

Interop Tokyo 2023 ShowNet の VPNベースバックボーン

ShowNet NOCチームメンバー / NTTコミュニケーションズ
上野 幸杜

Open Networking Conference 2023

自己紹介

- 上野 幸杜（うえの ゆきと）
- ShowNet NOCチームメンバー
 - 2014-2023 except 2022
 - 主にレイヤ2/レイヤ3技術を担当
- 所属：NTTコミュニケーションズ
 - 2016- R&D部門（イノベーションセンター）でレイヤ2/レイヤ3に関する技術・製品検証等を担当
 - 最近は伝送装置に入門中

Interop TokyoとShowNet

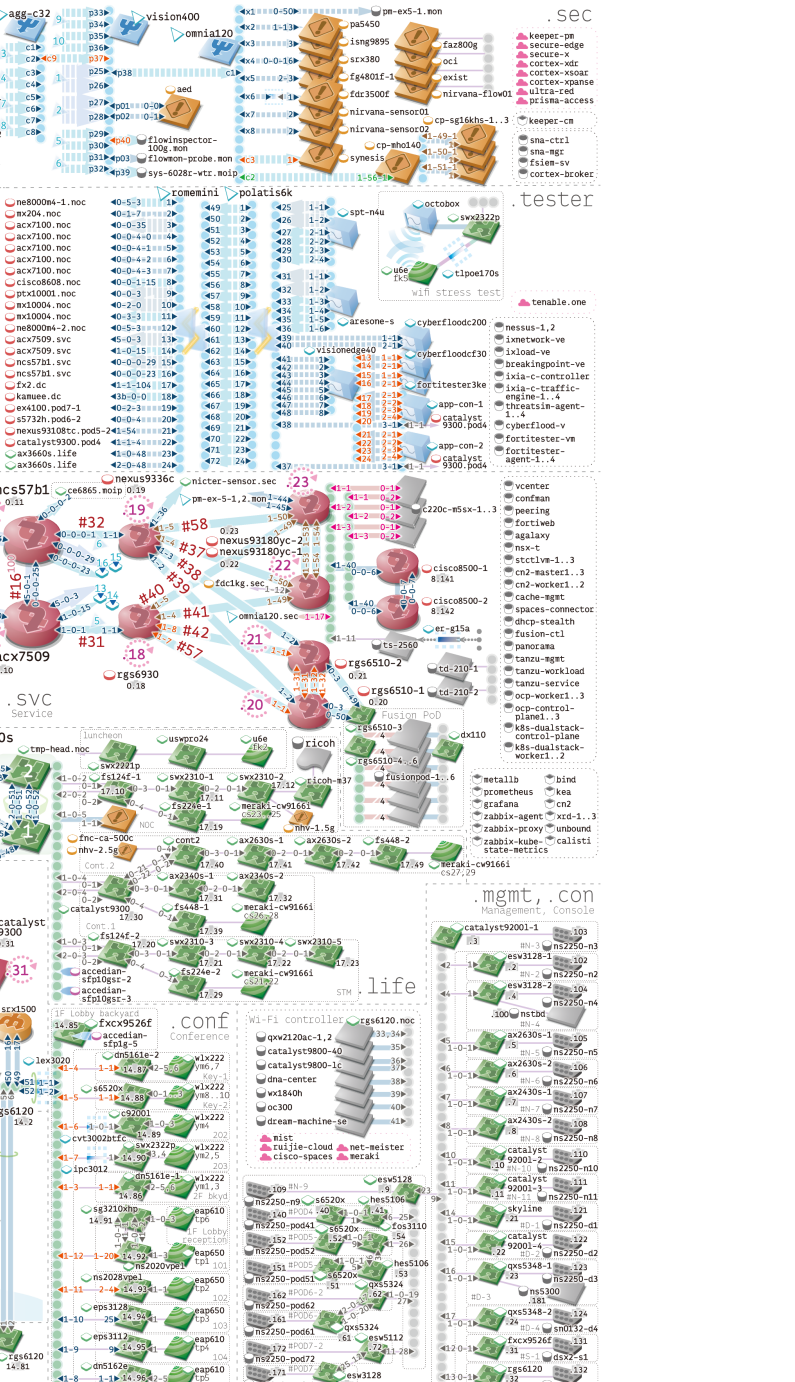
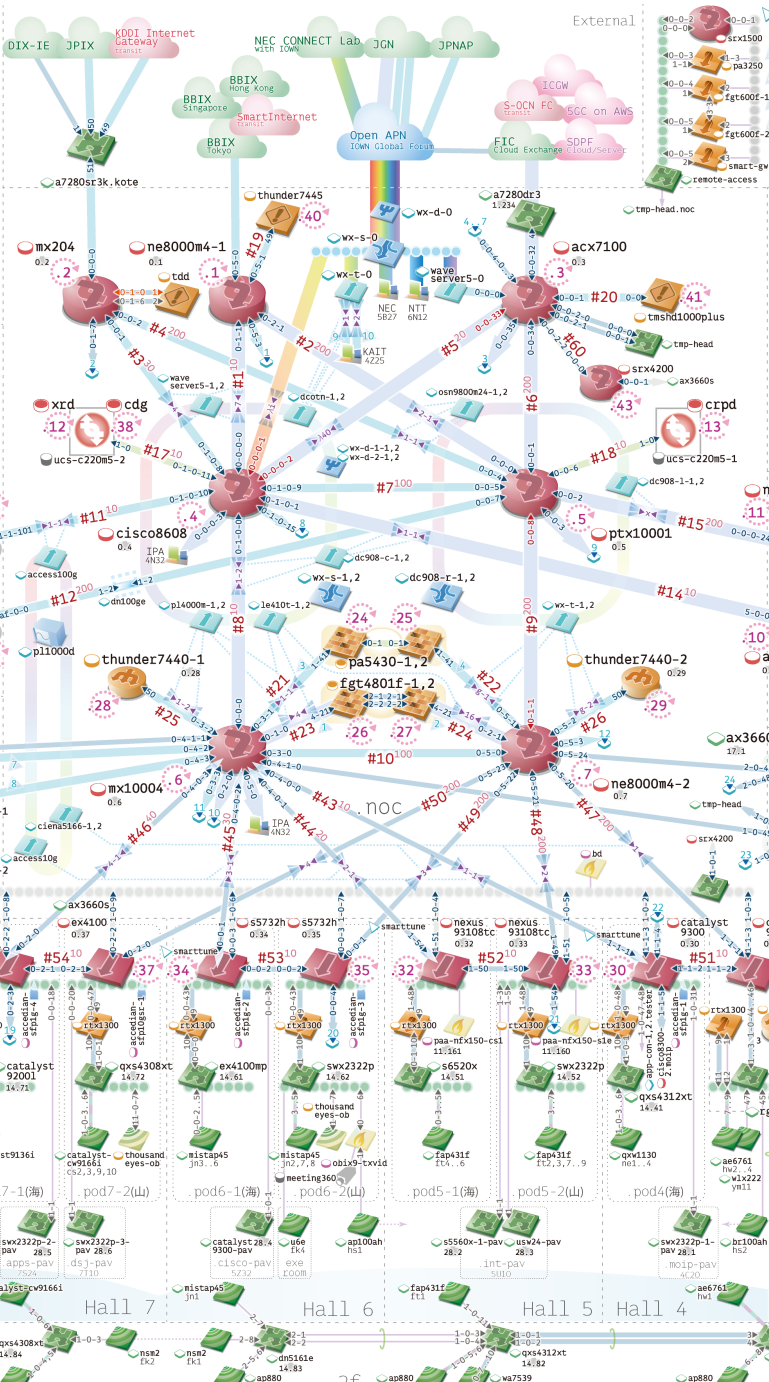
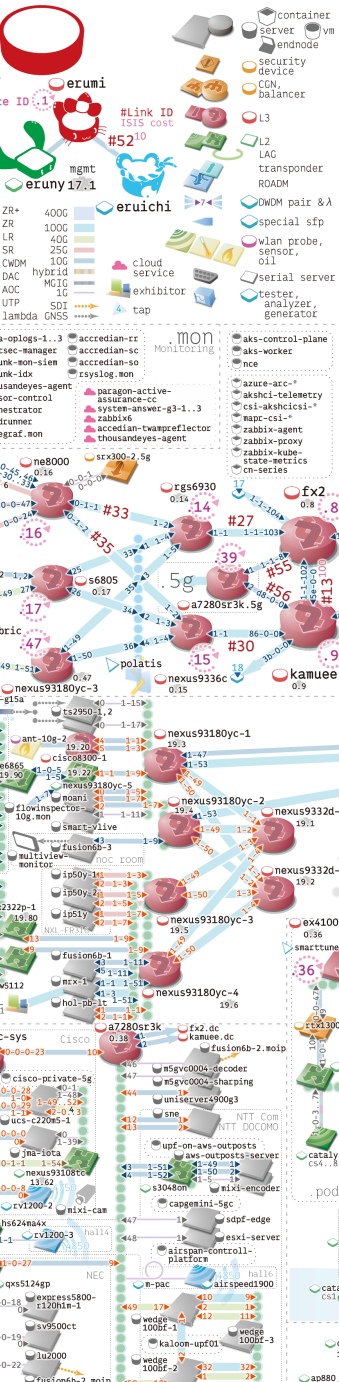
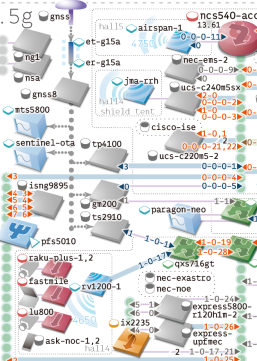
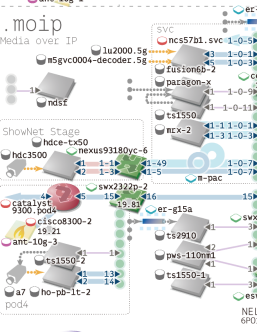
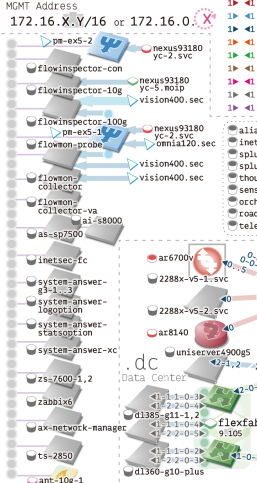
- Interop Tokyo: 世界最大のネットワーク機器と技術の展示会
 - 1986年米国モントレで開催されたカンファレンスイベント「TCP/IP Vendors Workshop」が始まり
 - Interop Tokyoは今年で30回目の開催
 - 来場者約15万人
- ShowNet: Interopで構築される世界最大のデモンストレーションネットワーク
 - 2年後、3年後に業界に浸透する技術に先駆けて挑戦
 - 様々な技術の相互接続性検証の場
 - 最新技術を実装しながら安定したサービスを出展ブース・来場者に提供



ネットワークの規模

- コントリビューション機器/製品/サービス台数: 約1600台
- 動員数: 675名
 - NOCチームメンバー: 28名
 - STM: 37名
 - コントリビュータ: 610名
- UTP総延長: 約20.0km 光ファイバー総延長: 約7.2km
- NOCラック及びPod総電気容量: 約128.0kW
 - 内訳(NOC 約117kW、Pod 約11.0kW)
- NOCラック及びPod総コンセント数(100V, 200V含む): 約310個
 - 内訳(NOC 約230個、Pod 約80個)

Device ID: **X**
Loopback: 45.0.VPN_ID.X/32
Address: 2001:3e8::VPN_ID.X/128
P2P Link ID:
VPN ID: 45.0.VPN_ID x 100 + Link_ID
v4: v4.0.(100 + VPN_ID).(Link_ID x 4)/30
v6: 2001:3e8::(VLAN_ID)::/64



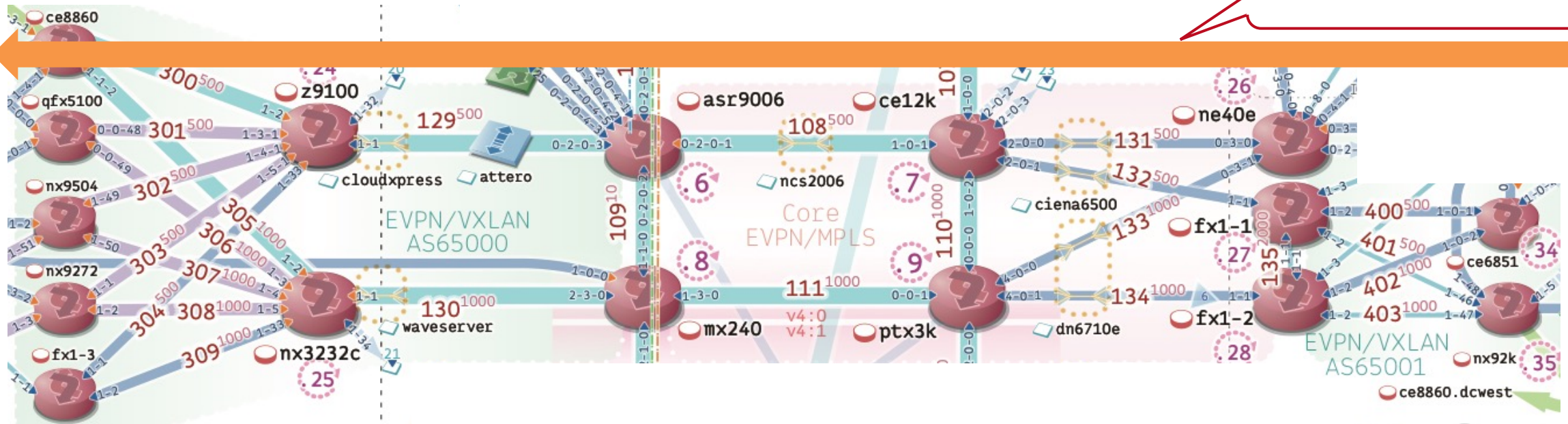


VPN化するShowNet

EVPN/MPLS - EVPN/VXLAN相互接続実験 in 2016

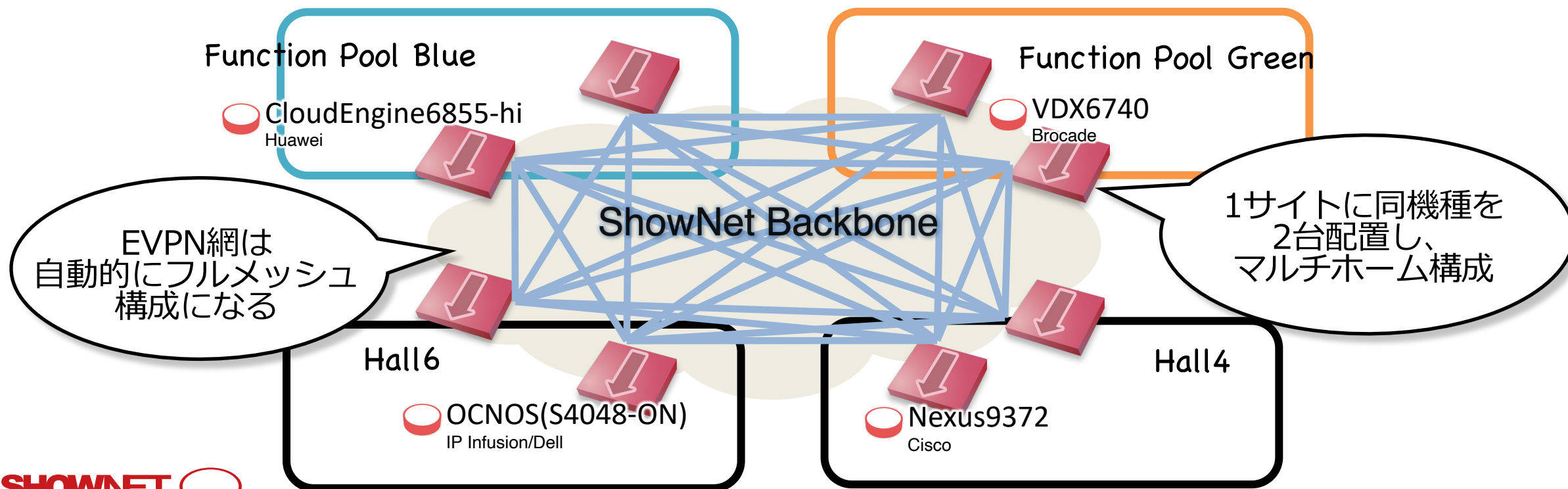
- MPLSバックボーン上でL2接続性を提供するEVPN/MPLSと東西DC内部のEVPN/VXLANを相互接続

L2セグメントの延伸



サービスチェイニング + EVPN/VXLAN in 2017

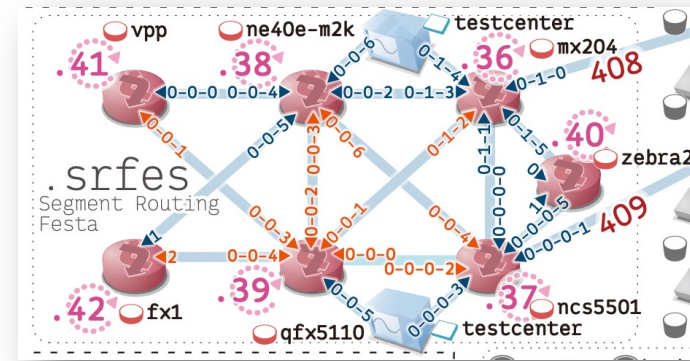
- 幕張メッセのホール・ファンクションプール単位で同一機種を使用
 - 異ベンダの場合、マルチホームの相互接続性に難があるため



Segment Routing Interrop in 2018

相互接続性検証を実施

- SR-MPLS: OSPFv2 Basic + TI-LFA, BGP Prefix SID, TE, SRTE&PCEP
- SRv6: ISIS, T.Encap, T.Insert



Step1 SR OSPFv2 Basic検証補足 SRGB(Segment Routing Global Block)

- SRGBが異なる環境でのSIDの広告
SRGB:16000-23999のルータとSRGB:800000-807999のルータの例

Step3 BGP Prefix SID

- RR向けにBGP LSのセッションを確立
 - 正しくLinkstate DBが広告できること
- BGP LU経由でBGP Prefix SIDを広告
 - BGP Prefix SIDがRIB/FIB/LFIBにインストール
 - ICMP及びTestCenterからのTrafficで疎通確認

SR-MPLS SR-MPLS出力(こんな感じに見えます)

SR-MPLS via OSPFv2 By Junos
SRTE LSP By Junos
TILFA(P node/Q node選出) By IOS-XR

Step2 SR OSPFv2 TILFA

- 通常のIP FRRを行う場合右図のようなコストではループが発生して失敗する
- 隣接するノードでは障害を検知していないため、最もコストの低いノードへ転送結果、ループしてしまふ
- 障害発生ノードでは2番目にコストの低い迂回路へ転送
- リングトポロジではよくある例
- これを避けるためにはループしないノードまでトンネルする必要がある

検証 Step6 SR

| STEP | 内容 |
|------|-------------------------|
| 6-1 | SRv6 Control plane ISIS |
| 6-2 | SRv6 T.Encap |
| 6-3 | SRv6 T.Insert |
| 6-4 | SRv6 End.DX6(VPNv6) |

- SRv6 SIDをISISでSIDを広報できること
- T.Encap方式で疎通ができること
- T.Insert方式で疎通ができること
- VPNv6経路をSIDと紐づけて処理ができること

SRv6 Service Chaining in 2019

- SRv6によるService Chaining

- SRv6 Service Programming

- このSRv6でユーザトラフィックを任意のネットワークサービスへ転送するデモンストレーションを実施

ShowNet 2019におけるSRv6サービスチェイニング

- 出展社・来場者のトラフィックは基本的にサービスチェインを通る
- チェイン中は全てIPv6で転送

fx201-1,2.noc (Furukawa Networks FX201) | kamuee.fp1 (NTT Communications kamuee) | ne40e-x2.fp2 (Hitachi NEC/IE-AZ) | ncs55a1.noc (Cisco NCS 55A1)

fx201.fp2 (Furukawa Networks FX201) | xdp-shownet-fs(1.2) (ShowNet) | vpp.fp2 (Cisco/IE-VPP) | seg6 (ShowNet)

Copyright © Interop Tokyo 2019 ShowNet NOC Team

テスターによるSRv6の様々な試験

Function Poolへの誘導

- Function Poolから広告されるSIDによってトラフィックを誘導

Copyright © Interop Tokyo 2019 ShowNet NOC Team

ShowNet 2019 Backbone

- SRv6によるサービスチェイニング
- IP RoutingにSegment Routingによる柔軟性の追加
- SR-ProxyによるSR-unawareファンクションへの対応

Copyright © Interop Tokyo 2019 ShowNet NOC Team

SRv6のモード

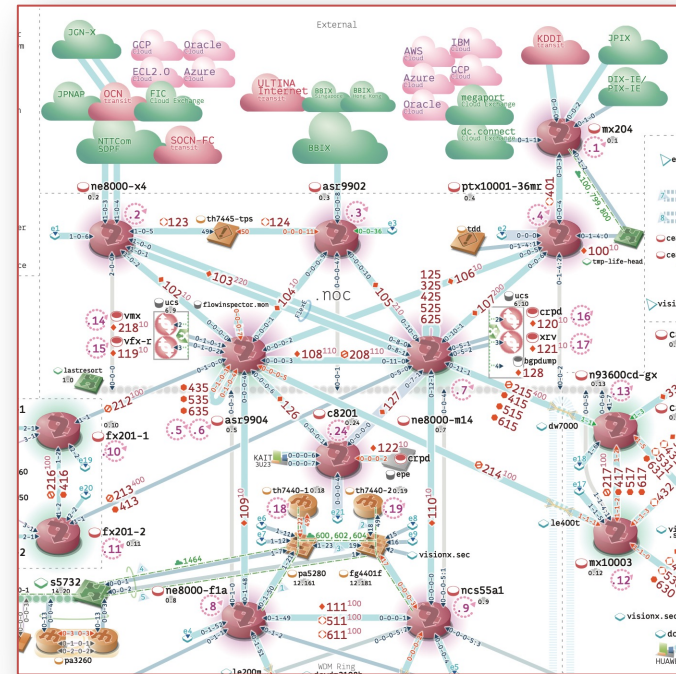
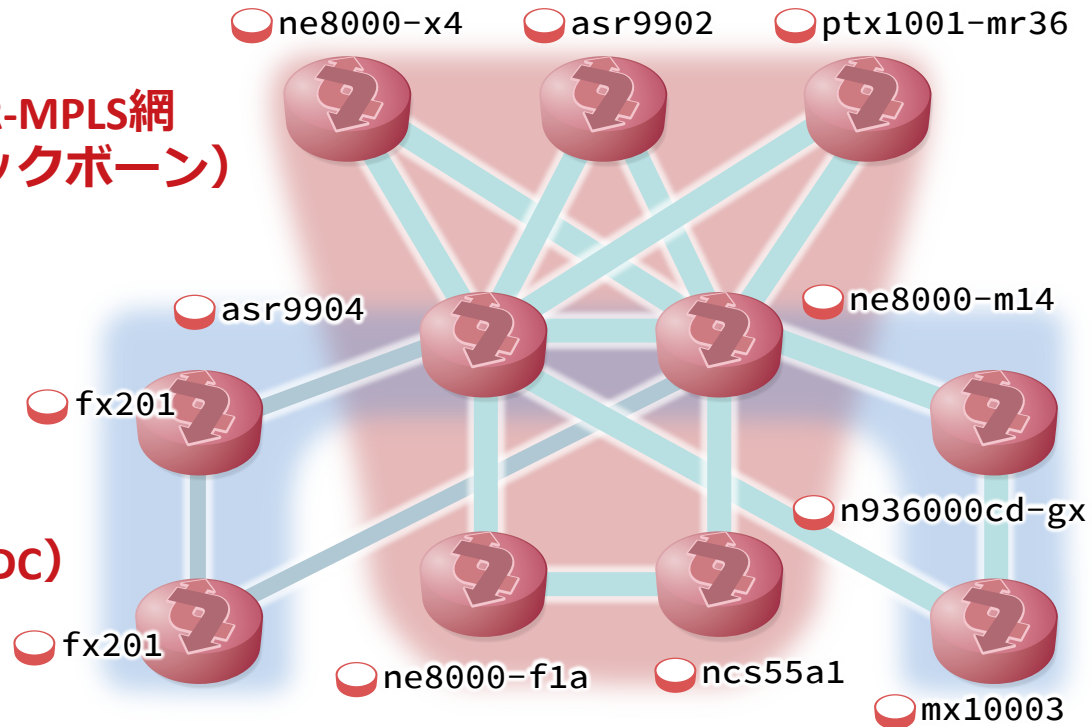
- SRv6には2つのモード (ヘダの付加方式) があり、ShowNet 2019ではアドレスタイプに応じて使い分け
- IPv6の場合はInsert mode
- IPv4の場合はEncap mode
- 決定要因はSRv6 Proxyの方式

Copyright © Interop Tokyo 2019 ShowNet NOC Team

SR-based Backbone in 2021

- コアをSR-MPLS、DC向けをSRv6で構築
- Flex-AlgoやEgress Peer Engineeringのデモンストレーション

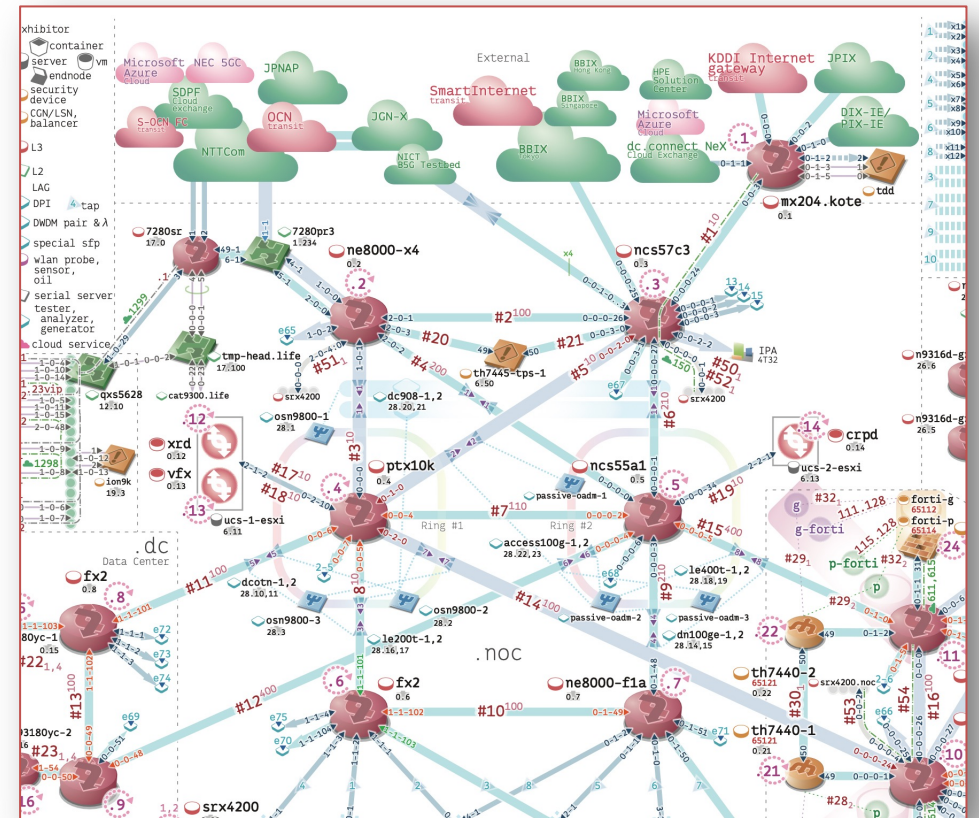
SR-MPLS網
(バックボーン)



SRv6 Single-Stack Multi-Service Backbone in 2022

• ShowNetバックボーンがフルSRv6化

- BackboneはIPv6のみ
- 全てのIPv4/IPv6トラフィックはSRv6 L3VPNで転送
- SRv6 Flex Algo
- uSID相互接続検証
- SRv6による動的な長距離伝送路構築実験



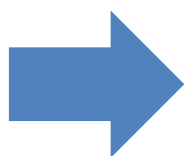
2023年のShowNetバックボーン

• フルVPN化

- VPNのメリットをさらに活用する方向に進化
- フルSRv6でのL3VPN網: 2022までの蓄積を活用
- EVPNベースのアクセス網: 今年の新たなチャレンジ

• 対外接続の進化

- Bright ZR+による400Gbps対外線収容
- RPKI
- DDoS対策



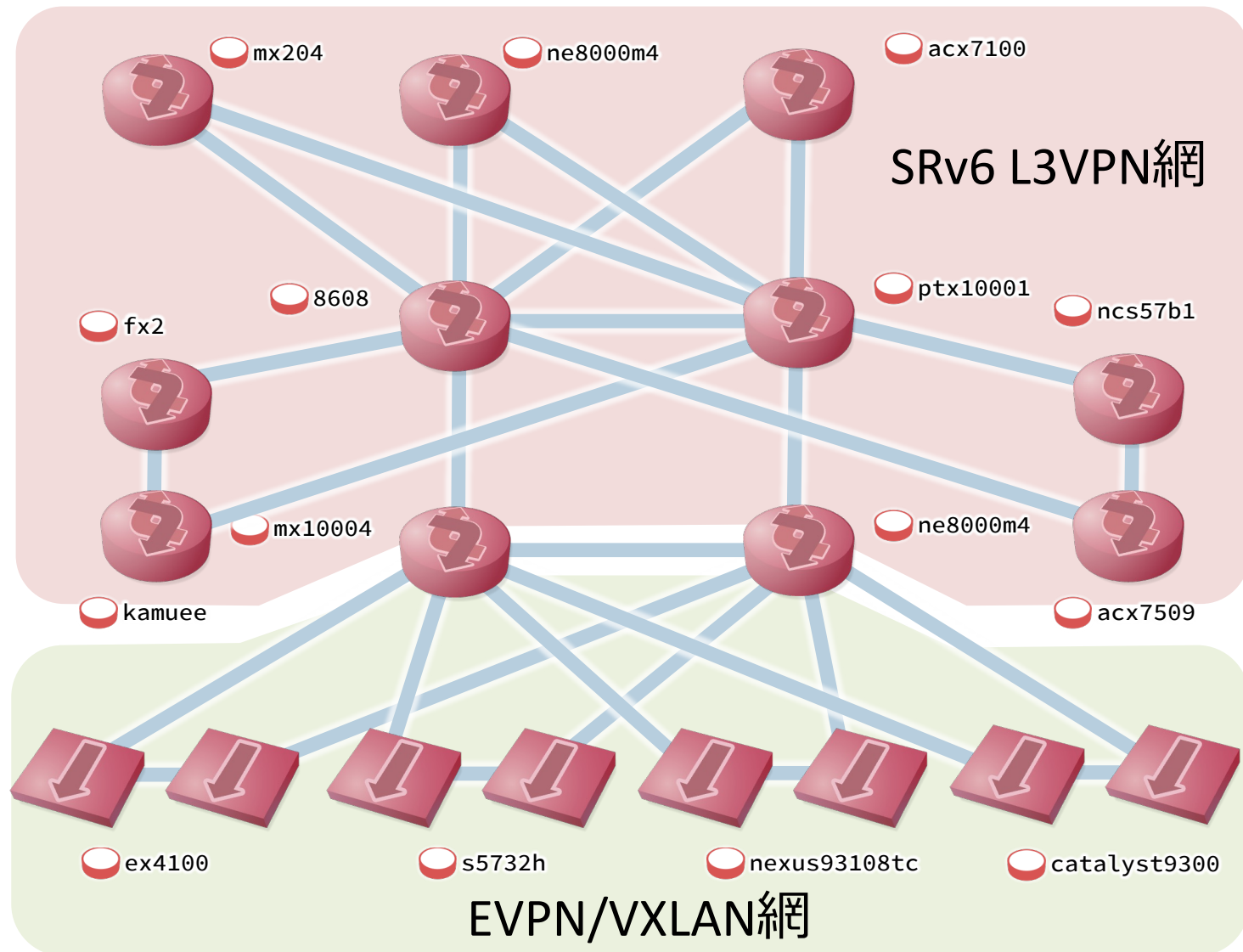
2023年というタイミングで遂に全てのパーツが揃い、
ShowNet全体をVPN化することに成功



フルVPNバックボーン @ShowNet 2023

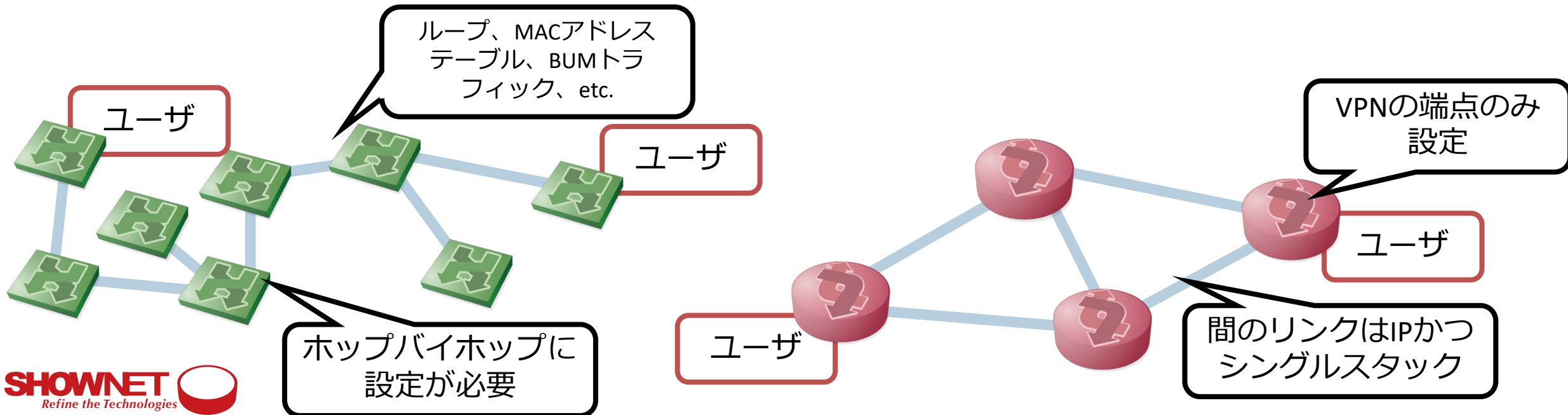
ShowNetのフルVPN化

- 今年>ShowNetは上から下まで全てがVPN
 - バックボーンはSRv6 L3VPN
 - アクセス網はEVPN/VXLAN



VPN化のモチベーション

- オペレーションコスト
 - アンダーレイの構成が単純に
 - IPv6 link local only / IPv4 single stack
 - ホップバイホップに設定を投入する必要がない
- スケーラビリティ
 - 広大かつ多ユーザなL2網をVLANのみで拡張していくのは難しい



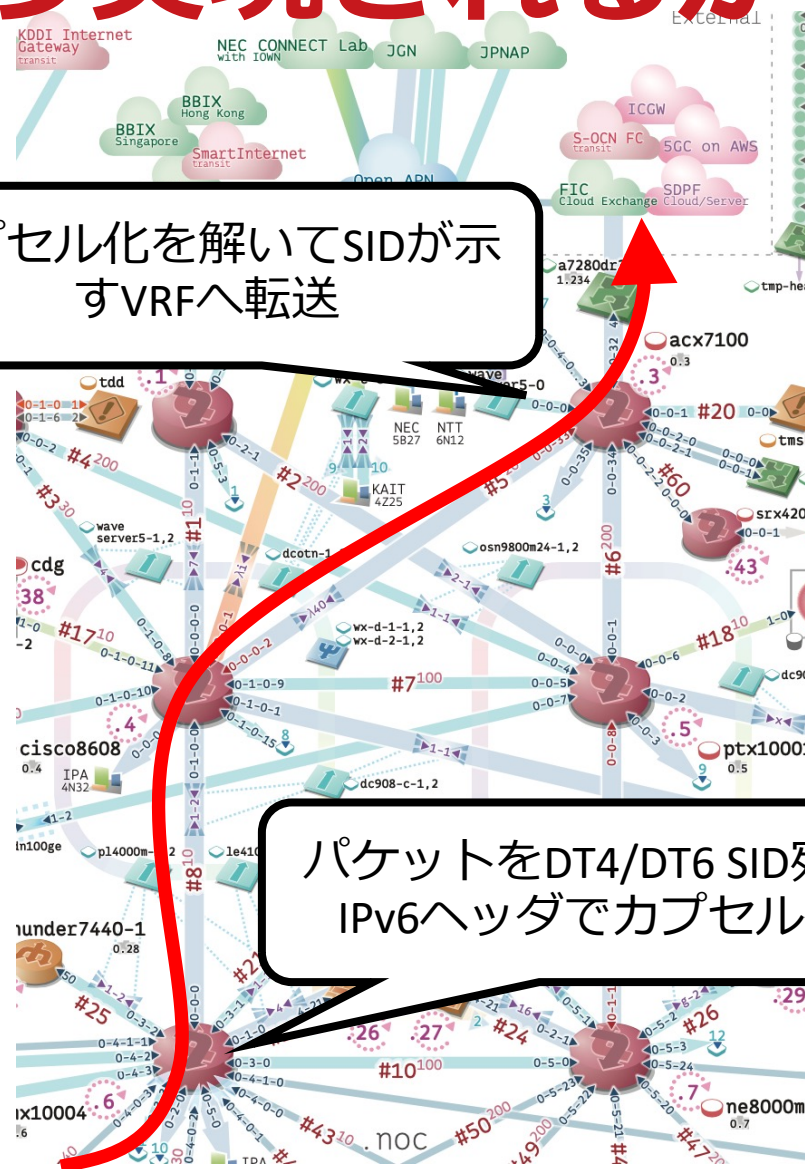
マルチテナントはどう実現されるか

- PEにVRFを作成し、VRF間をSRv6 End.DT4/DT6によって接続
- バックボーンのパケットはIPv6 Link Localアドレスのみで接続

```
interface FourHundredGigE0/0/0/24.14
description fhg-0-2-0.ptx10k.noc
mtu 9021
ipv6 enable
```

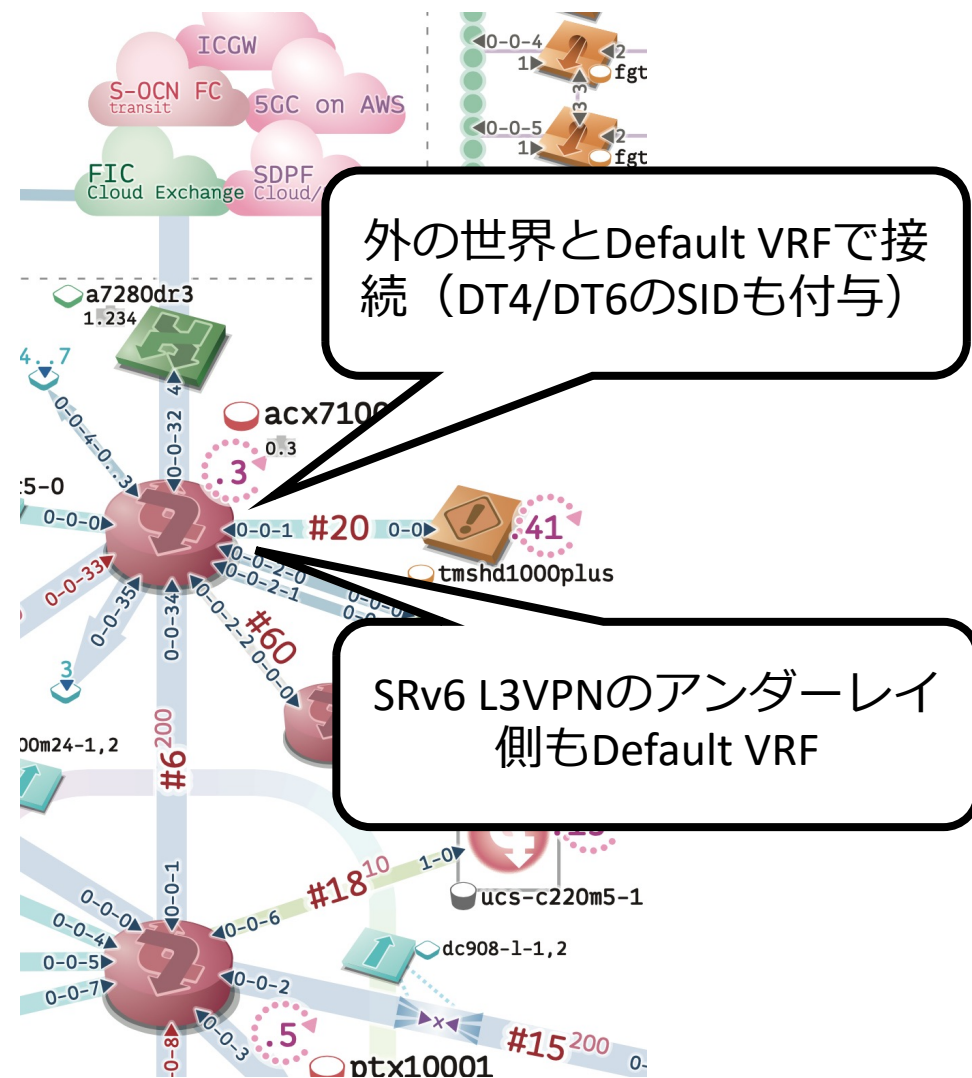
カプセル化を解いてSIDが示すVRFへ転送

パケットをDT4/DT6 SID宛のIPv6ヘッダでカプセル化



Default VRFをどう使うか問題

- インターネット接続面をVRFにする場合の問題
 - 装置によっては一部の機能がDefault VRFでしか動かない
 - アンダーレイとオーバーレイを別々に監視運用する必要
- 今年のShowNetでは、実験的にアンダーレイとインターネット接続面を両方ともDefault VRFに収容
 - FX2/kamueeはこのために追加開発

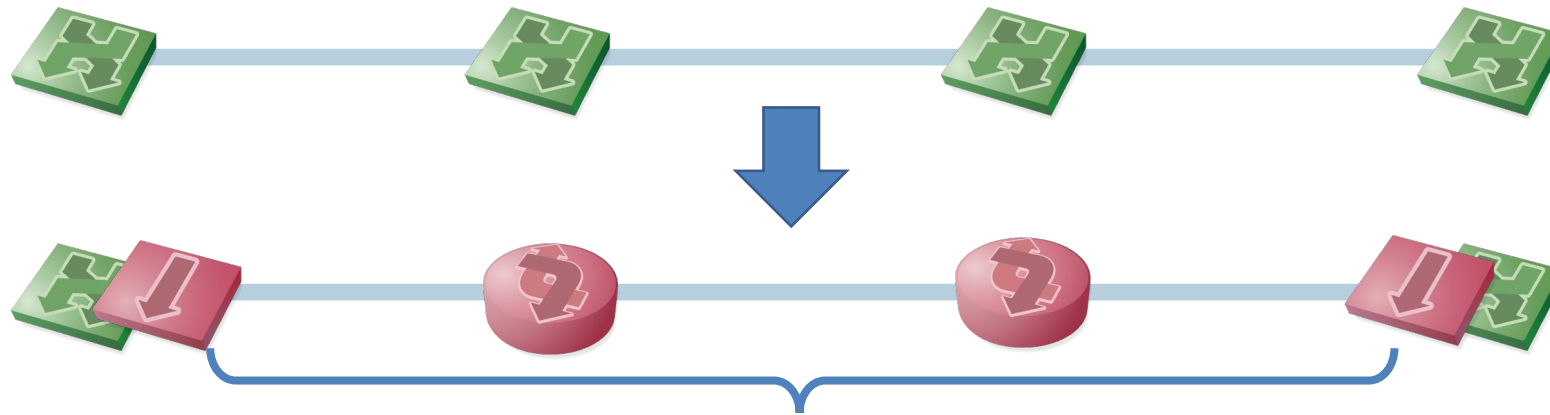


SRv6 L3VPN まとめと課題

- SRv6にしてよかったこと
 - アンダーレイが単純化 (IPv6 Link Local Only)
 - SR-MPLSではBGPによる経路交換のためにIP/MPLSの両方が必要
 - L3VPNの成熟したソリューション
 - 400Gbps程度であればきちんと性能も出る
 - Flex-Algo、uSIDなど拡張機能を使えばさらに柔軟な経路制御が可能
- 今後なんとかしたいこと
 - Network Programmabilityの活用

アクセスネットワークのVPN化

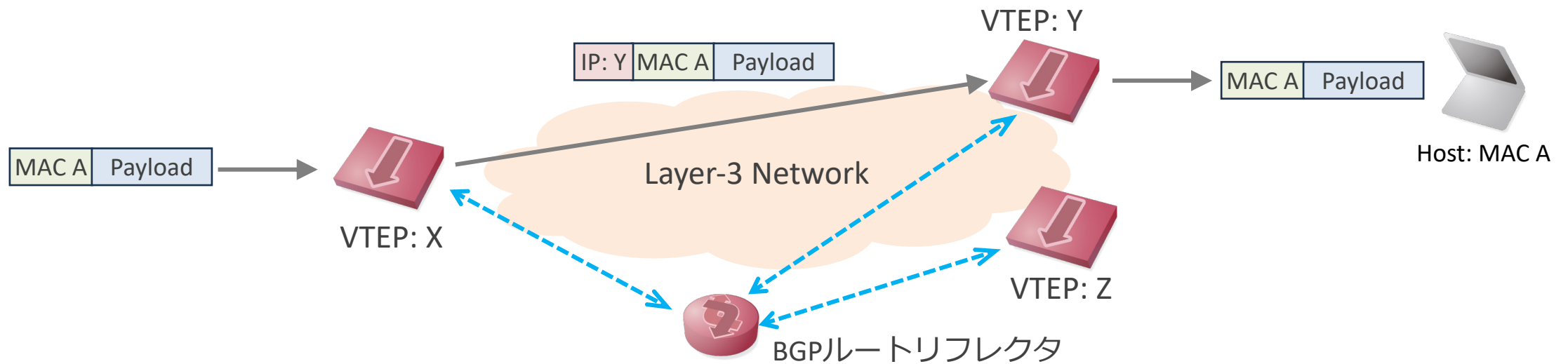
- エンタープライズ、キャンパスネットワークの主流はいまだVLAN
 - 運用コスト、スケーラビリティが課題
- VPNによるL2延伸技術がエンタープライズから注目
 - モチベーションはバックボーンのL3VPN化と似ている



間の区間をL3にすることで運用性・規模性を確保

EVPN-VXLANによるLayer-2の延伸

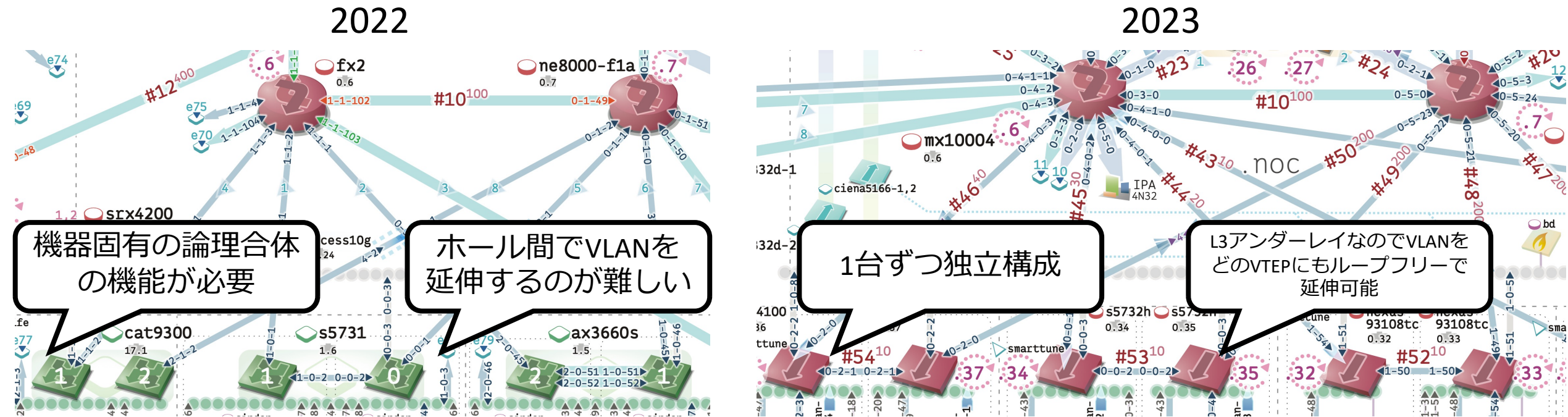
- VXLAN: EthernetフレームをIP越しに転送するオーバーレイ
- EVPN: VXLANの転送先を解決するためのBGPの拡張
 - VTEPとその配下にいる端末のMACアドレスのペアをBGPで交換



データセンターやキャリアでの利用を想定し標準化、実装が進んだ技術

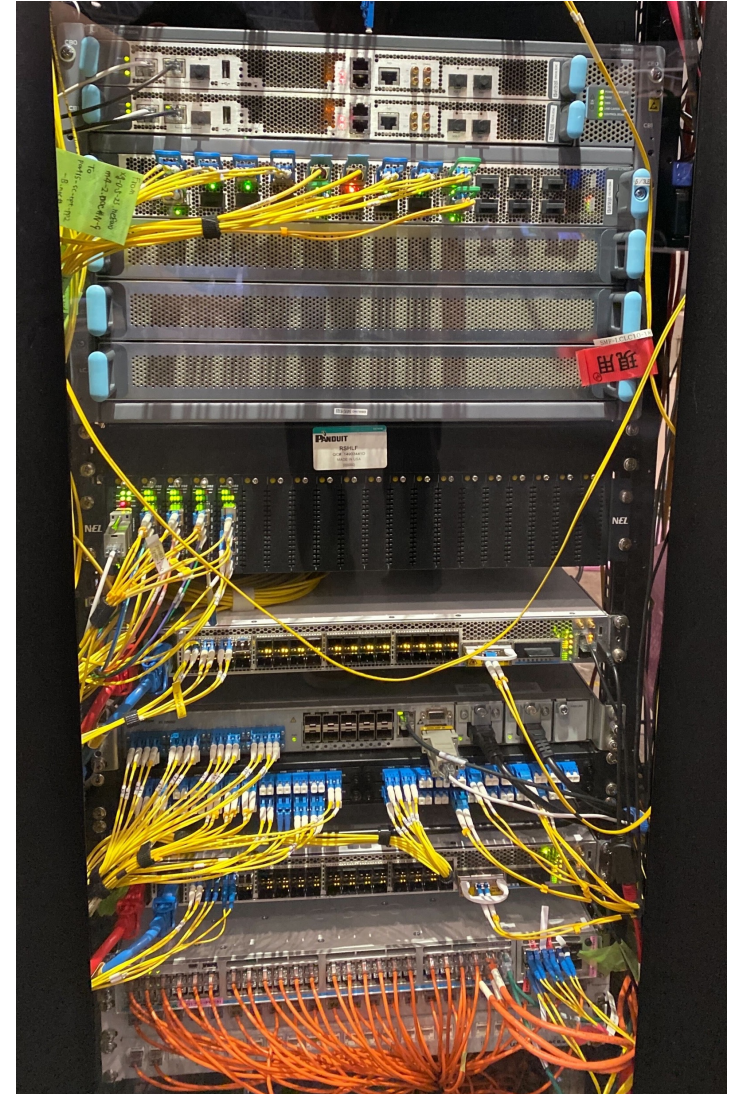
EVPN/VXLANの何が良かったか

- アンダーレイの構成がシンプルなL3になった
 - 冗長を取るのが楽
 - MC-LAGや筐体を論理的に統合する機能が不要
 - 仮に全体の規模が大きくなっても設定が必要なのはVTEPのみ



EVPN/VXLANの相互接続性

- マルチベンダでの相互接続にはハマりどころも多い
 - VLANの扱い（VLAN-Based / VLAN-Aware Bundle）
 - 設計思想の違いに起因する経路広告の内容と期待動作
（特にType2経路にIPアドレスを含めるか）
- それでも、本番では大きなトラブルなく運用することができた
 - EVPN/VXLANのゲートウェイとVTEPがマルチベンダで稼働

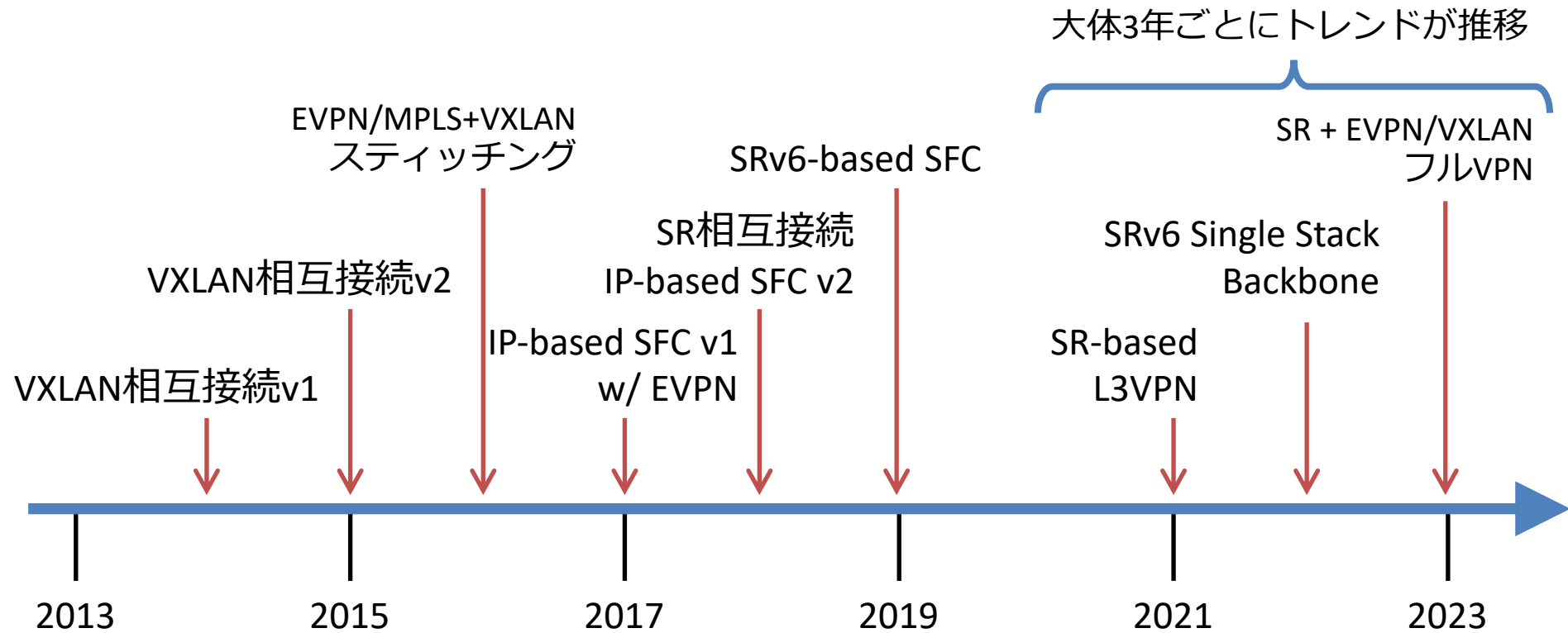




総括と今後の展望

ShowNetのトレンド今昔

- ShowNetバックボーンの基本技術はSRとEVPNに収束



※その年々に参加していた多数の企業・機関のアイデアを集約した結果なので、ShowNetのトレンドはある程度業界のトレンドを反映しているはず

SRとEVPNによって得たもの

- ネットワーク規模の拡張性
- アンダーレイの運用コスト軽減
- プログラマビリティ
- トラフィックエンジニアリング

主にVPN化の文脈で求められてきたもの

主にSDN/NFV/SFCの文脈で求められてきたもの

+ オーバーヘッドも解決されつつある

- ハードウェアでカプセル化を処理できるチップも増えた

所感

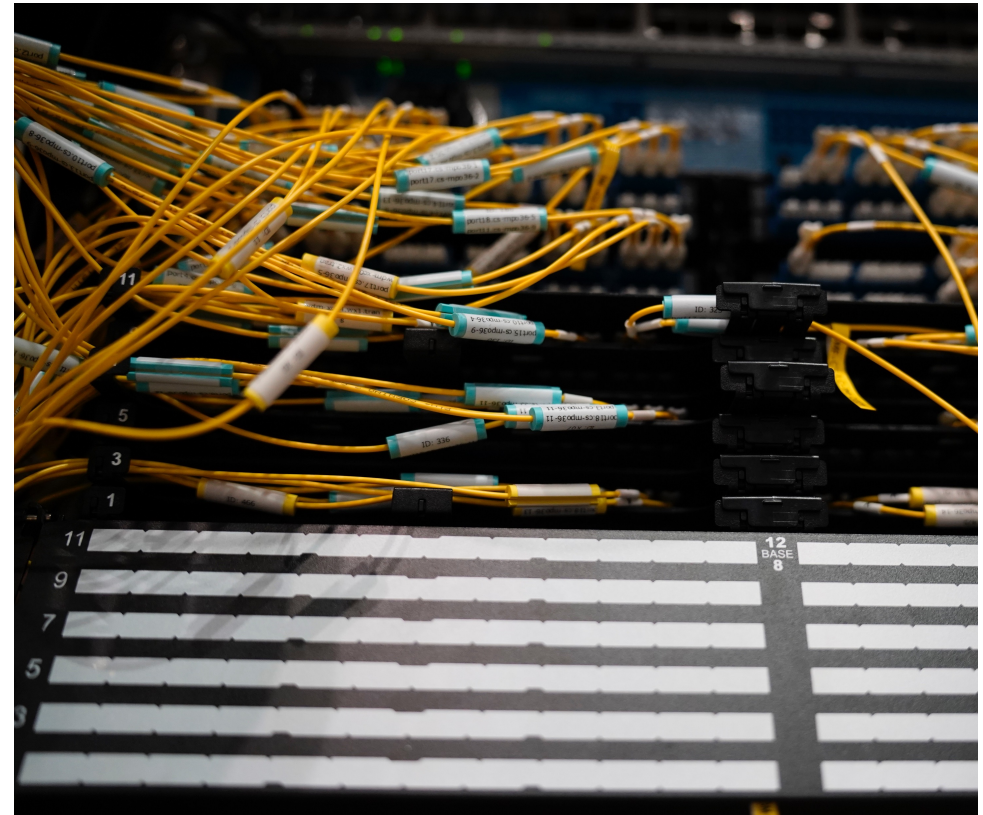
- 規模性/運用コスト低減といったメリットは容易に享受できる
 - アプリケーション透過であるため
 - しかし、ユーザからすると実感の薄いコストメリットに落ち着いてしまいがち
- 一方で、プログラマビリティ/トラフィックエンジニアリングをうまく活かすのは難しい
 - その時々アプリケーションにある程度特化する必要がある
 - その分、ユーザから見てできることを増やせる可能性がある

では、次のアプリケーションって？

- 個人的に特化する余地があると思うアプリケーション
 - Media over IP
 - モバイル
 - セキュリティ (w/ Service Chaining)
 - テレメトリ
 - In-Network Computing
- ぜひ、来年のShowNetに参加して一緒に議論しましょう！
 - <https://f2ff.jp/event/nom-exhibitor>

Interop Tokyo 2023 ShowNet

- 未来のネットワークの1つのカタチ
 - 10年先のインターネットをつくる
 - そのモデルを示すデモと検証
 - 相互接続性
 - ShowNetは異種ベンダー、異種機器間の相互接続で成り立つ
 - オープンな技術の上に成り立つ組み合わせの自由度
 - そしてコミュニティへの還元
- 2024年のSTM募集も開始しました！
 - <https://www.interop.jp/2024/shownet/>



Interop[®]23

Tokyo JUNE 14-16
MAKUHARI MESSE, JAPAN

SHOWNET 
Refine the Technologies

