amsix
amsterdam internet exchange

AMS-IX version 4

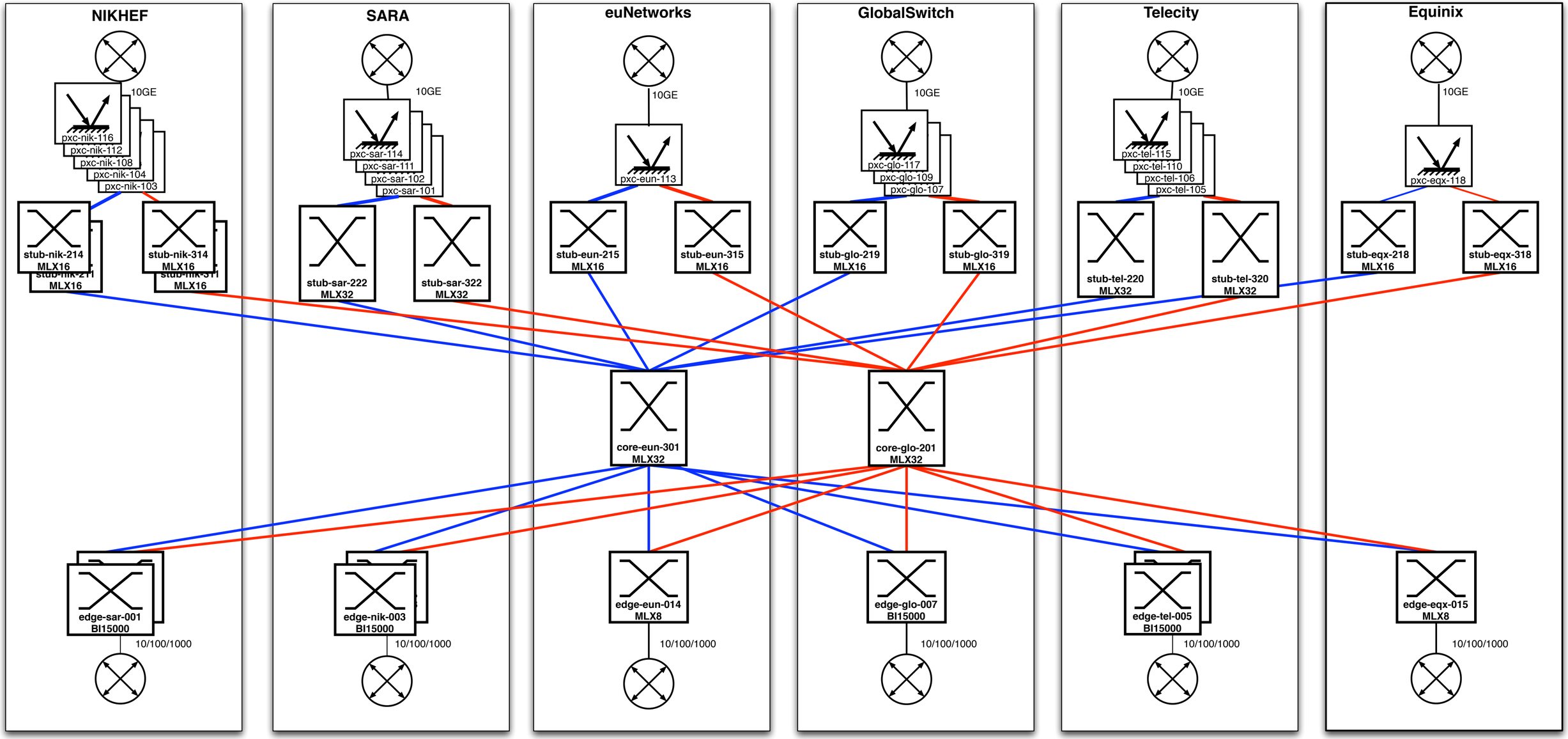
Migration to an MPLS/VPLS based platform

Henk Steenman
Henk.Steenman@ams-ix.net

Apricot 2010
March 2 2010

Overview

- AMS-IX version 3
  - Short overview
  - Bottlenecks and limitations
- AMS-IX version 4
  - The MPLS/VPLS platform
- AMS-IX v3 to v4 migration
  - Current status



June 2009 situation before start of migration

AMS-IX version 3

AMS-IX version 3

Characterization

- Two networks: one active at any moment in time
- Selection of active network by VSRP (Brocade proprietary
  - Inactive network core-switch blocks ports to prevent loops
- PSCD, photonic switch control daemon
  - AMS-IX developed software to act on VSRP traps and manage PXCs

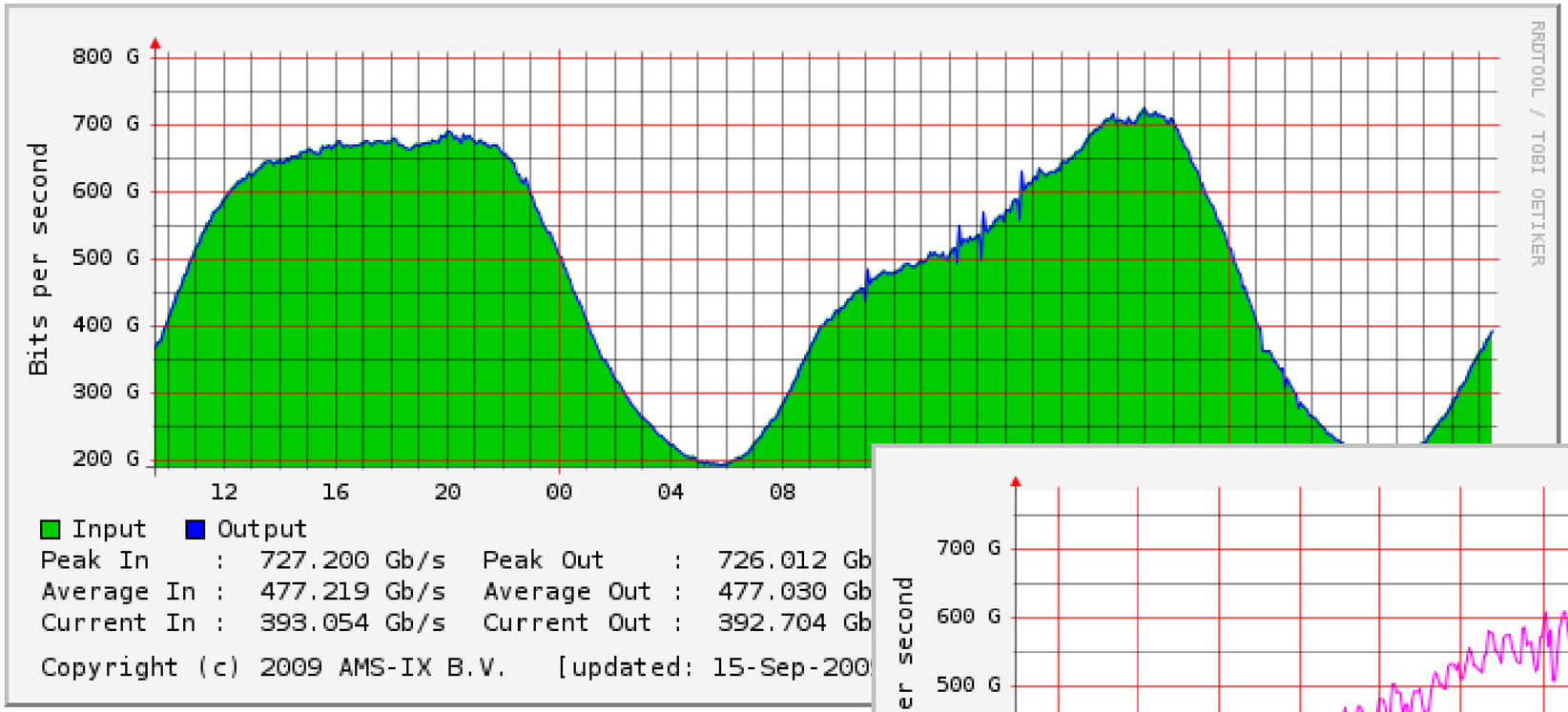
AMS-IX version 3

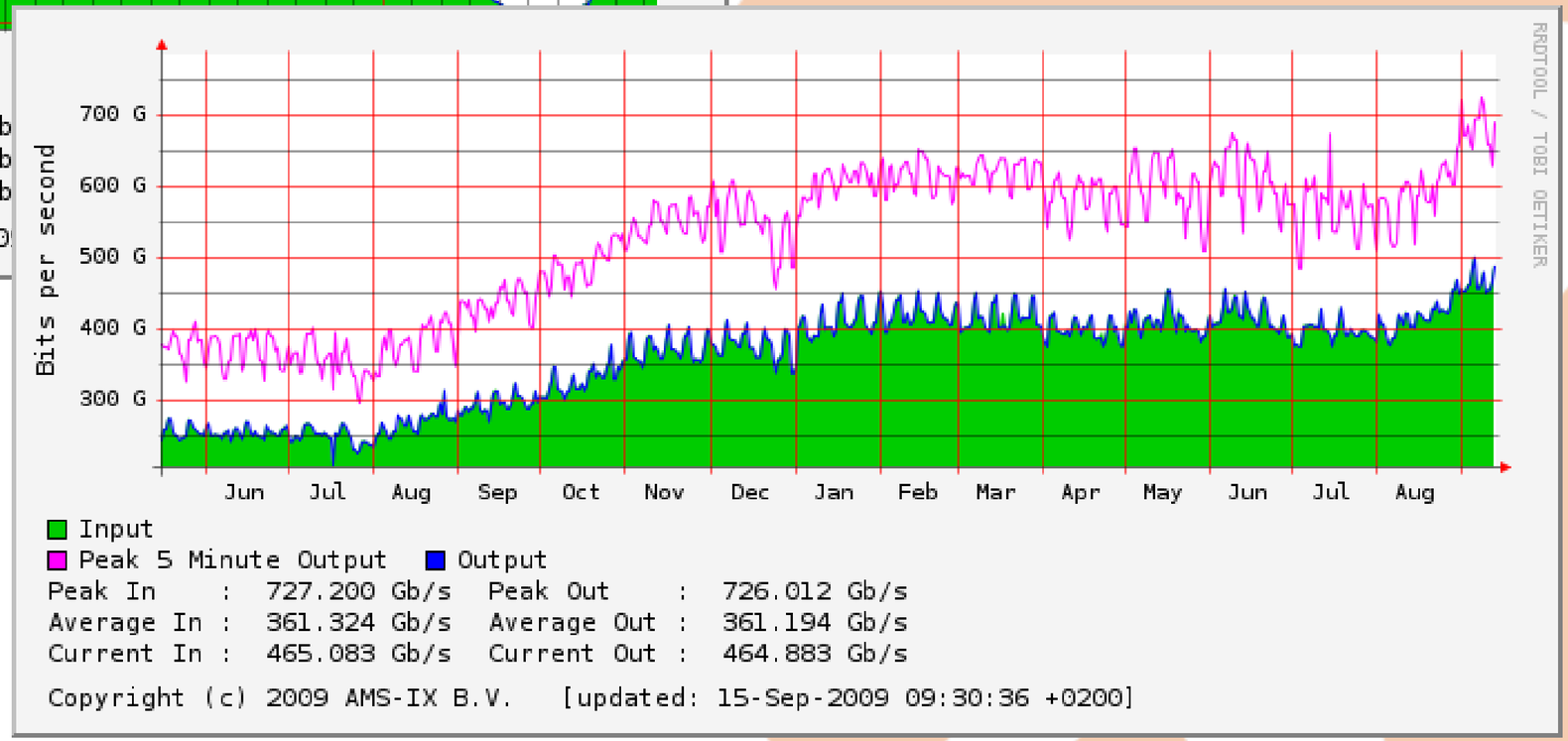
Characterization

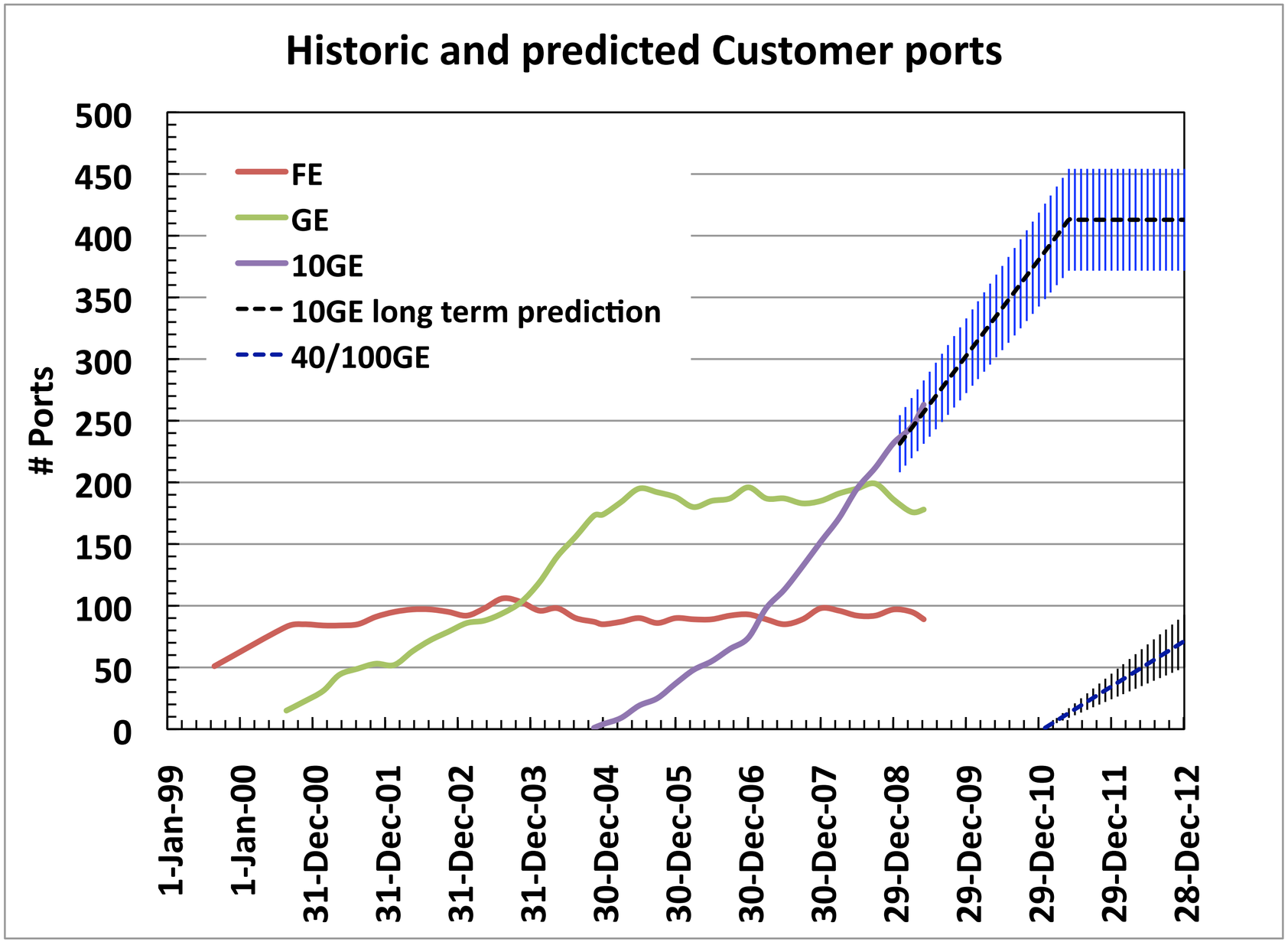
- E, FE and (N *) GE connections on BI-15k or RX8 switches
- (N *) 10GE connections resilient connected on switching platform (MLX16 or MLX32) via PXCs
- Brocade “port security” on customer interface to enforce one MAC per port rule for loop prevention

AMS-IX customer traffic

Daily and yearly traffic Aug 2009







Long Term 10 and 40/100G customer Port Predictions

Traffic and Port Prognoses

AMS-IX version 3

Bottlenecks and Limitations

- Core switches (MLX32, 128 10GE line rate) fully utilized
  - Limits ISL upgrade
  - Summer 2009 no substantial bigger switches on the market
- Platform failover introduces short link-flap on **all** 10GE customer ports. In few (but increasing) cases this leads to BGP flapping
  - With more and more 10GE customer ports (268 June 2009) impact on overall platform stability becomes larger and larger
- Growth of number of 10G connections and 10GE customer LAG size requires larger 10GE access switches
  - Smaller switches => less local switching => larger ISL trunks

AMS-IX version 4

An MPLS/VPLS based Exchange platform

AMS-IX version 4

Requirements

- Scale the core to at least double amount of ports (Q2/3 2009)
- Keep resilience in platform and 10GE access but reduce impact on failover.
- Increase amount of 10GE customer ports on access switches
  - More local switching
- Migrate to single architecture platform
  - Reduce management overhead
- Use future proof (3 to 5 years) hardware that allows upscaling to high-density 10GE (2010) and 40/100GE (end 2010, early 2011)

SARA
NIKHEF
euNetworks
GlobalSwitch
Telecity
Equinix
10GE
pxc-sar-114
pxc-sar-111
pxc-sar-102
pxc-sar-101
pxc-nik-116
pxc-nik-112
pxc-nik-108
pxc-nik-104
pxc-nik-103
pxc-eun-113
pxc-glo-117
pxc-glo-109
pxc-glo-107
pxc-tel-115
pxc-tel-110
pxc-tel-106
pxc-tel-105
pxc-eqx-118
stub-sar-222 MLX32
stub-sar-322 MLX32
stub-nik-221 MLX32
stub-nik-321 MLX32
stub-eun-215 MLX16
stub-eun-315 MLX16
stub-glo-213 MLX16
stub-glo-313 MLX16
stub-tel-220 MLX32
stub-tel-320 MLX32
stub-eqx-218 MLX16
stub-eqx-318 MLX16
core-eun-301 MLX32
core-eun-302 MLX32
core-glo-201 MLX32
core-glo-202 MLX32
edge-sar-001 MLX8
edge-nik-003 MLX8
edge-eun-008 MLX8
edge-glo-007 MLX8
edge-tel-005 MLX8
edge-eqx-009 MLX8
10/100/1000

Complete MPLS/VPLS topology

AMS-IX version 4

AMS-IX version 4

Overview

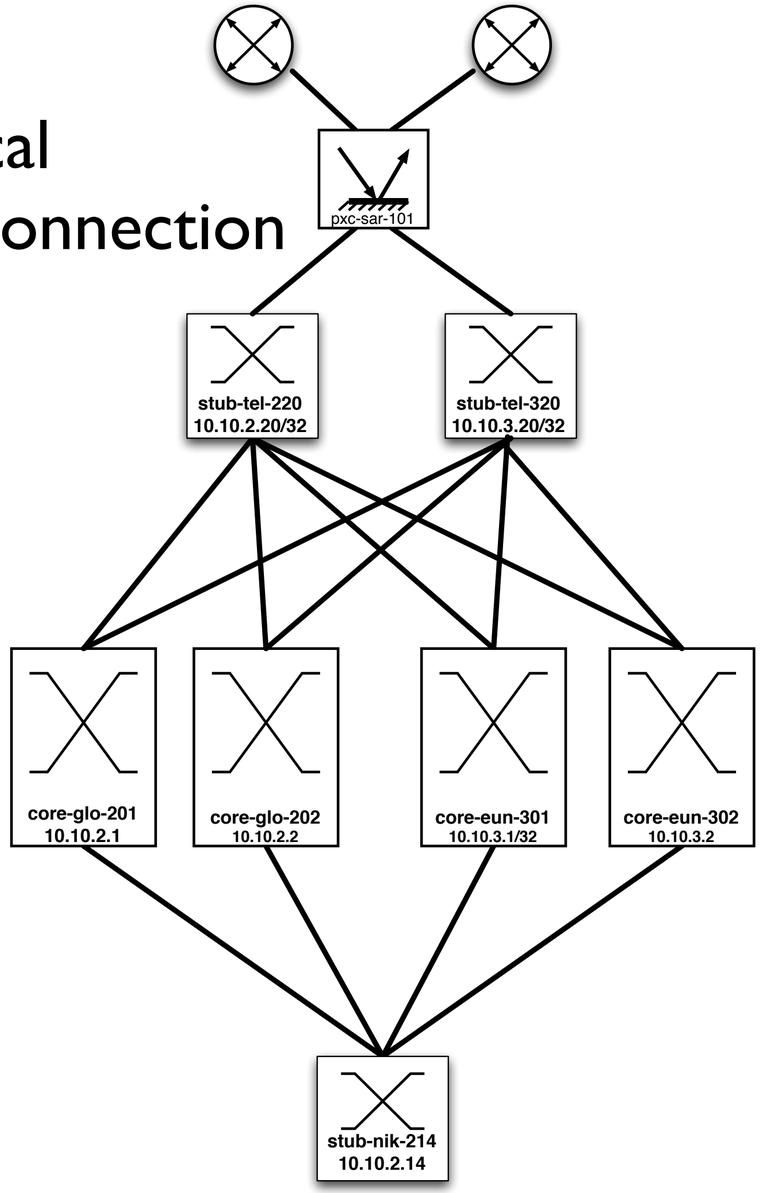
- MPLS/VPLS-based peering platform
- Scaling of core switches by adding extra switches in parallel
  - 4 LSPs between each pair of access switches
  - Load balancing of traffic over 4 LSPs between each pair of access switches
- Retain 10GE access switch resilience
  - Keep 10GE customer connection on PXC
  - No need for complete platform failover anymore
    - Local impact only (single pair of access switches on a site)

AMS-IX version 4

Characterization

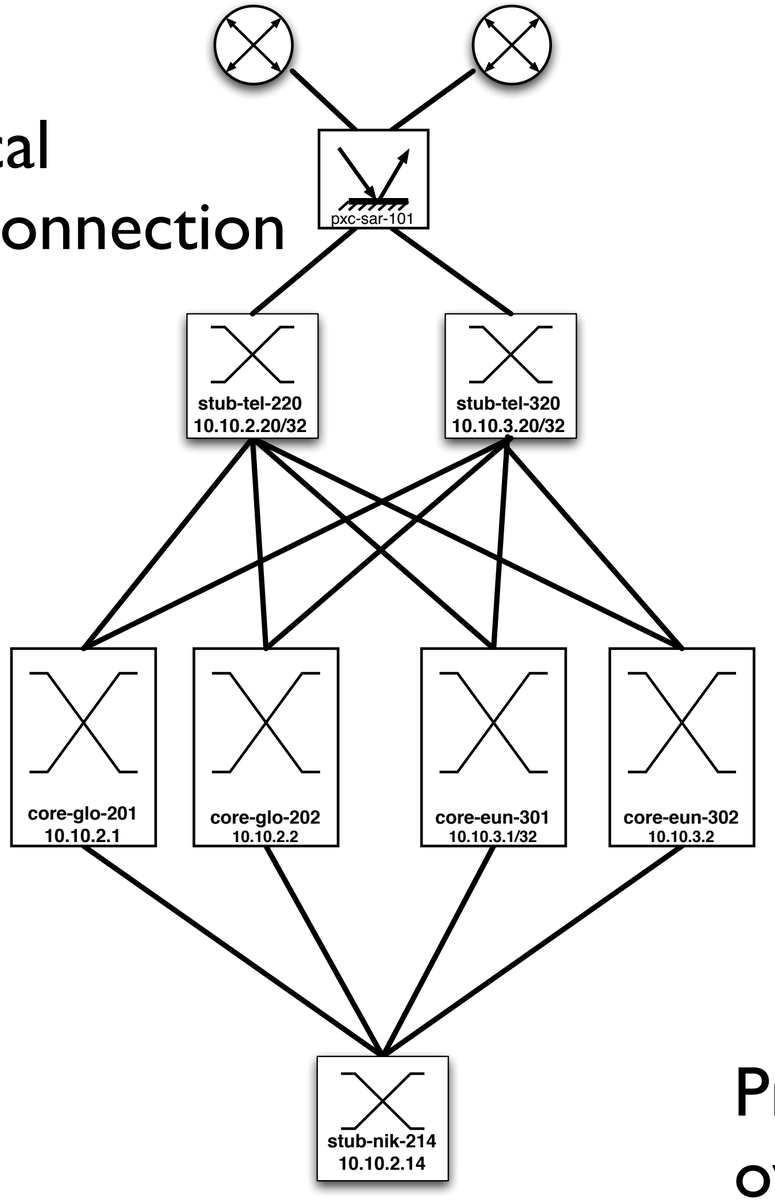
- OSPF
  - BFD for fast detection of link failures
- RSVP-TE signalled LSPs over predefined paths
  - primary and secondary (backup) paths defined
- VPLS instance per VLAN
  - Static defined VPLS peers (LDP signalled)
  - Load balanced over parallel LSPs over all core routers
- Layer 2 ACLs instead of Port Security
  - Manual adjustment for now

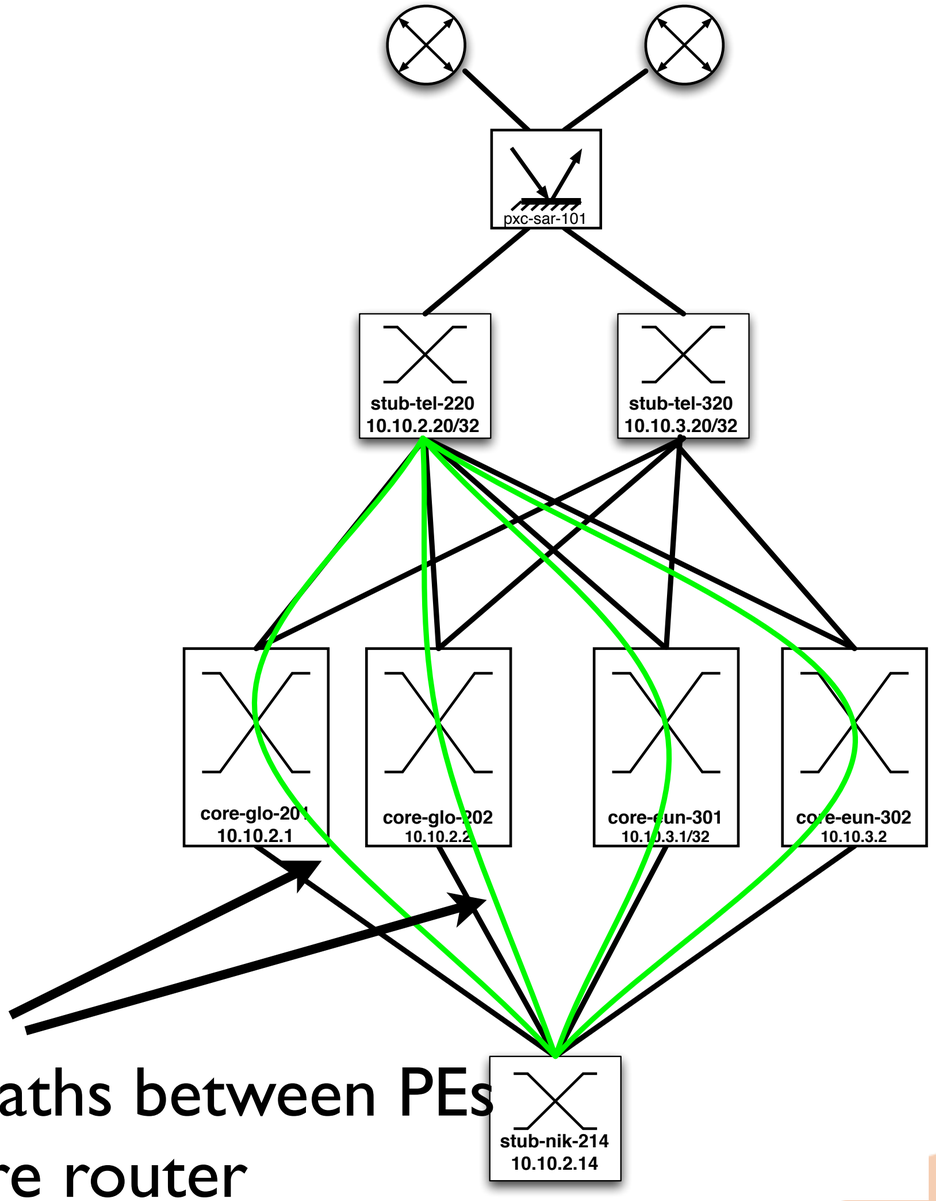
Physical Interconnection



MPLS/VPLS setup

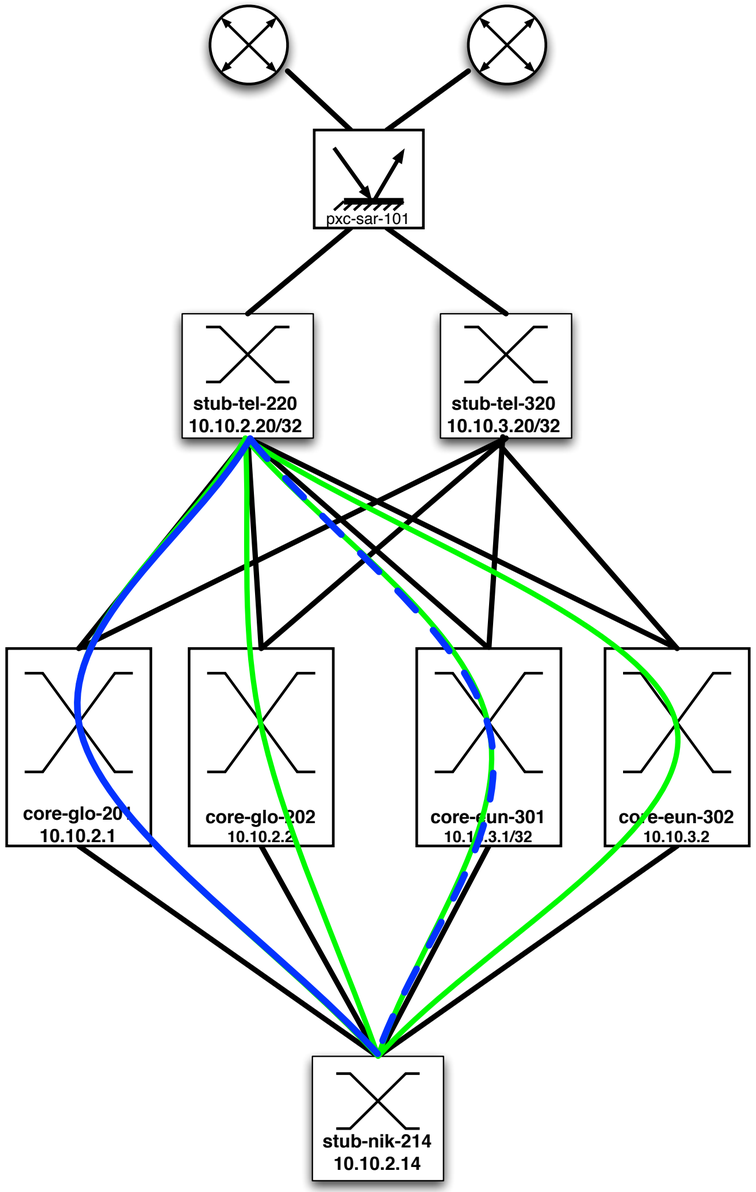
Physical Interconnection





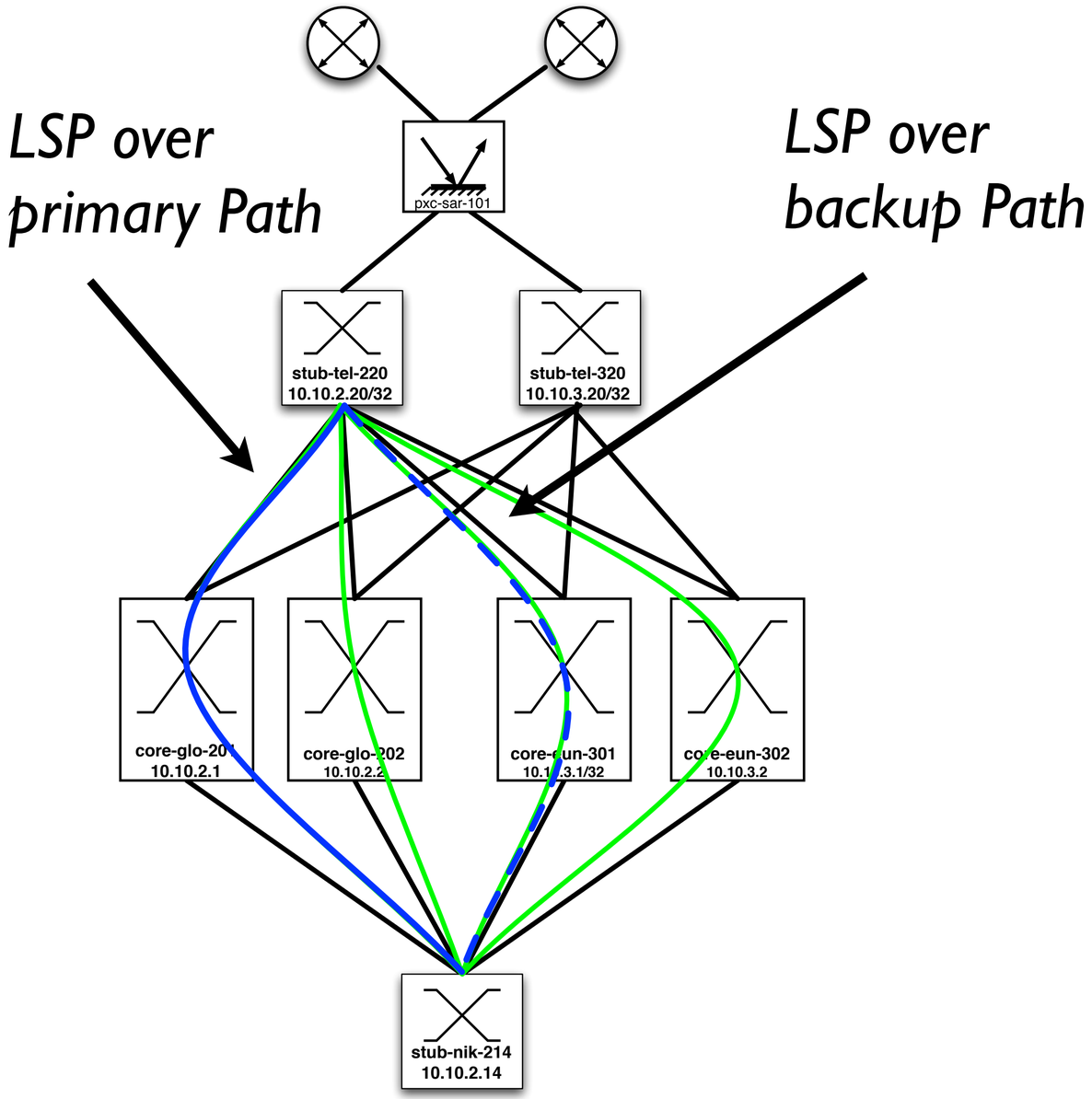
Pre-defined paths between PEs over each core router

MPLS/VPLS setup



MPLS/VPLS setup

Resilience

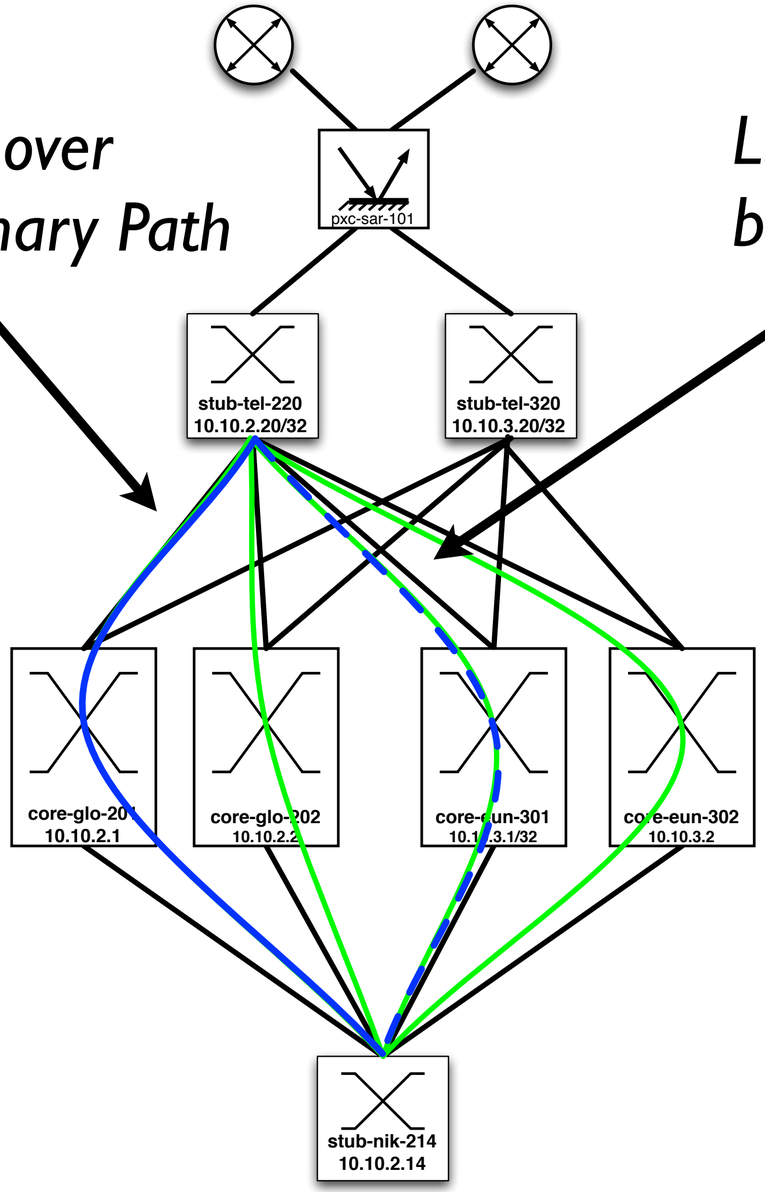


MPLS/VPLS setup

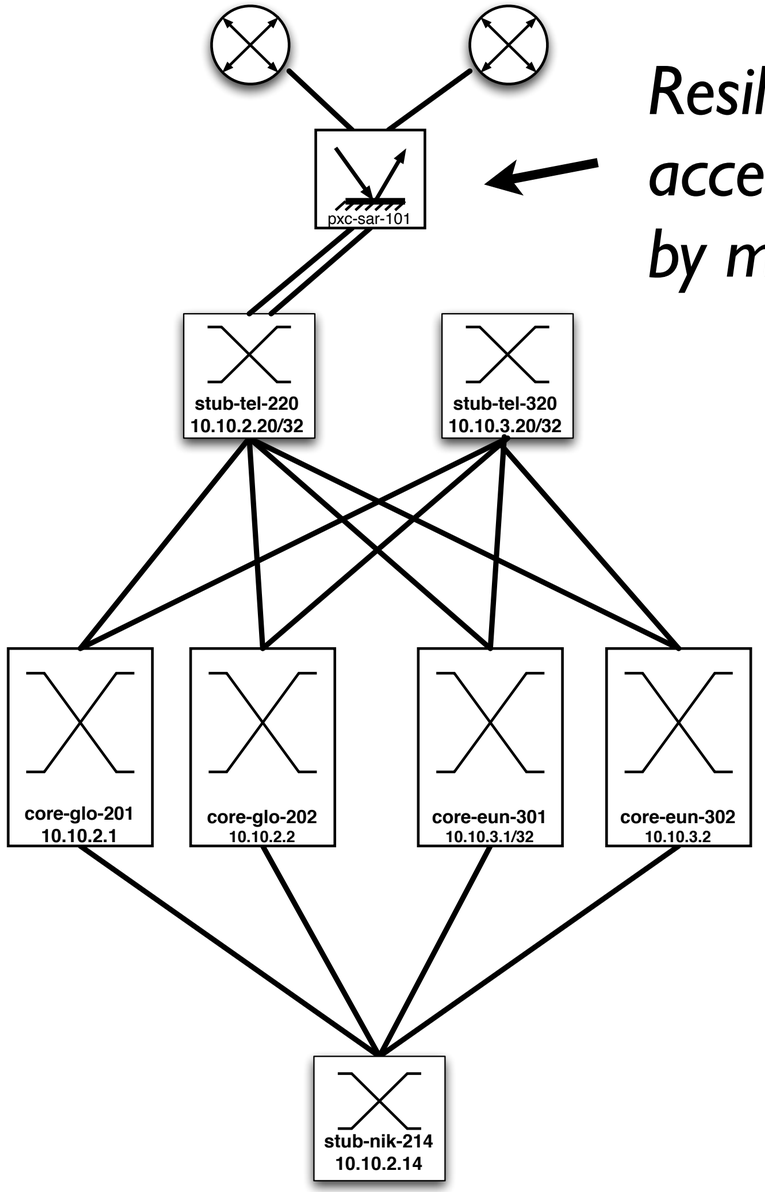
Resilience

LSP over primary Path

LSP over backup Path

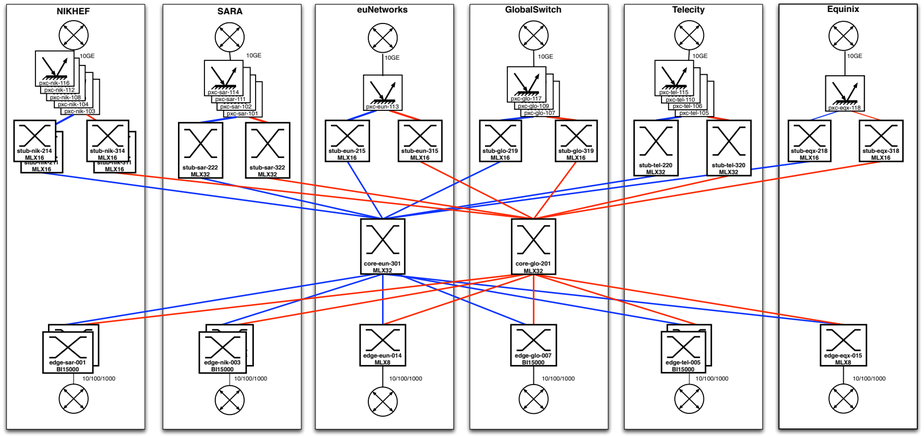


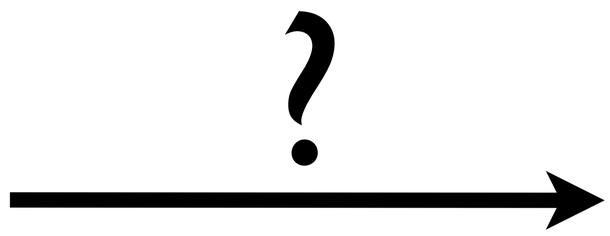
Resilience in access connection by means of PXC

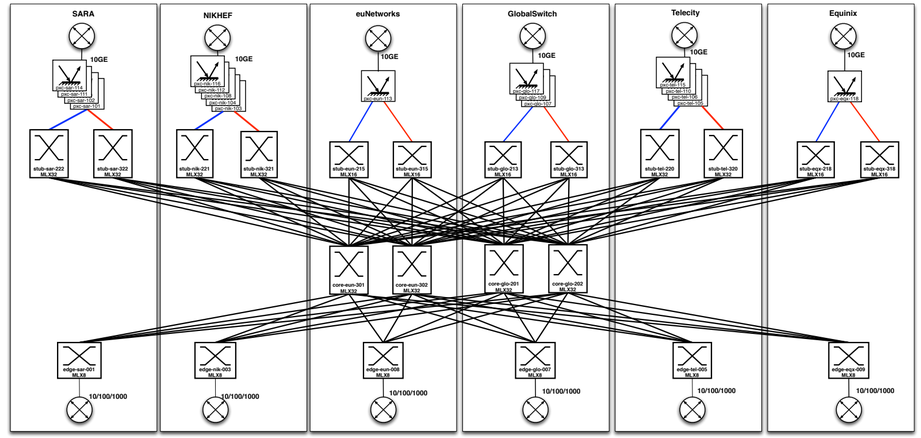


MPLS/VPLS setup

Resilience

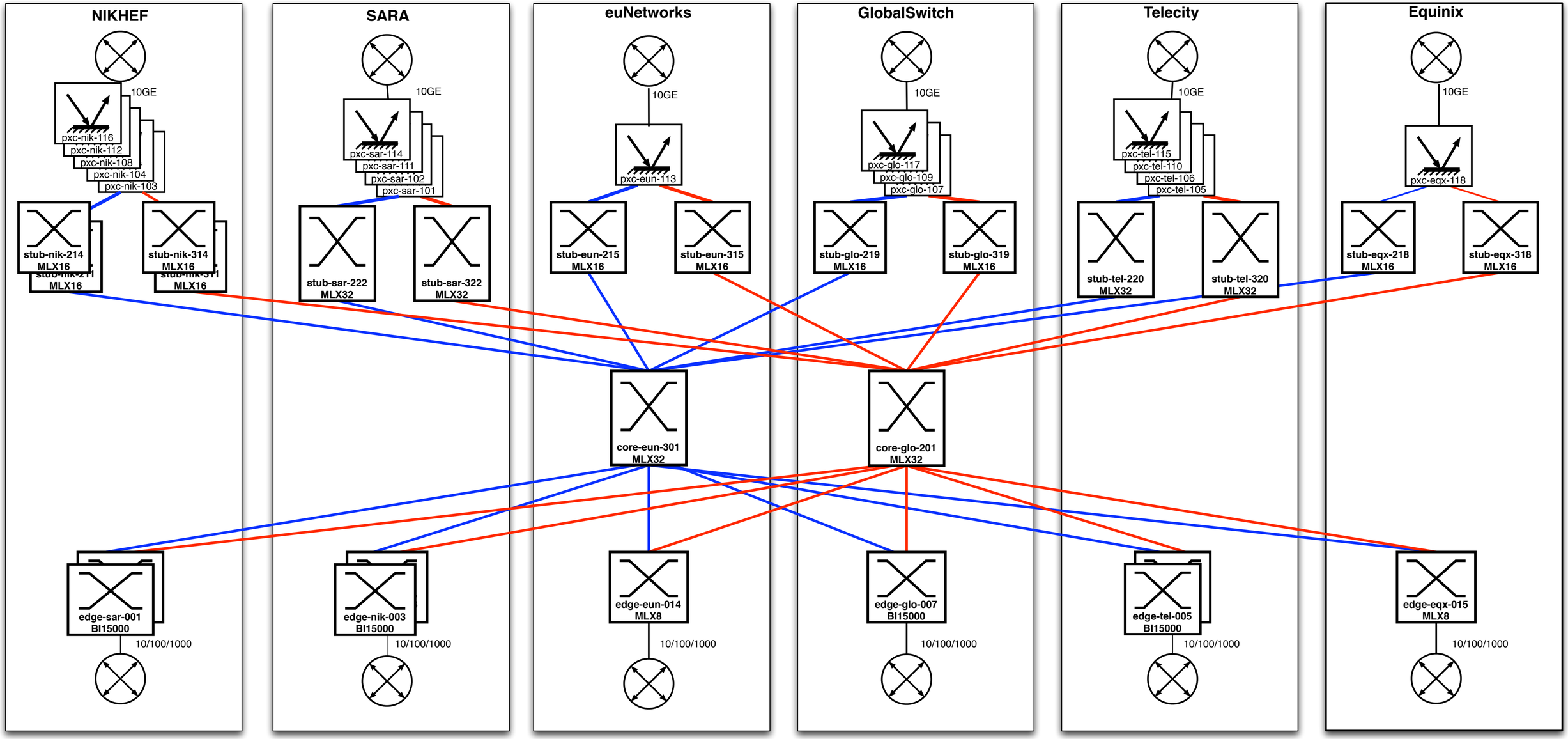






AMS-IX v3 to v4 migration

Requirement: No customer impact !



Migration steps: Initial situation

AMS-IX v3 to v4 migration

Platform Migration

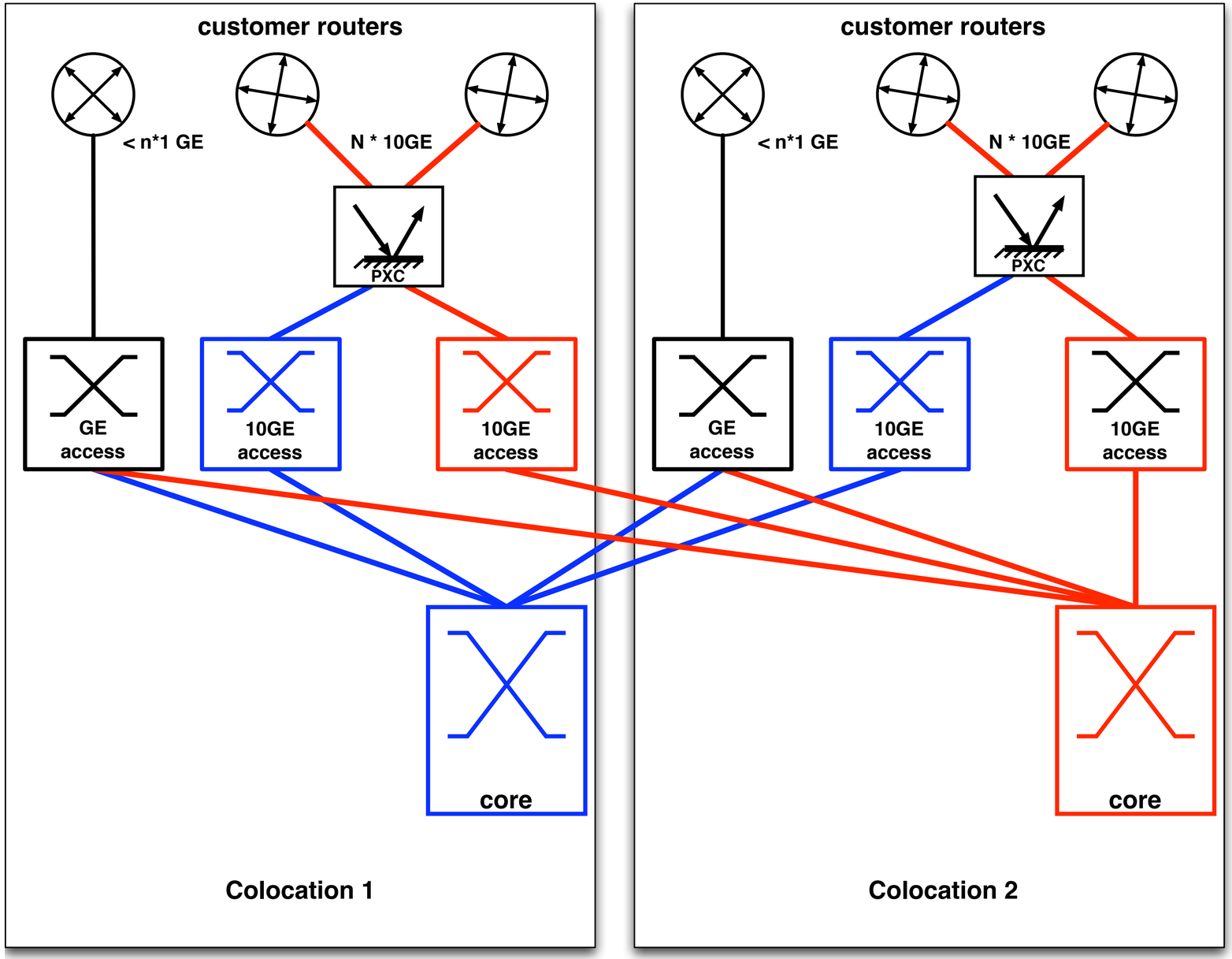
Preparation

- Build new version of PSCD (Photonic Switch Control Deamon)
  - No VSRP traps but LSP state in MPLS cloud
- Develop configuration automation
  - Describe network in XML, generate configurations from this
- Move non MPLS capable access switches behind MPLS routers and PXC as a 10GE customer connection
- Upgrade all non MPLS capable 10GE access switches
- Define migration scenario that would have no customer impact

2 Co-location sites only for simplicity

Double L2 network

VSRP for master slave selection and loop protection

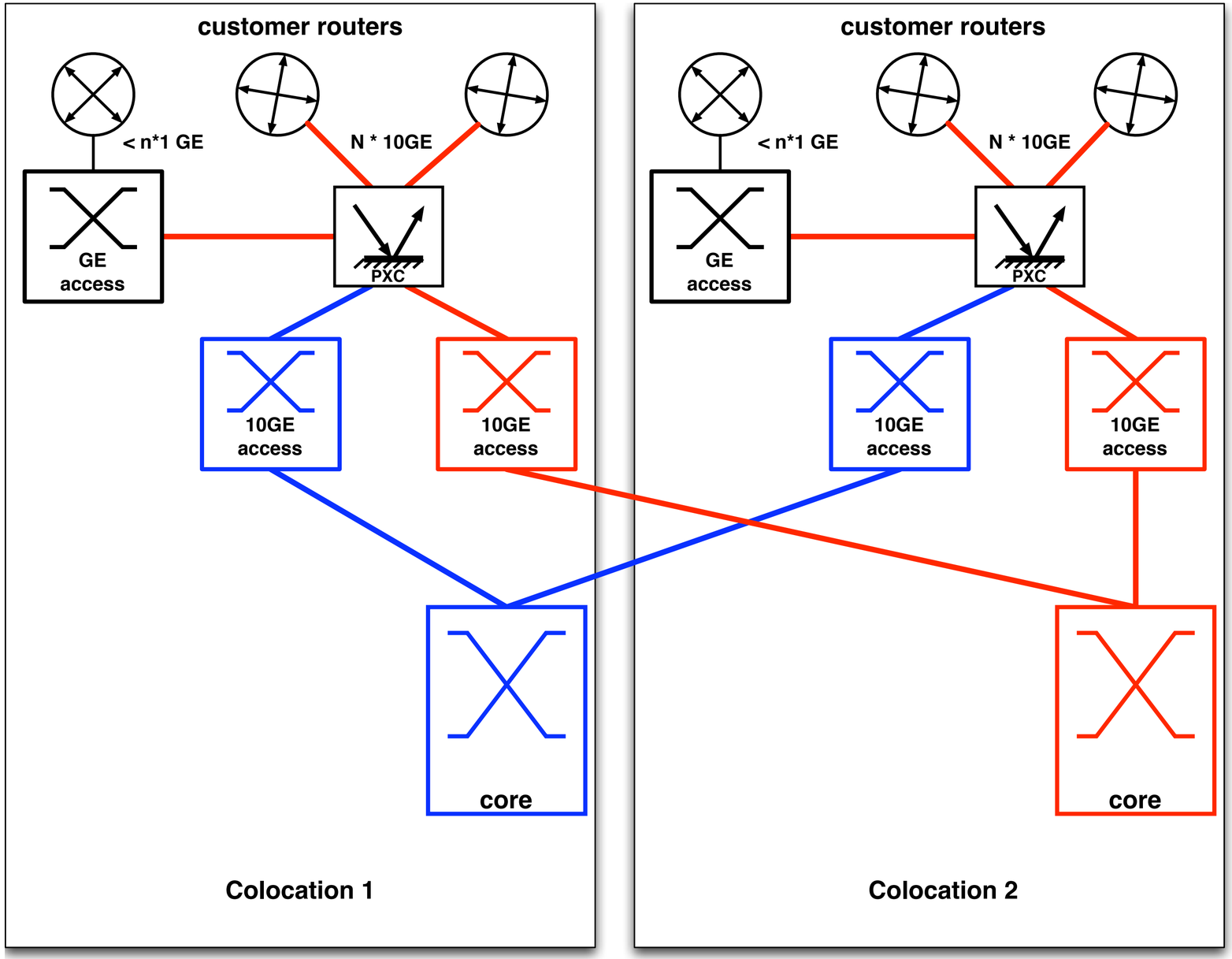


Migration steps: Initial situation simplified

AMS-IX v3 to v4 migration

Not possible to connect GE access switch to both MPLS/VPLS cloud and basic L2 network

Brocade BI-15K (end of life) not MPLS capable



Migration steps: move GE access behind PXC

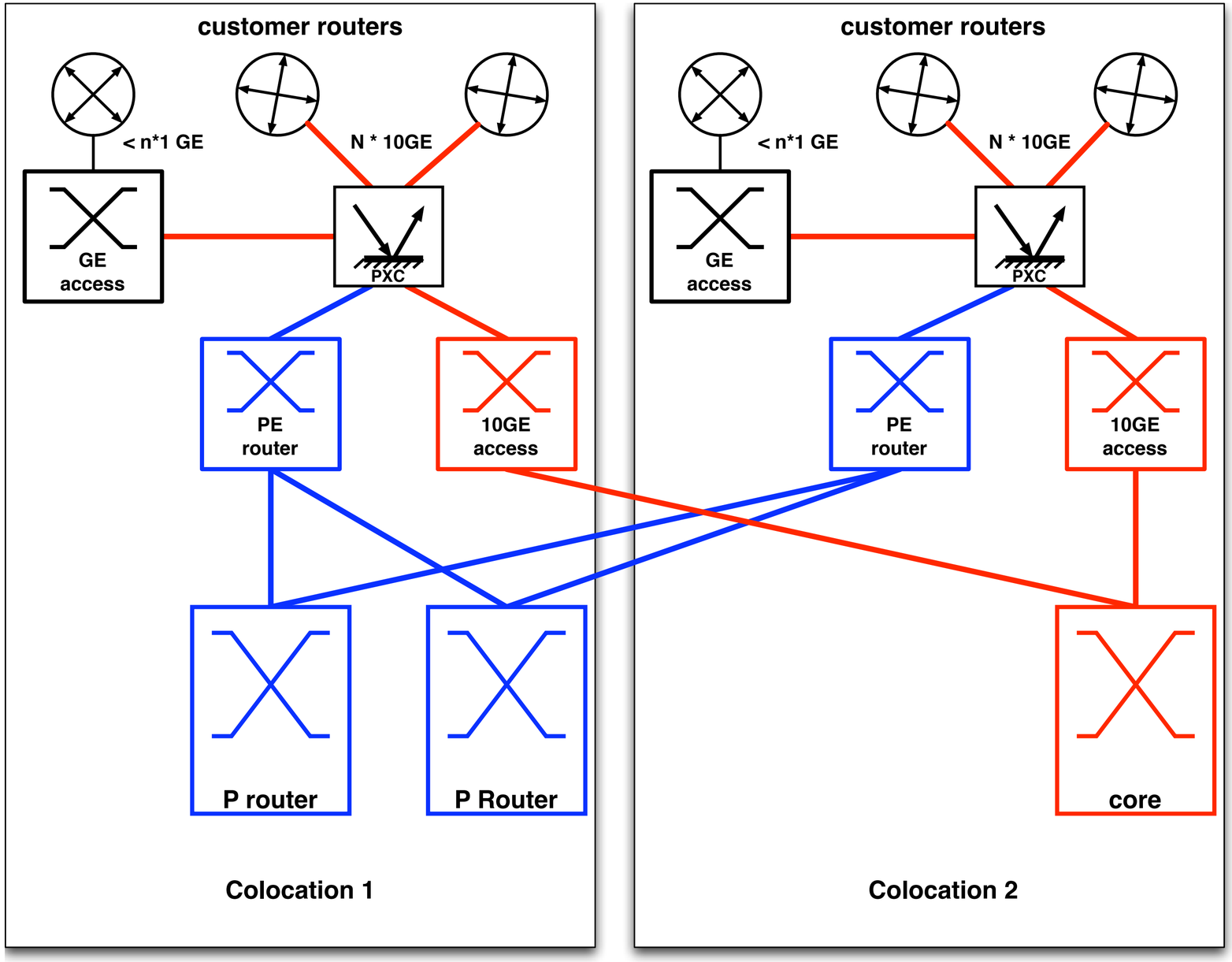
AMS-IX v3 to v4 migration

Production on L2 network (red)

Migrate blue network to MPLS/VPLS

Traffic between two PE routers load balanced over 2 LSPs, one over each P router

Test functionality and connections using test traffic sent by Anritsu traffic generators



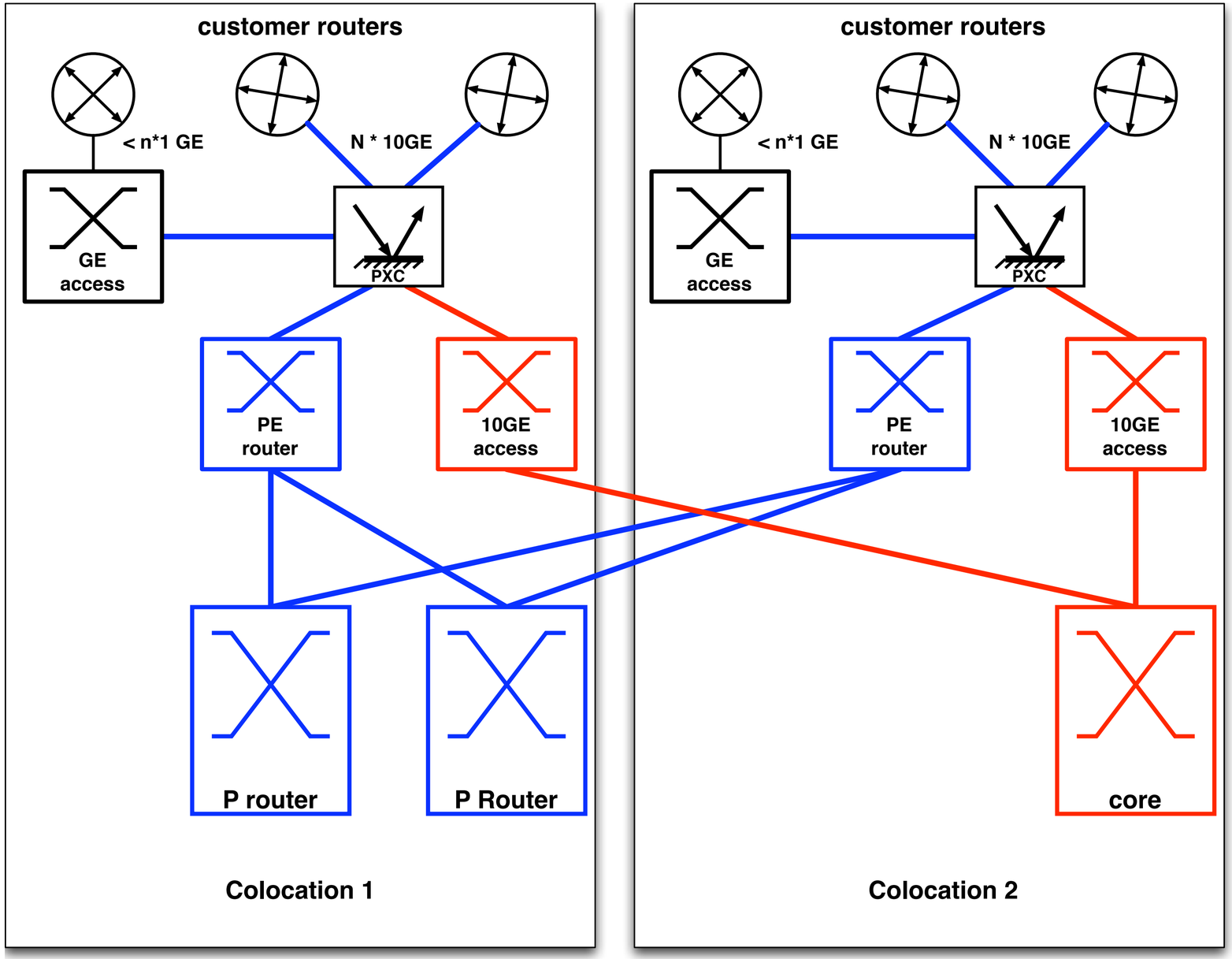
Migration steps: Migrate one half to MPLS/VPLS

AMS-IX v3 to v4 migration

Move production traffic to MPLS/VPLS cloud

- Use PXCs for failover
- New PSCD

Run production on MPLS/VPLS cloud for 6 weeks

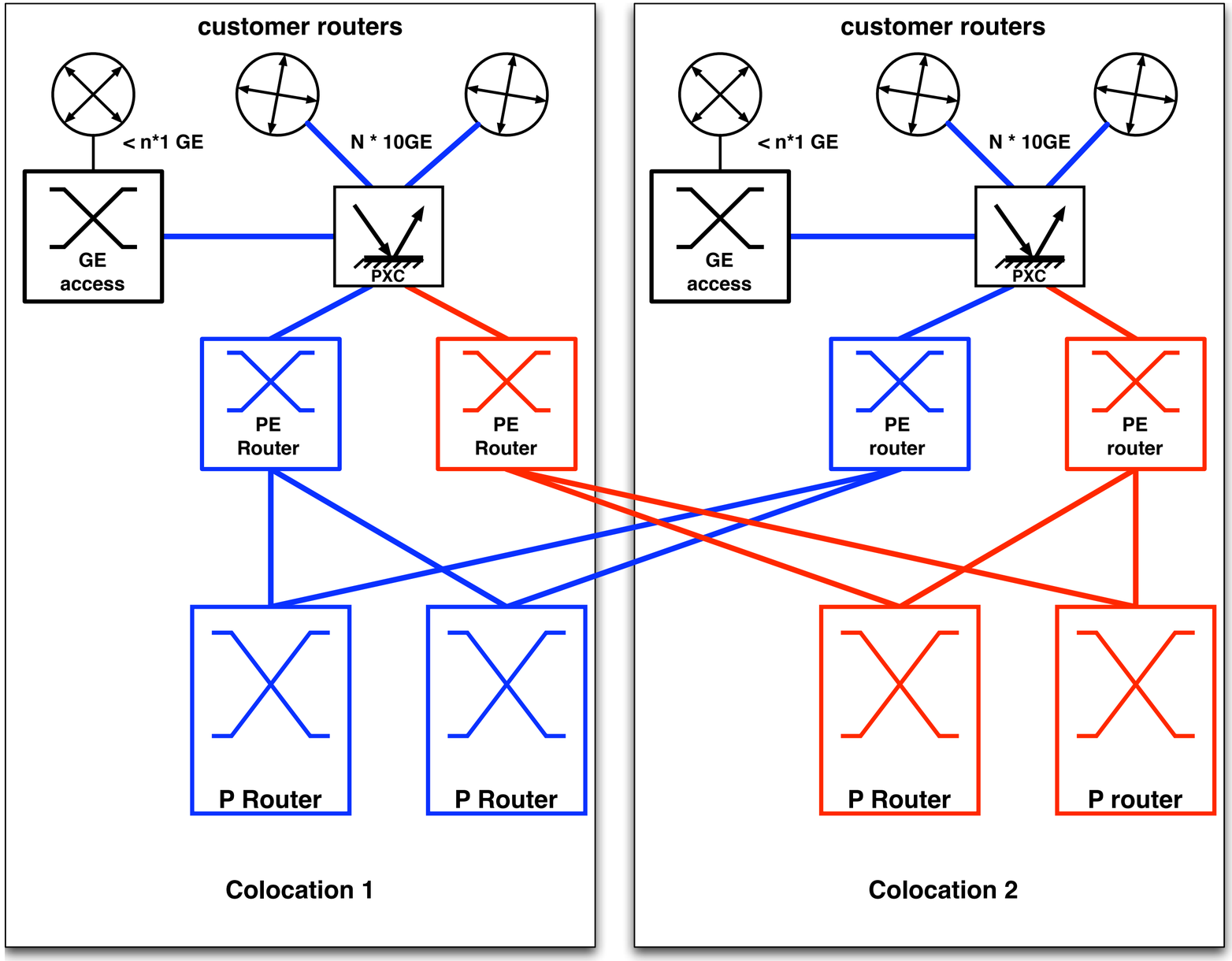


Migration steps: Production on MPLS/VPLS, L2 backup

AMS-IX v3 to v4 migration

Migrate second half of the platform to MPLS/VPLS

Test functionality and connections using test traffic sent by Anritsu traffic generators

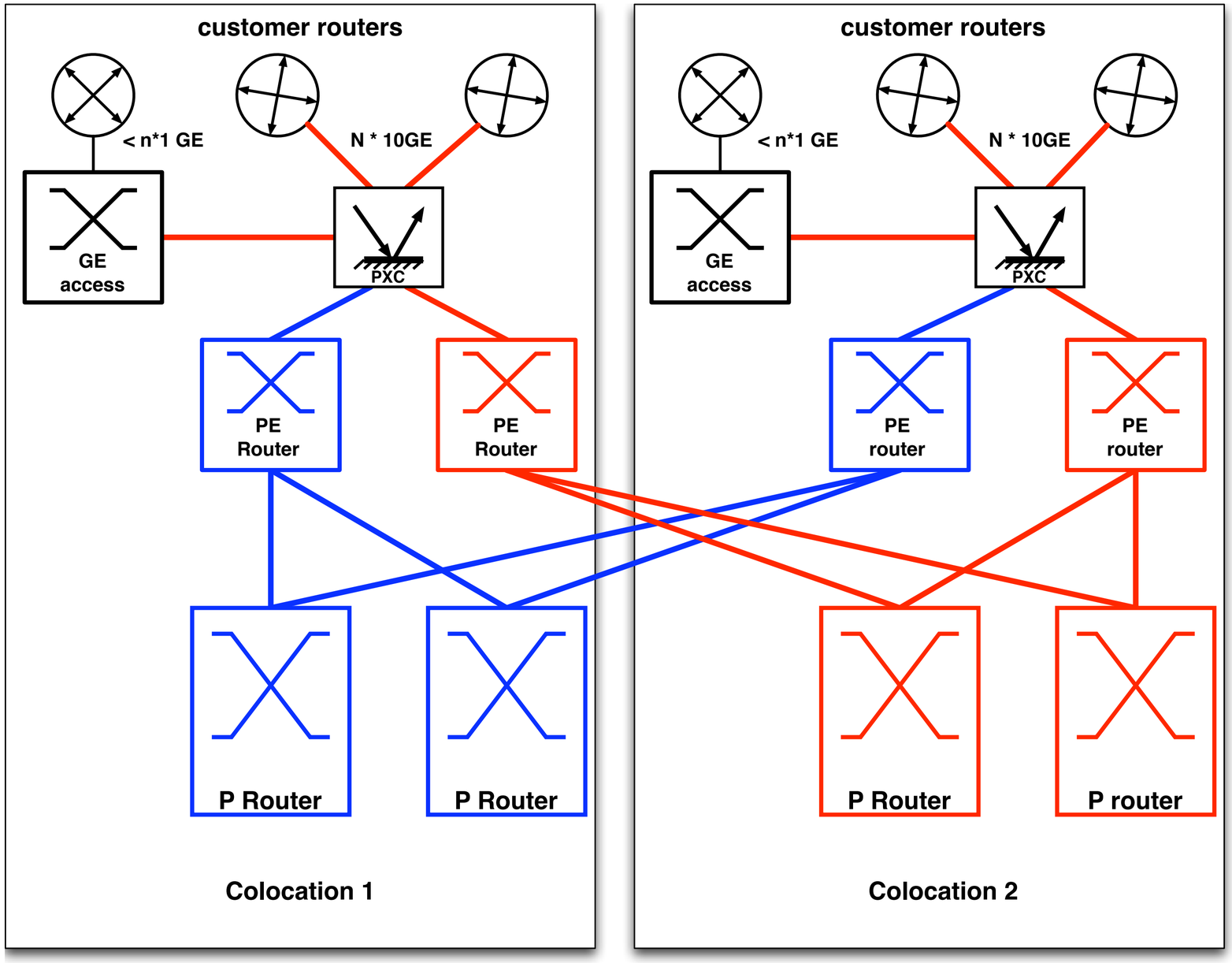


Migration steps: Two MPLS/VPLS platforms

AMS-IX v3 to v4 migration

Move production traffic to red MPLS/VPLS cloud using the newly developed version of PSCD to manage the PXCs

Still two separate networks, both MPLS/ VPLS based



Migration steps: production on second MPLS/VPLS platform

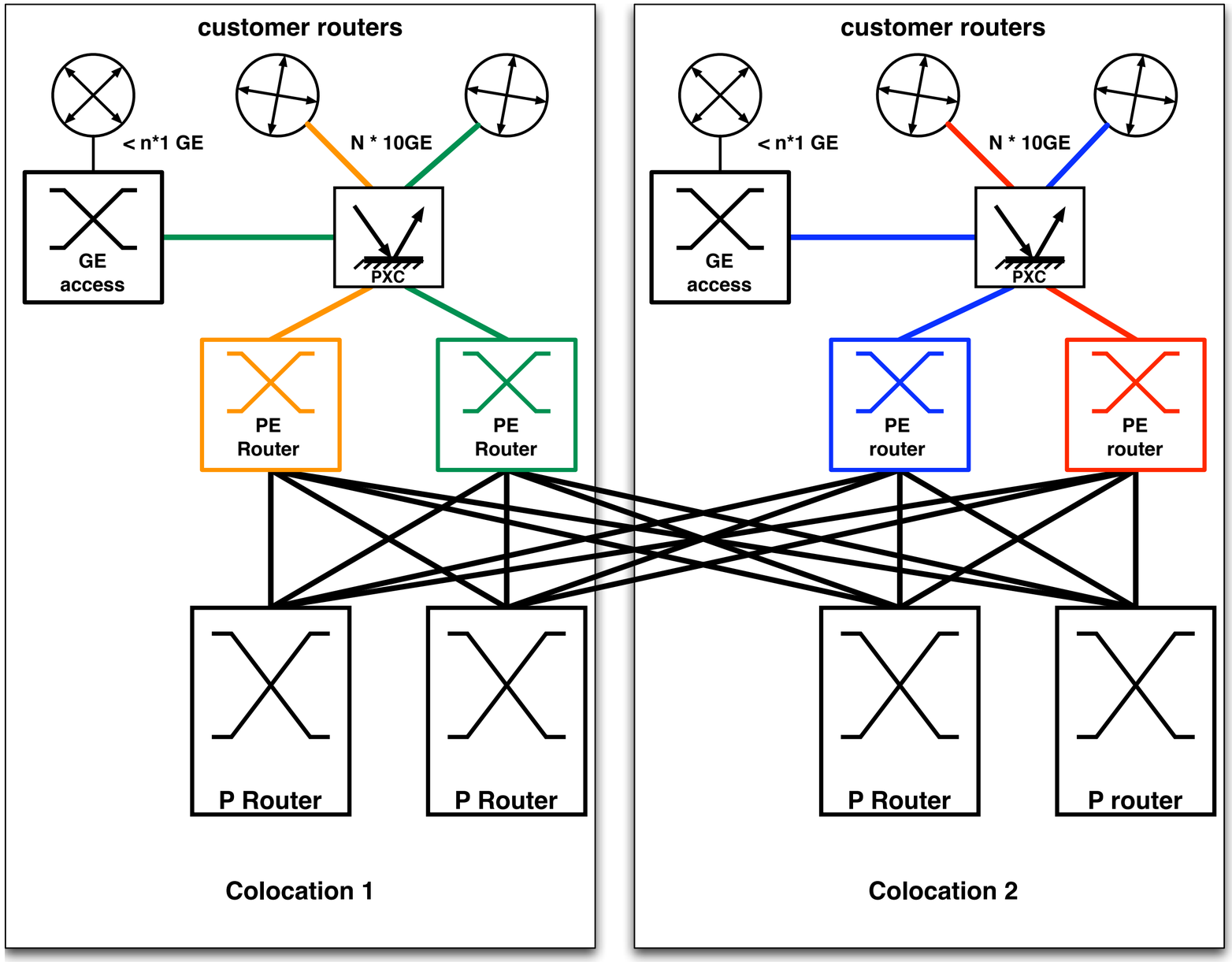
AMS-IX v3 to v4 migration

All PE routers connected to all P routers

- Between each pair of PE routers, 4 LSPs. One over each P router
- Traffic between each pair of PE routers load balanced over the 4 LSP

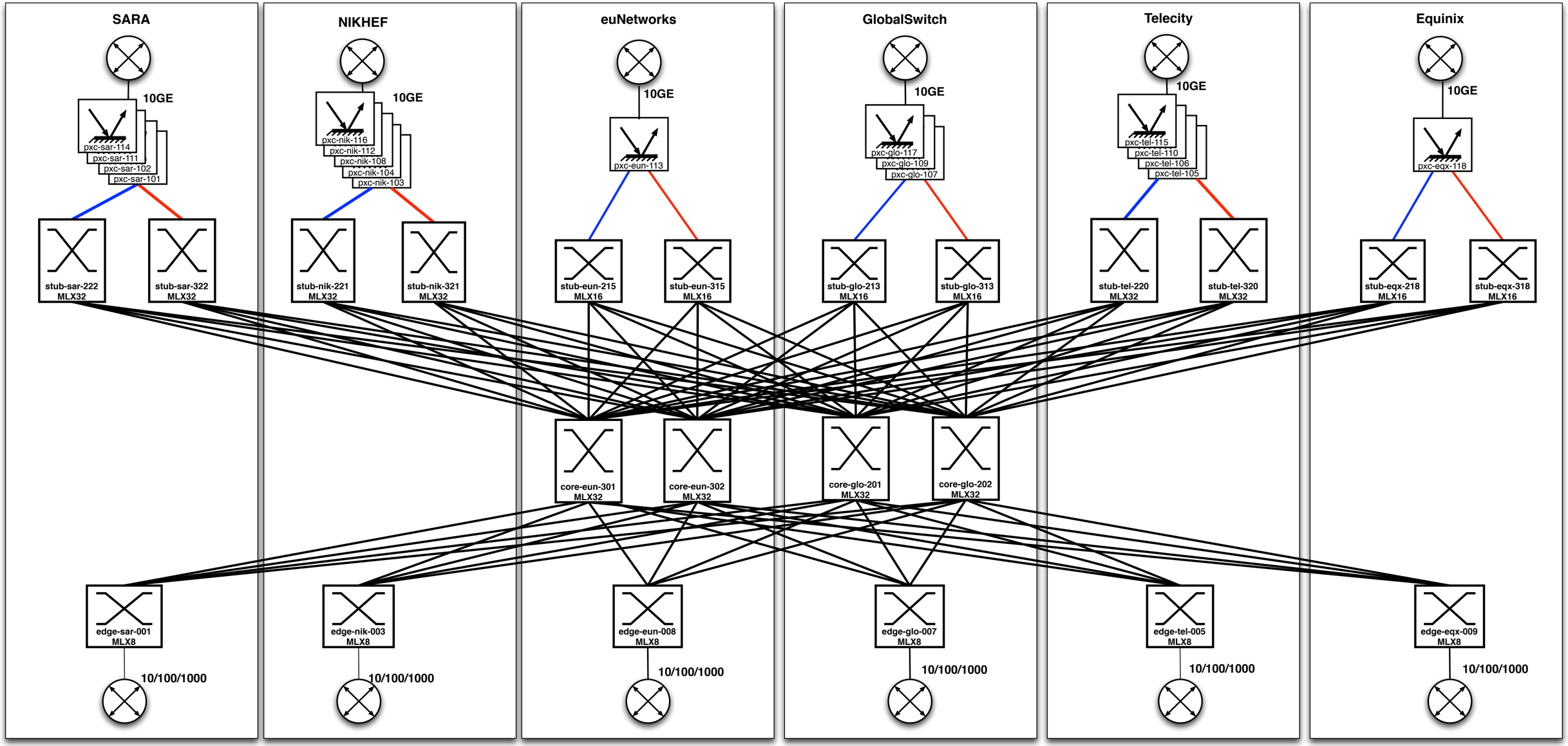
10GE customer connections distributed over local PE routers

- Resilience in 10GE customer connection to local PE router by means of PXCs
- Failover of customer connection to other PE router independent of other connections



Migration steps: integration to single MPLS/VPLS cloud

AMS-IX v3 to v4 migration



Final situation

AMS-IX v3 to v4 migration

Results

- Increased stability
  - Backbone failures handled by MPLS (not seen by customers)
  - Access switch failures handled for a single pair of switches
  - Phased relocation of traffic streams
  - Looped traffic filtered by L2 ACL => No effect on linecard CPU

Results

- Easier debugging of customer ports
  - Simply swap to different, active switch using Glimmerglass PXC
- Config generation
  - Absolute necessity due to size of MPLS/VPLS configuration
  - Fairly simple because of single hardware platform

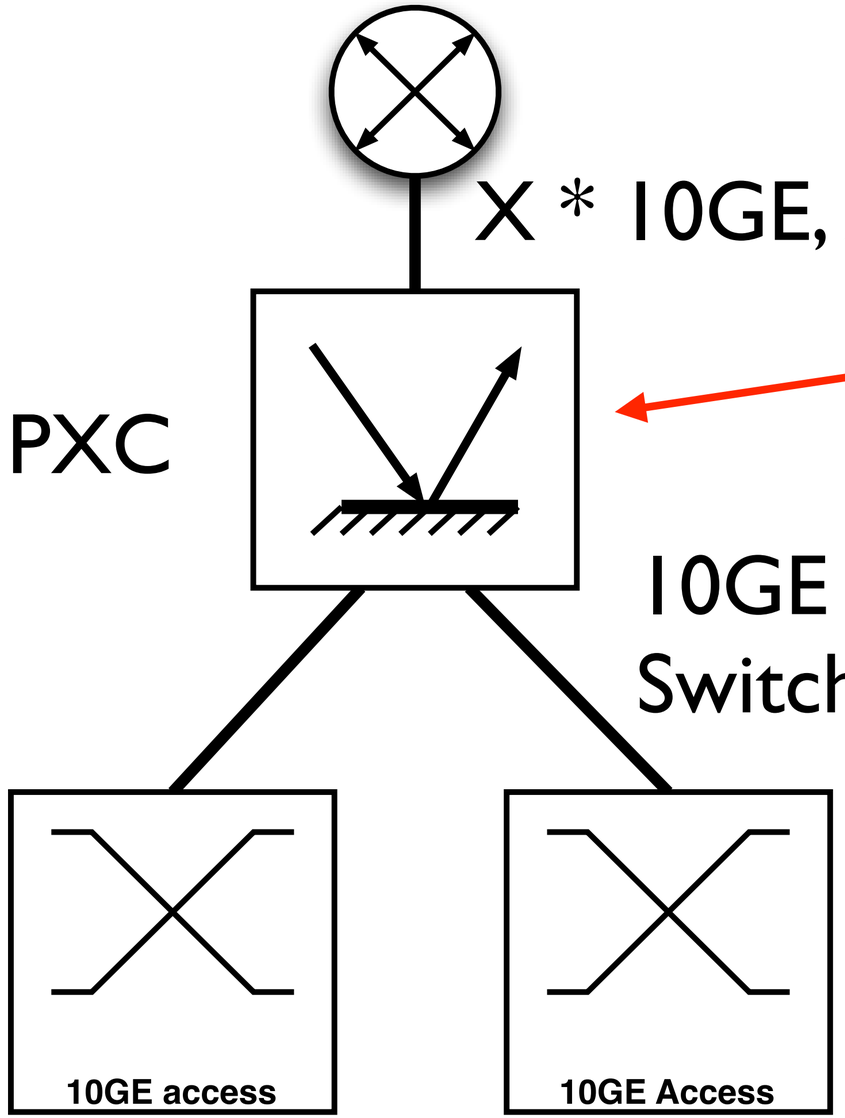
Results

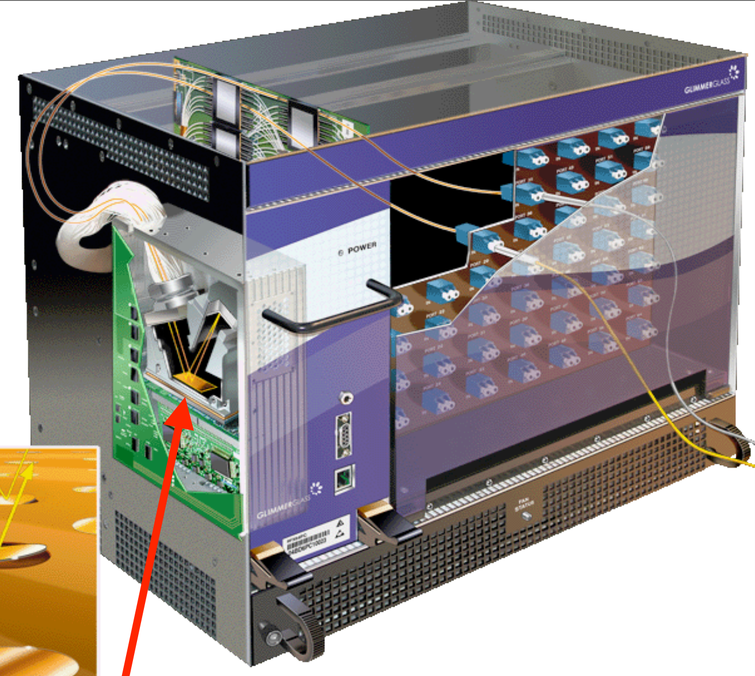
- Scalability (future options)
  - Bigger core devices
    - Do not need to be MPLS-capable
  - Load-sharing over > 4 cores
    - Pending feature request
  - Use of different cores for sets of PEs
  - Multiple layers of P-routers

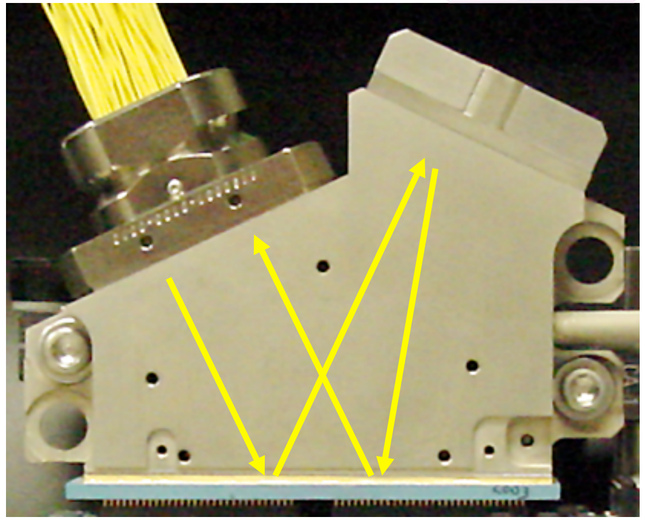
Questions ?

Backup Slides

SubTitle

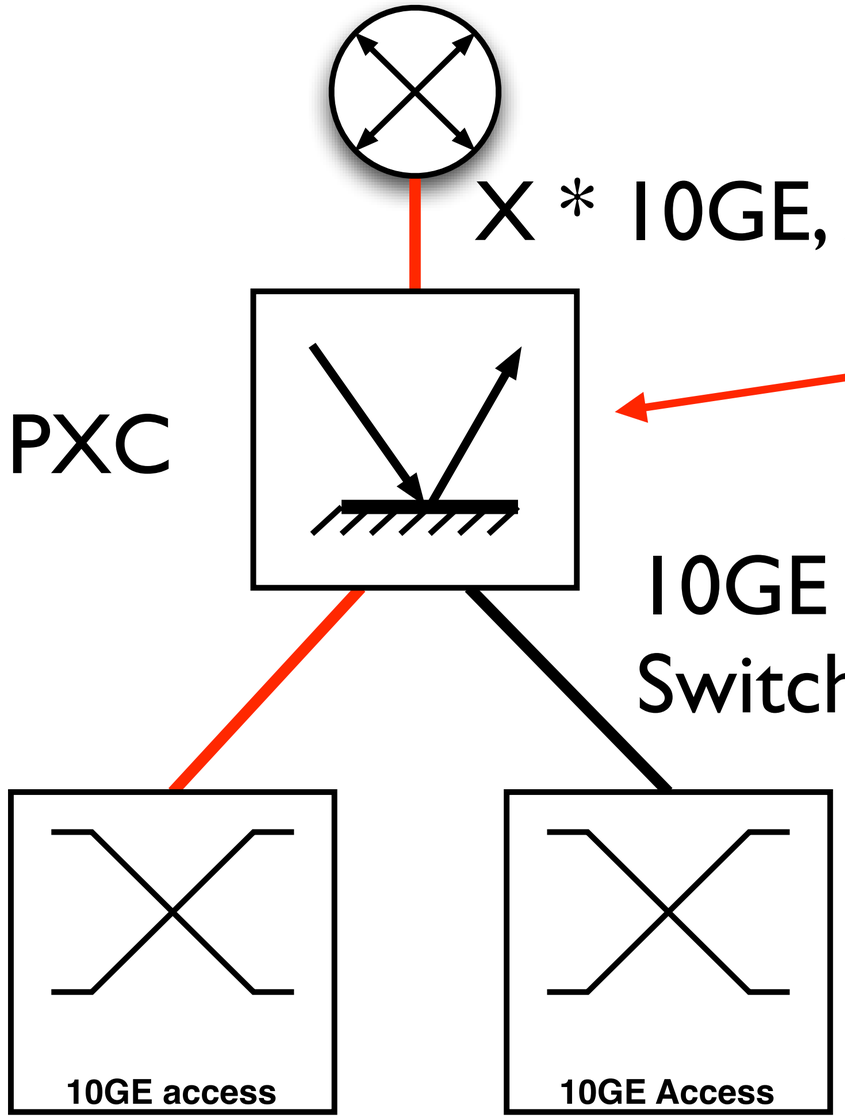


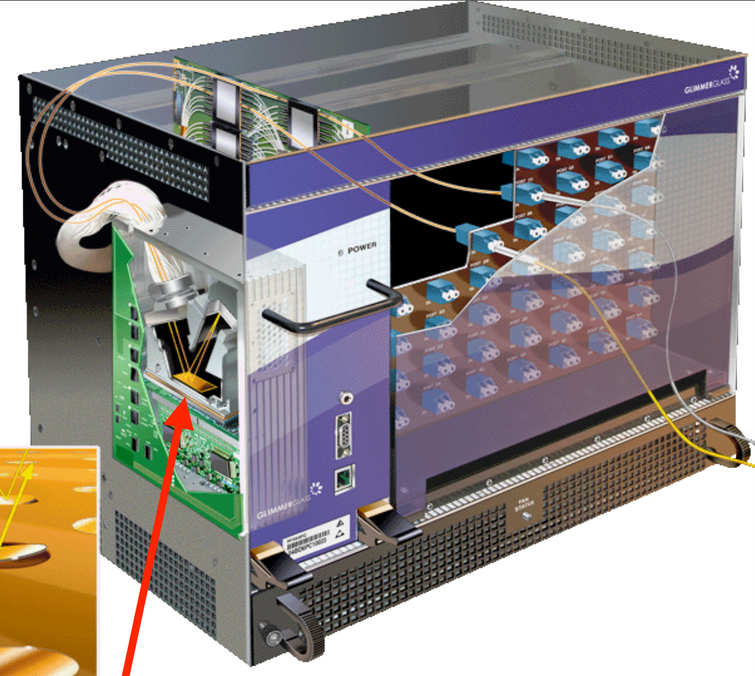


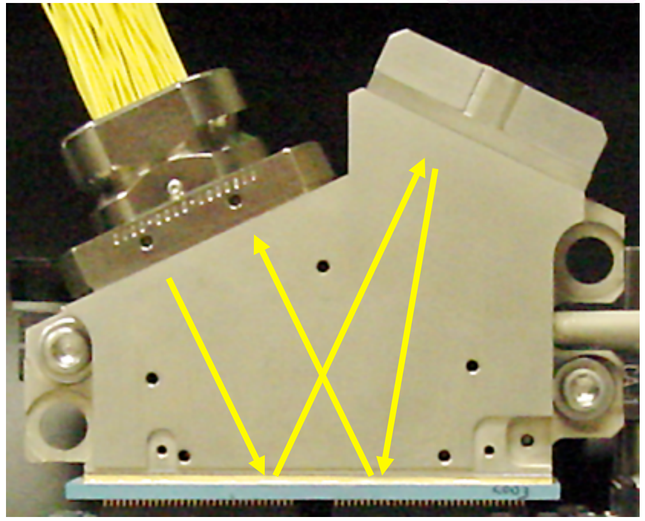


AMS-IX Photonic Cross Connects (PXC)

Resilience in 10GE customer access connection

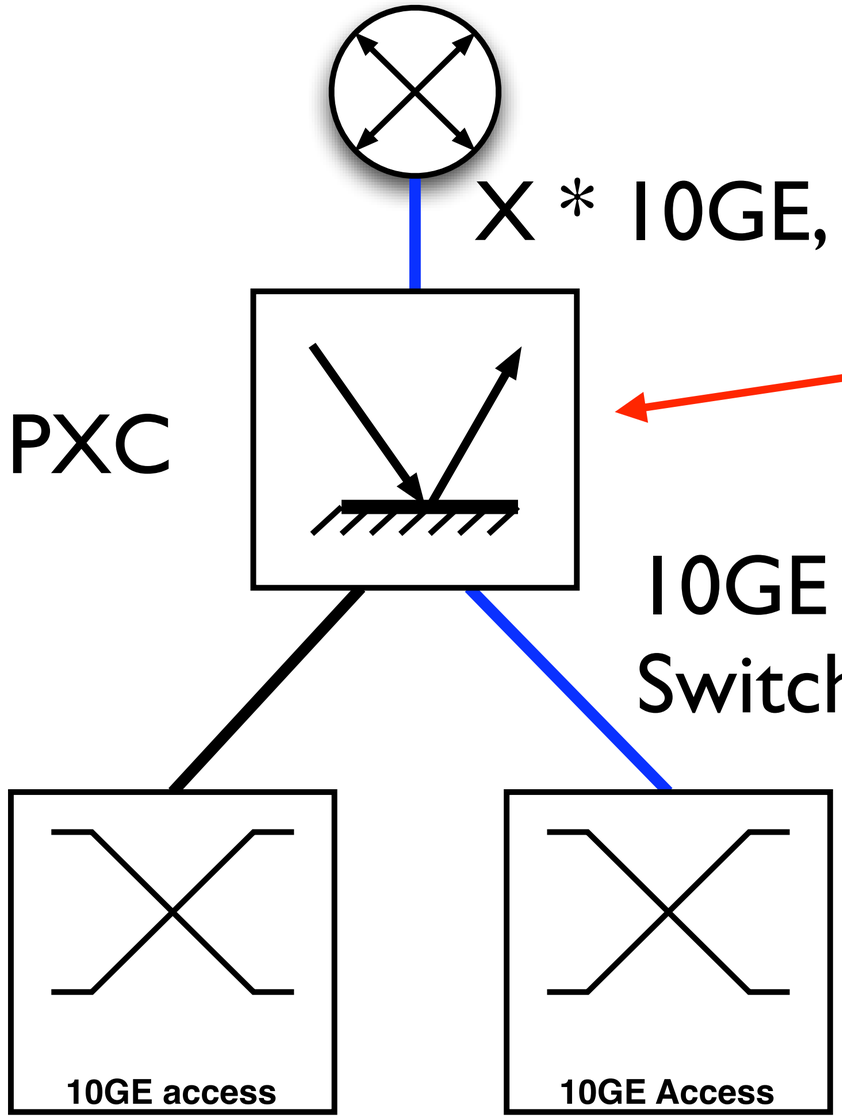


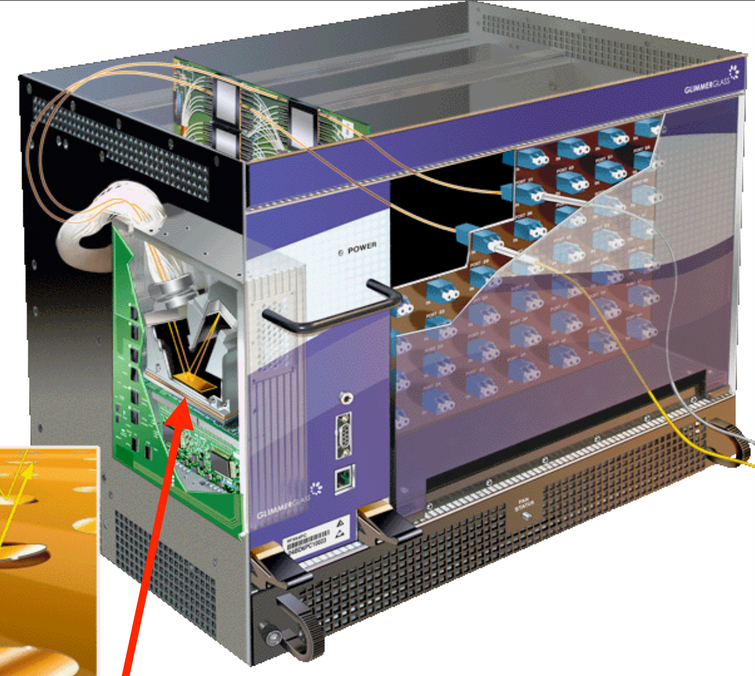


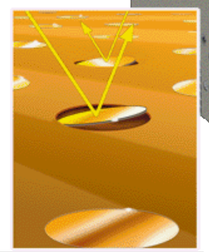


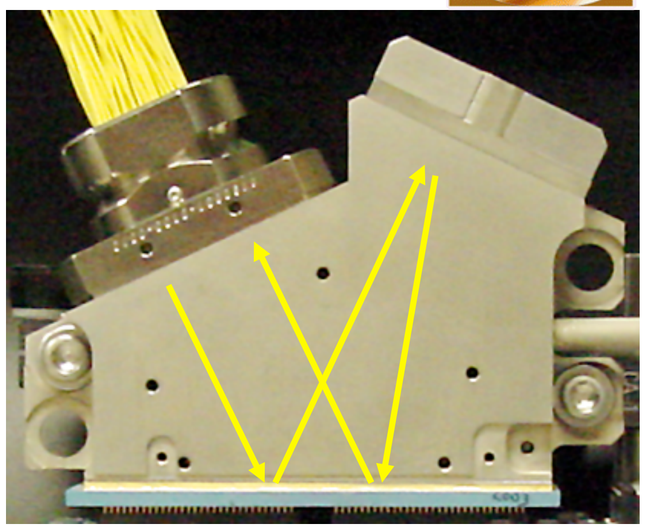
AMS-IX Photonic Cross Connects (PXC)

Resilience in 10GE customer access connection



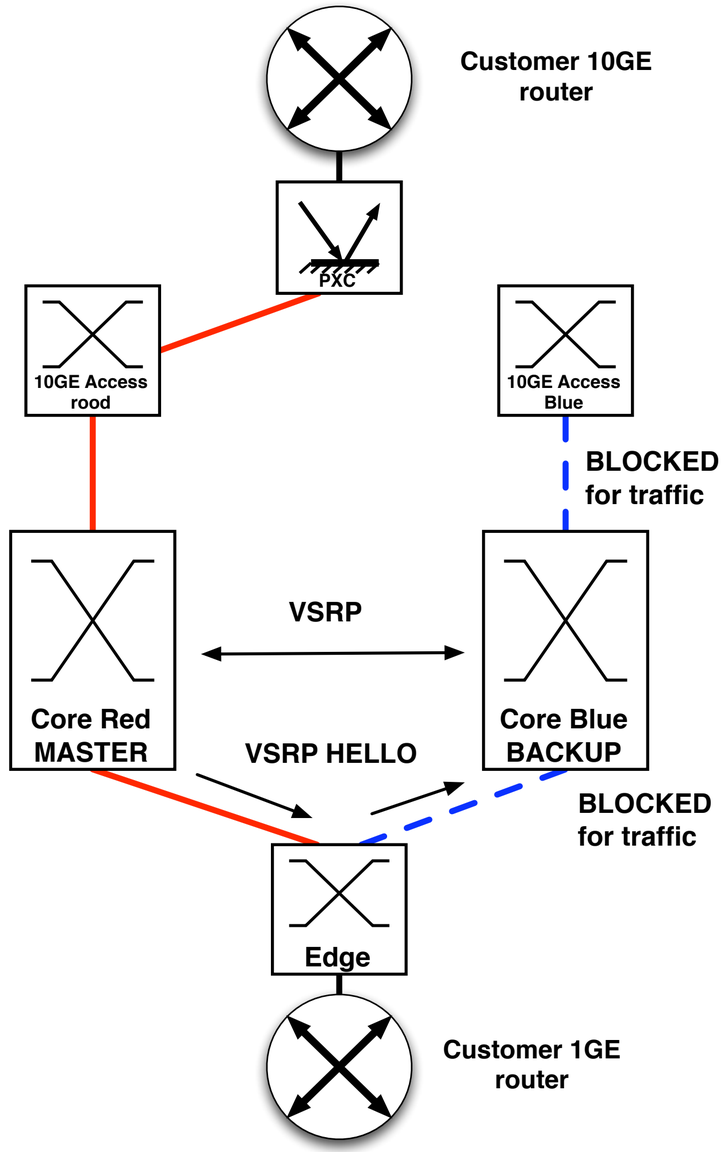






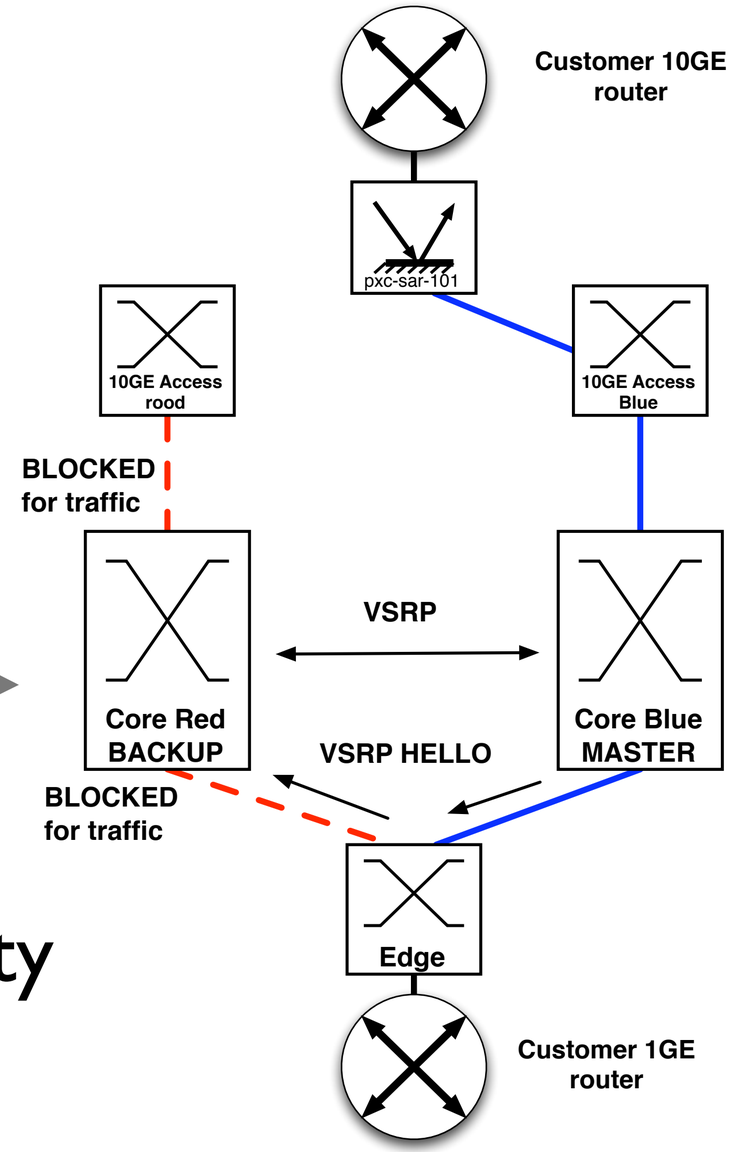
AMS-IX Photonic Cross Connects (PXC)

Resilience in 10GE customer access connection



Problem or maintenance in red network

VSRP Priority Red Master lower than the Blue priority



AMS-IX Version 3 Platform

Topology Failover