OSSでつくるyet another SD-WAN

ENOG49 Meeting

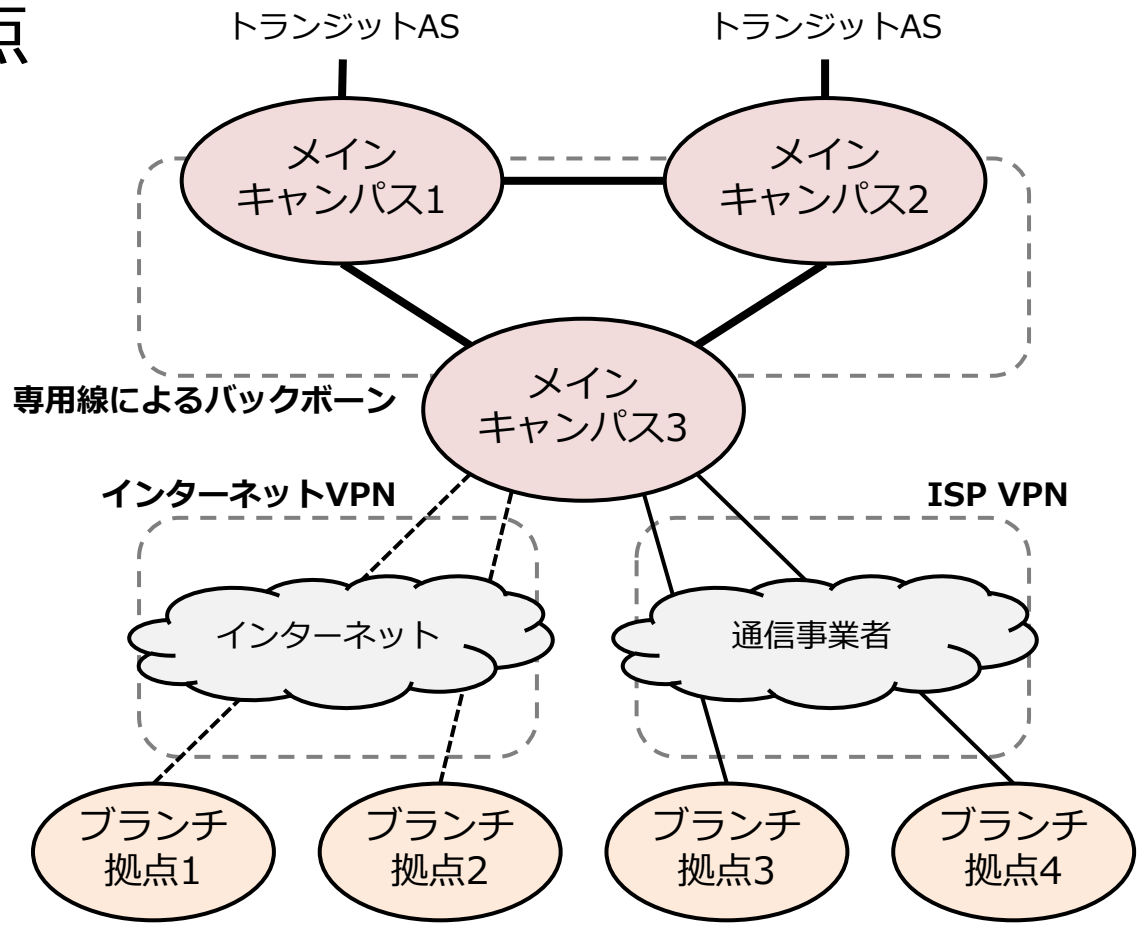
upa@haeena.net

自己紹介

- upa@haeena.net
  - https://github.com/upa
  - AS290, AS2500, AS2501
  - 東京大学 情報基盤センター ネットワーク研究部門 助教
    - 主な研究はネットワーク仮想化技術関連
    - 研究で書くコードはほとんどLinux kernelのnetworkまわり
  - Inteorp Tokyo ShowNet NOC team member
    - 2012年から現在
    - 主にL2/L3とSDNを担当

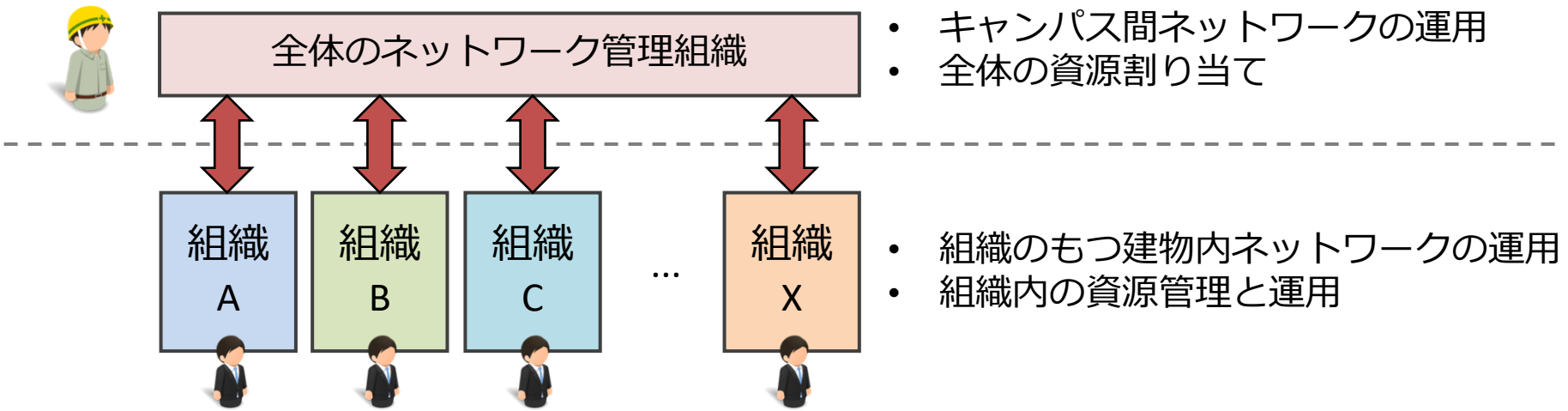
大学のキャンパスネットワーク

- メインキャンパスとブランチ拠点
  - 東京大学の場合
    - メインキャンパス:
      本郷、駒場、柏
    - ブランチ/遠隔拠点:
      白金、小石川、中野、神岡、
      東海村、富士、その他多数
- 接続方法
  - メインキャンパス間: 専用線
  - ブランチ拠点: VPN



キャンパスネットワークの管理

- 組織ごとに資源と運用管理を委譲
  - XX学部、YY研究科、ZZセンターなど
  - 資源: IPアドレスブロック、VLAN番号、ドメイン名など
- エンド側での管理と使い方
  - エンドユーザへのネットワーク疎通性の提供(ラストワンマイル)
  - IPアドレスによるアクセス制限
    - 計算機資源、研究データ、学内システム



- キャンパス間ネットワークの運用
- 全体の資源割り当て

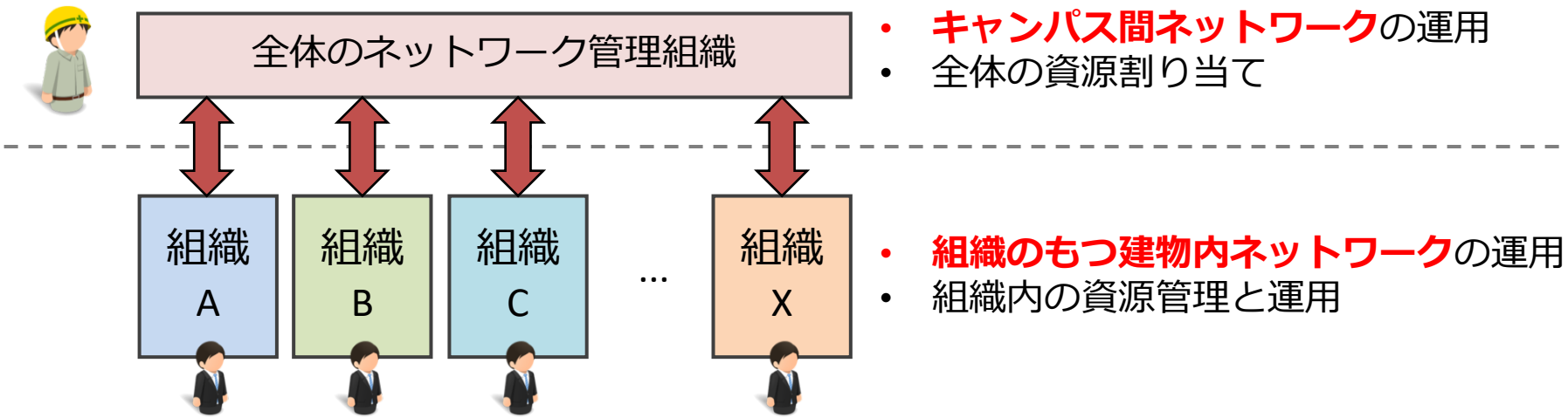
- 組織のもつ建物内ネットワークの運用
- 組織内の資源管理と運用

キャンパスネットワークの管理

- 組織ごとに資源と運用管理を委譲
  - XX学部、YY研究科、ZZセンターなど
  - 資源: IPアドレスブロック、
- エンド側での管理と使
  - エンドユーザへのネット
  - IPアドレスによるアクセス
    - 計算機資源、研究データ、学内シス

ル)

複数のキャンパスや
建物にまたがる組織の
ネットワークは？



- **キャンパス間ネットワーク**の運用
- 全体の資源割り当て

- **組織のもつ建物内ネットワーク**の運用
- 組織内の資源管理と運用

ある作業例: VLAN開通 xN

ユーザ  キャンパスAの建物XでキャンパスBのVLAN 10が使いたい

組織ネットワーク管理者  キャンパス間VLANは全学ネットワークの管理者に申請します

全学ネットワーク管理者  申請を受け付けました

 多重化した申請分の設定作業

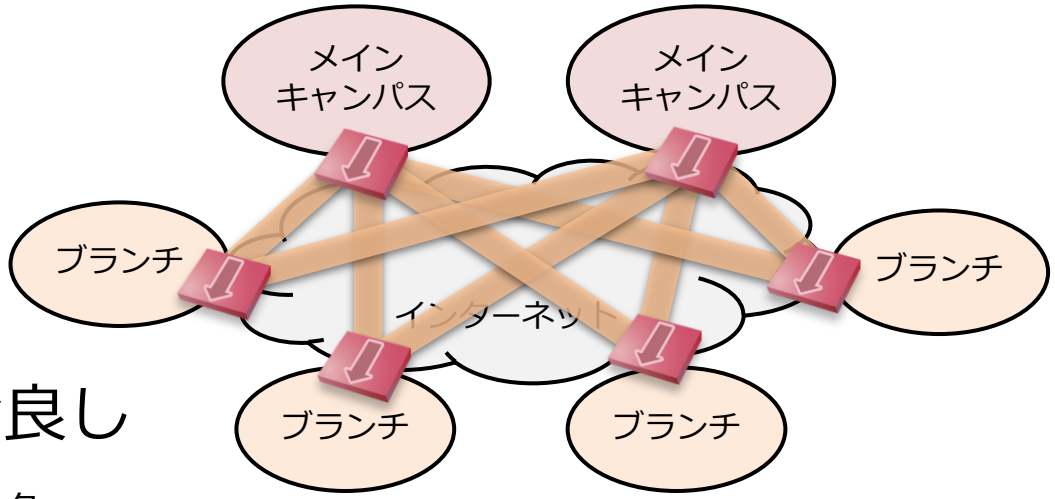
全学ネットワーク管理者  設定が完了しました

組織ネットワーク管理者  建物内はすでに設定済みです

ユーザ  やっとつかえるようになった(怒)

キャンパス間の運用をSD-WANで楽にしたい

- (本発表における)SD-WANとは*1
    - 暗号化されたオーバーレイによる複数拠点間接続
        - オーケストレータによる一括した網制御
- キャンパスネットワークへのSD-WANの適用
    - 複数のブランチ拠点を接続し、キャンパスネットワークを延伸する
    - 拠点間にまたがる設定変更を集中制御で自動化する
- 標準化されたプロトコル
    - オープンソース実装があるならなお良し



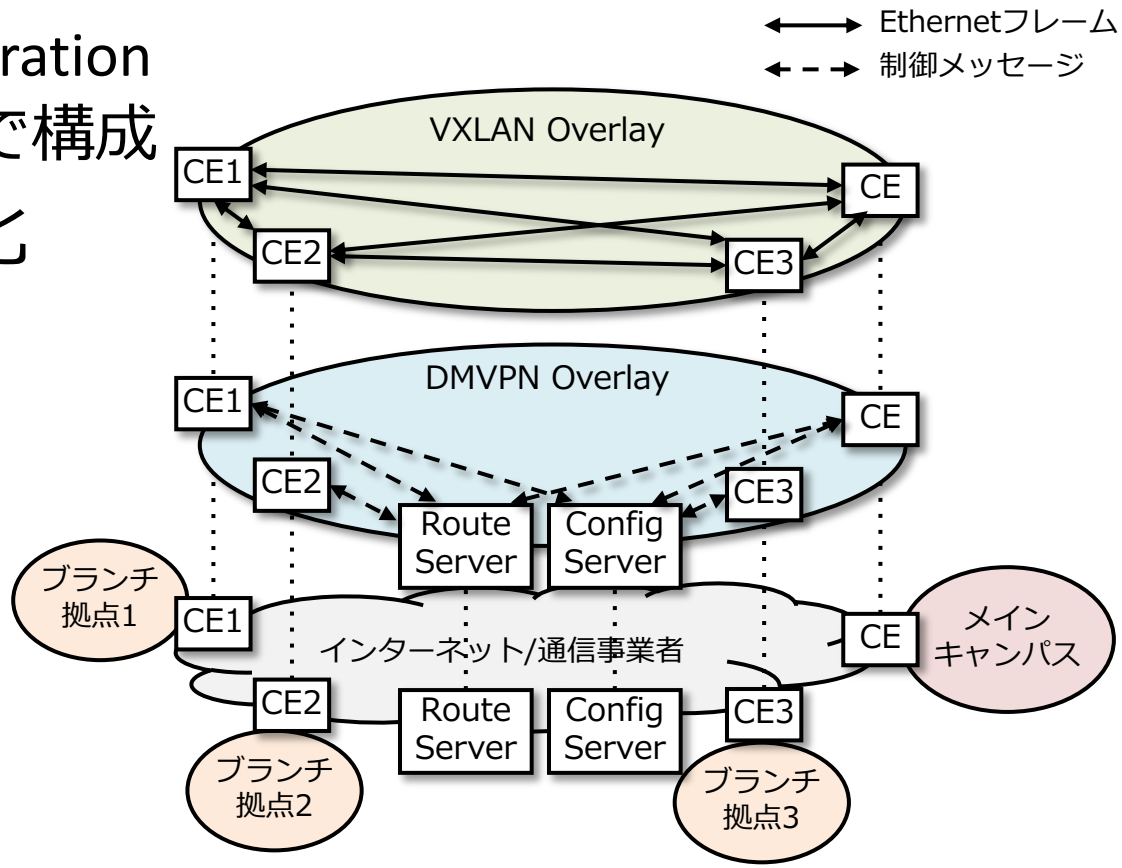
*1: 商用製品では、他にもDPI、可視化、冗長化、アプリケーショントラフィックの最適化、ZTP、Internet Breakoutなどなどなど多種多様な機能が存在

機能要件整理

- データプレーン
  - レイヤー2セグメントの延伸
  - NATを越えられること
  - トラフィックの暗号化
- オーケストレーション
  - 複数機器の一括制御
  - 迅速な設定変更の反映
  - アカウント管理

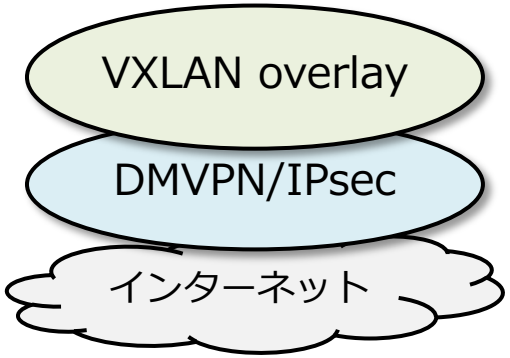
ASANO System

- SD-WANによる複数拠点間のネットワーク延伸
  - Advanced Service And Network Orchestration
  - エッジ装置(CE)とオーケストレータで構成
- 異なる2つのオーバーレイの多重化
  - VXLAN: Ethernet over IP
  - DMVPN/IPsec: IP over IP
- コンテナを用いた制御システム
  - 各制御コンポーネントをコンテナ化
    - Ubuntuを入れてdocker pull,
      docker runするだけでCEを追加



2つのオーバーレイネットワークの併用

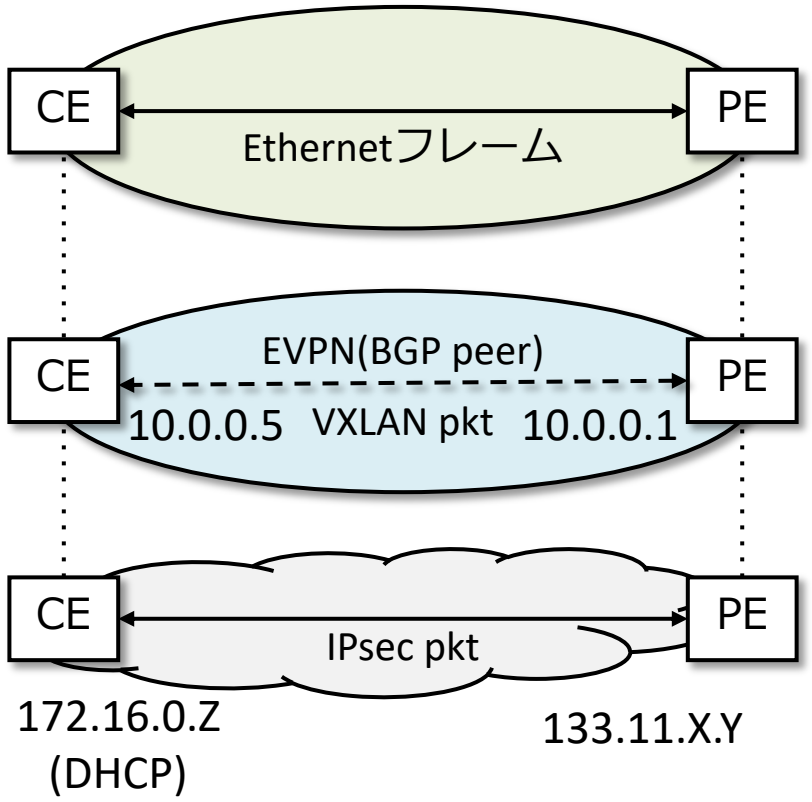
- VXLAN over DMVPN/IPsec
  - トンネルプロトコルはそれぞれの目的に合わせてデザインされている
  - それぞれの欠点を、利点を重ねることで補い合う
  - **NAT越えと暗号化できるマルチポイントL2オーバーレイ**
- 複数回カプセル化する欠点: スループットの低下
  - MTUの減少
  - パケット処理の増加



	L2 Encap	NAT越え	暗号化	多重化	マルチポイント
VXLAN	Yes	No	No	Yes	Yes
DMVPN/IPsec	No	Yes	Yes	No	Yes
VXLAN over DMVPN/IPsec	Yes	Yes	Yes	Yes	Yes

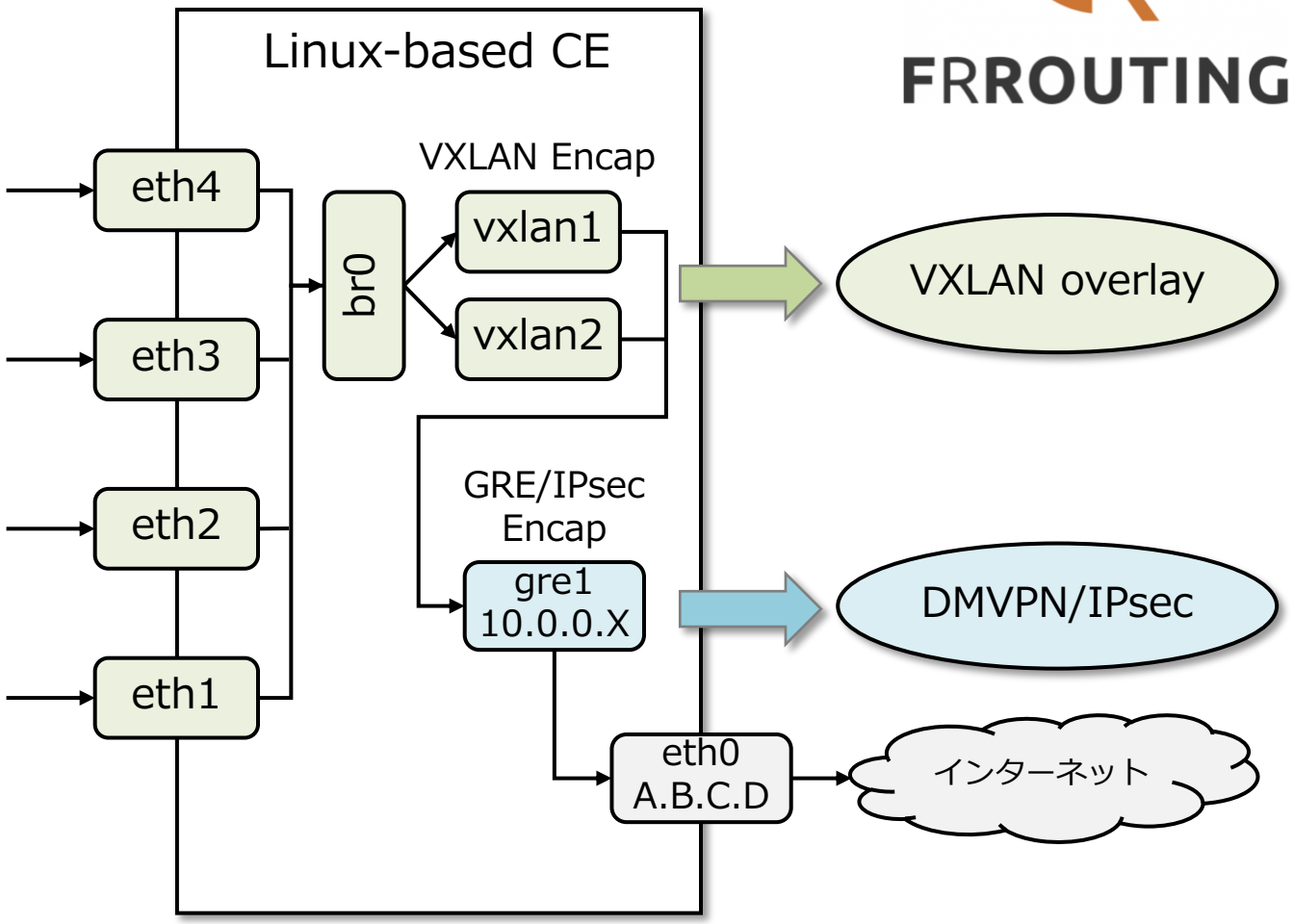
Control Plane

- NBMA Next-hop Resolution Protocol (NHRP)
  - DMVPN上のアドレスからWAN回線のアドレス解決
- Ethernet VPN (EVPN)
  - DMVPN越しにBGPをはってVXLAN用にMAC/VTEP経路を交換
- DMVPN上で、BGPを張りVXLANを通すと、制御に用いるIPアドレスを回線から分離できる
  - WAN回線のアドレスは変わりうるが、
  - ➢**DMVPN上では固定したIPアドレスを利用できる**



OSSでつくるASANO System CE

- OS
  - Ubuntu 17.10
- Software
  - FRRouting
    - commit 67c0a92
    - release 3.1を正座待機
  - StrongSwan
    - nhrp用patch付き: git.alpinelinux.org/user/teras/strongswan

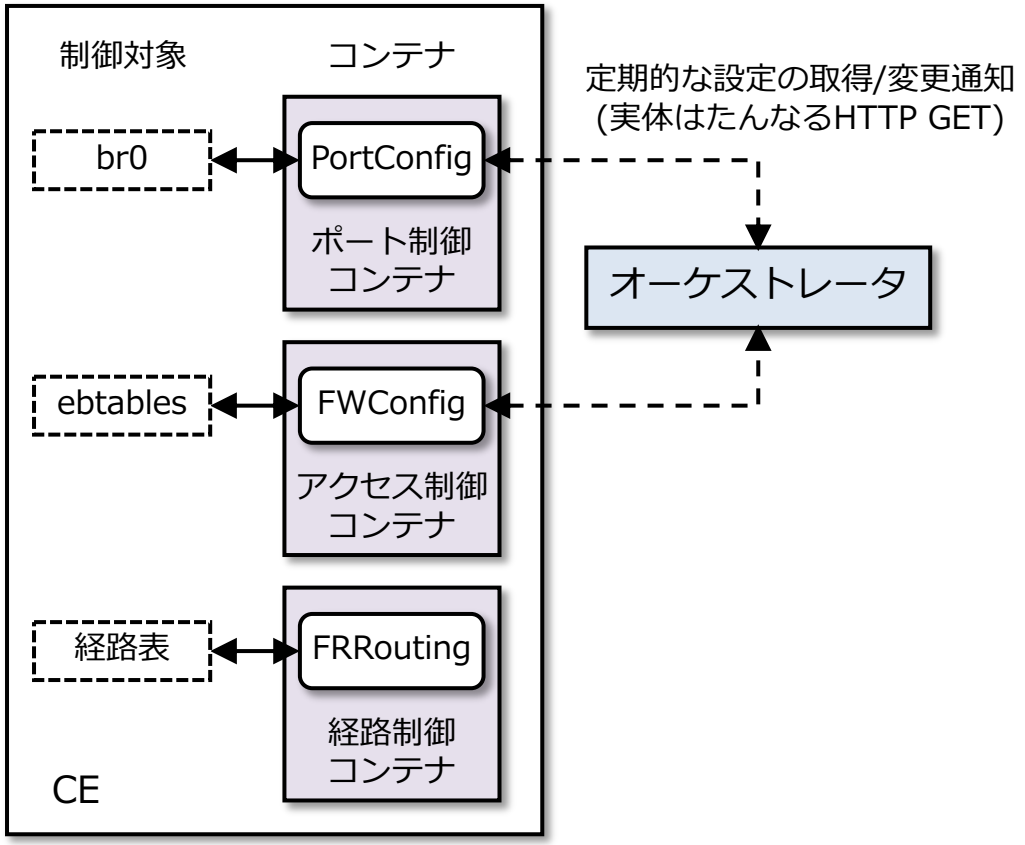


RRのconfigこんな感じ

```
router bgp 65000
 bgp router-id 10.0.0.1
 no bgp default ipv4-unicast
 bgp cluster-id 0.0.0.1
 neighbor asano-system peer-group
 neighbor asano-system remote-as 2501
 neighbor asano-system update-source gre1
 neighbor asano-system capability extended-nexthop
 neighbor 10.0.0.0 peer-group asano-system
 neighbor 10.0.0.1 peer-group asano-system
 neighbor 10.0.0.2 peer-group asano-system
 neighbor 10.0.0.3 peer-group asano-system
 neighbor 10.0.0.4 peer-group asano-system
 neighbor 10.0.0.5 peer-group asano-system
 neighbor 10.0.0.6 peer-group asano-system
 neighbor 10.0.0.7 peer-group asano-system
 neighbor 10.0.0.8 peer-group asano-system
 !
 address-family ipv4 unicast
  redistribute nhrp
  neighbor asano-system activate
  neighbor asano-system route-reflector-client
  neighbor asano-system soft-reconfiguration inbound
 exit-address-family
 !
 address-family l2vpn evpn
  neighbor asano-system activate
  neighbor asano-system route-reflector-client
  advertise-all-vni
 exit-address-family
 vnc defaults
  response-lifetime 3600
  exit-vnc
!
```

(なるべく簡単に)オーケストレーション

- コンポーネントごとにコンテナ化
  - 経路制御コンテナ
  - ポート制御コンテナ
  - アクセス制御コンテナ
- コンテナ化するメリット
  - バージョン管理
  - 環境を含めたパッケージング
  - => **CE作成の手間の削減**
- 設定を一箇所に集約
  - CEからの定期的な取得
  - 変更時はCEに通知



実装中の実機

- 機器
  - 上: Celeron J1900 1.99GHz
  - 下: Celeron J1800 2.41GHz
- Ubuntu 17.10 + Docker
- WiFi APとしても動作可能
  - 802.1x over DMVPMで認証連携
- メインキャンパスのVLAN集約部分は強めのVM

なんちゃってSD-WAN: Nante-WAN

- https://github.com/upa/nante-wan
  - ASANO Systemのオーバーレイ部分の実装
  - これ単体でLinux箱をSD-WAN箱にできる

```
docker run -dt --privileged --net=host
    -v nante-wan.conf:/etc/nante-wan.conf -v /dev/log/:/dev/log
    upaa/nante-wan-routing
```

```
docker run -dt --privileged --net=host
    -v nante-wan.conf:/etc/nante-wan.conf -v /dev/log/:/dev/log
    upaa/nante-wan-portconfig
```

nante-wan.conf

- コンテナの中で
  1. nante-wan.confを読み込み
  2. 必要なconfig(frr.confやipsec.conf)を生成(jinja2)して
  3. zebra/bgpd/nhrpd/ipsecを起動
- 機器を追加するときは
  1. dmvpn_addrを変えて
  2. wan_interfaceを適切に変え
  3. コンテナ起動

```
[general]
dmvpn_addr       = 10.0.0.10

[routing]
wan_interface    = enp1s0
dmvpn_interface = gre1
nhs_nbma_addr    = 10.0.0.1
as_number        = 65001
rr_addr          = 10.0.0.1
ipsec_secret     = hoge
gre_key          = 1
gre_ttl          = 64

[config_fetch]
timeout          = 5
interval         = 3600
failed_interval = 5

[portconfig]
br_interface     = bridge
json_url_prefix = http://10.0..0.1/portconfig
bind_port        = 8080
```

できること、できないこと

- Internet Breakout
  - L2なのでちょっと難しいがOFなどでひねればできなくもない
- アプリケーショントラフィック最適化
  - 適当にひねればできなくも(略
- storm-control
  - できあいの機能はおそらく無い? qdiscかrx_handlerかnf_hookらへんでコード書けばできな(略
- WAN回線冗長
  - nhrpで切り替えか、load balanceはEVPN multi-homeかipsecのとこコード書けば(略
- 可視化とかDPIとかセキュリティとか
  - Linuxなので好きなものをどうぞ
- Zero Touch Provisioning
  - Preseedでわりとがんばれる
- Full meshのrekey
  - 数が増えたら考える
- 大抵のことは(userlandでもkernelでも)コード書けばなんとかできなくはない
  - どこまでやるかは現在考え中

まとめ

- OSSと標準化されたプロトコルでつくるなんちゃってSD-WAN
  - VXLAN/EVPN over IPsec-DMVPN/NHRP
  - powered by FRRouting
  - managed by docker container
  - on Linux