Improving scalability of the AMS-IX network

Stéfan Deelen & Reinier Schoof

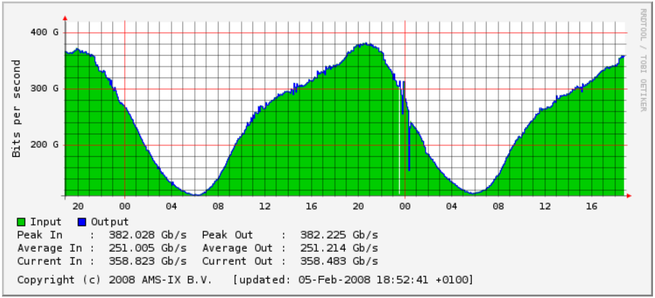
February 6, 2008

amsix
amsterdam internet exchange

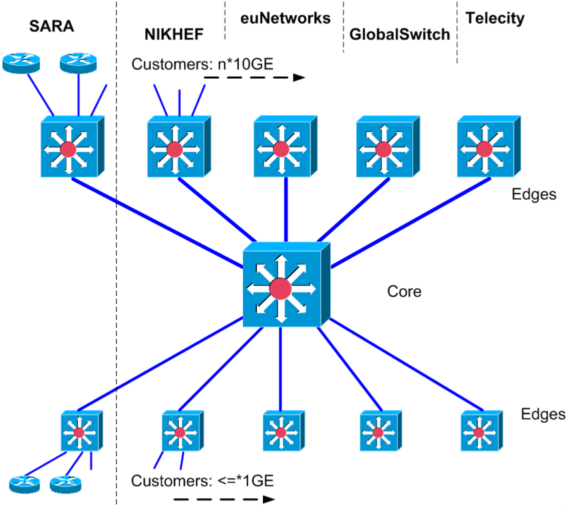
Improving scalability of the AMS-IX network

Amsterdam Internet Exchange

Amsterdam Internet Exchange
World's biggest IX
293 members (05 Feb 2008)
Peaks of over 400 Gb/s



AMS-IX Topology

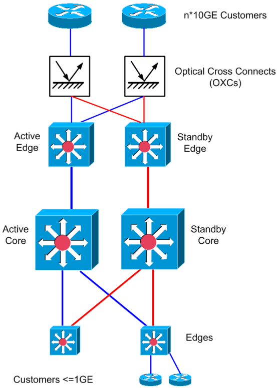


AMS-IX fail-over methods

Completely redundant network

Virtual Switch Redundancy Protocol

Fail-over in approx. 300ms



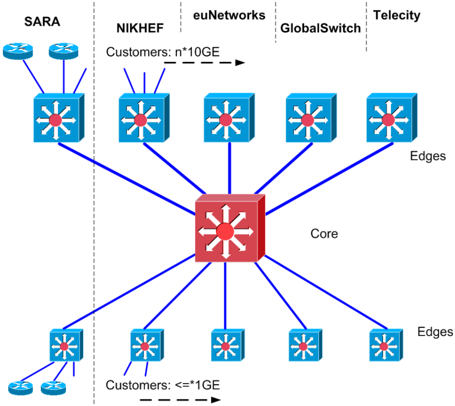
Scalability definition

The ultimate level of scalability for the AMS-IX would be to facilitate unlimited traffic exchange for an unlimited amount of members, with the only limits on throughput being either the capacity of the sending or the receiving party.

"AMS-IX is dedicated to offer non-blocking peering services over Ethernet infrastructure."

Bottleneck

75% to 80% of traffic through core



Problem definition

How can the scalability of the AMS-IX network be improved?

1. What other relevant researches have been conducted previously preceded and what is their relevance to the current research project?
2. Which potential solutions can be found to address AMS-IXs problem in scalability and what are their respective cons and pros?
3. Is there a solution which deserves preference?
4. How could this solution be deployed on the AMS-IX network?

Non approaches

Approaches we did not prefer:

- Up-scaling core switch:
  - No such hardware
  - Still hardware dependent
- Applying redundant links:
  - Loops in network
  - Need for STP
  - No balancing over links
- 'flow based forwarding':
  - Vendor specific feature
  - Not high performance

Full mesh of all customers?

All customers directly connected to each peering partner.

Pros:

- Most efficient offloading
- Fully decentralized

Cons:

- Not scalable
- Not transparent
- High layer 1 costs
- *Very* high port-cost

Full mesh of all edges?

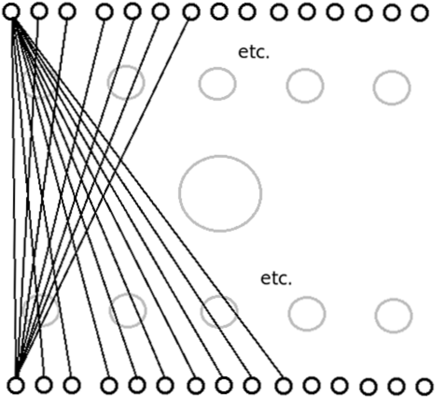
All edges directly connected to each other Pros:

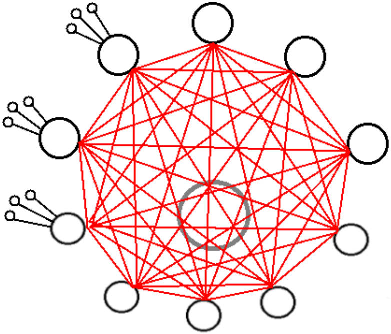
- Efficient offloading
- Transparent from customers point of view

Cons:

- High port-cost
- High layer 1 costs

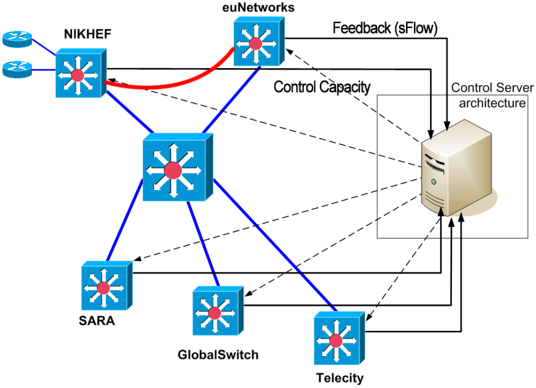
Full mesh



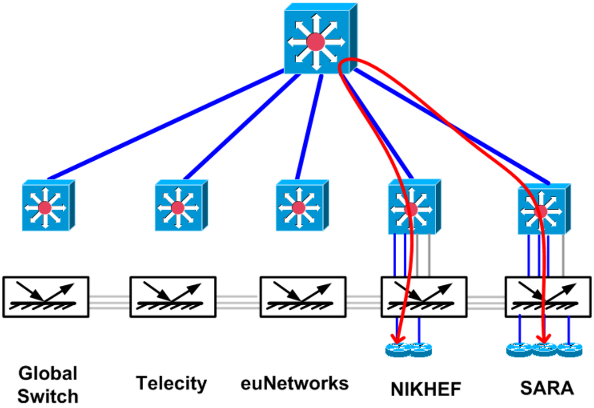


Control Architecture

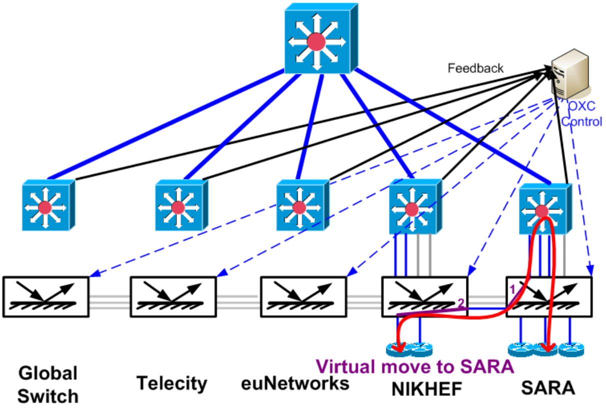
Provides dynamic network usage



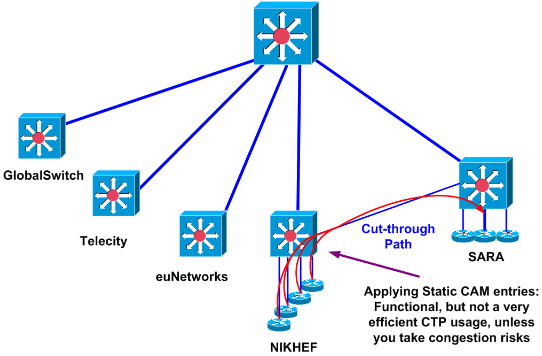
Increase locally switched traffic



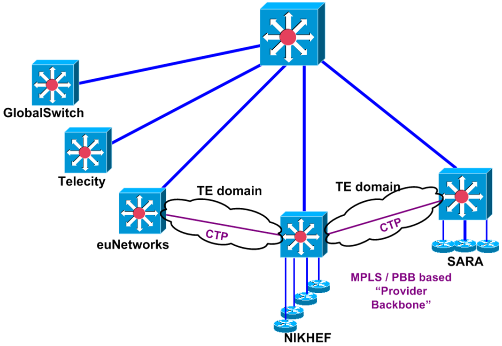
Increase locally switched traffic



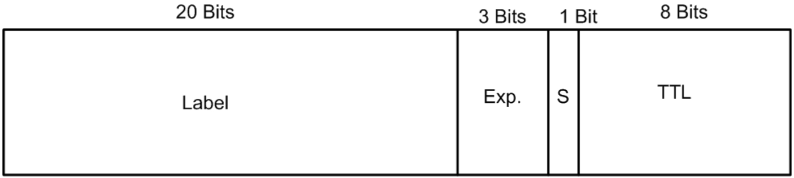
Static CAM entry in edge switch



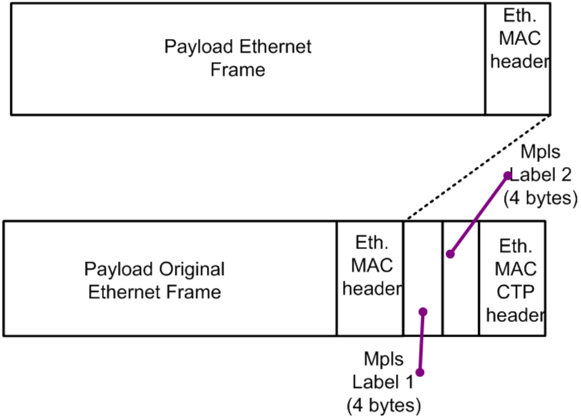
PBB/MPLS Overview



MPLS label



VPLS label


Payload Ethernet Frame
Eth. MAC header
Mpls Label 2 (4 bytes)
Payload Original Ethernet Frame
Eth. MAC header
Eth. MAC CTP header
Mpls Label 1 (4 bytes)

MPLS/VPLS Label

Source MAC (6bytes) + Destination MAC (6bytes) + Ethertype (2bytes) + MPLS-Label1 (4bytes) + MPLSLabel2 (4bytes) = 22 bytes.

Compared to a regular Q-tagged Ethernet header of 26 bytes, MPLS/VPLS adds 84% protocol overhead

GMPLS

Generalized Multiprotocol Label Switching

- MPLS traffic engineering extended with optical cross connect control
- Multi-layer control plane

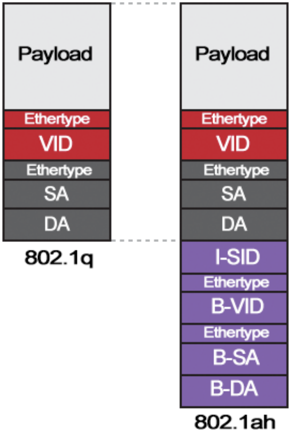
Provider Backbone Bridging

Provider Backbone Bridging (PBB)

- IEEE 802.1ah, but not standardized
- Encapsulates Ethernet frames in Ethernet headers (MAC-in-MAC)
- Forwarding method untouched for non-PBB devices
- Flow based traffic engineering

PBB Frame

Regular Ethernet frame compared to PBB encapsulated frame



PBB-TE

Provider Backbone Briding - Traffic engineering

- IEEE 802.1Qay, but not standardized
- Control plane for PBB
- Suite of several control protocols

Conclusion

Conclusion:

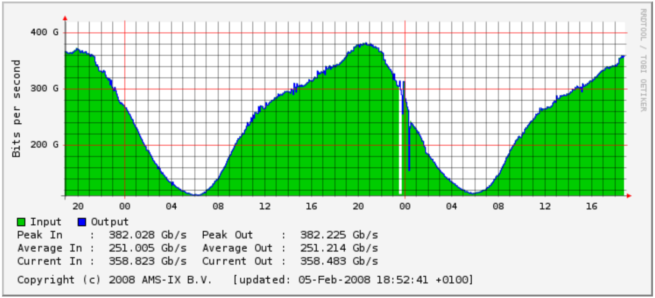
- Load adaption on layer 1
- Traffic engineering on layer 2
- Both PBB and MPLS are solid solutions

Future work:

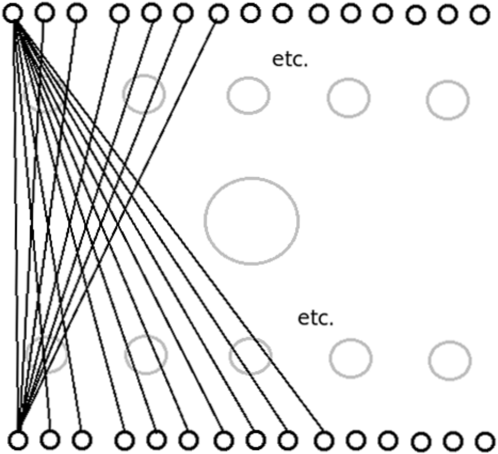
- Performance comparison PBB vs. MPLS
- Implementation of demand-based CTP preparation
- Implementation of control architecture

Questions & Thank you!

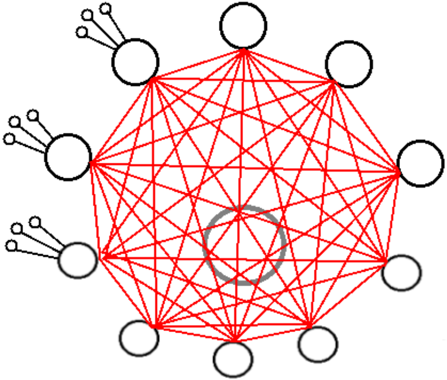
Questions?



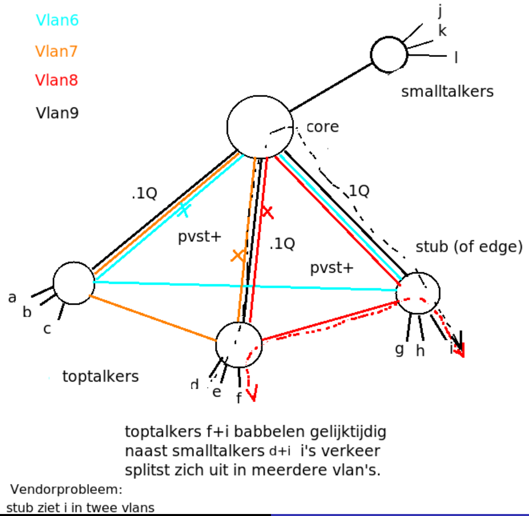
Full mesh of all customers



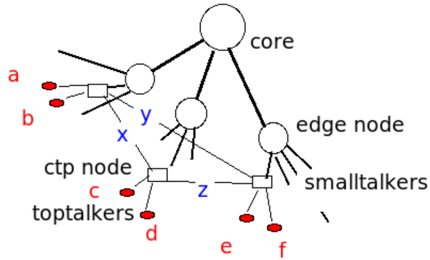
Full mesh of all edges



Full mesh of all edges



Full mesh of all edges



learning via de core wordt overruled door de statische cam entry in de accessnode

nadeel: 1) spof's! toptalkersaanname, 2) beperkt schaalbaar; extra access bridges

voordeel: beproefde techniek, realistisch

Distributed core