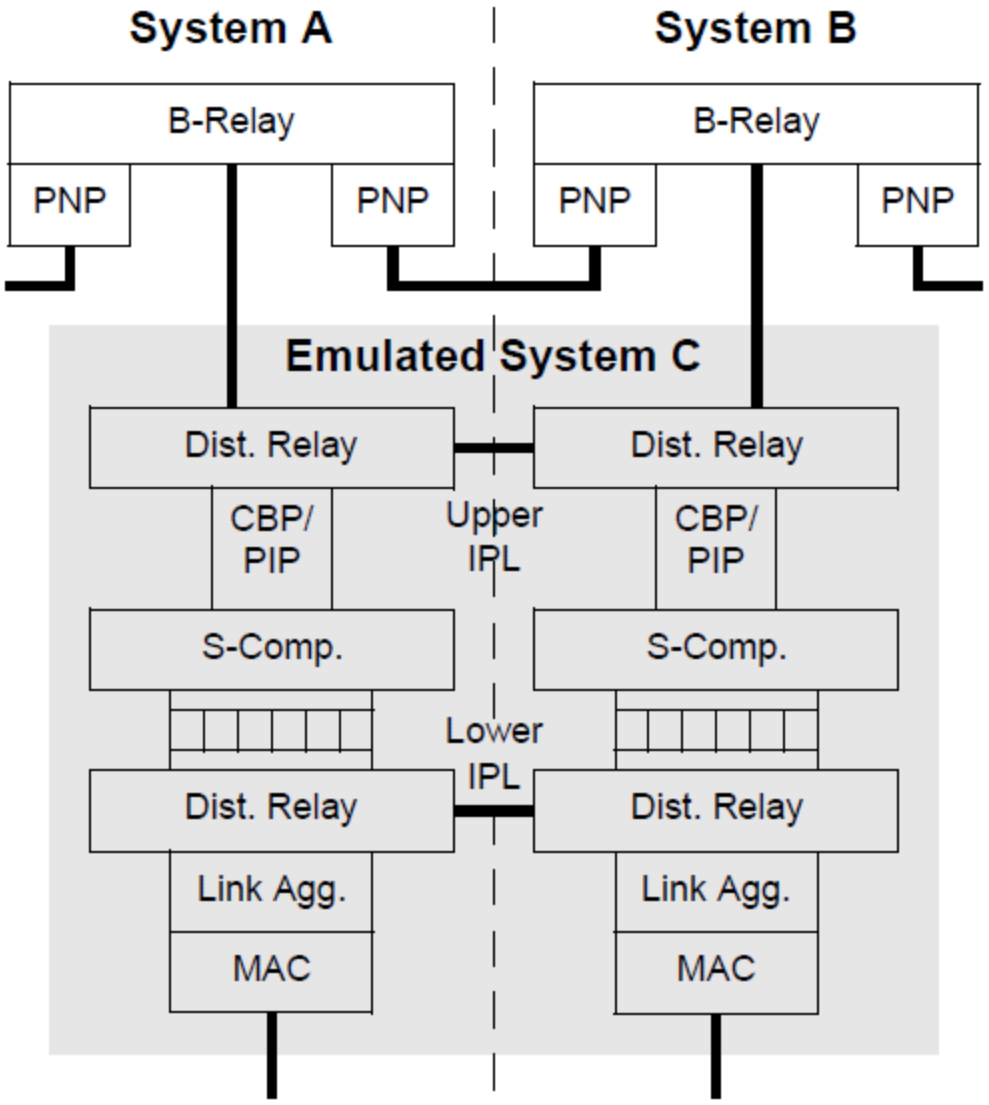
Finn model: IPL bandwidth

- Let us use a simpler version of the reference diagram:



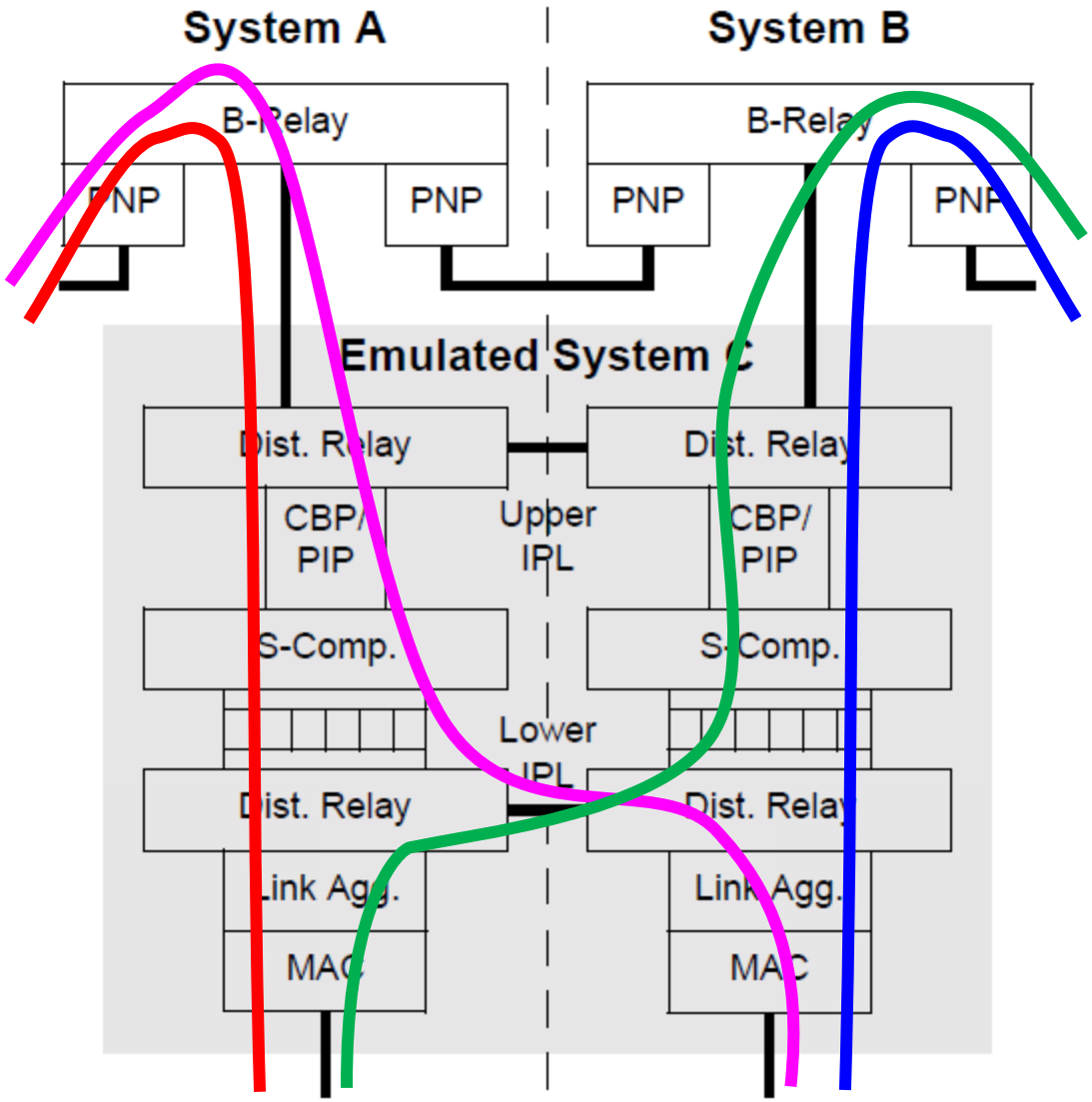
Finn model: IPL bandwidth

- Let us use a simpler version of the reference diagram:
- (We'll worry about the lower MEP placements later.)



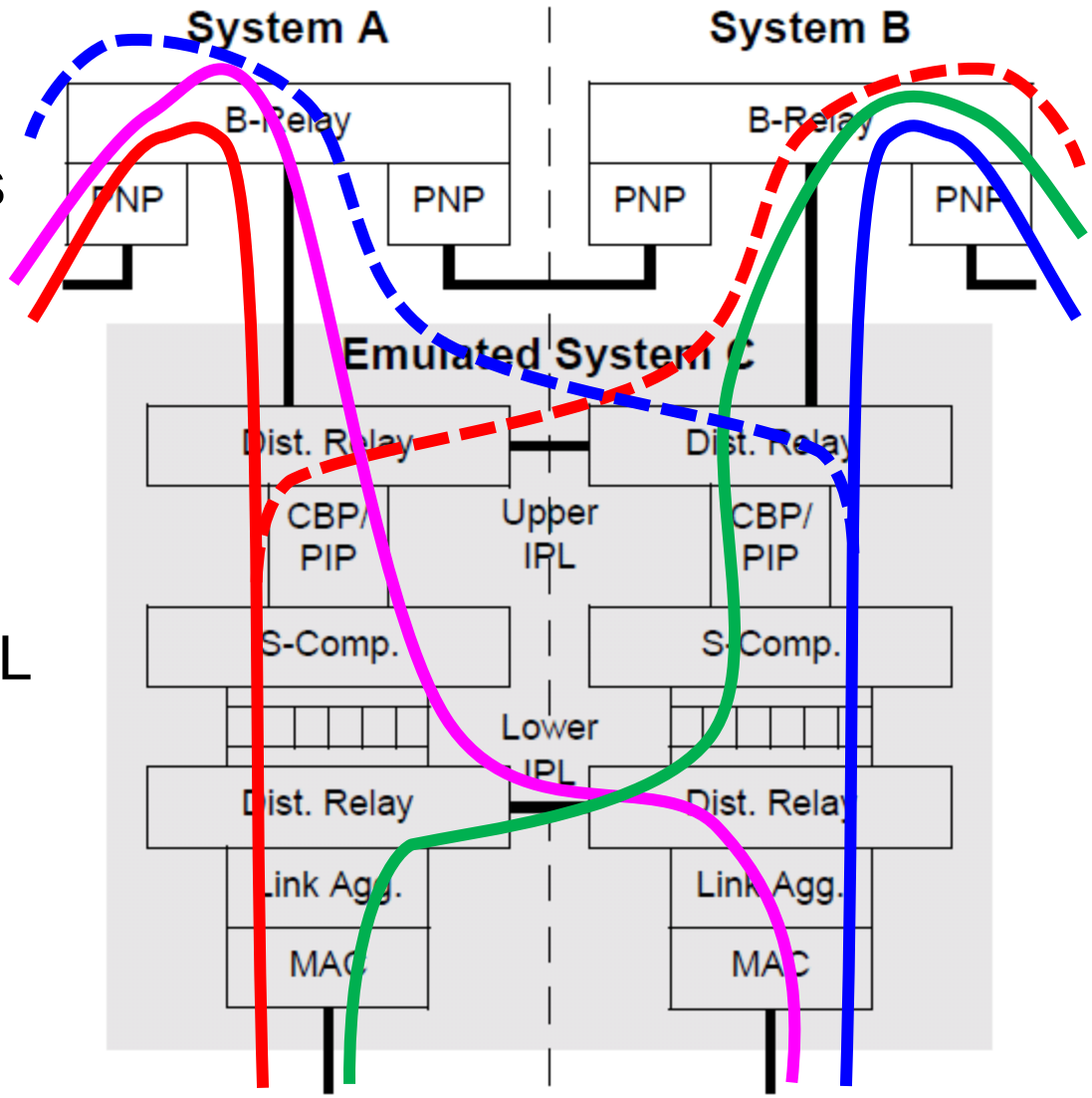
Finn model: IPL bandwidth

- You would normally configure **two CBPs** for:
  - **Left flows**
  - **Right flows**
  - **A-B criss-cross**
  - **B-A criss-cross**
- **Note that there is no excess IPL traffic.**



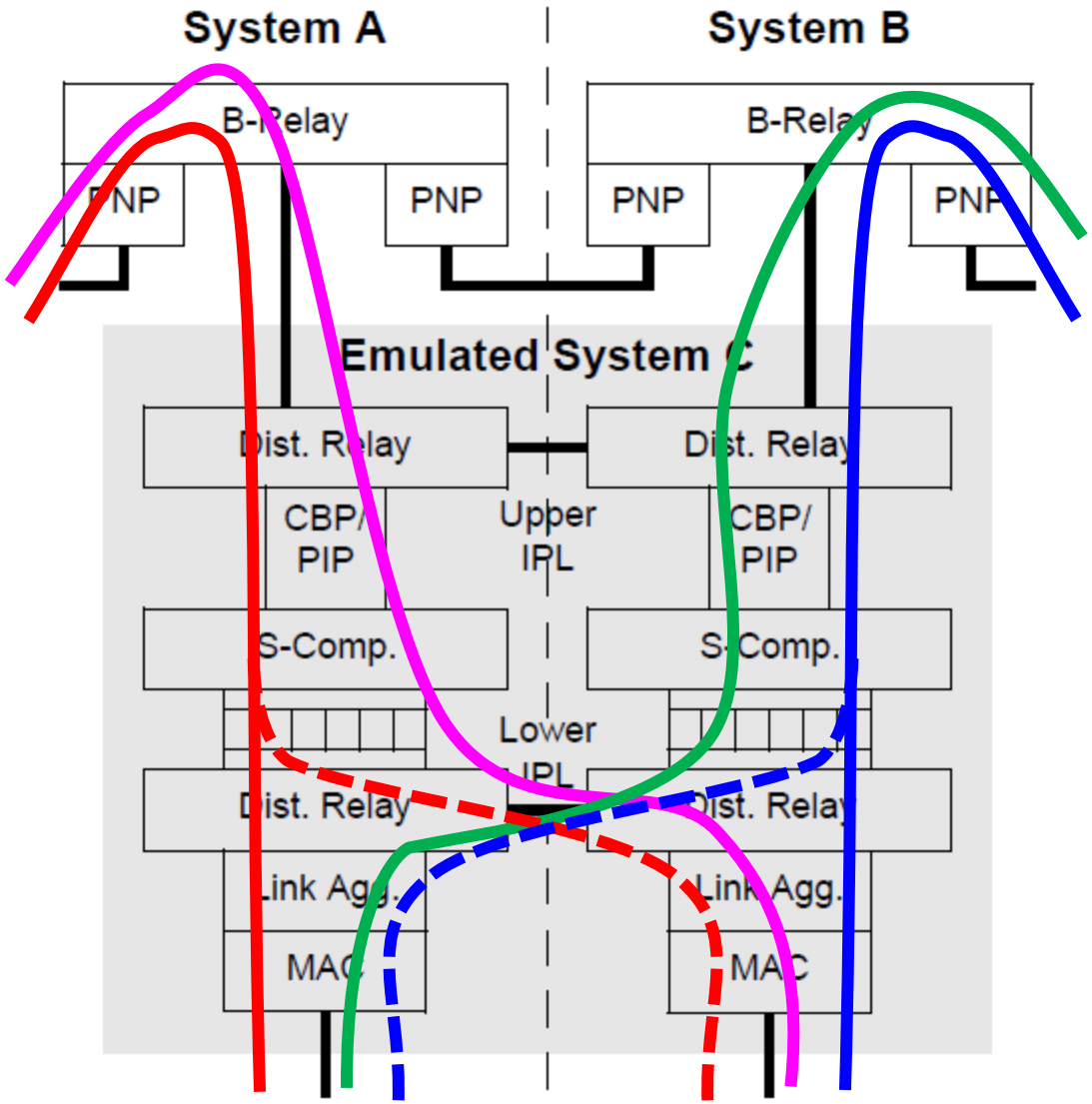
Finn model: IPL bandwidth

- A failure of one of the (presumably many) ESPs in the **Left** or **Right** flows would trigger the use of an alternate ESP that would use the Upper IPL.
- **This is not "excess bandwidth".** The one IPL trip is required for connectivity.



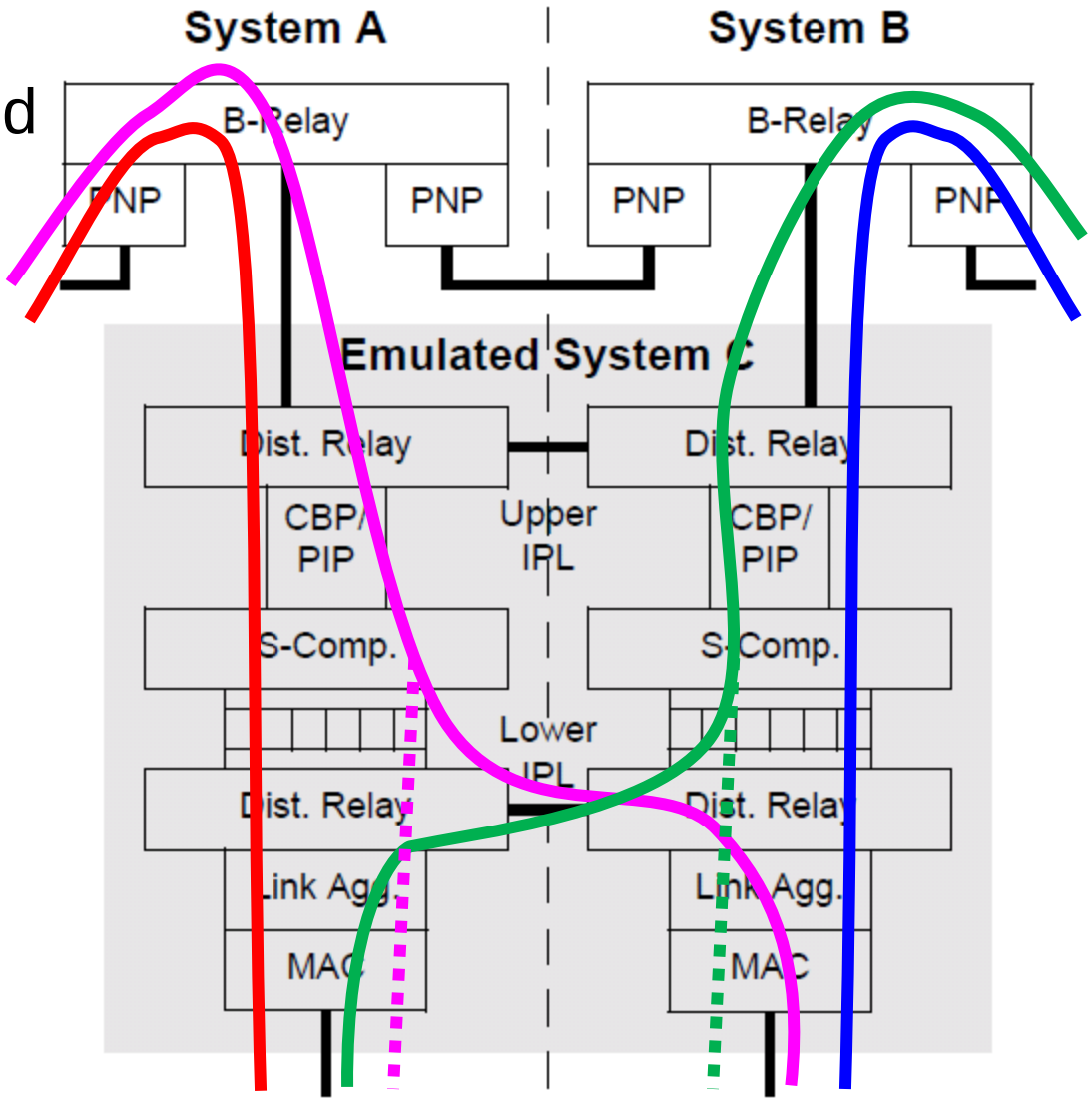
Finn model: IPL bandwidth

- Similarly, a failure of a DRNI link or a movement of an S-VLAN to another DRNI link on the **Left** or **Right** flows does not generate any excess IPL bandwidth.



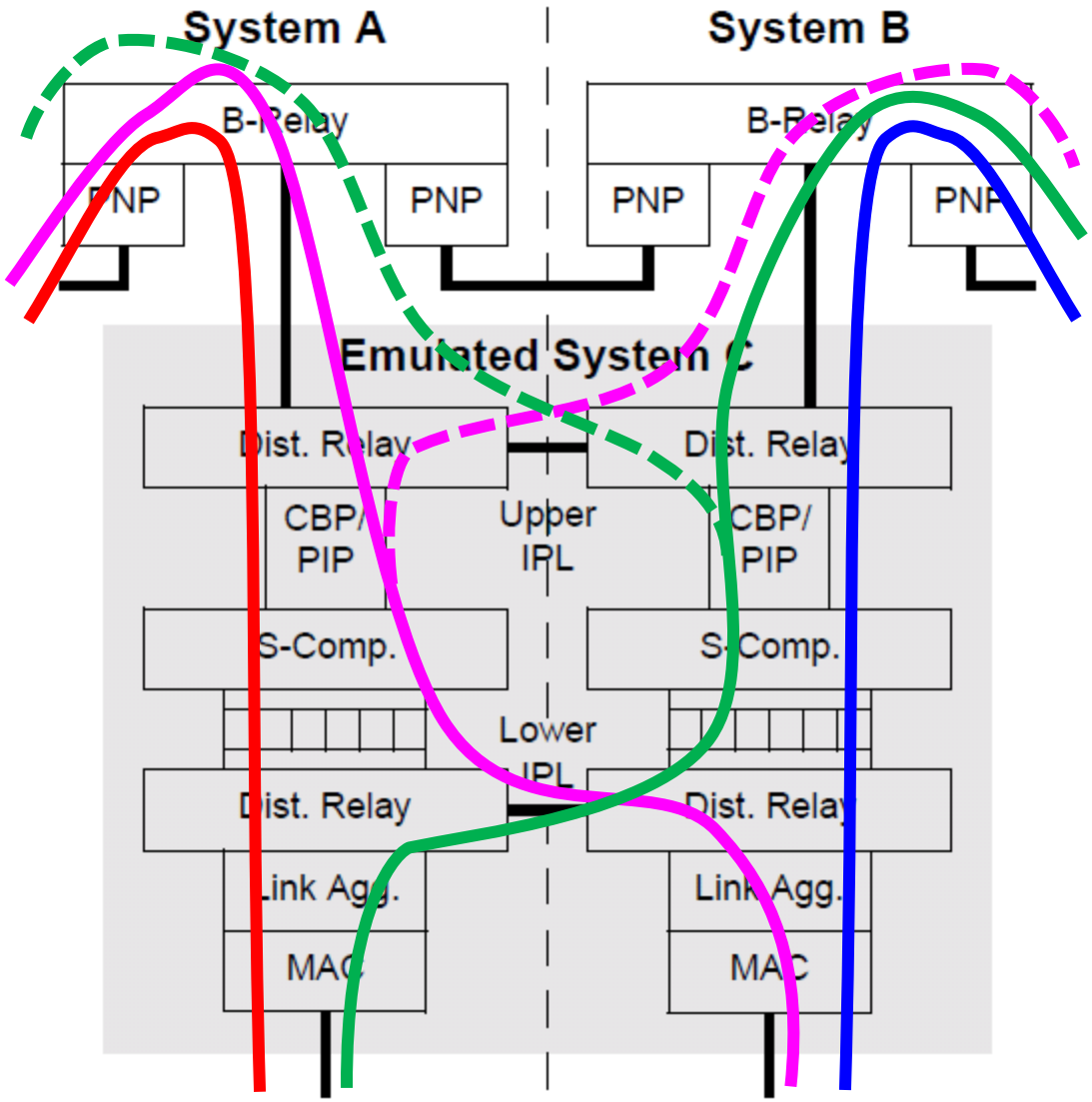
Finn model: IPL bandwidth

- A failure at the “better” end of a **criss-cross** link actually **improves** the situation.



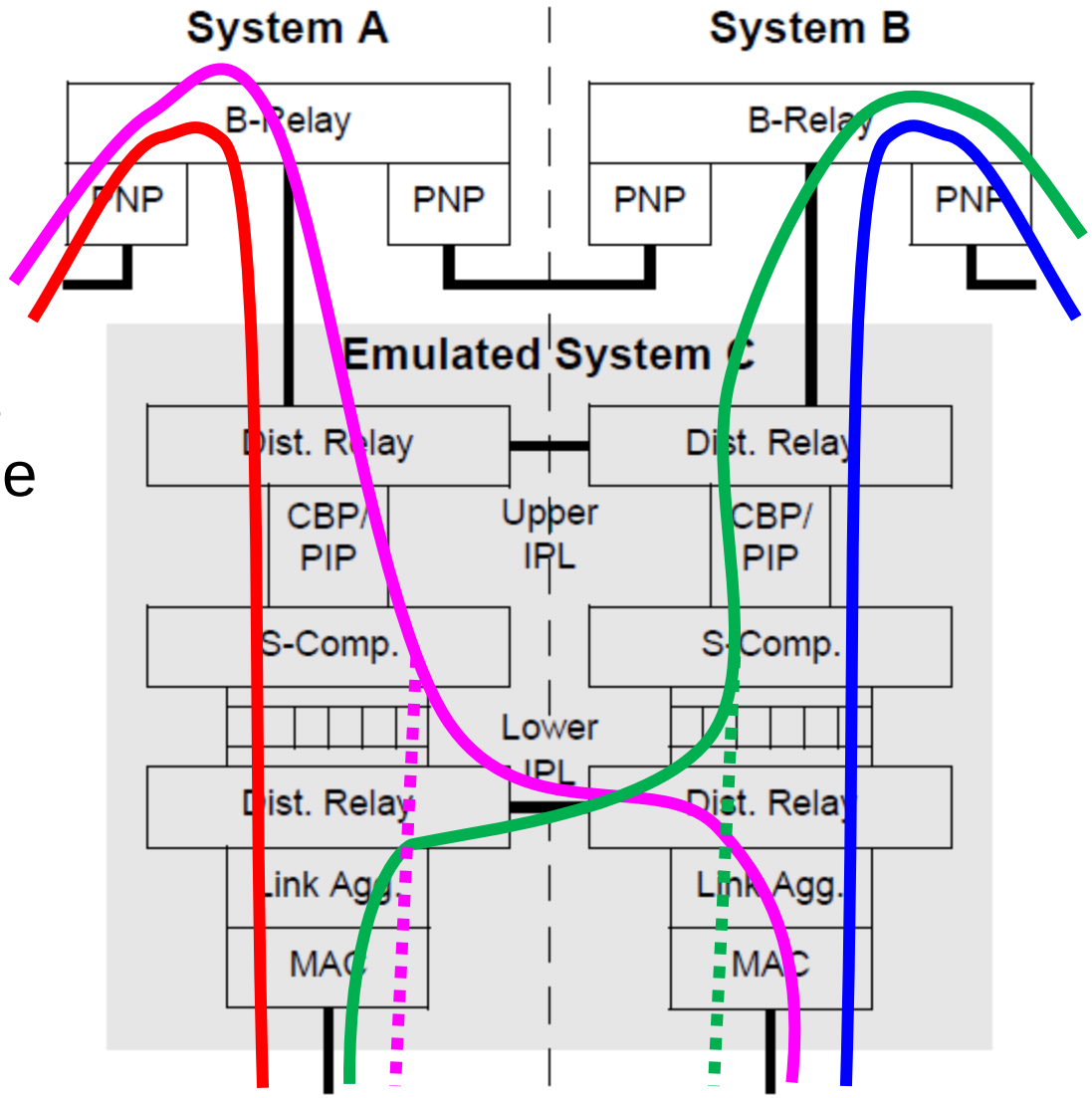
Finn model: IPL bandwidth

- **Only** a failure at the "worse" end of a **criss-cross link** creates **excess IPL bandwidth**, in that frames in the failed service make two trips between the Portal Systems.



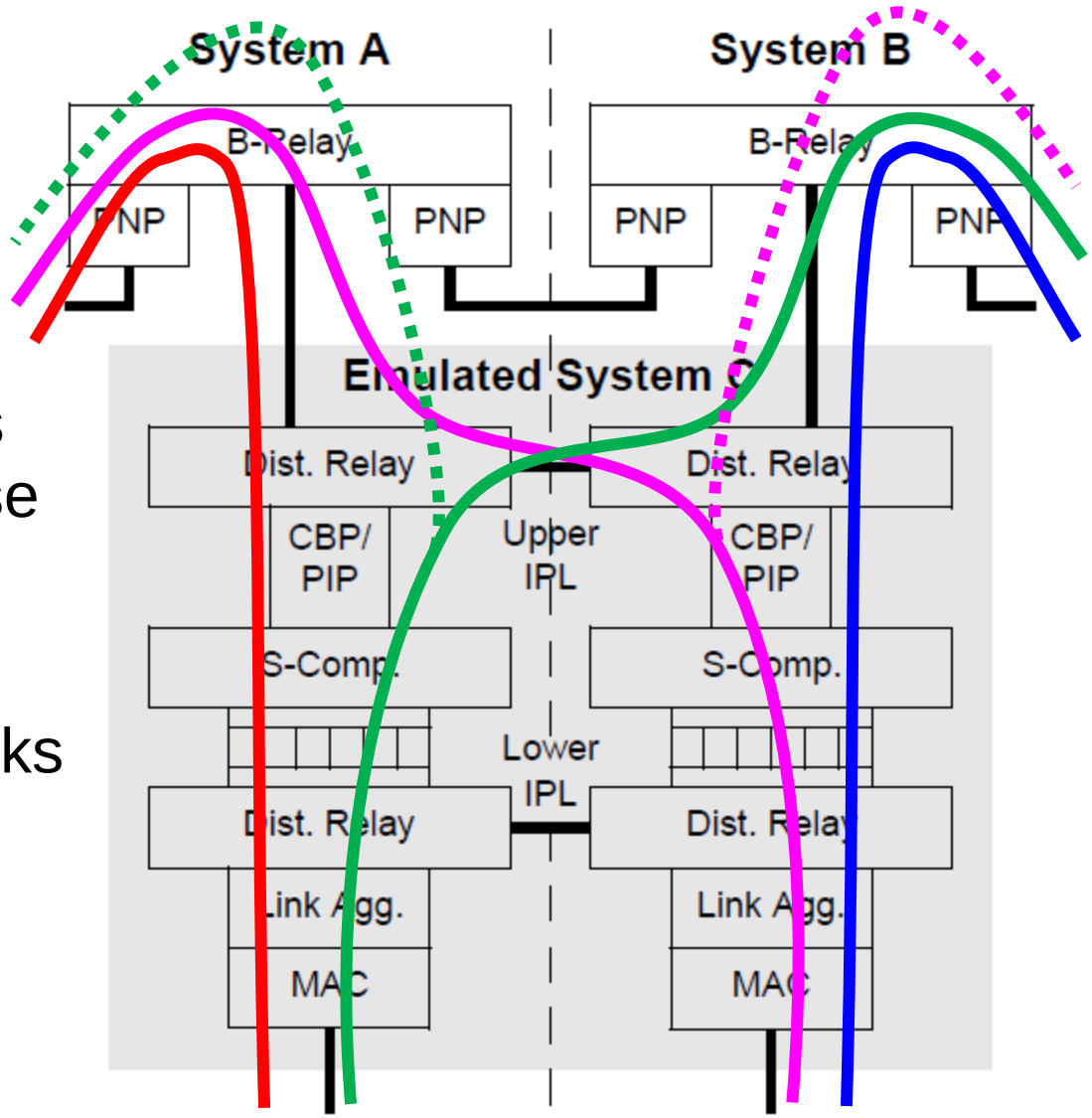
Finn model: IPL bandwidth

- If one expects far more failures at one end that at the other end (network end vs. LAG end), then one should arrange one's "**criss-cross**" CPBs to use the end that is expected to fail most often.



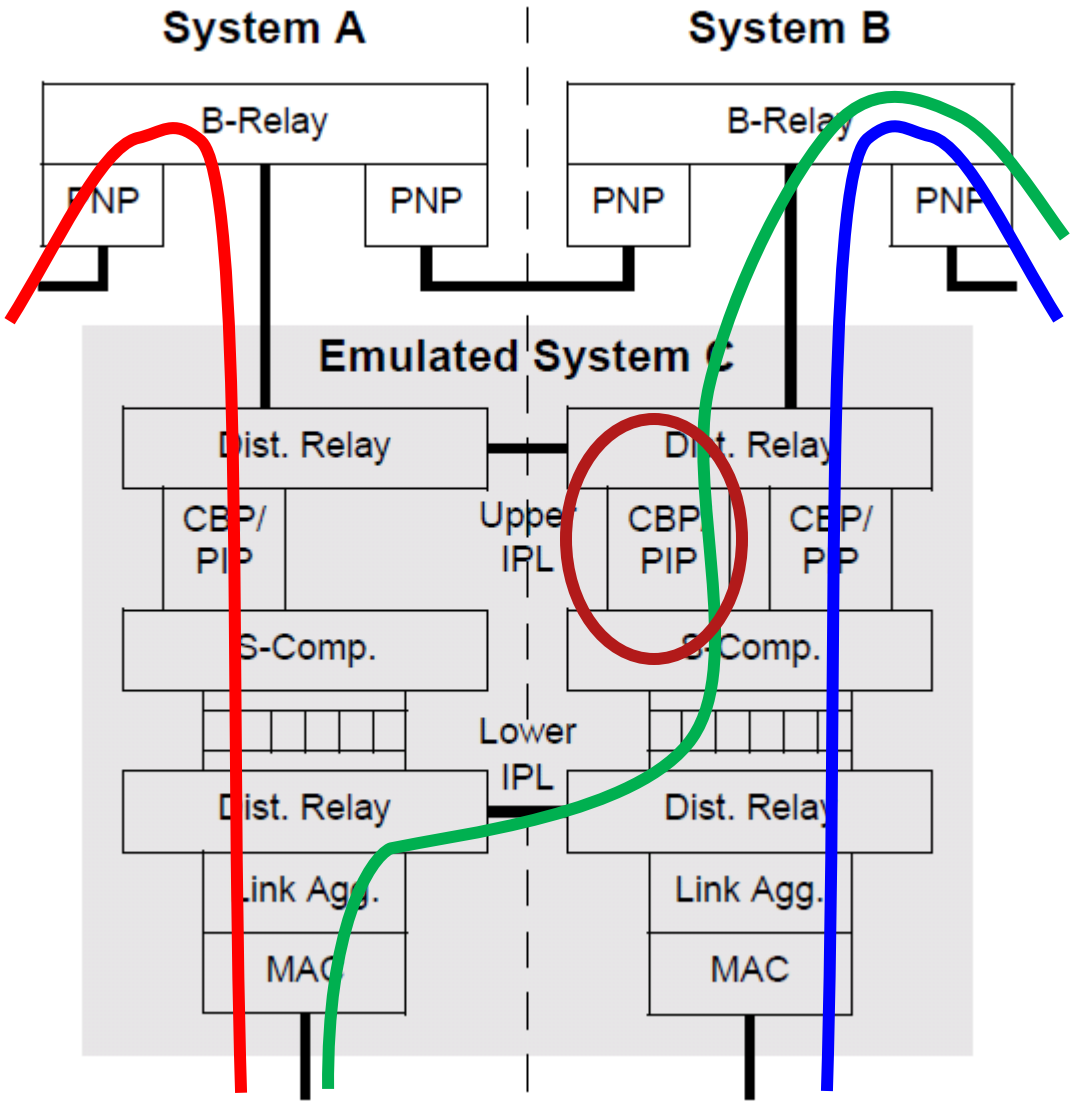
Finn model: IPL bandwidth

- If one expects far more failures at one end that at the other end (network end vs. LAG end), then one should arrange one's "**criss-cross**" CPBs to use the end that is expected to fail most often.
- Presumably, the DRNI links fail more often, so **this is not the typical picture**.



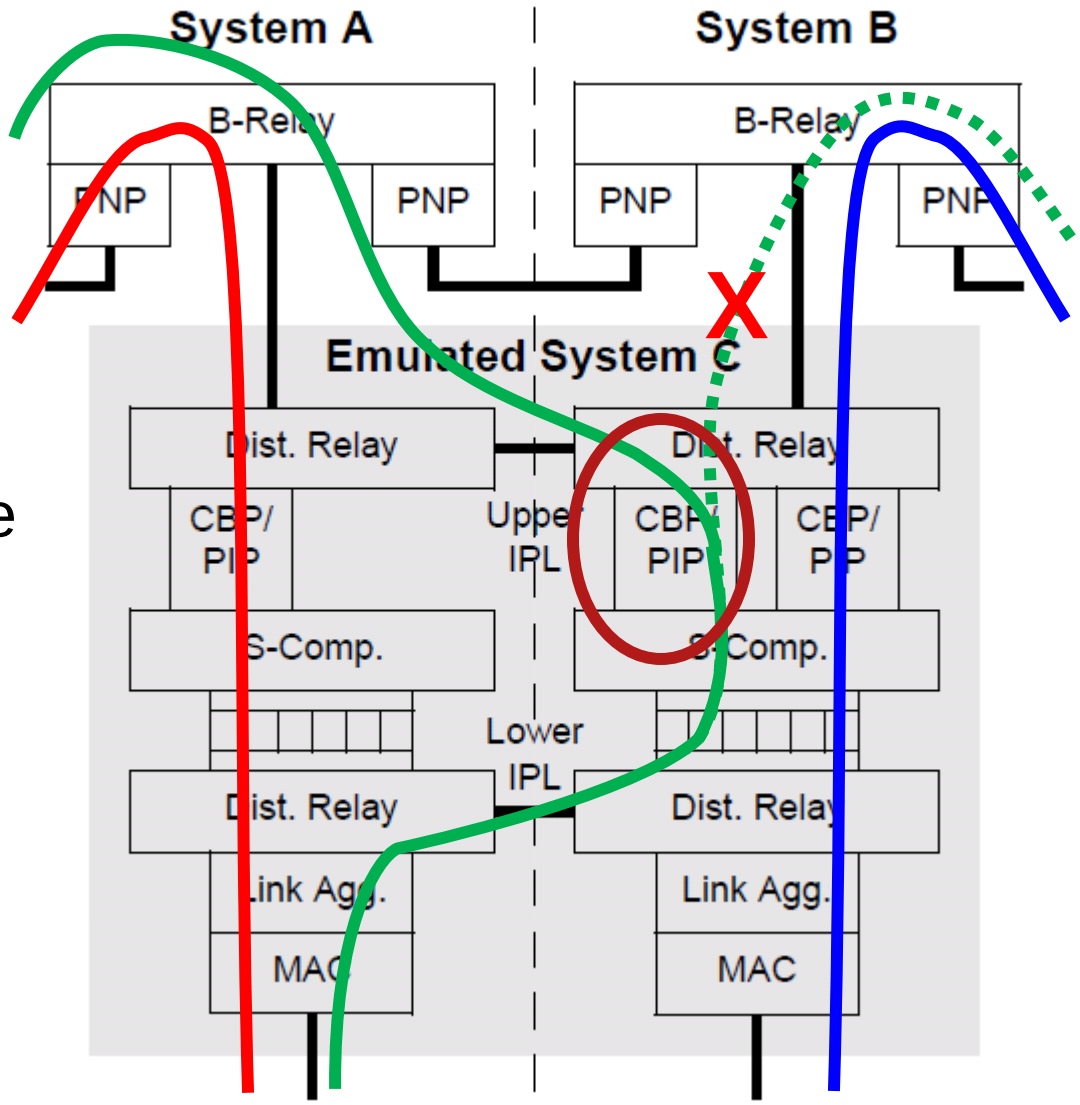
Finn model: IPL bandwidth

- If one is really worried about excess IPL traffic, one can employ **extra CBPs** for only the **criss-cross** traffic.



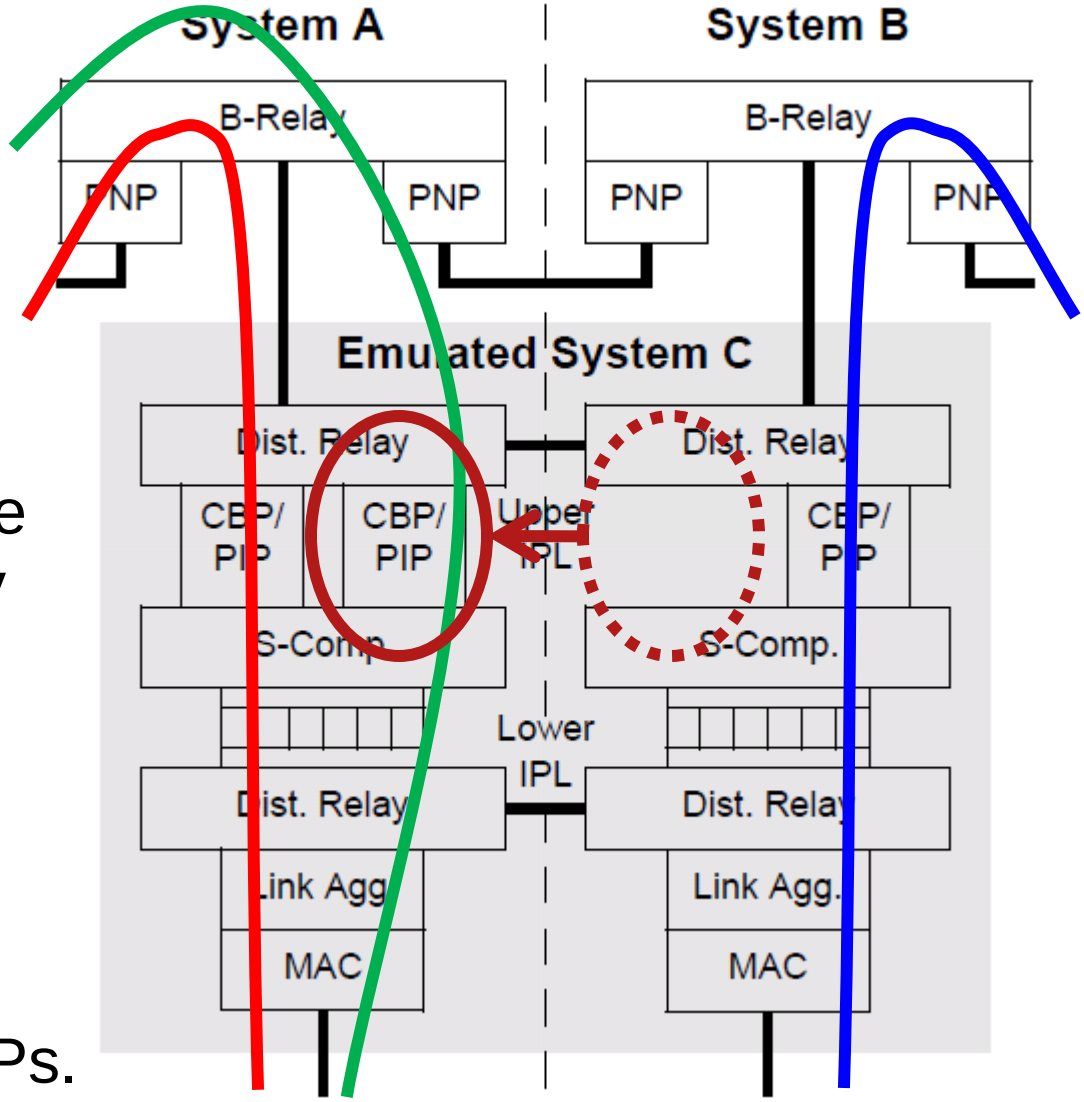
Finn model: IPL bandwidth

- If one is really worried about excess IPL traffic, one can employ **extra CBPs** for only the **criss-cross** traffic.
- If the traffic in one of these CBPs is making too many hops across the IPLs,



Finn model: IPL bandwidth

- If one is really worried about excess IPL traffic, one can employ **extra CBPs** for only the **criss-cross** traffic.
- If the traffic in one of these CBPs is making too many hops across the IPLs, switch over to a **Standby CBP+MEP** to correct the problem
- Of course, fine-grained control requires more CBPs.



Finn model: IPL bandwidth

- The other models avoid this situation, at the cost of Shared MEPs.
- However, we should note that, because there are no Shared mode MEPs, and because the sharing protocol necessarily adds to the failover time, the **failover times** of the Finn model do not require any more changes in state or alterations of the forwarding tables than for the non-DRNI case, and hence **will be faster** than for the other models.