ルートサーバの現状と
マルチラテラルピアリングの可能性

Eiichiro Watanabe
watanabe@mfeed.ad.jp

- ルートサーバとは？
- ここ数年で変わってきています。
- こんなんやろうと思いますがどう思う？

ルートサーバとは？

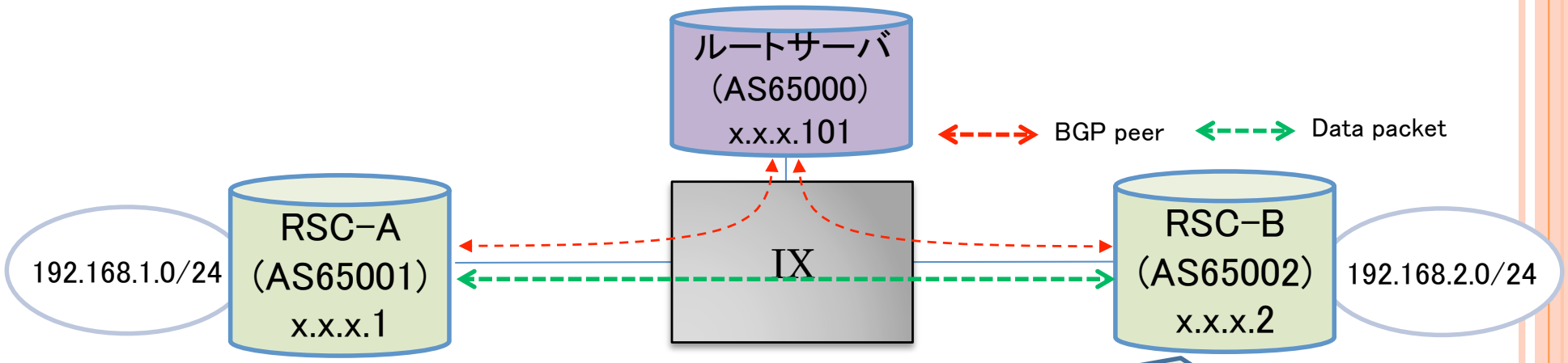
- もちろん、root (DNS) serverじゃないです…

  route server ≠ root server

- 広義の意味では以下のようなroute collecter(やlooking glass)も含まれますが、これも含めません。
  - RouteViews
  - RIPE NCC Routing Information Service
  - Team Cymru Bogon Route Server
  - …etc.

本セッションでは、ルートサーバを
「IX上でトラフィックを運ぶためのBGP経路交換を仲介する箱」
として説明します。

ルートサーバの基本動作



RSC-Bのルーティングテーブル

Prefix	next-hop	MED	AS_PATH
*192.168.1.0/24	x.x.x.1	100	65001 I

NEXT_HOP属性は変更されないため、フォワーディングはルートサーバを経由しない。

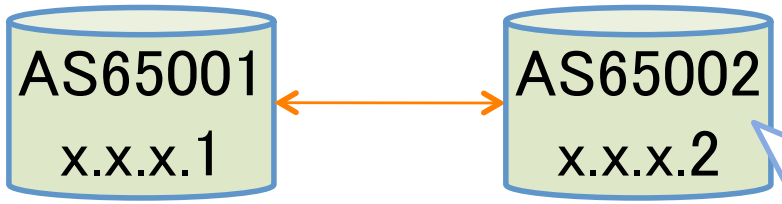
MED属性も変更されず、透過します。
(実はRFC4271違反)

AS_PATH属性も変更されず、ルートサーバのASは付加されない。
(実はRFC4271違反)

ルートサーバは基本的に受信した経路のBGP属性を変更せず配信

経路の見え方はほぼ同じ

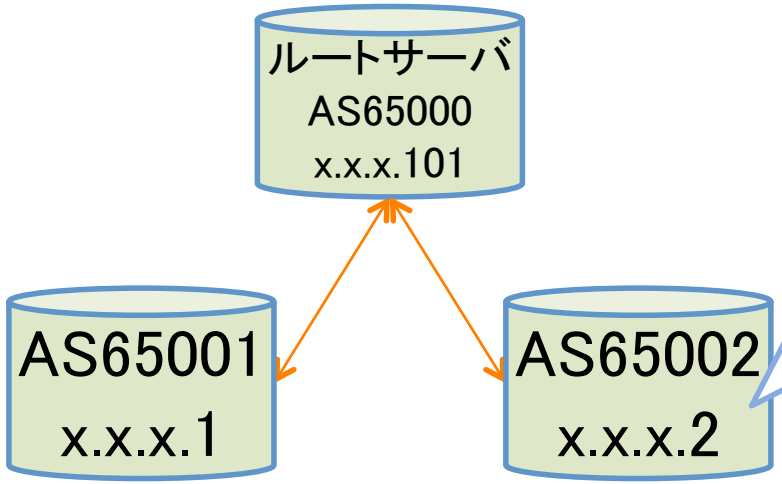
バイラテラルの場合




```
Paths: (1 available, best #1, table default)
  Not advertised to any peer
  Refresh Epoch 1
  65001, (received & used)
    x.x.x.1 from x.x.x.1 (x.x.x.1)
      Origin IGP, metric 0, localpref 100, valid, external, best
```

マルチラテラルの場合

remoteアドレスとrouter-idが違うだけ




```
Paths: (1 available, best #1, table default)
  Not advertised to any peer
  Refresh Epoch 1
  65001, (received & used)
    x.x.x.1 from x.x.x.101 (x.x.x.101)
      Origin IGP, metric 0, localpref 100, valid, external, best
```

- ルートサーバとは？
- ここ数年で変わってきています。
- こんなんやろうと思いますがどう思う？

標準化動向

IETFにて現在以下の2つのInternet-Draftが提案されている。
（現時点ではまだPersonal Draft）

■IDR(Inter-Domain-Routing) WG

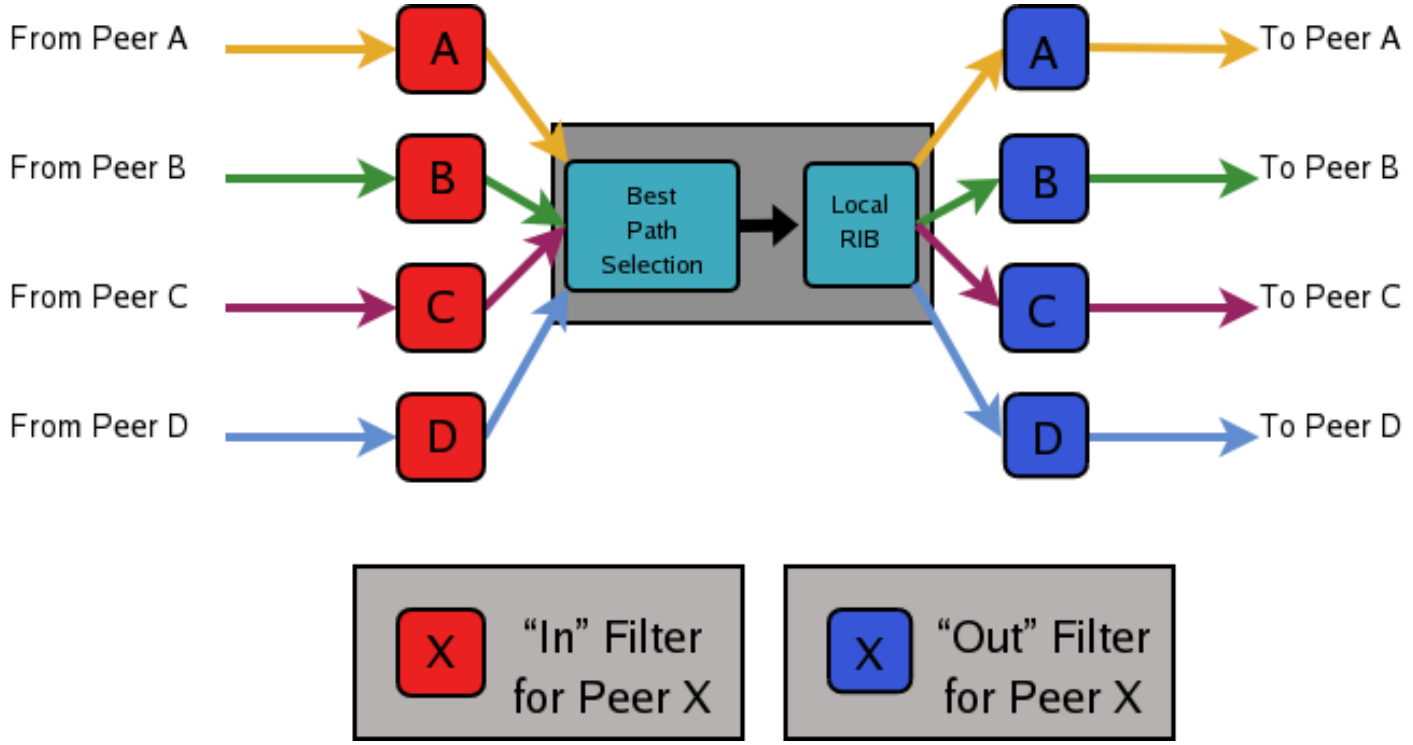
draft-jasinska-ix-bgp-route-server-03
http://tools.ietf.org/html/draft-jasinska-ix-bgp-route-server-03
Standards Trackとして提案中
ルートサーバの技術的仕様について言及

■GROW(Global Routing Operations) WG

draft-hilliard-ix-bgp-route-server-operations-00
http://tools.ietf.org/html/draft-hilliard-ix-bgp-route-server-operations-00
Informationalとして提案中
ルートサーバの運用上考慮すべきポイントについて言及

bgpdのmulti-RIB(per-client RIB)化

従来(Single-RIB)でのBGPプロセス(quaggaでの例)

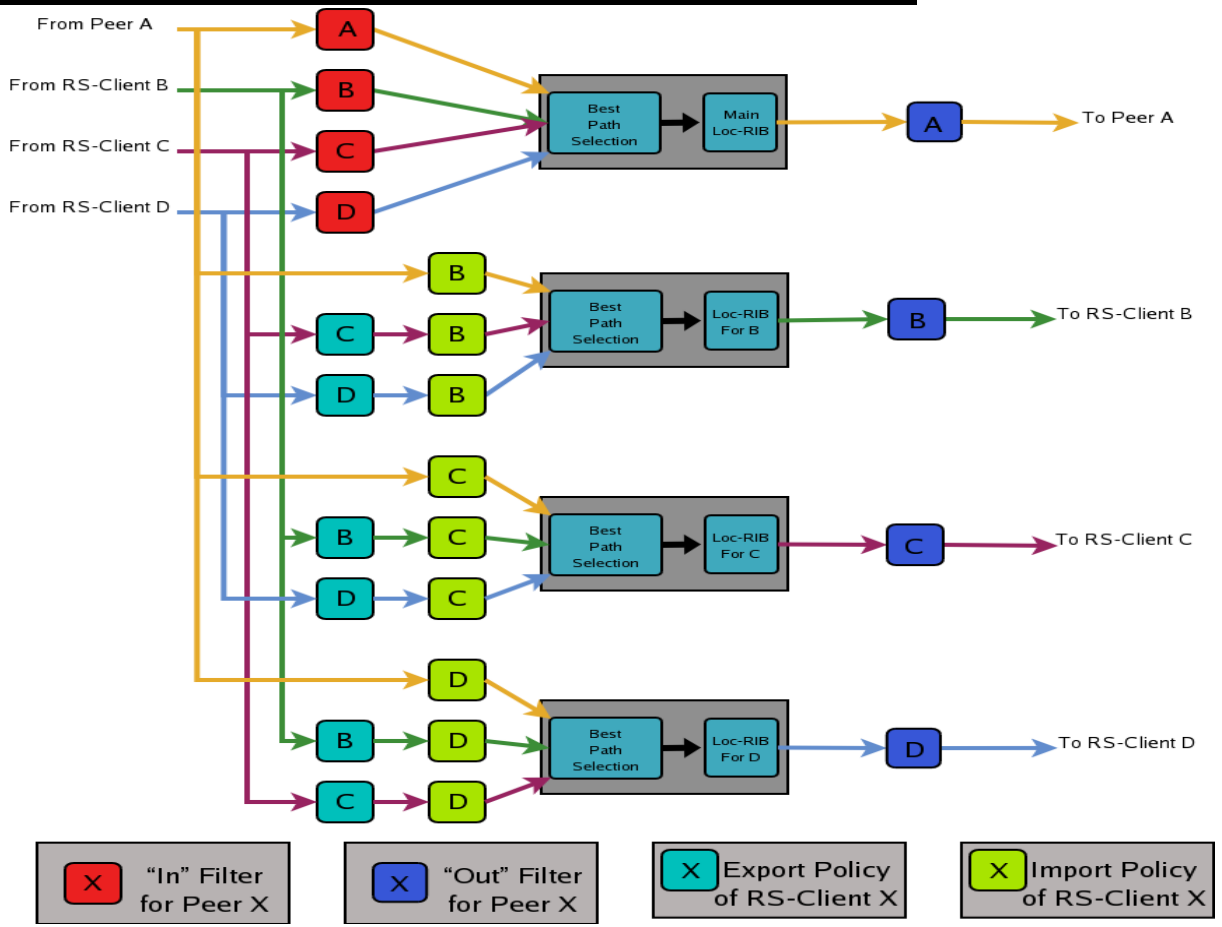


http://www.quagga.net/docs/fig-normal-processing.png

各Peerから受信した経路は、1つのRouting Table上で
ベストパス選択され、配信される。

bgpdのmulti-RIB(per-client RIB)化(つづき)

Multi-RIBでのBGPプロセス(quaggaでの例)



http://www.quagga.net/docs/fig-rs-processing.png

各PeerごとにRouting Tableが管理され、
各々でベストパス選択が可能。

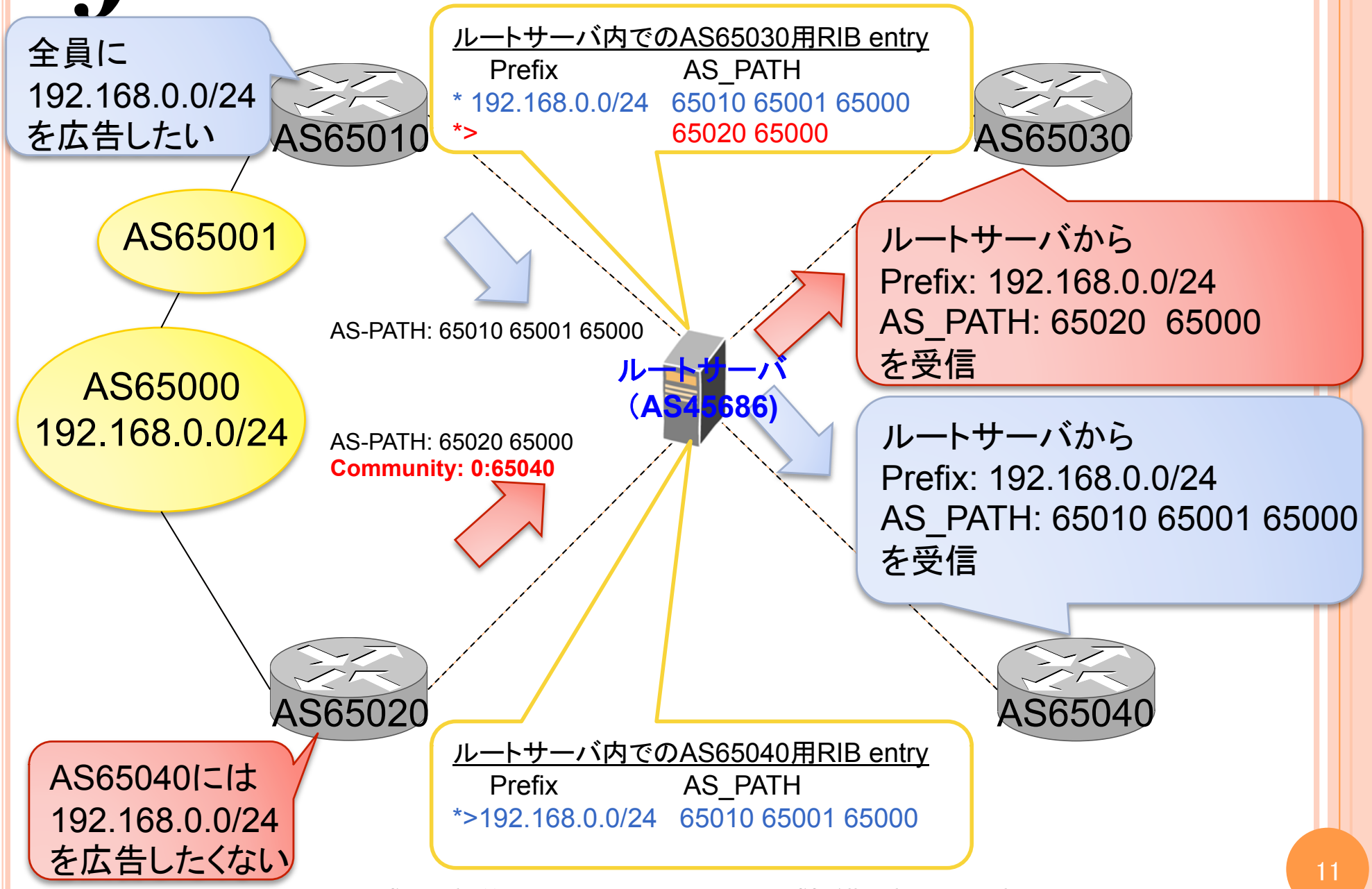
multi-RIB利用例

ルートサーバ利用者が自身の広報する経路にBGP Communityを付加することでPeer単位に経路広報をすることができる。

BGP Community定義(JPNAPの例)

BGPコミュニティ値	制御内容
0:<peer as>	<peer as>には配信しない
45686:<peer as>	<peer as>には配信する
0:45686	全てのASに配信しない
なし（標準設定）	全てのASに配信する

multi-RIB利用例(つづき)



BGP Community利用状況とその課題

BGP Community利用状況

- 現時点では、常時このCommunityを利用するユーザはほとんどいない。
- メンテナンスや帯域圧迫により一時的にトラフィックを逃がす目的で利用されている様子。

一方、ルートサーバ利用者の声として

- “経路を広報しない”という制御は強すぎ => 課題1
- BGP CommunityでOutgoingのトラフィック制御ってできないよね？ => 課題2

- ルートサーバとは？
- ここ数年で変わってきています。
- こんなんやろうと思いますがどう思う？

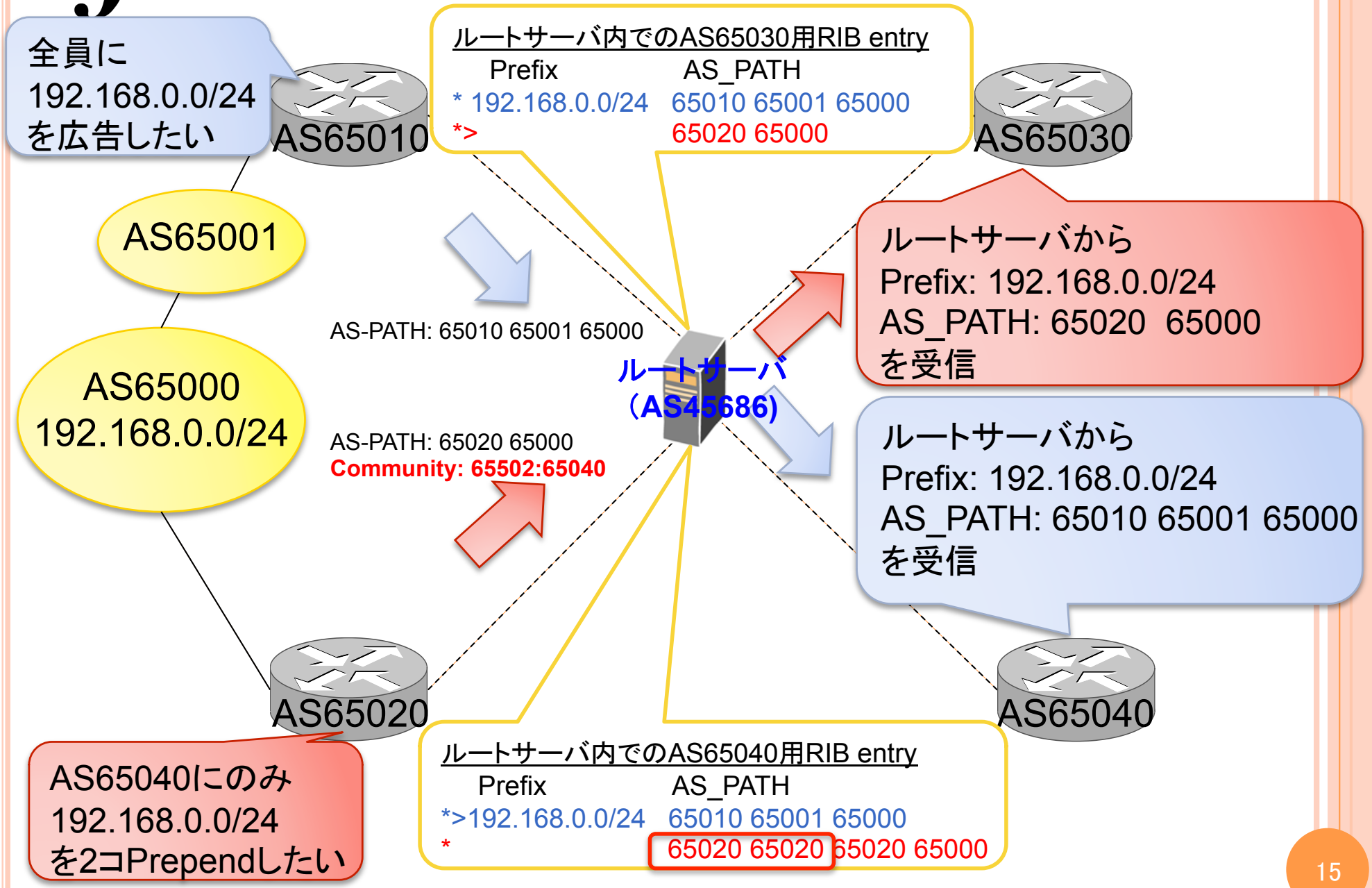
課題1の対策案

“特定のASに広報しない”BGP Communityだけでなく、AS_PATH prepend, MED加減算ができるようにする。

BGP Community定義(案)

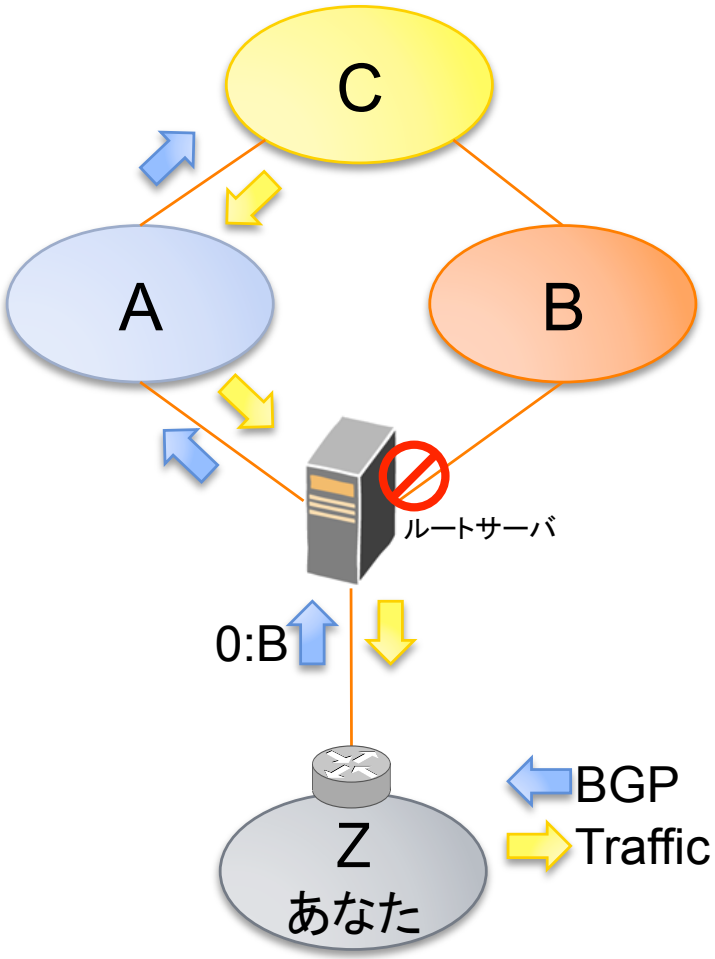
BGPコミュニティ値	制御内容
65501:<peer as>	<peer as>に対し送信元のASを1つprepend
65502:<peer as>	<peer as>に対し送信元のASを2つprepend
64999:<peer as>	<peer as>に対し、受信したMED値+1
65001:<peer as>	<peer as>に対し、受信したMED値-1

課題1の対策案(AS_PATH Prepend x2の例)



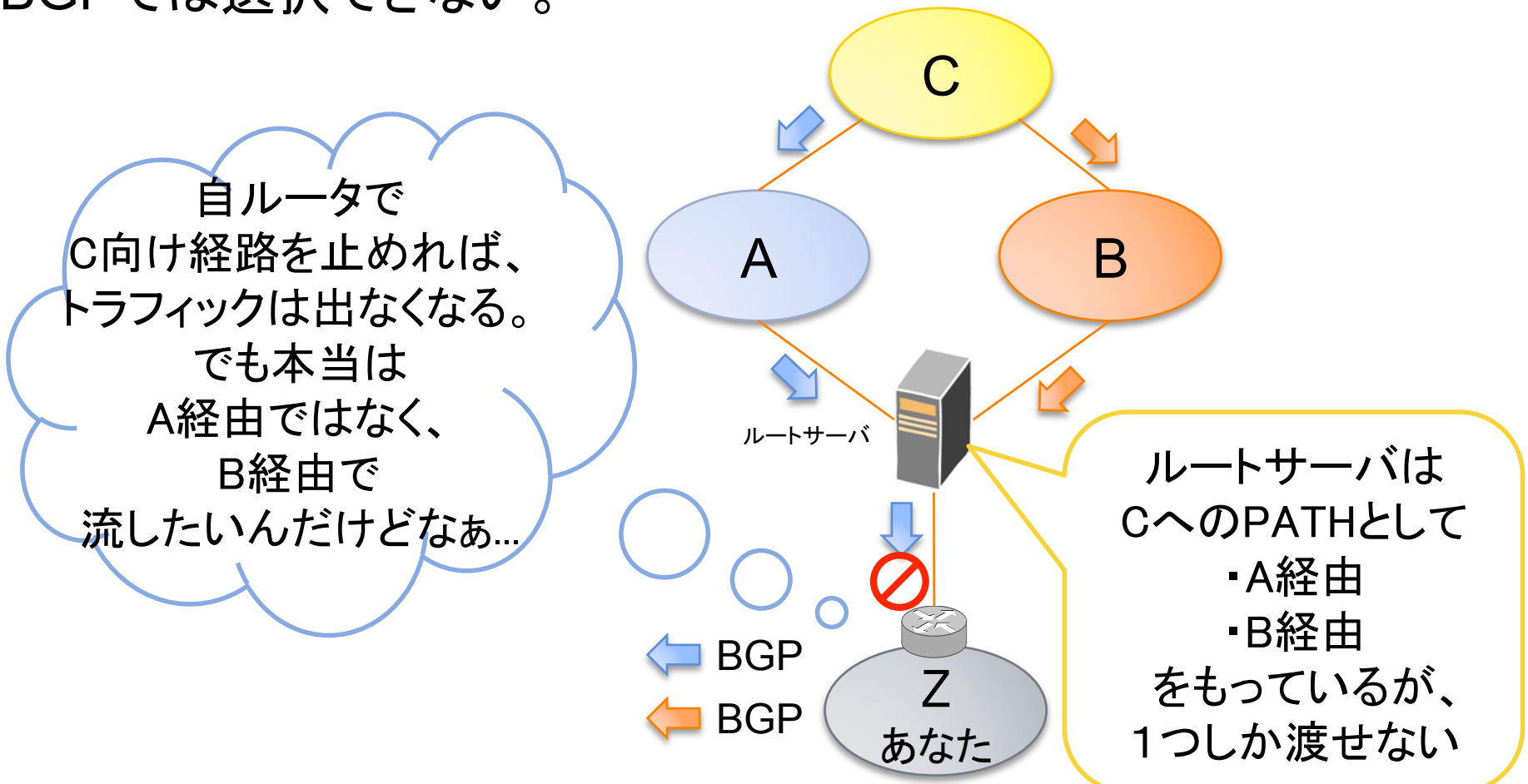
課題2（Incomingトラフィック制御）

前述のBGP Communityを用いれば、Incomingのトラフィック制御はある程度可能。



課題2(Outgoingトラフィック制御)

ルートサーバが経路を集約し、ルートサーバが判定したベストパスのみを渡すため、受信経路の選択はルートサーバ利用者側からBGPでは選択できない。



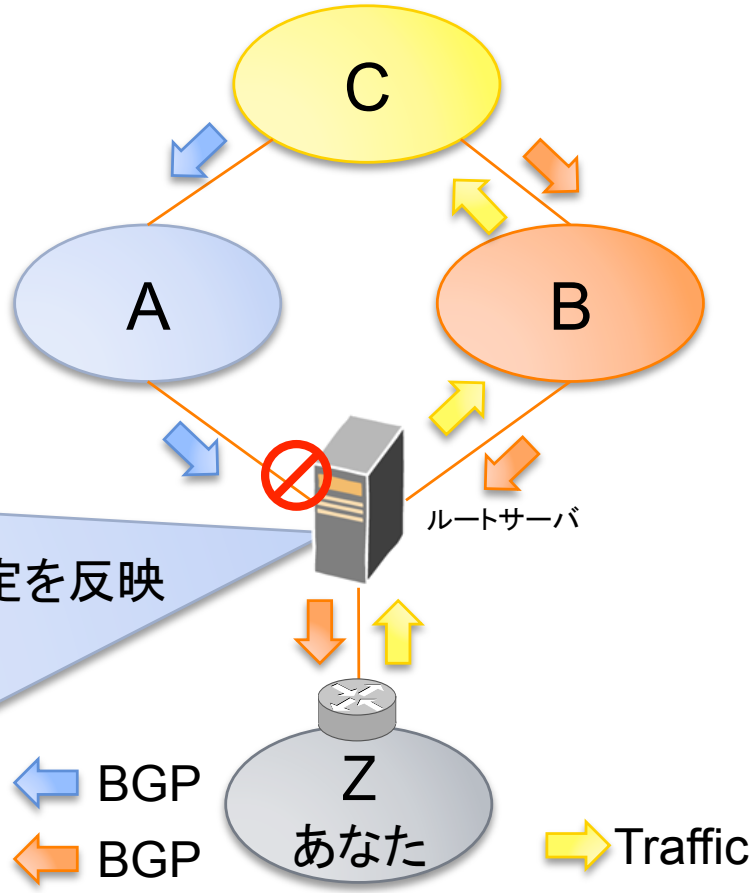
Outgoingトラフィックを制御させるには別の何かが必要…

課題2の対策案1（GUIでポチポチ）

カスタマポータルサイト上で個々にPeerする／しないを選択し、その内容をルートサーバの経路フィルタに反映。(下記はVienna IXでの例)

http://www.vix.at/vix_routeserver.html?&L=1

設定を反映



〇利用者はルータをさわらなくても、ポリシー制御が可能。

×提供するにはそれなりの開発コストが必要。

課題2の対策案2（IRRとの連携）

IRR(Internet Routing Registry)のimport/exportフィールドを見て、ルートサーバの経路フィルタに反映(AMS-IXの例)

IRRdb-Based Filtering

ANY

Send and receive prefixes to/from any RS participant:

```
aut-num: AS1200
descr: Amsterdam Internet Exchange (AMS-IX)
[...]
import:       from AS6777 accept ANY
export:       to AS6777 action community .= { 6777:6777 }; announce AS1200
[...]
```

ANY except

Send and receive prefixes to/from any RS participant EXCEPT AS1103 & AS12859:

```
aut-num: AS1200
descr: Amsterdam Internet Exchange (AMS-IX)
[...]
import:       from AS6777 accept ANY AND NOT <^[AS1103 AS12859]>
export:       to AS6777 action community .= { 6777:6777, 6777:1103, 6777:12859 };
                          announce AS1200
[...]
```

RESTRICTIVE

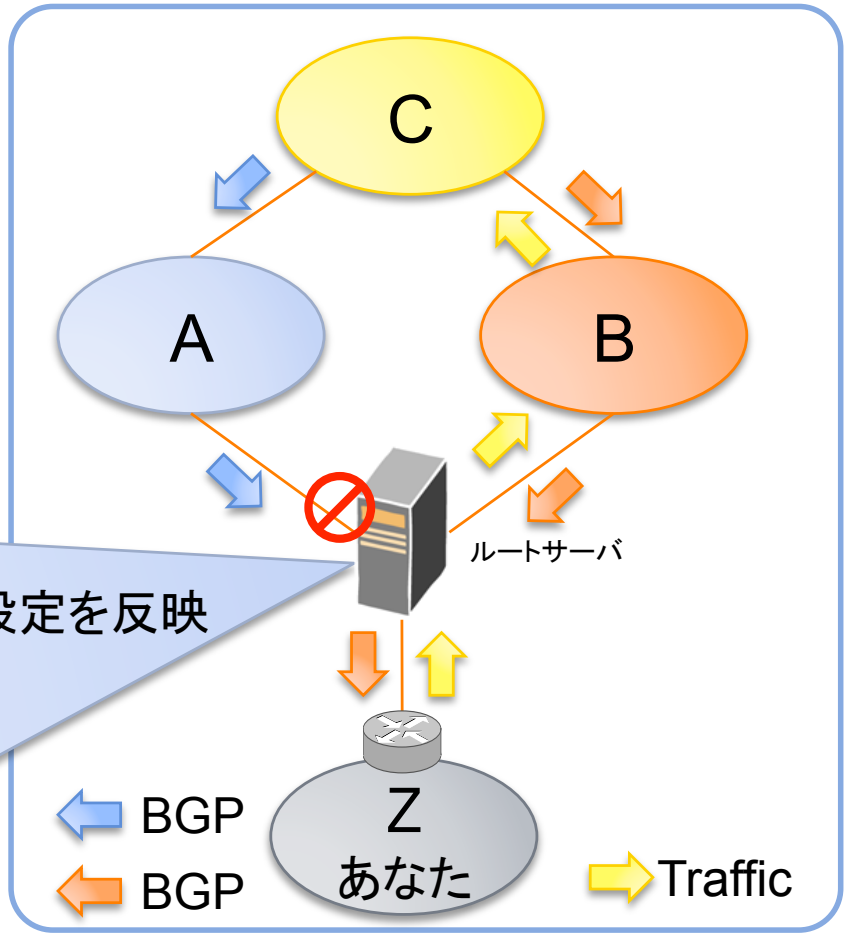
Send and receive prefixes ONLY to/from AS15703 :

```
aut-num: AS1200
descr: Amsterdam Internet Exchange (AMS-IX)
[...]
import:       from AS6777 accept <^AS15703>
export:       to AS6777 action community .= { 6777:15703 }; announce AS1200
[...]
```

Of course also AS-SETs can be used in the import/export rules. All ASes participating in the Route Server peering are grouped into the AS-SET: AS-AMS-IX-RS.

http://www.ams-ix.net/ams-ix-route-servers/

設定を反映

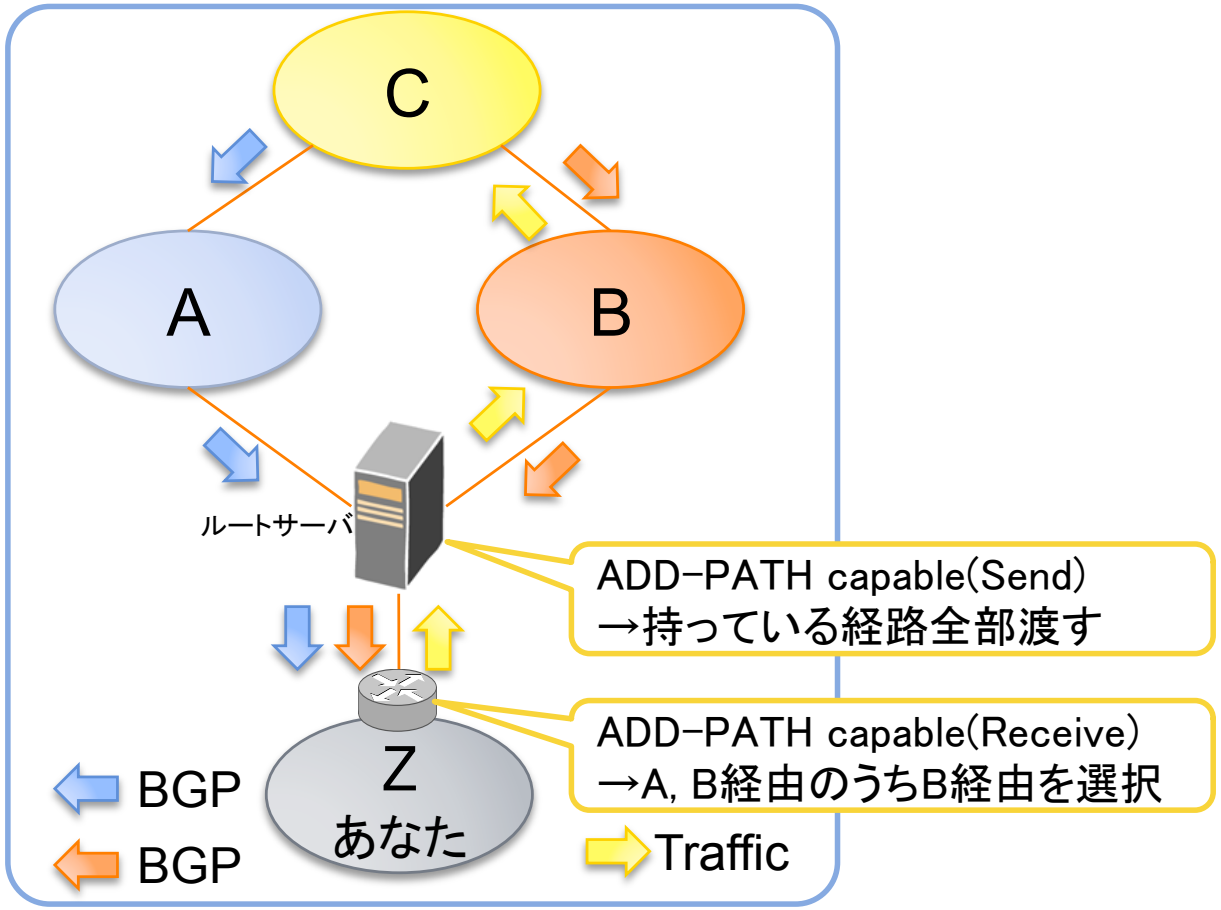


○BGPオペレータの比較的なじみのあるIRRでポリシー制御できる。

×ポリシーモロ出し....日本では不向きか。

課題2の対策案3（BGP ADD-PATH）

BGP ADD-PATHを使ってルートサーバの持つ経路をすべて渡す。
利用者はそのうち好きな経路を選択する。



○バイラテラルとほぼ同等？コンバージェンス早くなる？
×双方でBGP ADD-PATHの実装が必要。

ルートサーバ提供側が思うこと

たぶん、

- BGPエンジニアたくさんいるぜ！
- Peering交渉なんて楽勝！
- バイラテラルで各Peerごとに細かくポリシー制御が俺の仕事！

ってのがもちろんBGP運用としては最強！

でも、

- BGPエンジニアが少ない．．．っていうか俺一人。
- Peering交渉大変。トラフィックさっさと流したいんだけど。
- 日々のBGP運用大変

というひとたちに対して、マルチラテラルピアリングは有効

みなさんに聞いてみたいこと

さらに、バイラテラルっぽい経路制御をルートサーバ利用者に対して提供することでユーザの利便性を向上させたいなと思っています。

制御方法	ルートサーバ提供側として考えている対象
GUIでポチポチ	BGPエンジニアが少ないお客様向け 現ユーザの要望多い
BGP Community	BGPエンジニアがいるお客様向け ？
IRRとの連携	
BGP ADD-PATH	

使う？ 使わない？

ご清聴ありがとうございました。

Q & A