

ASを越えたE2E TEを実現する Multi-AS SR アーキテクチャの提案と検証

三島 航, 田島 照久

NTTコミュニケーションズ イノベーションセンター

w.mishima@ntt.com, teruhisa.tajima@ntt.com

概要

このセッションでは下記について紹介・議論したい

- ASを超えたTEの必要性と課題
- SR-MPLSやSRv6がAS単位で接続されたアーキテクチャの実現
- 我々の検証で確認できたことと今後の課題

本発表内容は弊社の環境で行った独自の検証結果に基づくものです。
いかなるサービスの実装に言及するものではなく、
また各ルータの性能を保証するものではありません。

登壇者紹介

■ 三島 航

(みしま わたる)

 watal_i27e

 <https://github.com/watal>



■ 略歴

- 2019年～ NTTコミュニケーションズ 技術開発部/イノベーションセンター

■ 仕事：NTT Com 全社検証網の技術検証・企画・設計・構築・運用（つまり全部）

- 2 ASes (AS38639, AS37900)： 20+ routers, 200+ switches, …
- その他スタンドアロン環境： ルータ・スイッチ・伝送装置・サーバ多数

■ 田島 照久

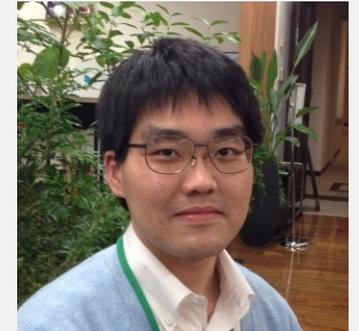
(たじま てるひさ)

■ 略歴

- 2015年～ SIer R&D
- 2019年～ NTTコミュニケーションズ 技術開発部/イノベーションセンター

■ これまでのJANOG

- JANOG40 BoF ネットワークテスト
- JANOG47 LT ネットワークトポ図

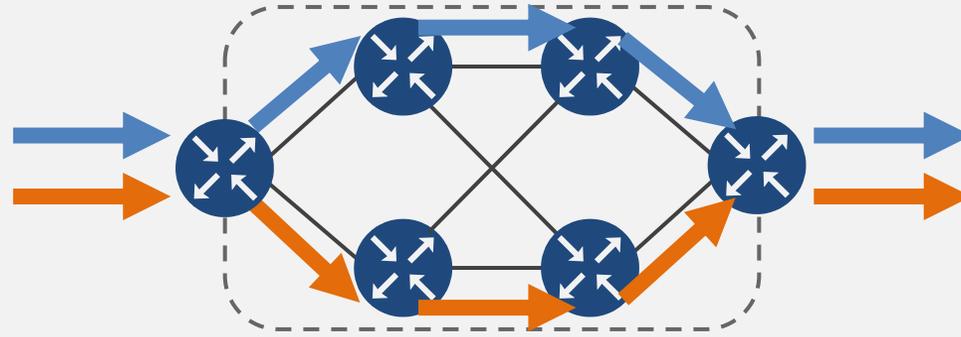


Multi-AS Segment Routing

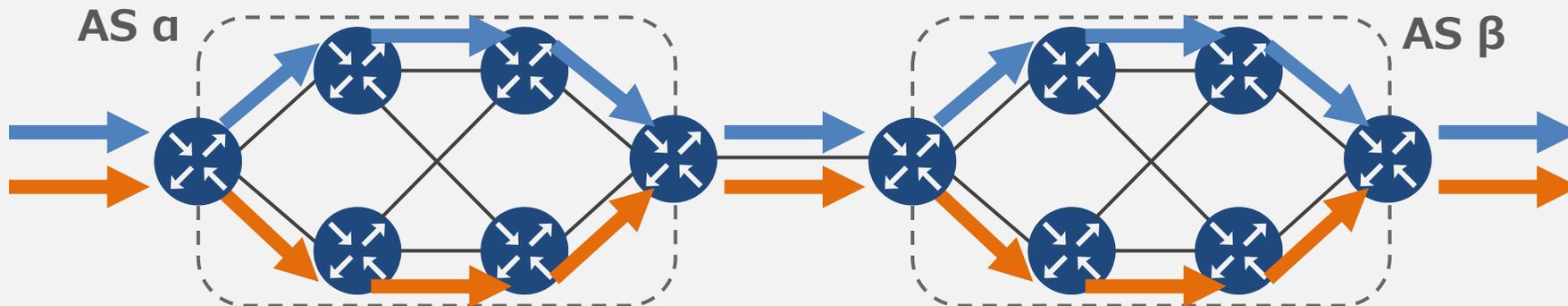
- モチベーションとアーキテクチャ -

ASを超えたEnd-to-EndのTraffic Engineering (1/2)

- 従来のTEはAS内の経路制御が対象

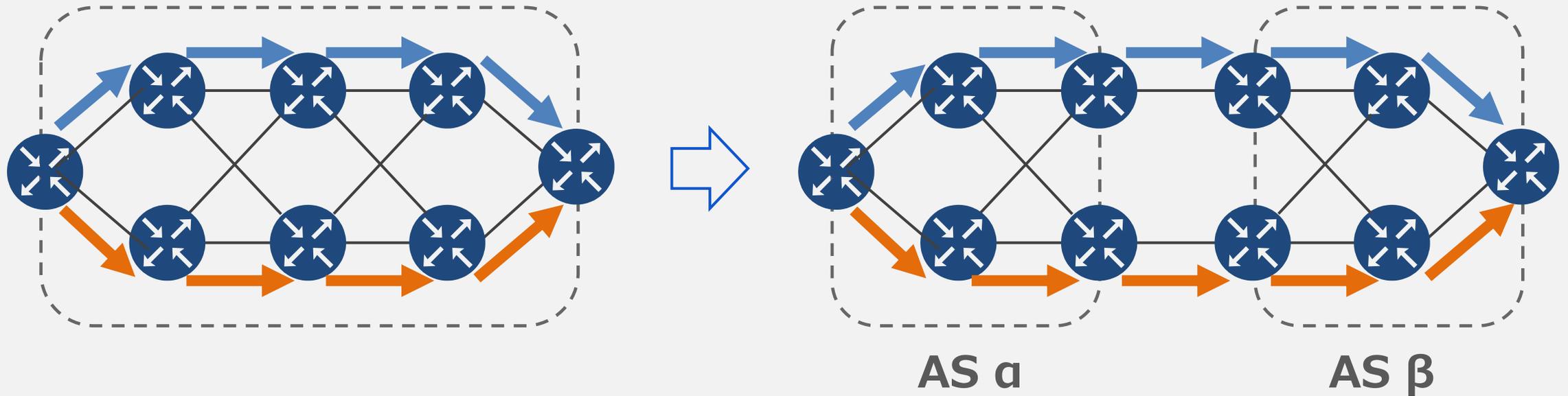


- 複数のASを経由するE2E経路の制御により、新たな価値を生み出せる
 - 例：モバイル網の機能ごとのNW・Public Cloud等と連携したサービス提供



ASを超えたEnd-to-EndのTraffic Engineering (2/2)

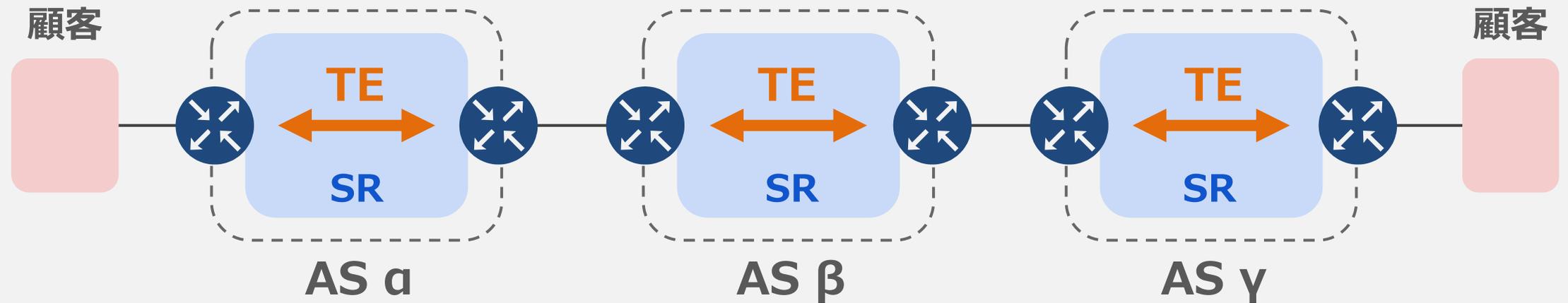
- 規模の拡大で経路制御が複雑になったASの分割ができるはず



→ 接続・分割のどちらのケースでもASを超えたE2EのTEが必要

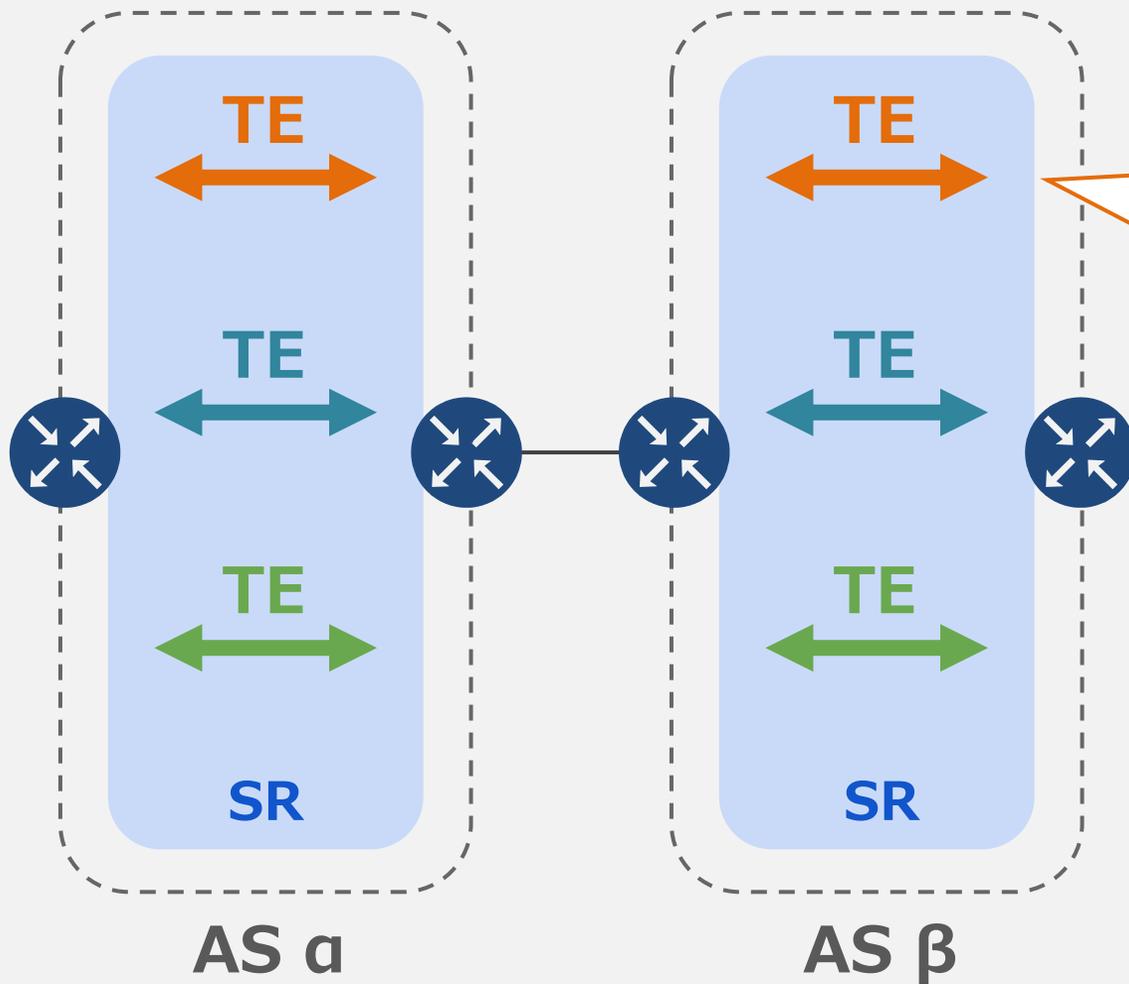
目指すアーキテクチャ : Multi-AS Segment Routing

- 複数のASを通して顧客同士が通信するモデル
- 各ASでdomain内のTEを行い、それらをつなぎ合わせる



TEをつなぎ合わせるとは？

- 各AS内でTEメニューを実装し、ASBRで選択する



例えばQoSの差別化以外に、Service Function Chainingのように経路を曲げて途中に処理を挟むTEメニューもあり

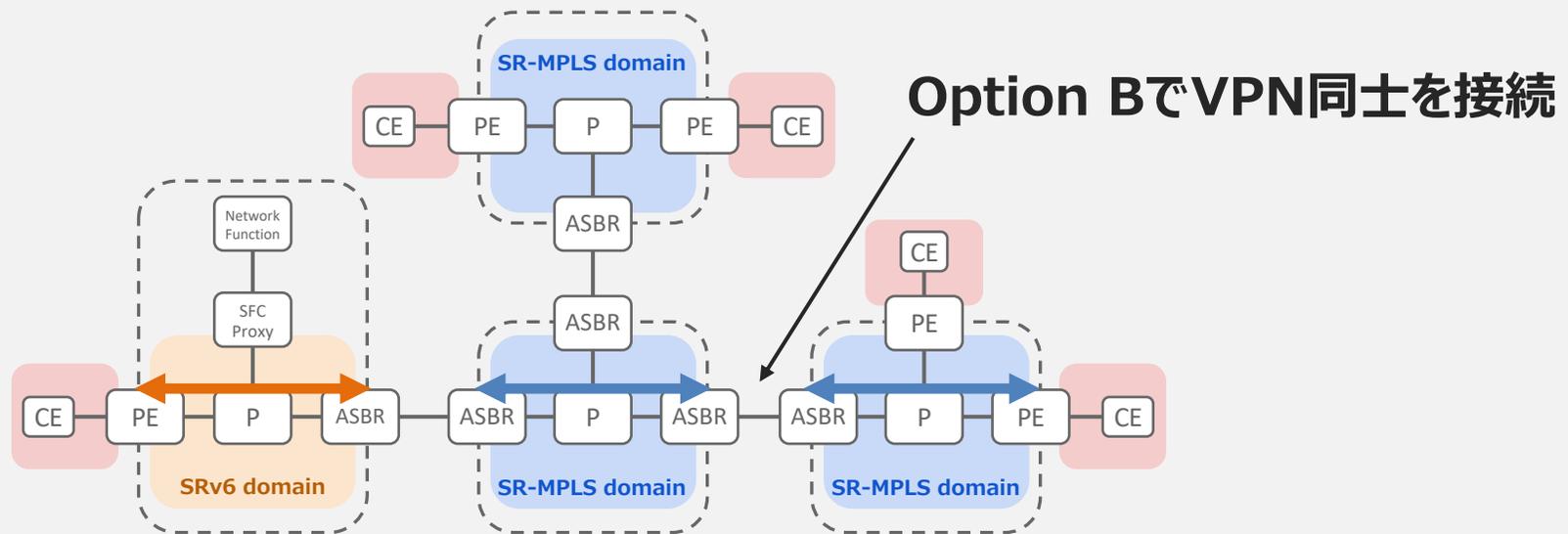
ASを超えるTEの難しさ

- ASの疎結合な構成と相互接続
 - ASごとのUnderlay/Overlay独立と、VPN/TEの連携の両立
- TEのメニュー提供手段
 - AS間を疎結合に構成しつつ、対向ASにメニューをリクエスト
- VPNの相互接続
 - VPNv4/VPNv6/EVPNそれぞれをMulti-ASに構成
- SR-MPLS/SRv6相互接続
 - 異なるプロトコル同士を疎結合に接続し、付加価値を享受

AS内とAS間それぞれの実装方針

- AS内：2種類のASを使い分ける
 - SR-MPLS：実装が比較的安定。VPN/TEを提供するASとして構築
 - SRv6：実装が足りない部分はあるが新機能を提供するASとして期待
- AS間：MPLS Inter-AS Option Bを採用

複数のSR-MPLS/SRv6で構成 + 相互接続



我々の取り組みと SR-MPLS/SRv6の実装状況概要

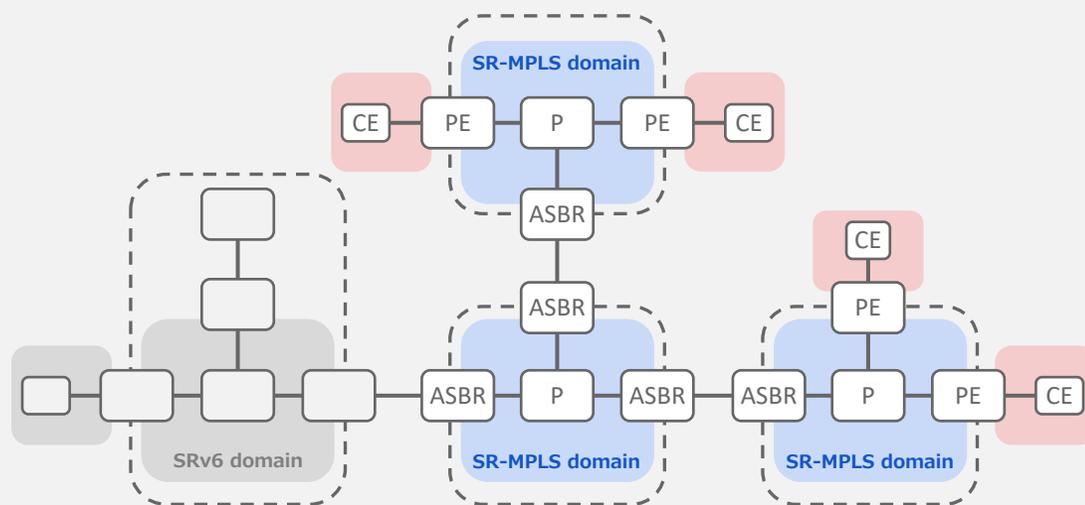
我々のSegment Routingの取り組み

- SR-MPLSを活用したインフラを構築・運用中
- SRv6の機能検証を実施中
- TE管理のためのコントローラ開発中
 - 自作Stateful PCE & PCEP libraryの開発/OSS公開

- 参考
 - [SR-MPLSの導入事例と今後の展望について \(MPLS JAPAN 2019\)](#)
 - [マルチベンダASかつマルチベンダにおけるSR-MPLS TEの実現 \(MPLS JAPAN 2021\)](#)
 - [Segment Routing 用 Stateful-PCE をフルスクラッチで開発した話 \(ENOG74\)](#)

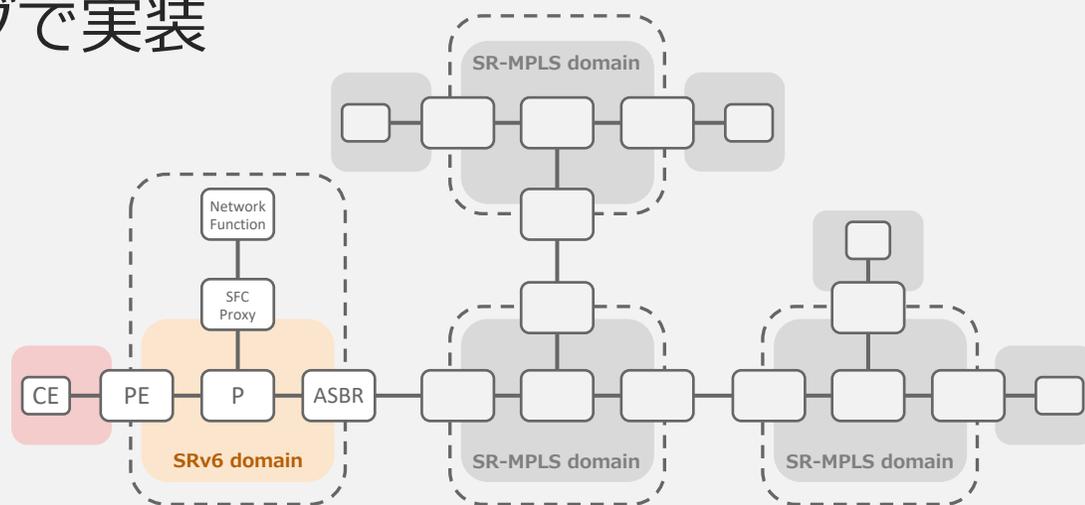
Single-ASにおけるSR-MPLSの状況

- VPN/TEメニューを提供する網として社内向けに運用中
- 各社ルータへの実装も実用も進んでいる
 - IGP : IS-ISで実装済、OSPFは実装途中
 - L3VPN (VPNv4/v6) : 実装済
 - L2VPN (EVPN) : 実装済
 - TE Explicit-Path : 実装済
 - Flex-Algo : 実装済
 - TWAMP (Delay-based TE) : 実装済
 - PCEP : シングルベンダでは市販製品あり



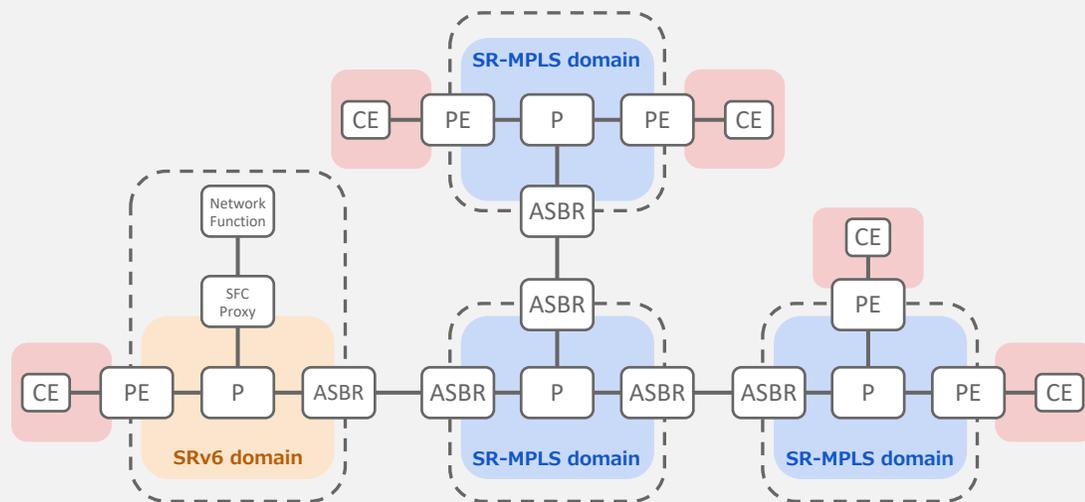
Single-ASにおけるSRv6の状況

- VPN/TEに加え、SFCを実現する網として検証中
 - 一部SR-MPLSより足りない実装もあるが、SR-MPLSにはない強みも存在
- 実装状況 ※ベンダごとに機能の差も多数存在
 - IGP : IS-IS拡張 (Locator広告) が実装済
 - L3VPN (VPNv4/VPNv6+End.DT4/DT6) : 実装済
 - L2VPN (EVPN+End.DX2) : 一部ベンダで実装
 - TE Explicit-Path : 一部ベンダで実装
 - Flex-Algo : 一部ベンダで実装
 - PCEP : 既存実装なし ([Internet-Draft](#))



中間まとめ

- ASを超えたTEの実現を目指し、Multi-AS Segment Routingを提案
 - AS同士を疎結合に設計 & 相互接続
 - SR-MPLS/SRv6の双方を実装
- ここからは実現に向けた考慮ポイントを紹介
 1. VPNのユースケース
 2. TEのユースケース
 3. SFC（網全体としてのユースケース）



1. VPNのユースケース

- AS間の共有リソース検討 -

VPNの考慮ポイント：RD/RT設計

- 要件：全リソースをAS内に閉じ、疎結合に設計&運用
- RDは機器ごとに自由に採番可能
 - EVI RD：自動RD（RFC7432）、VRF RD：AS番号:VRFごとのID
- RTはAS間でサービス固有のRTを共有 or ASBRで対向ASのRTに変換
 - L2VPN/L3VPN共に下記フォーマットを採用

AS番号 or サービスごとの共通ID

VRF/EVIごとのID

64999:100

2. TEのユースケース

- 各種ユースケースとSR Policy -

実現したいTEのユースケース

2-1. 経路ごとに制御可能なTEメニュー提供 (Color-Based Steering)

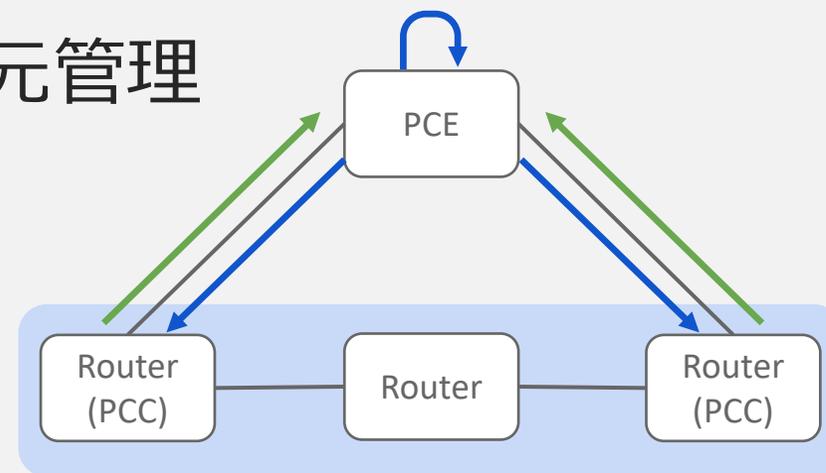
- ASを超えたTEの相互接続

2-2. Intent-basedな実装方式 (Flexible Algorithm)

- TEメニューを効率的に設計 & 適用

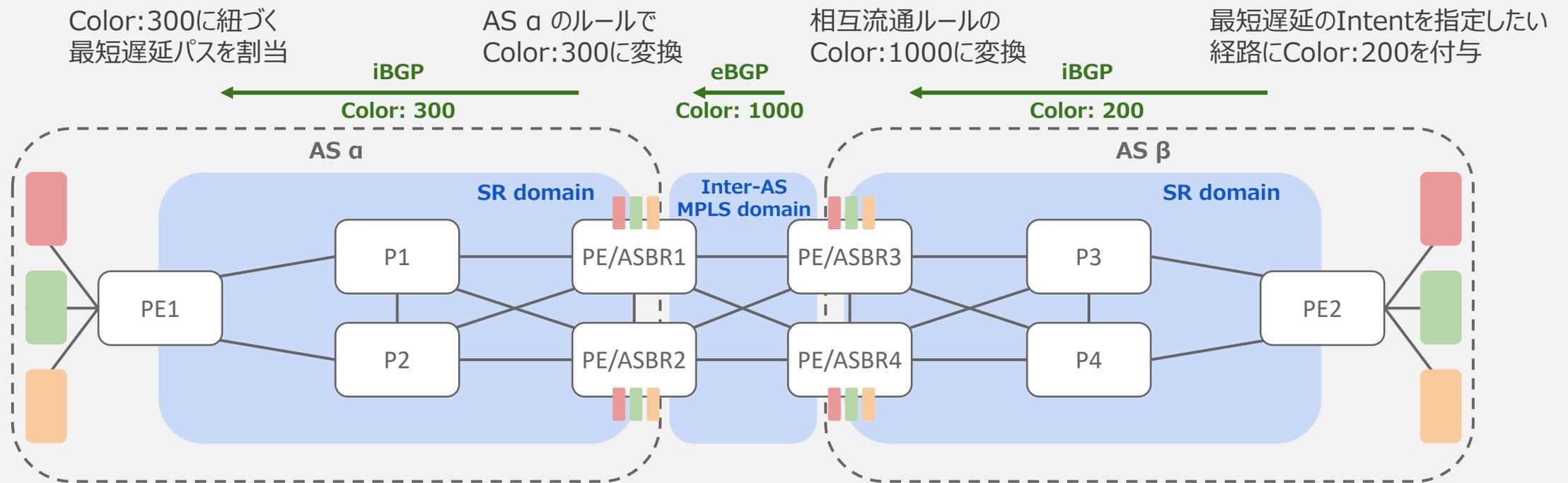
2-3. PCEを用いたDynamic SR policyの一元管理

- 網全体のIntentや発行済みPolicyを管理



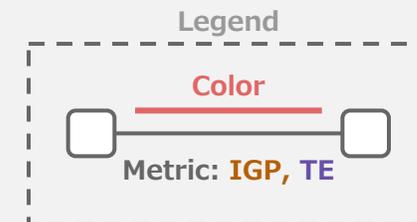
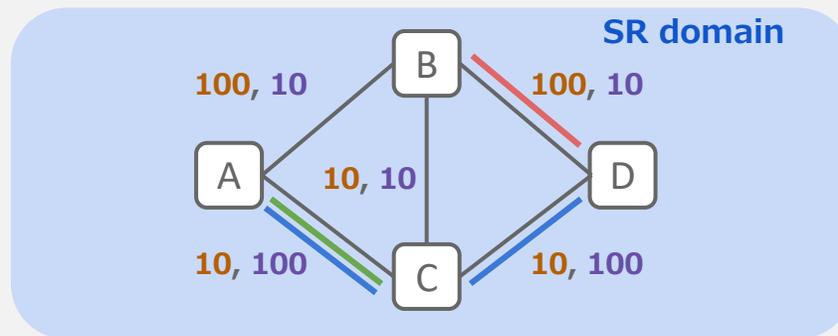
2-1. Color-Based Steering

- BGP Extended Community Colorを利用し、AS間でTEを接続
 - 対向ASに経路ごとのTEメニューをリクエスト
 - SR policyのColorと1:1対応させ適用

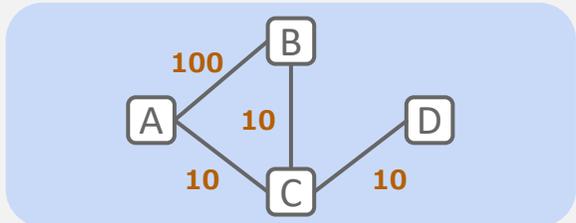


2-2. Flexible Algorithm (Flex-Algo)

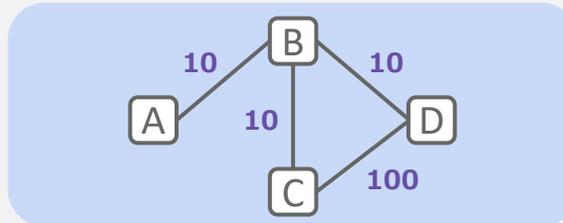
- 柔軟な経路計算を目的とし、IGPの経路計算面を多重化する技術
 - 効率的なTEメニューの設計・運用が可能
 - QoS実現のために多段のSegment Listが不要
 - TWAMPによる計測値に基づき、遅延ベースのスライシングを実現



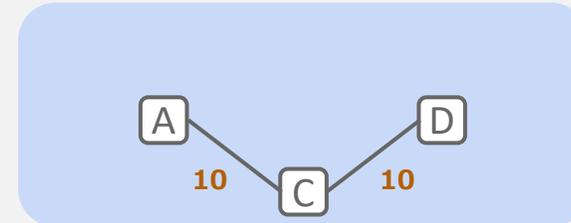
Algorithm 128: Exclude RED IGP Metric



Algorithm 129: Exclude GREEN TE Metric



Algorithm 130: include-all Blue IGP Metric



Explicit/Dynamic SR Policy

- SR policyの生成手法により2種類に分類
 - explicit SR policy: 明示的に指定したSegment Listを適用
 - dynamic SR policy: 制約 (Intent) を満たすよう最適なポリシーを計算
 - 機器内でポリシーを計算する分散型と、中央で計算する集中型が存在
 - 参考: [draft-ietf-spring-segment-routing-policy](#)
- SR-MPLS/SRv6でそれぞれのポリシーを検証済
 - explicit SR Policy : SR-MPLSは全ベンダOK、SRv6は一部実装にのみ存在
 - dynamic SR Policy : SR-MPLSでのみ利用。要件に合わせて自作PCEを開発

2-3. Pola PCE - Stateful PCE & PCEP LibraryのOSS実装

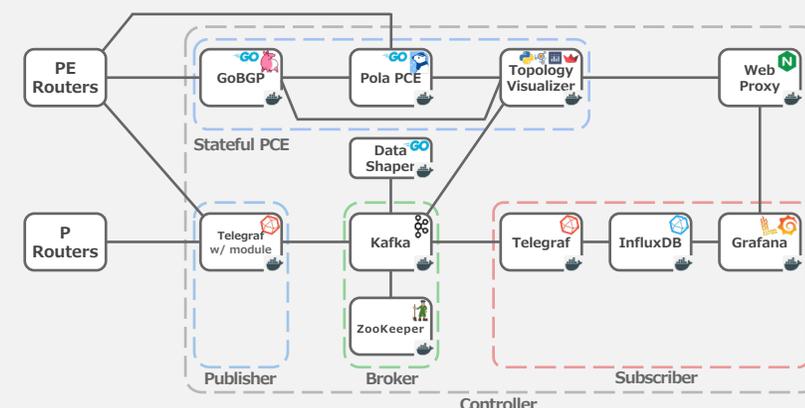
■ Multi-AS SRで求める要件を満たすため開発 & OSS公開

- 拡張性：SRv6やMulti-AS機能の追加を視野
- マルチベンダ：各種機器との相互接続
- マイクロサービス：gRPC APIを提供。自作コントローラやGoBGPとの連携



■ 参考資料

- GitHub：<https://github.com/nttcom/pola>
- Pola PCE 公式ドキュメント：<https://nttcom.github.io/pola/>
- [大規模SR網の運用を効率化するネットワークコントローラの開発 \(NTT Tech Conference 2022\)](#)
- [Segment Routing 用 Stateful-PCE をフルスクラッチで開発した話 \(ENOG74\)](#)



SR policyの適用単位と適用条件

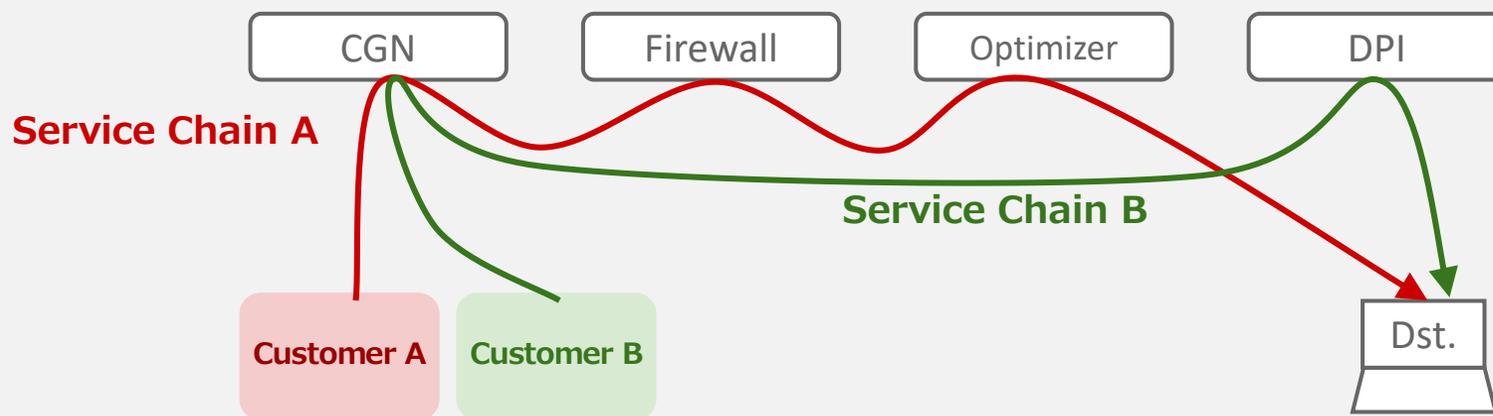
- SR policyを適用する粒度は実装により選択可能
 - Endpoint > VRF > Prefix > flow (5-tuple) の順に制御単位が大きい
 - 制御単位の粒度は上位互換関係ではなく、ユースケースに応じて選択すべき
- ToS/CoSを利用し、パケットに付与した値をもとにTE可能
 - CEからPEに対するSR policyのリクエスト等に利用
 - AS間のサービス提供にも利用可能
 - BGP ColorやBSIDによる接続と比べ、ToS/CoSルールの管理コストが必要

3. SFCのユースケース

- SR-MPLS/SRv6相互接続と更なる拡張 -

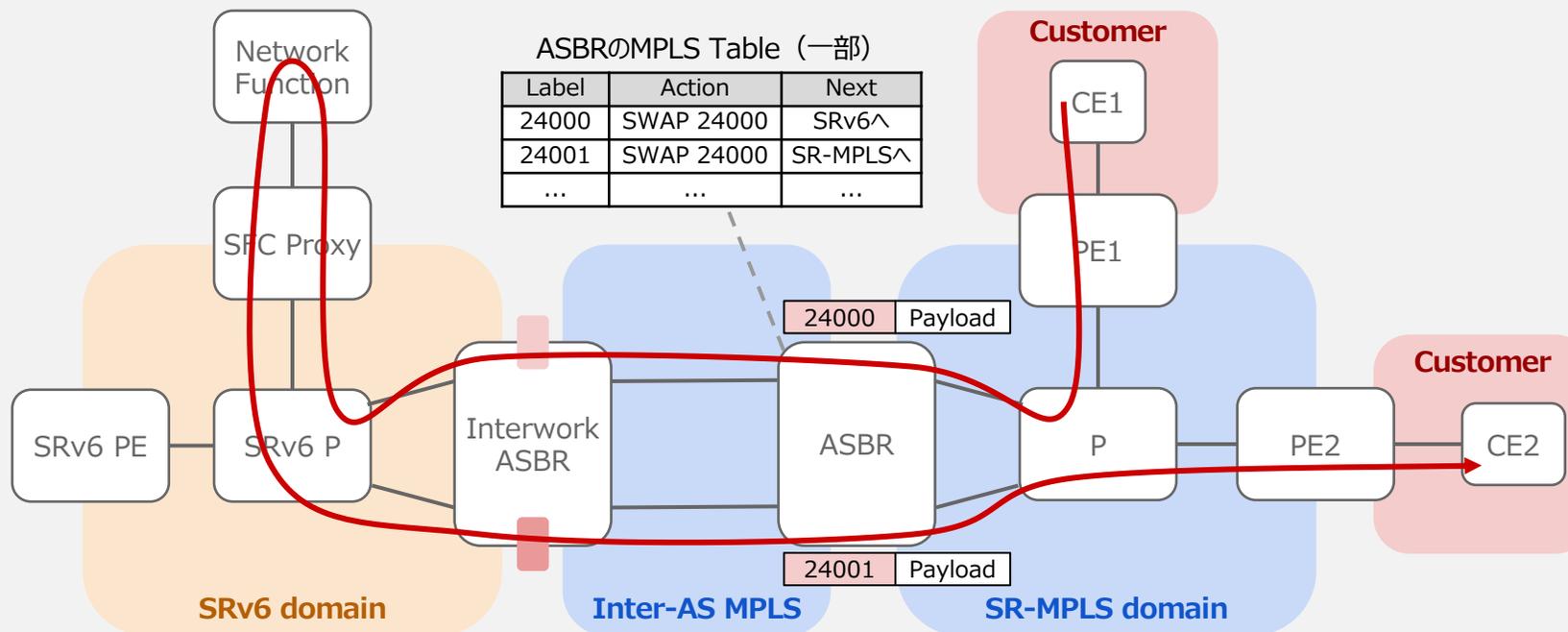
Service Function Chaining

- NW上の機能を特定の順序で繋ぎ、サービスとして提供する技術
 - SRv6の主なユースケースの1つ
 - 先進的なSFC手法を検証中、Multi-ASの利点を活かし外部ASとの共同実験も
 - SR-MPLS/SRv6相互接続により、SR-MPLSにもSFCを提供！
 - 新規性として、SR-MPLS網のSRv6折り返しを実現



SR-MPLS/SRv6 相互接続の実装

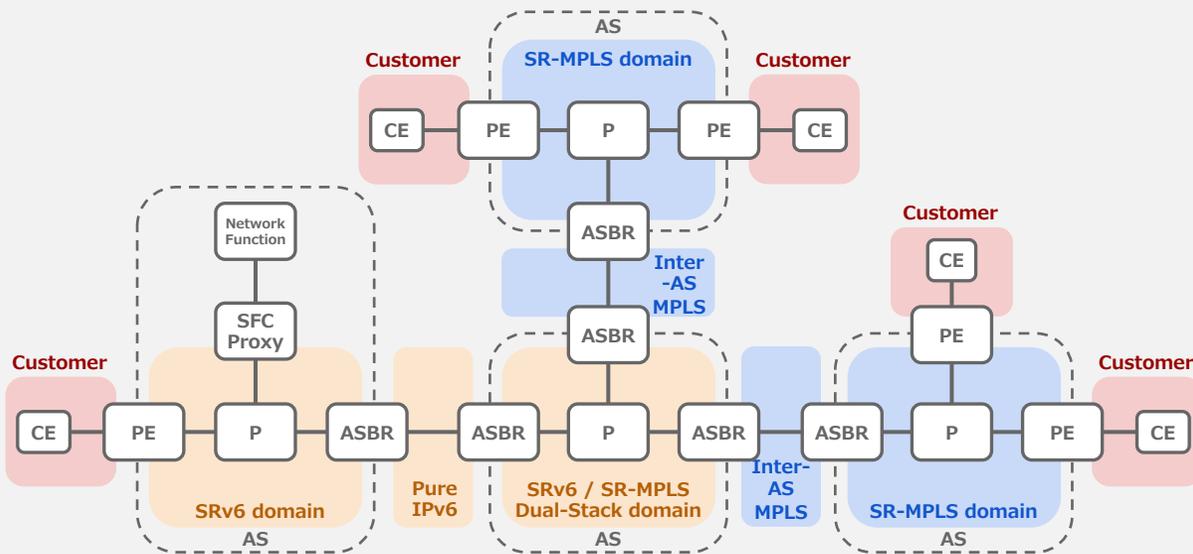
- 折り返しのため、送受信双方のVRFを作成
 - 各VRFでSR-MPLS/SRv6のVPNを接続
- 全EndpointにSRv6を用いた相互接続なども可能
 - 各ASをどこまで疎結合にするか、またTE要件や機器の実装状況を考慮し設計



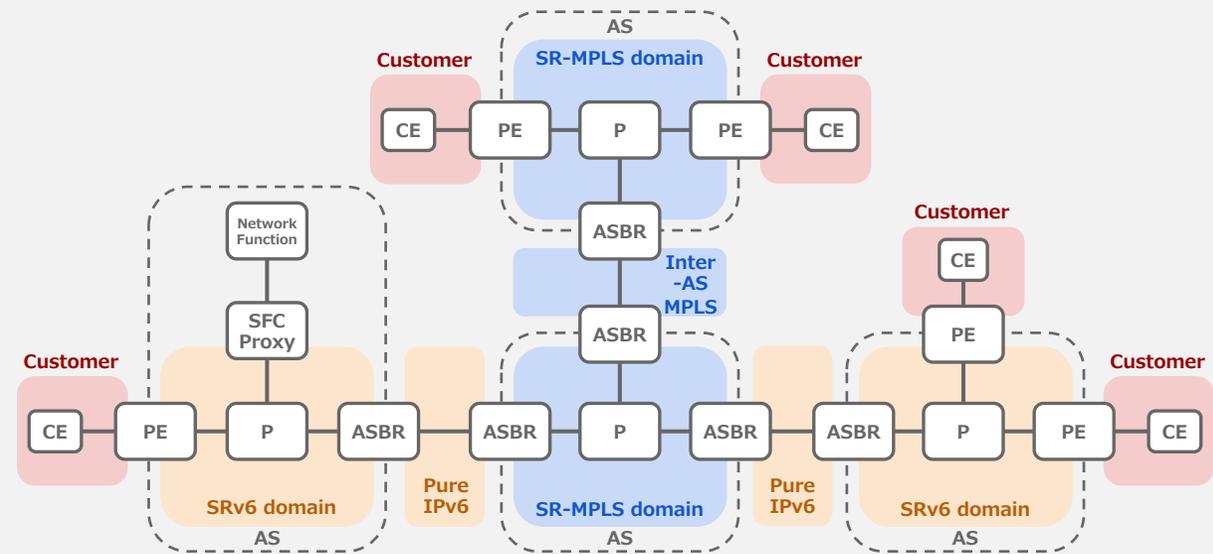
SR-MPLS/SRv6 Migration

■ ユースケース：一部ASでの最新機能の検証 & 先行利用

- SRv6網の機能をSR-MPLSにも提供可能に！
- 先進技術の Dogfooding → Trial Release → Migration の流れを実現
 - Multi-AS SRの強みを活かし、他ASに影響を与えず一部ASに導入 & 相互利用



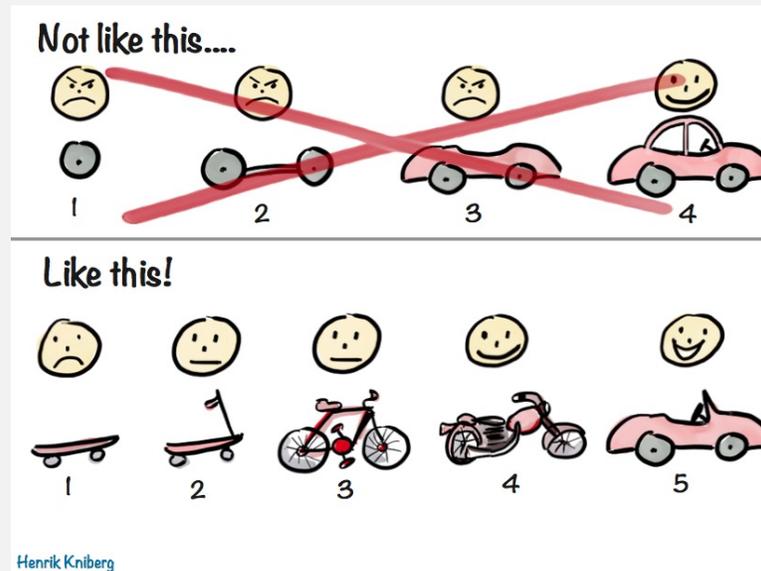
隣接するASをSRv6 Migrationしていく例



離れたASをMigrationし、IPv6 underlayで接続する例

SRv6導入に向けての考慮ポイント

- 基本的なVPN/TE等は実装が先行するSR-MPLSに優位性あり
 - SRv6の強みとなるSFCや独自Functionによる付加価値が移行の鍵
- まずは安定したSR-MPLSでサービスを提供、必要に応じてSRv6化
 - Migrationはウォーターフォールに対するアジャイルの発想



画像 : <https://twitter.com/henrikkniberg/status/691932339354599425>

検証・実装時の悩みなど

- マルチベンダの課題：仕様の違いや実装状況の差
 - 特にSRv6では機器により未実装の機能が多数
 - Explicit-Path・VPN C-plane・コントローラ連携など
 - SR-MPLSでも実装中の機能があり、まだ枯れ切っていない
 - IS-IS TLV実装に一部違いが存在
 - EVPNのInter-AS Option-B等を実装中の部分が存在
 - 一部機器ではSRGB変更後にRebootが必要
 - ZTP時も含まれるため自動化に工夫が必要
- (Multi-AS/Single-AS双方の) 動作検証や障害時の切り分けの難しさ
 - SR-MPLS/SRv6共にOAMに関する知見が少ない&一部存在しない

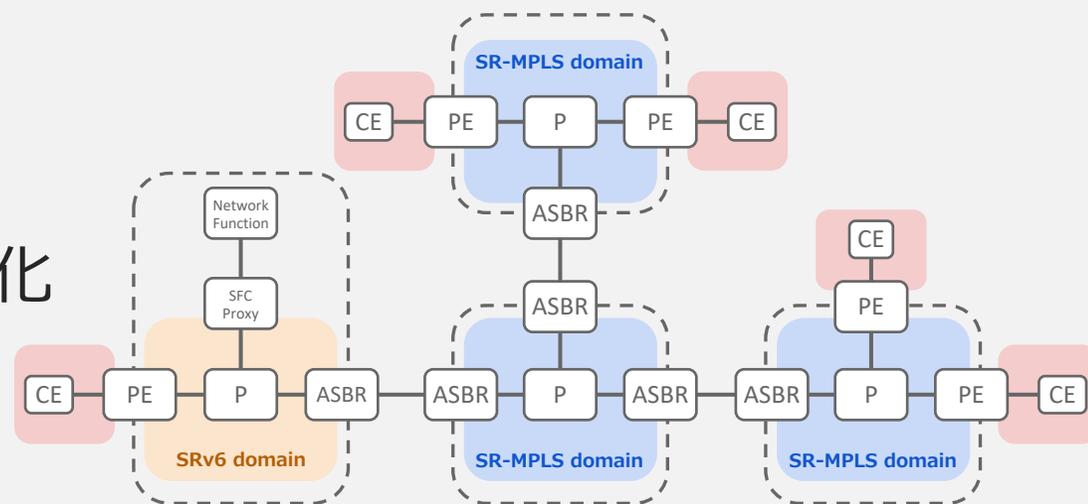
まとめ と Future Work

まとめ : Multi-AS SRアーキテクチャと取り組み

- 複数のASを疎結合に接続し、一連のTE/VPNを実現
 - 運用者の異なるASの相互接続による通信への付加価値提供
 - 用途ごとのAS分割によるスケーラビリティ向上
 - Option Bによる接続とColorによるTEメニューの提供

- 各ASでSRの機能を検証・実装

- 様々な粒度でのPolicy適用/Flex-Algo
- SR-MPLS/SRv6相互接続による価値提供
- コントローラによるSR Policy管理の効率化



Future Work

- Multi-AS SRをより大規模な環境にデプロイし、TEを提供したい！
 - 複数拠点に跨がるSR-MPLS網の大規模展開
 - 複数の機能を持つネットワーク同士の接続
 - 複数の運用者が持つネットワークの相互接続

- PCEの機能を充実させ、大規模TEでも集中管理可能にしたい！
 - Multi-AS固有のサービスメニュー等を可視化・相互接続・セルフマネージ化
 - FastReRouteなどの状態把握を含めた帯域保証

参考情報

■ Multi-AS SRの技術解説を連載中

- [nttcom engineers blog](#) で検索
- 隔週月曜日に連載予定
- 連載目次 : <https://engineers.ntt.com/entry/2022/06/13/090000>



■ Pola PCEのContributor募集中！

- 大規模なSR網を制御したい方
- RFC-compliantなOSSを実装したい方など
- お気軽にPR下さい！ : <https://github.com/nttcom/pola>



議論トピック（もちろんこれ以外も大歓迎です！）

- そもそもASを超えたTEについて
 - コンテンツを流している方 > ISPにやって欲しい制御はありますか？
 - ISPの方 > 流れるユーザパケットの属性で転送制御したいですか？

- 複数のASで構成するデザインについて
 - キャリア網でDC-NWのPodのようにASを増やしていく際の課題は？