# komro

Mehr Freiraum. Mehr Leben.

komro
Mehr Freiraum. Mehr Leben.

# Von Legacy zu Leaf-Spine: Unsere EVPN/VxLAN-Migration

# Unternehmen

komro
Mehr Freiraum. Mehr Leben.

- Über 26.000 Privat- und Geschäftskunden
- Regionaler Telekommunikations-Dienstdienstleister
  für Internet & Telefonie, TV, Standardfestverbindungen sowie IoT Lösungen
- 100% Tochterunternehmen der Stadtwerke Rosenheim
- Aktuell 55 Mitarbeiter
- Highspeed-Glasfaser-Kabelnetz flächendeckend in Rosenheim

komro
Mehr Freiraum. Mehr Leben.

# Datacenter Dienste

Was für Dienste hostet ein ISP?

1

# ISP Dienste

komro
Mehr Freiraum. Mehr Leben.

- ✓DNS Resolver
- ✓DHCP Server
- ✓CGNAT Maschinen
- ✓CDN Caches
- ✓VoIP
- ✓Speedtest™ Server
- ✓Flowanalysetools
- ✓DDoS Detection
- ✓WiFi Controller / Router
- ✓Storage
- ✓Monitoring/Alerting
- ✓Linux Package Mirror
- ✓CCTV
- ✓Ticketsystem
- ✓VPN Gateway
- ✓Interne Dienste

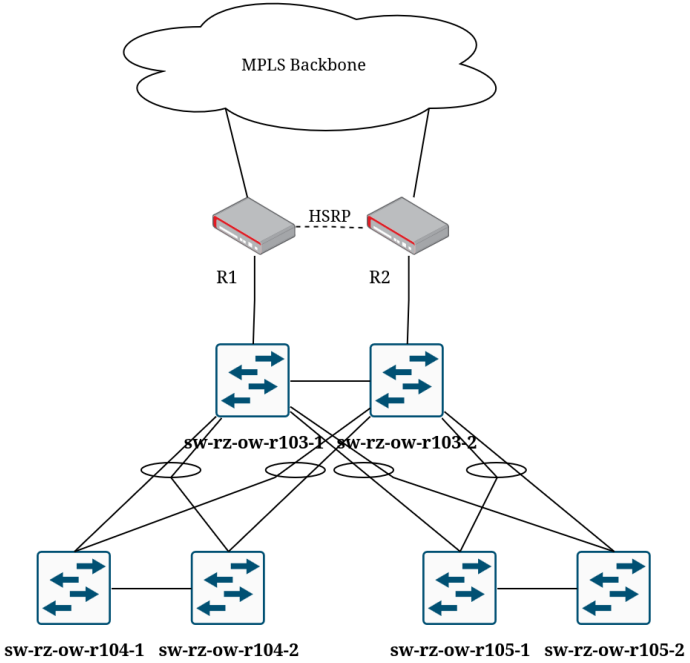
komro
Mehr Freiraum. Mehr Leben.

# Ausgangslage

2

komro
Mehr Freiraum. Mehr Leben.

# Ausgangslage




- ✓ 2 Georedundante Datacenter mit Cisco n3k Switches, vPC Topologie
- ✓ 2 Cisco asr920 als IP/MPLS Router mit HSRP pro Standort

# Pain Points

- ✓ Hardware EoS/EoL
- ✓ HSRP IPv6 nicht unterstützt
- ✓ Große L2 Domains - MAC Learning
- ✓ MAC Aging bug in Nexus (CSCul82233)
- ✓ Max 10gbit/s Output pro Datacenter
- ✓ Max 10gbit/s LAG Members

komro
Mehr Freiraum. Mehr Leben.

# EVPN/VxLAN

3

komro
Mehr Freiraum. Mehr Leben.

# EVPN/VxLAN

## VxLAN

Overlay-Technologie zur Erweiterung von Layer-2-Netzen über Layer-3 (IP)

Nutzt MAC-in-UDP-Encapsulation (VXLAN-Header + UDP/IP-Transport)

Segmentierung via 24-Bit VNI (Virtual Network Identifier) für bis zu 16 Mio. Logische Netze

## EVPN

Control-Plane-Protokoll (BGP-basiert) für automatisiertes MAC-/IP-Adress-Learning (vs. klassisches Flood and Learn)

ARP und ND Cache bieten die Möglichkeit, Broad- und Multicastlast zu verringern

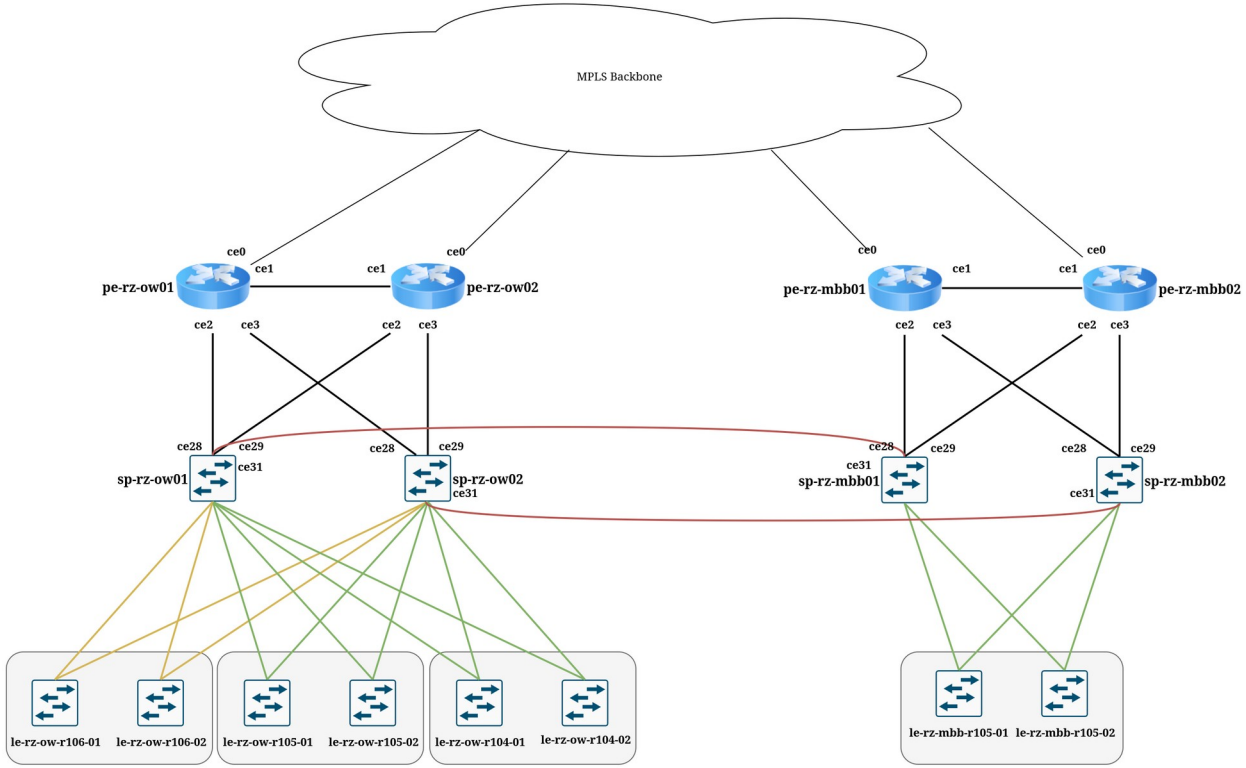
Kombiniert VXLANs dynamische Skalierbarkeit mit EVPNs Steuerungsintelligenz

Verschiedene Route Typen [1]

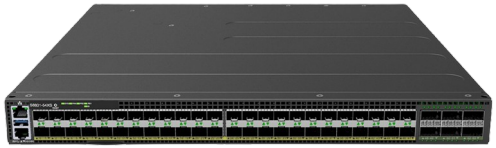
[1]: https://www.bgphelp.com/2017/05/22/evpn-route-types/

komro
Mehr Freiraum. Mehr Leben.

# Zielarchitektur



# Hard-/Software



## Leaf
UfiSpace s8901

- ✓ Throughput: 2Tbps
- ✓ Uplink: 6x 40/100G
- ✓ Downlink: 48x 1/10/25G
- ✓ ASIC: Trident3-X5 BCM56770

## Spine
UfiSpace s9110

- ✓ Throughput: 3.2Tbps
- ✓ Up-/Downlink: 32x 40/100G
- ✓ ASIC: Trident3-X7 BCM56870

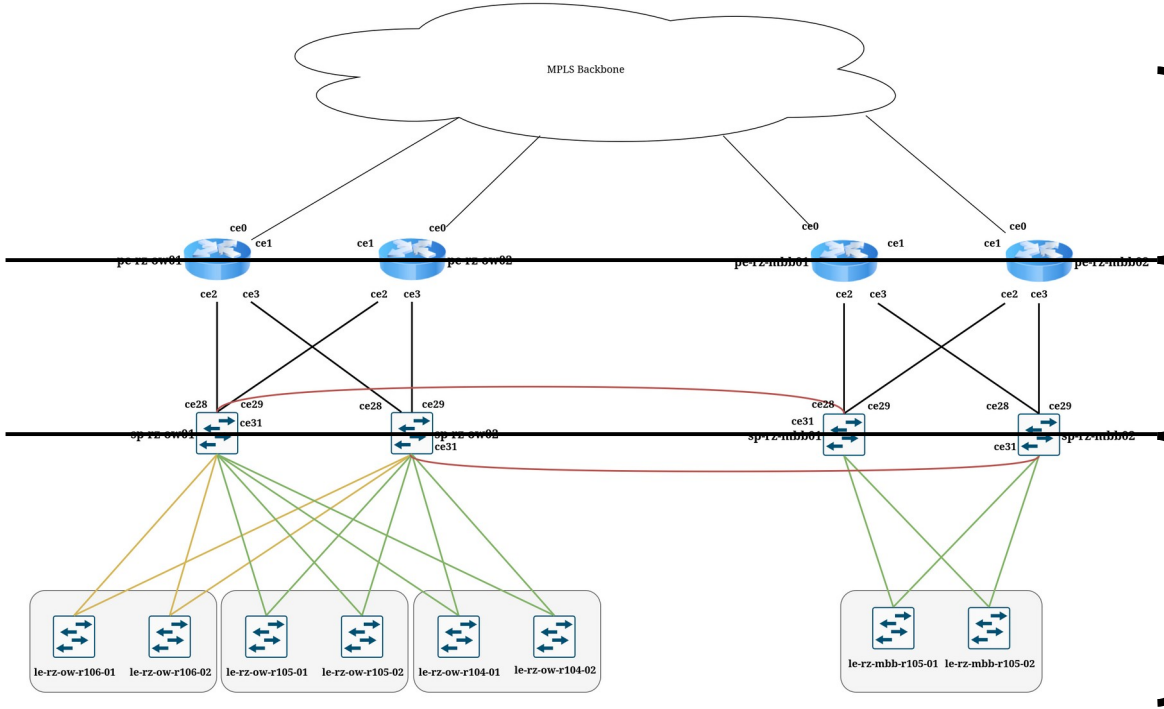
## Border Leaf
UfiSpace s9501

- ✓ Throughput: 800gbps
- ✓ Uplink: 2x 100/400G + 2x 40/100G
- ✓ Downlink: 24x 1/10/25G
- ✓ ASIC: Qumran2a BCM88483

ipinfusion™ OcNOS

komro
Mehr Freiraum. Mehr Leben.

# Netzwerkstack




MPLS

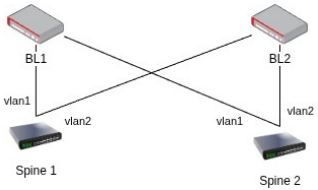
BGP/VLAN

EVPN/VxLAN + OSPF

komro
Mehr Freiraum. Mehr Leben.

# Services

- ✓ Alle Dienste sind VRFs. Die Routen werden auf den Spines von den Border Leaves über VLAN interfaces + BGP gelernt.




⚡ Jeder Service bedarf 4 Transferlinks/vlans, inkl. /31 bzw. /127 IP Adressen, BGP + BFD Sessions

Die Konfiguration davon übernimmt Ansible, die Dokumentation Netbox

```
- name: SERVER-MGMT #vrf in evpn
  vnid: 128 #1-2046, unique
  irb_ips:
    - 10.2.1.1/24
    - 10.2.14.1/26
  irb_ips6:
    - 2A02:2488:2:6::1/64 #anycast gateway ipv6
    - FDE3:7FF9:7A68::1/64
  transfernet: 10.20.4.72/29 #ipv4 transfer between spine-borderleaf
  transfernet6: fdd4:cfb1:138d::10/125 #ipv6 transfer between spine-borderleaf
  dhcp_server: 10.2.1.2
  export_vrf: management
  irb_shut: false
```

Die Transfernets definieren die Netze für die BGP Sessions spine-borderleaf

export_vrf definiert das VRF aus dem Backbone, in dem der DC Service läuft.

komro
Mehr Freiraum. Mehr Leben.

# Services

✓ Können auch reine L2 Domains sein:

💡In den L2 VxLANs werden die VLAN Tags mit transportiert.

**inner-vid-disabled** in der VxLAN definition würde das unterlassen.

```
- name: GPON-VPLS
  vnid: 4103
```

CLI wie “rewrite ingress tag pop 1 symmetric”

💡💡 Ein VxLAN kann auch mehrere VLANS transportieren!

komro
Mehr Freiraum. Mehr Leben.

# Service Access

Für den Service benötigen wir ein Access Interface.
Das kann ein Ingressport (untagged) oder eine Port+VLAN Kombination sein.

## Tagged:

```
#Vlan 129 von xe44 in vxlan 129 mappen:

nvo vxlan access-if port-vlan xe44 129
 map vnid 129
```

## Untagged:

```
#Port xe45 in vxlan 129 mappen:

nvo vxlan access-if port xe45
 map vnid 129
```

```
leaf-rz-ow-r106-01#sh nvo vxlan access-if brief

                          Inner                          Admin   Link
Interface  Vlan           vlan  Ifindex    Vnid          status  status
-------------------------------------------------------------------------
xe44       3              ---   0x7a13d    3             up      up
xe44       4              ---   0x7a126    5000          up      up
xe44       11             ---   0x7a13e    4111          up      up
xe44       12             ---   0x7a13f    4112          up      up
xe44       14             ---   0x7a135    14            up      up
xe44       16             ---   0x7a12c    5515          up      up
xe44       20             ---   0x7a140    20            up      up
xe44       21             ---   0x7a12f    21            up      up
xe44       22             ---   0x7a12e    22            up      up
xe44       23             ---   0x7a133    23            up      up
xe44       24             ---   0x7a13a    24            up      up
xe44       30             ---   0x7a136    4333          up      up
xe44       31             ---   0x7a12d    5531          up      up
xe44       103            ---   0x7a13c    4103          up      up
xe44       128            ---   0x7a130    128           up      up
xe44       129            ---   0x7a138    129           up      up
xe44       130            ---   0x7a124    130           up      up
xe44       131            ---   0x7a125    131           up      up
xe44       150            ---   0x7a12b    5556          up      up
xe44       199            ---   0x7a123    199           up      up
xe44       255            ---   0x7a131    255           up      up
xe44       256            ---   0x7a132    256           up      up
xe44       311            ---   0x7a134    311           up      up
```

komro
Mehr Freiraum. Mehr Leben.

# BGP Services

✓ Einige Services brauchen BGP

Dafür werden auf den Leaves **eindeutige** IPs konfiguriert, die vom Anycast Gateway abweichen.

Für die BGP Session wird dann die eindeutige IP verwendet, per route-map die Next-Hop IP auf die Anycast IP gesetzt.

⚠️ Man kann beliebig viele (eindeutige) secondary IPv4 Adressen aus einem Subnet auf einem Interface konfigurieren. Für IPv6 kann man aber nur eine IP Pro Subnet konfigurieren (OLS-13591), was dazu führt das alle BGP6 Sessions Multihop sind und am Client (mindestens) eine statische IPv6 Route brauchen.

komro
Mehr Freiraum. Mehr Leben.

# Port Sniffer

👍 OcNOS DC kann erspan. Eine Kopie des Traffics aus einem Port wird mit GRE Headern versehen und an einen Collector geschickt.
Der Header kann am Collector mit xdp entfernt werden, um das Filtern mit libpcap zu ermöglichen [1]

⚠️ Traffic von der CPU wird nicht gemirrort, debugging von Protokollen wie ARP/ND schwierig. Feature Request FAE-1511.

[1]: https://github.com/komronaut/xdp-erspan-stripper

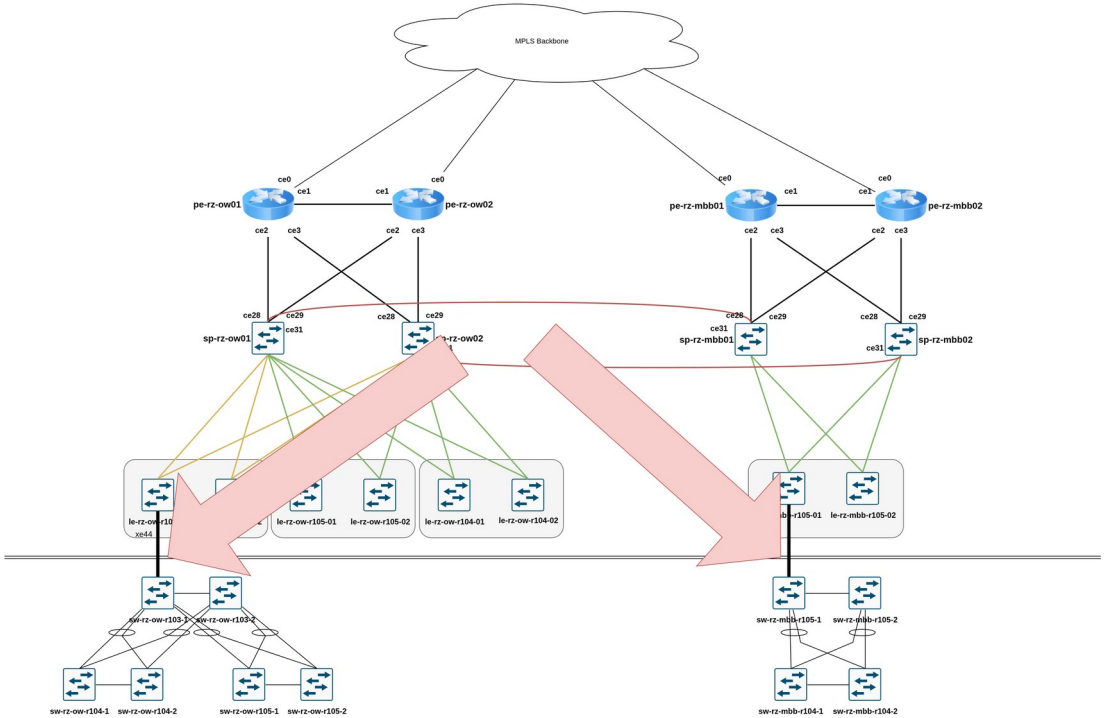
# Migration

4

komro
Mehr Freiraum. Mehr Leben.

# Migrationspfad

Crossconnect zwischen “Alter” und “Neuer” Welt:



komro
Mehr Freiraum. Mehr Leben.

# Migrationspfad

- ✓ L2 Services konfigurieren, Access Ports mappen
- ✓ L3 Services konfigurieren, irb Interface auf shutdown, Access Ports mappen. L3 Interface am legacy Router Shutten, und irb No-Shutten. Nicht erreichbarkeit der Dienste <5s
- ✓ Hosts, dessen Services alle auf der “neuen” Welt existieren, können im Wartungsfenster Umgesteckt  werden.

komro
Mehr Freiraum. Mehr Leben.

# Lessons Learned 

- ✓ Komplexe repetitive Konfiguration, Automatisierung hilft Fehler zu vermeiden (Ansible)
- ✓ Migrationen sind Zeitaufwändig
- ✓ Nicht immer ist die CP mit der DP sync. Bei der Fehlersuche, den Hardwarelayer beachten
- ✓ Dekonfiguration mittels Ansible schwer. Atomic commits wären hilfreich
- ✓ Gute Erfahrung mit IPI TAC, Antwortrate gut, bugfixes werden zeitnah implementiert.

  **Danke dafür!!!** 🙇

# QNA

komro
Mehr Freiraum. Mehr Leben.

👍 Vielen Dank für eure Aufmerksamkeit!