オーバーレイSDN最新技術動向

日本アルカテル・ルーセント　鹿志村　康生
yasuo.kashimura@alcatel-lucent.com

Alcatel·Lucent

IETF NVO3 WG : NETWORK VIRTUALIZATION OVERLAYS (OVER L3)

http://datatracker.ietf.org/wg/nvo3/charter/

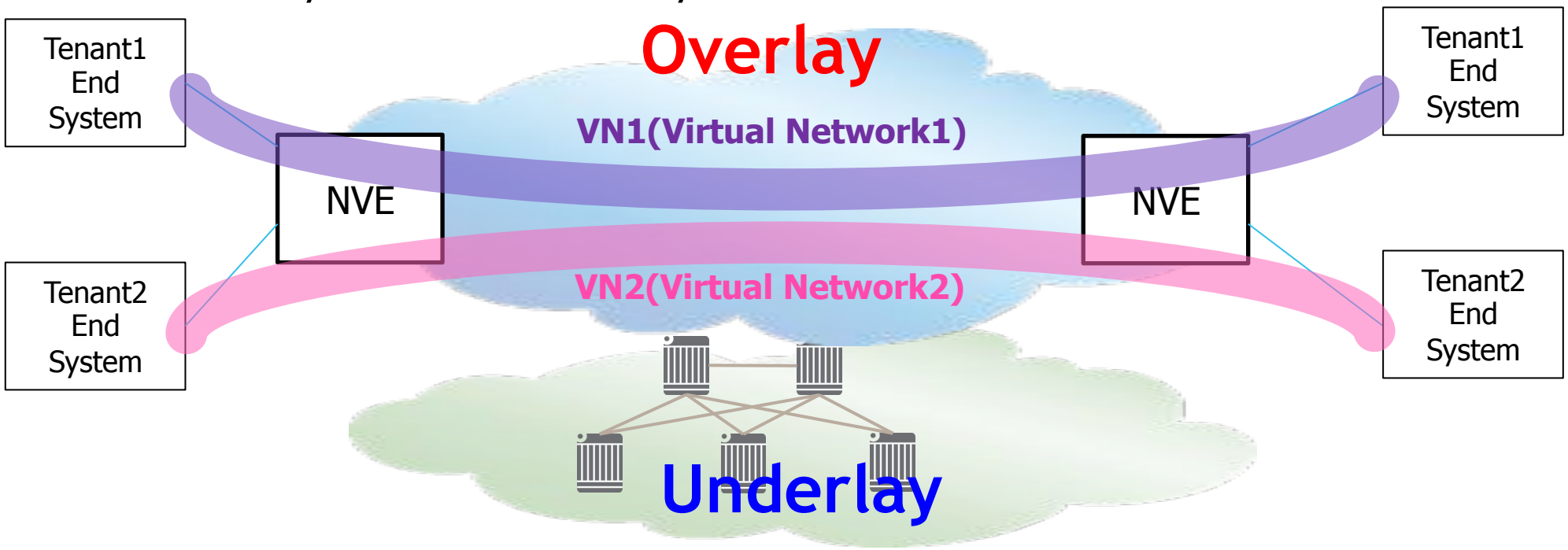
IPベースのUnderlayネットワーク上で、マルチテナンシーやモビリティーを可能とするLayer2/Layer3の仮想ネットワークサービスを提供するためのプロトコル/プロトコル拡張の開発を目指す。

- RFC 7364 : Problem Statement : Overlays for Network Virtualization
- RFC 7365 : Framework for Data Center (DC) Network Virtualization

Feb 2015	Data Plane Requirements submitted for IESG review
Feb 2015	Control Plane Requirements submitted for IESG review
Feb 2015	Operational Requirements submitted for IESG review
Feb 2015	Security Requirements submitted for IESG review
Apr 2015	Architecture submitted for IESG review
Apr 2015	Use Cases submitted for IESG review
Oct 2015	NVE - NVA Control Plane Solution submitted for IESG review
Oct 2015	End Device - NVE Control Plane Solution submitted for IESG review
Oct 2015	Data Plane Solution submitted for IESG review
Dec 2015	Recharter or close working group

NVO3 FRAMEWORK/ARCHITECTURE OVERVIEW (1)

Overlay Network:L3 overlayにより仮想ネットワーキング機能を提供

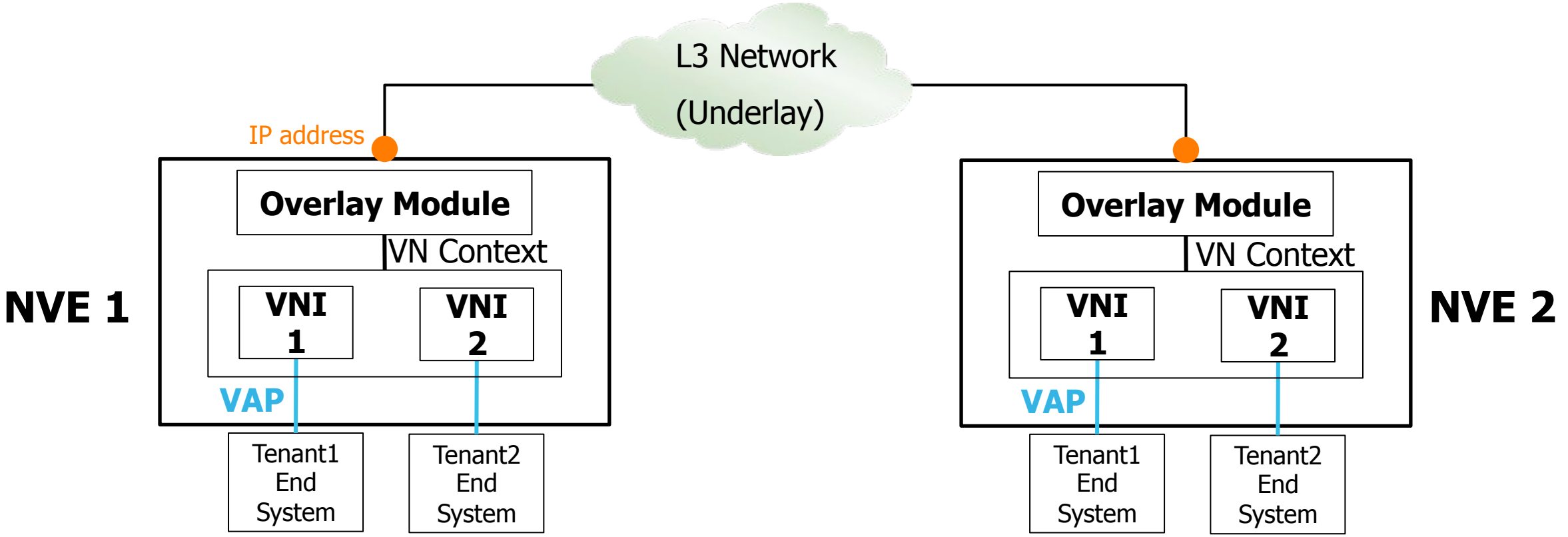


Underlay Network：NVE間のIP Reachabilityを提供, Overlayのstate等は管理しない。IP Tunnelingの機能

NVE(Network Virtualization Edge)：OverlayによるL2/L3仮想ネットワーク機能を提供するエッジ

L2サービス(Ethernet LAN-Like)、L3サービス(IP VPN-Like)両方のサービスを想定

NVO3 FRAMEWORK/ARCHITECTURE OVERVIEW (2)

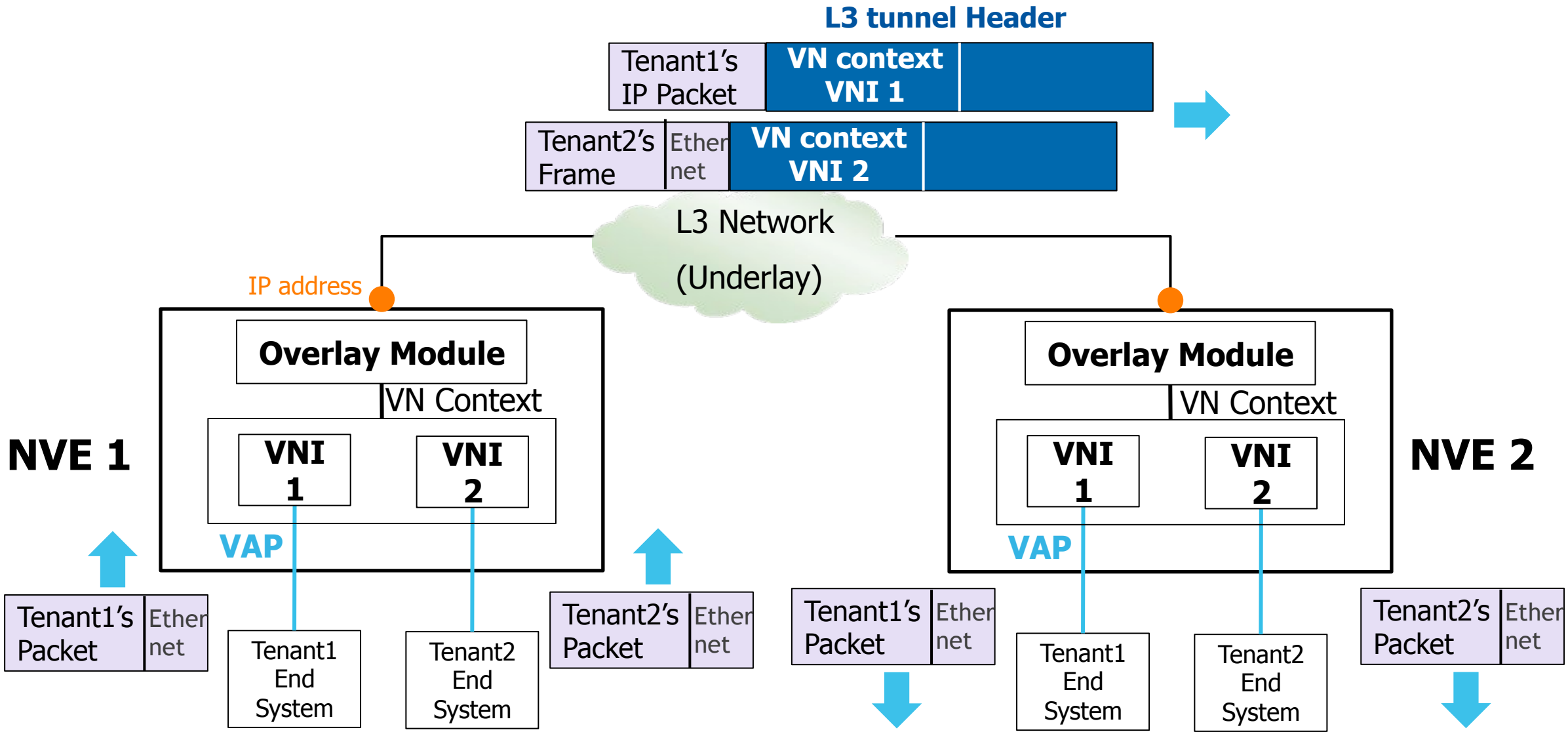


VNI (Virtual Network Instance): Virtual Networkの特定のインスタンス

VAP（Virtual Access Point）: テナントシステムを接続するためのポート(物理ポート/仮想ポート)

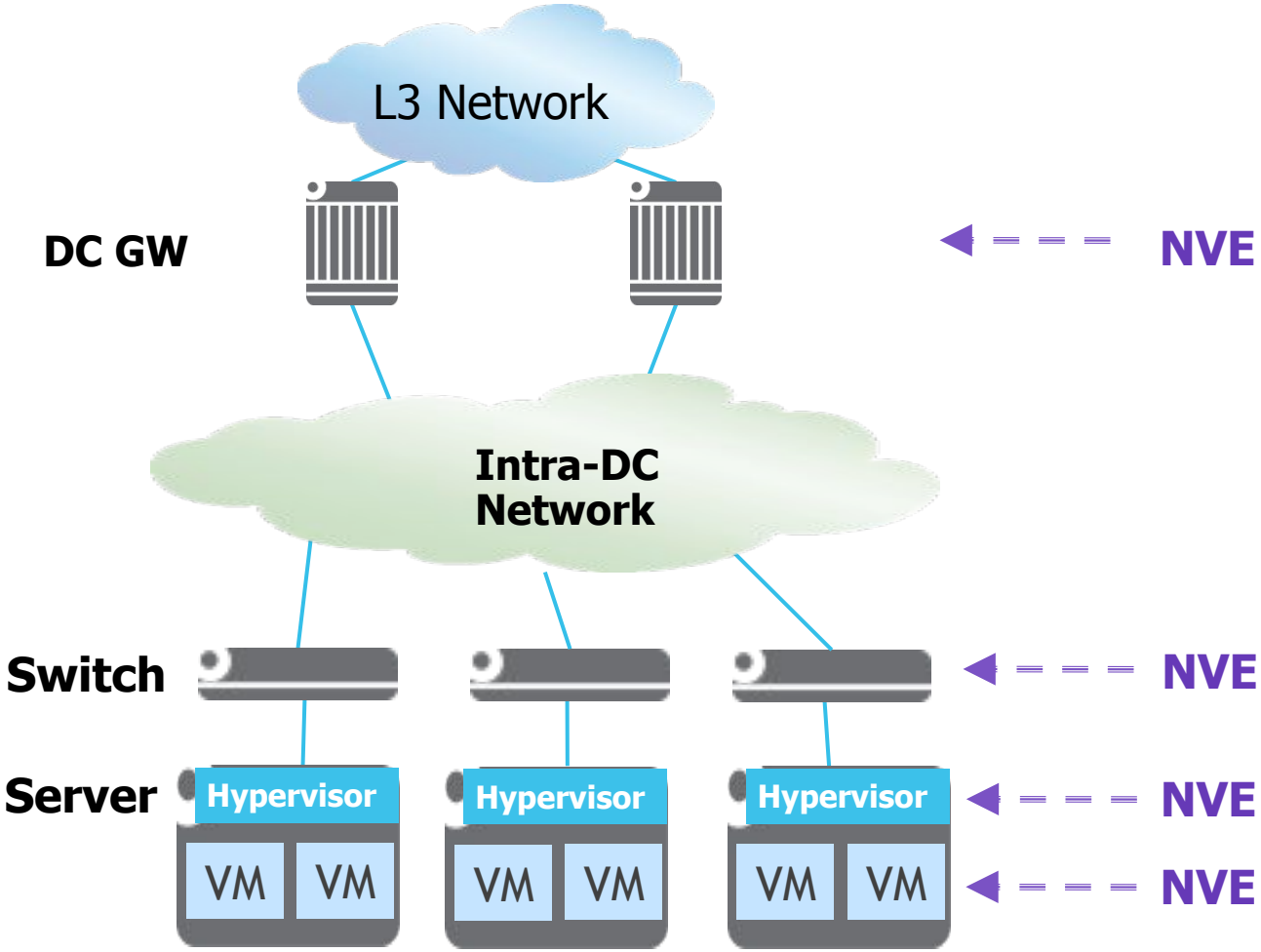
VN Context Identifier: VNの識別子

NVO3 FRAMEWORK/ARCHITECTURE OVERVIEW (3)



NVO3 FRAMEWORK/ARCHITECTURE OVERVIEW (4)

NVE の機能配置



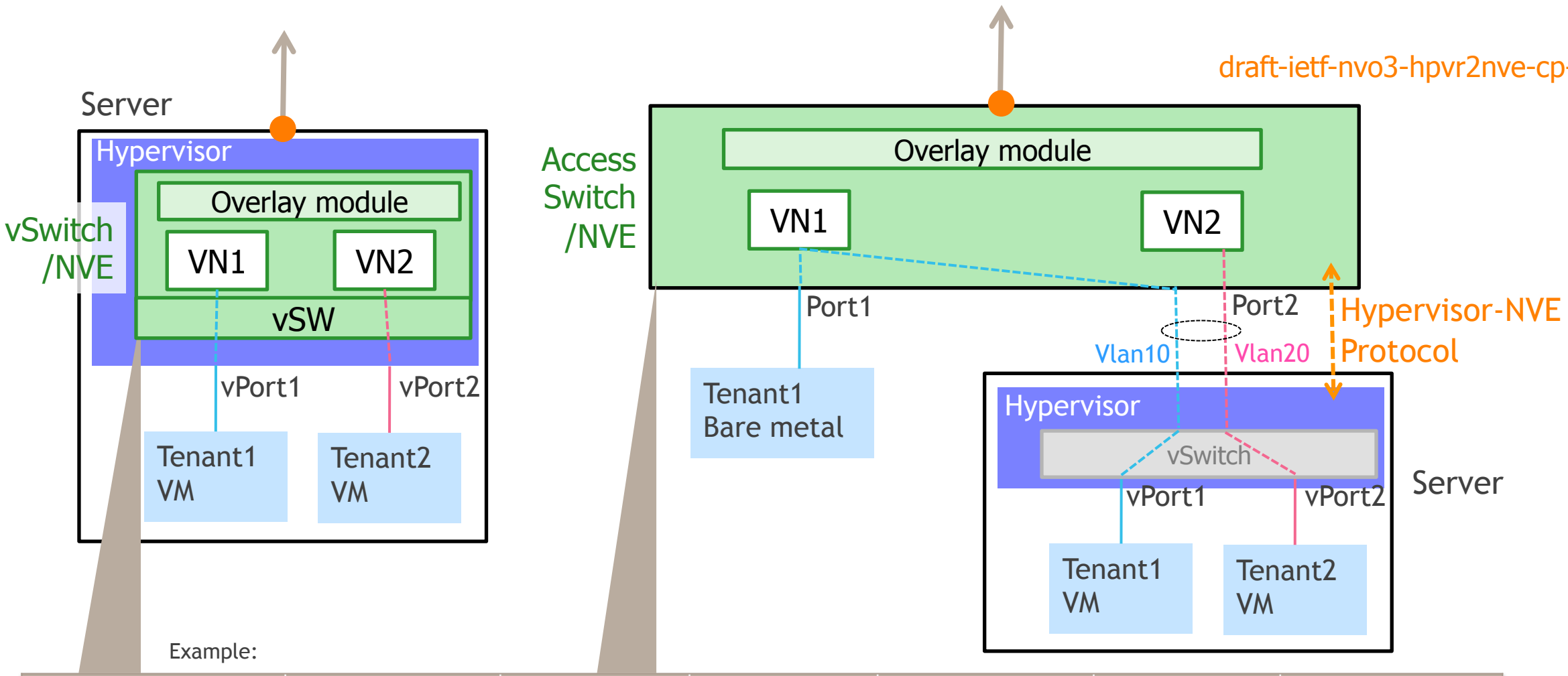
NVEの機能配備レイヤを決める際に考慮すべきこと

- 処理能力とメモリの要求
  - データパス(lookup/filtering/encap/decap)
  - Control plane(routing/signaling/OAM)の配置
- FIB/RIBのサイズ
- Multicastのサポート
  - プロトコル/Replicationポイント
- Fragmentation
- QoS
- Resiliency

NVO3 FRAMEWORK/ARCHITECTURE OVERVIEW (5)

NVEの配備例とVNの識別

※NVE機能が分割して配置されるケースもあり。



draft-ietf-nvo3-hpvr2nve-cp-req

Example:

vPort/ Port(and Vlan)	VN Context	Src Addr	Dest Addr	Encap Type	Local-NVE-Addr	Remote-NVE-Addr

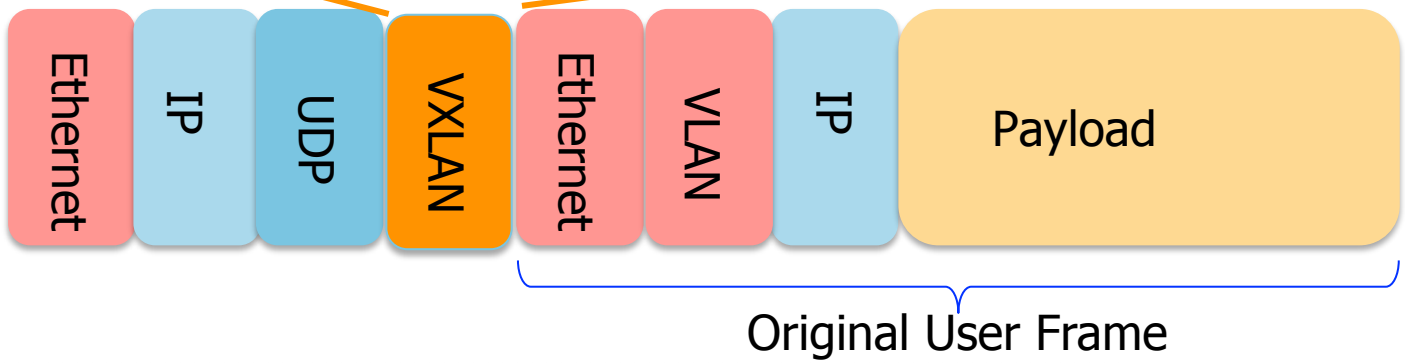
Data-Plane 代表的な Encapsulation
VXLAN: RFC 7348

- UDP/IPでL2フレームをカプセル化
- 24-bitのVNIDによるVN識別
- Innerヘッダの情報を元にHashによりOuter UDP SRC-PORTを決めるとMulti-pathが良い感じに。

  (分散のための十分なエントロピー、Inner Flow毎の経路の統一性)

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R|R|R|R|I|R|R|R|                Reserved                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                VXLAN Network Identifier (VNI) |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

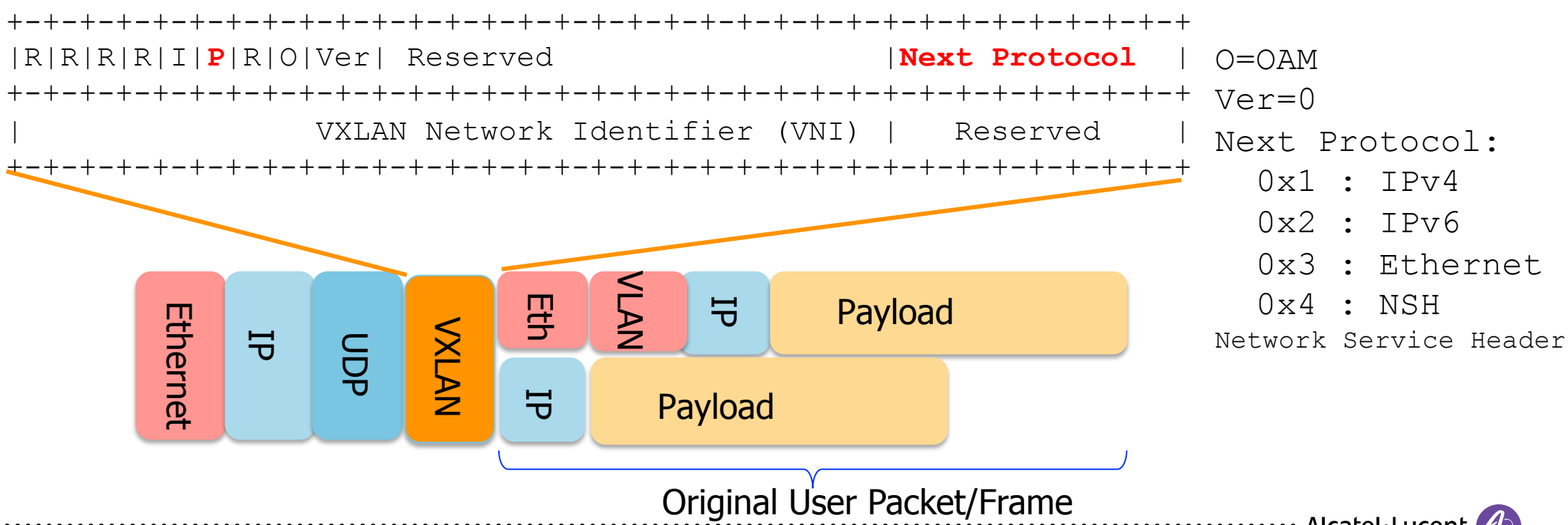
I=1



Data-Plane 代表的な Encapsulation
VXLAN GPE(Generic Protocol Extension)

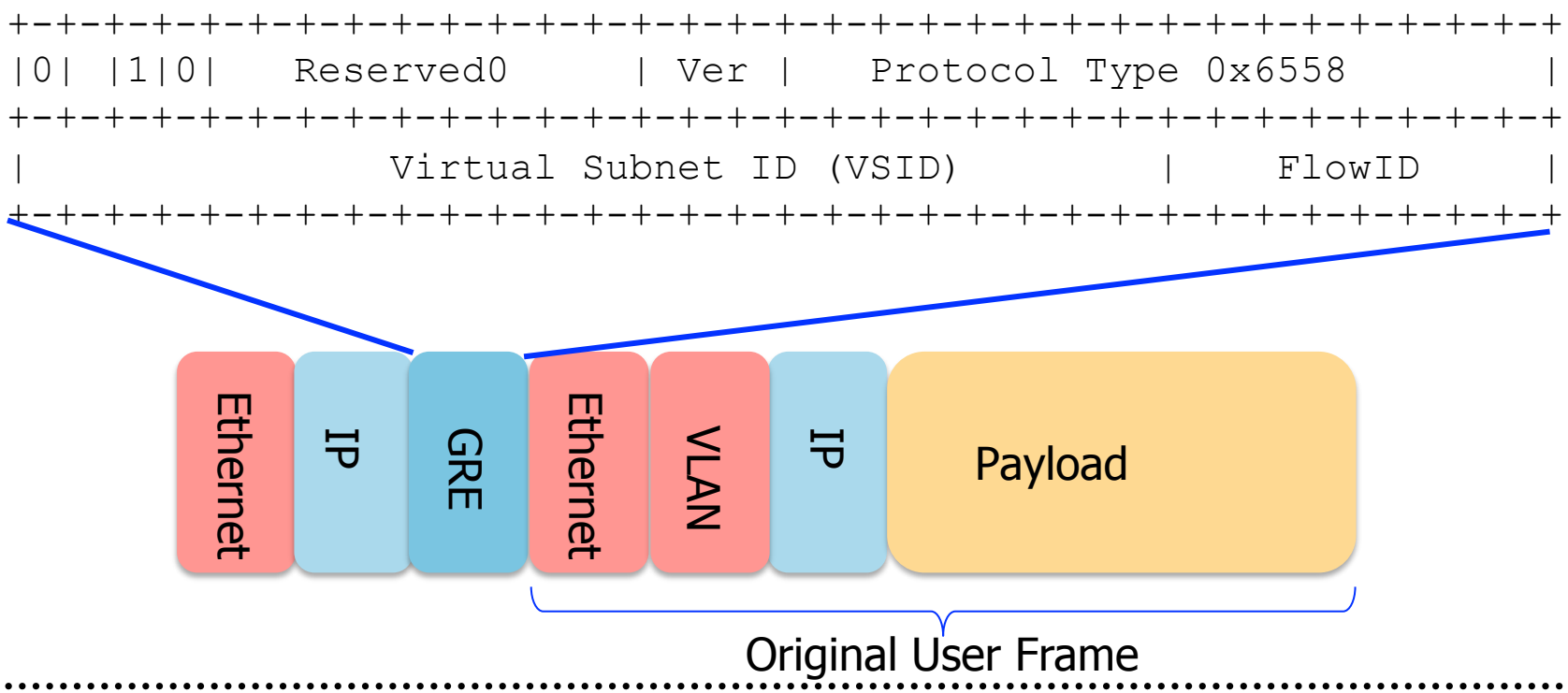
draft-quinn-vxlan-gpe

- VXLANをEthernetだけでなくMulti ProtocolのTransportとして拡張
- Next Protocolフィールドを使用する場合はP=1、P=0の場合はRFC7348 VXLANに準ずる



Data-Plane 代表的な Encapsulation
NVGRE : draft-sridharan-virtualization-nvgre

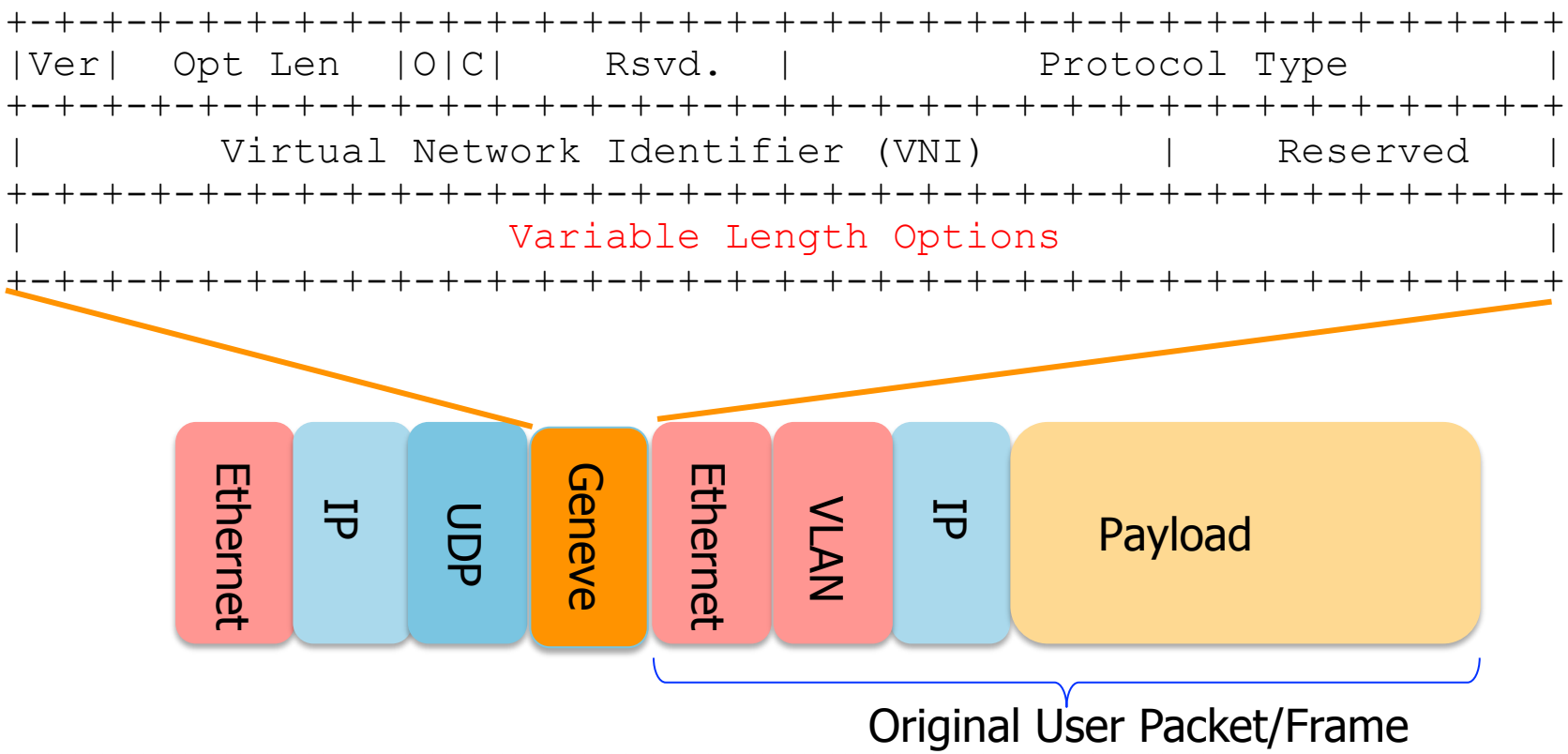
- GRE/IPでL2フレームをカプセル化
- 24-bitのVSID(Virtual Subnet ID)によるVN識別
- FlowID: Flowエントロピーのために使用、Encapsulationする側がセット
- Multipathを考慮する場合はOuter IPヘッダ+GREヘッダを見る必要あり



Data-Plane 代表的な Encapsulation
Geneve (Generic Network Virtualization Encapsulation)

draft-gross-geneve (VMware/MS/RH/Intel/Broadcom/Arista/Cumulus)

- ヘッダに含められる情報の拡張性(Options/Tunnel Options、メタデータ等の付随情報など)
- 拡張性とHWでの実装のしやすさの両立を目指す。(Intel NICで実装済み)



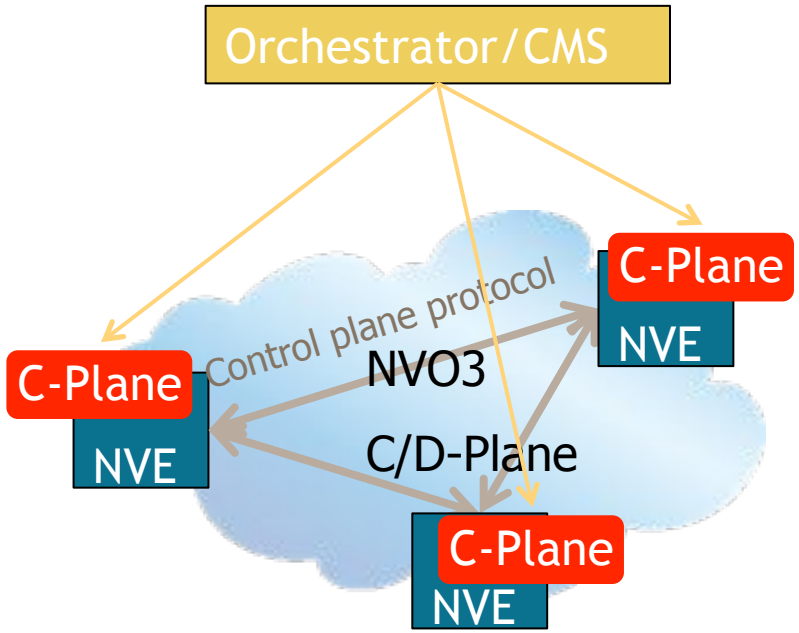
その他のEncapsulation方式

- MPLS over GRE：RFC4023
  - IP-VPN like L3サービスで使用する実装あり。多くのルータでサポートしている。
- STT：draft-davie-stt
  - VMwareが提案しているL2oL3の方式。TCP-like。

NVO3 FRAMEWORK/ARCHITECTURE OVERVIEW (6)
コントロールプレーンの機能配備

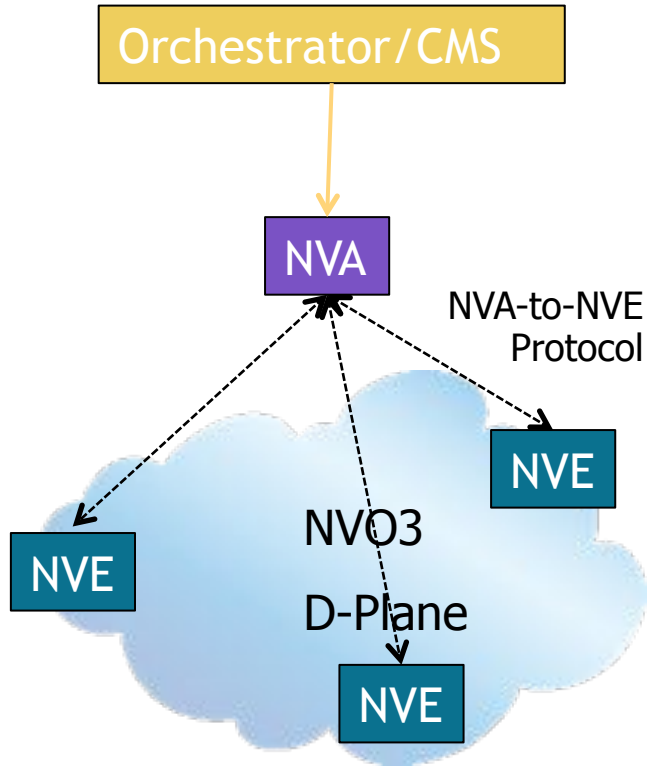
NVEにコントロールプレーンを実装

Full Distribute型



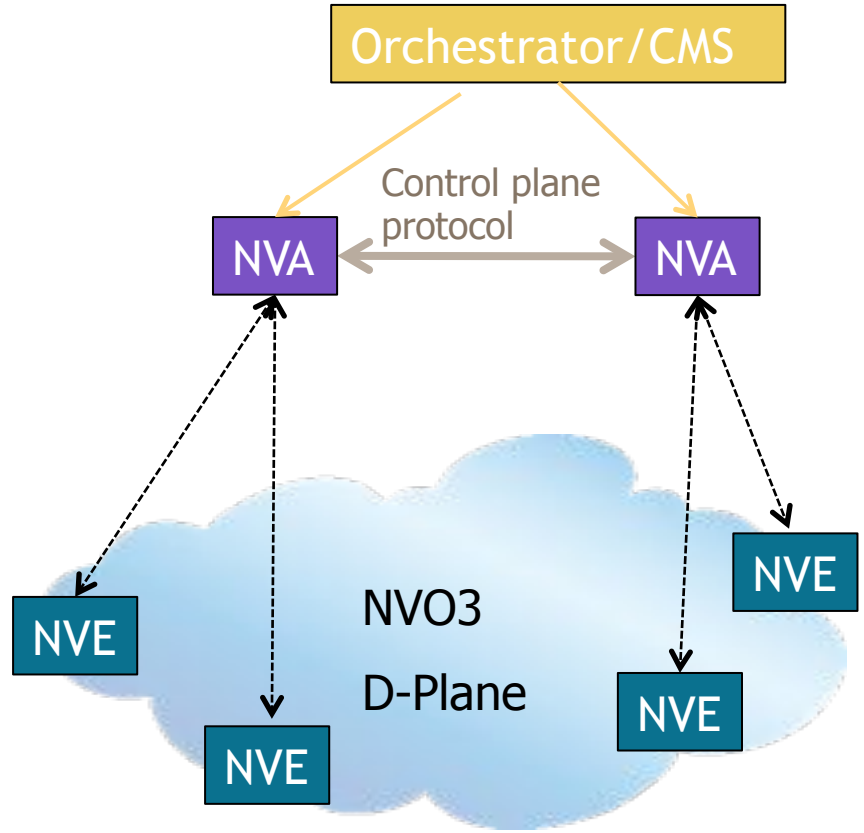
NVAにコントロールプレーンを実装

Centralized型



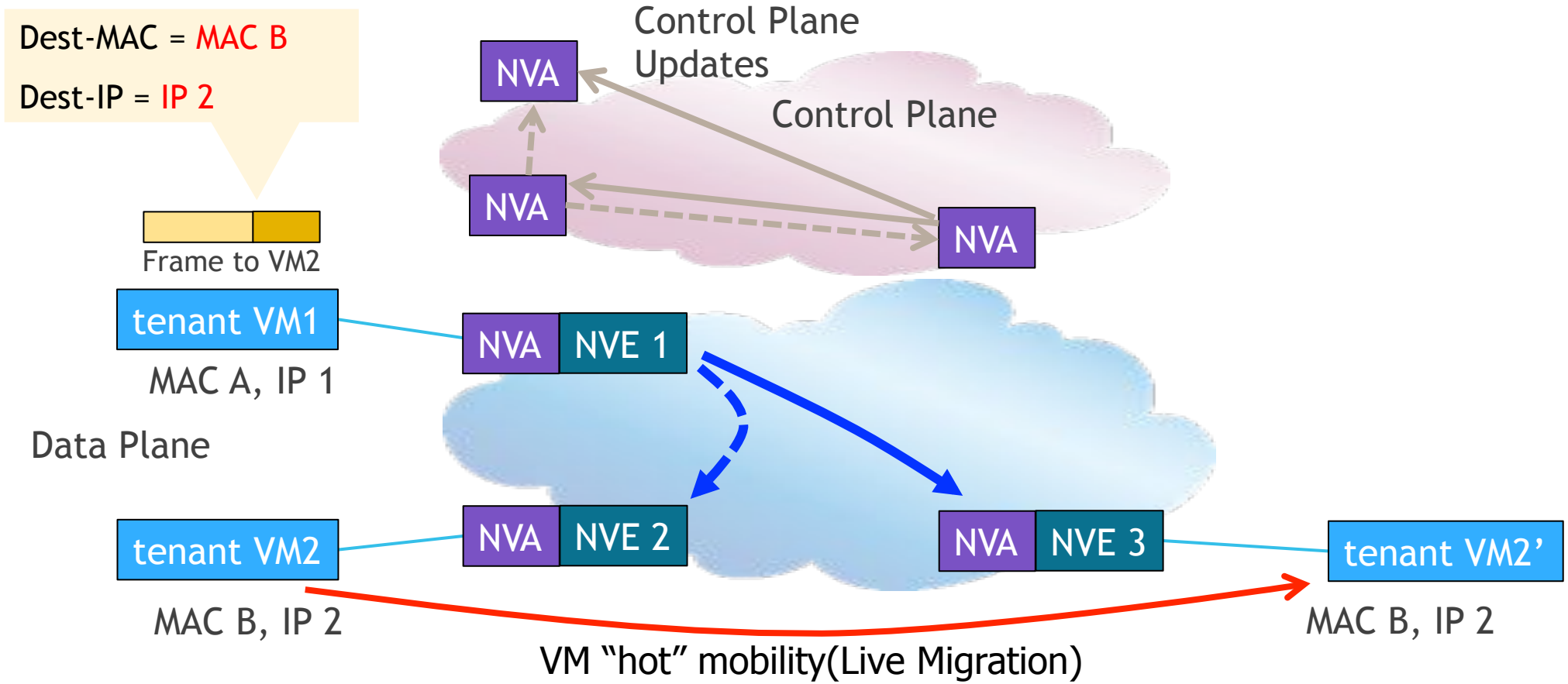
NVAにコントロールプレーンを実装

Hybrid型



NVA(Network Virtualization Authority):NVEにアドレスマッピング等の転送に必要な情報を与える外部エージェント。

VM MOBILITYへの対応



L2: MACアドレスの継続性、L2 Table更新

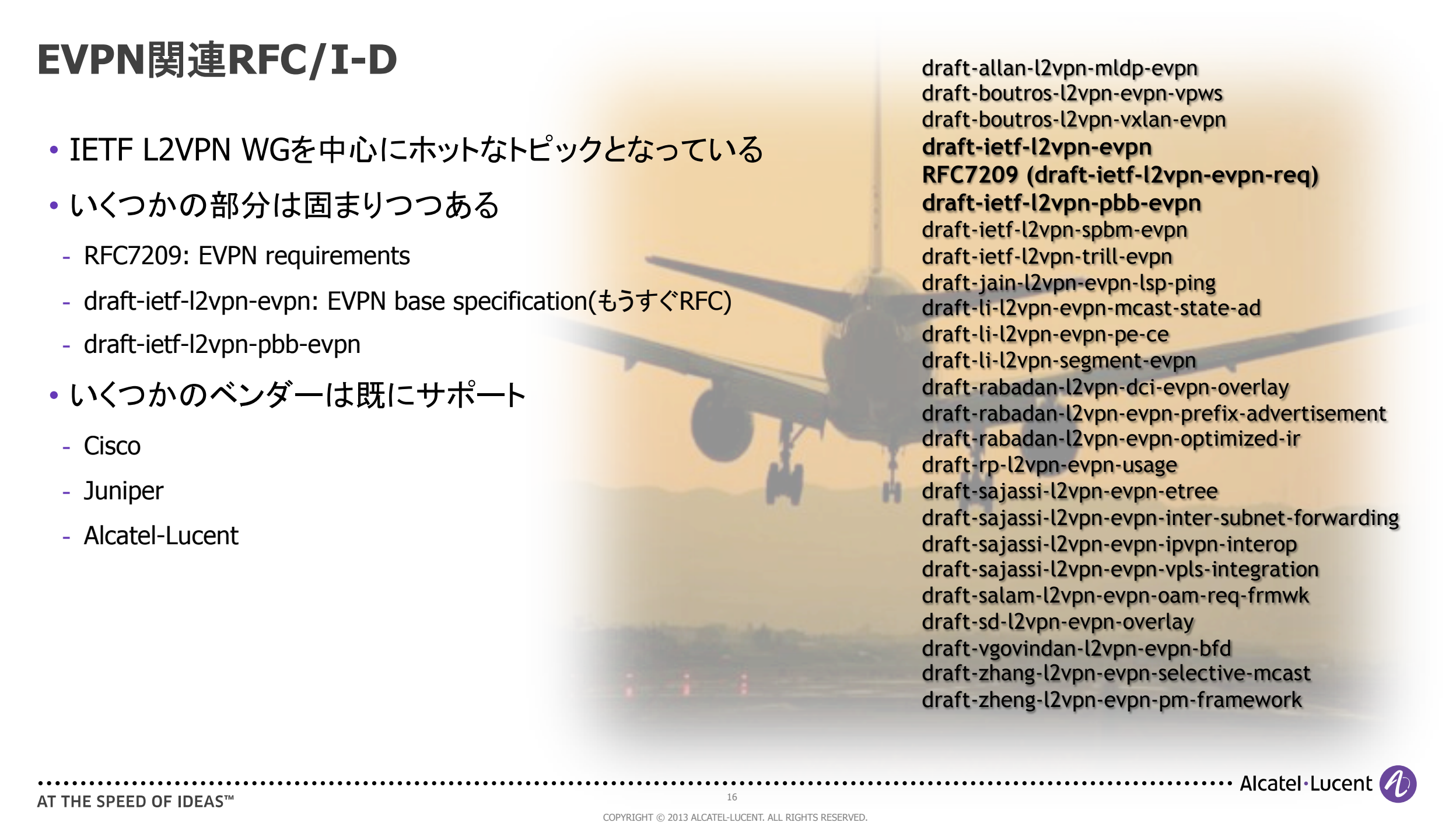
L3: IPアドレスの継続性、L3 Table更新、ARP

L2 MACラーニングってどうしてる？

- ユーザフレームトリガ(普通のL2NWと同様)
  - ARP/ND等BUMの処理
    - Underlay Multicastを利用 = UnderlayでPIM等Multicast Protocol動作要
    - 頑張ってIngress Replication = 量によっては負荷に。
- CMS and/or Hypervisorと連携し教えてもらう(or 検知する)
- L2の情報を運べるコントロールプレーンプロトコルを使用 (ex. EVPN)

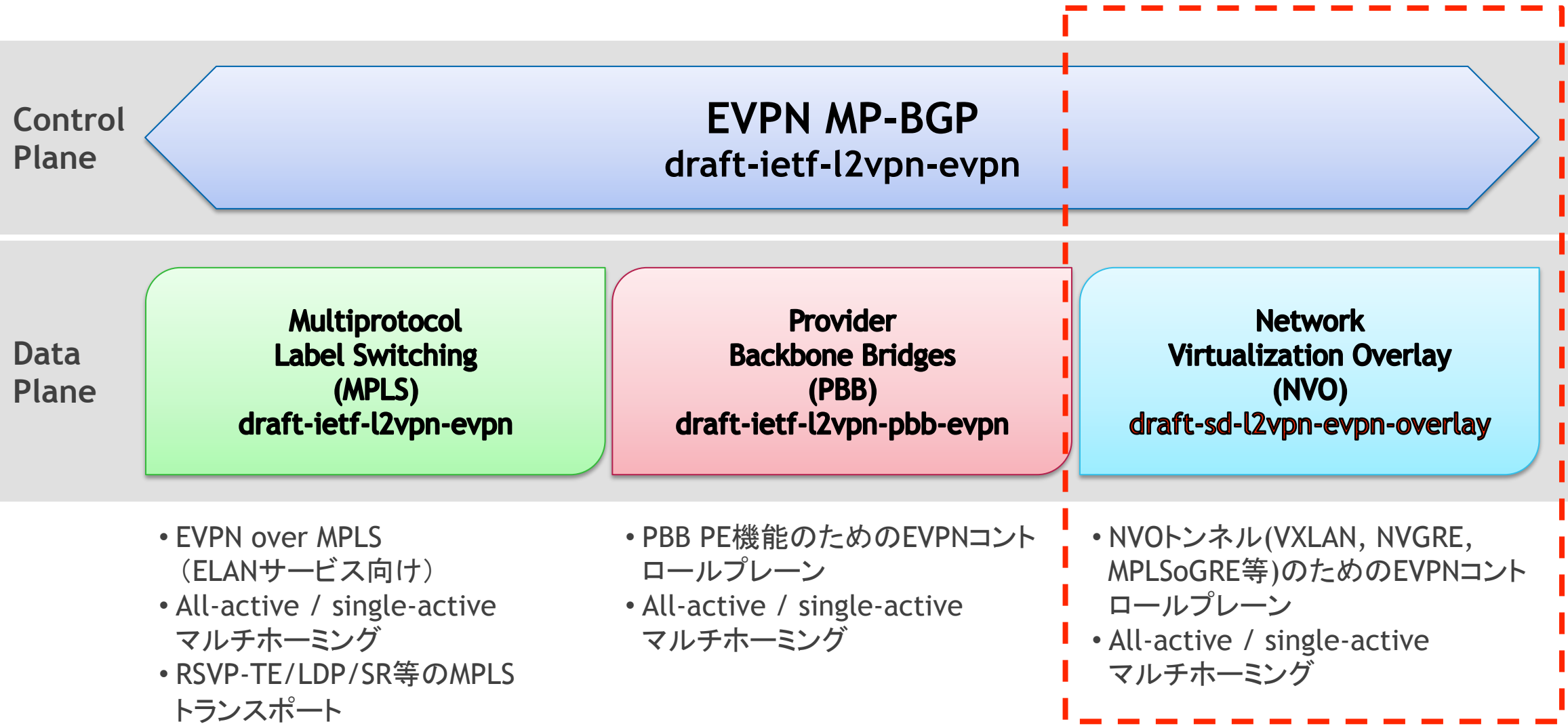
EVPN関連RFC/I-D

- IETF L2VPN WGを中心にホットなトピックとなっている
- いくつかの部分は固まりつつある
  - RFC7209: EVPN requirements
  - draft-ietf-l2vpn-evpn: EVPN base specification(もうすぐRFC)
  - draft-ietf-l2vpn-pbb-evpn
- いくつかのベンダーは既にサポート
  - Cisco
  - Juniper
  - Alcatel-Lucent



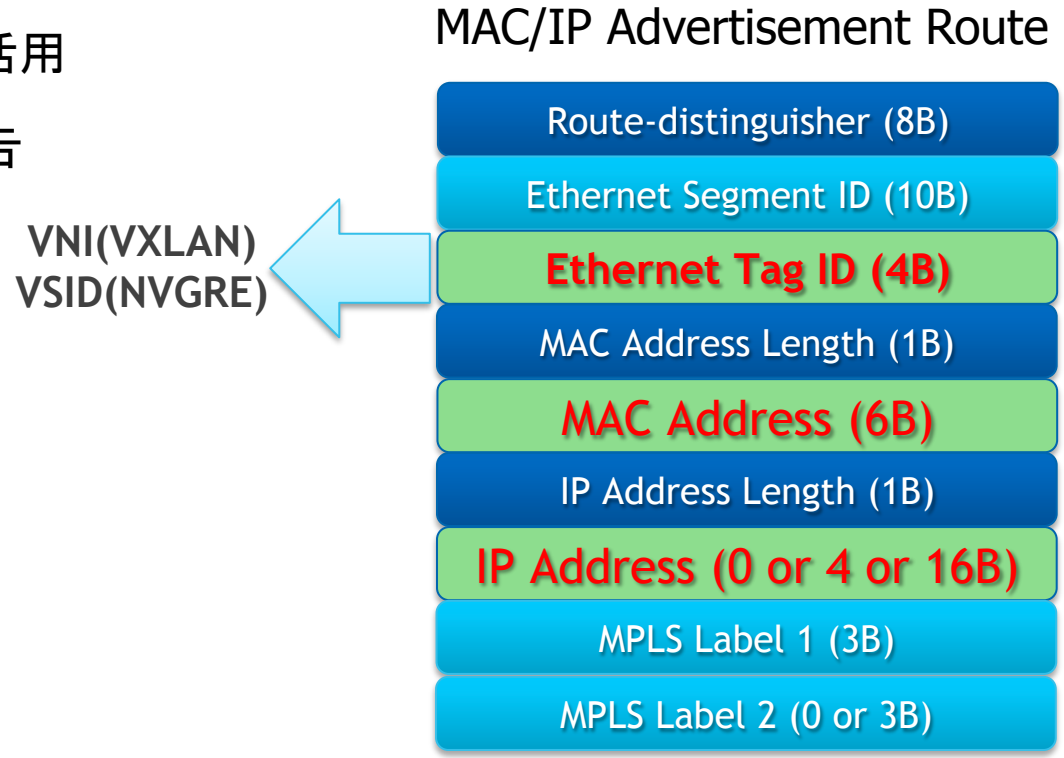
draft-allan-l2vpn-mldp-evpn
draft-boutros-l2vpn-evpn-vpws
draft-boutros-l2vpn-vxlan-evpn
draft-ietf-l2vpn-evpn
RFC7209 (draft-ietf-l2vpn-evpn-req)
draft-ietf-l2vpn-pbb-evpn
draft-ietf-l2vpn-spbm-evpn
draft-ietf-l2vpn-trill-evpn
draft-jain-l2vpn-evpn-lsp-ping
draft-li-l2vpn-evpn-mcast-state-ad
draft-li-l2vpn-evpn-pe-ce
draft-li-l2vpn-segment-evpn
draft-rabadan-l2vpn-dci-evpn-overlay
draft-rabadan-l2vpn-evpn-prefix-advertisement
draft-rabadan-l2vpn-evpn-optimized-ir
draft-rp-l2vpn-evpn-usage
draft-sajassi-l2vpn-evpn-etree
draft-sajassi-l2vpn-evpn-inter-subnet-forwarding
draft-sajassi-l2vpn-evpn-ipvpn-interop
draft-sajassi-l2vpn-evpn-vpls-integration
draft-salam-l2vpn-evpn-oam-req-frmwk
draft-sd-l2vpn-evpn-overlay
draft-vgovindan-l2vpn-evpn-bfd
draft-zhang-l2vpn-evpn-selective-mcast
draft-zheng-l2vpn-evpn-pm-framework

EVPN：コントロールプレーンとデータプレーンの分離
用途に応じて各データプレーンと動作可能

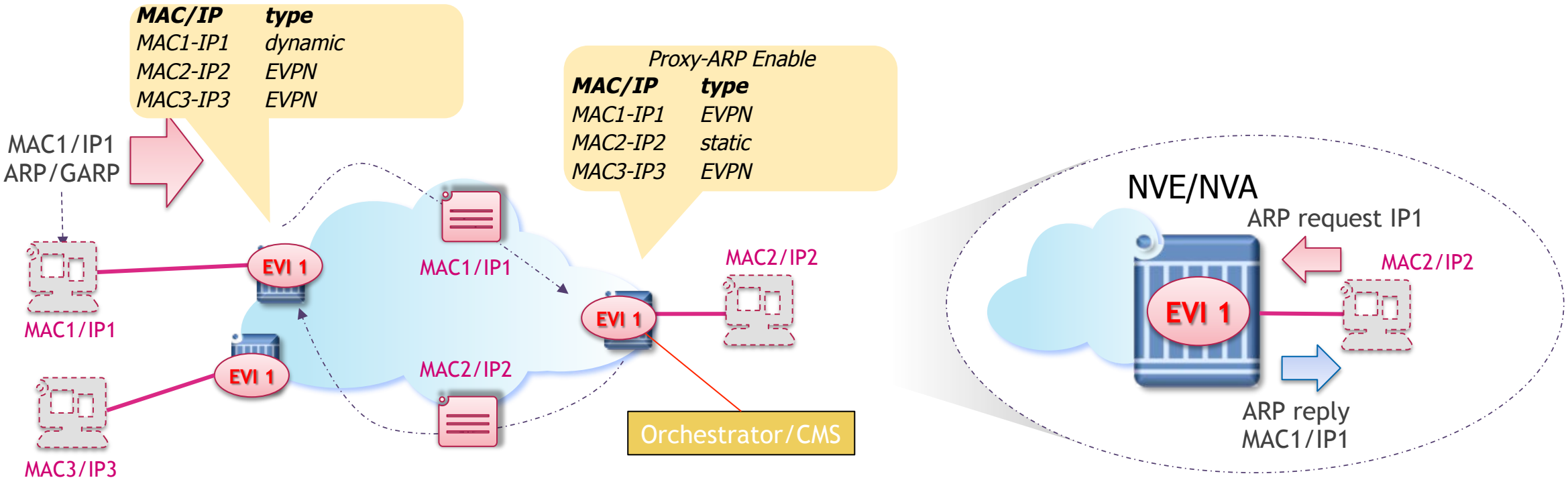


EVPN：MACアドレスの情報をMP-BGPで広告

- スケーラブルなBGPの仕組みをMACルートの広告に応用
  - Flooding の回避
  - L2にもコントロールプレーンを。RRや様々なBGP機能を活用
- MACアドレスとIPアドレスの情報をEVPN NLRIで広告
  - AFI = 25 (L2VPN) / SAFI = 70 (EVPN)
- MACラーニングをコントロールプレーンで行う
  - ラーニングポリシー等も適用可能
- EVPN L2インスタンス(EVI)間のIsolation



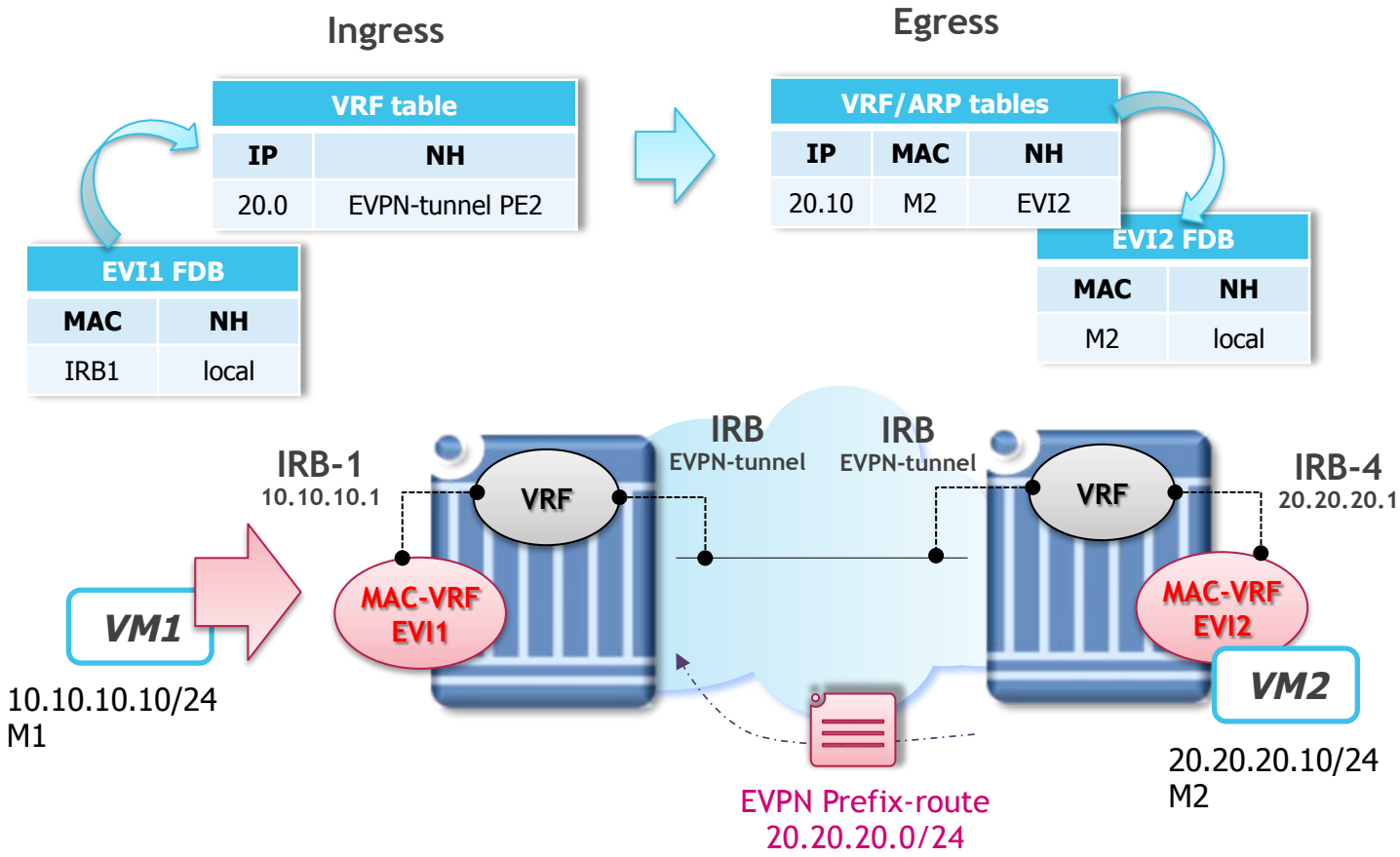
EVPN: Proxy-ARP/NDによるFlooding抑止



Dynamic(ARP/NDトリガ)、又はStatic(CMSやポリシーサーバ連携)で学んだLocalルートを他のControl Plane ノードにEVPNにより広告

右側のNVE/NVAは既にEVPNによりMAC1/IP1の情報を学習しているため、LocalホストからIP1のARPが来た場合にARPに答えられる。

EVPN: L2/L3 forwardingへの拡張
draft-rabadan-l2vpn-evpn-prefix-advertisement



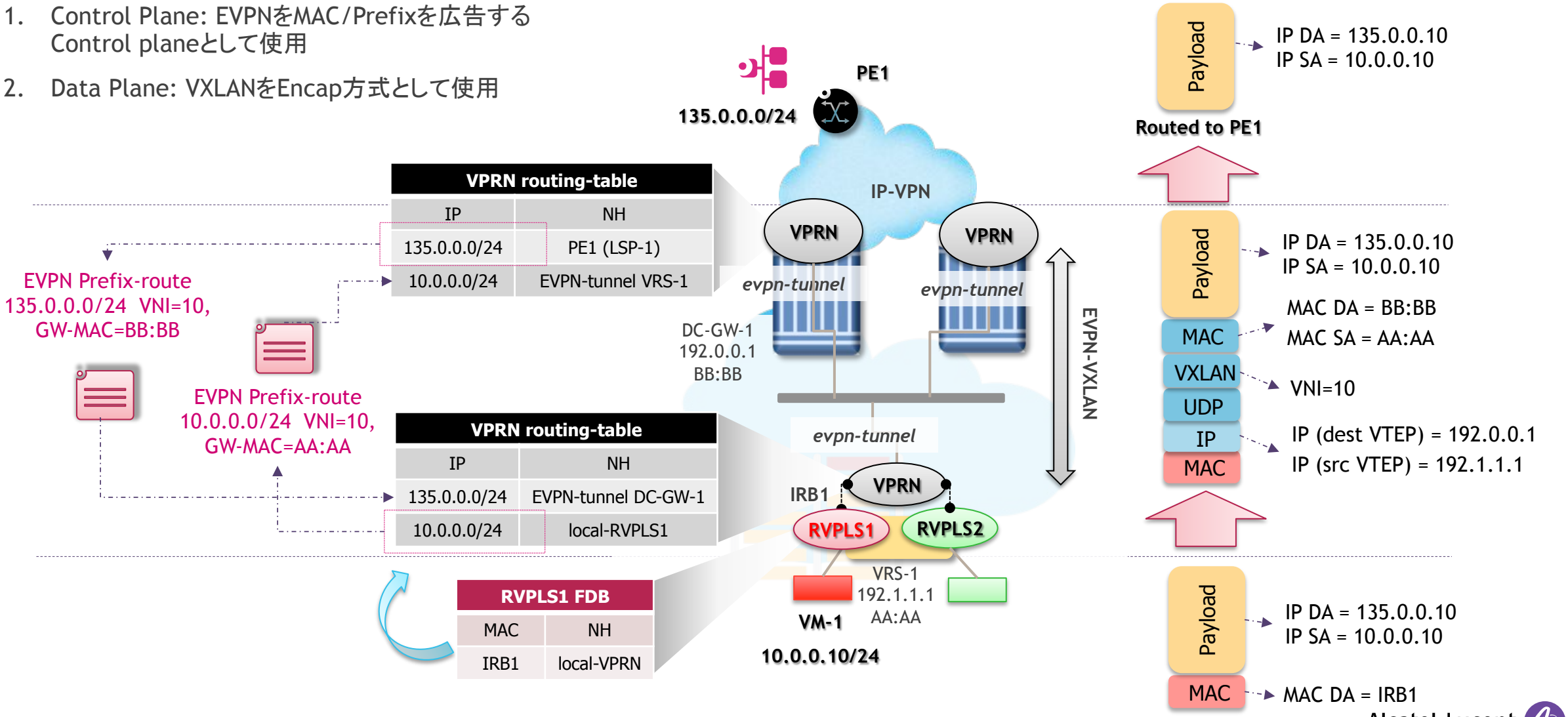
- EVPNでPrefix-routeを広告する拡張
- EVIはサブネット内にホストが居るところにのみ存在する
- PE上のVRFはLocalのMAC-VRFとEVPN-Tunnel向けにIRB(Integrated Routing and Bridging) インタフェースを持つ
- Localに居る以外のEVIについてはRemote PEからのHost MAC/IPはimportしない
- EVPNによりVRFのrouting tableからimportされたprefixも広告可能

EVPNをcontrol planeとして使用し、IP-VRF内のsubnet間ルーティングが可能となる:

- ingress PE
  - FDB lookupによりIRBインタフェースへ
  - Routing lookupにより remote PEへ
- egress PE
  - Routing/ARP lookupによりlocal EVIへ
  - FDB lookupによりlocal ACへ

EVPN: DC-GWとのVXLAN-EVPNによる接続 ALUの実装例

1. Control Plane: EVPNをMAC/Prefixを広告する Control planeとして使用
2. Data Plane: VXLANをEncap方式として使用

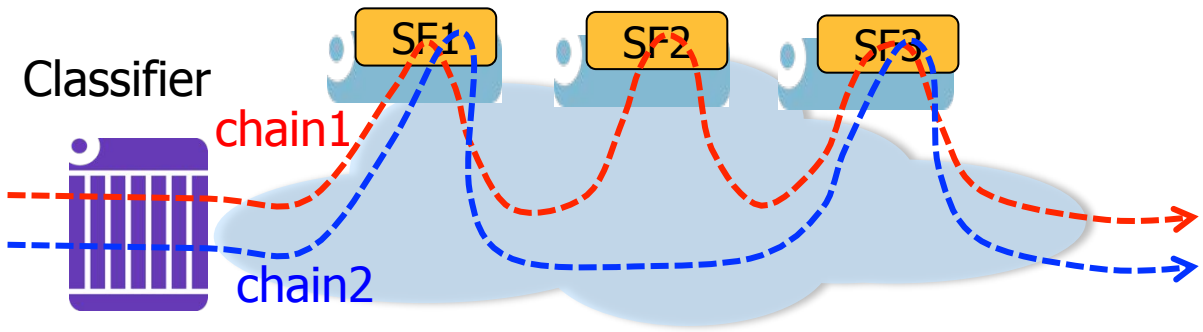


IETF SFC WG : Service Function Chaining

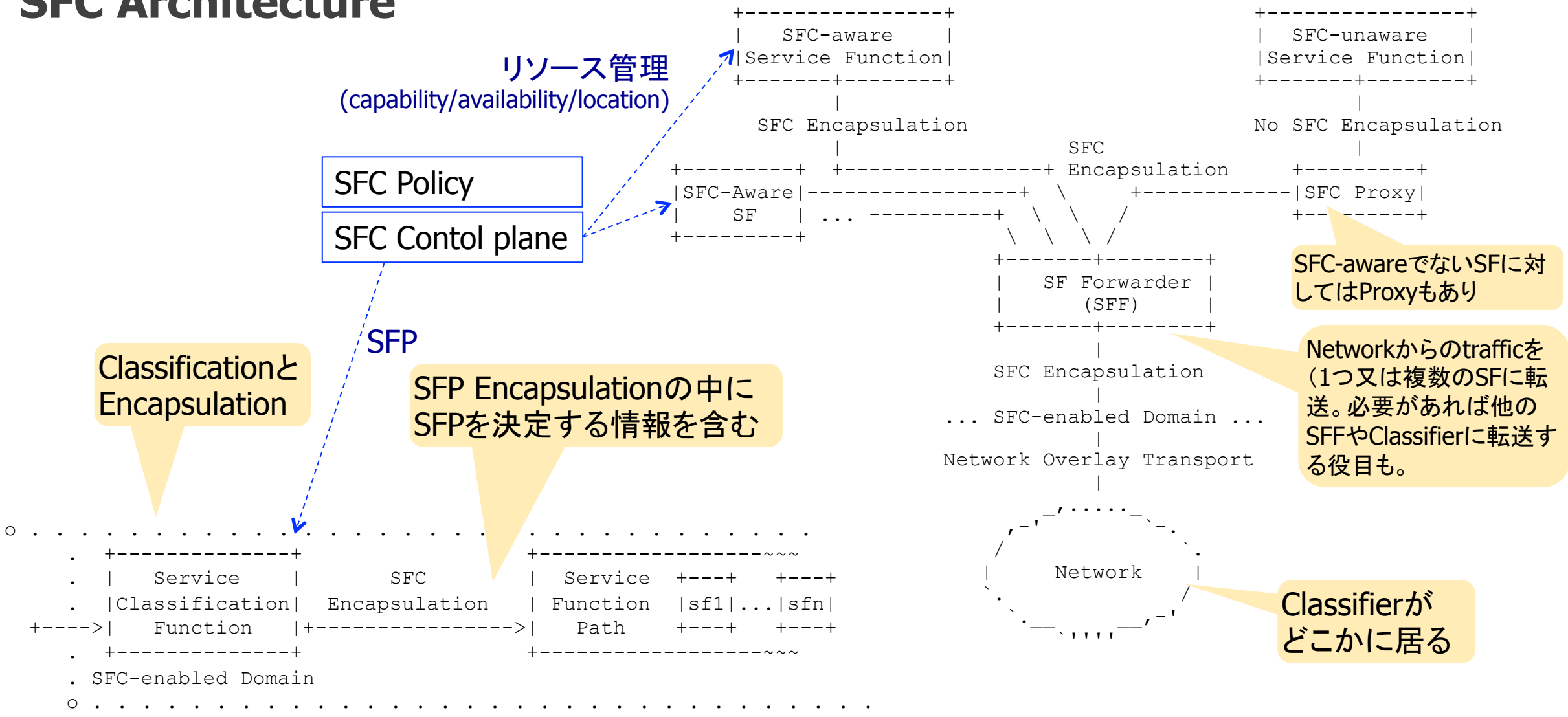
http://datatracker.ietf.org/wg/sfc/charter/

物理/仮想で提供される複数のService Functionをデータパス上に繋げて行く(Chainさせる)ためのアーキテクチャとEncapsulation,必要となるControl planeの確立を目指す。

- draft-ietf-sfc-problem-statement (submitted to IESG)
- draft-ietf-sfc-architecture
- draft-ietf-sfc-dc-use-cases
- draft-ietf-sfc-long-lived-flow-use-cases
- draft-ietf-sfc-use-case-mobility



SFC Architecture



SFC Encapsulation

異なる2つの提案方式

- draft-quinn-sfc-nsh (NSH：Network Service Header)
  - Mandatoryな固定長のContext header
  - OVS data plane及びOpenDaylight control planeでの実装が始まっている
- dra3-zhang-sfc-sch (SCH：Service Chain Header)
  - 可変長のContext header

マージの可能性に向けた議論

まだ固まるまでには多少時間が必要か?

まとめ

<NVO3>

- データプレーンについてはかなり多くの実装が出ており、Commercial Deploymentも多くなってきている。
- コントロールプレーンはまだ選択肢が色々あるが、EVPN関連はかなり活発な提案が進んでいる。
- L2/L3 combinedサービスの実装も。
- かなり広くインプリされたVXLANを、将来Geneveが取って代わる時が来るのか？

<SFC>

- Architecture / Usecaseは固まってきた。
- やはりヘッダ問題の方向性が決着しないと。。

www.alcatel-lucent.com