Cisco Expo 2009

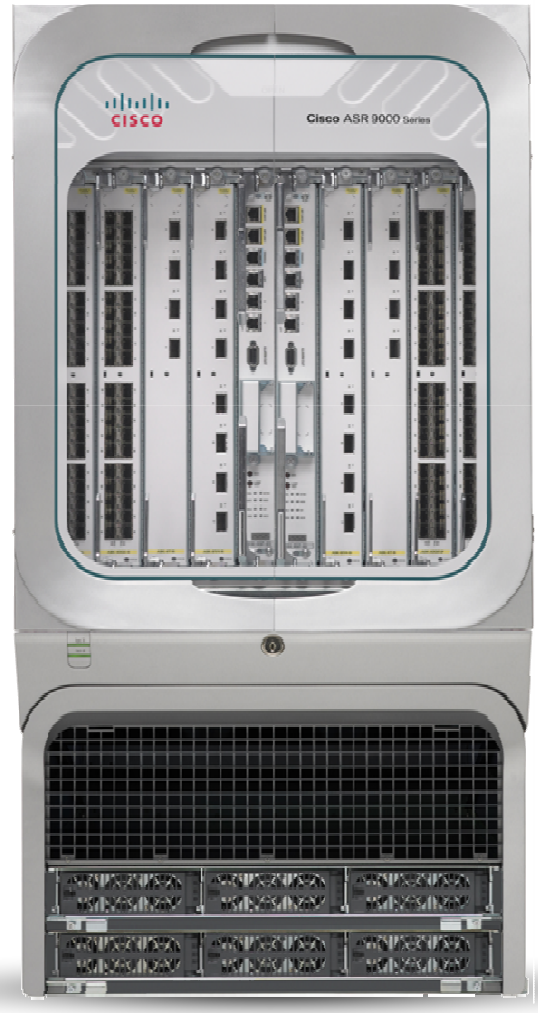
Cisco ASR 9000

Đorđe Vulović, Systems Engineer, CCIE #16582

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

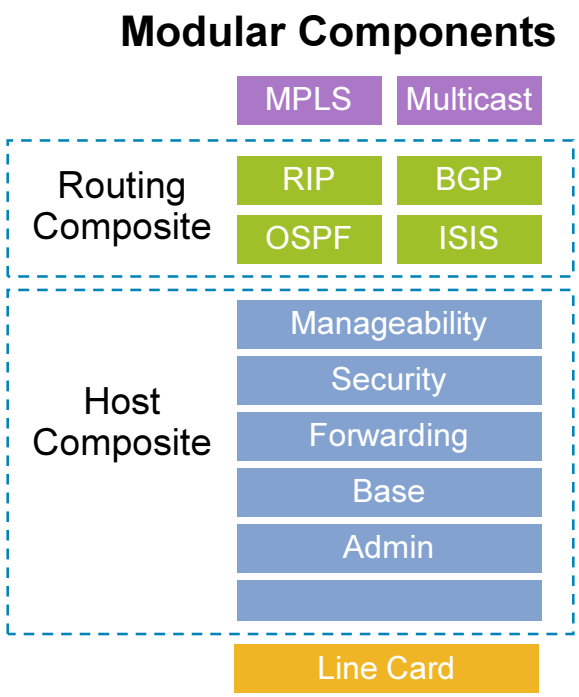
ASR 9000 – na prvi pogled

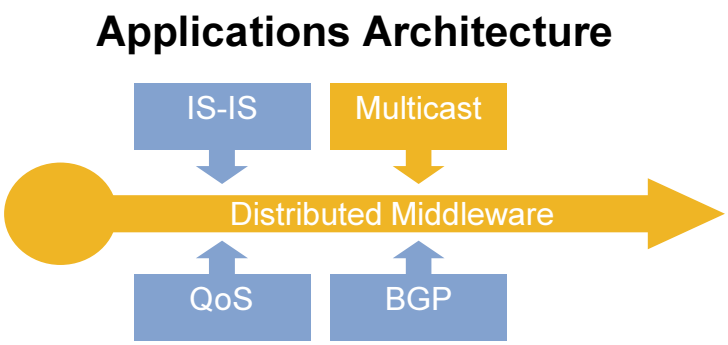
- Optimizovan za L2/L3 10GE i 100GE agregaciju
- Dizajniran za 400Gbps propusnog opsega po slotu
- Baziran na IOS XR operativnom sistemu
- Nadzor i upravljanje sa ANA sistemom

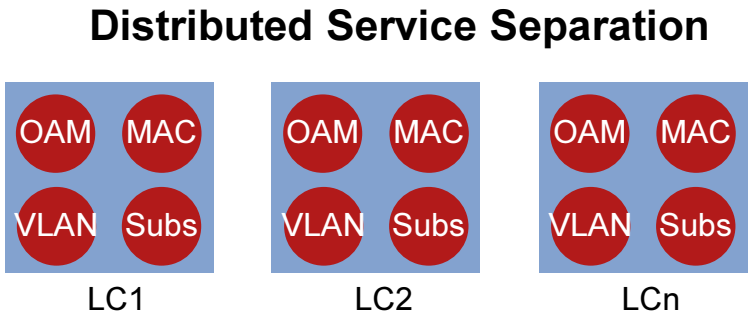


 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Carrier Class OS – IOS-XR







Next Generation Architecture

Designed for Scale, HA, and Performance For IP NGN Applications

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Cisco Expo 2009

ASR 9000

Šasije

© 2008 Cisco Systems, Inc. All rights reserved.

ASR 9010

- 10 slotova
  - 8 linijskih kartica
  - 2 RSP-a
- Visina 21 RU
- AC i DC napajanje
- Protok vazduha spreda prema pozadi

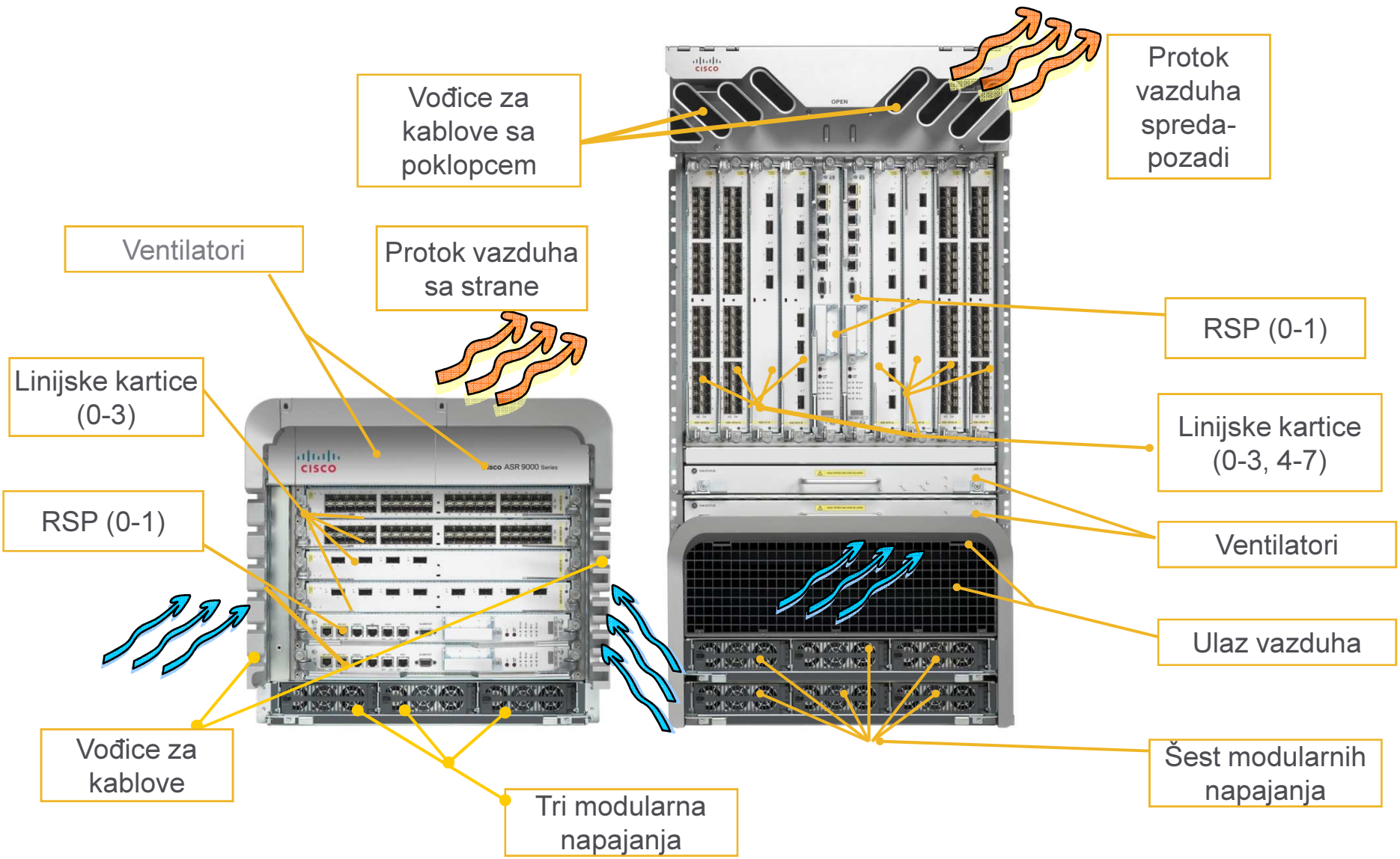
 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

ASR 9006

- 6 slotova
  - 4 linijskih kartica
  - 2 RSP-a
- Visina 10 RU
- AC i DC napajanje
- Protok vazduha sa strane

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Elementi ASR 9010 i ASR 9006 šasija



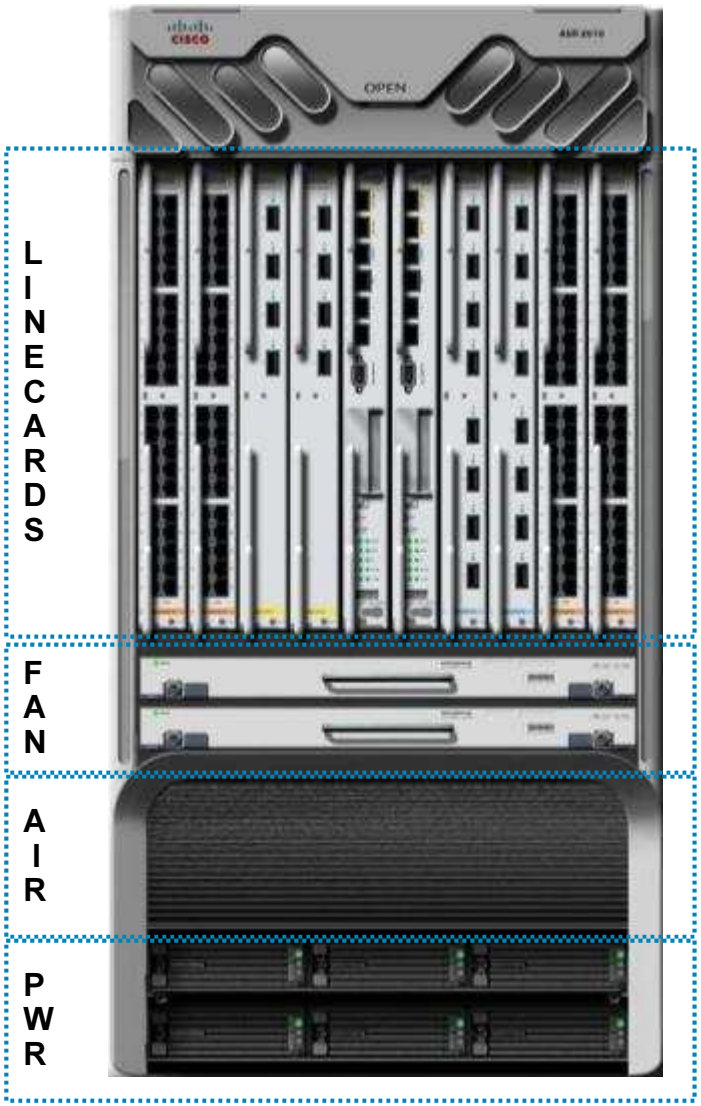
 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

ASR 9010 napajanje i hlađenje

- Kapacitet hlađenja po slotu od 750W
- Dvostruki horizontalni ventilatori ispod linijskih kartica
- AC napajanja od 3kW
- DC napajanja od 2kW ili 1.5 kW
- Konfiguracije napajanja:

  Predviđeno 5+1 DC ili 3+3 AC

  Trenutno 3+1 DC ili 2+2 AC



 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Redundansa ASR 9000 hardvera

- Redundantni RSP/Fabric
- Redudantna napajanja
- Redundantni ventilatori
- Hardverska podrška za Master/Slave RSP arbitraciju
- Hardverska podrška za notifikaciju RSP *switchover*-a ostatku sistema
- Redundantne EOBC (~upravljački Ethernet) putanje; svaka linijska kartica ima po jedan GE link prema svakom RSP-u;
- OIR LK/RSP/Ventilatora nema uticaja na operaciju *fabric*-a tako da nema gubitka paketa

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Cisco Expo 2009

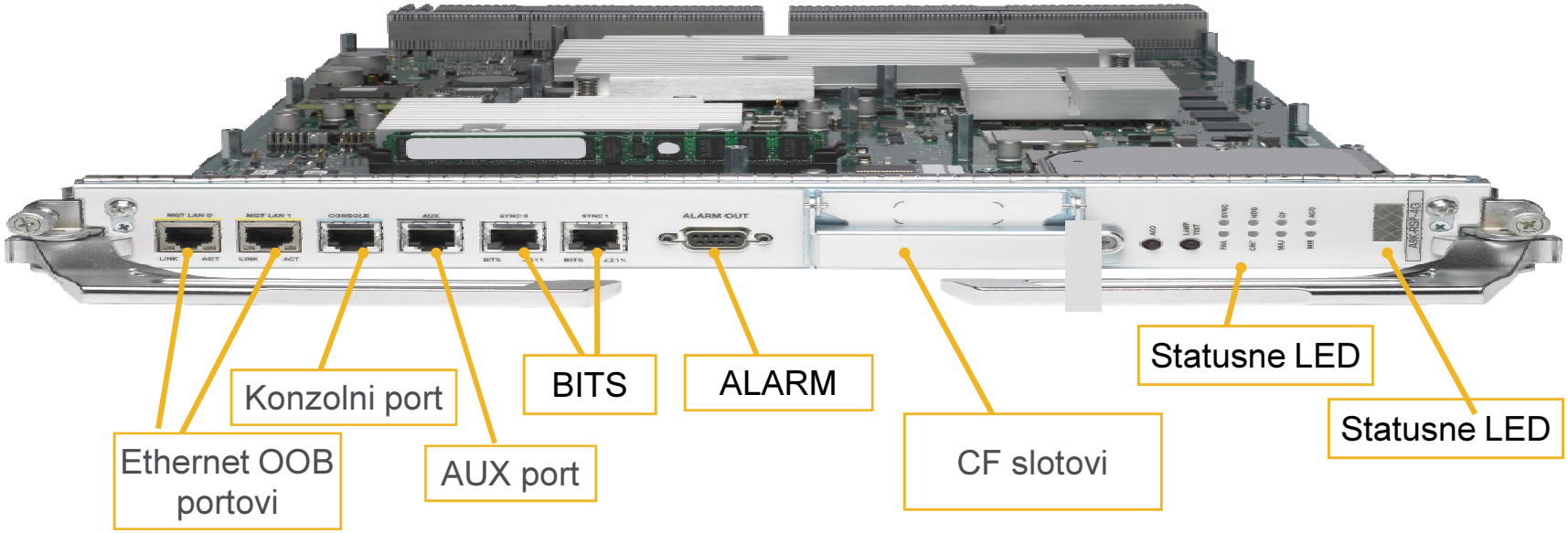
ASR 9000

RSP & *Fabric*

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

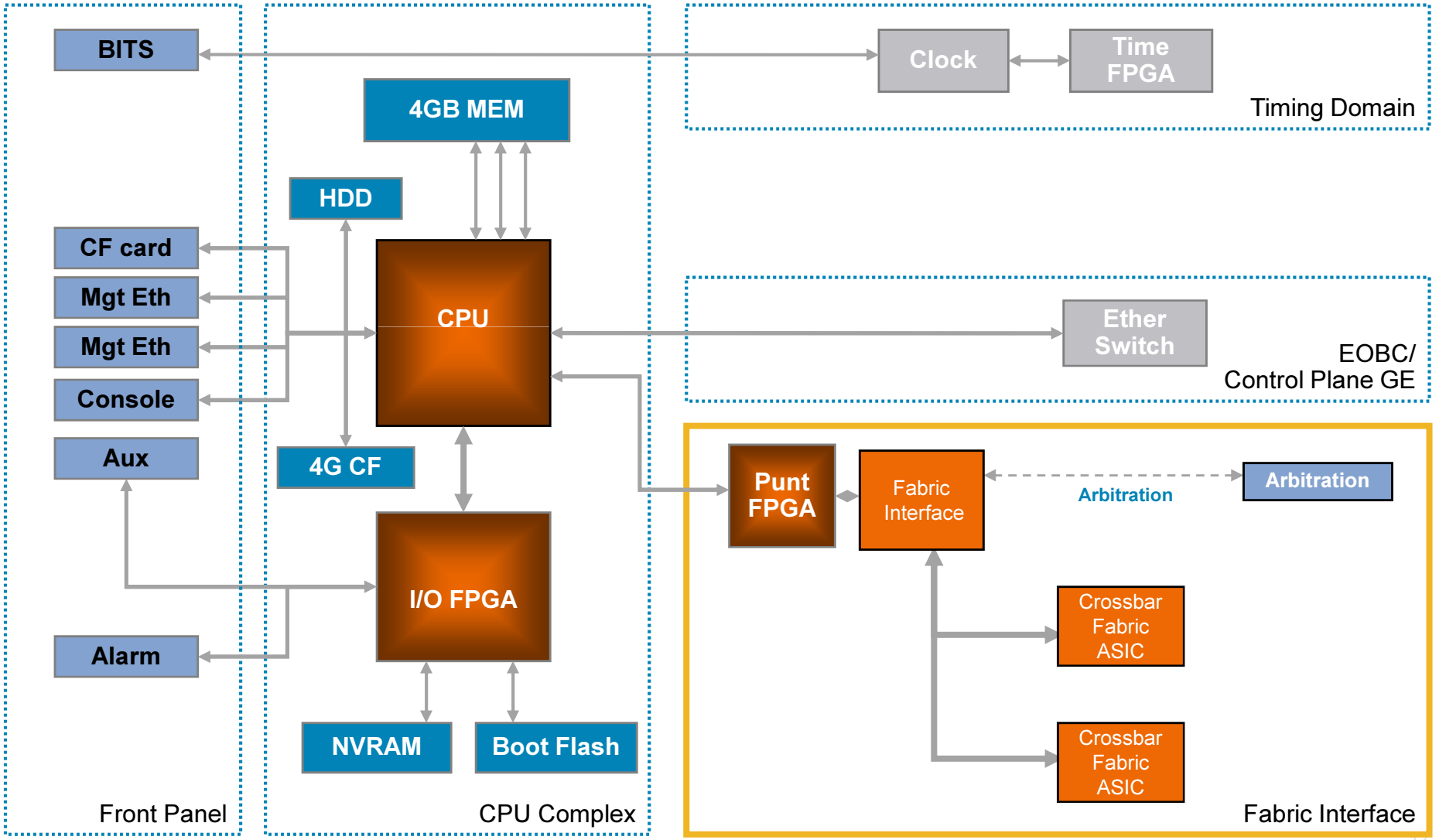
RSP

- Vrši funkcije kontrolne ravni (*control plane*) i upravljanja
- Dual Core CPU sa 4GB DRAM-a
- 2MB NVRAM, 2GB interni *bootdisk*, 2 eksterna CF slota
- Dva *Out-of-band* 10/100/1000 interfejsa za nadzor i upravljanje
- Konzolni i AUX port
- USB portovi za prenos fajlova
- Hard disk od 40GB



 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Arhitektura RSP-a



Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

ASR 9000 *fabric*

- Fizički se nalazi na RSP-u ali je logički potpuno odvojen
- Sadrži centralni arbitar za upravaljanje pristupa *fabric*-u
- Četiri klase za diferencijaciju saobraćaja na ulasku u *fabric*:
  - Dve prioritetne klase (P1, P2)
  - Dve *best effort* klase (BE)
- Sprečava HOLB (*Head Of Line Blocking*) pomoću VOQ mehanizma
- Razdvojeni redovi za *unicast* i *multicast* saobraćaj

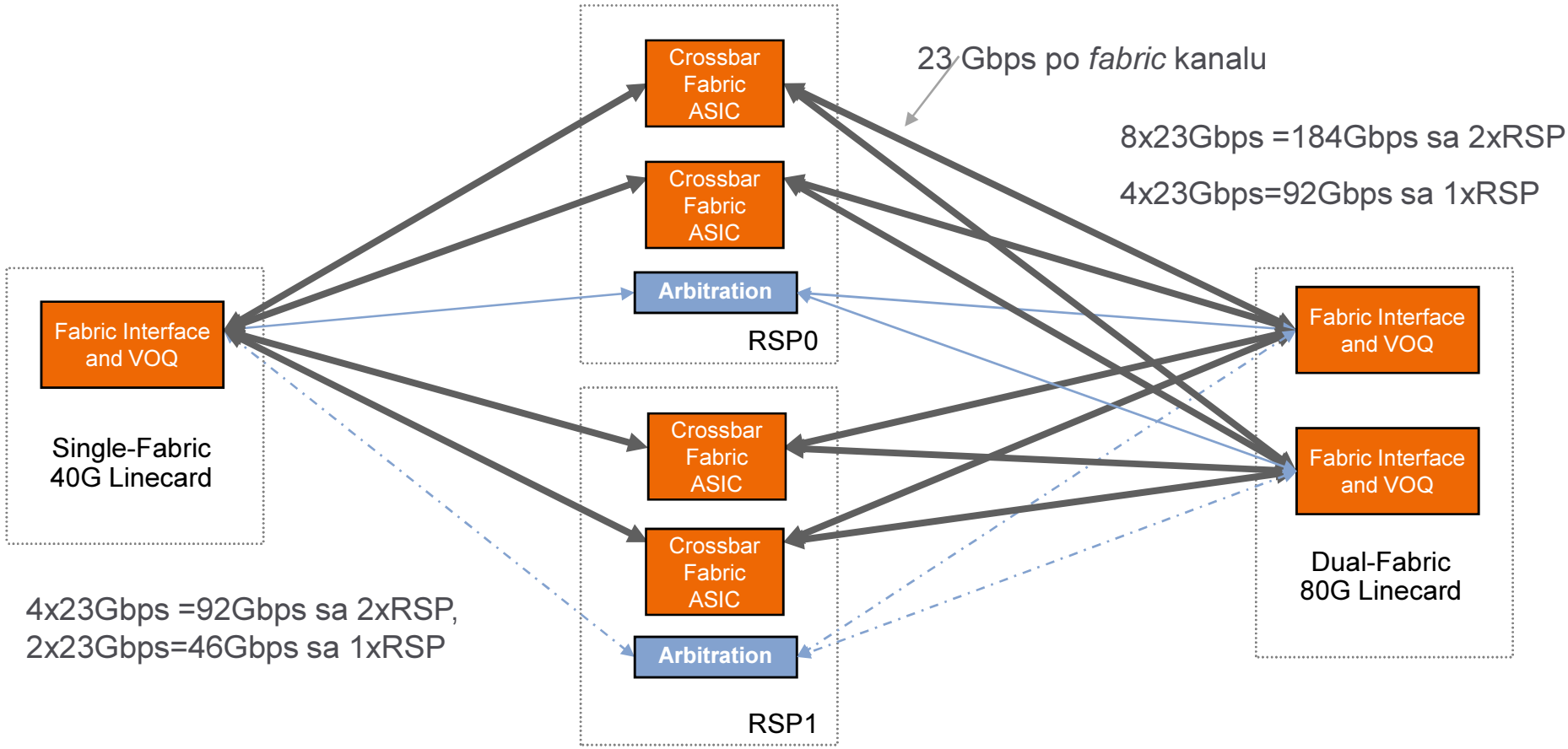
 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

ASR 9000 *fabric* ASIC-i (1)

- F*abric* se sastoji od dva Crossbar Fabric ASIC-a;
- 40Gbps linijska kartica i RSP ima jedan ASIC za interfejs sa *fabric*-om (Fabric Interface ASIC);
  80 Gbps linijska kartica će imati dva takva ASIC-a
- Svaki Fabric Interface ASIC ima po jedan 23 Gbps kanal prema svakom Crossbar Fabric ASIC-u

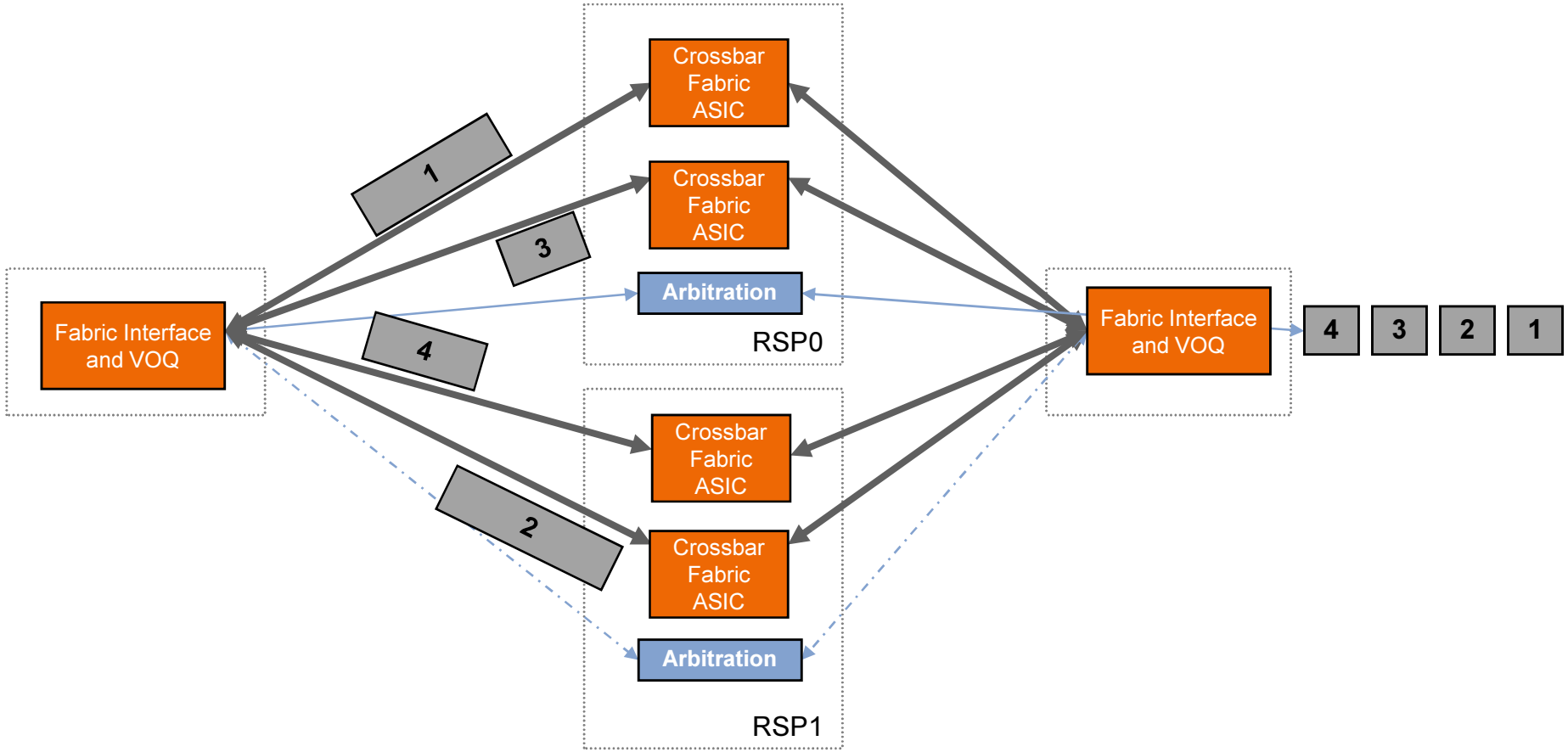
 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

ASR 9000 *fabric* ASIC-i (2)



 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Fabric Load Sharing

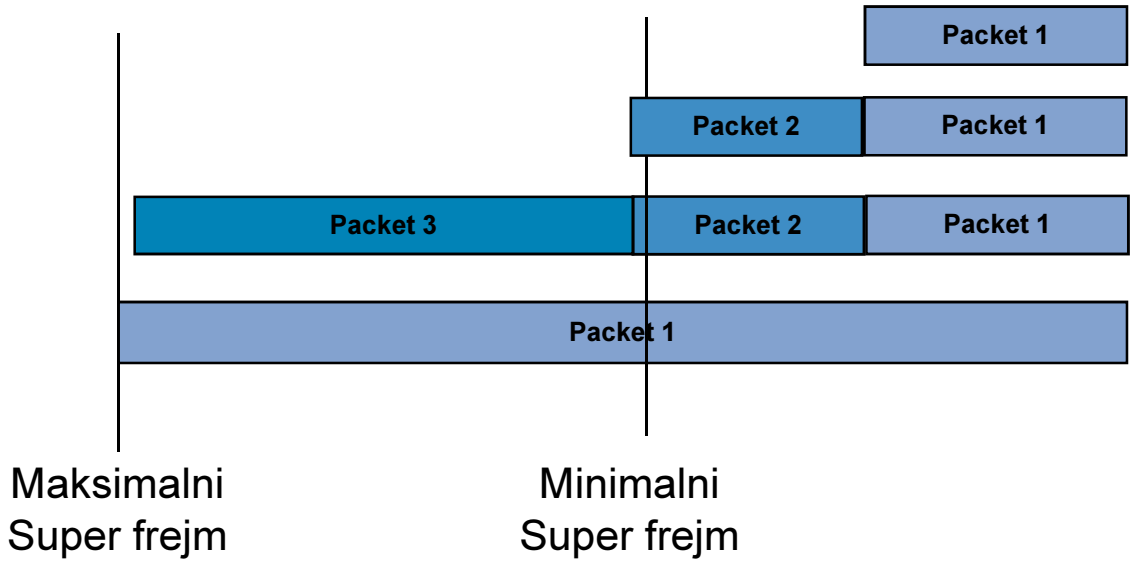


- Unicast saobraćaj se šalje kroz prvi dostupan link prema f*abric-u*
- Svaki ASIC za interfejs prema *fabric-u* sadrži logiku za re-sekvenciranje frejmova; kašnjenje u resekvenciranju se meri u nanosekundama

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Super-frejming mehanizam

- Više *unicast* frejmova (do 32) prema istom odredištu se može agregirati u jedan super frejm
- Značajno poboljšava ukupan protok prema *fabric*-u



 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Arbitracija u *fabric-u*

- Pristup *fabric*-u se kontroliše pomoću centralnog arbitracionog ASIC-a
- U jednom trenutku samo jedan arbitracioni ASIC vrši arbitraciju
- Algoritam arbitracije prepoznaje QoS i obezbeđuje da da P1 klasa ima prioritet prema P2 klasi pre čemu obe te klase imaju prioritet prema neprioritetnim klasama
- Arbitracija obezbeđuje jednak pristup propusnom opsegu svim linijskim karticama
- Arbitracija se vrši relativno prema izlaznom (*egress*) NPU na izlaznoj linijskoj kartici

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Virtualni Izlazni Redovi (*Virtual Output Queues* – VOQs)

- Kapacitet *fabric* kanala prema izlaznom modulu se reprezentuje sa Virtualnim Izlaznim Redovima (*Virtual Output Queues* – VOQs) na ulazu u *fabric* sa ulaznog modula
- Izlazni modul sadrži po jedan VOQ za svaku od četiri klase saobraćaja svakog NPU-a na izlaznom modulu:

  Jedan NPU obuhvata:

  Jedan fizički10G port (neblokirajuća linijska kartica)

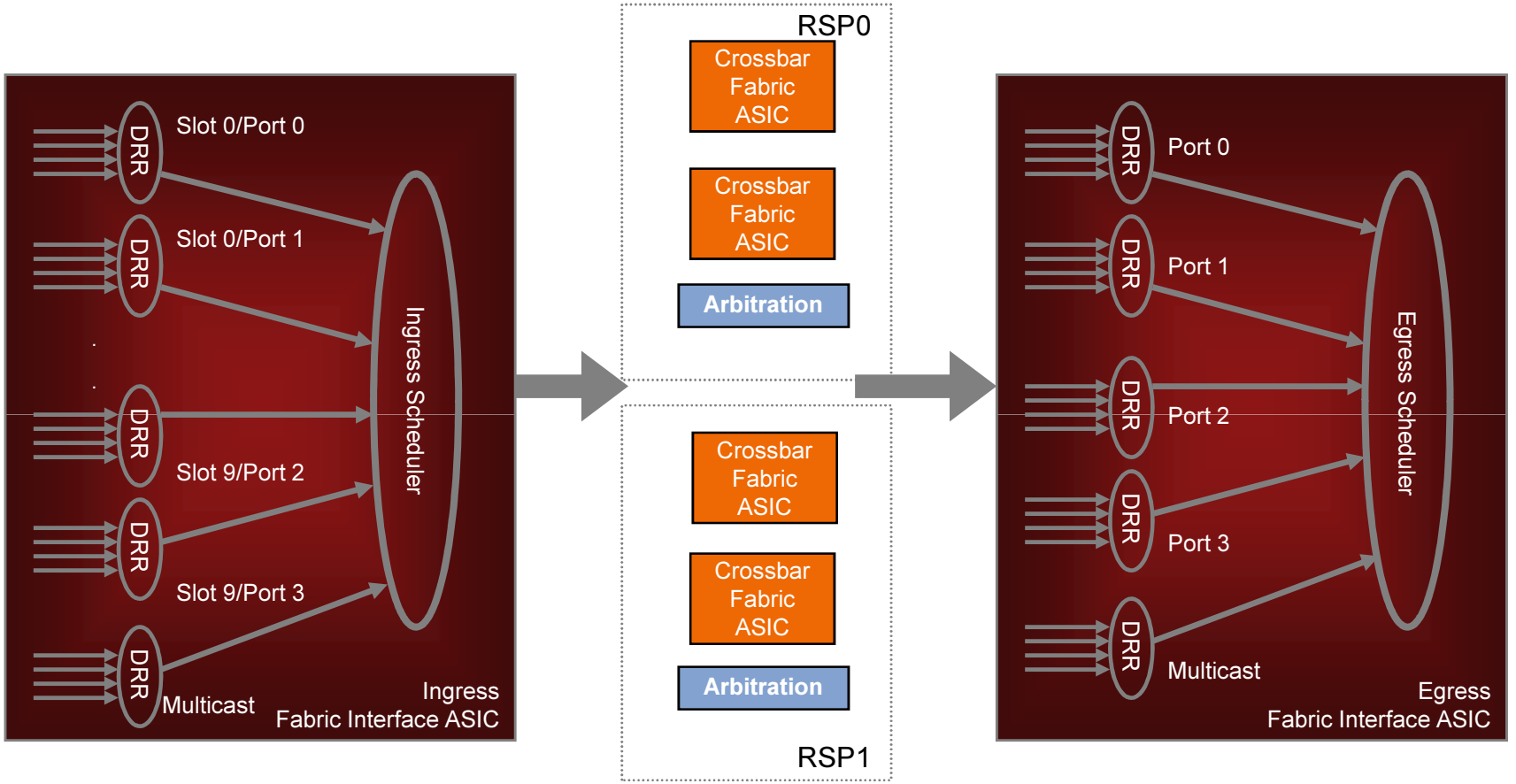
  Dva fizička10G porta (blokirajuća linijska kartica)

  Deset 10/100/1000 portova

- Za svaki primljeni paket, arbitar određuje da li za njega ima mesta u odgovarajućem VOQ ili ne

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

VOQ-evi i izlazni *fabric* redovi



136 ulaznih VoQ-ova:

- 8 odredišnih LK * 4 10G portova/LK * 4 klasa/port** == 128 VoQ-ova
- 2 odredišna RSP-a * 1 10G port/RSP * 4 klasa/port == 8 VoQ za RSP-eve
- 4 reda za multicast

20 izlaznih redova:

- 4 klasa/port * 4 porta/LK (unicast) == 16
- 4 multicast klasa == 4

Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Cisco Expo 2009

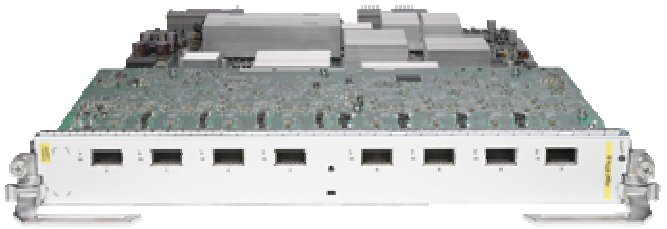
ASR 9000

Linijske kartice

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

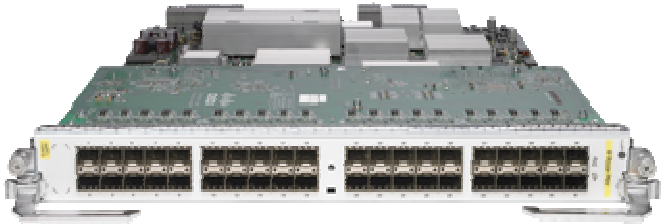
ASR 9000 linijske kartice (1)

- Trenutno dostupne linijske kartice:
  - 4x10GE
  - 40x1GE
  - 8x 10GE (2:1 *oversubscription*)
- IP, Ethernet, MPLS L2VPN/L3VPN servisi
- 1M IPv4 ruta
- 1M MAC adresa
- 32000 L2/L3 interfejsa
- 384k redova



A9K-8T/4-E

A9K-4T-B



A9K-40GE-B

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

ASR 9000 linijske kartice (2)

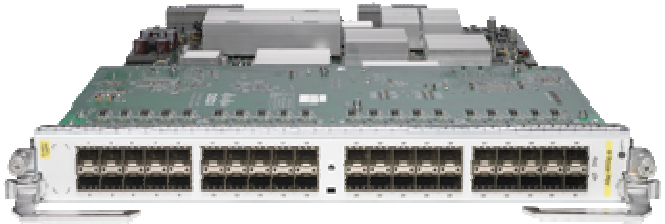
- Memorijske varijante:

  Base

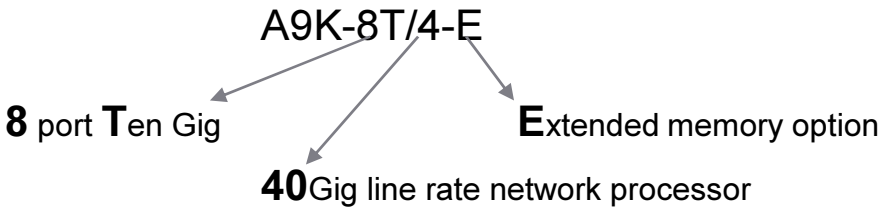
  *Extended (*isti broj ruta, 2x više interfejsa, 3x više redova*)*

A9K-8T/4-E

A9K-4T-B



A9K-40GE-B



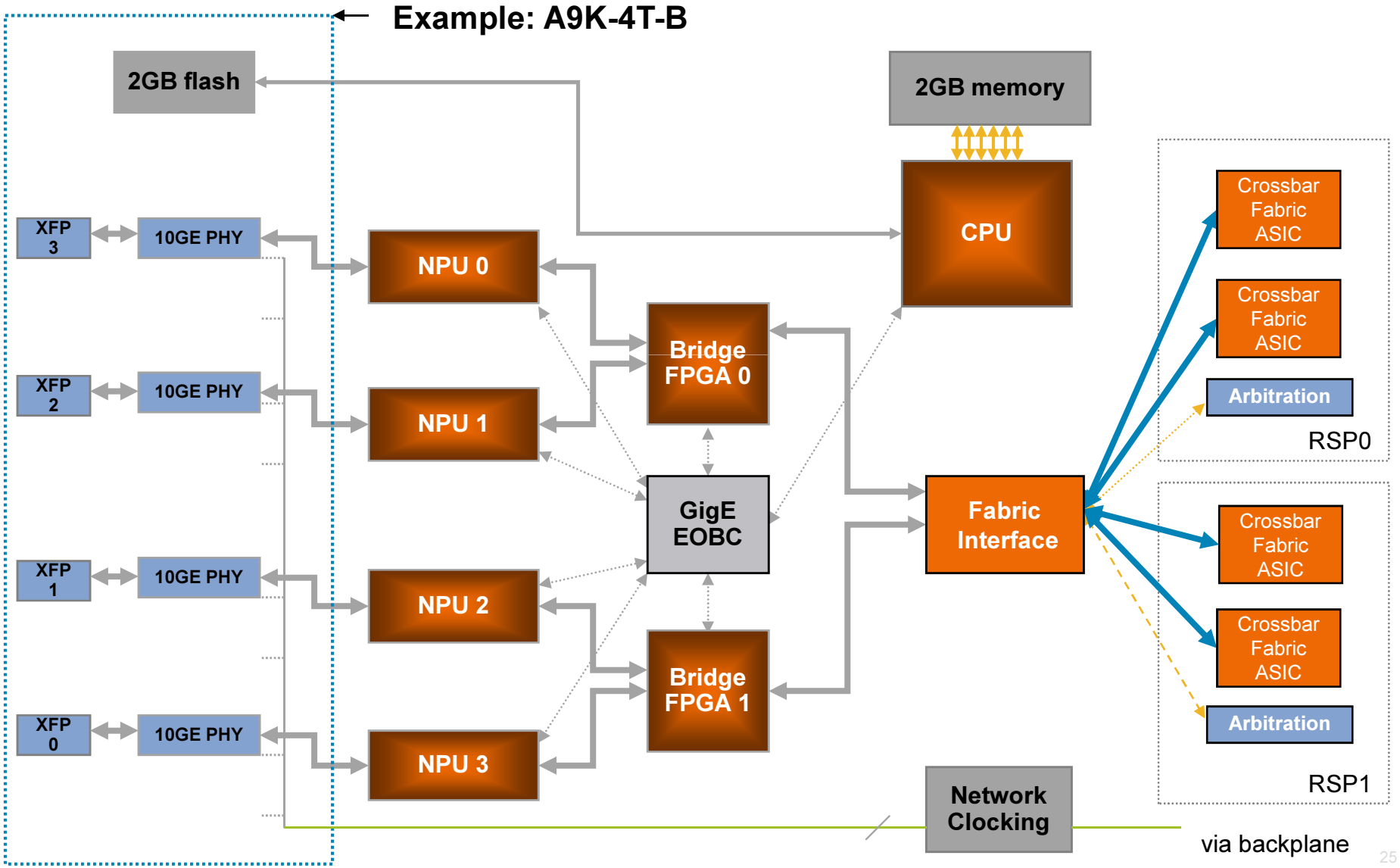
 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Komponente ASR 9000 linijskih kartica

- ASIC za interfejs prema *fabric-u*
  - Sadrži jedan bidirekcioni 23 Gbps kanal prema svakom *Crossbar Fabric* ASIC-u
  - Sa dva RSP-a u sistemu, svaka linijska kartica ima 4x23 Gbps=92 Gbps propusnog opsega kroz *fabric*
  - Sadrži tabelu za replikaciju multicast saobraćaja i vrši replikaciju prema *Bridge FPGA*
- *Bridge FPGA*
  - Povezuje NPU i ASIC za interfejs prema *fabric-u*; vrši konverziju između NPU zaglavlja i *fabric* zaglavlja
  - Sadrži hardverske redove za sistemski QoS
  - Sadrži tabelu za replikaciju *multicast* saobraćaja i vrši replikaciju prema NPU-ovima
- Trident NPU (*Network process unit*)
  - Glavni procesor paketa, vrši L2 i L3 *lookup* kao i replikaciju *multicast* saobraćaja
- CPU
  - Istovetan CPU kao na RSP-u
  - Određeni protokoli kao BFD, CFM, ARP se distribuiraju na CPU-ove linijskih kartica
  - Lokalni_proces dobija FIB tabelu sa RP-a i programira hardverske tabele na NPU, *Bridge FPGA*-u i Fabric interface ASIC-u

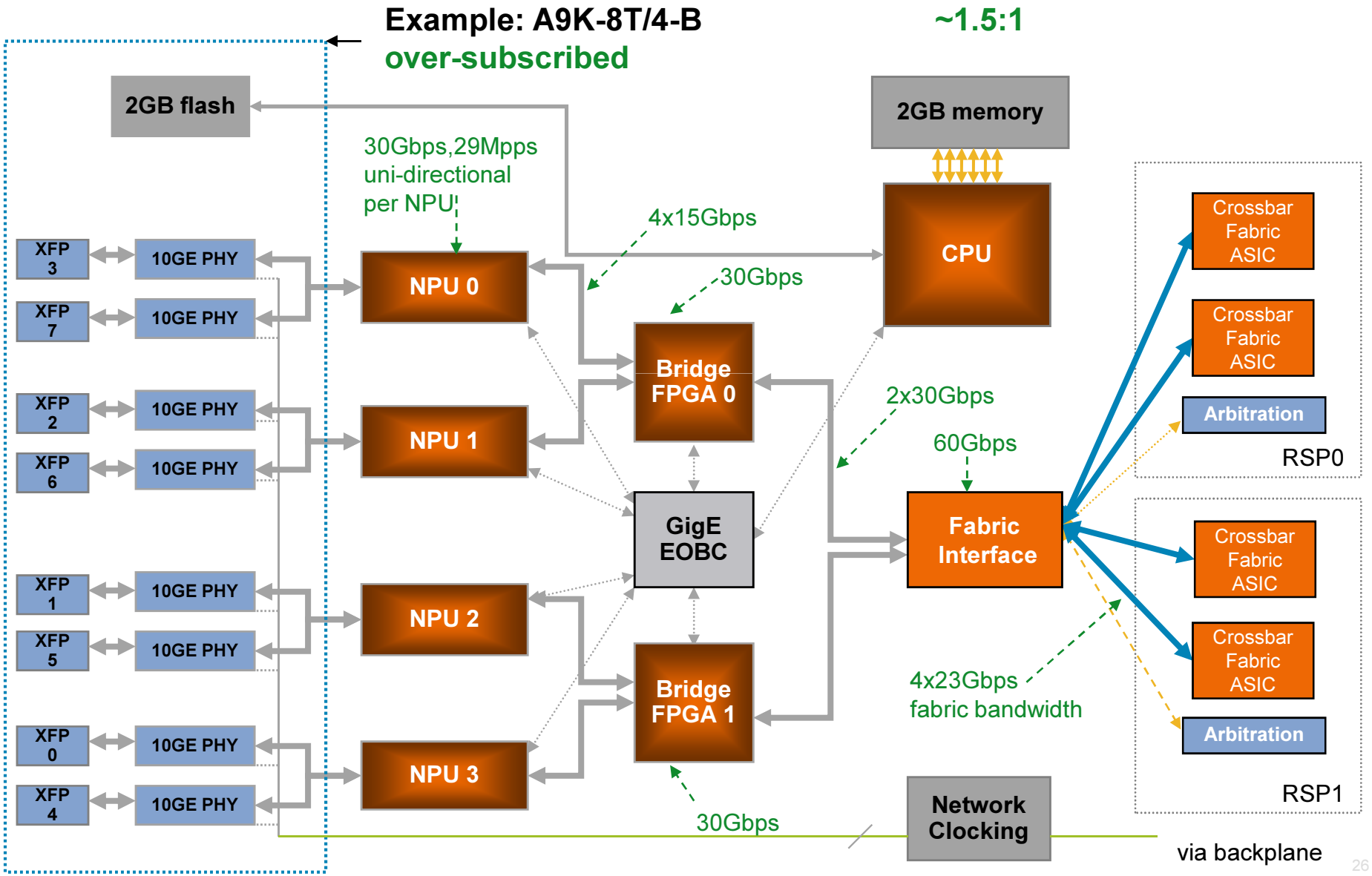
 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Arhitektura A9K-4T-B linijske kartice



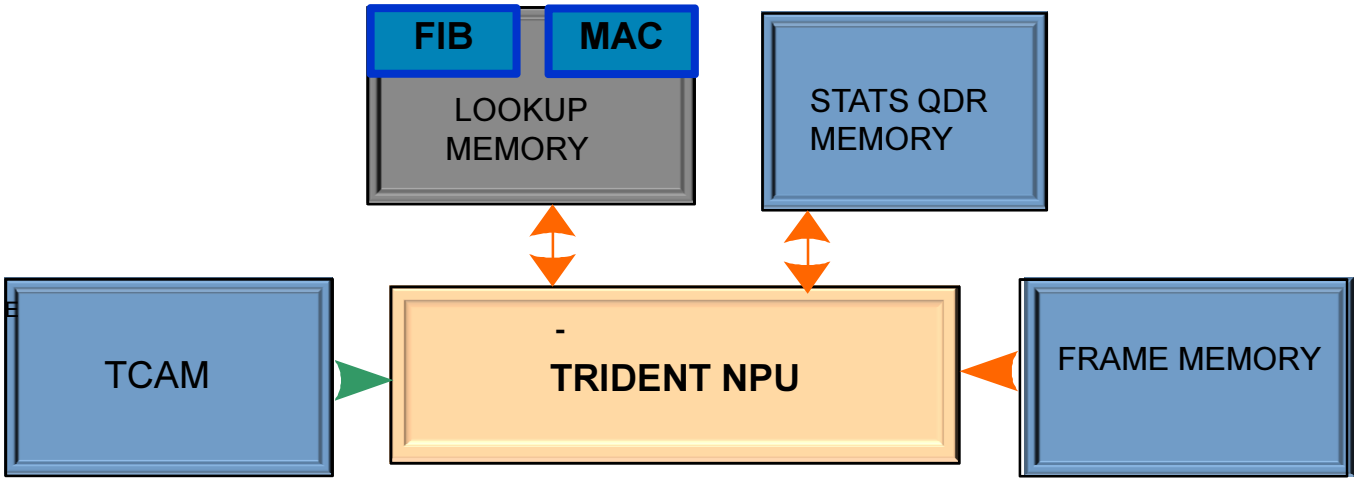
© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Arhitektura A9K-8T/4-B linijske kartice



© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Trident Network Process Unit (NPU)



- 10 Gbps bidirekcioni protok sa uključenim servisima
- Četiri pridružene memorije:
  - TCAM memorija se koristi za VLAN tag, QoS i ACL klasifikaciju
  - *Lookup* memorija čuva FIB tabelu, tabelu MAC adresa i CEF susedstva
  - Stats QDR memorija čuva statistike interfejsa itd.
  - *Frame* memorija je bafer za sve redove
- *Base* i *Extended* memorijske opcije imaju istu veličinu *Lookup* memorije (tj. maksimalni broj ruta) ali različite veličine ostalih memorija

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Cisco Expo 2009

ASR 9000

Arhitektura prosleđivanja saobraćaja

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Prosleđivanje paketa na ASR 9000

- Prosleđivanje paketa je u potpunosti distribuirano na linijske kartice
- *Two-stage prosleđivanje:*

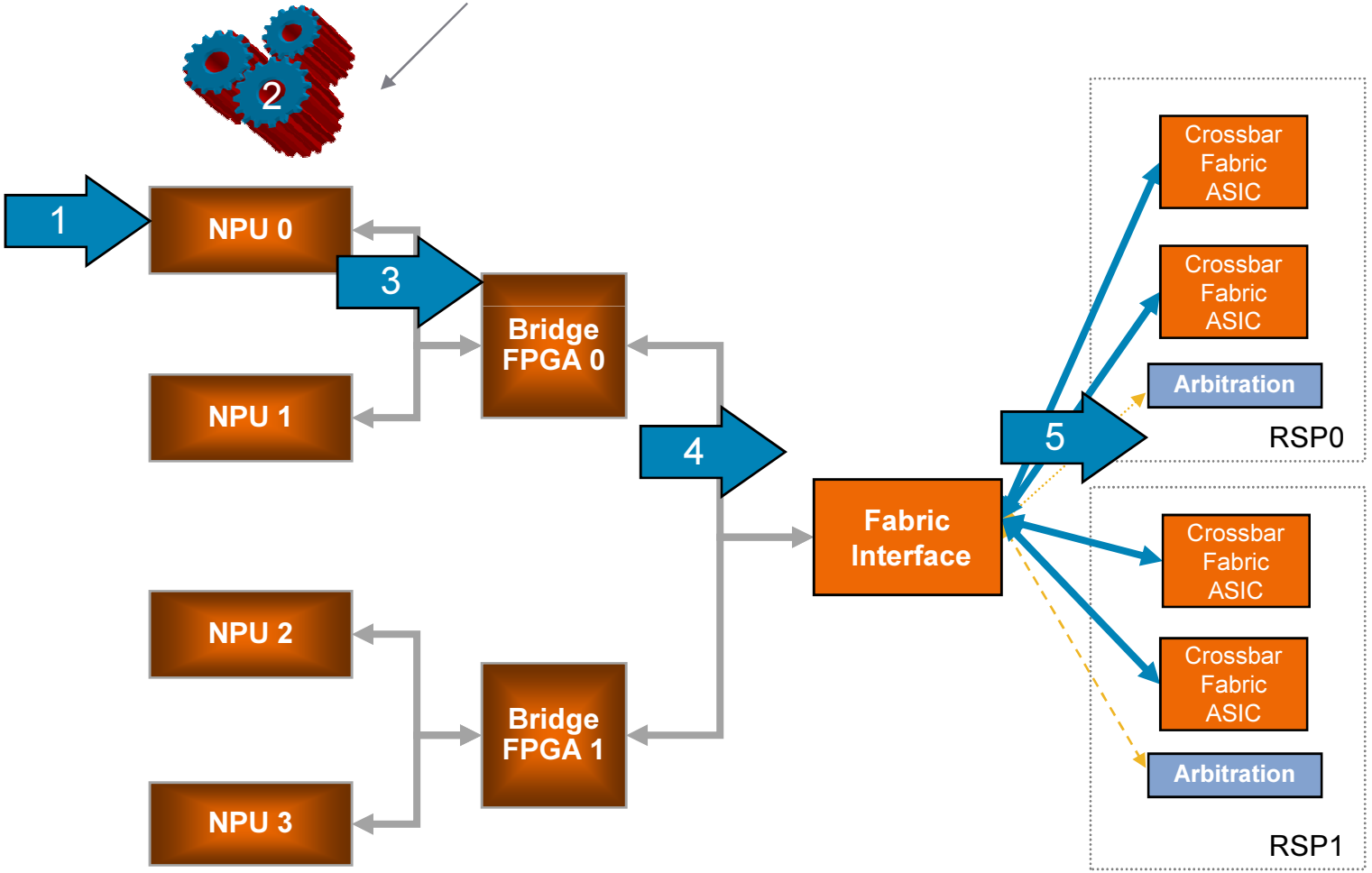
  Prvi (ulazni) *lookup* se radi na ulaznoj linijskoj kartici i daje izlazni *Switch Fabric Port* (SFP)

  Drugi (izlazni) *lookup* se radi na izlaznoj linijskoj kartici i daje izlazni fizički port
- ARP/CEF Adjacency informacija je lokalna na kartici

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Put paketa – ulazna linijska kartica

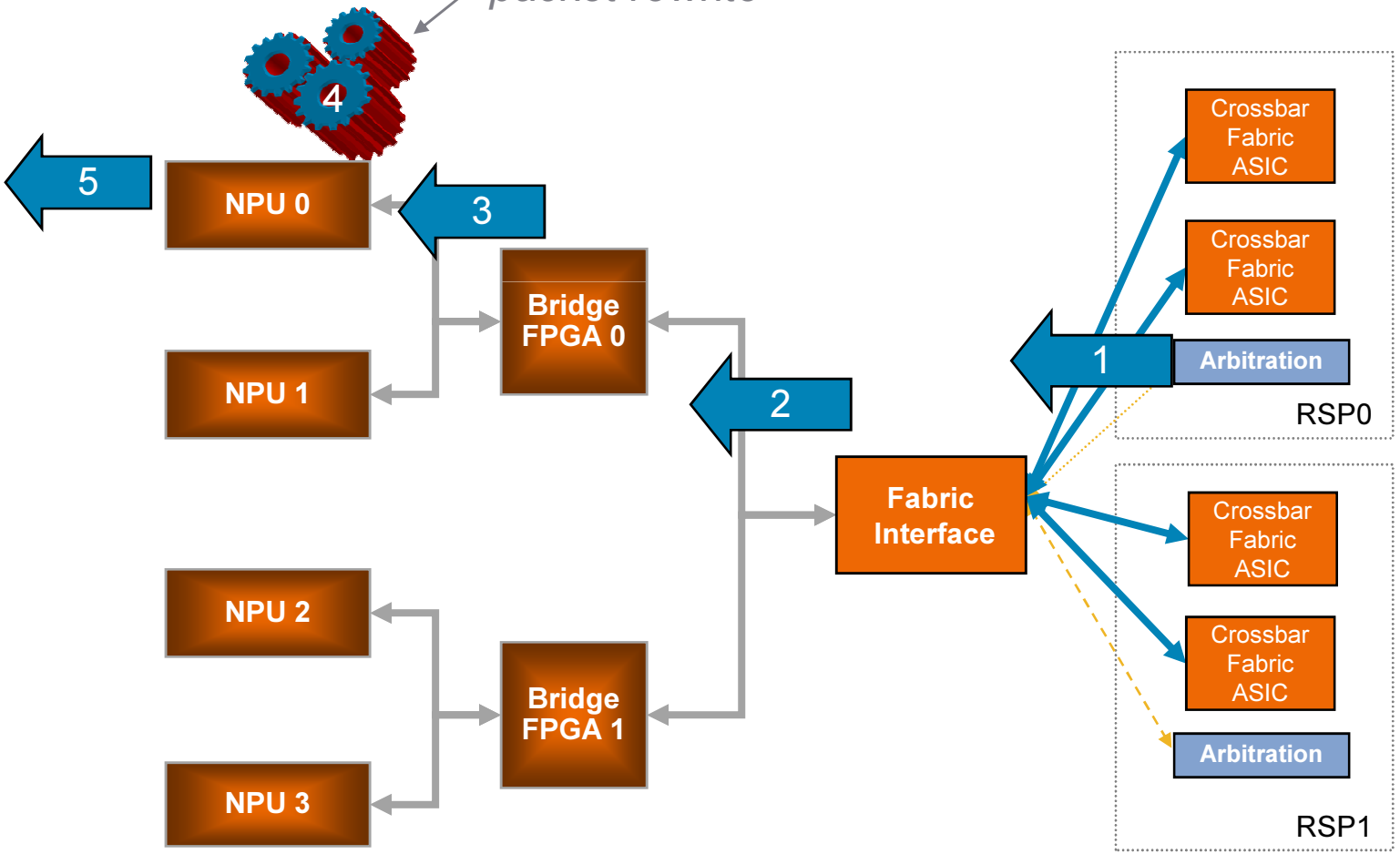
Prvi *lookup* na ulaznom NPU-u; vraća odredišni NPU (Switch Fabric Port – SFP)



 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Put paketa – izlazna linijska kartica

Drugi *lookup* se vrši na izlaznom NPU; vraća odredišni port kao i informacije za *packet rewrite*



 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

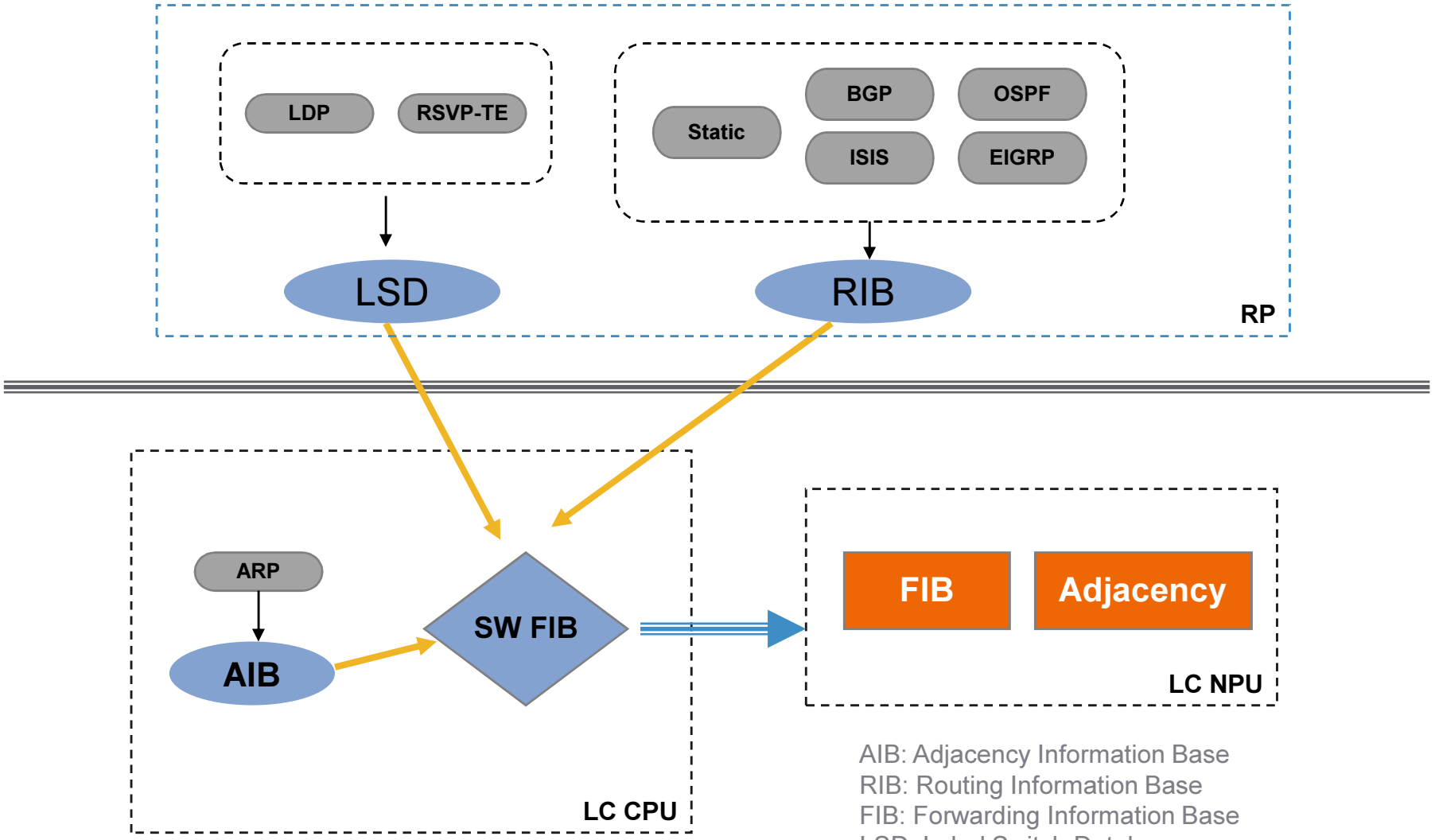
Cisco Expo 2009

ASR 9000

Prosleđivanje L3 paketa

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

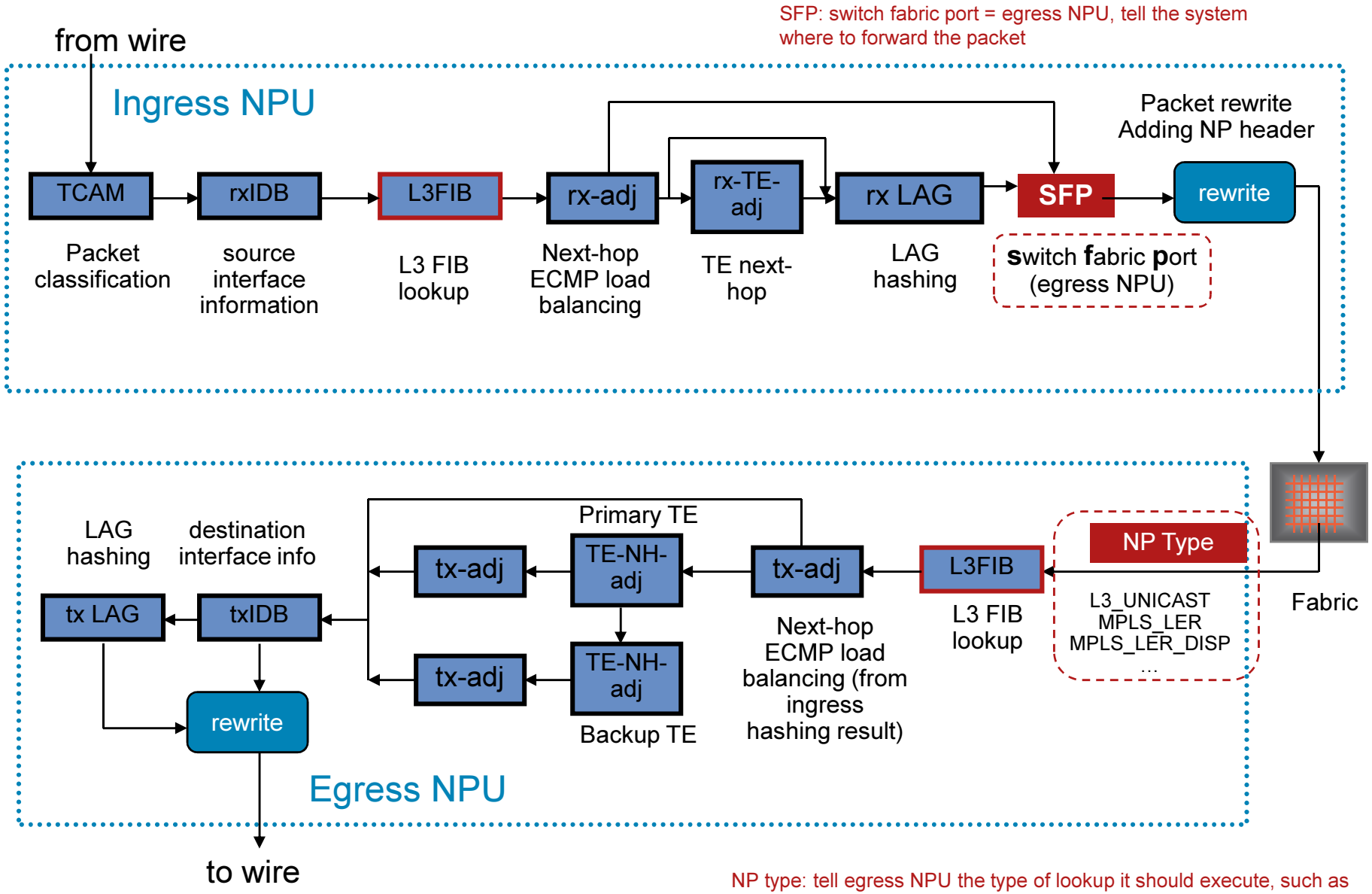
Pregled L3 kontrolne ravni



AIB: Adjacency Information Base
RIB: Routing Information Base
FIB: Forwarding Information Base
LSD: Label Switch Database

Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Prosleđivanje L3 paketa



NP type: tell egress NPU the type of lookup it should execute, such as L3 unicast lookup, MPLS label switch, MPLS label disposition, etc

Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Cisco Expo 2009

ASR 9000

Prosleđivanje L2 paketa

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Prosleđivanje L2 paketa

- L2 servisi:

  Point-to-point; prenos frejmova između dva logička porta bez učenja MAC adresa

  Multi-point; bridžovanje između dva logička porta bazirano na naučenim MAC adresama; logički port je lokalni L2 interfejs ili VPLS/EoMPLS PW

- L2 strukture:

  XID – logički interfejs; loklani XID je lokalni L2 (pod)interfejs ; globalni XID je PW ili *link bundle*

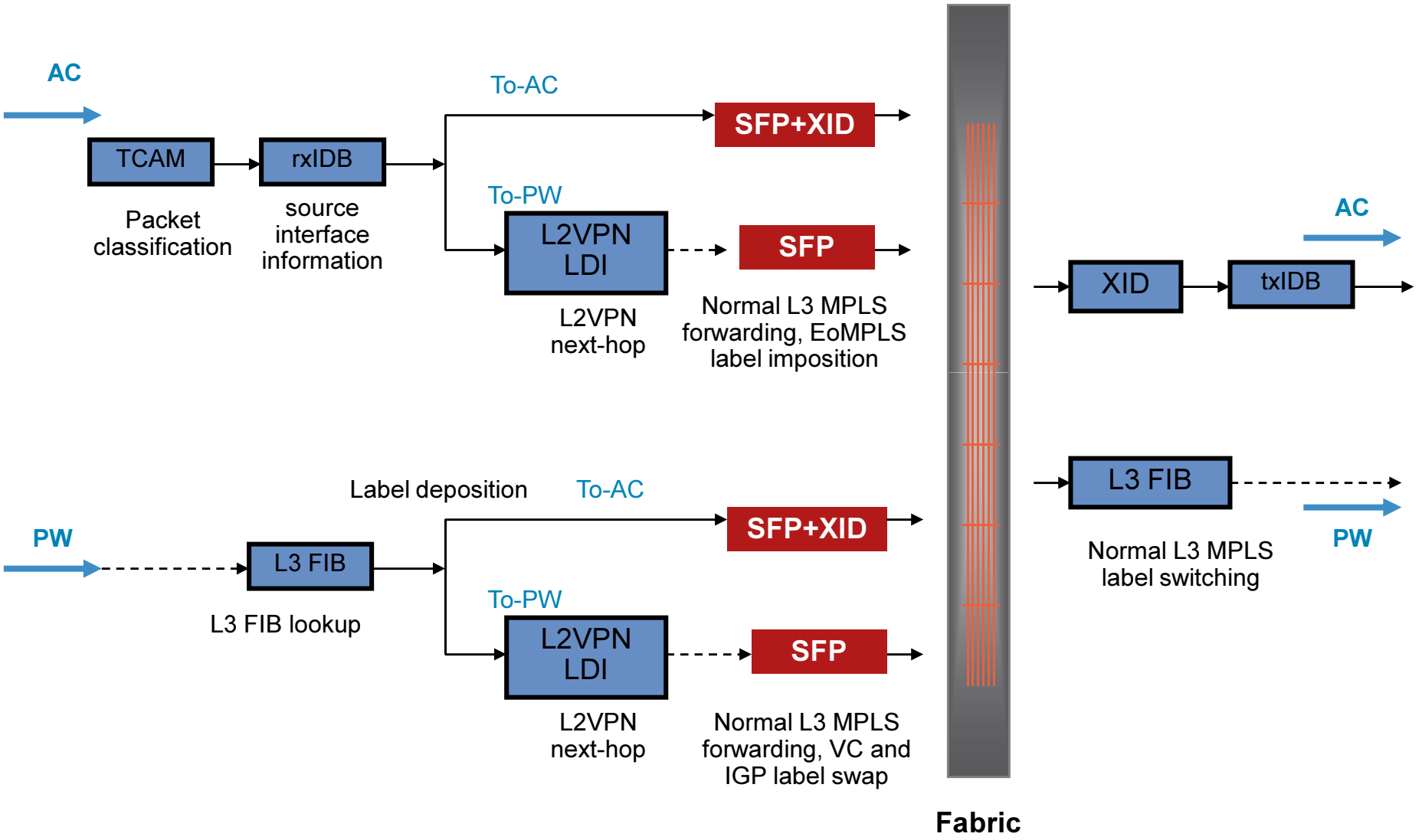
  L2VPN LDI (loadinfo); ako je logički port PW onda mu se pridružuje L2VPN LDI za impoziciju MPLS labele i prosleđivanje

  Tabela bridž domena; sadrži listu logičkih interfejsa na NPU-u i pointer na tabelu MAC adresa

  Tabela MAC adresa

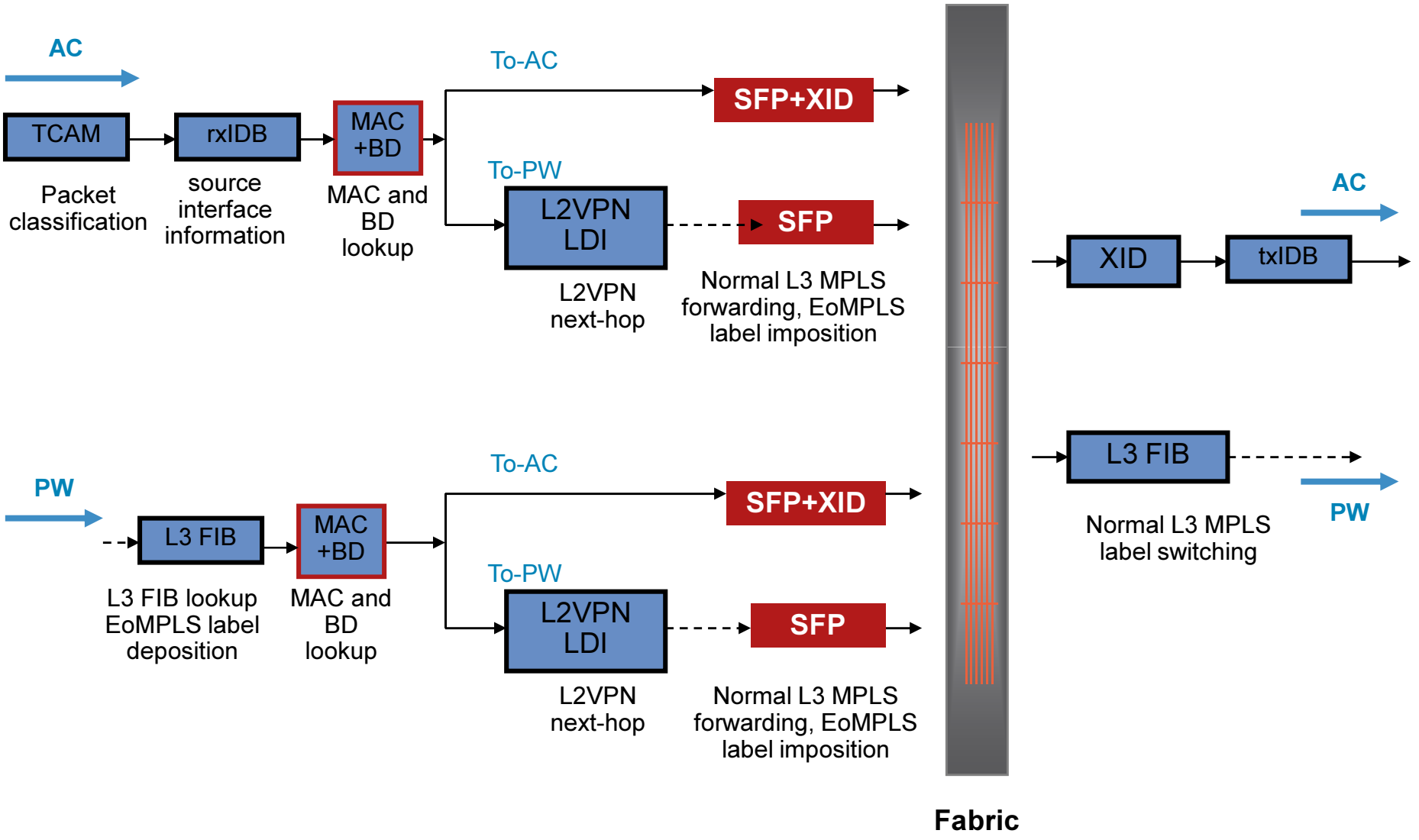
Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Prosleđivanje L2 paketa u *point-to-point* servisima



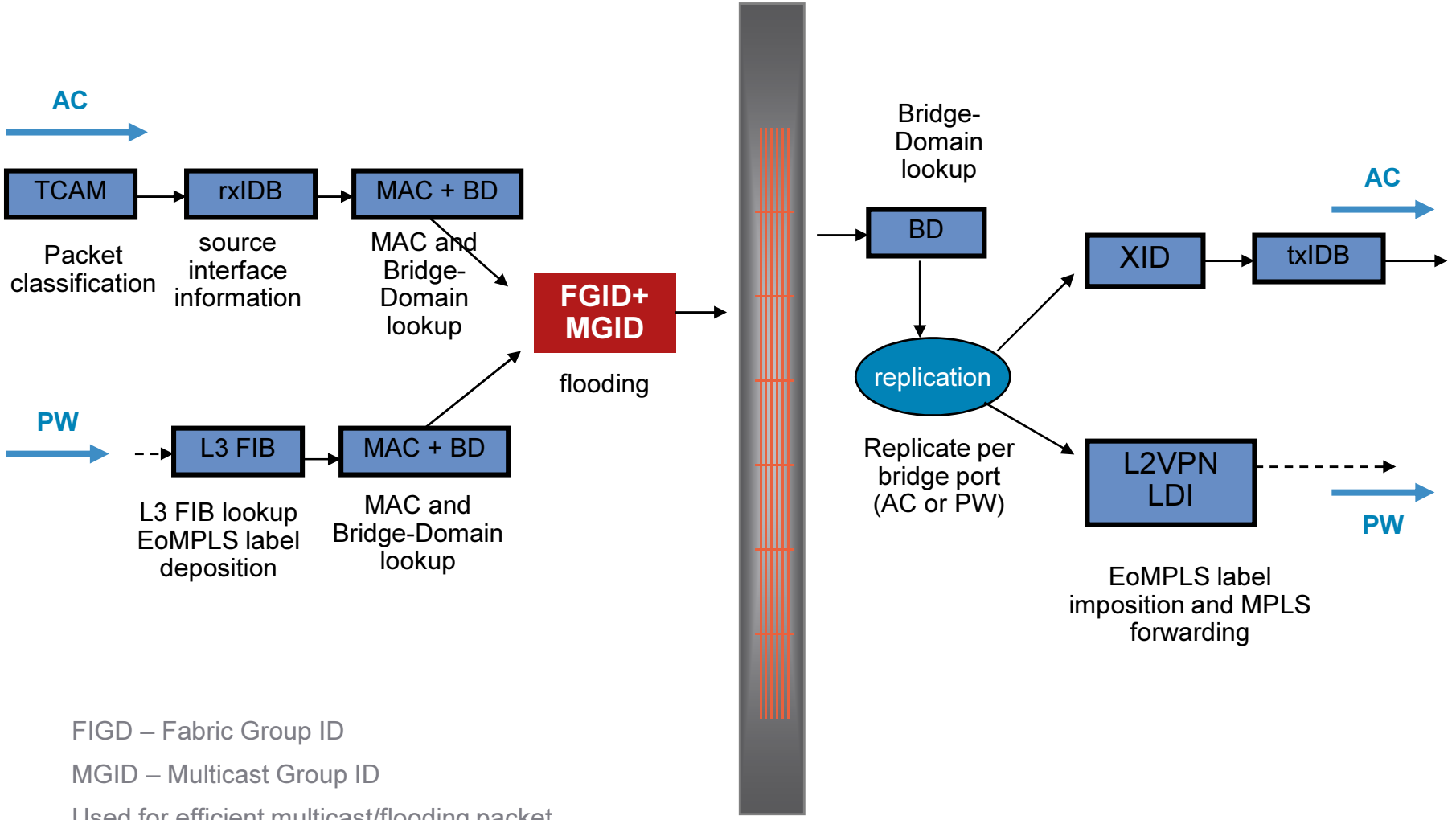
Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Prosleđivanje L2 paketa u *multi-point* servisima



Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Flooding L2 paketa u *multi-point* servisima



FIGD – Fabric Group ID

MGID – Multicast Group ID

Used for efficient multicast/flooding packet replication, refer to “multicast architecture”

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Cisco Expo 2009

ASR 9000

Prosleđivanje *multicast* paketa

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

ASR 9000 m*ulticast* arhitektura

- Kontrolna ravan
  - Svi *multicast* protkoli se izvršavaju na RP-u
  - CPU na linijskoj kartici obrađuje posebne *multicast* pakete (softversko prosleđivane)
- Ravan prosleđivanja podataka
  - *Multicast* prosleđivanje je u potpunosti distribuirano na linijske kartice
  - *Two stage* prosleđivanje
- Hardverska replikacija paketa
  - Replikacija *multicast* paketa se vrši unutar *fabric*-a odnosno izlazne linijske kartice; ulazna linijska kartica nikad ne vrši replikaciju
  - *Multicast* paketi se repliciraju isključivo u hardveru; *multicast* je *line rate* i u slučaju potpuno popunjene šasije
- L2 vs. L3 multicast
  - L2 i L3 *multicast* imaju razdvojene kontrolne ravni ali zajedničku ravan podataka

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Pregled L3 *multicast* kontrolne ravni

- PIM i IGMP paketi se šalju direktno na RSP CPU
- RSP CPU ažuira MRIB tabelu
- MRIB tabela se distribuira na LK gde se zapisuje u MFIB tabelu; na osnovu MRIB tabele se kreira FPOE tabela na *fabric-u* koja definiše replikaciju kroz *fabric*
- CPU na linijskoj kartici programira MGID (Multicast Group ID) tabele na Fabric ASIC-u, FPGA-ovima i NPU-ovima koja definiše replikaciju na tim elementima

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

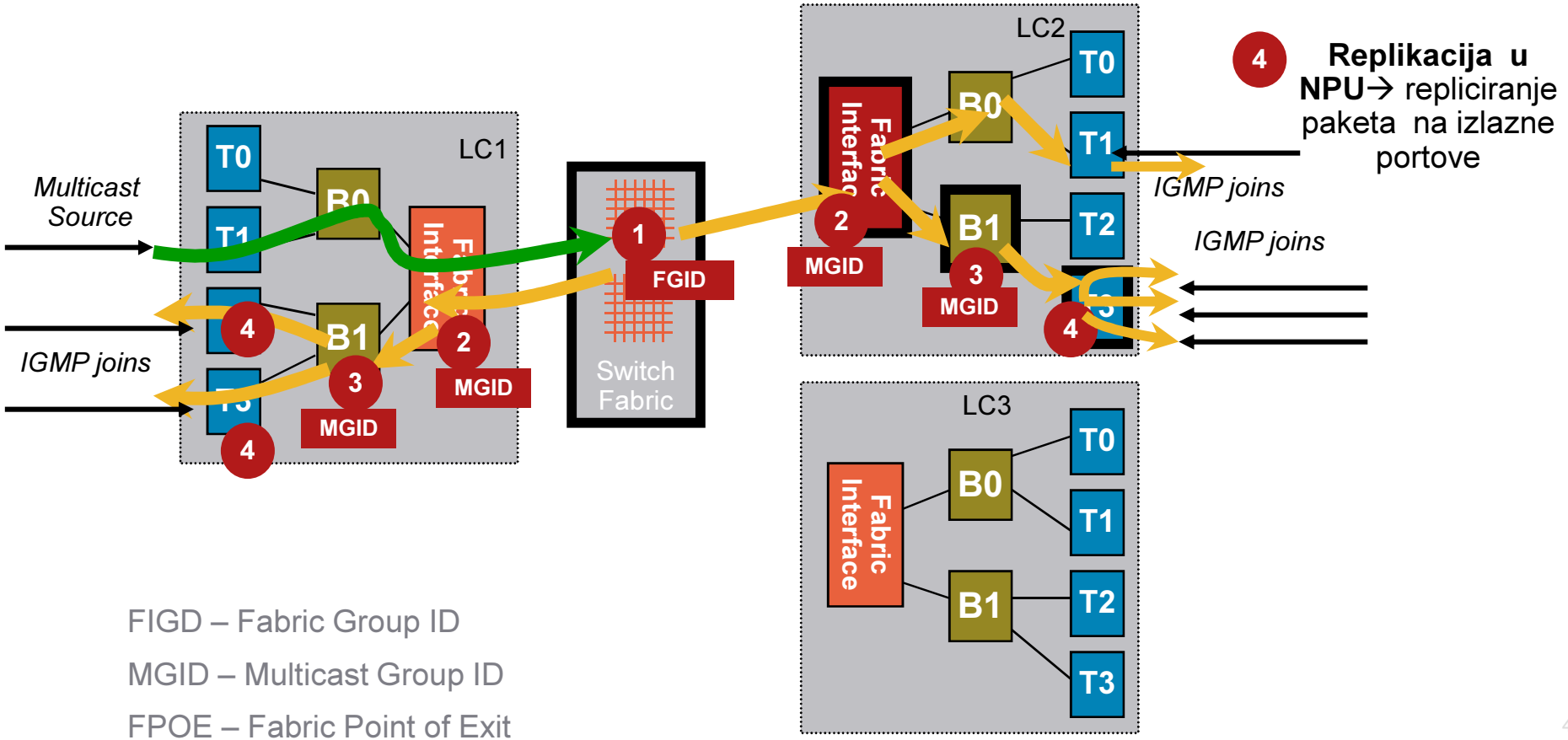
Replikacija multicast paketa

1 **Replikacija u Fabricu** → repliciranje paketa na izlaznu LK prema FGID & FPOE tabelama

2 **Replikacija u Frabric interface ASIC-u**→ repliciranje paketa na Bridge FPGA prema MGID tabeli

3 **Replikacija u Bridge FPGA** → repliciranje paketa na NPU prema MGID tabeli

4 **Replikacija u NPU**→ repliciranje paketa na izlazne portove



FIGD – Fabric Group ID

MGID – Multicast Group ID

FPOE – Fabric Point of Exit

© 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Miercom ASR9000 Test Report

Competitive Performance Testing Results

Carrier Class Ethernet Services Routers

Cisco Aggregation Services Router 9000 Series

v

Juniper MX 960 Ethernet Services Router

27 August 2009

http://www.miercom.com/dl.html?fid=20090827&type=report

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

Cisco Networkers
25-28. januar 2010.
Barselona
Registrujte se

Presentation_ID © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential

 © 2008 Cisco Systems, Inc. All rights reserved. Cisco Confidential