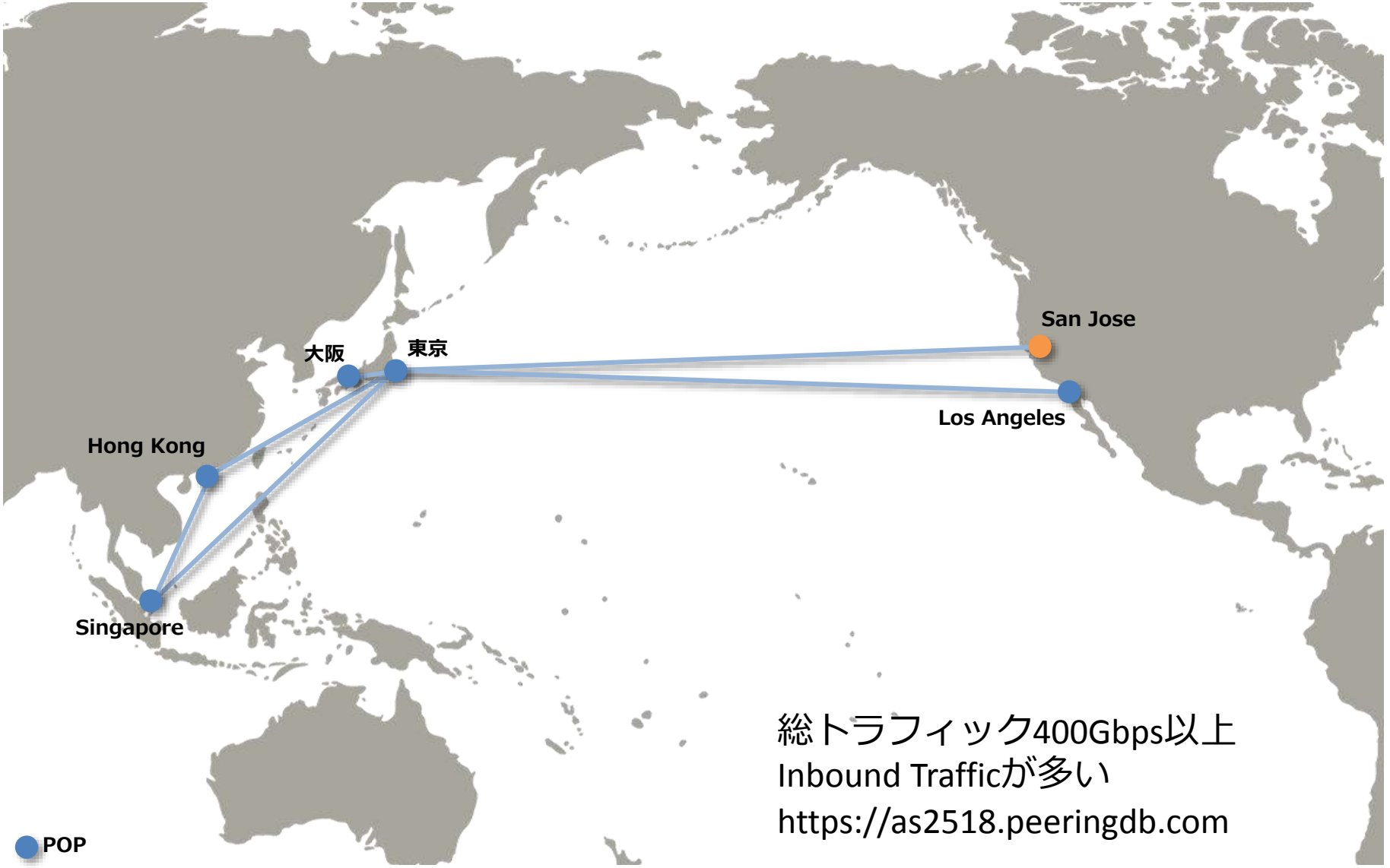
ISPのトラフィック制御と BGPコミュニティの使い方

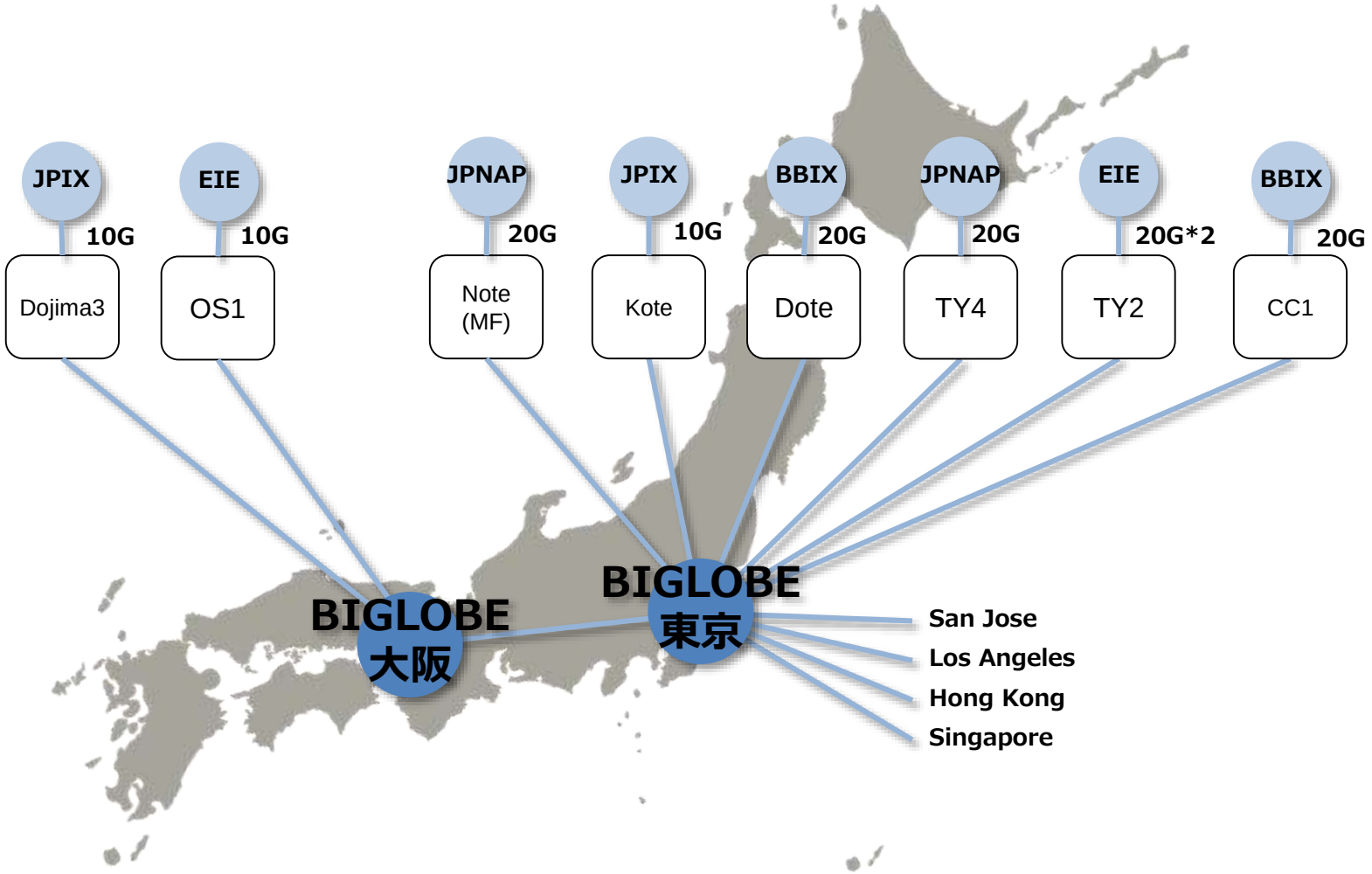
BIGLOBE Inc.
川村 聖一

はじめに：BIGLOBEのネットワーク



総トラフィック400Gbps以上
Inbound Trafficが多い
https://as2518.peeringdb.com

はじめに：BIGLOBEのネットワーク



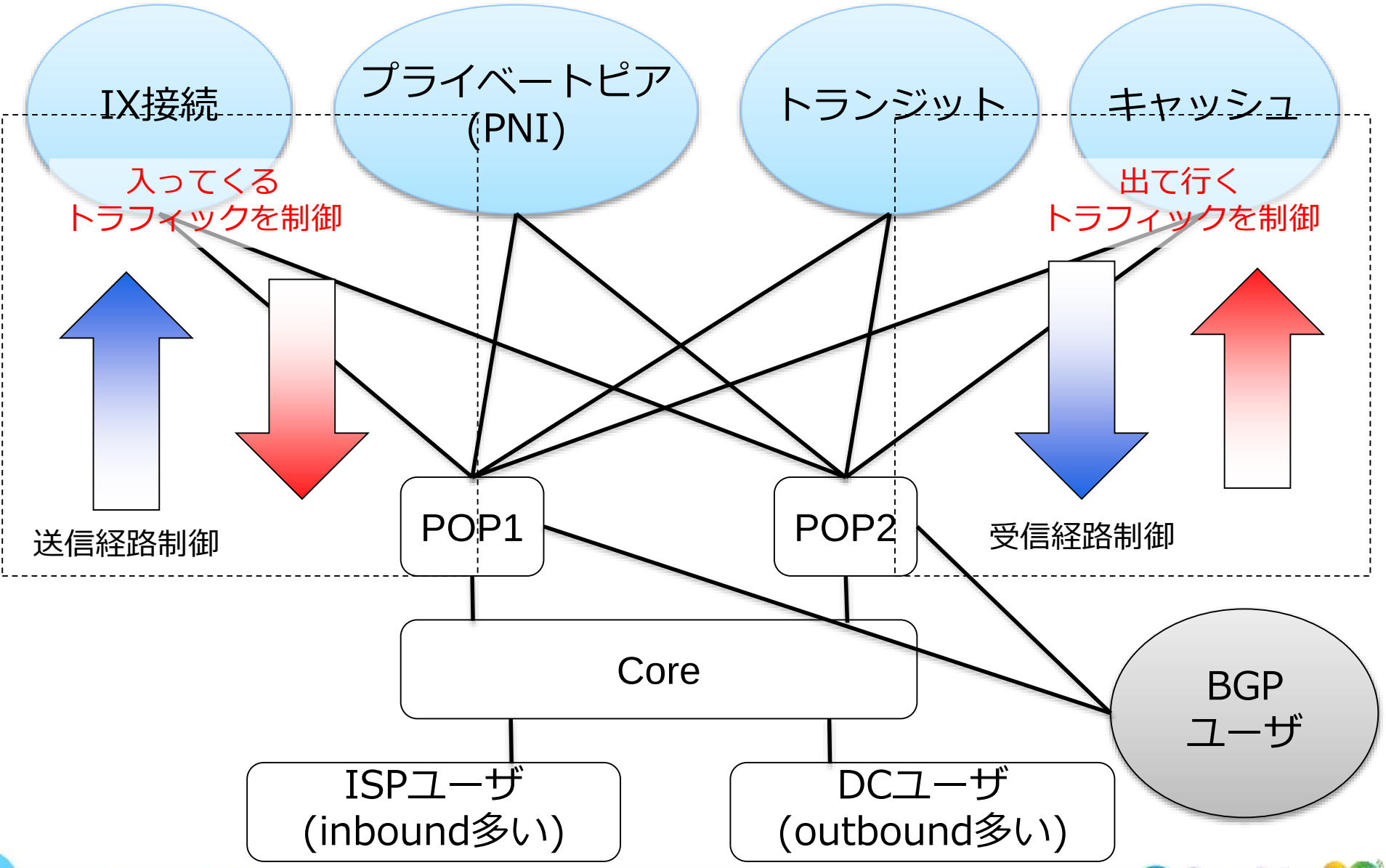
Upstream ISPは3つ
AS2914、AS6453、AS1299

トラフィック制御の考え方

芯：事業に最適なトラフィック設計を行う

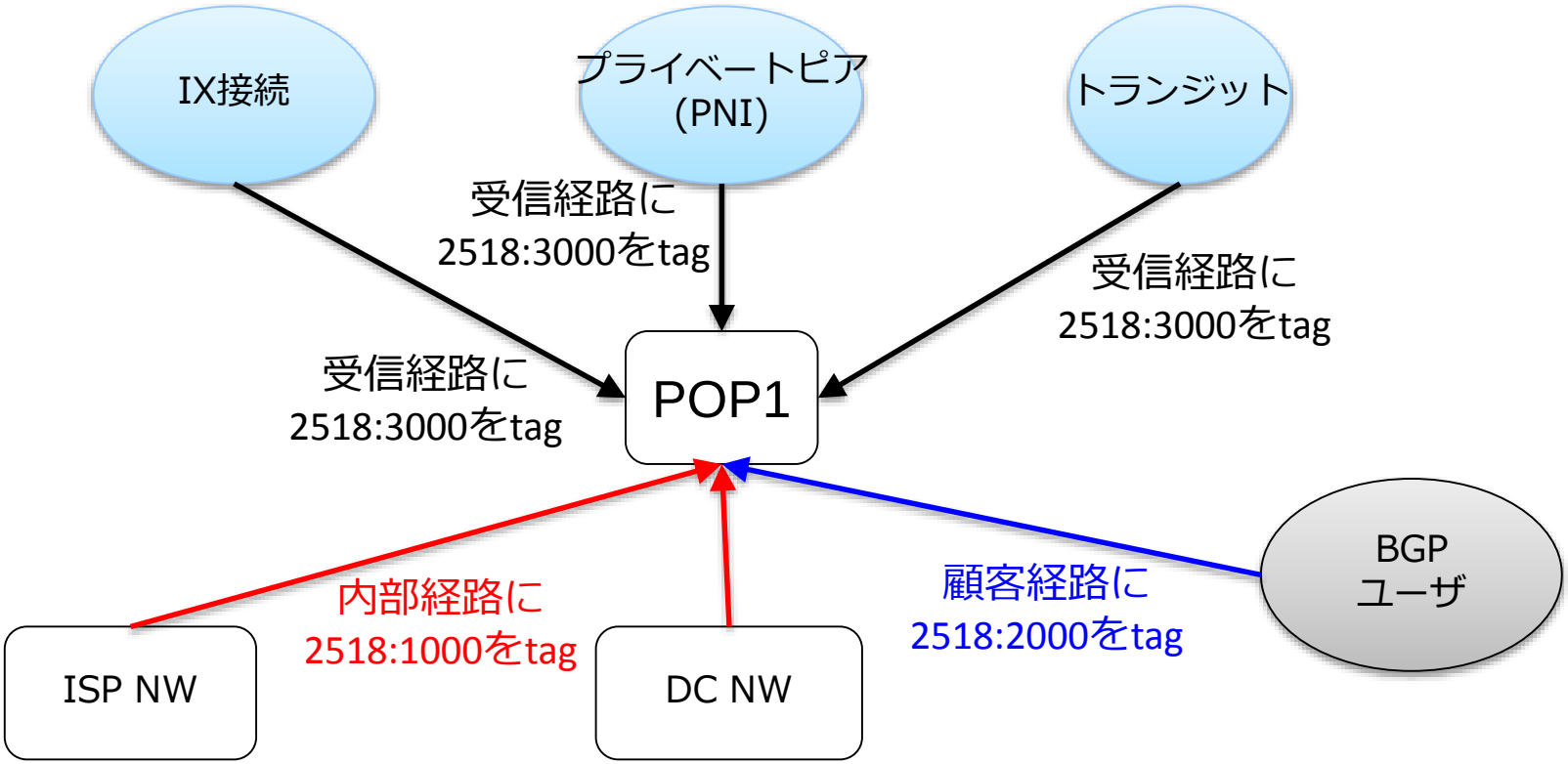
- 具体的に考慮する事：
  - ISP/MVNO事業としてのコスト、遅延
  - ハウジング/ホスティングのコスト、遅延
  - 資本関係やパートナーシップ
- やらない事：
  - 事業にそぐわない/無視した設計
  - インターネット上の他社に迷惑をかけてしまう設計

対外接続設計：基本要素



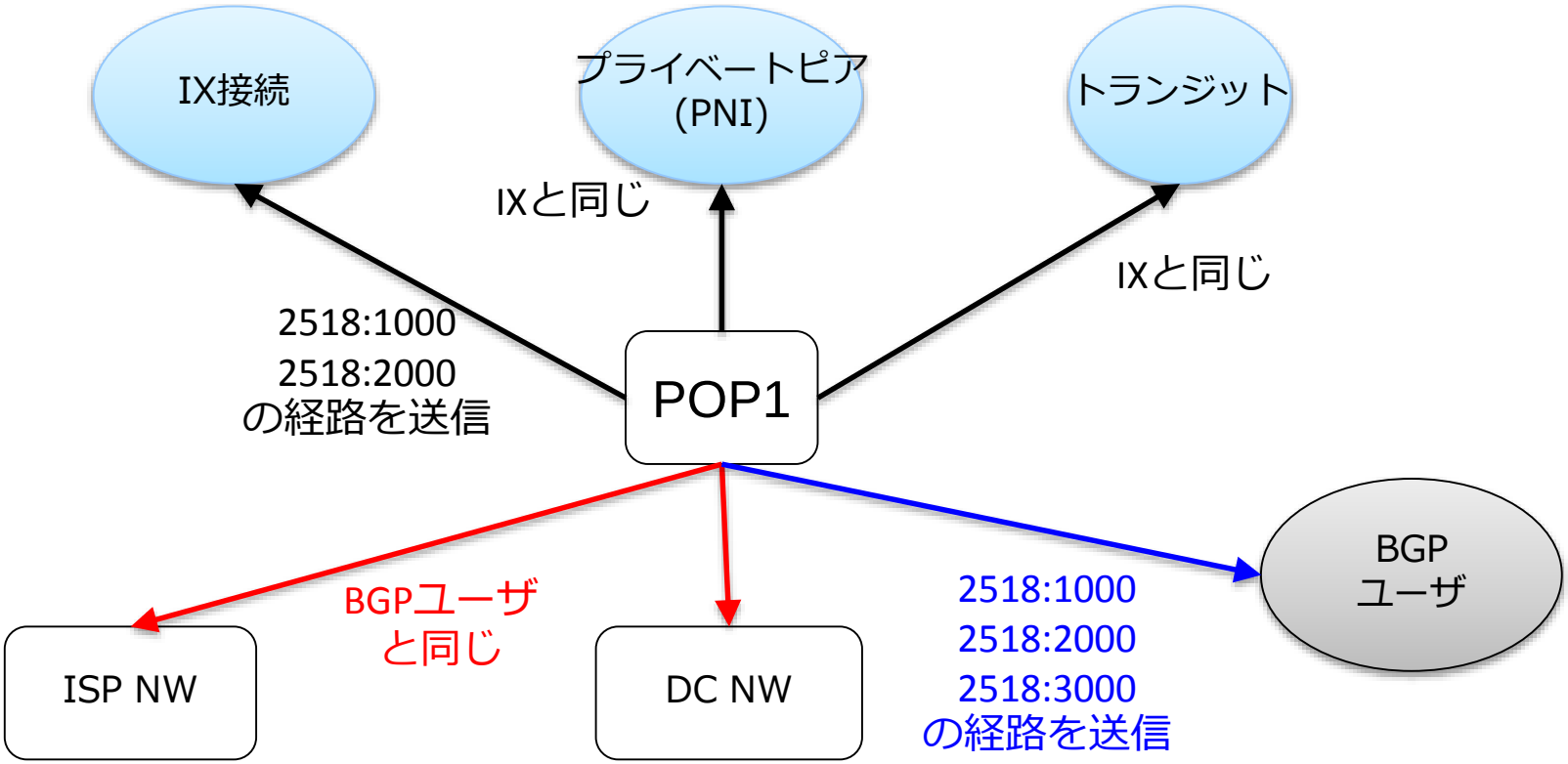
シンプルな実装(昔のAS2518の例)

- 難しいトラフィック制御は行わず、接続の種別に応じて送信経路を変更
- 受信した経路にCommunityをTag



シンプルな実装

- 送信経路はCommunity値で制御
  - Peerとトランジットには、Peerから受けた経路を流さない法則をCommunityで守る



シンプルな実装

- 間違い防止のために
  - Peerへの経路送信route policy、Transitへの送信route policyをあらかじめ定義しておき、それ以外はなるべく使わない

IOS-XRの例

```
route-policy peer-export
  apply do-not-send-these
  if community matches-any internal or community matches-any customer then
    set med 100
    delete community in any
    done
  endif
  drop
end-policy
```

基本フィルター

Peerの場合、コミュニティをつけて送っても意味が無い場合が多いので消す

基本のdrop。重要。

すごくシンプル！

シンプルな実装

- 相手がCommunityをつけてきても
  - Communityで操作する事を許していない相手の送ってくるコミュニティは見なくてもいい
  - AS2518では、Peer毎にポリシーを作ってる

IOS-XRの例

```
route-policy as65535-import
  apply do-not-receive-lists
  if as-path in as65535-aspath1 then
    set med 0
    set local-preference 300
    set community peer
  done
  endif
end-policy
```

基本フィルター

“additive”が付いて無いのでcommunity値2518:3000で上書きする

すごくシンプル！

シンプルな実装

- フルルート送る設定も簡単！

IOS-XRの例

```
route-policy full-out
  apply do-not-send-these
  if community matches-any internal or community matches-any customer or
  community matches-any external then
    set med 0
    delete community in any
    done
  endif
  drop
end-policy
```

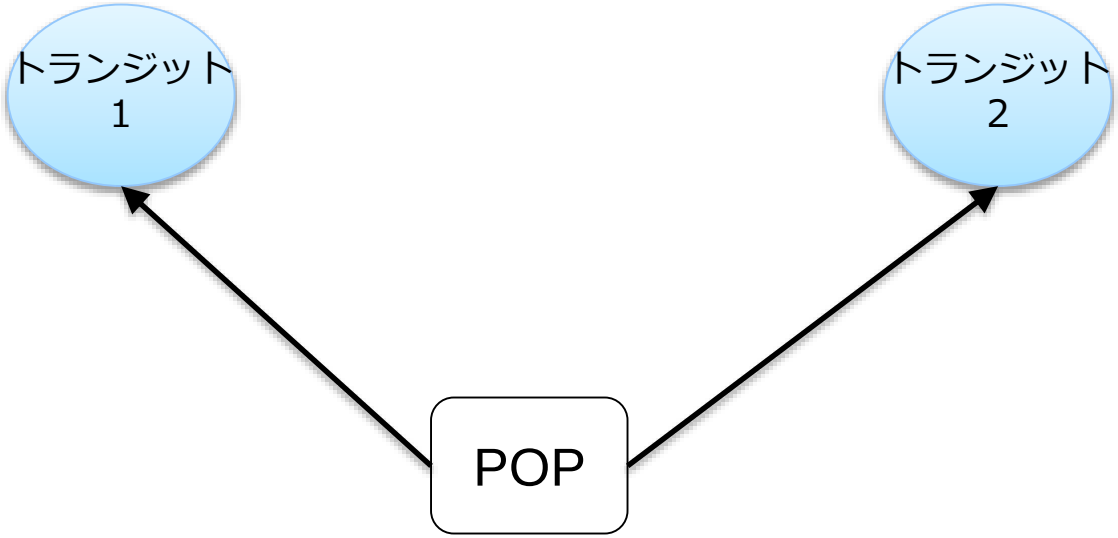
例では全部Communityを削除しているが、BGP Community機能を提供する場合は「何を送信して良いか」は定義した方が良い

すごくシンプル！

このシンプルな実装を元に、Prefix filterで
トラフィック制御をしていく

例えばTransit2社のトラフィックバランス

- トランジット1は標準設定だけど、トランジット2はinトラフィックを少し減らしたい場合



代表的な手法：
1) 一部送信経路にprependする(1-3回程度)
2) トランジッターの提供するBGP Communityを使う

#より細かい経路をトランジット1側に送信する事で減らす事もできますがこの構成ではあまりお勧めしません

1)の設定例

```
prefix-set transit2-prepend
  192.0.2.0/24
  end-set
route-policy transit2-export
  apply do-not-send-these
  if community matches-any internal or community matches-any customer then
    set med 100
    pass
  endif
  if community matches-any internal
    if destination in transit2-prepend then
      prepend as-path 2518
      pass
    endif
    pass
  endif
 if community matches-any internal or community matches-any customer then
    delete community in as2518-any
    done
  endif
  drop
end-policy
```

2)の設定例

```
prefix-set transit2-prepend
  192.0.2.0/24
  end-set
route-policy transit2-export
  apply do-not-send-these
  if community matches-any internal or community matches-any customer then
    set med 100
    pass
  endif
  if community matches-any internal
    if destination in transit2-prepend then
      set community 1299:5881 additive
      pass
    endif
    pass
  endif
 if community matches-any internal or community matches-any customer then
    delete community in as2518-any
    done
  endif
  drop
end-policy
```

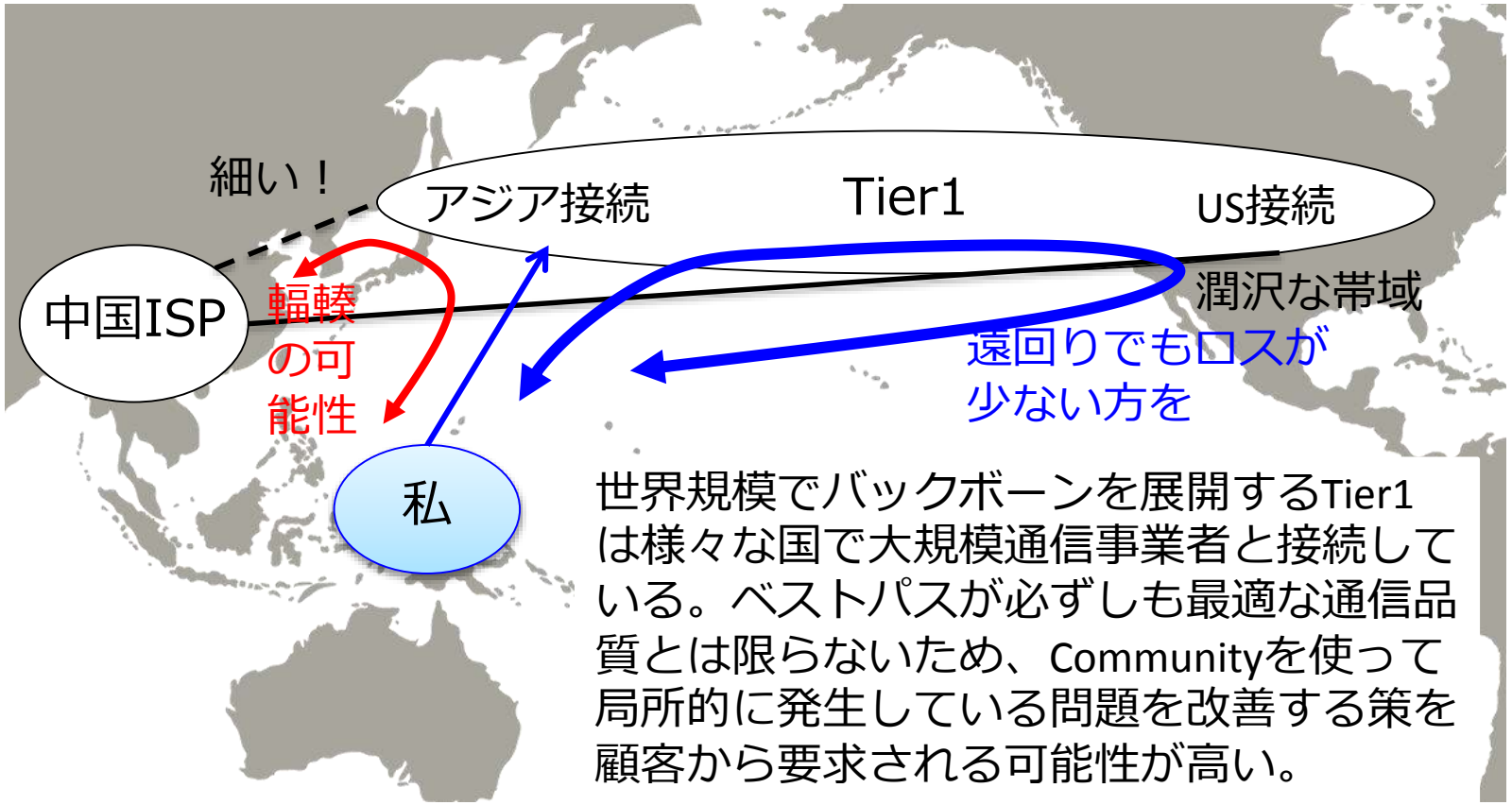
ここが前ページと違う
特定のISP向けは1回
prependしてから
トランジット2は送信

トランジッターのBGP Community利用

- Tier1は全事業者Community機能を提供している
  - whoisで公開している事業者、Webで公開している事業者、pdfなどファイルで情報提供する事業者
  - https://www.gtt.net/services/internet-services/ip-transit/bgp-communities/
  - https://www.us.ntt.net/support/policy/routing.cfm
- 注意！
  - RTBH機能を使う場合(65535:666)、IPv4は/32単位で受けてもらう事が一般的だが基本フィルターで/24よりも長いprefixはフィルターする設定になっている場合はコミュニティ値が付いている経路だけは/24以上で経路広告できる設定にする必要がある
- Tier2以下は対応がTier1に比べると対応機能が少ない

なぜTier1はBGP Communityが必要かの例

- Tier1はUSでの相互接続が多い(歴史的にも、インターネットのトラフィック的にも)
- BGPは近いところから出ようとする性質があるが、近いところが必ずしも最適とは限らない



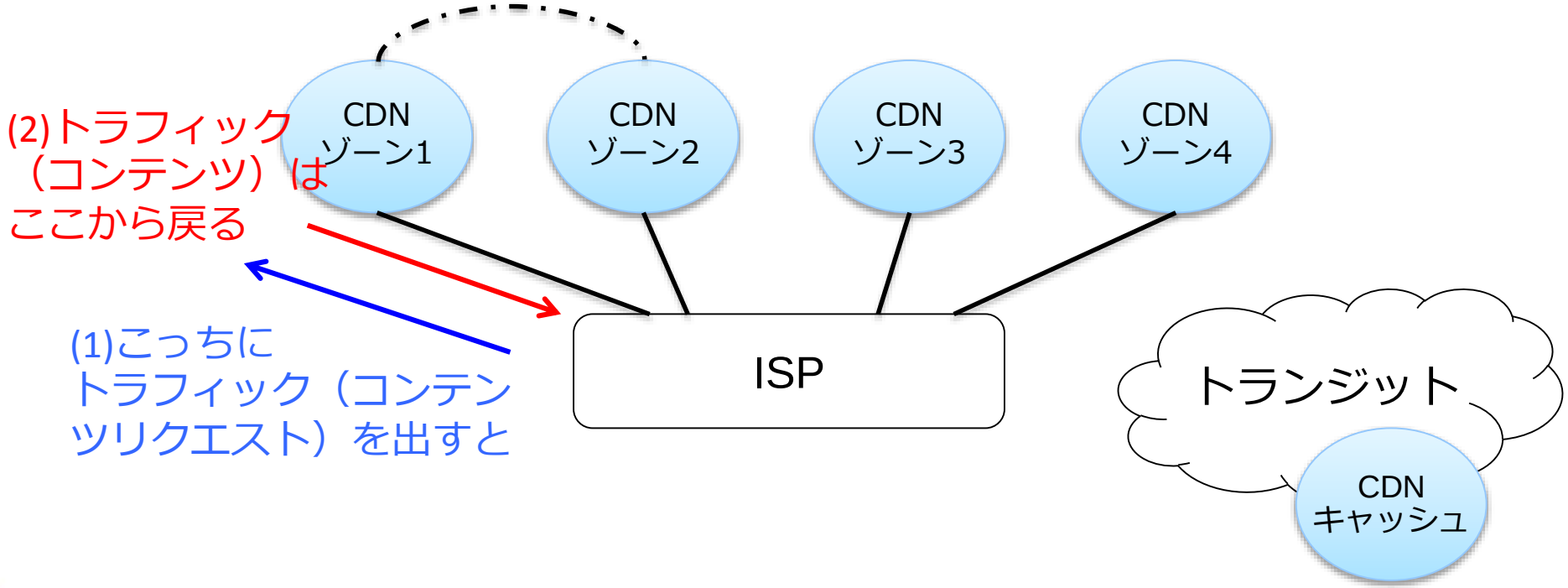
世界規模でバックボーンを展開するTier1は様々な国で大規模通信事業者と接続している。ベストパスが必ずしも最適な通信品質とは限らないため、Communityを使って局所的に発生している問題を改善する策を顧客から要求される可能性が高い。

CDNトラフィックの考え方

- CDNは、バックボーンを持っていない(配信サーバ間が内部ネットワークで接続されていない）場合が多く、独特な考え方が必要
- 2種類のメジャーなコンテンツ配送方式
  - Anycast方式
  - DNS方式

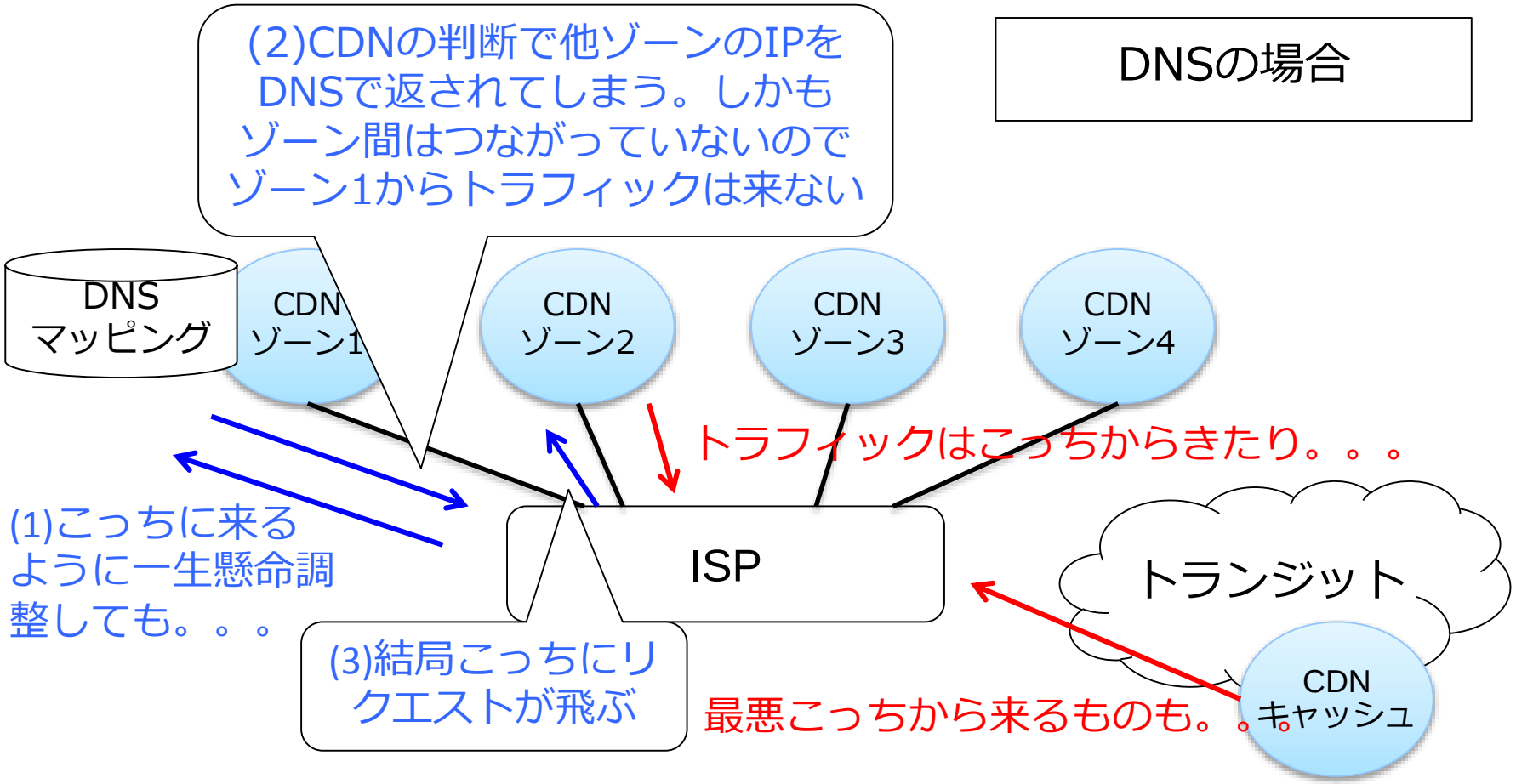
Anycastの場合

ゾーン間をつなぐネットワークが無く、インターネット経由で通信する



CDNトラフィックの考え方

- Anycastトラフィックは前ページのとおり送信したところからトラフィックが戻ってくる、しかしDNSベースでCDNマッピングされてしまう場合は送信トラフィックを制御するだけでは効かない



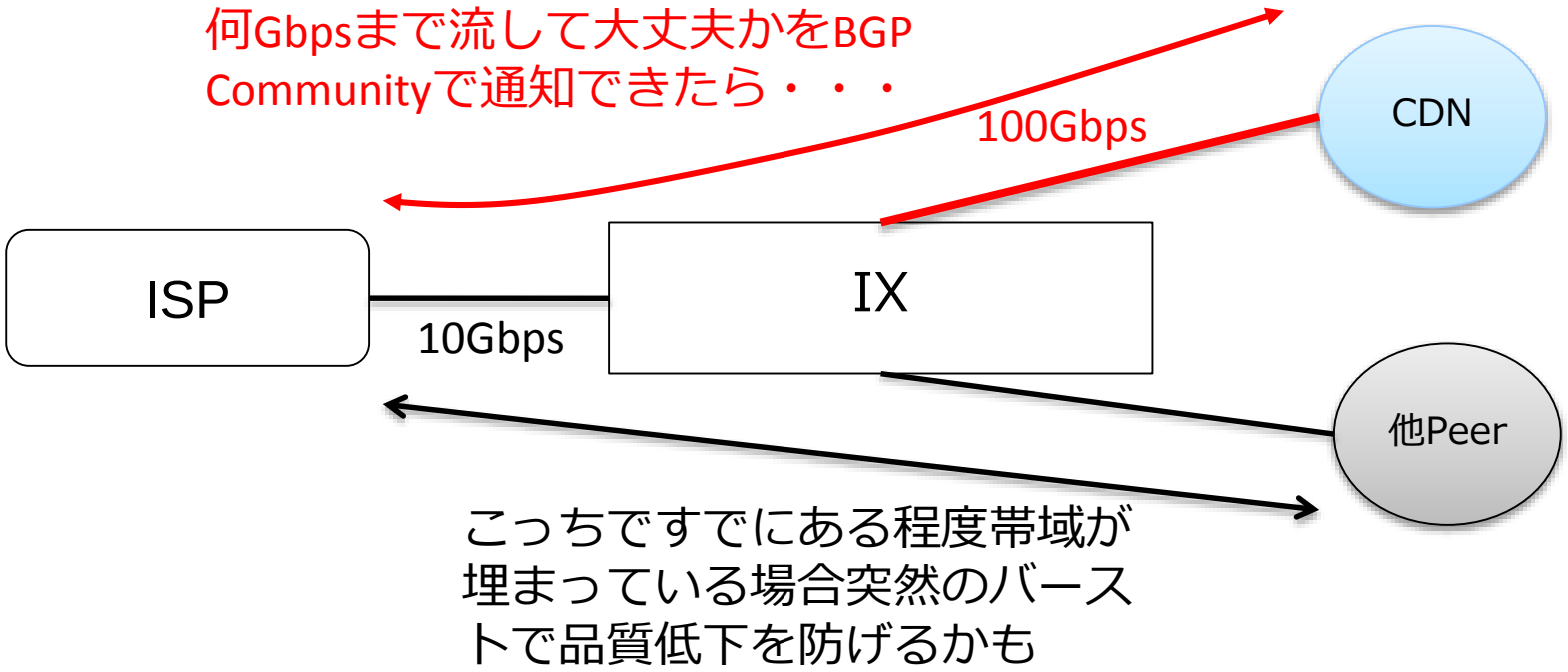
ISPでできる事

1) トラフィックが多い回線は、深く考えず増強する
   - 難しい操作をせず増速で対応。これがベスト
   - （プライベートピアの場合）CDNの多くは輻輳検知できるのであふれても品質劣化にはならない
   - CDNは通信遅延を測定し、最もコストがかからずかつ近いところから出そうとする傾向があるのでそれに任せる。ただしコストは重要ファクターなので必ず近いところ、というわけではなさそう
2) やりたい事をCDN会社に説明して協力をお願いする
   - 多くの場合協力的な姿勢
3) トラフィックを減らしたい回線に対してPrependする
   - MEDはほとんど効かない
   - 送信経路削減はよく考えて実施しないと通信断リスクがある

参考：https://conference.apnic.net/data/39/04-akamai-mj-traffic-engineering-apricot_1425633293.pptx
https://www.peeringforum.asia/files/file/CDN_AnycastxDNS_Mapping-APF2016_Kams.pdf

こんな事できたら面白そうだけど。。。実際はできません

- (IXなど共通リンクでPeerしている場合)BGP Community値で何Gbps受けられるか通知できたり・・・



Peerに対して提供するBGP Community

- BGP Communityは今までTransitがCustomerに提供する機能、もしくは内部利用が一般的だった
- しかし、DDoS対策のためにPeer同士でCommunityを利用する事も議論は出てきている
  - Peering Policyに記載している会社もある（厳密な要求ではない）
    http://www.riotgames.com/peering-policy
- トラフィックの流れ方が変わってくると、その調整方法や必要な技術も変わってくるため、動向には注視しておく必要がある

AS2518の現在：内部で使っているBGP Community

- 海外のPOPが増えたので地域識別を追加
  - どの国で拾った経路か識別したい
    - トランジット顧客への情報共有手段
    - 通信障害対応
      - (やったことないけど)リージョン毎のプリファレンス操作など
- トランジットサービスを始めたので顧客向け制御機能追加
  - 特定のISPやTransitへ経路を出さないコミュニティ
  - 今後prependも追加予定
- Black Hole Routing機能（内部向け、顧客向け）は昔から
  - 顧客向けに2518:666
  - RFC値に今後対応予定

今後検討するトラフィック制御手法

- Large Communities (RFC7999)対応
  - 4byte-ASのPeerやトランジット顧客が増えてきたため急務

- BGP Flowspecは実装や仕様で課題が多いがBlackholeでもなくScrubbing(DDoSのインテリジェントな防御）の中間策にもなるため何とか導入したい

まとめ

- トラフィック制御の考え方は、事業によって異なる
- トラフィック制御が必要になった場合、BGP Communityを使った制御は有効
- Tier1は多様なBGP Community機能を提供している
  - 使わないと困るケースもある
- CDNのトラフィックは通常のPeerと違う考え方が必要
- BGPのトラフィック制御は今後も進化しそう