Arhitekture omrežij za podatkovne centre

Zasnovane s stikali zgrajenimi s standardnimi SoC

peter.reinhardt@xenya.si

XENYA

Novi trendi v Podatkovnih centrih (PC)

- Virtualizacija aplikacij in storitev
  - Zagotavlja neodvosnost od HW, enstaven backup, optimizacija porabe resursov ....
- Centralizacija aplikacij in storitev (cloud, private cloud ...)
- Zmanševanje porabe energije
- Neodvisnost od proizvajalcev strojne opreme, več virov strojne opreme
- Enovita skladiščna in podatkovna omrežja
- Avtomatizirani postopki upravljanja (orkestracija)

XENYA

Novi trendi v izvedbi in upravljanju omrežij

- Dinamično prilagajanje omrežja aplikacijam
  - Dinamično kreiranje VLANov hkrati s kreiranjem/selitvijo VM
- Orkestracija - Odprte rešitve, neodvisne od strojne opreme, ki omogočajo poenostavitev upravljanja standardnih nalog in integracijo upravljanja strežnikov in omrežja
- SDN – Funkionalna Delitev Nadzorne in Podatkovne ravni v omrežjih, centralizacija Nadzorne ravni v zunanji SW krmilnik, aplikativni nadzor obnašanja omrežja
  - OpenFlow, OpenAPI
- Virtualizacija omrežnih komponent – Del funkcij omrežja se seli na strežnike v obliki virtuelnih stikal, usmerjevalnikov, požarnih pregrad, load balancerjev ...., Omrežje dobi dodatne konfiguracijske (aplikativne) podatke iz virtuelnih stikal.

XENYA

Novi trendi v izvedbi IP omrežij PC

- Velike hitrosti prenosa in prepusnosti komunikacijskega omrežja
  - Vmesniki proti klijentom 10G, 40G in 25G, vmesniki v hrbtenici 40G in 100G
  - Velika prepustnost za komunikacjo znotraj podatkovnega centra - horizontalna komunikacija (Oversubscription od 1:6 do 1:1)
  - Majhne zakasnitve in ultra majhne zakasnitve, ...
- Enotna struktura omrežja v celotnem Podatkovnem centru (PC)
  - “Fabric” omrežja – enaka pasovna širina, enake zakasnitve med priključki
  - Podpira enostavno optimizacijo uporabe resursov
- Zagotavljanje redundance z izborom arhitekture & protokolov v PC, ne z uporabo kompleksnih komponent, ki že same zagotavljajo redundanco
  - Boljše razmerje sposobnosti/cena, Hitrejše prilagajanje novim tehnologijam, nižja poraba energije, nijža cena ...
- Nove arhitekture - (spine&leaf), virtuelne komponente, SDN ...
- Overlay protokoli
  - npr. za tuneliranje L2 nad L3 (TRILL, VXLAN, NVGRE...)
- Novi protokloli za podporo enovitim omrežjem - učinkovito prenašajo klasične mrežne protokole in skladiščne protokole (FCoE) – DCB

XENYA

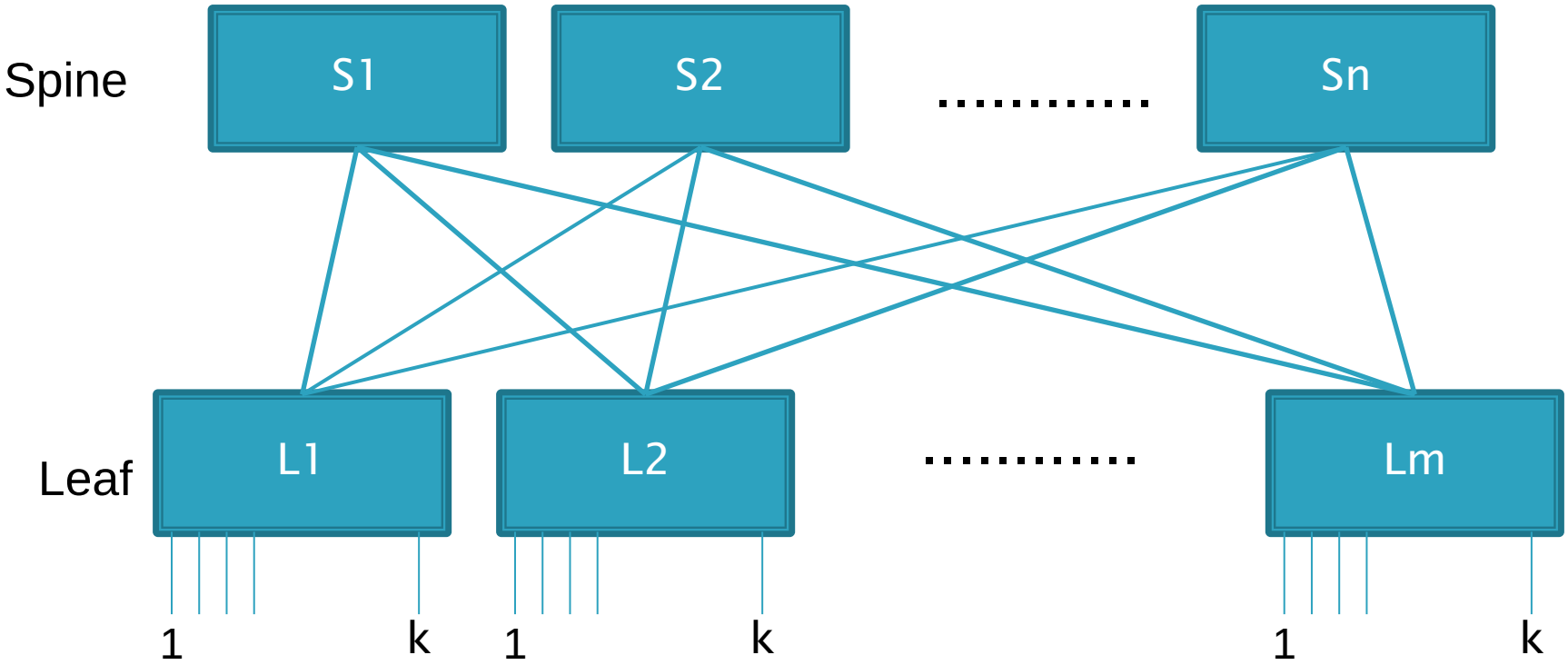
Fizični del omrežja pod. centra

- Lastnosti Spine/Leaf arhitekture (primerne za več kot 4 stikala)
  - Izpolnjuje vse zahtevane komunikacijske lastnosti:
    - visoka prepustnost, vsi priključki imajo isto pasovno širino, iste zakasnitve (Fabric)
  - Enostavno vzdrževanje:
    - Ena ali dva tipa standardnih stikal sestavljajo celo omrežje
  - Redundantna zasnova
    - Omejen vpliv okvare na povezavi/stikalu
  - Razširljivo do zelo velikih omrežij ....
  - Nizka poraba energije
- Glavni parametri za konstrukcijo:
  - Hitrost priključka
  - Število priključkov
  - Razmerje agregirane pasovne širine vseh priključkov klijentov proti agregirani pasovni širini povezav navzgor – Oversubscription
  - Izjemoma tudi zahtevana Latenca

XENYA

Arhitekture Spine/Leaf

- Spine / Leaf



Število priključkov: $N_C=m*k$

Oversubscription: $OSR= K*S_C /n*S_{up}$

Pasovna širina: $S_T=S_{up}*n$

XENYA

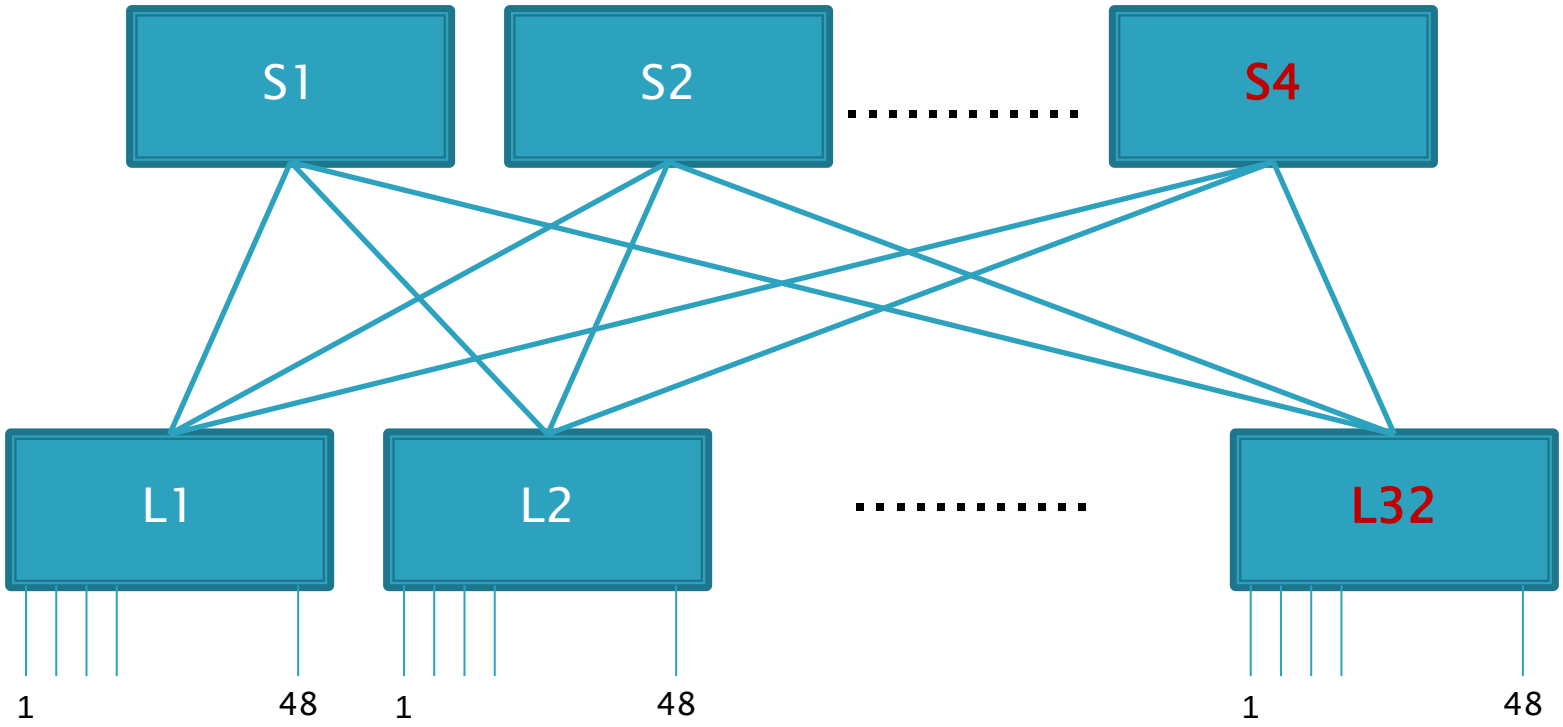
Nekaj Primerov izvedenih s 3 tipi stikal

- **LY2R**
  - 48 x .....1x10Gb (SFP+) ali 1x1Gb (SFP) in
  - 4 x .....1x40Gb ali 4x10Gb (QSFP+)
- **LY6**
  - 32x .....1x40Gb ali 4x10Gb (QSFP+)
- **IX1**
  - 32 x 1x100Gb, 2x50Gb, 2x40Gb, 4x25Gb (QSFP28)

XENYA

Spine /Leaf – Primer 1

Stikala: Leaf: (LY2R) 48x1/10Gb+ 4x40Gb
Spine: (LY6) 32x40Gb



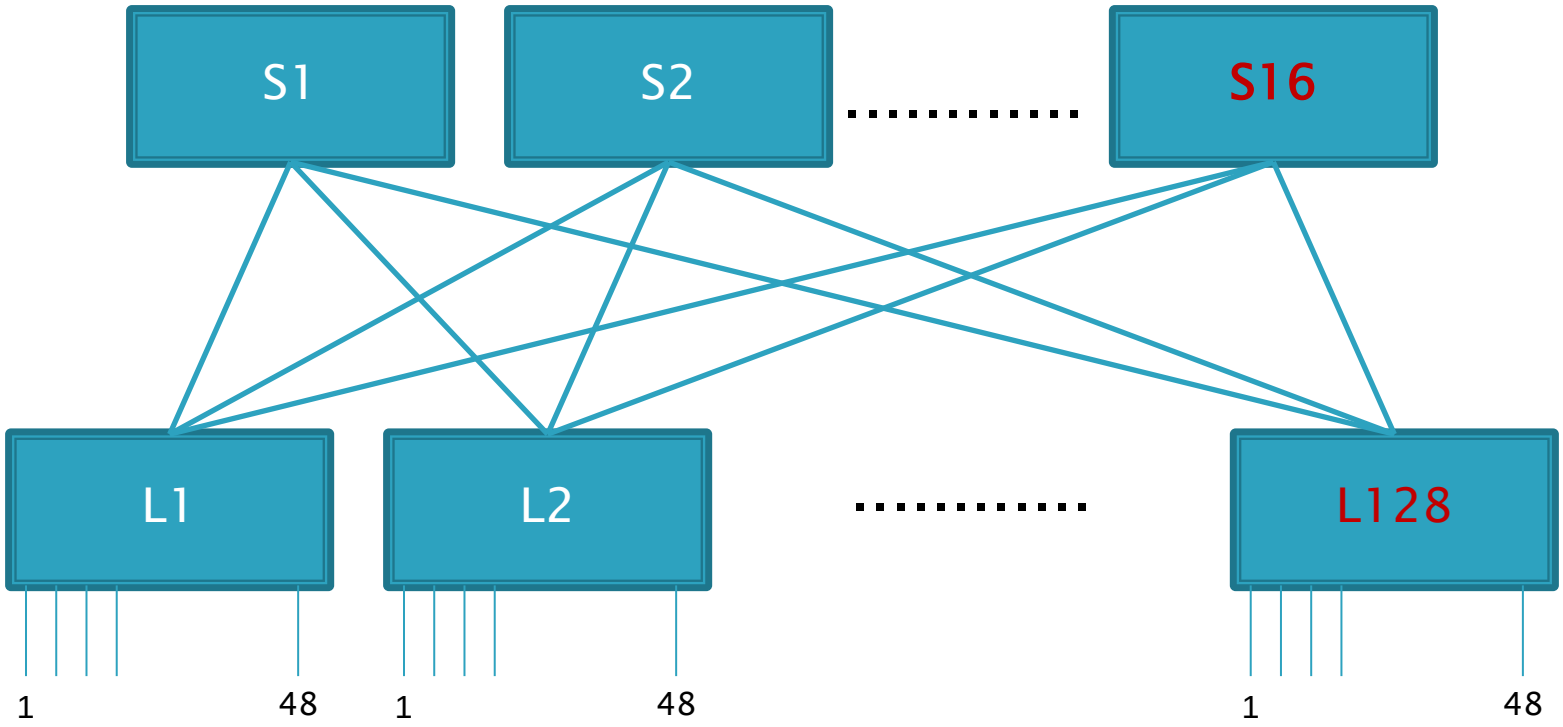
Število priključkov (10Gb) N_C=32*48=1536
Oversubscription OSR= 48*10/4*40=3
Pasovna širina S_T=160Gb

XENYA

Spine /Leaf – Primer 2

(enaka stikala kot v primeru 1, drugačna vezava)

Stikala: Leaf: (LY2R) 48x10Gb + 4x40Gb uporabljeni kot 16x10Gb (LY1R)
Spine: (LY6) 32x40Gb kot 32x4x10Gb



Število priključkov	N_C=128*48=6144
Oversubscription	OSR= 48*10/16*10=3
Pasovna širina	S_T=160Gb

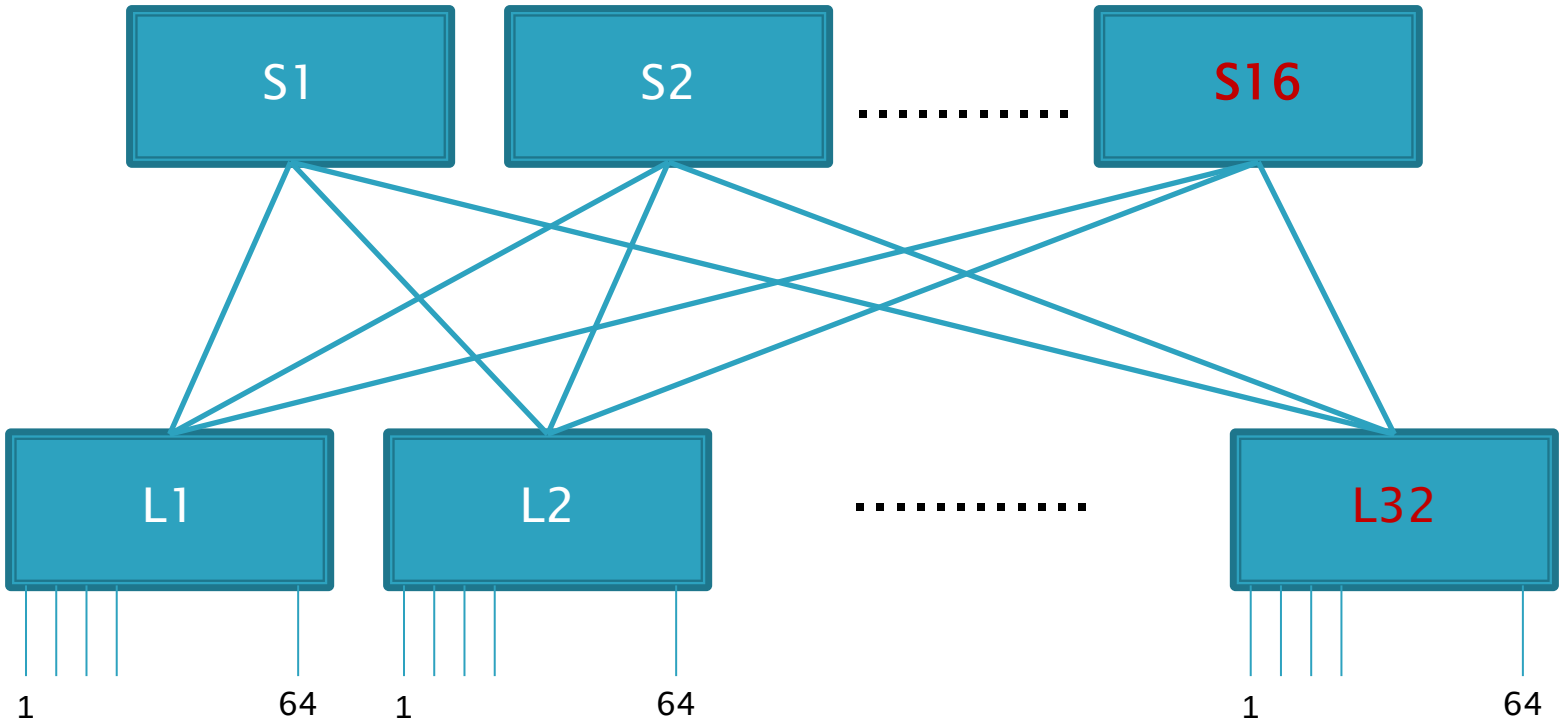
XENYA

Spine /Leaf – Primer 3

No Oversubscription

Stikala: Leaf: (LY6) 32x40Gb uporabljeni kot 16x40Gb + 64x10Gb

Spine: (LY6) 32x40Gb



Število priključkov (10Gb) N_C=32*64=2048

Oversubscription OSR= 64*10/16*40=**1**

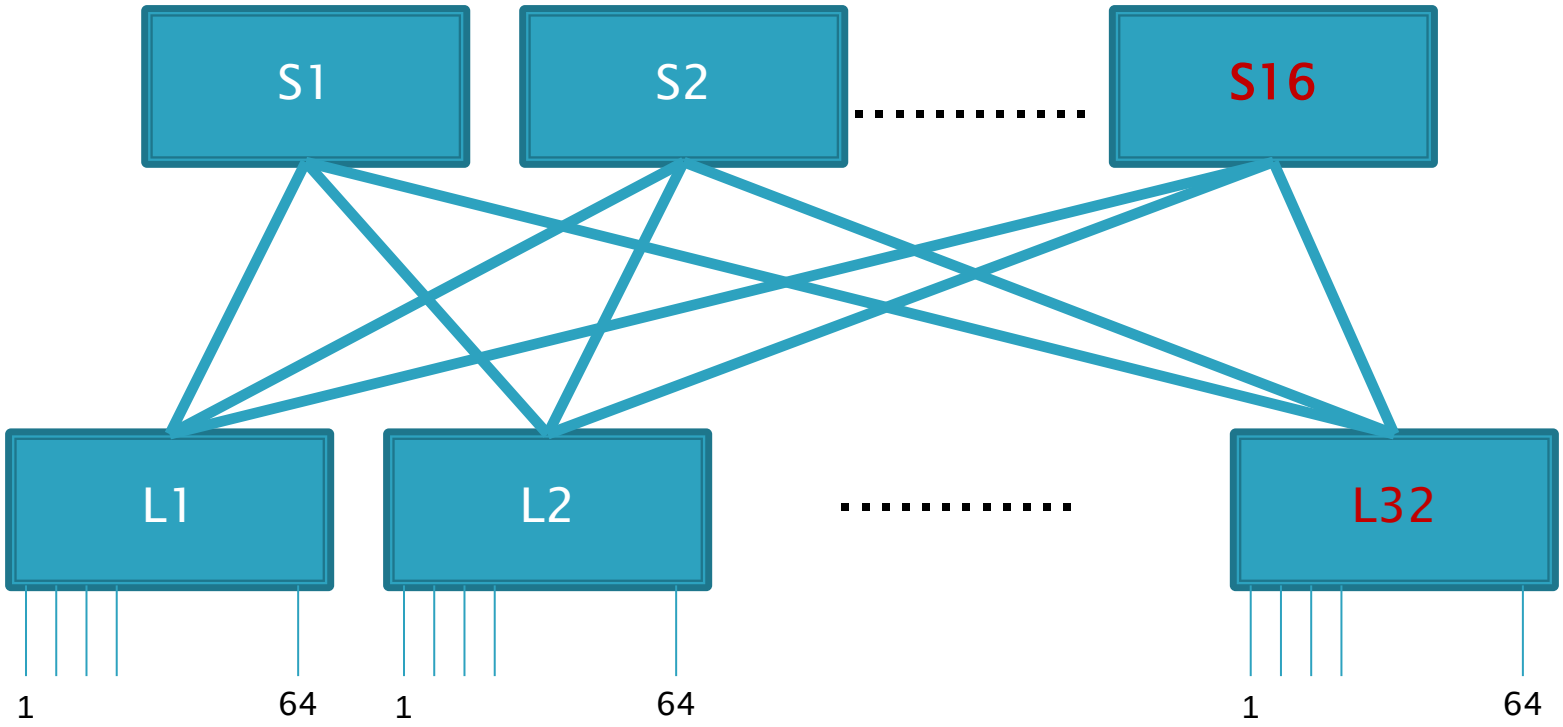
Pasovna širina S_T=640Gb

XENYA

Spine /Leaf – Primer 4

No Oversubscription

Stikala: Leaf: (IX1) 32x100Gb uporabljeni kot 16x100Gb + 64x25Gb
Spine: (IX1) 32x100Gb



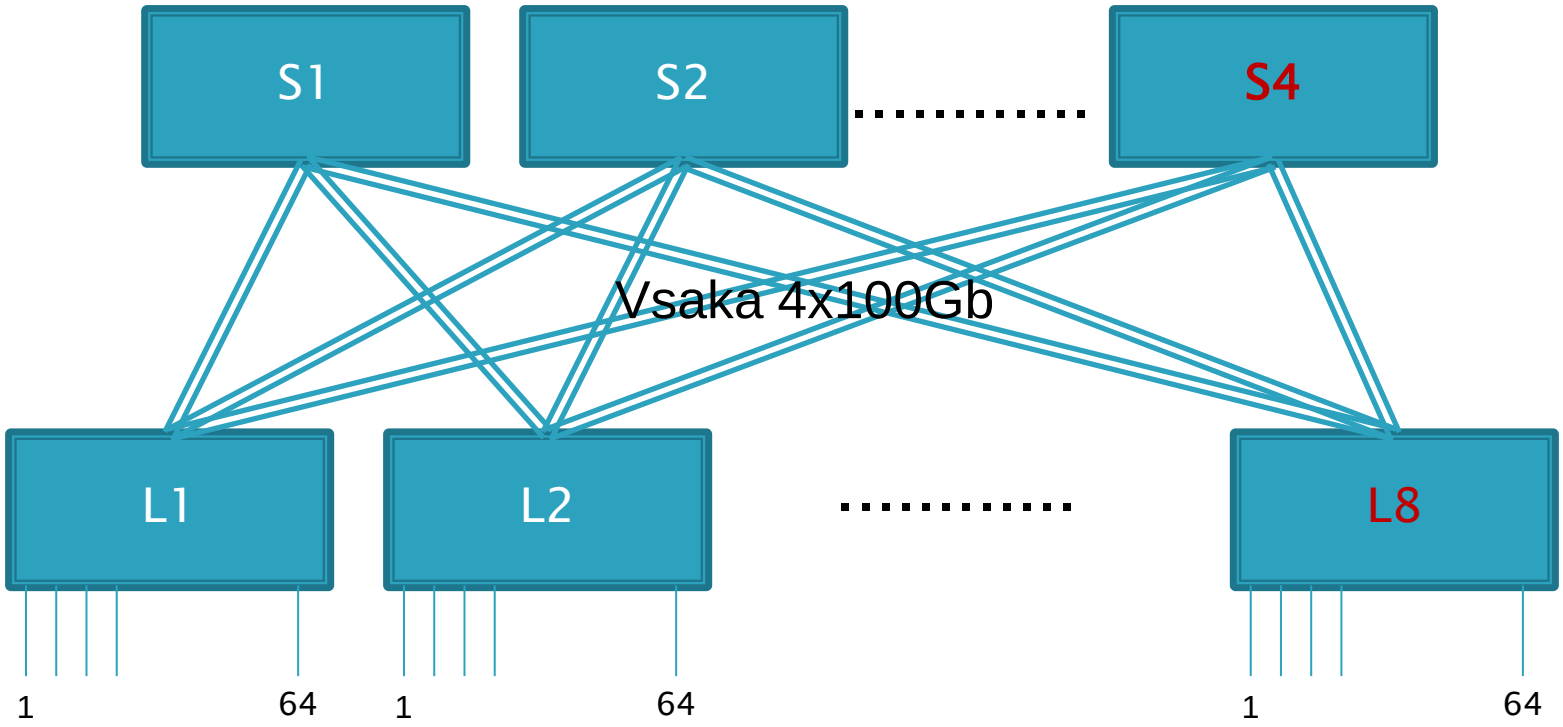
Število priključkov (25Gb)	32*(4*16)=2048
Oversubscription	OSR=64*25/16*100=**1**
Pasovna širina	S_T=1.6Tb
Zakasnitev	<1.2us

XENYA

Spine /Leaf – Primer 4a

No Oversubscription

Stikala: Leaf: (IX1) 32x100Gb uporabljeni kot 16x100Gb + 64x25Gb
Spine: (IX1) 32x100Gb



Število priključkov (25Gb)	8*(4*16)=512
Oversubscription	OSR=64*25/16*100=**1**
Pasovna širina	S_T=1.6Tb
Zakasnitev	<1.2us

XENYA

Redundanca omrežja

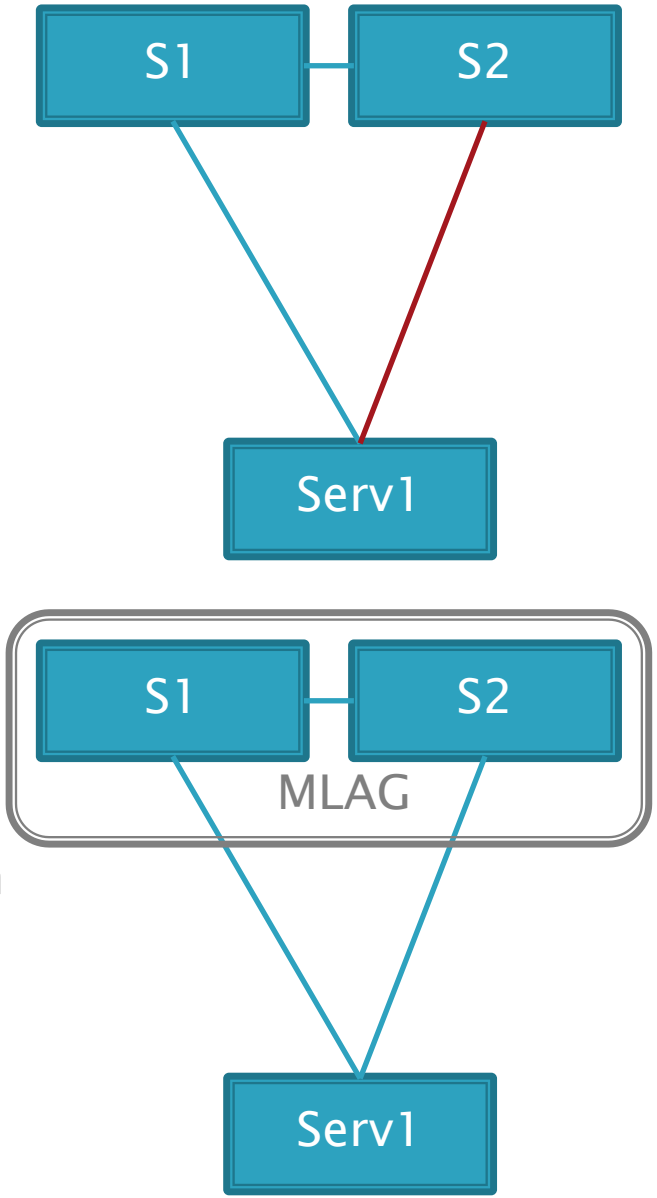
- Fizični nivo
  - Redundanca povezav
  - Redundanca Stikal
  - Redundanca napajanja

- Komunikacijski nivo
  - ECMP, dinamično usmerjanje OSPF, ISIS, BGP
    - Tipično do 64 enakovrednih smeri
  - LAG, MLAG

XENYA

Redundanca na L2 nivoju

- Namesto STP Spanning Tree Protocol-a (STP 802.1D, RSTP 802.1w, MSTP 802.1s)
  - Kompleksna konfiguracija odvisna od HW.
  - Počasna konvergenca (3 do 90s).
  - V močno redundantnih (n x n) topologijah polovica povezav ne prenaša podatkov.
  - Vsaka sprememba topologije načeloma povzroči blokado prenosa v celotni L2 domeni.
- MLAG – Multiple Chasis LAG (Link AGregation)
  - Enako kot navadni LAG prenaša podatke **prek vseh povezav**.
  - Hitra konvergenca, še posebno ob podpori LACP
    - **Standardni protokol** na strani klijenta (strežnika)



XENYA

Kreiranje MLAG

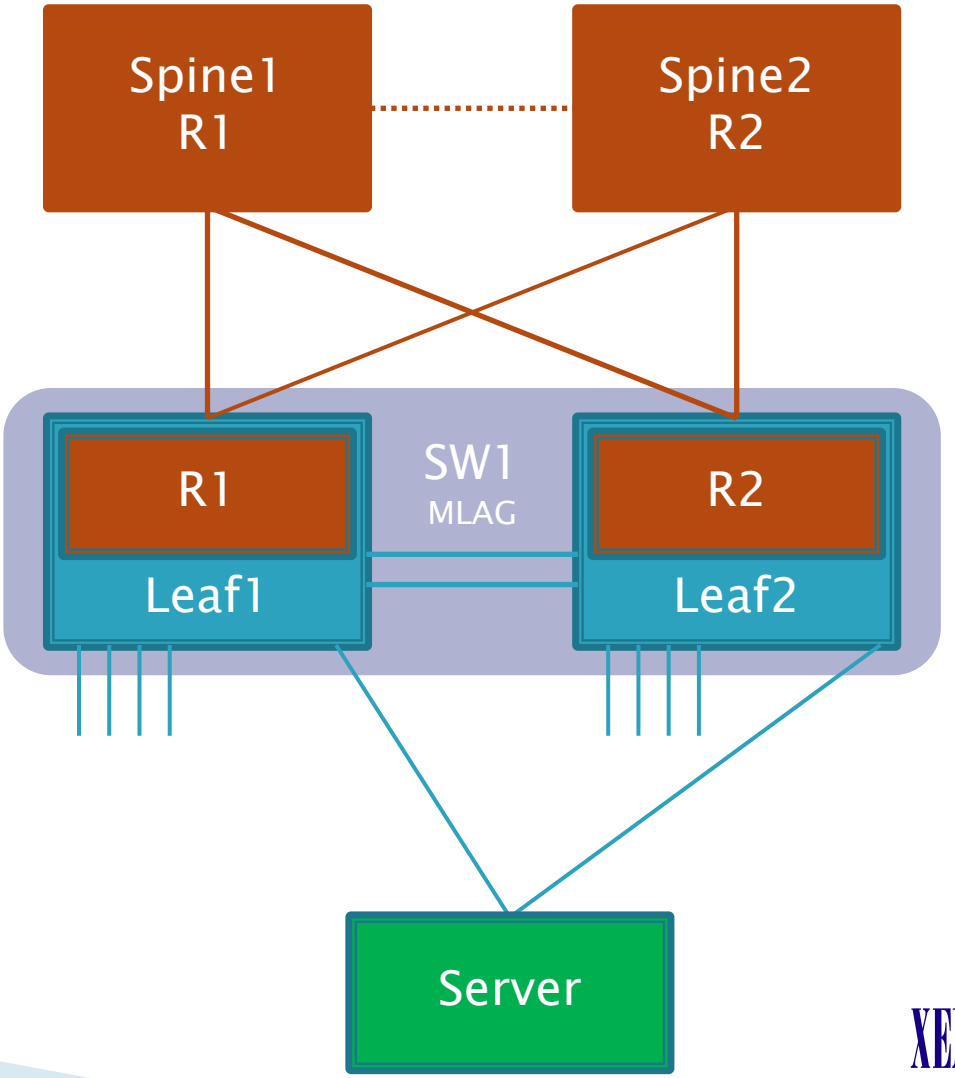
- Kreiranje agregiranih povezav (port-channel)
- Dodajanje fizičnih priključkov v agregirane povezave
- Kreiranje MLAG
  - Nastavitev MLAG domene, mora imeti vrednost različno od ostalih
  - Izbor agregirane povezave, ki povezuje oba stikala v MLAG domeno
  - Opcijsko določitev rezervne poti za nadzorni promet v MLAG domeni
- Določitev MLAG ID za vsako agregirano povezavo. Ta mora biti enak na povezavah v isti agregirani skupini na obeh napravah v MLAG domeni.

XENYA

MLAG in usmerjanje

- Kreiramo MLAG
  Dobimo
  - Eno logično stikalo in
  - Dva usmerjevalnika

  - VRRP
  - mlag peer-gateway



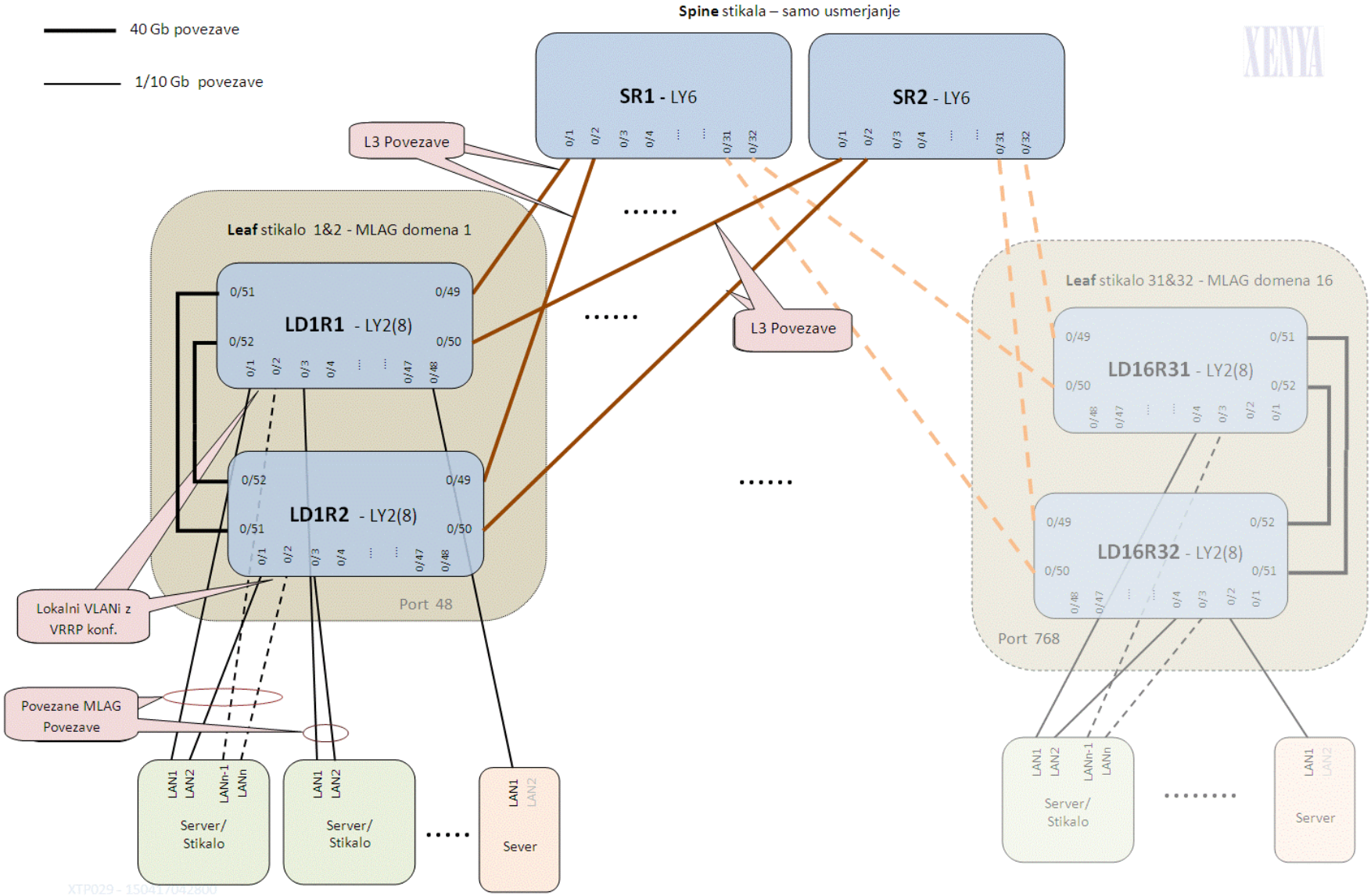
XENYA

Arhitektura 1

- “Fabric” topologija samo za L3 promet
- Redundanca in razporejanje prometa na hrbteničnem nivoju zagotovljene z ECMP usmerjanjem
- Na vsalem Leaf stikalu ločeno podomrežje za klijente
- Na L2 strani (proti uporabnikom) redundanco izvedemo z MLAG vezavo
- Kombinacija MLAG in usmerjanja zahteva uporabo VRRP na vmesnikih proti klijentom na Leaf stikalih
- Se lahko razširi do zelo velikega števila priključkov klijentov

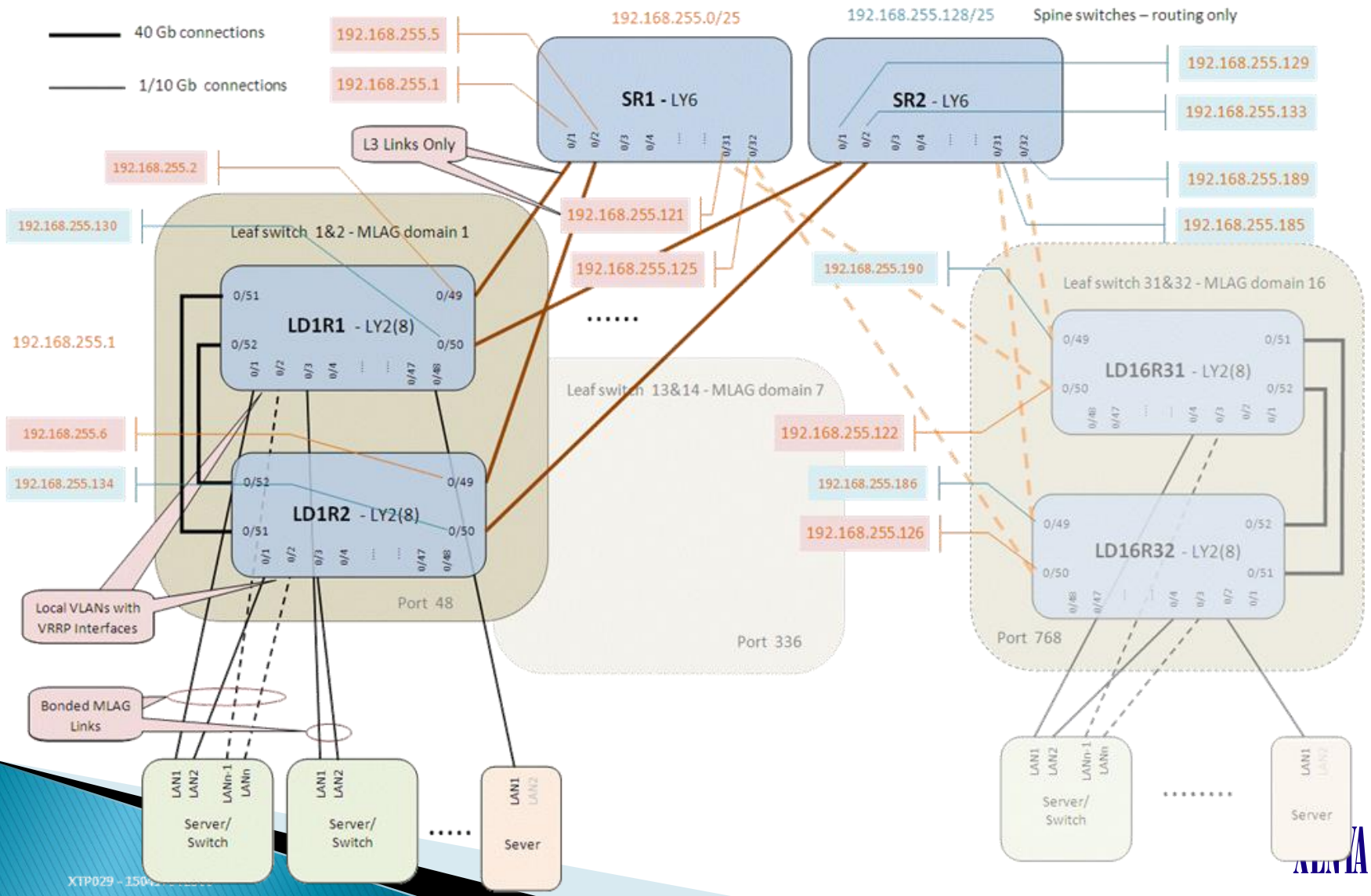
XENYA

Spine/Leaf Arhitektura A1 Praktična izvedba

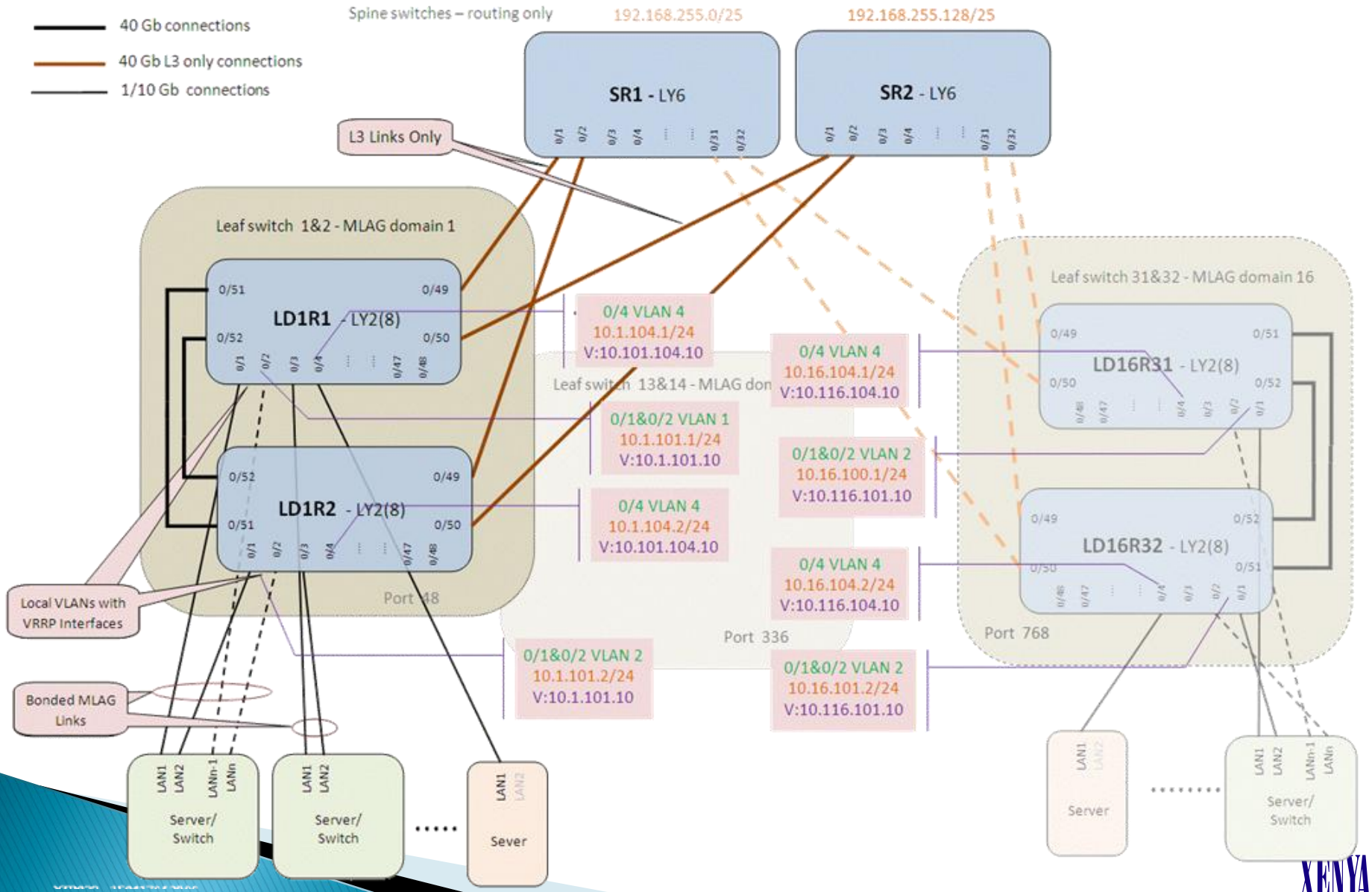


Arhitektura A1 – Primer klasične Spine/Leaf arhitekture z L3 ECMP usmerjanjem in MLAG redundanco na L2 povezavah do strežnikov

Spine/Leaf Arhitektura A1 Naslovi Uplink Povezav



Spine/Leaf Arhitektura A1 klijentni VLANi in naslovi



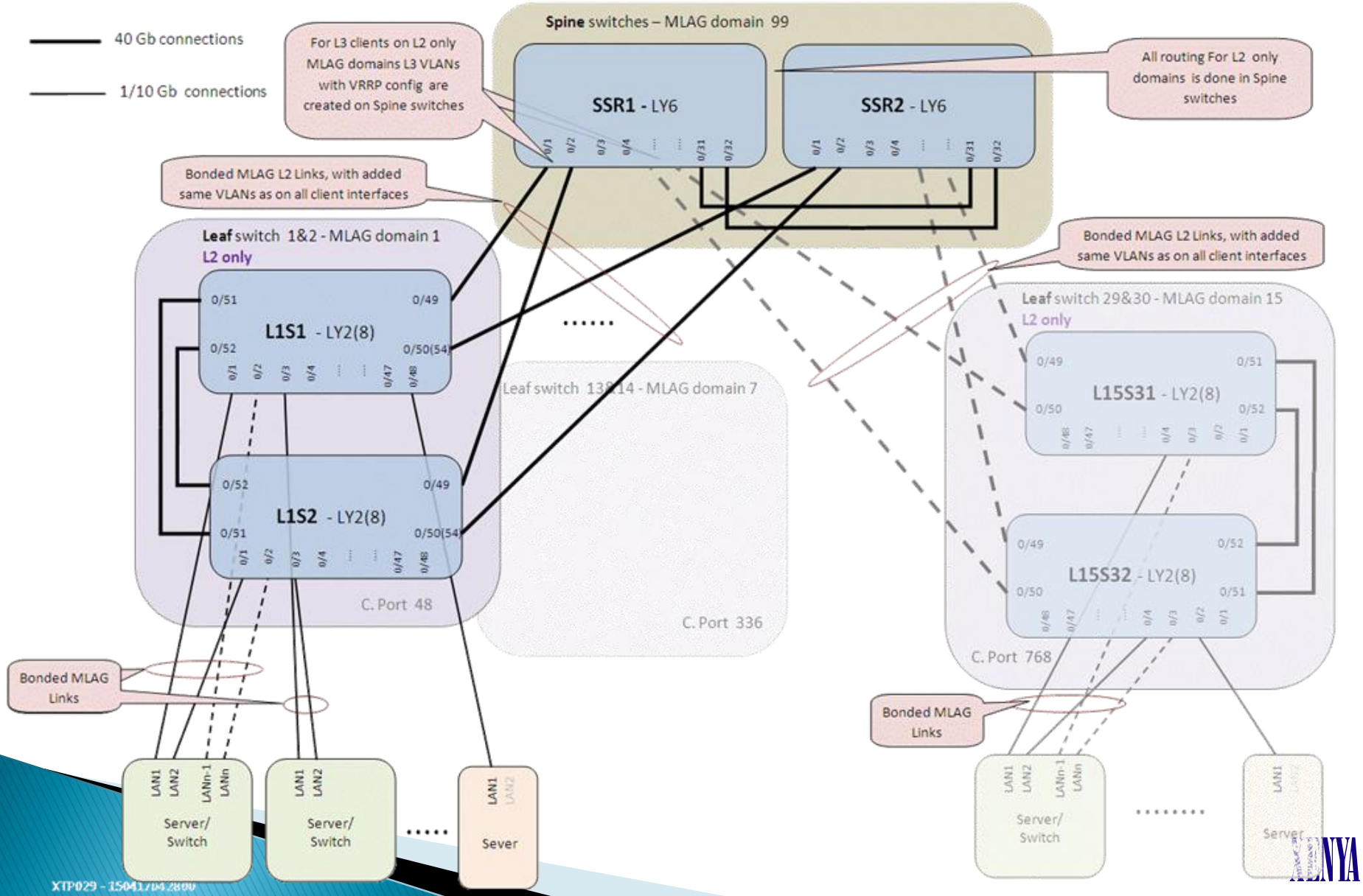
XENYA

Arhitektura 2

- Hrbtenični nivo podpira tudi L2 promet
- L2 in L3 promet
- Usmerjanje samo v Spine Mlag-u
  - Vsi vmesniki klijentov definirani na Spine stikalih
  - Leaf stikala samo premoščajo
- Velikost omejena z eno Spine MLAG domeno
- Nujni ukrepi za omejevanje posledic zank na L2 nivoju
  - Port guard, Storm control, MAC moving ...., opcijsko DHCP relaying, DAI, IP Souce Guard...

XENYA

Spine/Leaf Arhitektura A2 L2/L3 izvedba



Nadgradnja omrežij za podatovne centre

- **Datacenter Bridgeing Protokoli** – Omogočajo implementacijo enovitega omrežja, kjer se lahko prenaša tudi skladiščne podatke brez izgub paketov

- Protokoli za dinamično segmentacijo na stikalih (kreiranje VLANov)
  - Razširitve L2 arhitekture (zgrajene tudi s Trident+ SoC)
    - **VMTracer** – sprejema informacije o aktivnih VLANih direktno iz virtuelnih stikal. Enostavna konfiguracija, podprta na quanta & arista stikalih.
  - Razširitve standardne L3 arhitekture zgrajene s sikali na osnovi Trident II SoC
    - Overlaying protikoli
      - **VXLAN**, **NVGRE, ...**

XENYA

Dodatni protokoli

- Z DCB protokoli lahko zagotavljamo ločeno obravnavo različnih kategorij prometa (ethernet, storage):
  - Priority-based Flow Control (PFC): IEEE 802.1Qbb
  - Enhanced Transmission Selection (ETS) v IEEE 802.1Qaz
  - Congestion Notification (CN) 802.1Qau
  - Data Center Bridging eXchange (DCBX) v IEEE 802.1Qaz
- Fibre Channel over Ethernet(FCoE): T11 FCoE
- Overlay protokoli:
  - **VXLAN – L2 tuneliranje prek L3 z GRE protokolom**
  - NVGRE – alternativna izvedba L2 tuneliranja prek L3

XENYA

Alternativne izvedbe

- **Cumulus Linux -** Linux distribucija za stikala na osnovi Debian Linuxa instalirana na stikalo (bare metal, white box):
  - Priključki stikala predstavljeni kot mrežni vmesniki: swp1,....., swpn
  - Krmiljenje prek standardnih ukazov za krmiljenje omrežja na linuxu:
    - etc/network/interfaces datoteka: Iface,ifup, Ifdown; brctl...
  - Orkestracija integrirana v stikalo – upravlajnje mreže enako kot upravljanje strežnikov
    - Pupett, Chef, Ansible ...
  - VMWare podpora – NSX krmilnik direktno krmili tudi fizična stikala enako kot virtuelna
    - Registrira L2 storitev fizičnih stikal v VMWare NSX Controller-ju (prek NSX Service Node komponente instalirane na stikalu)
    - Tudi overlaying protokol VXLAN in konverzija v VLANe na stikalu (VTEP) je tako krmiljena direktno prek hypervisorja.
- FW tretjih ponudnikov – dobavljiv kot samostojen produkt
  - Juniper JUNOS,.....
- Zunanji SDN krmilnik
  - Stikala tudi s std, FW podpirajo OpenFlow 1.0 do 1.3, OpenAPI ....
- Open source FW rešitve

XENYA

Zaključek

- Standardne komponente & Nove arhitekture & Novi protokoli
  - Modularne, Redundantne, Skalabilne rešiteve
  - Z nizko porabo energije
  - Razpršene – z možnostjo rasti omrežja po rasti potreb
- Cenovno ugodne rešitve
- Rešitve z uporabo overlay protokolov še bolj fleksibilne od prikazanih.
- Podpirajo bodoče širitve v dinamična omrežja, SDN ...

XENYA