

# マルチベンダ環境における EVPN構築のノウハウ

～Interop Tokyo 2016 ShowNetでの  
相互接続検証を元に～

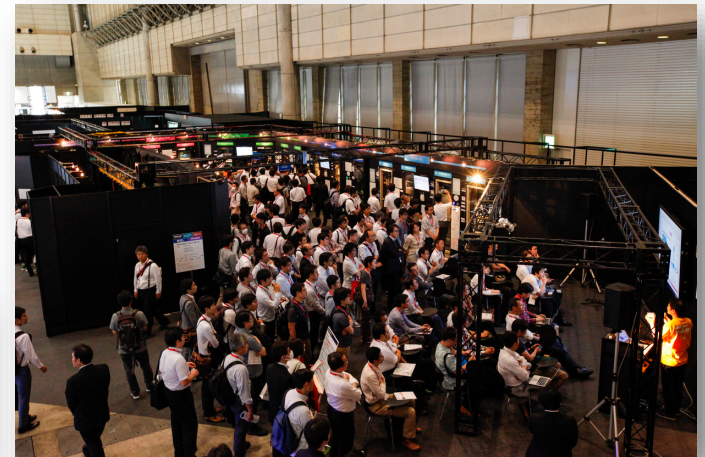
Interop Tokyo 2016  
ShowNet NOCチームメンバー 大久保 修一

## 免責事項

- 本発表では、個別製品の実装状況についてのご説明、およびご質問への回答ができません。実装の詳細については、改めてご自身での動作確認をお願いいたします。
- 本発表は参考情報の提供を目的としております。発表者、Interop Tokyoの主催者および運営事務局、ならびにInterop Tokyo ShowNet NOCチームでは、各製品導入に関する結果責任を負いかねます。導入に関する最終決定は利用者ご自身の判断において行われるようお願いいたします。
- 本発表では、わかりやすさを優先するために、一部詳細説明を省いております。正確な仕様については、末尾参考文献中のRFCやInternet Draft原典を参照ください。
- 記載内容は検証当時の結果となります。現時点、および将来異なる可能性がありますので、ご注意ください。

# Interop Tokyo

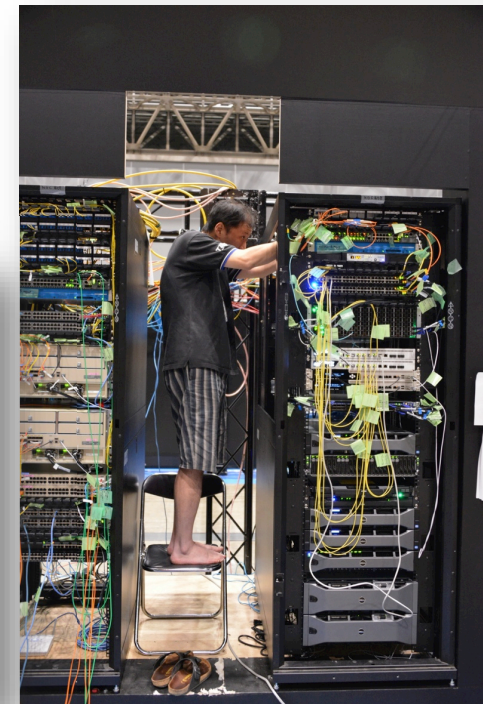
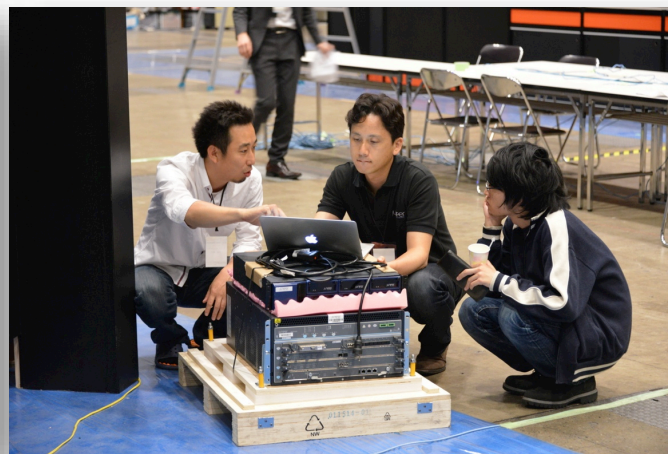
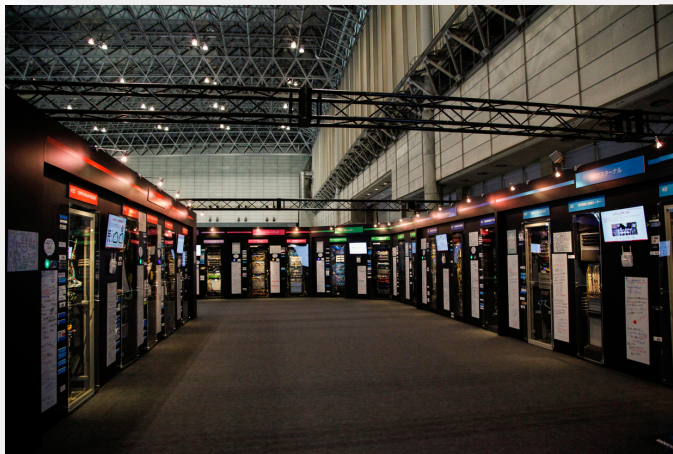
- 世界最大のネットワーク機器と技術の展示会
  - 2017年は第24回
  - 毎年6月に幕張メッセで開催
  - 来場者数約14万人



# .# show $\Delta$ net ←

Scratch and Re-build the Internet

- “I know it works because I saw it at Interop”
  - 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
  - 最新の技術で10年先のインターネットを構築する
  - 様々な技術の相互接続生検証の場
  - 出展者や来場者へのネットワーク提供





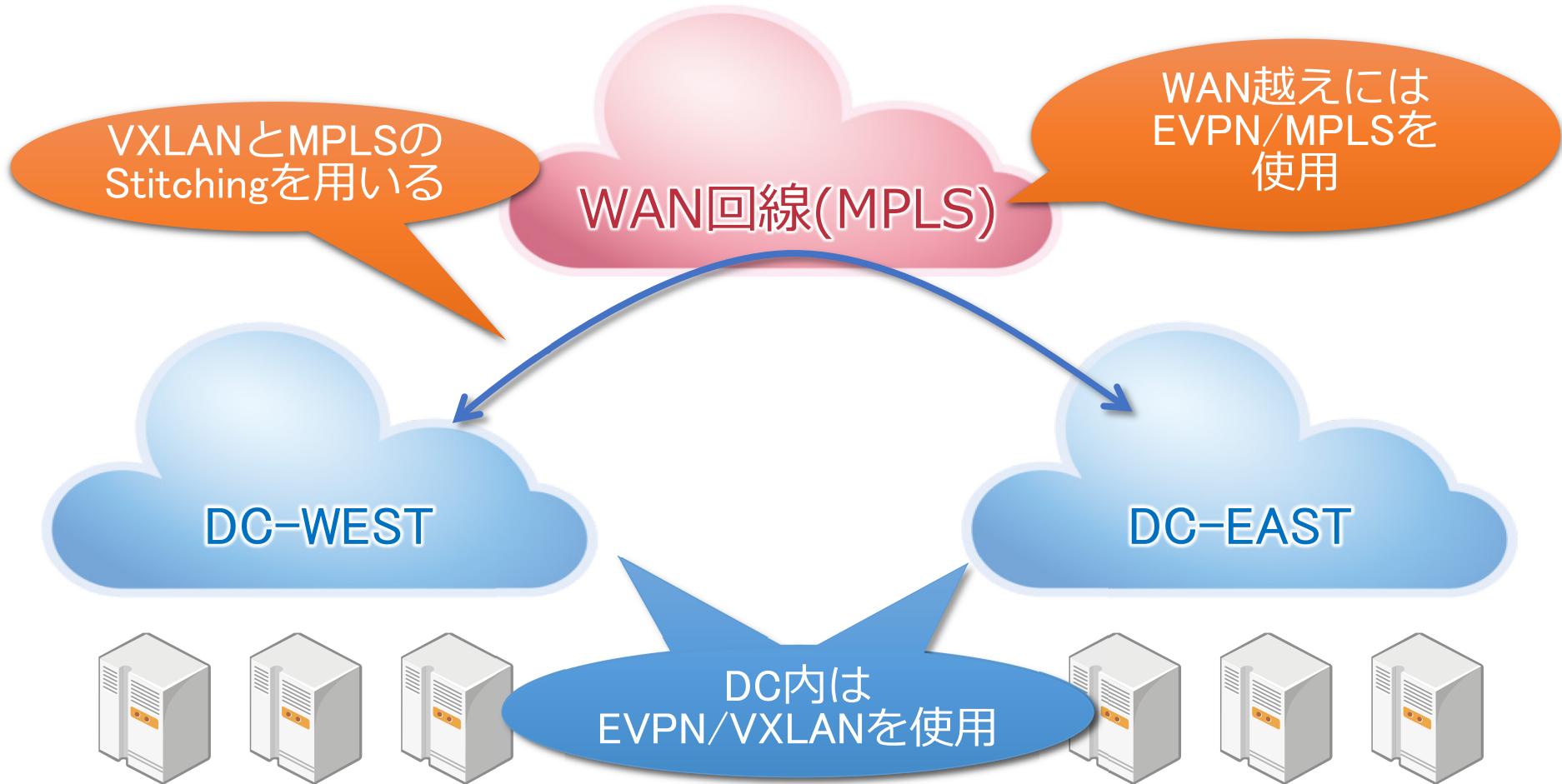


# EVPNの相互接続検証



# EVPN相互接続検証ネットワークモデル

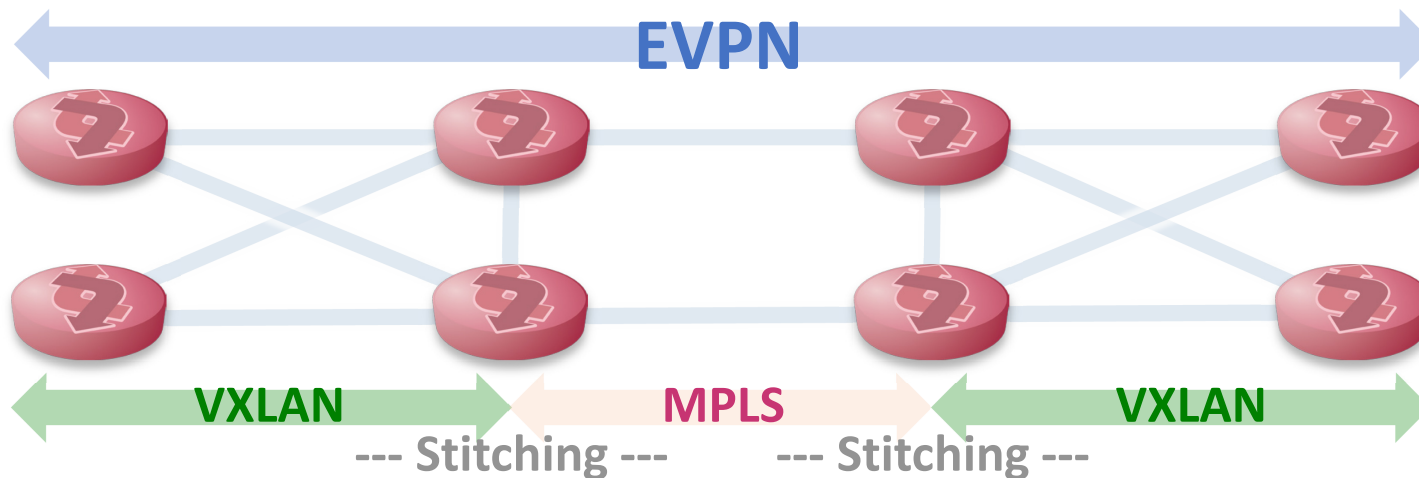
離れた複数のデータセンター間をMPLS WAN回線を通じて、共通のコントロールプレーン(EVPN)を用いてLayer2のVPNを提供する





# 検証コンセプト

- マルチベンダ間でEVPN相互接続検証を実施
- データプレーンは異なるカプセル化技術を適用
- データプレーン間を”ステッチング(縫い合わせ)”し, 更なる閉域網延伸を試みる検証を実施



## 検証実施概要

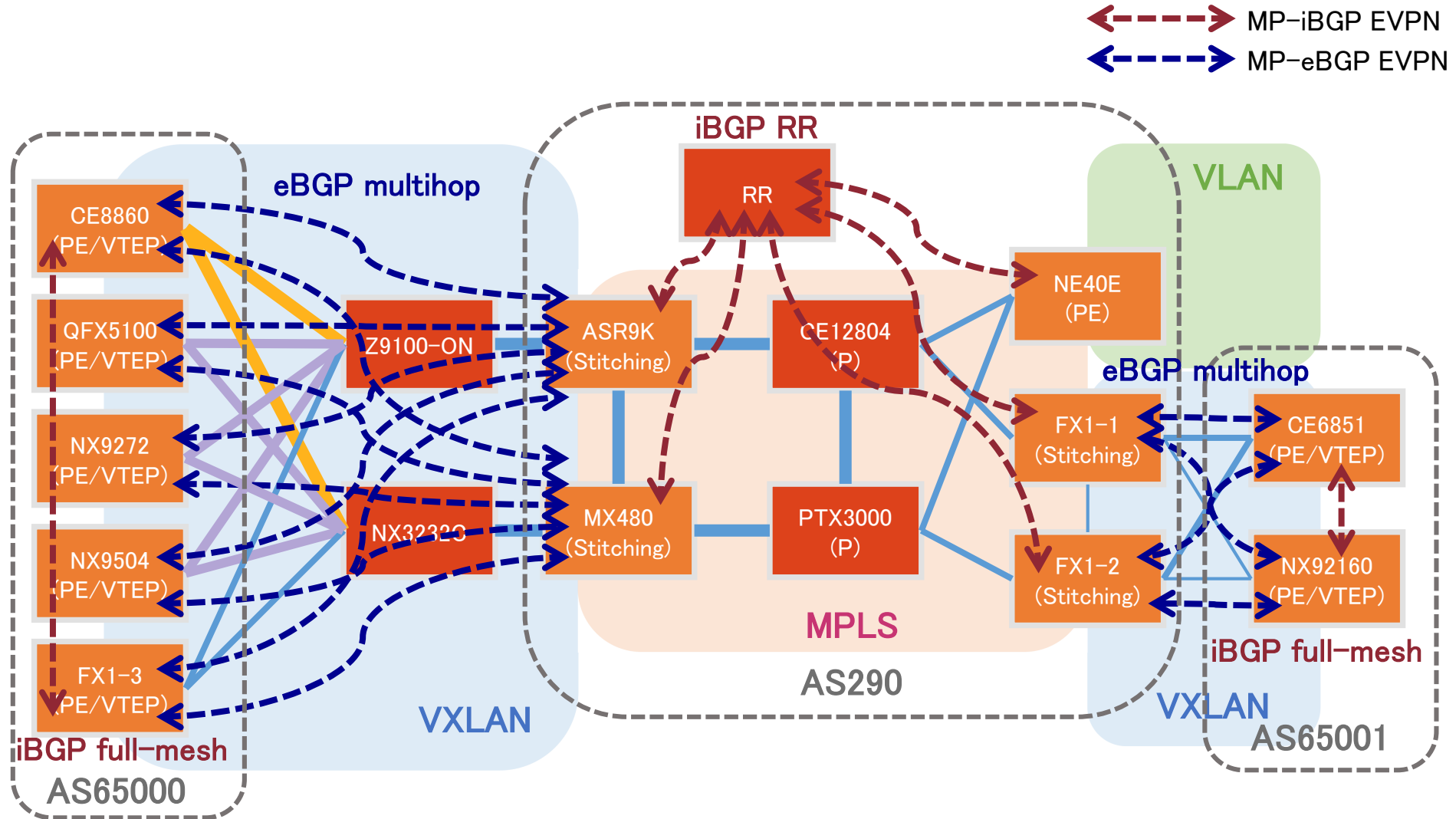
- 日程  
2016年5月27日～2016年6月10日
- 場所  
Interop Tokyo 2016 ShowNet ホットステージ会場  
(幕張メッセ)
- 参加企業(五十音順)  
シスコシステムズ様  
ジュニパーネットワークス様  
ファーウェイ・ジャパン様  
古河ネットワークソリューション様

ご協力いただいた  
皆さま、ありが  
うございました。

# 検証参加装置

EVPN/MPLS EVPN/VXLAN Stitching	Cisco ASR9006
	Juniper MX240
	Furukawa Electric FITELnet FX1
EVPN/MPLS	Huawei NE40E
EVPN/VXLAN	Cisco Nexus 9504
	Cisco Nexus 9272
	Cisco Nexus 92160
	Furukawa Electric FITELnet FX1
	Juniper QFX5100
	Huawei CE8860
	Huawei CE6851

# 検証構成



# 検証項目

STEP	内容
1	AS290 MPLS LSP
2	AS65001 iBGP VXLAN
3	AS65000 iBGP VXLAN
4	AS290 iBGP MPLS
5	AS290 マルチホーミング
6	AS290 RR運用に切替え
7	Stitching(AS290+AS65001)
8	Stitching(AS290+AS65000)
9	AS65000 マルチホーミング
10	Stitching(AS65000+AS290+AS65001)
11	Stitching + マルチホーミング(AS65001+AS290)



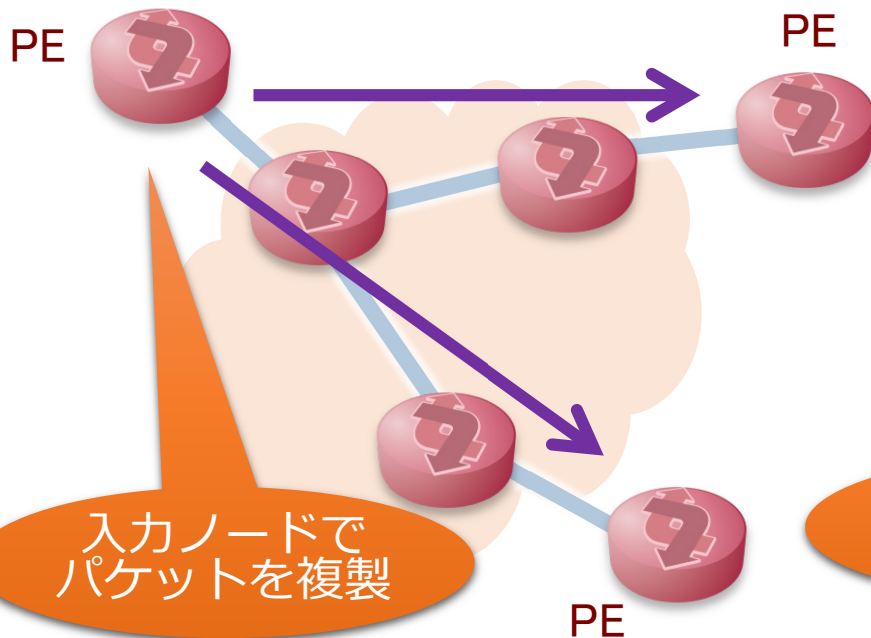
# マルチベンダ環境における EVPN実装確認ポイント

# 1. VXLAN BUM転送方式

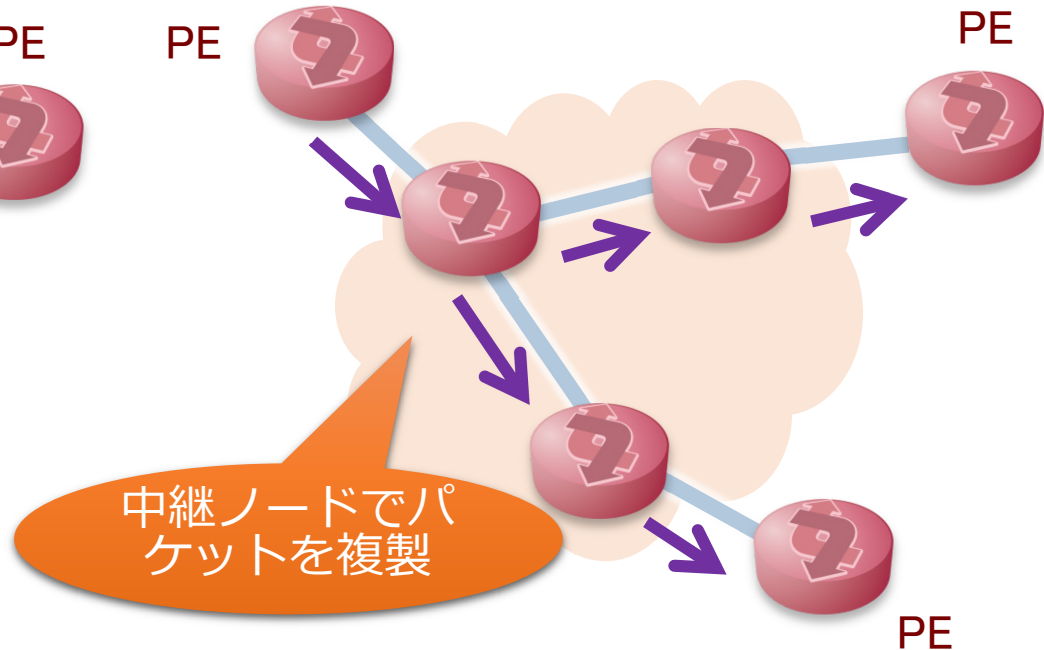
# VXLAN BUM転送方式

- BUM = Broadcast, Unknown unicast, Multicast
- オリジナルのVXLANはIPマルチキャスト(商用環境での導入のハードル)
- EVPNではingress replicationが定義 → 本検証で使用

## ingress replication



## IP multicast





# VXLAN BUM転送方式

- 本検証では相互接続性の問題は見つからなかった
- 実装によってはreplication先ノード数に制限があるため要確認
- 補足：本検証におけるIPファブリックは、ジャンボフレームが通せる環境を構築

## 2. Service Interface Type種別

# Service Interface Type種別

- 本検証ではVLAN-BasedとVLAN-Aware Bundleの2種類で行った
- VLAN-Based
  - 1つのEVIにつき1つのブロードキャストドメイン
- VLAN-Aware Bundle
  - 1つのEVIにつき複数のブロードキャストドメイン
- 対象機器の実装状況はまちまちで、この2種類の両方、もしくはどちらかであった

# VLAN-BasedとVLAN-Awareの違い

## VXLANの場合

### VLAN-Based

BD毎にRD,RTのアサインが必要

EVI RD:10.0.0.1:1 RT:290:100	EVI RD:10.0.0.1:2 RT:290:200
BD VNI:11	BD VNI:12

Label = VNIとする  
E-TagID = 0とする

### VLAN-Aware Bundle

1つのRD,RTで複数のBDを扱える

EVI RD:10.0.0.2:1 RT:290:300	
BD VNI:21	BD VNI:22

Label = VNIとする  
E-TagID = VNIとする

E-TagIDの違い

EVI = EVPN Instance  
E-TagID = Ethernet Tag ID  
RD = Route Distinguisher

BD = Broadcast Domain  
VNI = VXLAN Virtual Network Identifier  
RT = Route Target Community

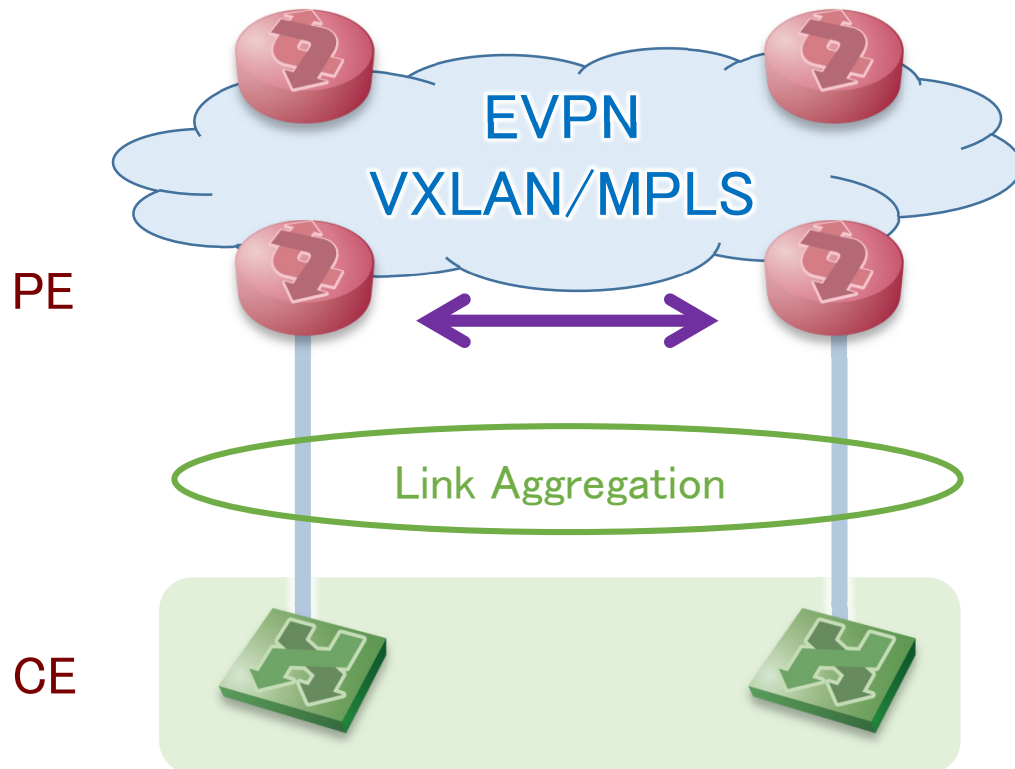
# Service Interfaceの相互接続性

- 基本的に、VLAN-BasedとVLAN-Awareでは接続できない。
- ただし、VLAN-Awareの装置が生成するE-TagIDが0でない経路を、VLAN-Basedの装置が受信するものと無視するものがあり、前者では一部条件下において、通信可能なパターンが存在した。  
(もちろん商用環境ではおすすりめしません)

## 3. マルチホーミング機能

# EVPNマルチホーミング機能

- マルチホーミングは、簡単に言うとEVPNをコントロールプレーンに使用してマルチシャーシLAGを組めるようになるための仕組み



EVPNによるMACテーブルの同期

LAGの対向装置は、1台でも、プロプライエタリなVirtual ChassisでもOK

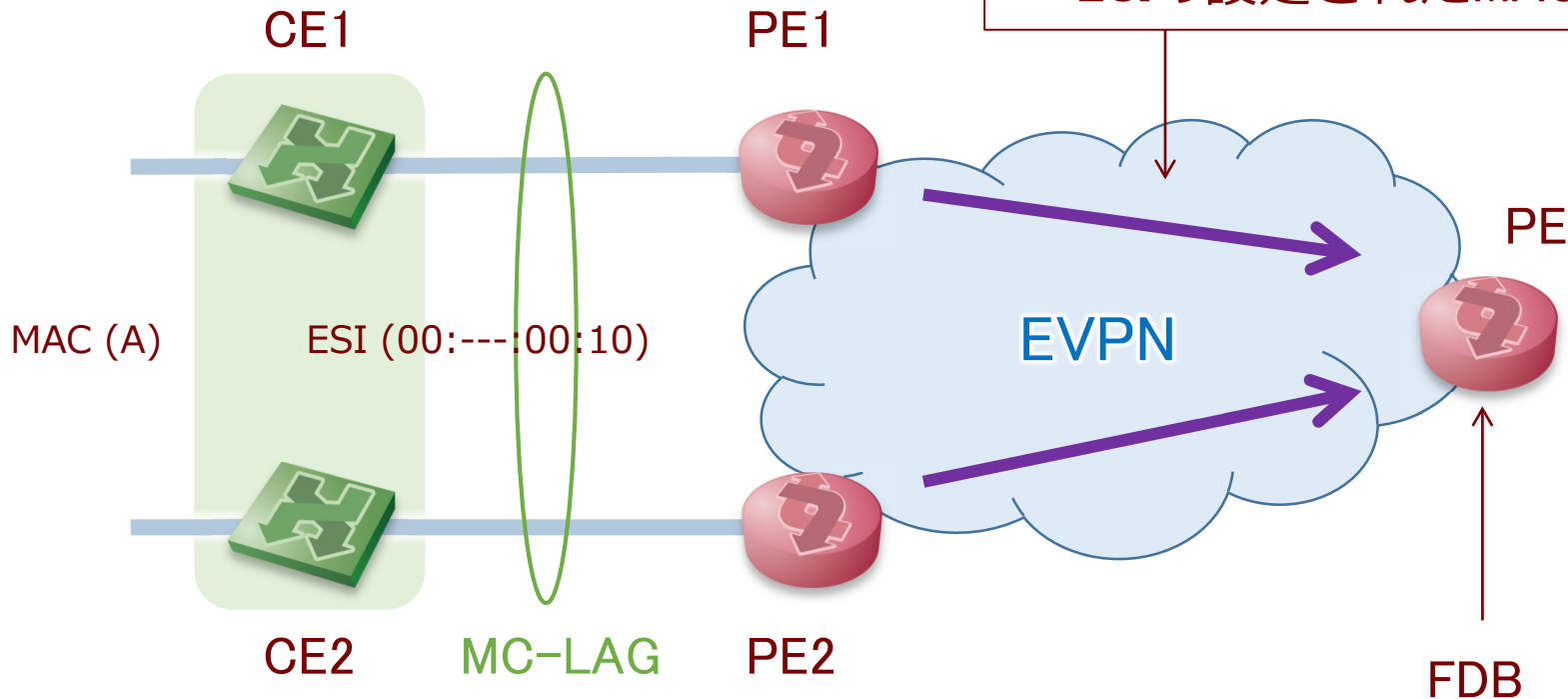
# マルチホーミング機能の相互接続性

- マルチホーミング機能を実装済、未実装の機器が存在
- また、VLAN-BasedとAwareの差異により、相互接続可能な組み合わせが1パターンのみであったが、マルチベンダでの相互接続に成功
- 未実装の機器を混在するとそもそも通信ができなかったり、BUMが重複して届くなどの不安定な挙動が発生した(次ページにて補足)
- LAGの方式は実装がまちまちなので要確認
  - Static LAGのもの
  - LACPを使用するもの
  - どちらか選択可能なもの



# マルチホーミング機能の仕組み

ESIにて同一LAG内のリンクを識別



両方のPEが経路広告

- EAD/ES EAD/EVI IMET
- ESIの設定されたMAC Route

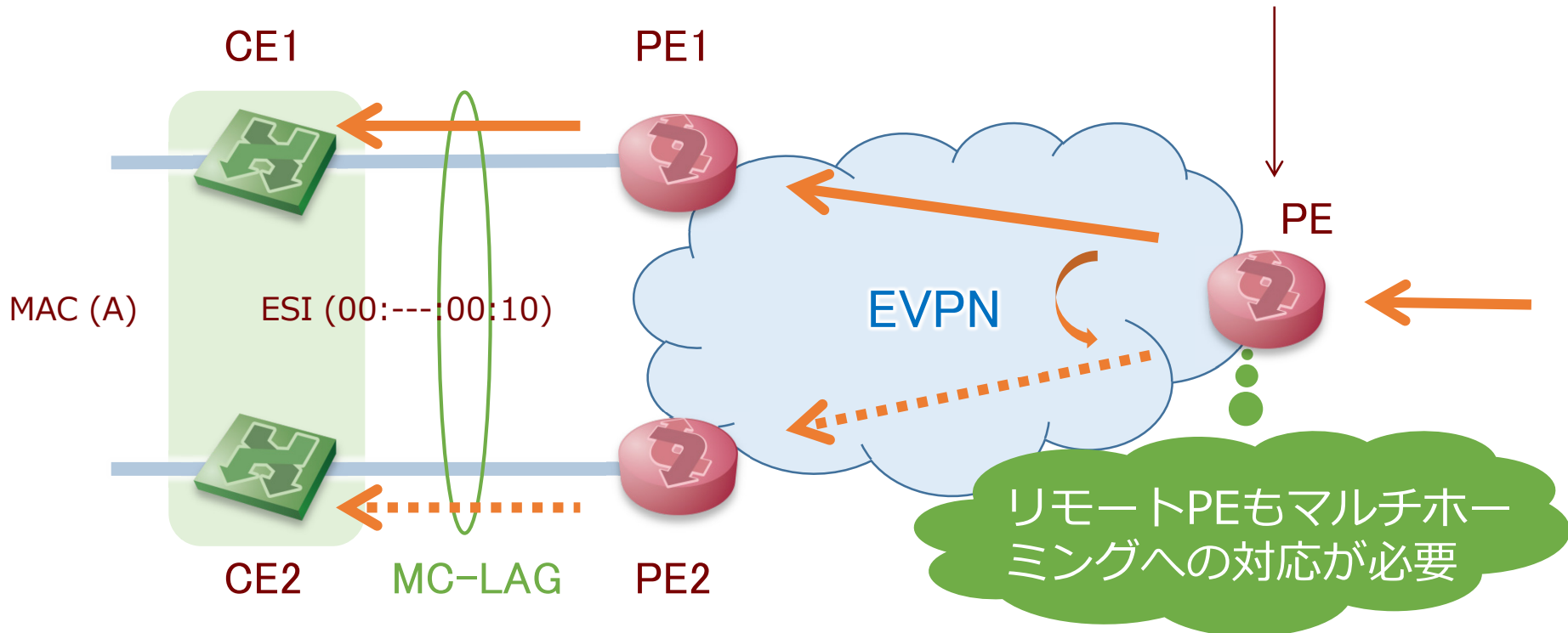
EAD/ES = Ethernet Auto-Discovery per ES  
EAD/EVI = Ethernet Auto-Discovery per EVI  
IMET = Inclusive Multicast Ethernet Tag

MAC (A) → ESI (00:---:00:10) → PE1,2  
※ MACアドレスをESIにマッピング

# マルチホーミング機能の仕組み

トラフィック送信

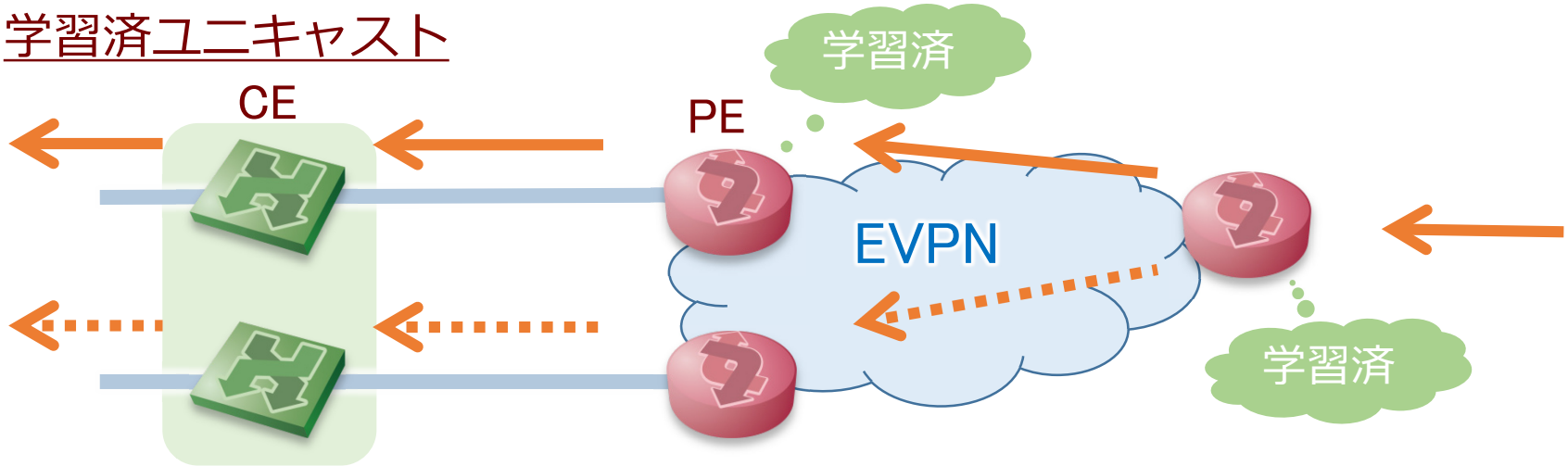
- 同じESIを持ったPEにロードバランス
- BUMは両方のPEに送付



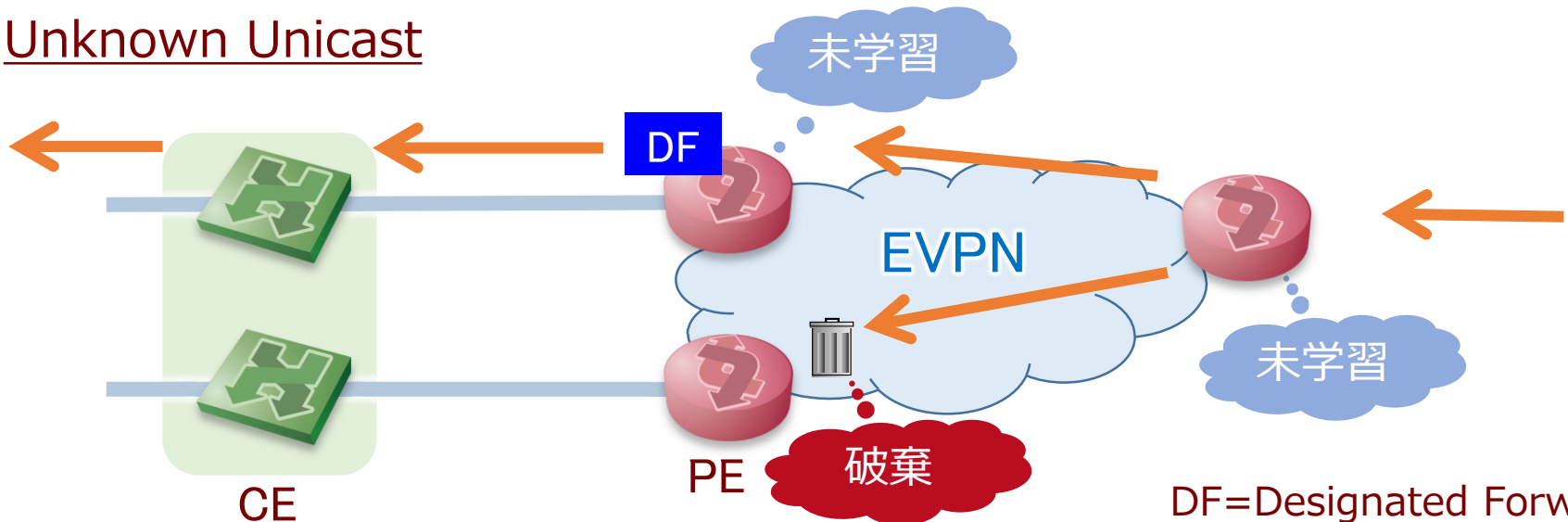
※ BUMはDesignated ForwarderであるPEのみがCEに転送

# 学習済/未学習宛先への配送方式の違い

## 学習済ユニキャスト



## Unknown Unicast

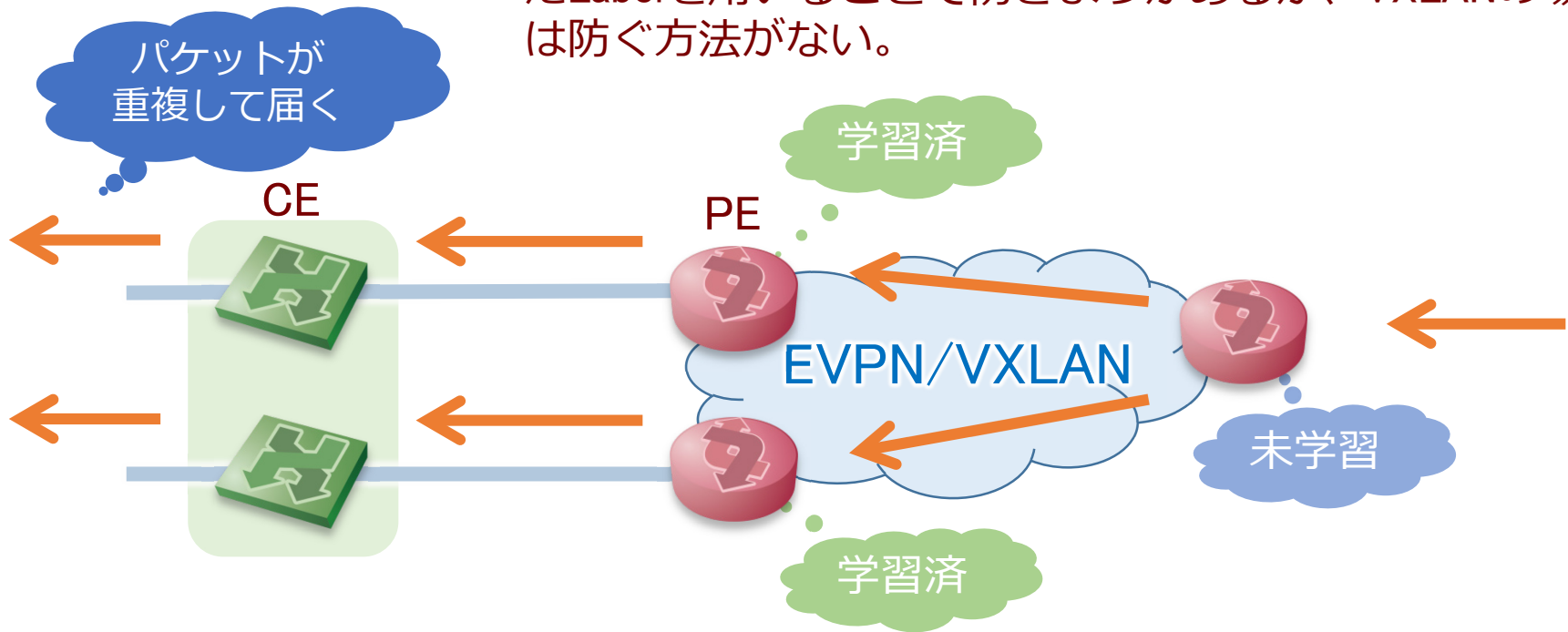


DF=Designated Forwarder

# MACテーブル不整合が生じると？

## Unknown Unicast

EVPN/MPLSの場合は、Ingress Replication用に割り当てられたLabelを用いることで防ぎようがあるが、VXLANの場合は防ぐ方法がない。



通常時、EVPNでは全てのMACアドレスを全ノードが同期するため問題ないが、clear mac等のコマンド実装不備でリモートMACアドレスが欠けたりすると、このような事象が発生する。

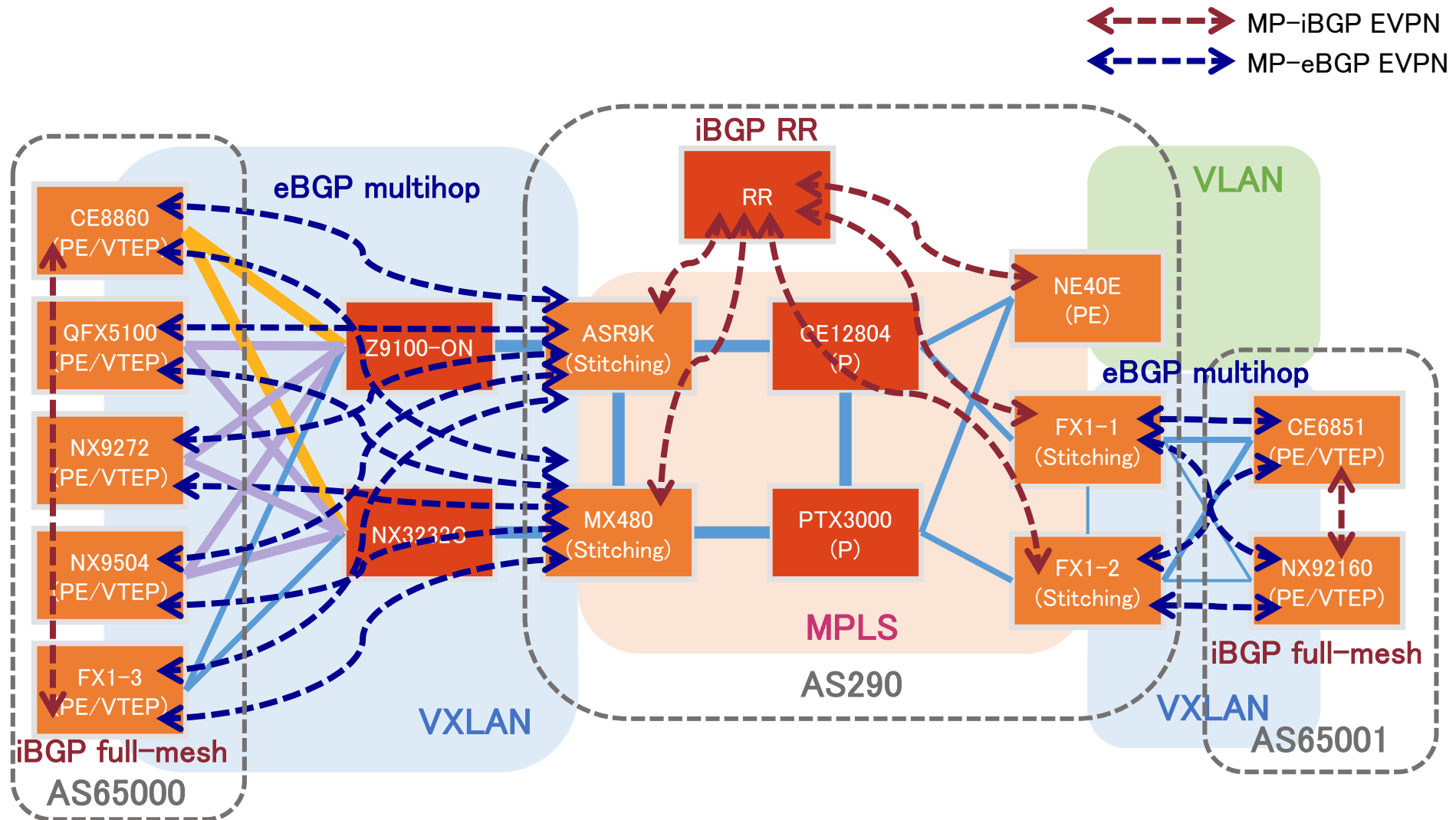
## 4. Stitching機能

# Stitching機能について

---

- Stitching = 縫い合わせ
- EVPN/VXLANとEVPN/MPLSのデータプレーンを変換する仕組み

# 検証構成(再掲)



# Stitching機能の動作確認結果

- 以下に挙げるような注意点はあったが、マルチベンダでの相互接続に成功
- EVPN/MPLSにおけるVLAN-Aware実装/未実装
  - 実装機と未実装機を混在すると疎通不良が発生。VLAN-Basedだと問題なし。
  - ※ Stitching検証の際に発見したissueだが、本来Stitching動作には関係ない
- EVPN/VXLAN側で受信したBGPメッセージ中の「encapsulation tunnel type=vxlan」が、EVPN/MPLS側のBGPピアにそのまま伝搬する問題
  - ある機種で発生したが、その場で修正いただき問題は解消した
- Stitching機能実装の違い(configの思想の違い)
  - メーカー1：一旦Bridge Domainで終端し橋渡しする方式
  - メーカー2：テーブルが共通化されている方式



# EVPN/MPLSでのVLAN-Based

- EVI = Broadcast Domain
- EVI毎にRD, RT, ラベルが異なる
- ラベルでBroadcast Domainを識別
- フレームフォーマット



Broadcast Domain毎に異なるラベル

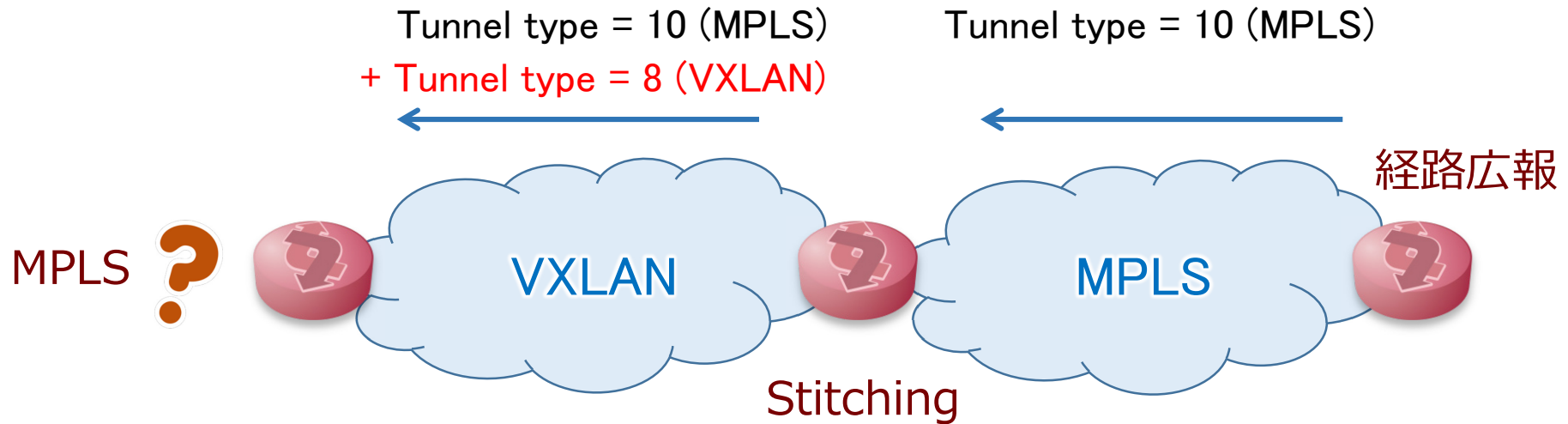
# EVPN/MPLSでのVLAN-Aware

- EVI = 複数Broadcast Domain
- 1種類のRD, RT, ラベル
- Broadcast Domainの識別は、
  - コントロールプレーン(BGP): E-TagID
  - データプレーン: 802.1qタグ (VLAN ID=E-TagID)
- フレームフォーマット



VLAN IDでBroadcast Domainの識別

# encapsulation tunnel typeの問題

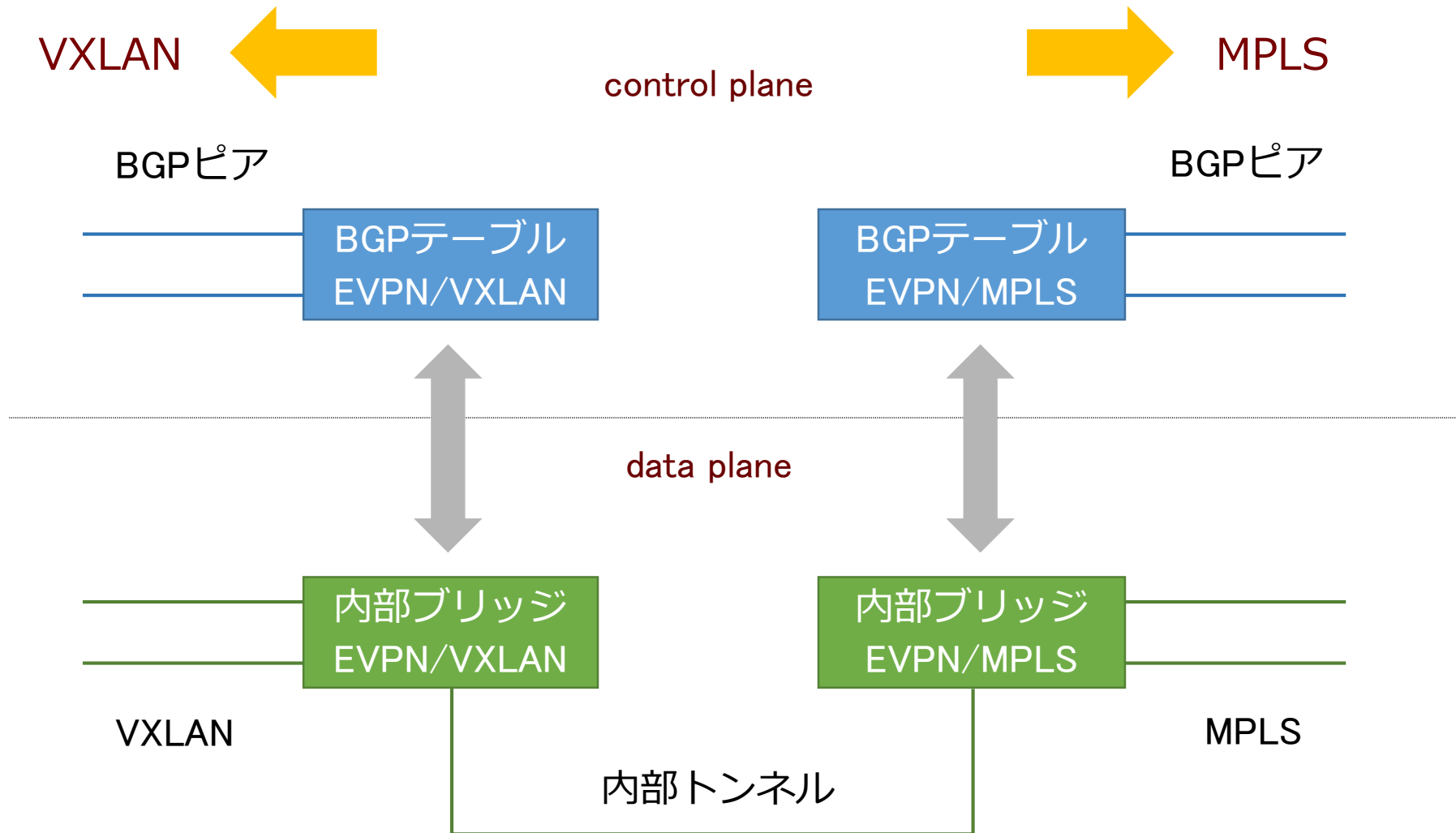


Tunnel Typeは、draft-ietf-bess-evpn-overlay にて定義されており、元のEVPN仕様(RFC7432)では定義されていない。

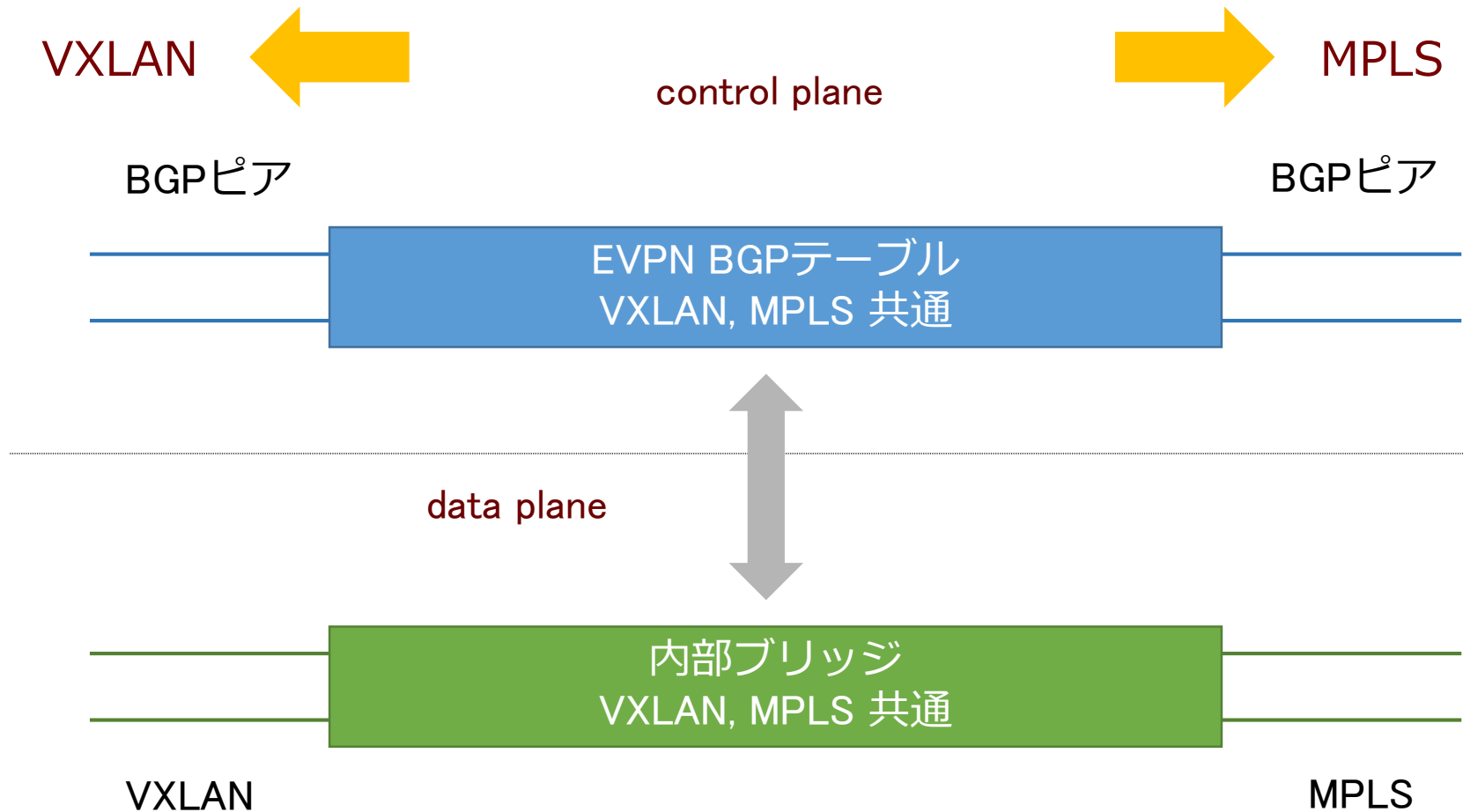
VXLANとMPLS両方に対応している機器では、Tunnel Typeを適切に設定するが、MPLSにしか対応していない機器では解釈できないことも。

本検証では、MPLS側にEVPN経路を広報する際、Tunnel Typeを付与しないように改修いただくことで回避した。

# テーブルとブリッジが分離された実装



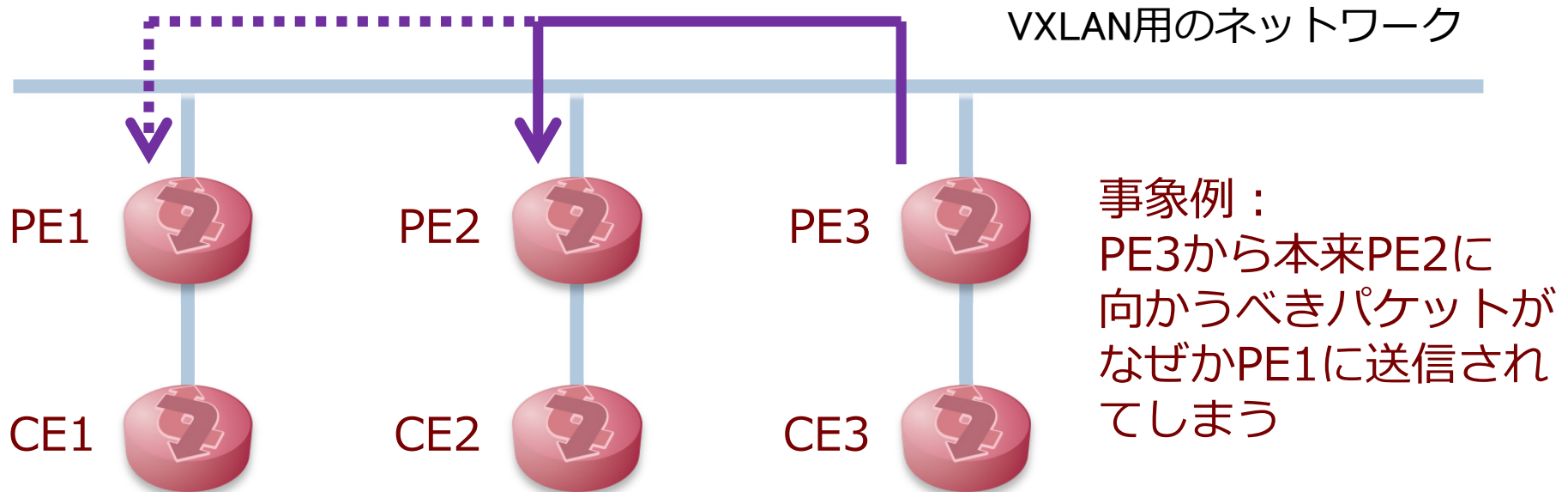
# テーブルとブリッジが同一である実装



## 5. デプロイ構成の注意点

# 一部製品におけるネットワーク構成の制限

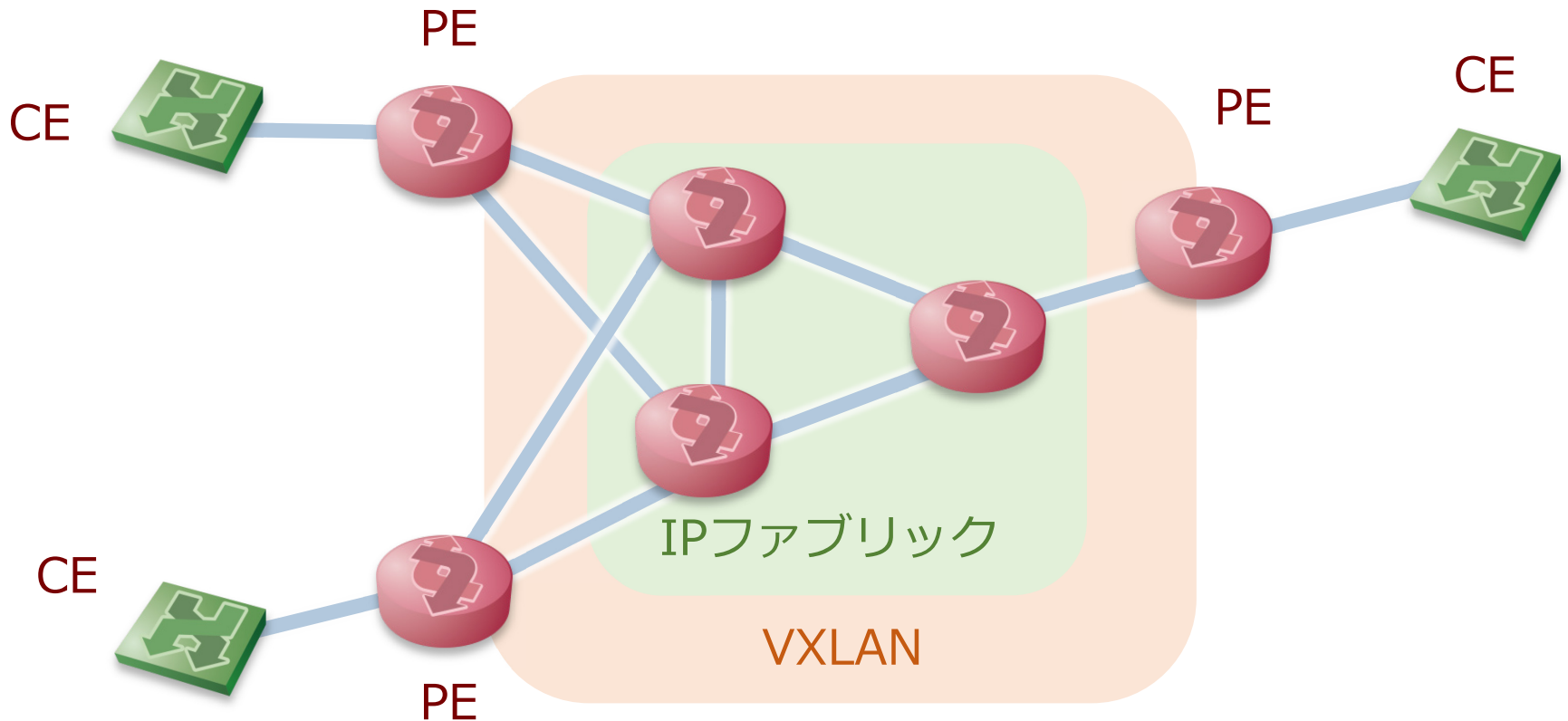
- 発生条件：VXLAN側インターフェイスと同一サブネットにリモートPEが複数存在するケース
- 事象：一部数機種にて、本来送信すべきPEとは別のPEにVXLANパケットを送信してしまう



※ EVPNは関係なくVXLANデータプレーンの問題

# 一部製品におけるネットワーク構成の制限

- IPファブリック環境を想定した製品もあるため導入構成に注意



PEのVXLAN側各IFの先に宛先MACアドレスが1つしか存在しないように構成



## まとめ

- 各製品のEVPN実装はかなりこなれてきており、シングルベンダでの構築では、ほぼ問題ない状況となっている。
- 今回ご紹介した確認事項も、ほとんどはシングルベンダであればそもそも気にする必要がない。
- 異なるメーカー製品によるダイバーシティを取りたい場合、EVPNマルチホーミング機能を用いたマルチシャーシLAGの実装は有力な選択肢となる。

# Special thanks!!



## 参考文献

---

- BGP MPLS-Based Ethernet VPN (RFC7432)
  - <https://tools.ietf.org/html/rfc7432>
- A Network Virtualization Overlay Solution using EVPN (I-D)
  - <https://tools.ietf.org/html/draft-ietf-bess-evpn-overlay>
  - 本検証実施時点では -03 を参照



INFINITE  
CHALLENGE



show<sub>△</sub>net ←