Internet week 2015

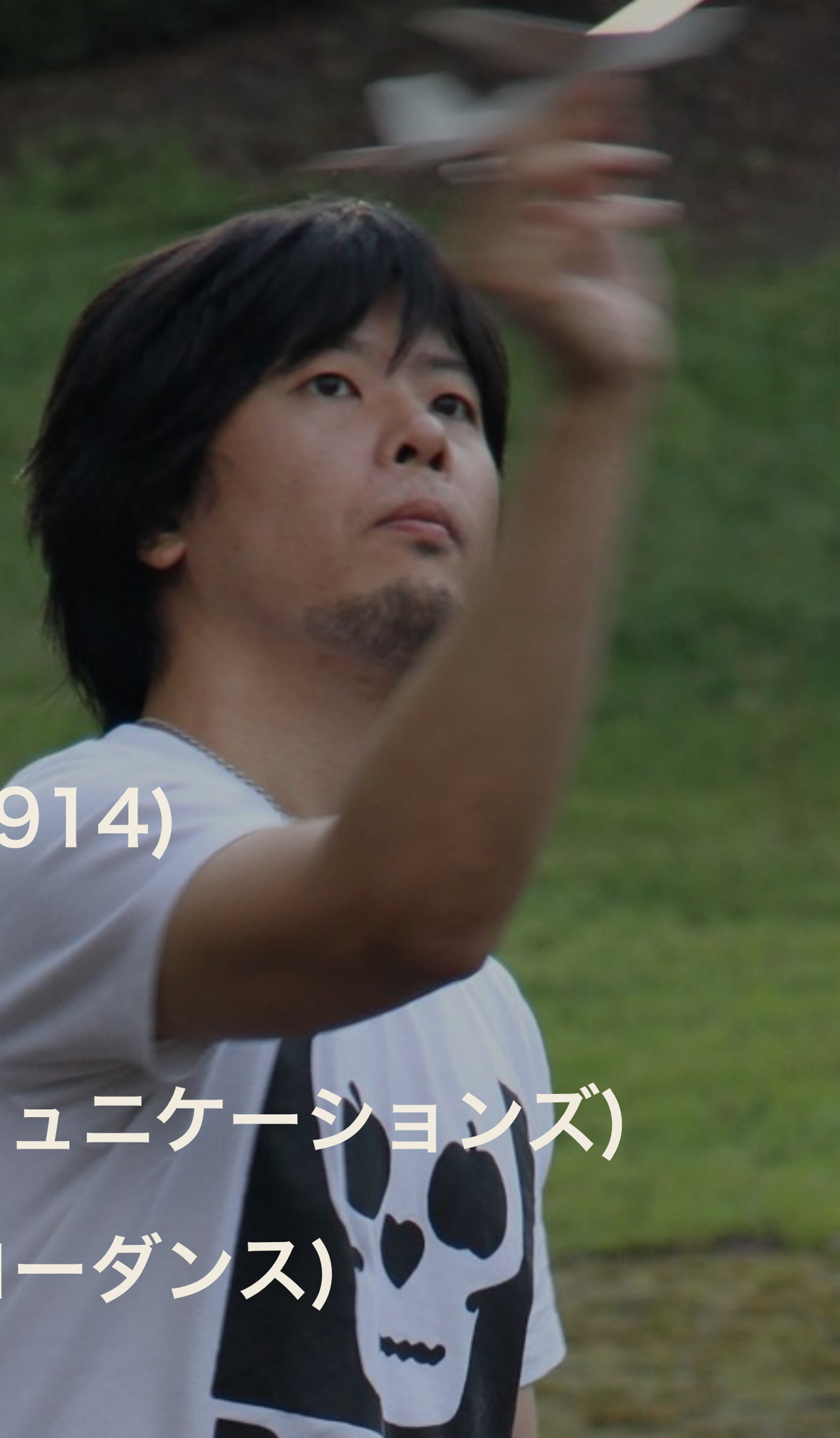
できる網設計

網設計のための BGP入門

小島 慎太郎

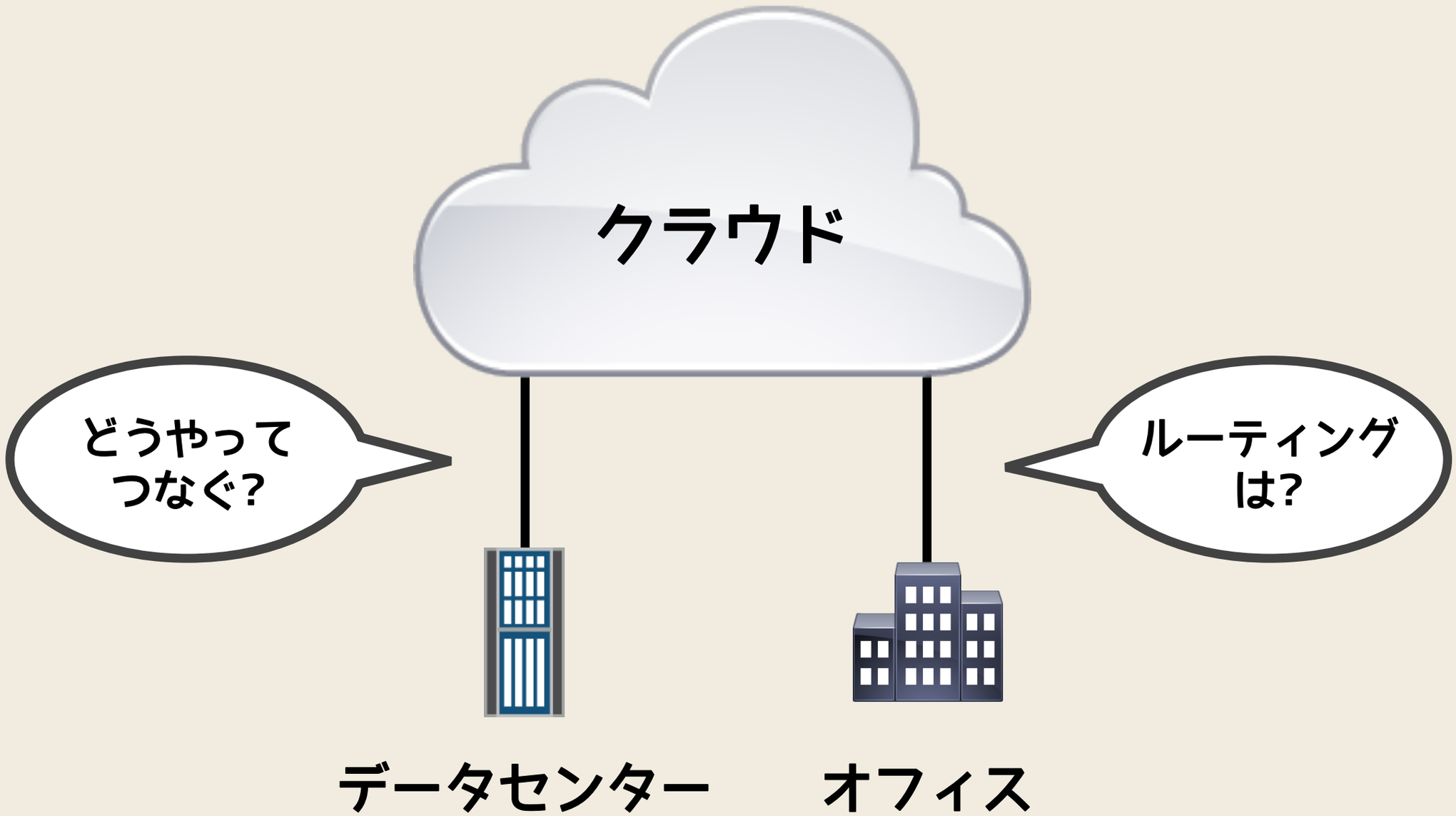
codeout

http://about.me/codeout



- ISP: 5年 (ntt.net / AS2914)
- IX:　4年 (JPNAP)
- クラウド:　1年 (NTT コミュニケーションズ)
- ネットワーク全般: 1年 (コーダンス)

本パートの話題



✋ 早速ですがアンケートです

Q1. クラウドサービス、使ってますか？

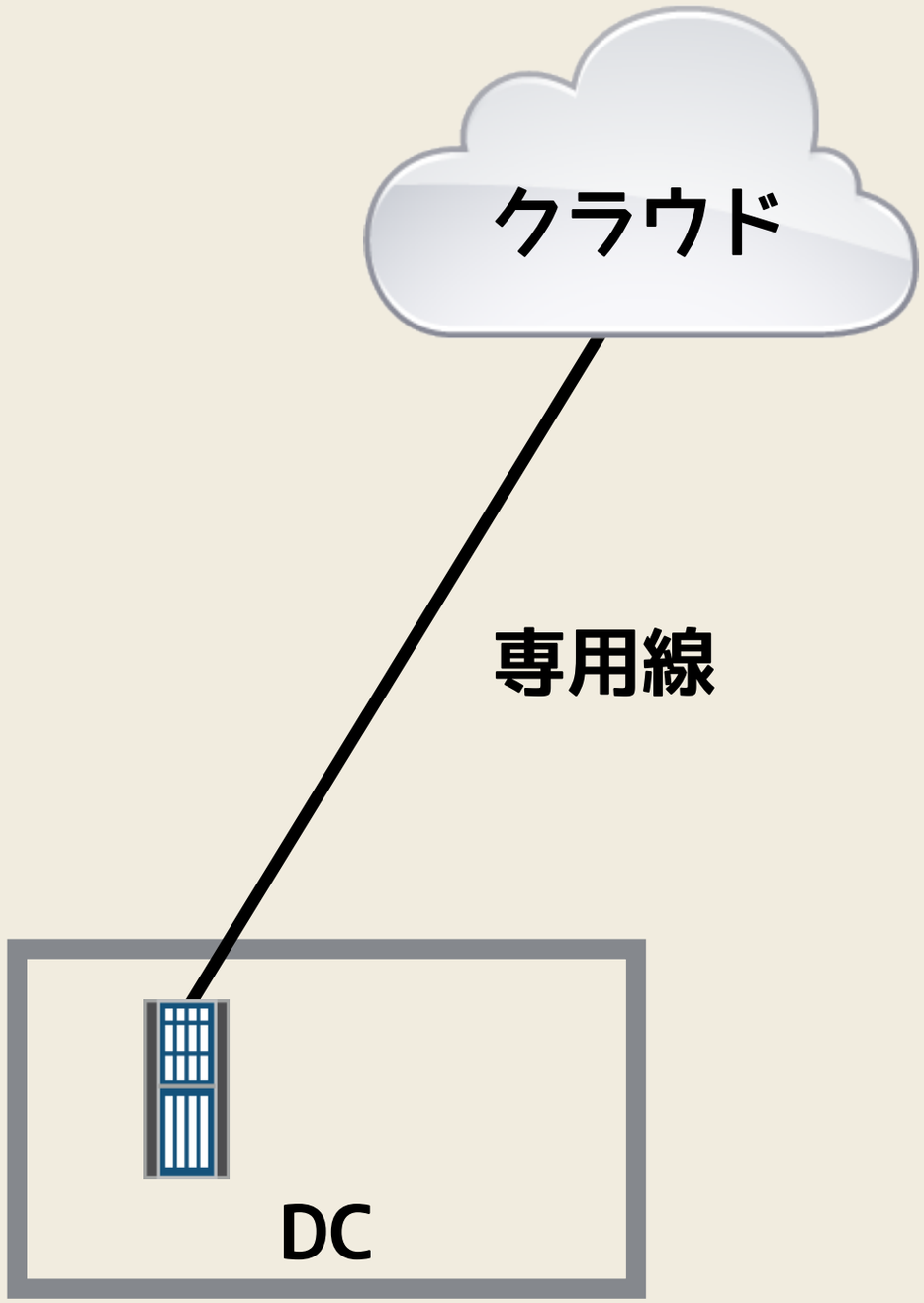
- パブリック クラウド
- ホステッド プライベート クラウド

✋ 早速ですがアンケートです

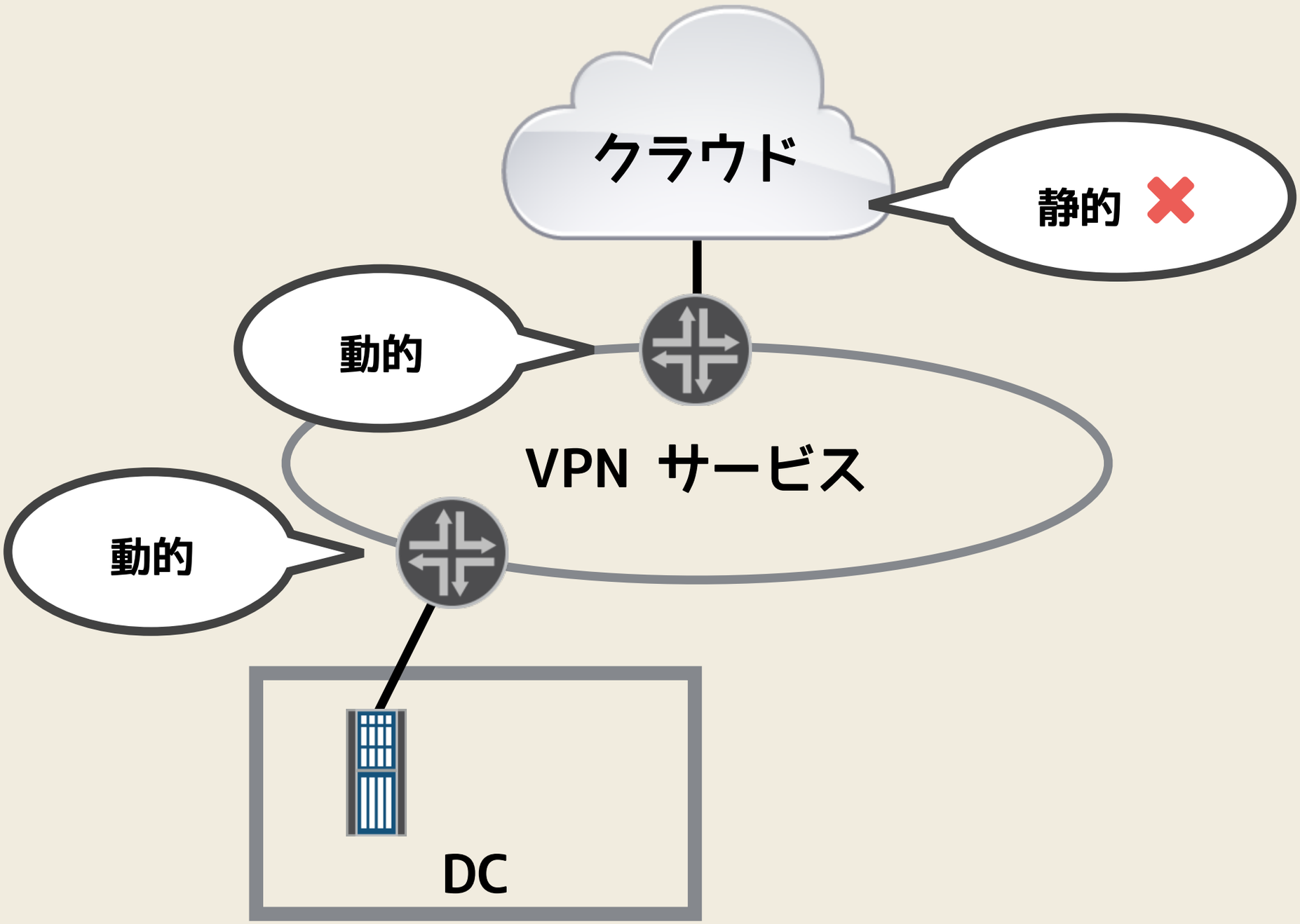
Q2. クラウドとのルーティングは？

- 静的
- 動的

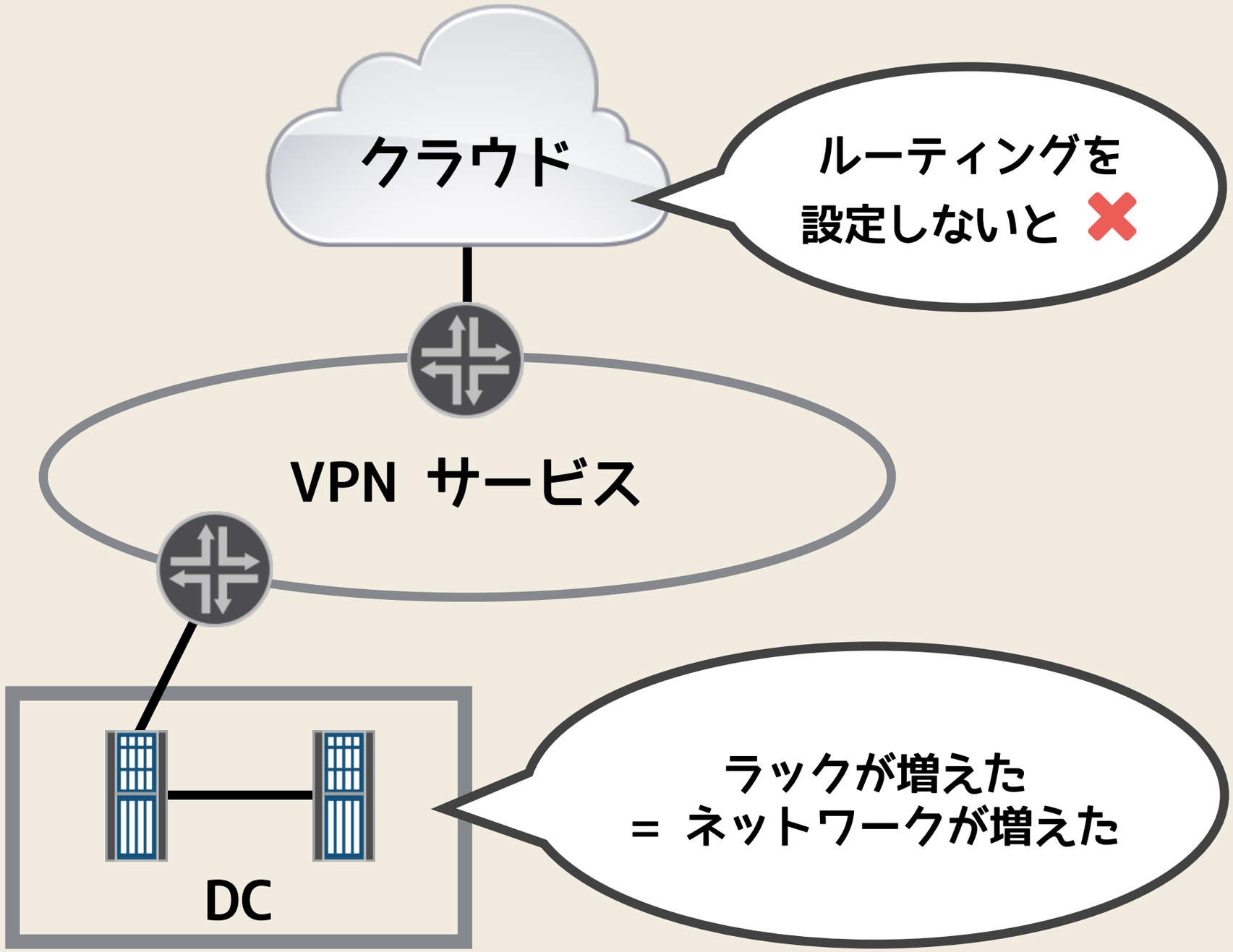
プライベートクラウドとの接続



プライベートクラウドとの接続

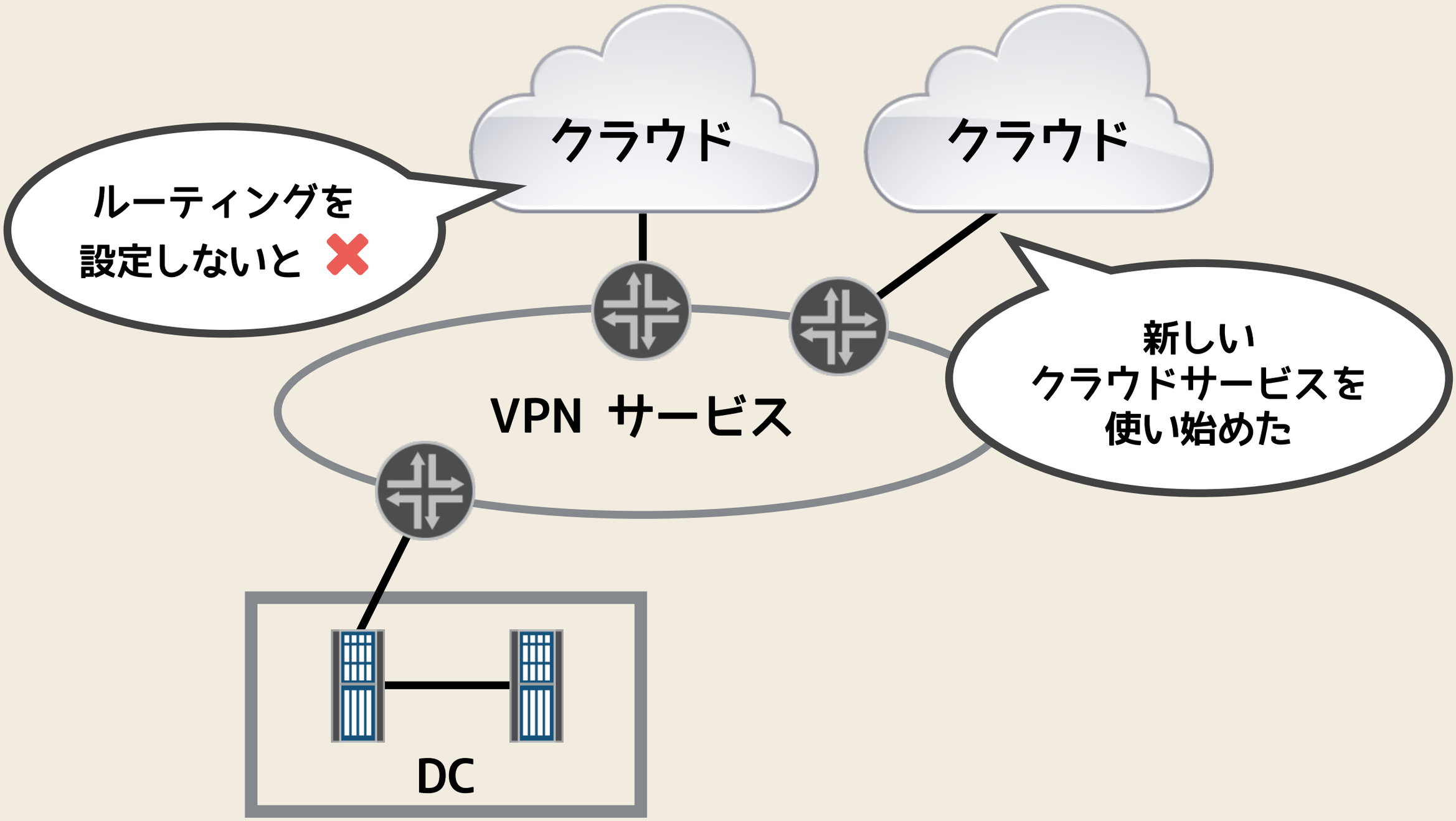


プライベートクラウドとの接続



めんどくさい

マルチクラウドアクセス



めんどくさい

ここが
本パートの
スタート地点です

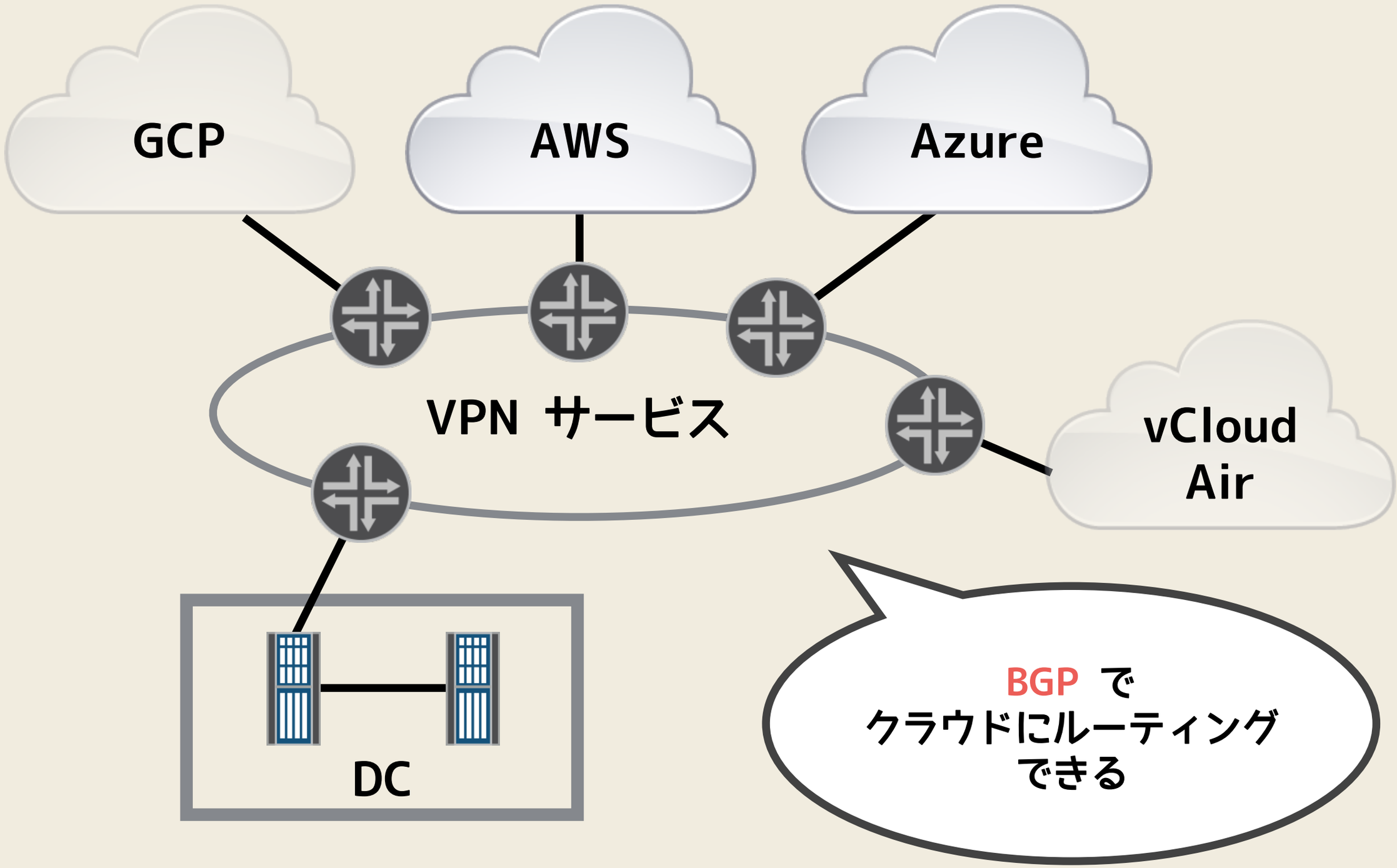
リーダーたちは

どうやっているか

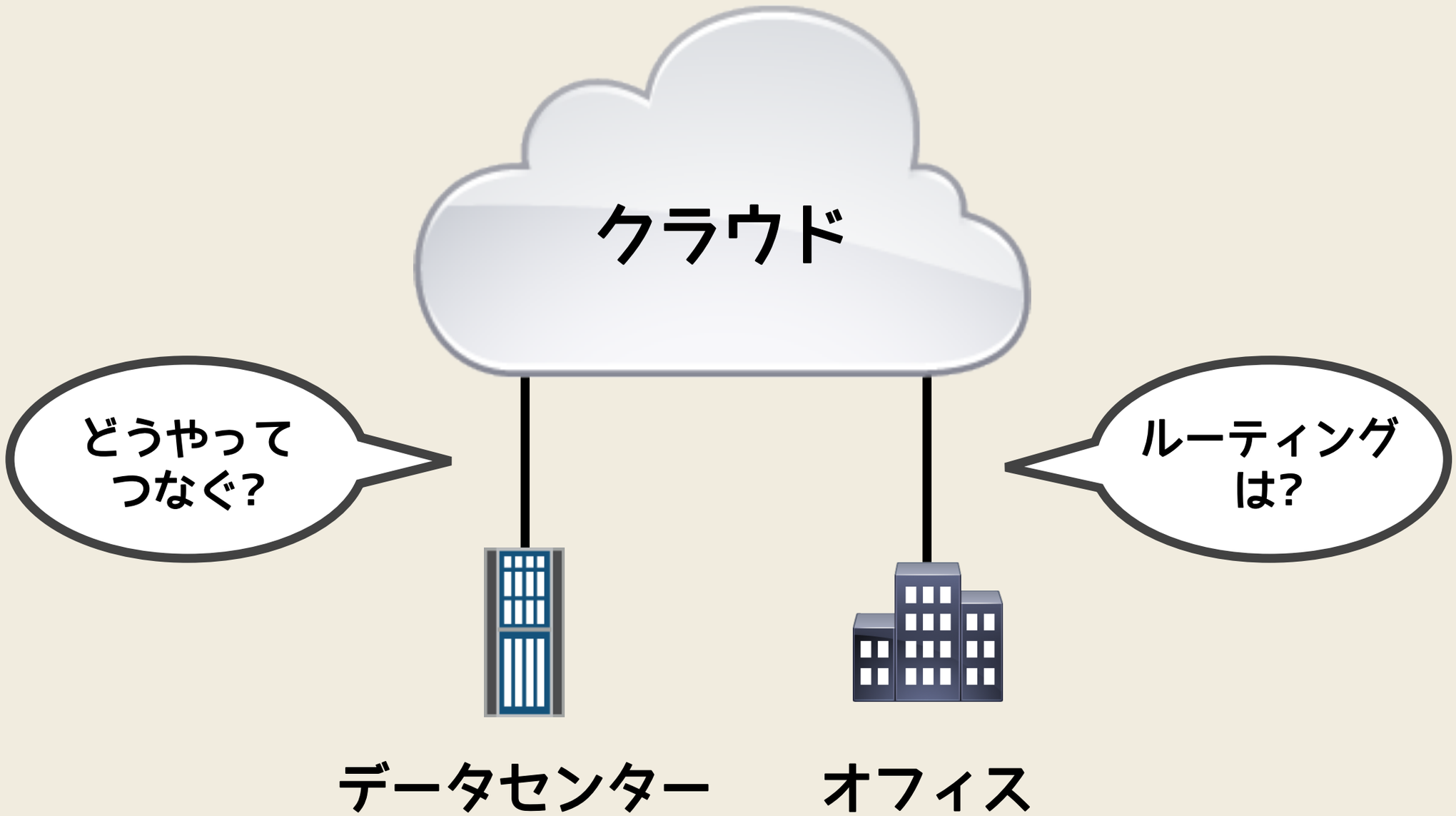
2015 Magic Quadrant for Cloud Infrastructure as a Service, Worldwide

https://aws.amazon.com/jp/resources/gartner-2015-mq-learn-more/

クラウドへのルーティングを動的に



本パートの話題



スケールする
クラウド接続方法
について
理解を深めたい

Agenda

- クラウドとの接続方法
  - ルーティング 編
    - BGP
    - クラウドへの応用
  - オンプレミスとの接続 編
    - L3
    - L3VPN
    - L2VPN
  - 応用 編

クラウドとの接続方法
(ルーティング編)

BGP

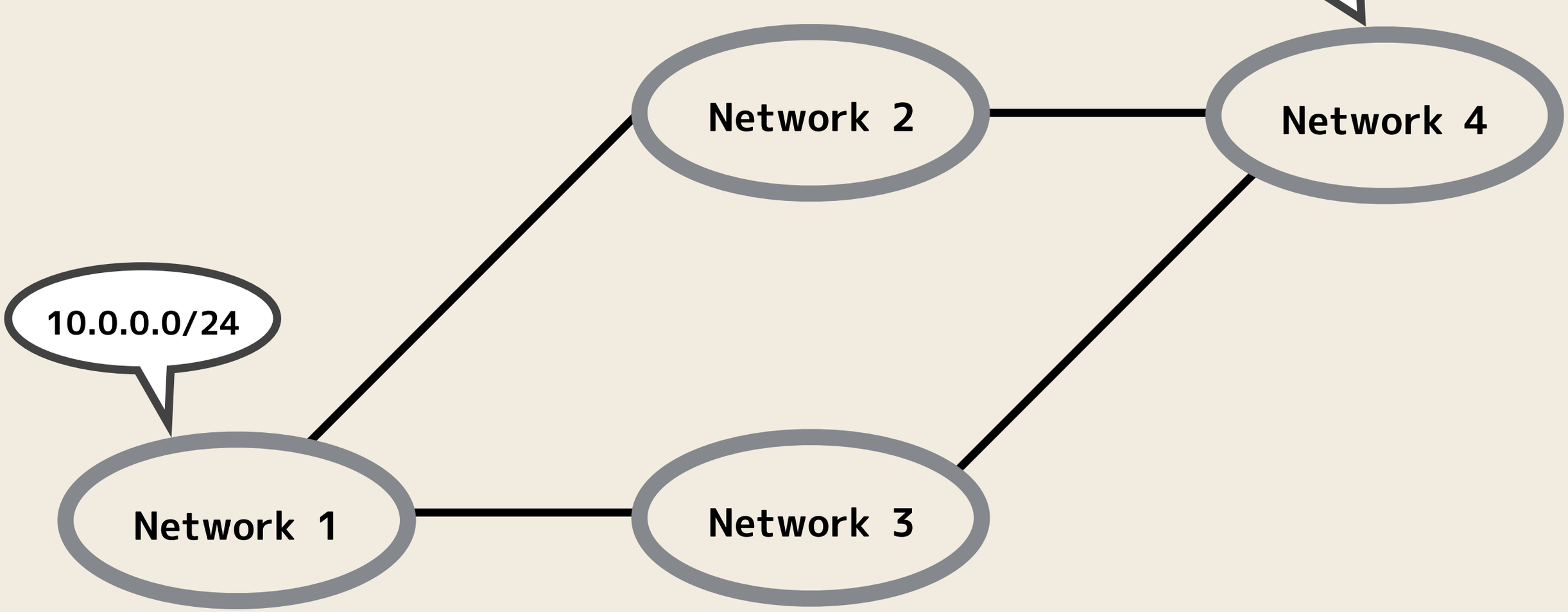
BGP とは

- RFC4271
  (A Border Gateway Protocol 4)
  - インターネットを制御するためのルーティングプロトコル
  - インターネット = ネットワークを相互接続したもの

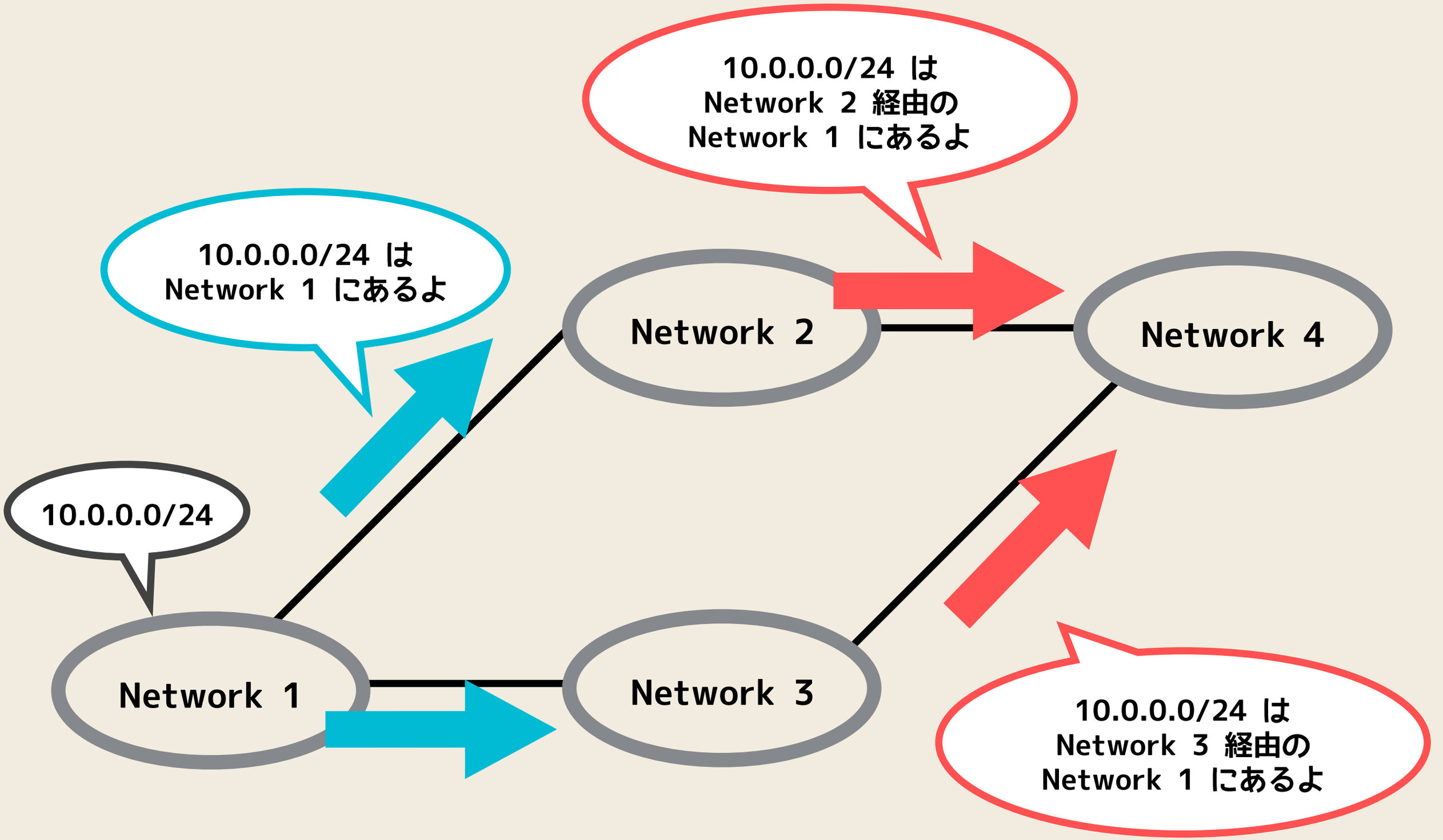
BGP は
パケット転送先を探すプロトコル

ルーティング

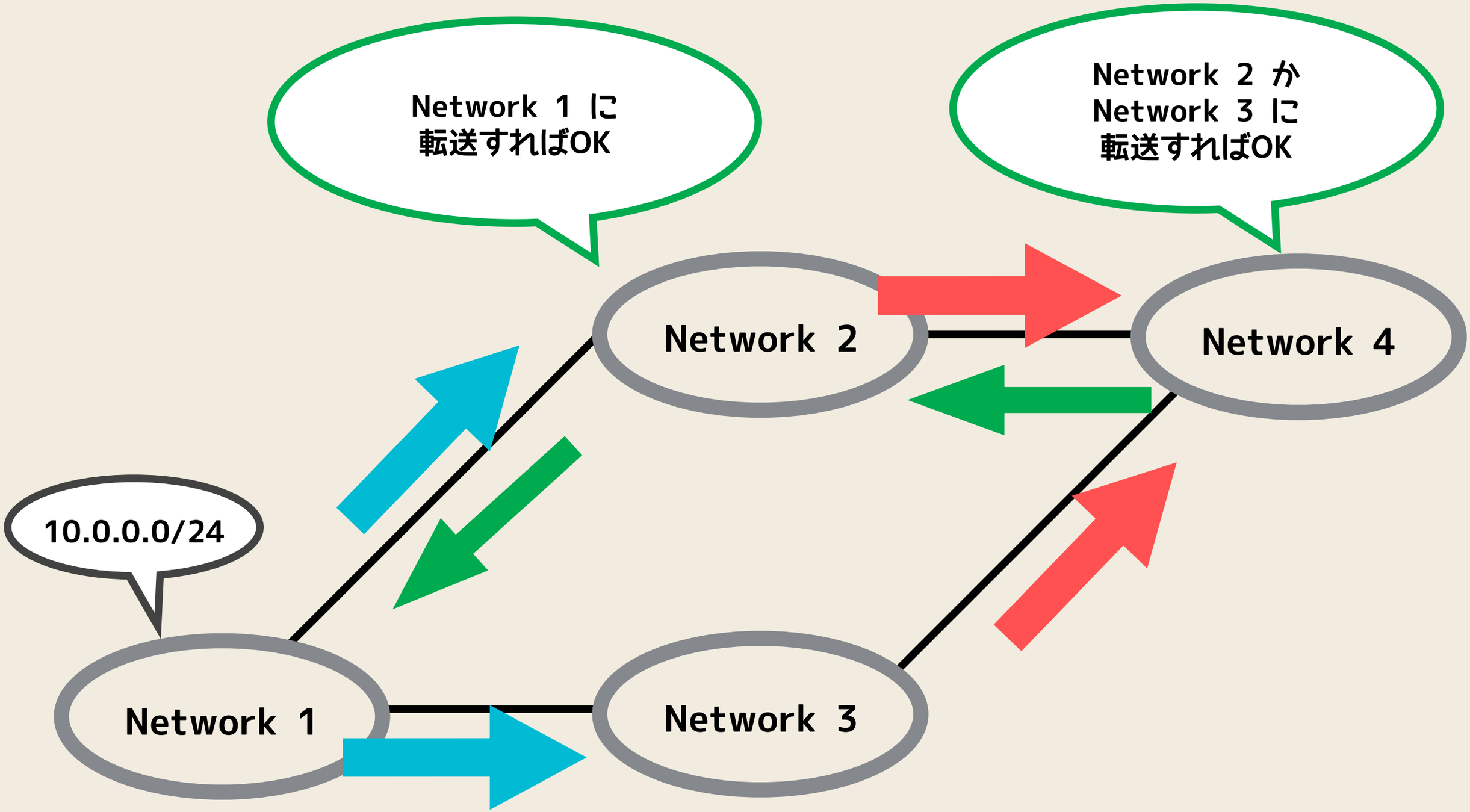
10.0.0.0/24 に
パケットを送りたい。
どこに転送すれば？



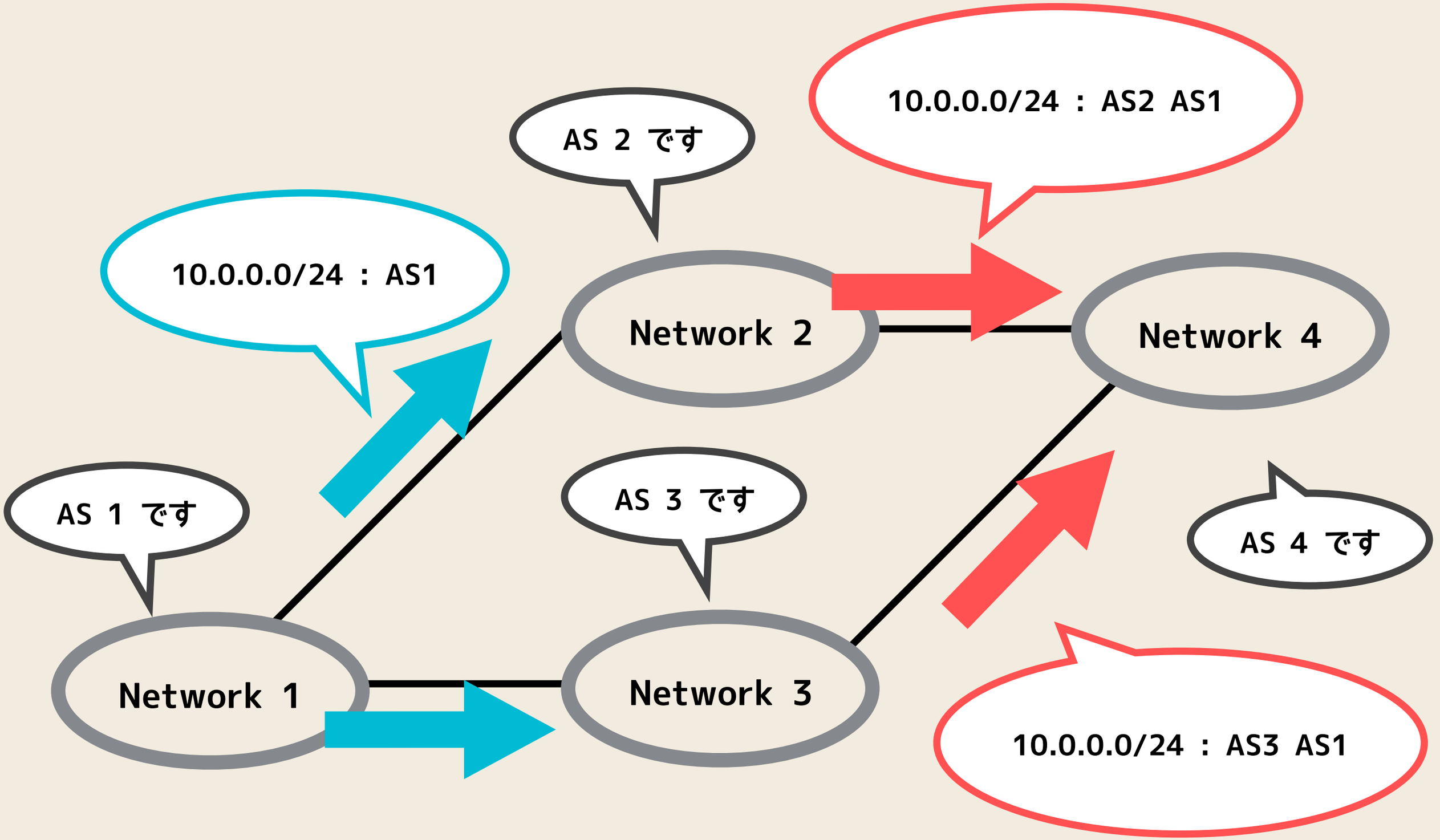
「経路」を交換する


10.0.0.0/24 は
Network 1 にあるよ
10.0.0.0/24 は
Network 2 経由の
Network 1 にあるよ
10.0.0.0/24
Network 1
Network 2
Network 3
Network 4
10.0.0.0/24 は
Network 3 経由の
Network 1 にあるよ

「経路」をたどってパケット転送



ネットワーク
= Autonomous System (AS)

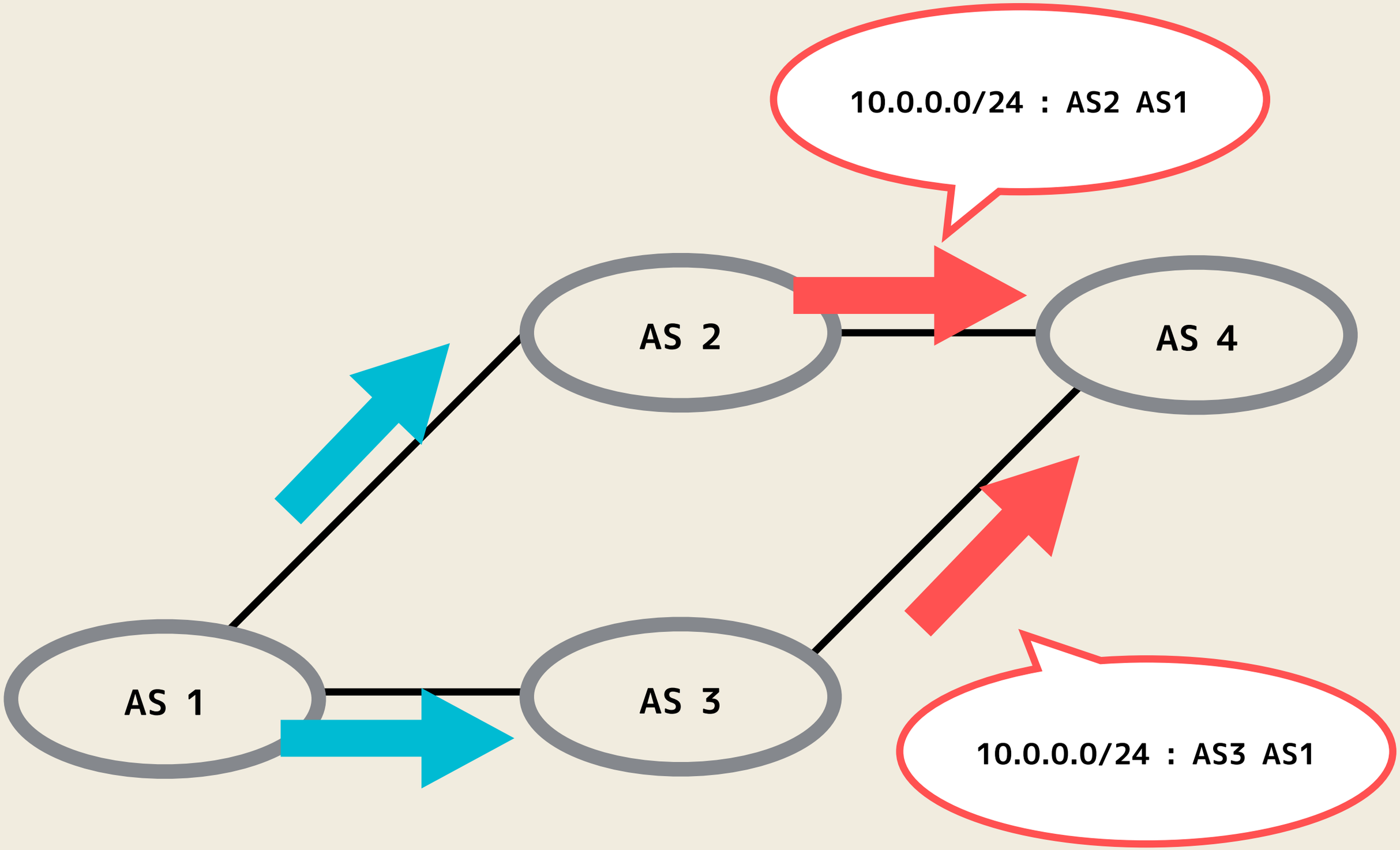


AS_PATH

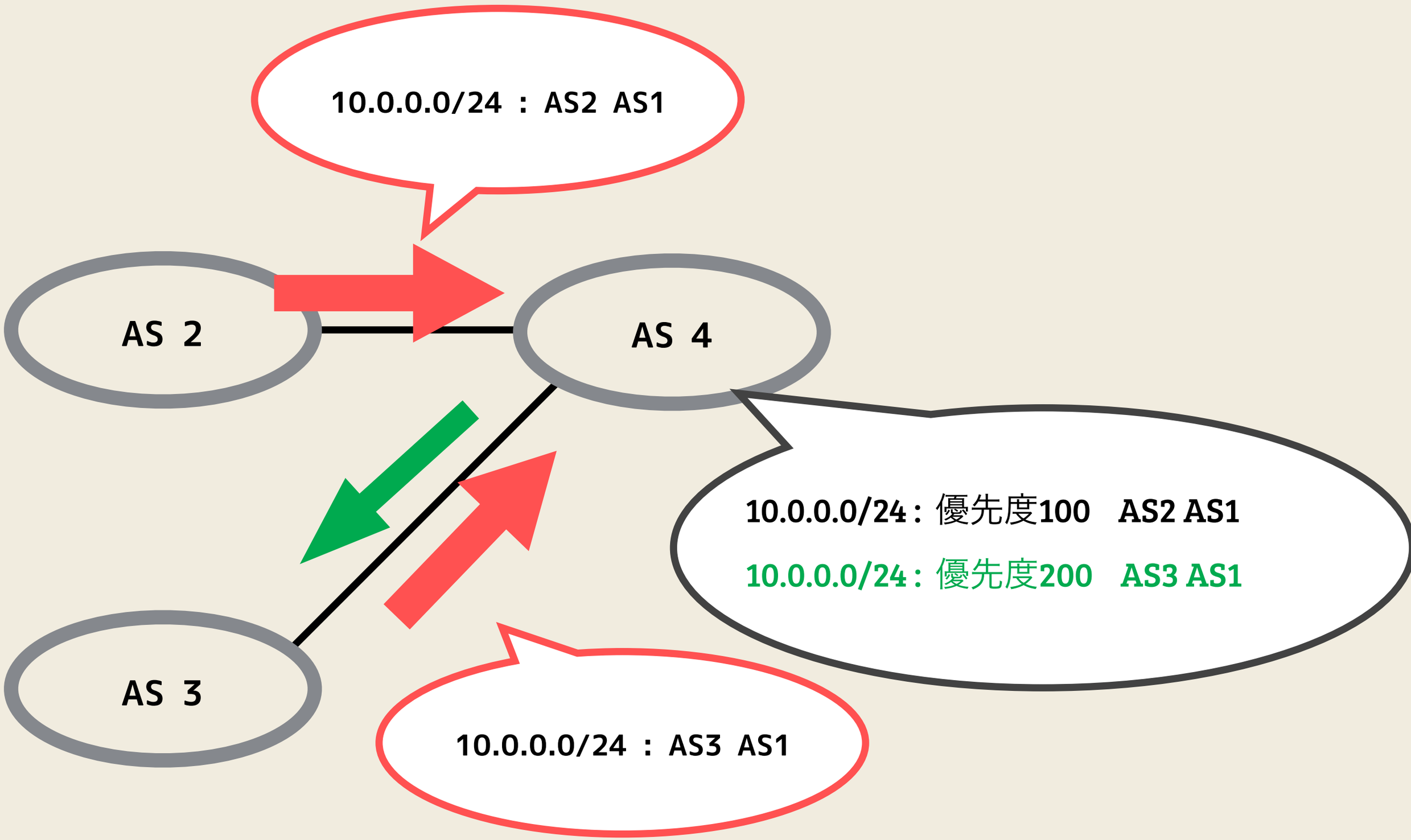
10.0.0.0/24 : AS2 AS1

- AS_PATH と呼ぶ
- AS_PATH が短い = 近い = そっちに転送したほうがよい
- AS_PATH のような経路の属性のことをパスアトリビュート と呼ぶ

AS_PATH が同じ長さだったら?



AS4 が受信時に優先度づけ

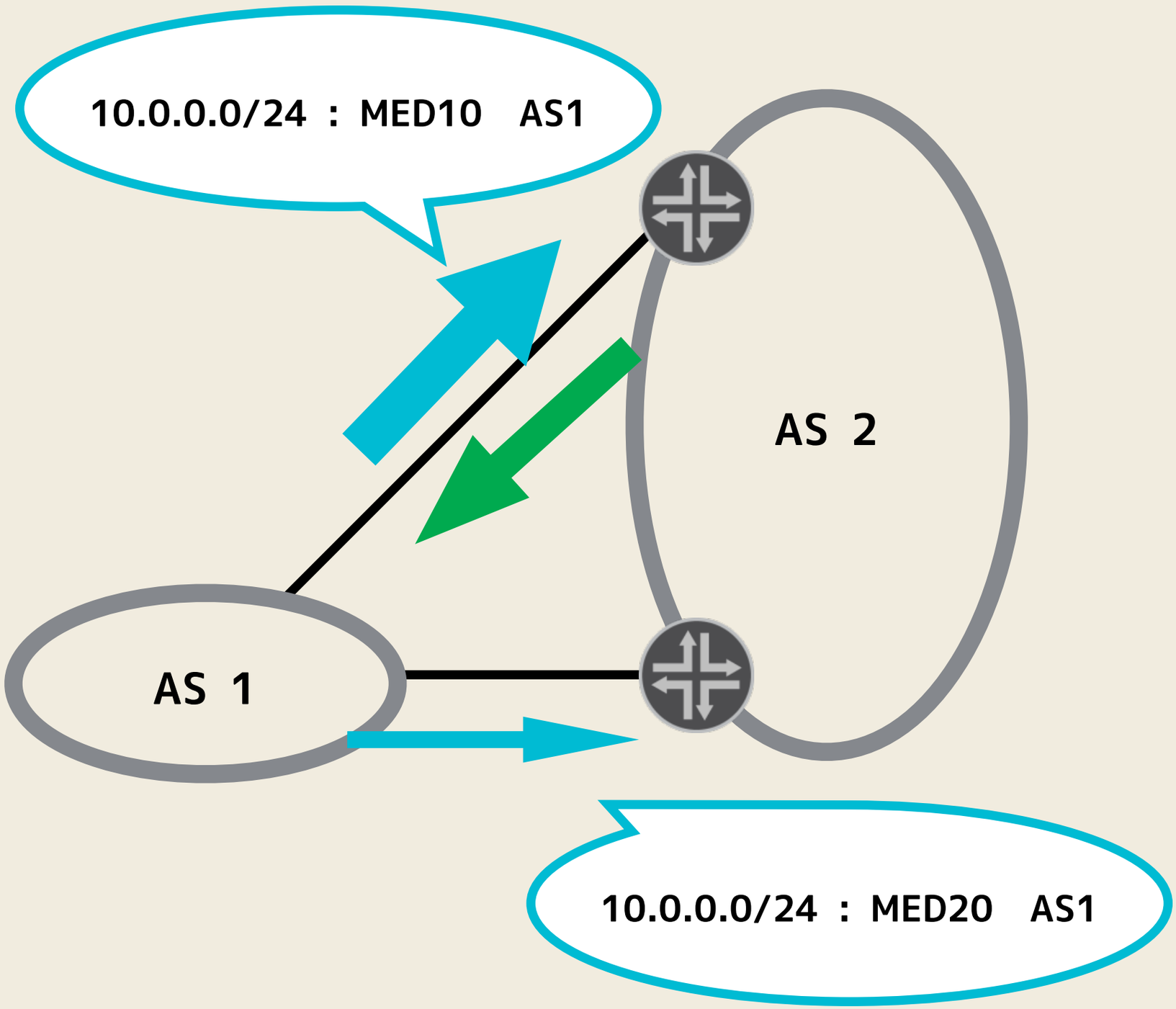


LOCAL_PREF
(Local Preference, LP)

10.0.0.0/24 : 優先度200　AS2 AS1

- LOCAL_PREF と呼ぶ
- LOCAL_PREF が大きい = 優先度高い = そっちに転送したほうがよい
- ほかのAS には転送されない。AS 内部でのみ有効

その他のパスアトリビュート
MULTI_EXIT_DISC (MED)

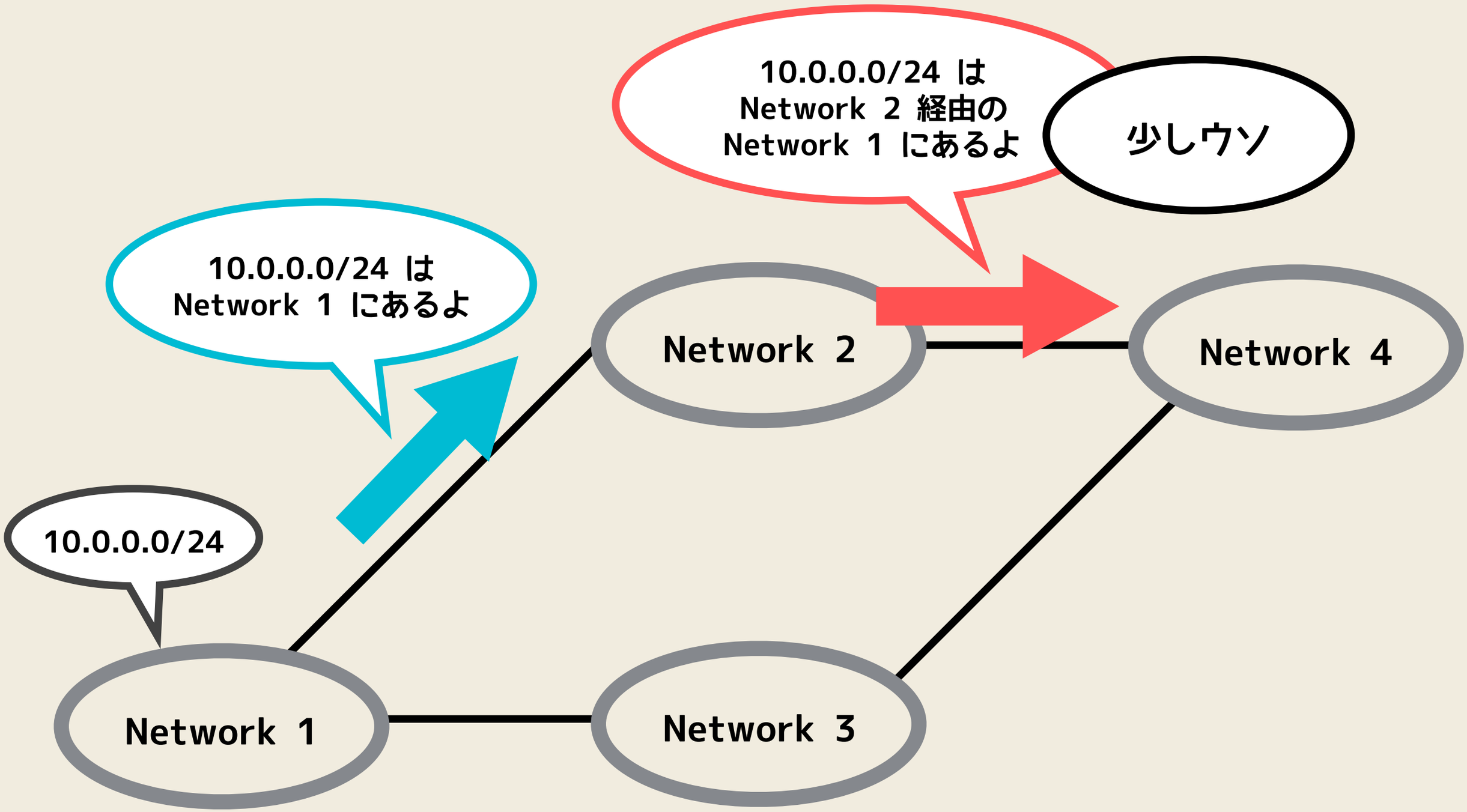


その他のパスアトリビュート
MULTI_EXIT_DISC (MED)

- MED と呼ばれることが多い
- MEDが小さい = 優先度高い = そっちに転送したほうがよい
- 同じAS から受け取ったMED だけが比較される
  - 例: AS2 が AS1 にパケット転送したいが、どちらのリンクに転送するか決めるとき
  - 「AS4 が AS2 かAS3 のどちらにパケット転送するか決めるとき」には無視される

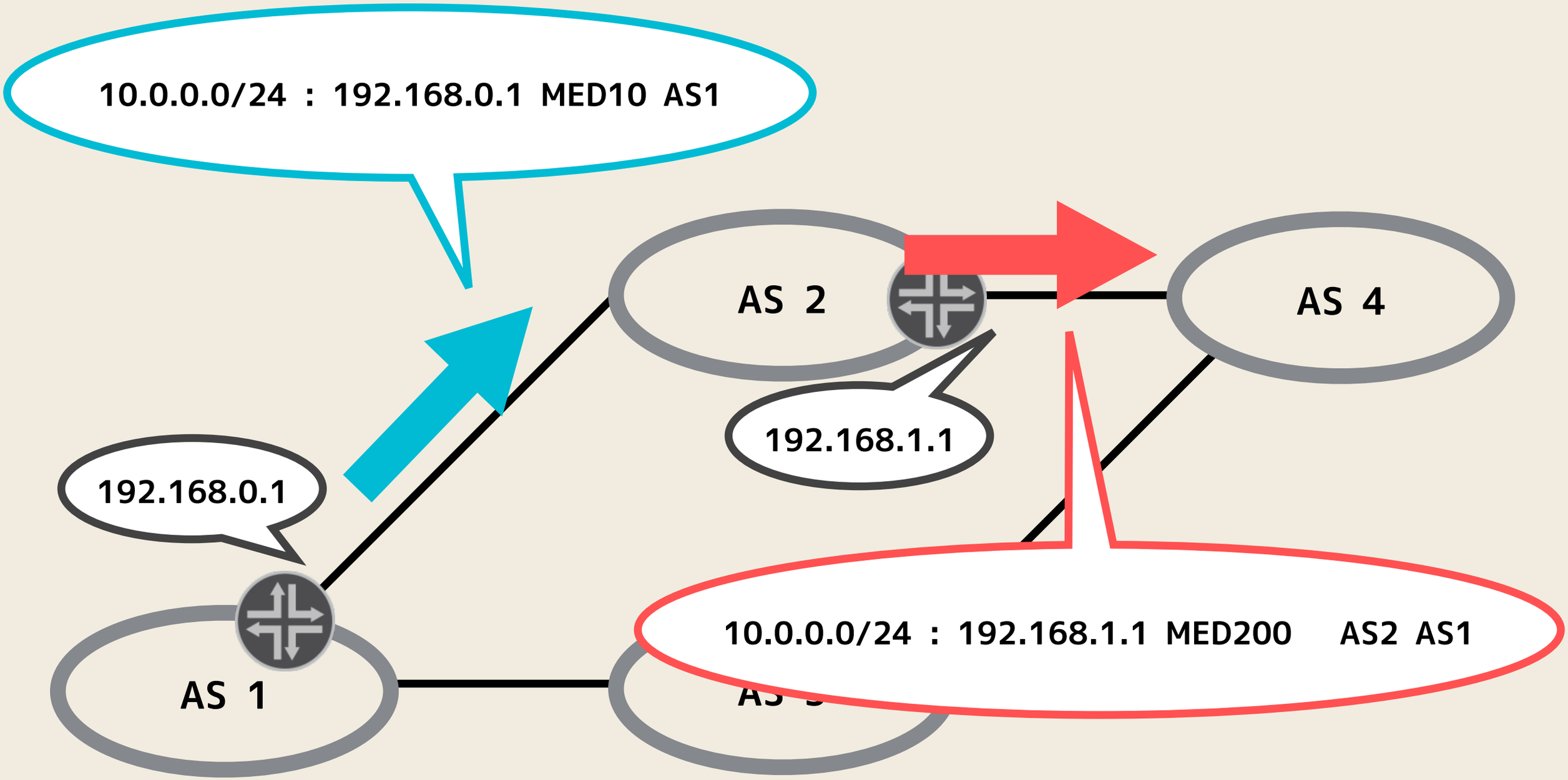
その他のパスアトリビュート

NEXT_HOP



その他のパスアトリビュート

NEXT_HOP



・実際は、接続しているルーターのIP アドレスを伝える

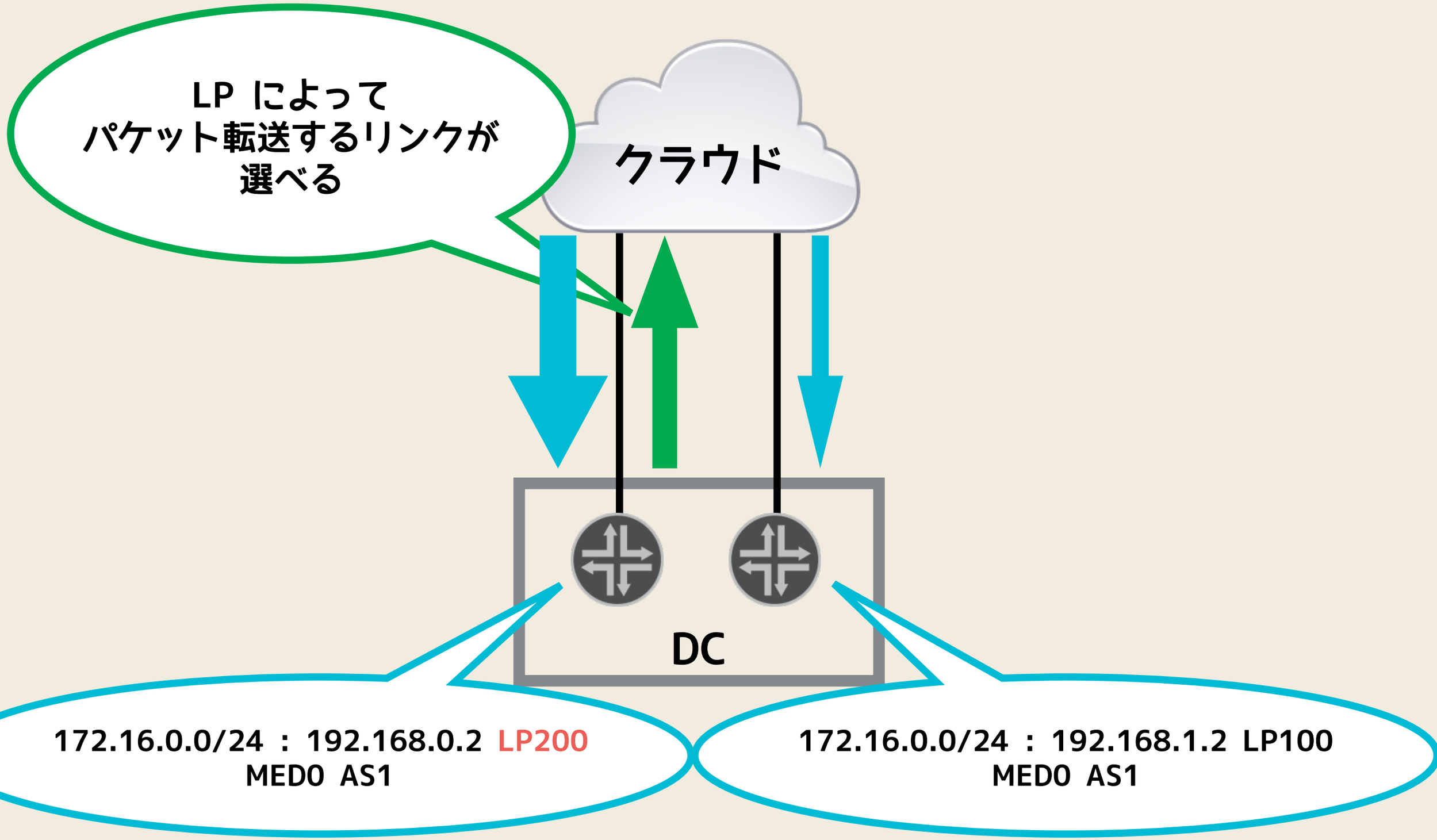
BGP まとめ

- インターネットを制御するためのルーティングプロトコル
- パスアトリビュートによって転送先 (NEXT_HOP) を優先づける
  - AS_PATH
  - LOCAL_PREF
  - MED
- パスアトリビュートは、経路送信時 / 受信時に変更できる

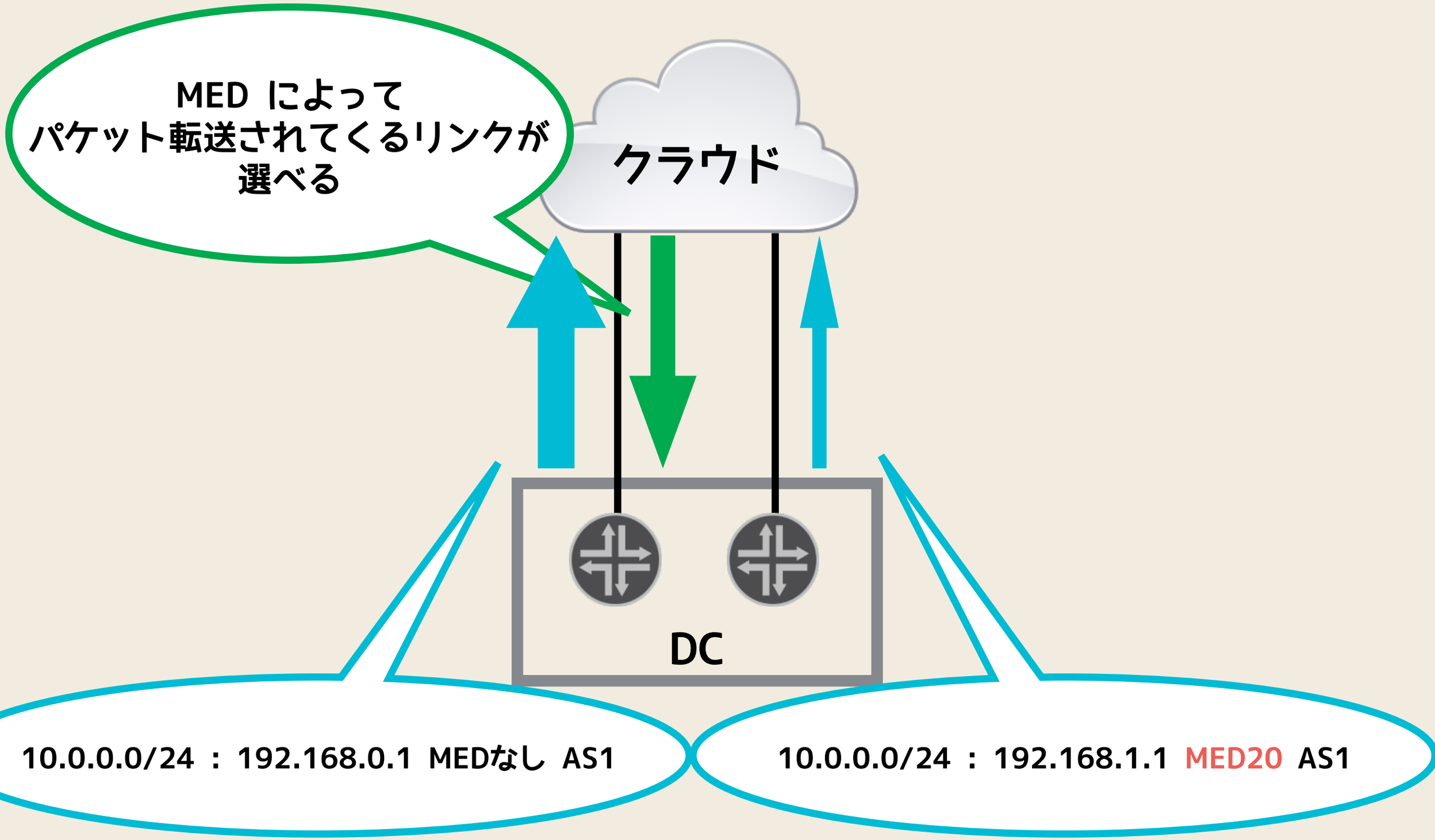
クラウドとの接続方法
(ルーティング編)

クラウドへの応用

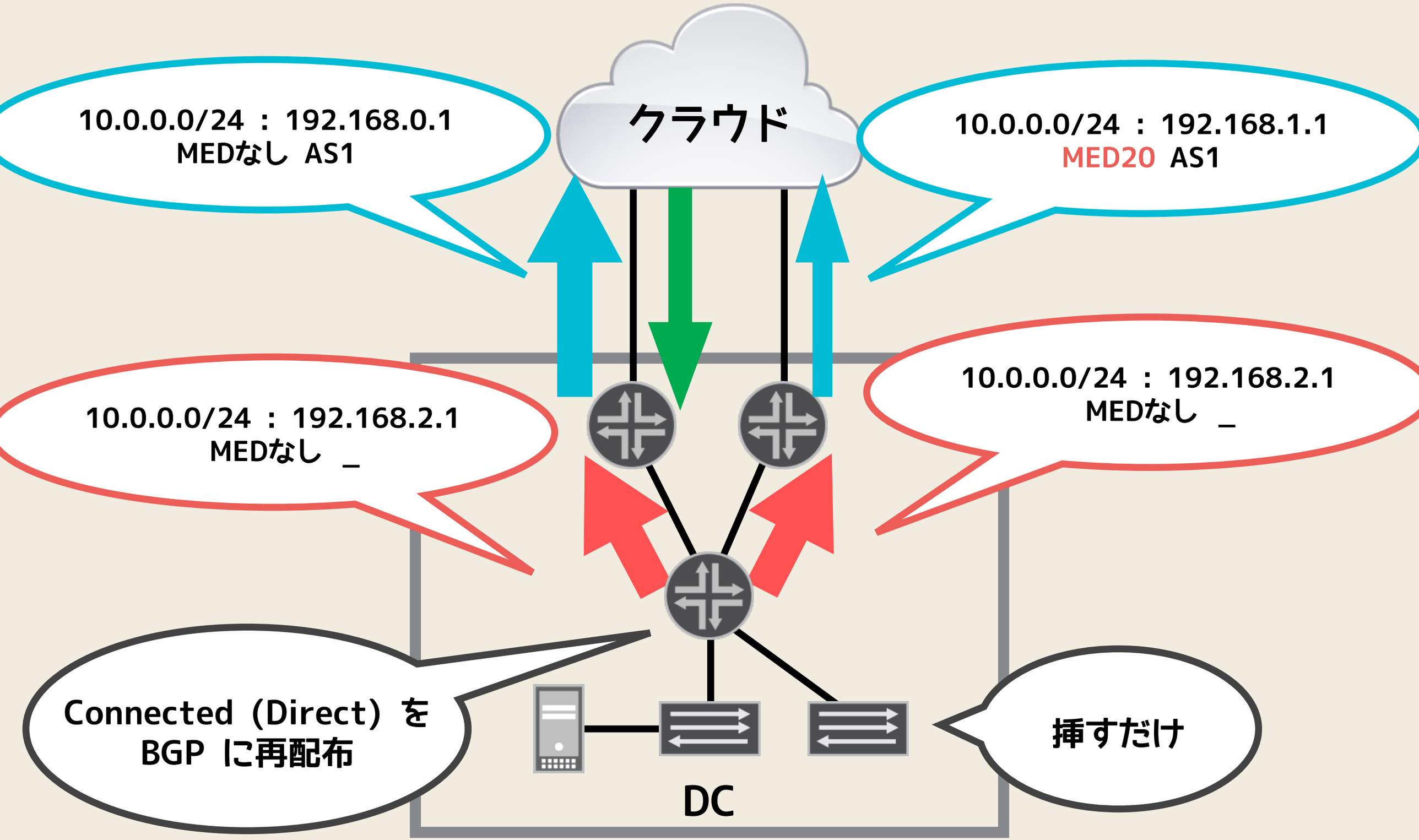
クラウド接続に応用すると



クラウド接続に応用すると



AS内部でもBGPを使う



クラウドへの応用 まとめ

- クラウドの設定を変えずに、パケット転送を制御できる
- ネットワークが増えても、送信する経路を増やすだけ

BGP によって、ネットワーク変更時に変えるべきポイントを減らす = スケールするネットワークが作れる

オススメ図書

インターネット ルーティング入門

- IP の基本
- ルータの設定
- OSPF
- RIP
- BGP
- MPLS
- 仮想ネットワーク
- etc…

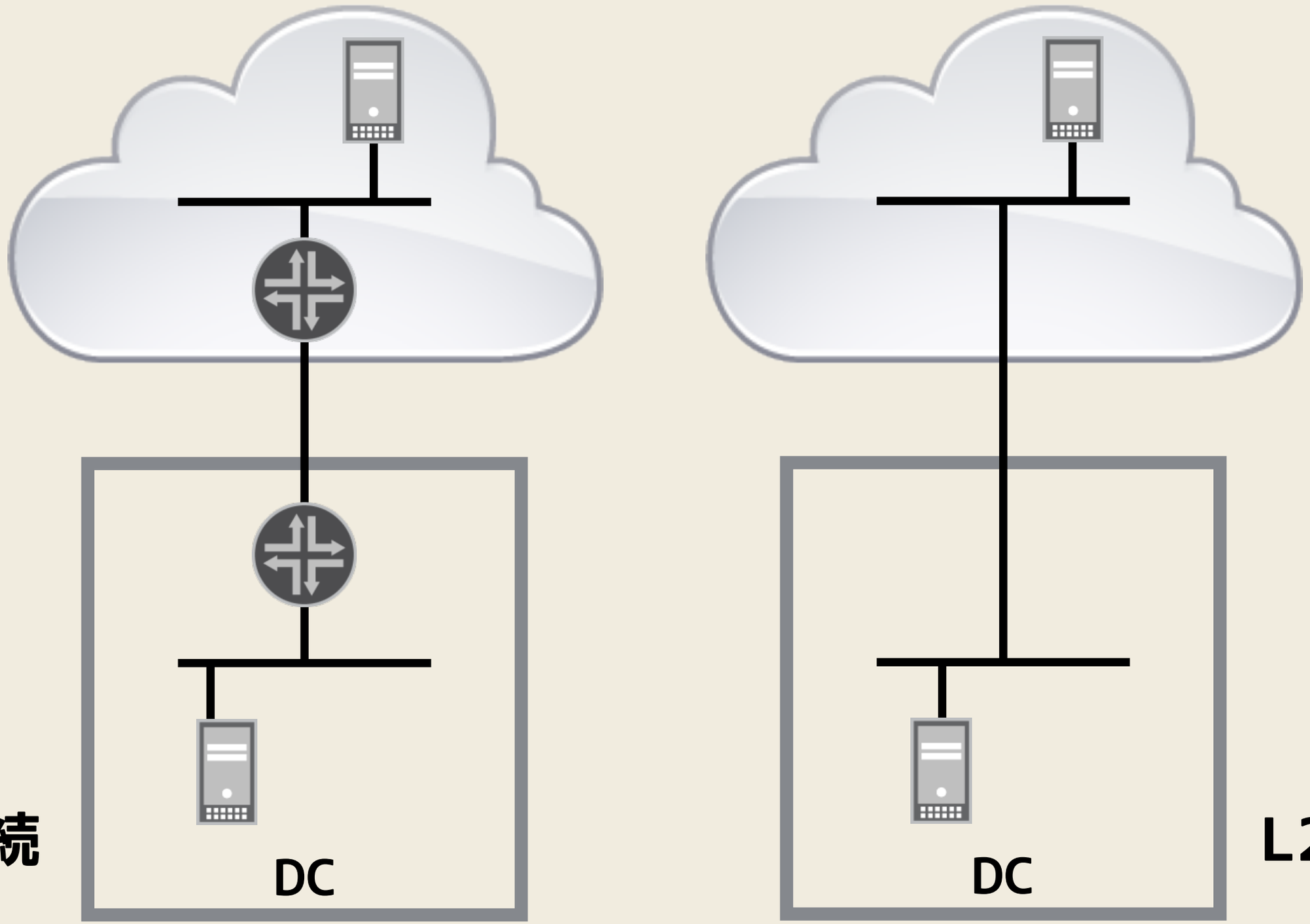
なるほど、BGP が便利
そうなのはわかった。

でも、クラウドとの物
理接続によるのでは？

Agenda

- クラウドとの接続方法
  - ルーティング 編
    - BGP
    - クラウドへの応用
  - オンプレミスとの接続 編
    - L3
    - L3VPN
    - L2VPN
  - 応用 編

クラウドのインスタンス、
どう見せたい？



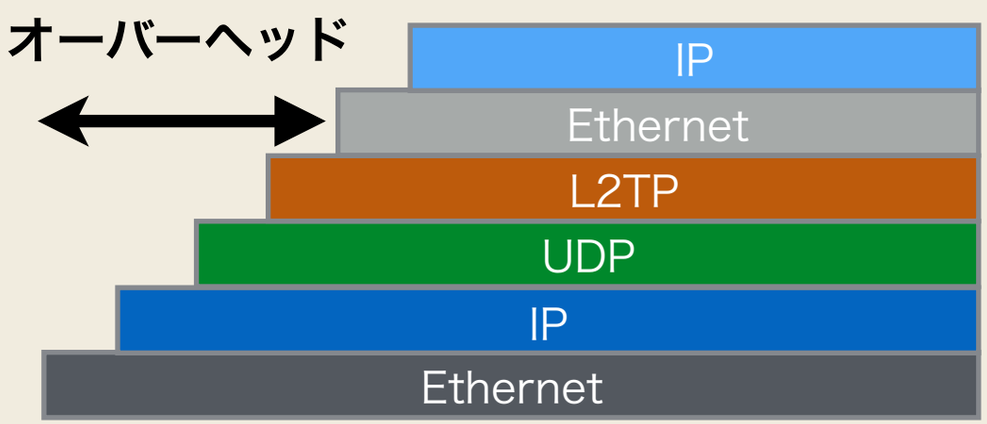
L3接続

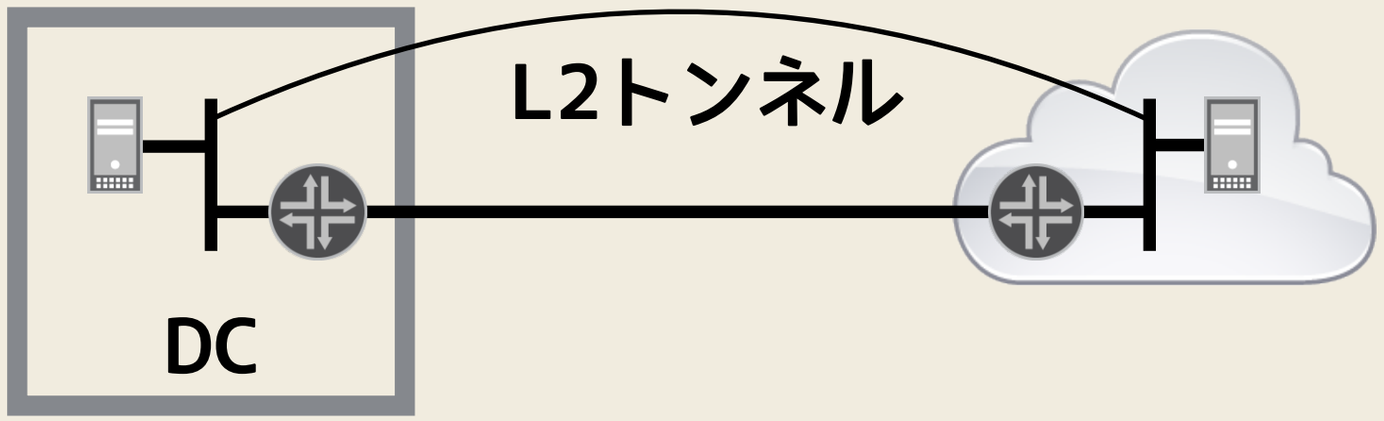
L2接続

L3接続 vs. L2接続

クラウドのゲートウェイが通常L3 なので…

- L3 → 作りやすい ○
  - シンプルな設計
  - オーバーヘッドなし
- L2 → L2TP などのトンネルプロトコル
  - オンプレミス側に終端装置が必要
    (PC で終端することもできる)
  - オーバーレイ
    - オーバーヘッドが大きい (50B~)
  - スループット問題





L2接続のモチベーション

- Live Migration
- 同じL2ネットワークのみで動く業務アプリ？
- PC接続環境としてはよいかもしれない
- メリット薄い
  - 技術的ハードルが高い
  - Ethernet 自体のリスク
    - スケールしないL2プロトコル
    - 脆弱なプロトコル構造
    - ループの可能性

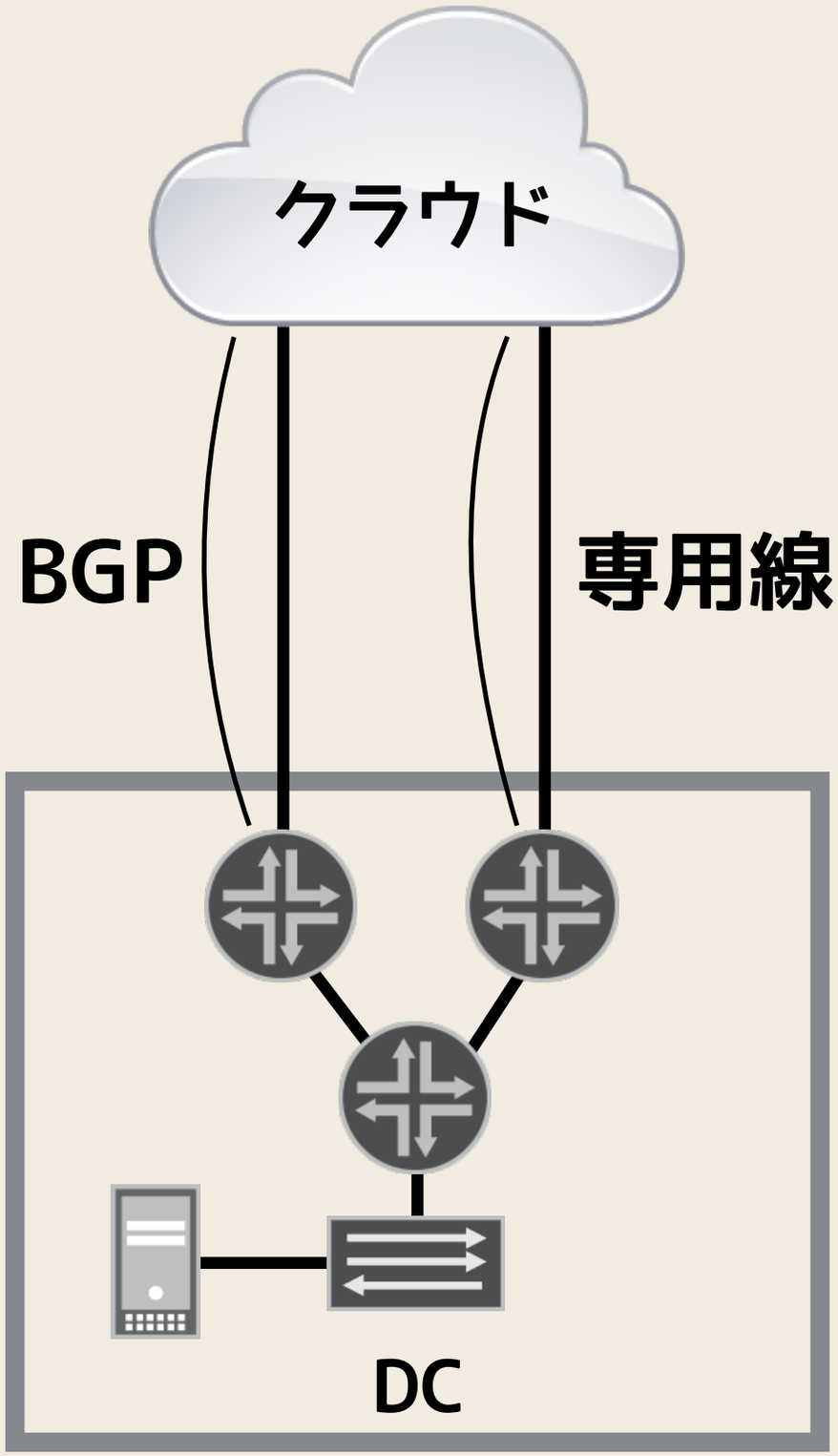
クラウド-オンプレミス 接続

	L3接続	L2接続
物理	専用線	L3接続 + L2TP or PPTP
VPN	IP-VPNサービス IPSec-VPN	L2TP or PPTP / 専用線 L2TP or PPTP / IP-VPNサービス L2TP or PPTP/ IPSec-VPN

クラウドとの接続方法
(オンプレミスとの接続編)

L3接続

L3接続 - 専用線



○ 専用線による安定した接続環境

- クラウド-オンプレミス間の直接
  BGP による細やかな経路制御

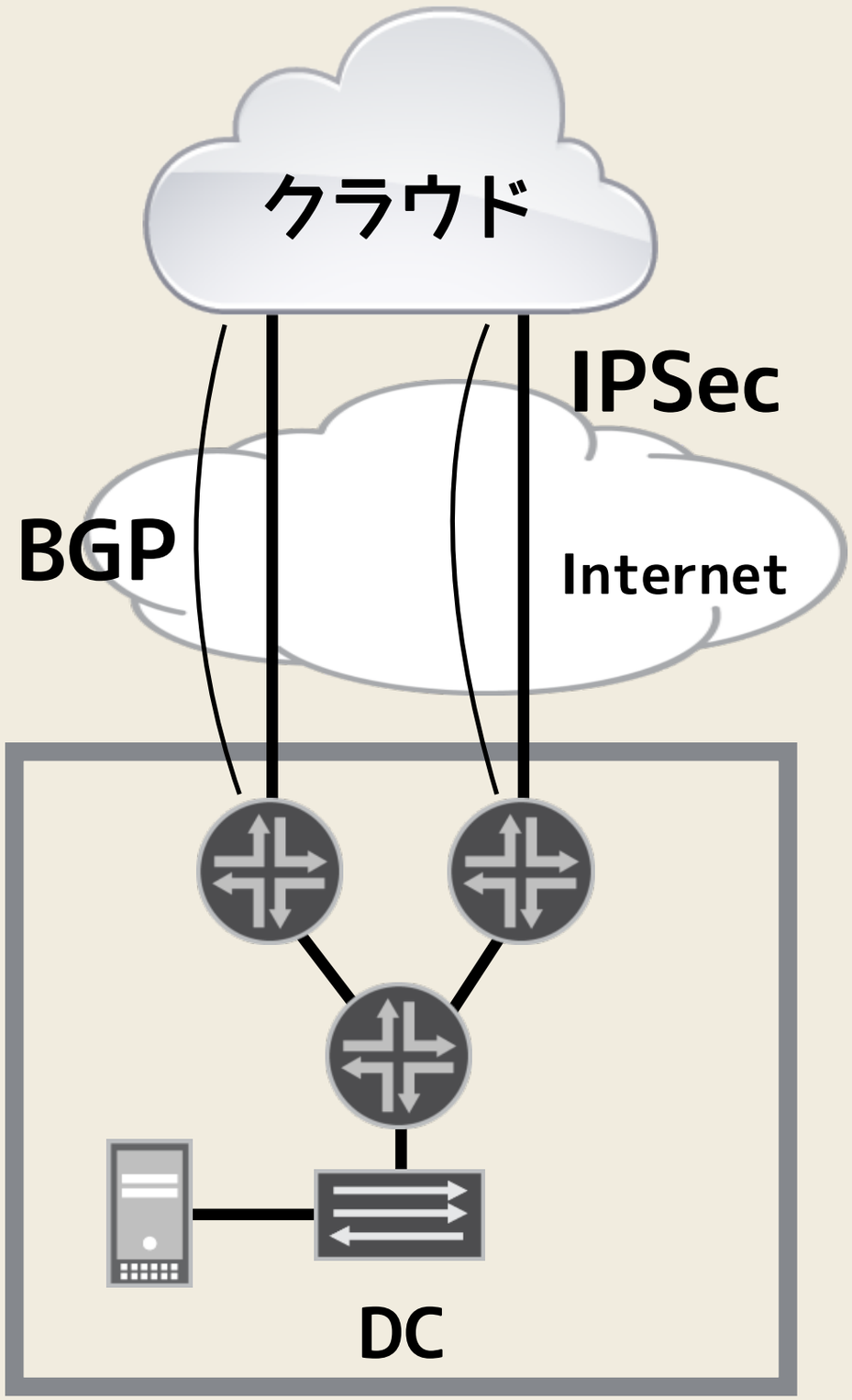
✕ 高コスト
(DCがクラウド接続サービスを
提供していれば安価に)

- クラウドトラフィックが多い場
  合向け

クラウドとの接続方法
(オンプレミスとの接続編)

L3VPN接続

L3VPN接続 - IPSec-VPN



✖ Internet の通信品質に依存

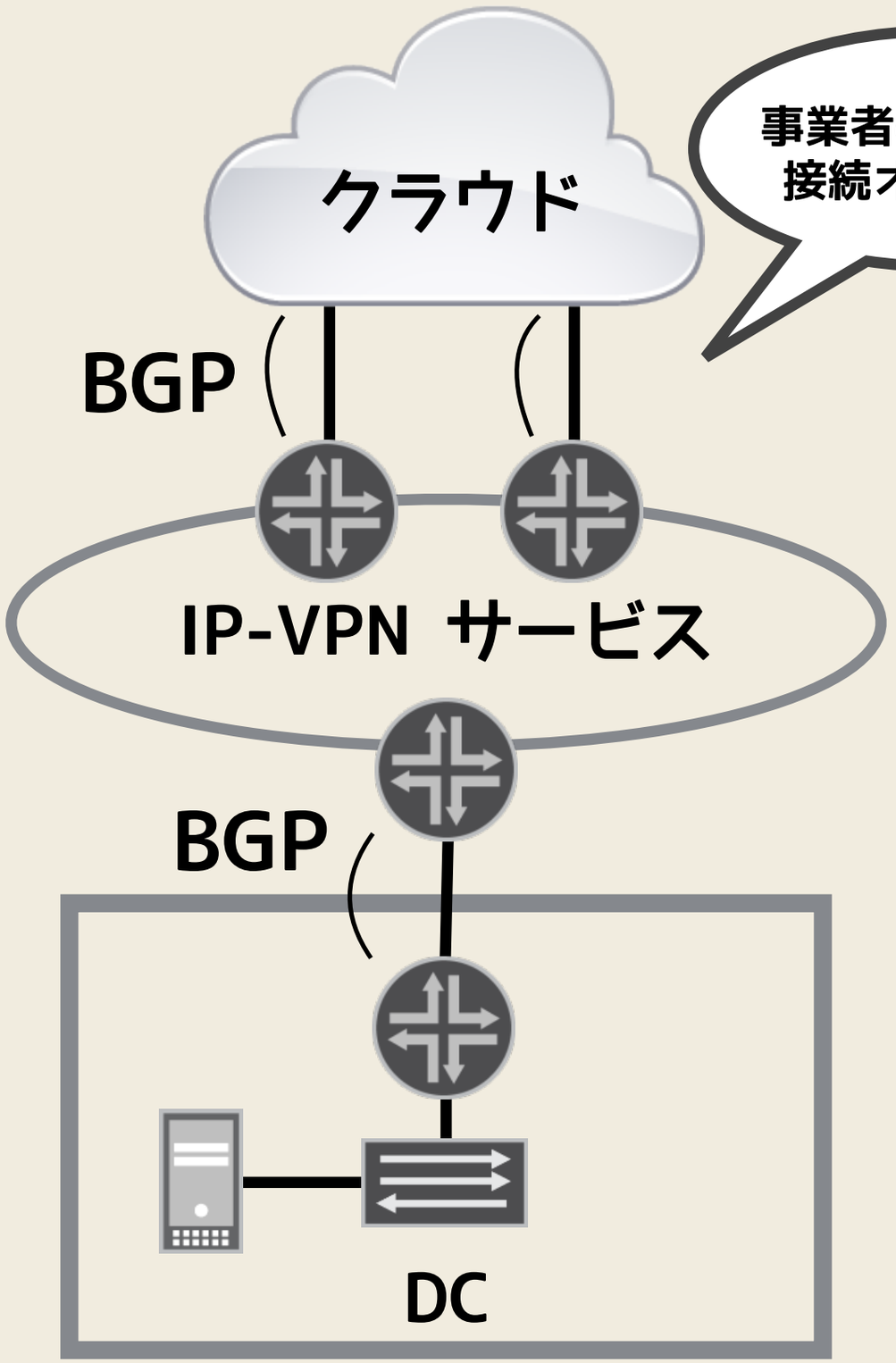
・ クラウド-オンプレミス間の直接
BGP による細やかな経路制御

○ 低コスト

・ スモールスタート向け

・ 他の方式のバックアップにも

L3VPN接続 - IP-VPNサービス



- VPN による良好な通信品質
  - クラウド-オンプレミス間接続はVPN事業者に任せられる
- 低コスト
  - 既にIP-VPN サービスを利用している場合向け

クラウドとの接続方法には、多数のオプションがある

- クラウドのトラフィック量
- 要求される、ネットワークの安定度
- トラフィックコントロールの必要性
- オンプレミス拠点はどこか？ (オフィス？DC？)
  - そこにクラウド事業者 や VPN事業者 がいるか？
- VPN サービスを既に利用しているか？
- VPN サービス-クラウドの相互接続はあるか？
- 予算
- …

要件にあう接続方式と要素技術を選んで、ハイブリッド構成をつくる

クラウド+オンプレミスの
ハイブリッド構成に
したときの、運用まで
イメージしてください

ハイブリッド構成の現実

- 断時間が許されない
- 部署ごとにネットワークを分離したい
- DR 用なのでコストをかけられない
- クラウドに持っていけないハードウェア、ソフトウェアがある
- オンプレミス ネットワークが複雑

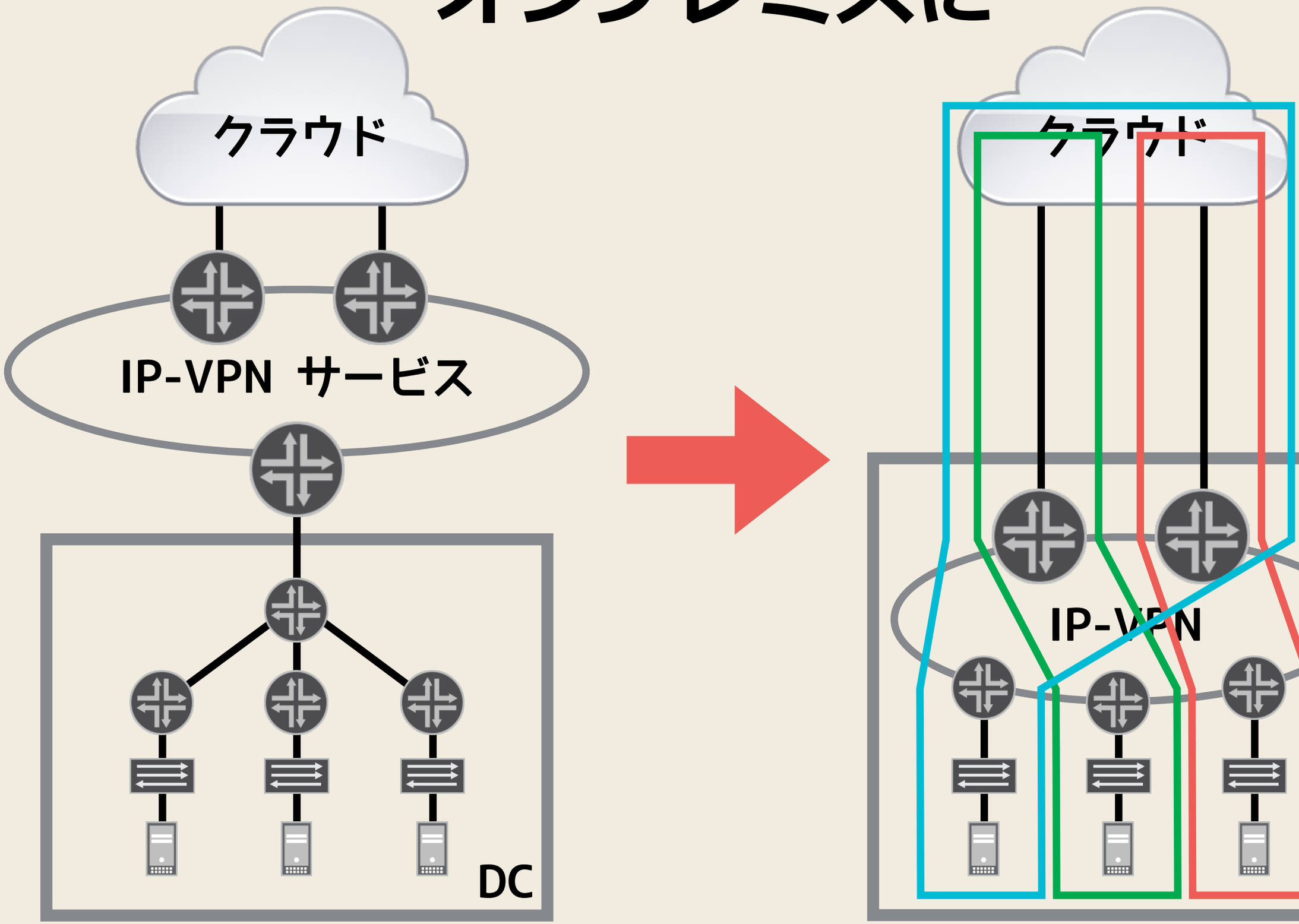
要件やオンプレミス側の事情により、クラウドの柔軟性が活かせない。全体として運用がラクにならない

→ オンプレミス側にも手を入れるべき局面

Agenda

- クラウドとの接続方法
  - ルーティング 編
    - BGP
    - クラウドへの応用
  - オンプレミスとの接続 編
    - L3
    - L3VPN
    - L2VPN
  - 応用 編

アイデア: クラウドの柔軟性を
オンプレミスに



IP-VPN に必要な技術

- MPLS
  - LDP
  - MPLS-VPN
    - Inter-AS Option-A
    - Inter-AS Option-B
- MP-BGP

→ 複雑だが、オンプレミス側でも実現できる

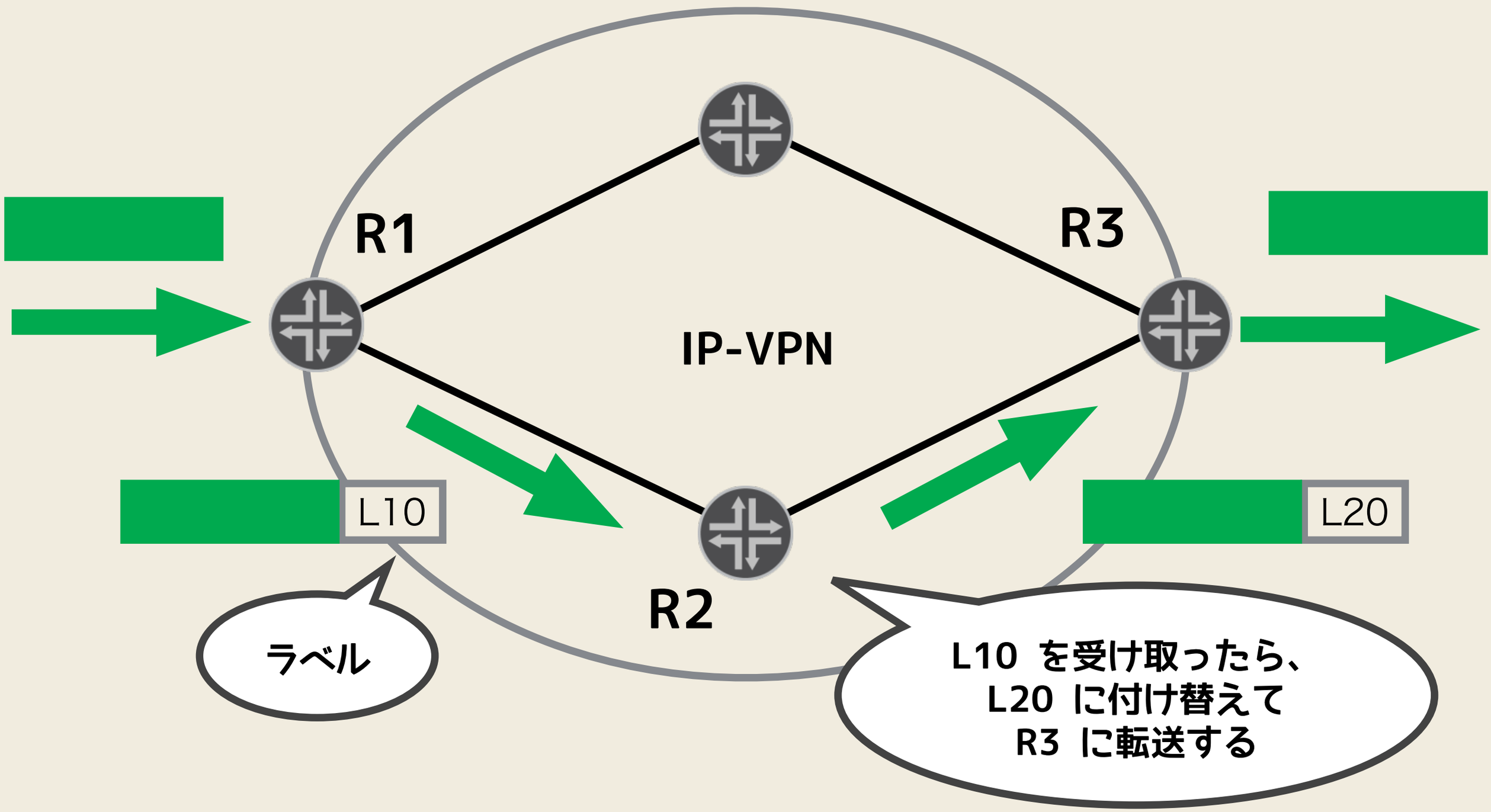
クラウドとの接続方法
(応用編)

MPLS

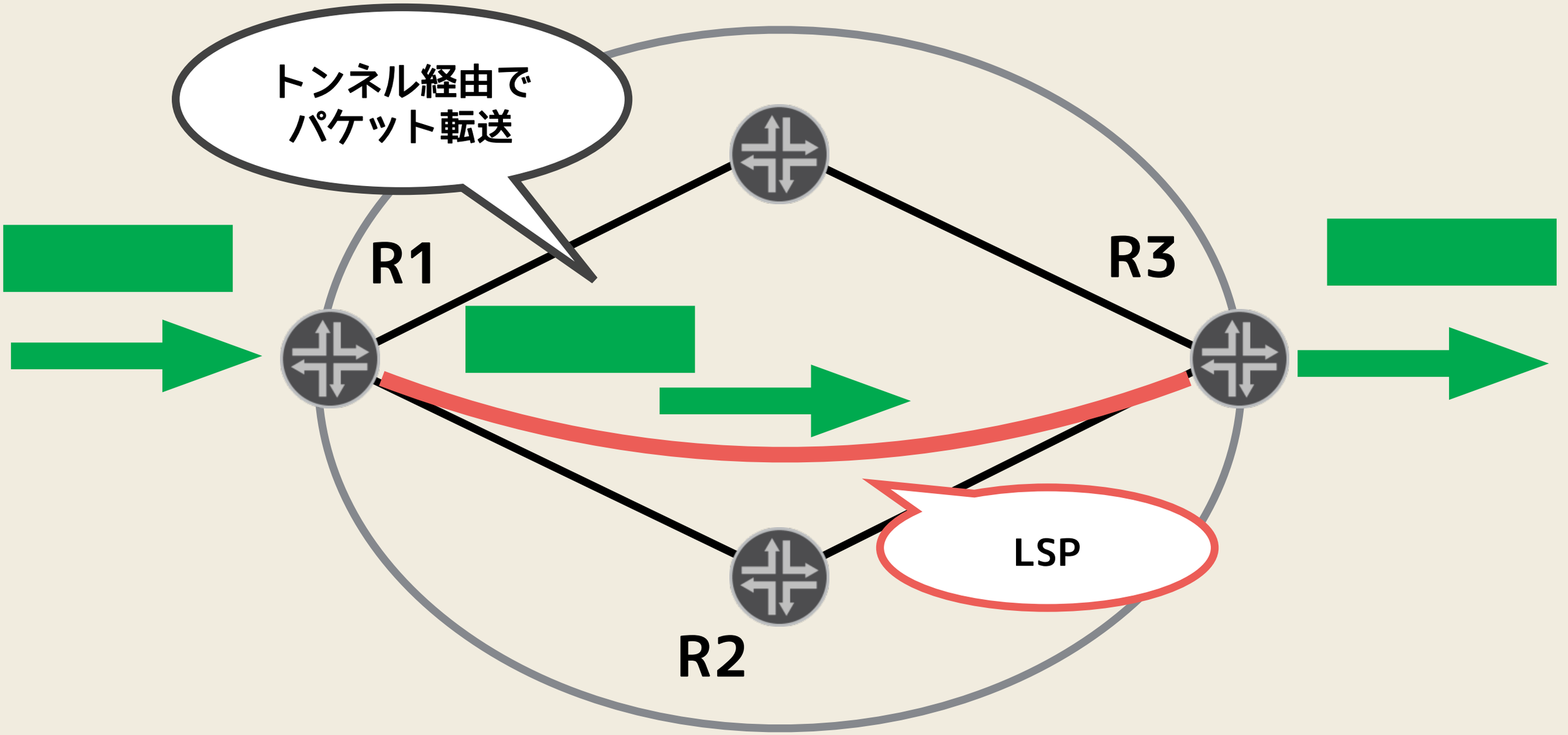
MPLS

- Multi Protocol Label Switching
  - RFC3031, 3032
  - AS内ルーター間をトンネル(LSP) で接続し、VPN、TE を実現する技術
- LDP
  - RFC3036 (Label Distribution Protocol)
  - LSP を張るためのラベル情報を交換するプロトコル

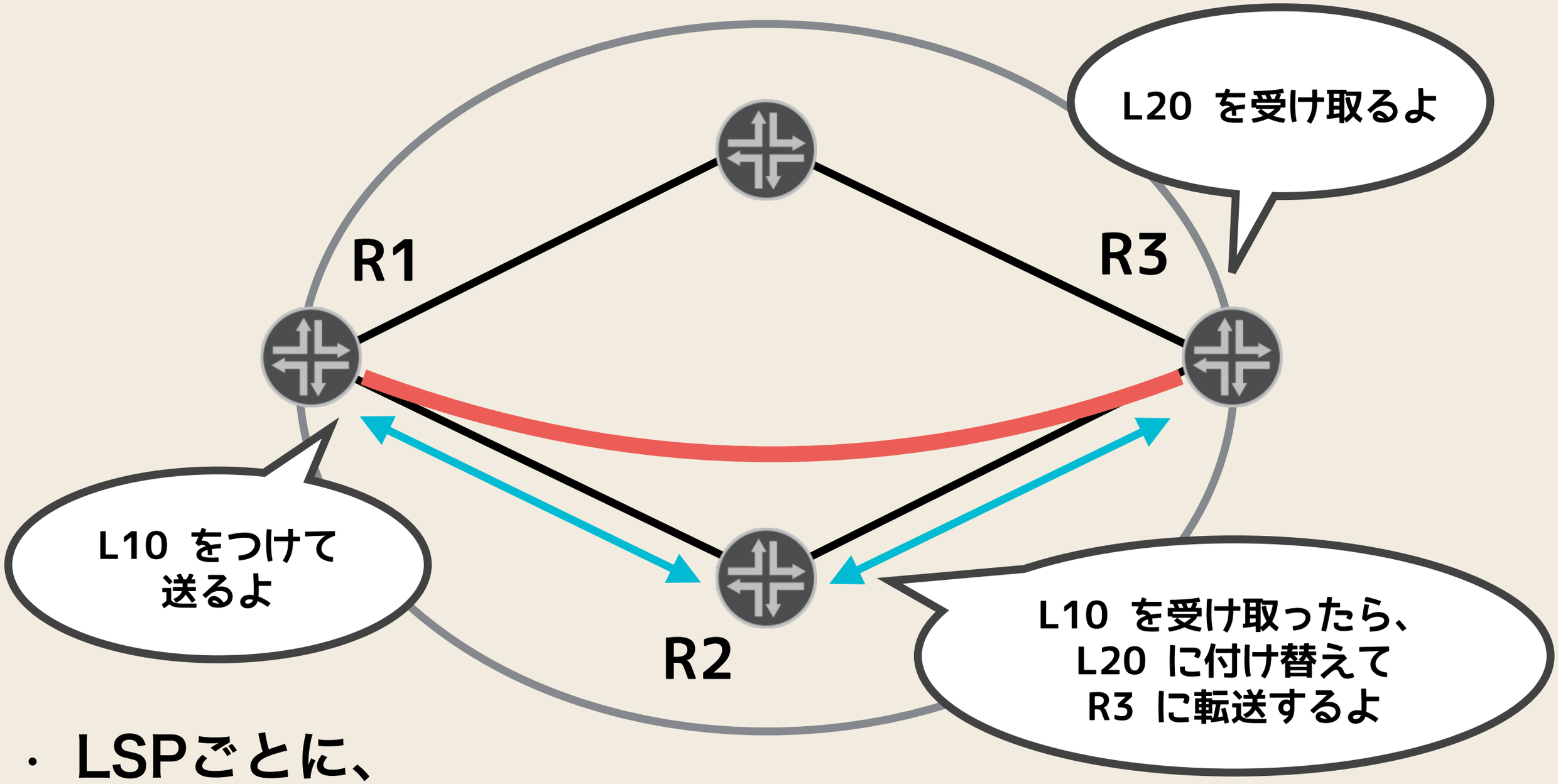
MPLS



MPLS

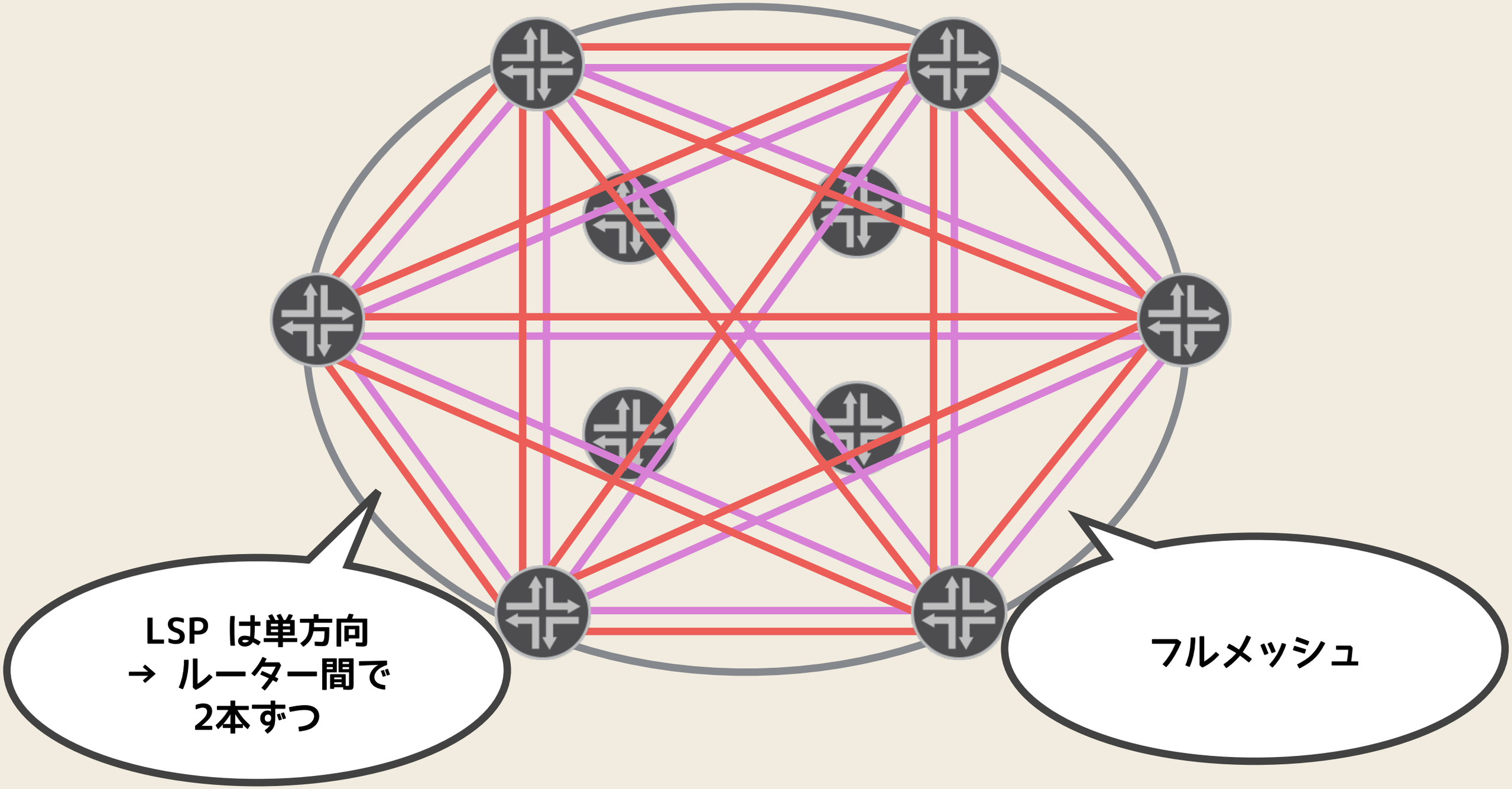


LDP



- LSPごとに、
  ラベルに関する情報を交換しておく

MPLS



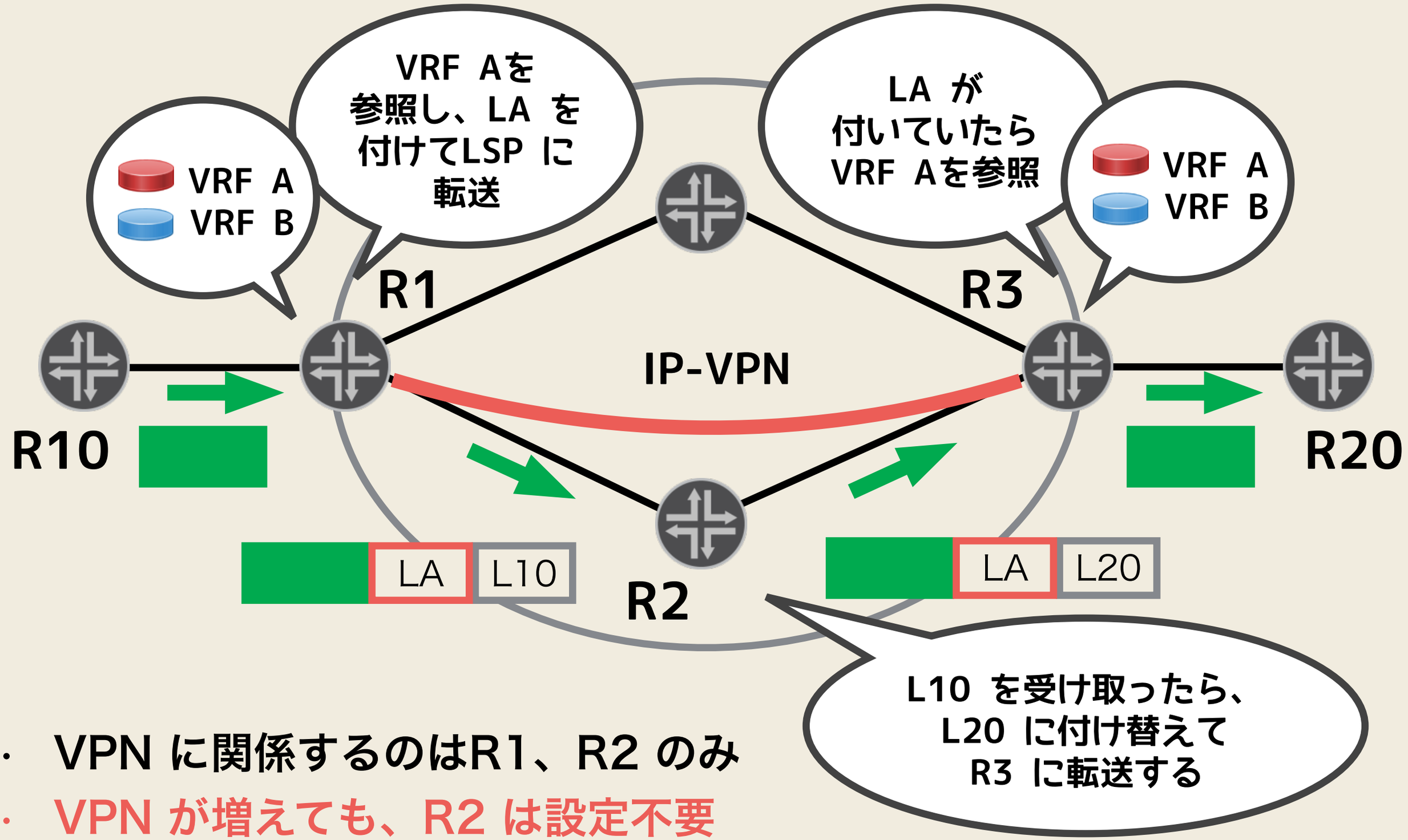
クラウドとの接続方法
(応用編)

MPLS-VPN

MPLS-VPN

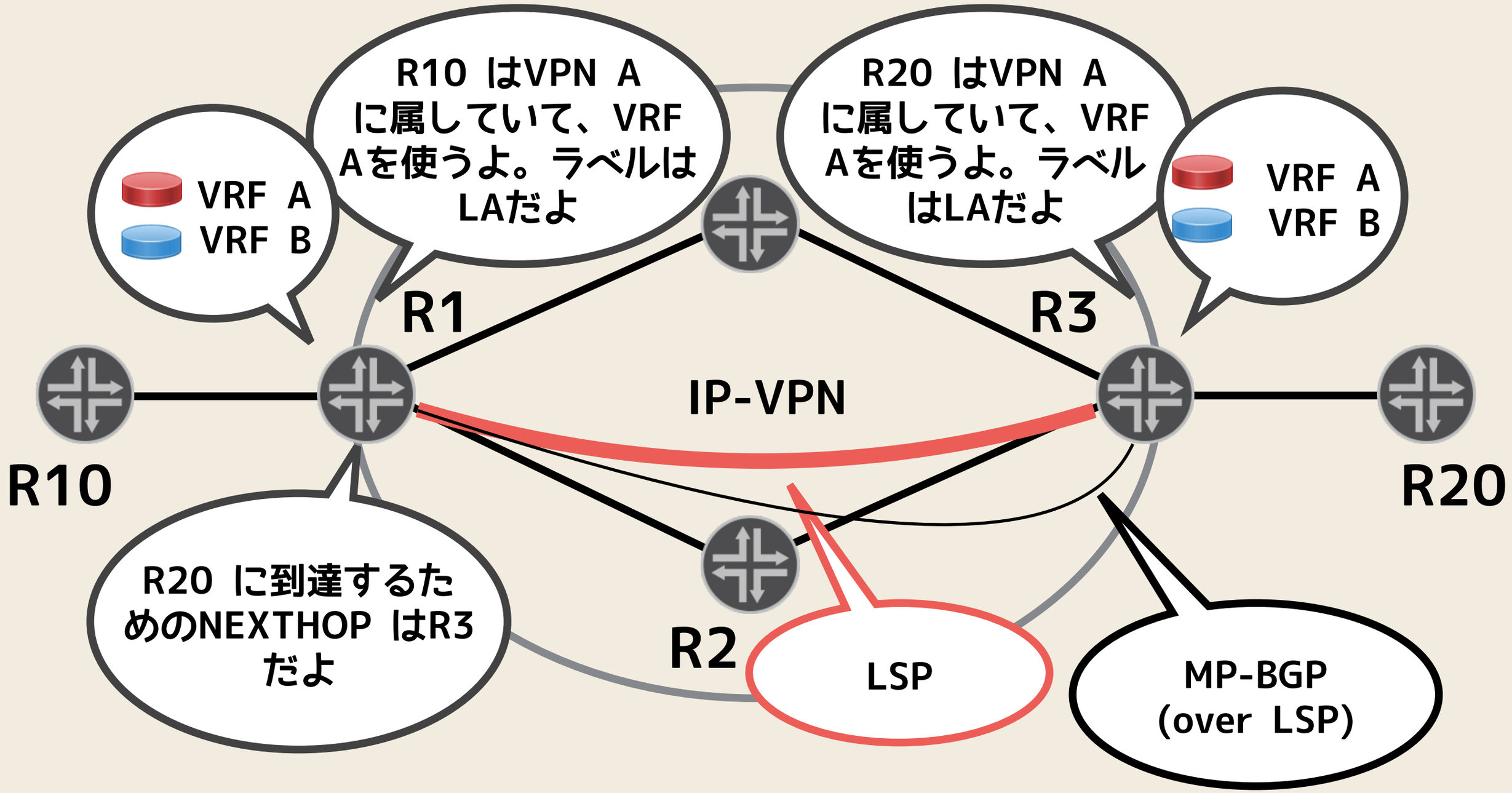
- MPLS-VPN
  - RFC4364
    (BGP/MPLS IP Virtual Private Networks)
    - MPLS を用いてVPN を実現する技術
      - AS(事業者)間の経路交換について、オプションA~C がある
- MP-BGP
  - RFC4760(Multiprotocol Extensions for BGP-4)
  - さまざまなネットワークレイヤーの経路情報を交換できるよう、BGP-4 を拡張した
    - VPN、IPv6

MPLS-VPN



- VPN に関係するのはR1、R2 のみ
- VPN が増えても、R2 は設定不要

MP-BGP



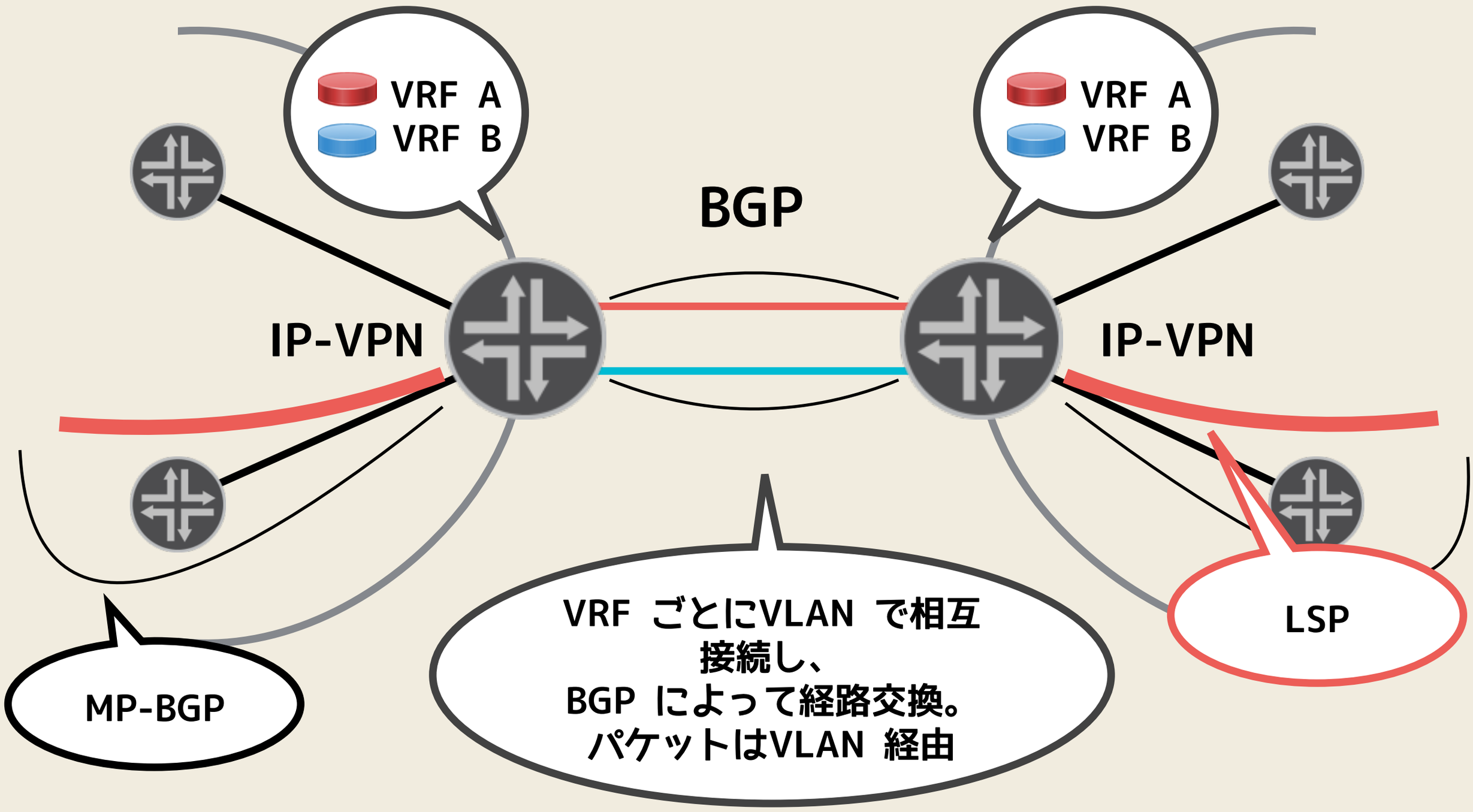
- VPN ごとに
  ラベルに関する情報を交換しておく

クラウドとの接続方法
(応用編)

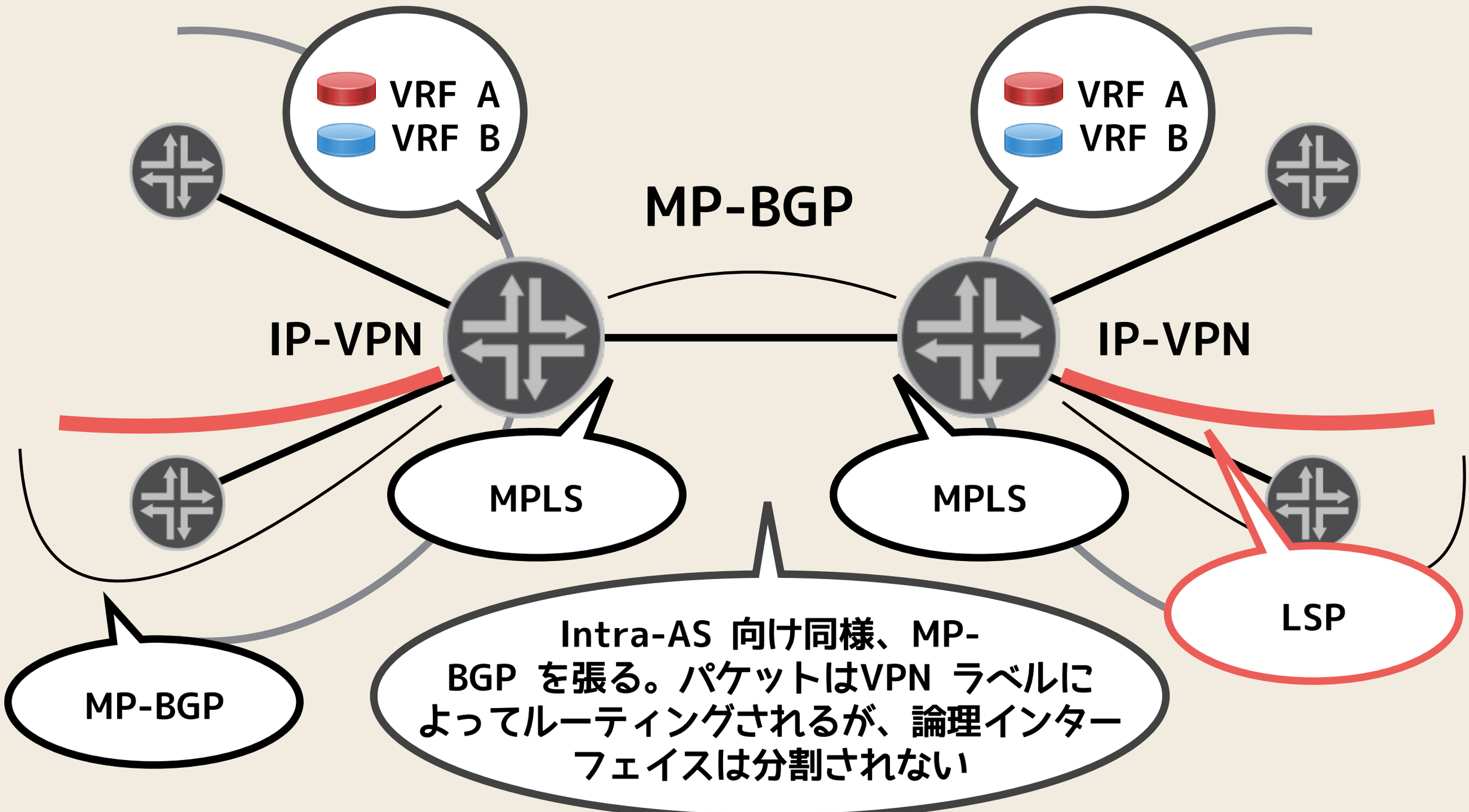
MPLS-VPN

Inter-AS

MP-BGP Inter-AS Option-A



MP-BGP Inter-AS Option-B

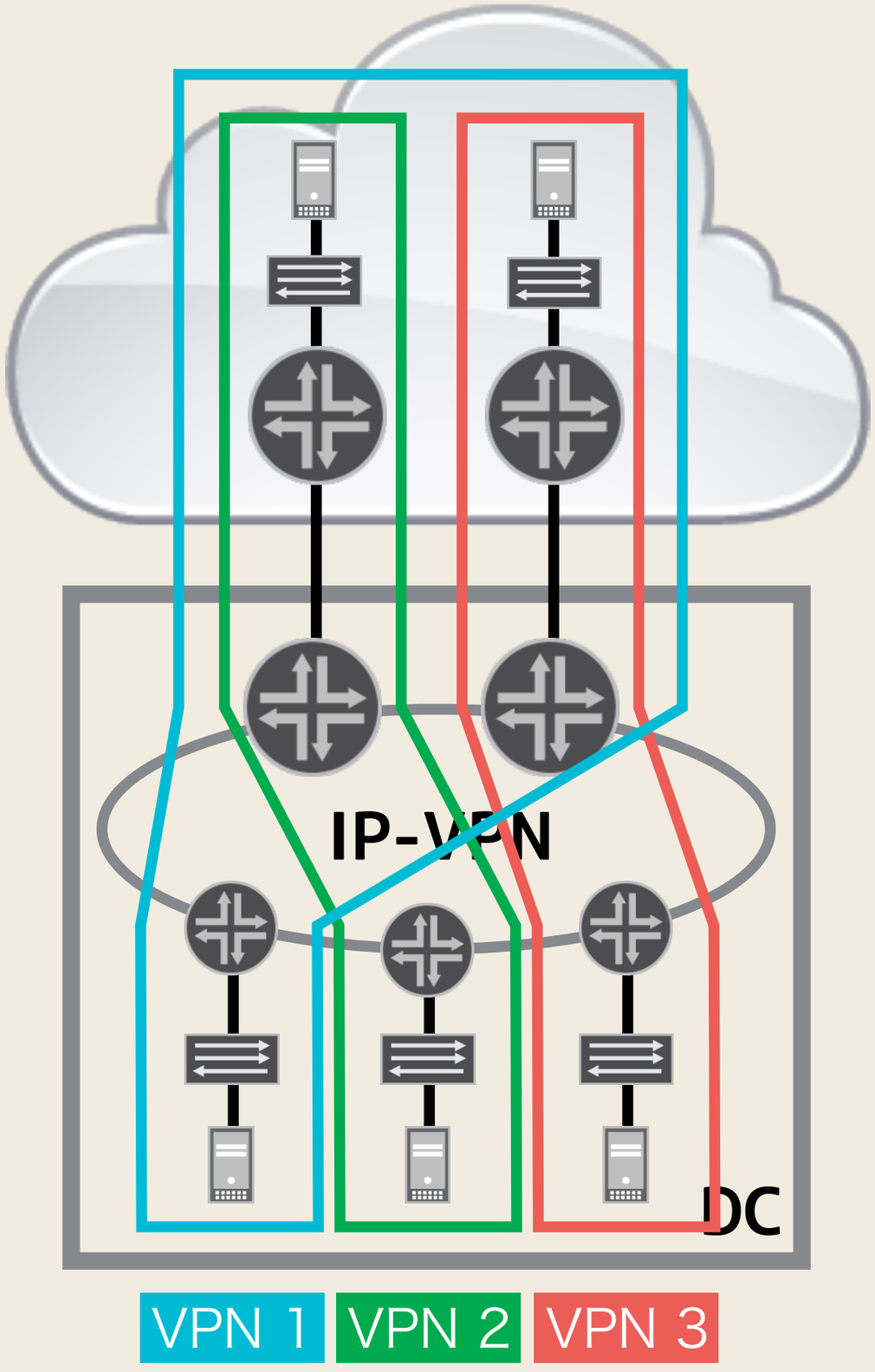


MP-BGP Inter-AS
Option-A vs. Option-B

・どちらかが特に優れる、ということはない。適材適所

	Option-A	Option-B
VPN ごとにトラフィック管理ができる	○	✕
Route Target (VPN 識別子) を 各AS で決められる	○	✕
トラフィックバランスやパス最適化が可能	○	✕
VPN 数に対してスケーラブル	✕	○
MPLS を対外インターフェイスで動かさな くてよい (高信頼モデルが不要)	○	✕

クラウド + MPLS-VPN



- ○ スケーラブルなVPN、DC 内のVLAN削減
- ○ 単純なデータプレーン / リーンコア

- ✖ MPLS によるオーバーレイ
- ✖ ネットワーク機能(NAT、FWなど) を挟みにくい

クラウドの柔軟性と
BGP、MPLS-VPN は相性がいい。

SDN まで行かずとも、
要素技術をうまく選ぶことで
クラウドネイティブなネットワー
クを作ることができる

まとめ

- コントロールプレーンの話
  - クラウド-オンプレ接続にBGP が使われ始めている
  - 動的プロトコルはスケールするから
  - BGP 入門編
- データプレーンの話
  - 専用線 or VPN サービス or IPSec-VPN 構成の紹介
- 社内MPLS-VPN の可能性

Questions ?