Internet week 2016

想いが伝わるBGP運用

BGP Communityの基本設計

小島 慎太郎

codeout

http://about.me/codeout

2004 ISP ntt.net

2014 フリーランスの AS手伝い

2009 IX JPNAP

本日話すこと

- BGP Community がどのように使われてるか
  - 自分たちで決めるパターン
  - ほかで決まったものをつかうパターン
- 自分たちで設計してみよう
  - なぜこの設計なのか、私の経験から
- 設計したBGP Community をほかに公開するべきか考えてみよう

BGP Community

BGP Community

- 経路につけるタグ
- 2バイト：2バイト

- Optional Transitive
  (あってもなくてもOK & ASまたぎで伝搬する)
- 基本的には勝手に決めてOK
  - NO_EXPORT、NO_ADVERTISE などの
    定義済みCommunity と重複しないかぎり

BGP Community

Global Administrator
ふつうは自分のASNに固定

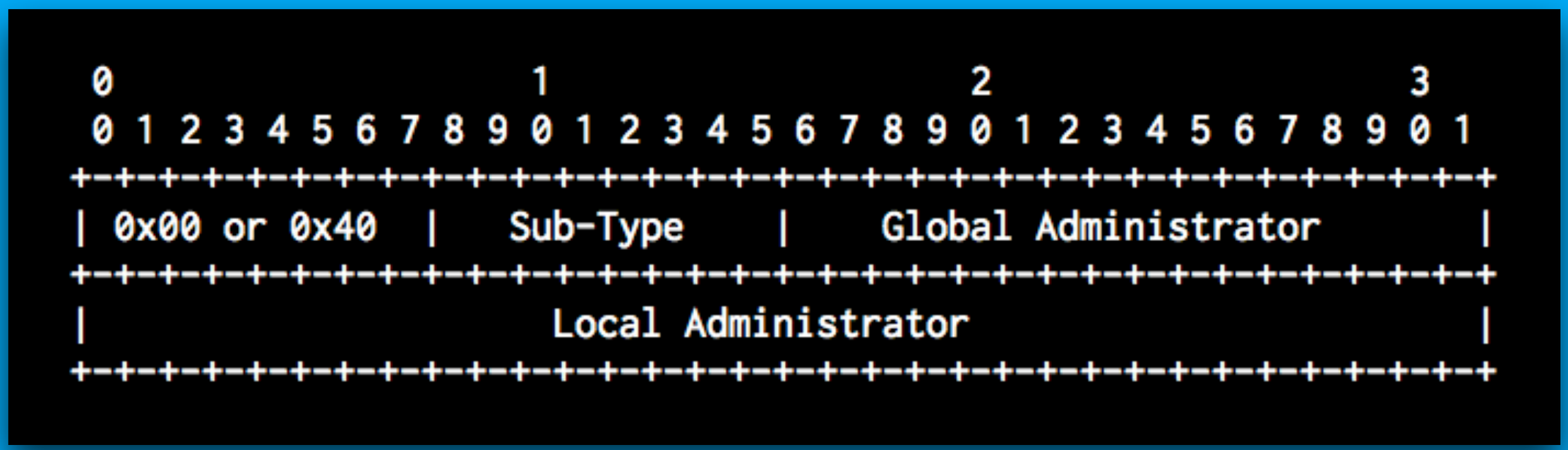
2914 : 2518

自由 🙌🎉🙌

2バイト：2バイト だけ？

・4バイト:2バイト、2バイト:4バイト のパターンもある

・RFC4360: BGP Extended Community



微妙 

...4バイトAS のひといますか？ 

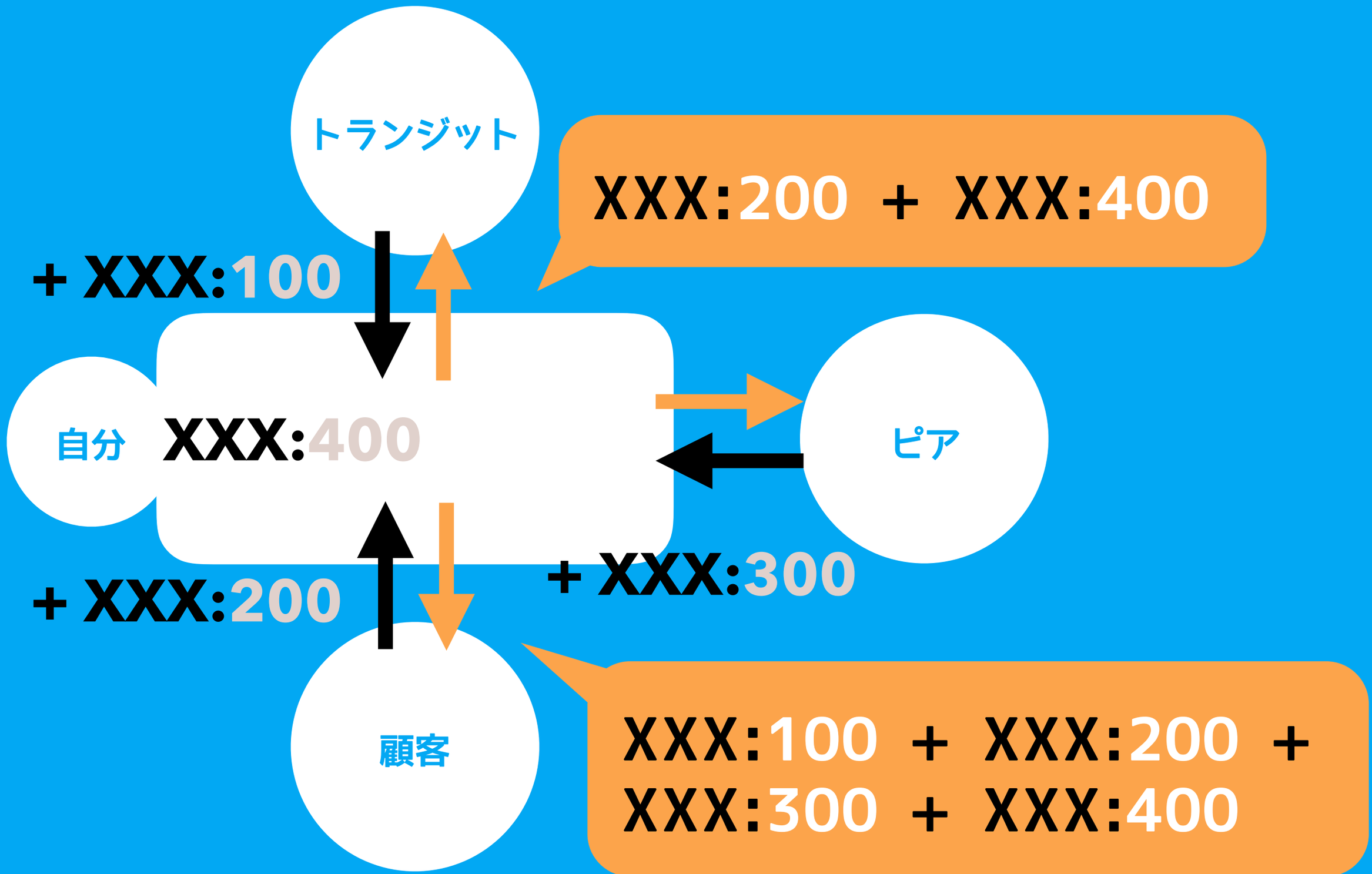
- 4バイト:2バイト、2バイト:4バイト
- 合計6バイトしか使えない
  - 何につかえるかは RFC7153 を
  - 前半4バイト + 後半4バイト使いたかったらアウト
  - もちろん無理やり使えるよ？
    - 後半4バイト → 2バイトにマッピングするとかね？

      バッドノウハウ臭...
- BGP Large Community に期待

  くわしくは吉村さんのパートで

使われかた

トランジット、ピア、顧客 の区別



ntt.net

2914:460	98	peer backup
2914:470	100	peer
2914:480	110	customer backup
2914:490	120	customer default
2914:666		blackhole

Customers wanting to alter their route announcements to other customers.

NTT Communications BGP customers may choose to prepend to all other NTT Communications BGP customers with the following communities:

Community	Description
2914:411	prepends o/b to customer 1x
2914:412	prepends o/b to customer 2x
2914:413	prepends o/b to customer 3x

Customers wanting to alter their route announcements to peers.

NTT Communications BGP customers may choose to prepend to all NTT Communications peers with the following communities:

Community	Description
2914:421	prepends o/b to peer 1x
2914:422	prepends o/b to peer 2x
2914:423	prepends o/b to peer 3x
2914:429	do not advertise to any peer
2914:439	do not advertise to any peer outside region

Note: 2914 is the ASN prepend in all cases. If used, 654xx:nnn overrides 655xx:nnn and 2914:429, 655xx:nnn overrides the 2914:42x communities.

Customers wanting to alter their route announcements to selected peers.

NTT Communications BGP customers may choose to prepend to selected peers with the following communities, where *nnn* is the peer's ASN:

Community	Description
65400:nnn	do not advertise to peer nnn in North America
65401:nnn	prepends o/b to peer nnn 1x in North America
65402:nnn	prepends o/b to peer nnn 2x in North America
65403:nnn	prepends o/b to peer nnn 3x in North America
65410:nnn	announce to peer nnn in North America, disregards 2914:429 and 65500:nnn

Get More Information

- Product Collateral
- Case Studies
- White Papers
- Audio & Video

Get Started

To find out which solutions will best benefit your business, contact one of our Account Managers.

- Click Here to Get Connected
- Call us at **1-877-868-8638**

Stay Connected

- Follow Us on Twitter
- Friend Us on Facebook
- Join Us on LinkedIn

https://www.us.ntt.net/support/policy/routing.cfm

ntt.net

2914 ： 480
LP をちょっと下げる

2914:460	98	peer backup
2914:470	100	peer
2914:480	110	customer backup
2914:490	120	customer default
2914:666		blackhole

Customers wanting to alter their route announcements to other customers.

NTT Communications BGP customers may choose to prepend to all other NTT Communications BGP customers with the following communities:

Community	Description
2914:411	prepends o/b to customer 1x
2914:412	prepends o/b to customer 2x
2914:413	prepends o/b to customer 3x

Customers wanting to alter their route announcements to peers.

NTT Communications BGP customers may choose to prepend to all NTT Communications peers with the following communities:

Community	Description
2914:421	prepends o/b to peer 1x
2914:422	prepends o/b to peer 2x
2914:423	prepends o/b to peer 3x
2914:429	do not advertise to any peer
2914:439	do not advertise to any peer outside region

Note: 2914 is the ASN prepend in all cases. If used, 654xx:nnn overrid
2914:42x communities.

Customers wanting to alter their route announcements to selec

NTT Communications BGP customers may choose to prepend to selected
peer's ASN:

Community	Description
65400:nnn	do not advertise to peer nnn in North America
65401:nnn	prepends o/b to peer nnn 1x in North America
65402:nnn	prepends o/b to peer nnn 2x in North America
65403:nnn	prepends o/b to peer nnn 3x in North America
65410:nnn	announce to peer nnn in North America, disregards 2914:429 and 65500:nnn

Get More Information

- Product Collateral
- Case Studies
- White Papers
- Audio & Video

Get Started

To find out which solutions will best benefit your business, contact one of our Account Managers.

- Click Here to Get Connected
- Call us at **1-877-868-8638**

Stay Connected

- Follow Us on Twitter
- Friend Us on Facebook
- Us on LinkedIn

65500 ： 2518
2518 には広告しない

https://www.us.ntt.net/support/policy/routing.cfm

EQUINIX ルートサーバー

EQUINIX | INTERNET EXCHANGE PORTAL

SERVICES HELP

MLPE BGP COMMUNITIES

All Equinix MLPE route servers support the ability for participants to more granularly control outbound announcements of their routes using BGP Communities. The base actions are to allow or deny all routes, with exceptions for both, and to allow the prepending of your ASN up to three times.

There is no requirement to use BGP communities in conjunction with the MLPE service, however they are commonly used by large peers to avoid taking traffic from other large peers over the MLPE.

Community information and Cisco configuration examples follow below. Please contact peering@equinix.com with any questions or comments.

Example Communities

Action	Community
Default Open (Announce to All)	24115:24115
Default Open Except AS12345	24115:24115 0:12345
Default Closed (Announce to None)	0:24115
Default Closed Except AS12345	0:24115 24115:12345
Include AS24115 in BGP Path to all participants	65501:24115
Prepend 1x to AS12345	65501:12345
Prepend 2x to AS12345	65502:12345
Prepend 3x to AS12345	65503:12345

Example Configurations

https://ix.equinix.com/ixp/mlpeCommunityInfo

EQUINIX ルートサーバー

EQUINIX | INTERNET EXCHANGE PORTAL

SERVICES HELP

MLPE BGP COMMUNITIES

All Equinix MLPE route servers support the ability for participants to more granularly control outbound announcements of their routes using BGP Communities. The base actions are to allow or deny all routes, with exceptions for both, and to allow the prepending of your ASN up to three times.

There is no requirement to use BGP communities in conjunction with the MLPE service, however they are commonly used by large peers to avoid taking traffic from other large peers over the MLPE.

Community information and Cisco configuration examples follow be

Example Communities

Action	Community
Default Open (Announce to All)	24115:24115
Default Open Except AS12345	24115:24115 0:12345
Default Closed (Announce to None)	0:24115
Default Closed Except AS12345	0:24115 24115:12345
Include AS24115 in BGP Path to all participants	65501:24115
Prepend 1x to AS12345	65501:12345
Prepend 2x to AS12345	65502:12345
Prepend 3x to AS12345	65503:12345

Example Configurations

0:24115 + 24115:2518
2518 にだけ広告する

https://ix.equinix.com/ixp/mlpeCommunityInfo

BGP Community の使われかた

- 他社の決めかたは めちゃくちゃ参考になる
  - 他社のBGP Community まとめサイト
    https://onestep.net/communities/

- 外から見えているものがすべてではない
  - 内部管理のためのBGP Community もある

自分たちでも決めてみよう

基本戦略

大量のものは扱いにくい → タグをつけて扱いやすく
扱いやすい粒度で、扱いやすいタグをつける

1. BGP Community のマッチ方法を知る
2. 自分たちにとって便利なようにタグを決める
   - どんな粒度、どんな方法でTEしているか
   - 一発で決めなくていいよ！

     ベストな設計に至るには時間が必要 💩👌
3. それは顧客にとっても便利なのでは？
   - 自分が便利に使っているものを公開していく
   - ここは慎重に 🚧👷🚧

BGP Community のマッチ方法を知る

BGP Community を定義する前に、どのようにマッチできるか知っておく必要がある。マッチしづらいグループ分けは避けないといけない。だいたい正規表現が使える。範囲指定できるものも。

正規表現がないとかなりつらい

Juniper

`XXX:1..$`	`XXX:*`	`[XXX:100 XXX:200]`

IOS

`XXX:1..$`	`XXX:.*`	`XXX:100 XXX:200`

IOS-XR

`XXX:1..$`	`XXX:.*`	`XXX:100 XXX:200`	`XXX:[100-120]`

BGP Communnity の意味を考える

どんなCommunity にどんな意味を持たせるか

- まず単なるラベルとして
  - 経路の種別
    - 他社の経路: トランジット or ピア or 顧客
    - 自分の経路: サービス別
- ロケーション
  - 国内 or 海外、関西 or 関東
  - 事業規模によって粒度がちがう

BGP Communnity の意味を考える

- 経路操作 (パスアトリビュート)
  - Local Preference
  - AS_PATH Prepend
  - 無条件に操作する or 特定のNeighbor に出すときだけ操作

全部に効く / 一部に効くなど、粒度パターンが必要かも

- 経路操作 (広告する or しない)
  - 特定のNeighbor だけに出す / 出さない
- トラフィック アカウンティング
  - SCU / DCU (Juniper)
    - http://www.juniper.net/techpubs/en_US/junos15.1/topics/concept/source-class-usage-options-junos-nm.html
  - BGP Policy Accounting (Cisco)
    - http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13760-38.html

えっ、なんかめちゃくちゃ多いんですけど…

- なので、基本は 自分たちに便利なように
  - ドッグフーディング重要
  - 「顧客にとって便利だろうか」は一旦忘れましょう
- 粒度で悩んだら、とりあえず細かく
  - 10 刻みの階段を後で細分化 → 厳しい 😵
  - 10 刻みの階段を後でまとめる → OK 🙂
- 直交するように

- ダメな例 ❌
  - 関西のトランジット　→ XXX:100
  - 関東のトランジット　→ XXX:110
  - 関西のピア　→ XXX:200
  - 関東のピア　→ XXX:210

- いい例 ⭕
  - 関西　→ XXX:100
  - 関東　→ XXX:110
  - トランジット → XXX:300
  - ピア　→ XXX:310

関西のトランジット ＝
XXX:100 XXX:300

これは たぶんわかりやすい 😭

直交してない

- じゃあ、これは？
  - ASN Y に広告しない → 65000:Y
  - アジアでは、ASN Y に広告しない → 65001:Y

- こうでは？
  - ASN Y に広告しない → 65000:Y
  - アジアでは、ASN Y に広告しない → 65000:Y + XXX:100
  - XXX:100 = 効果をアジアに限定する

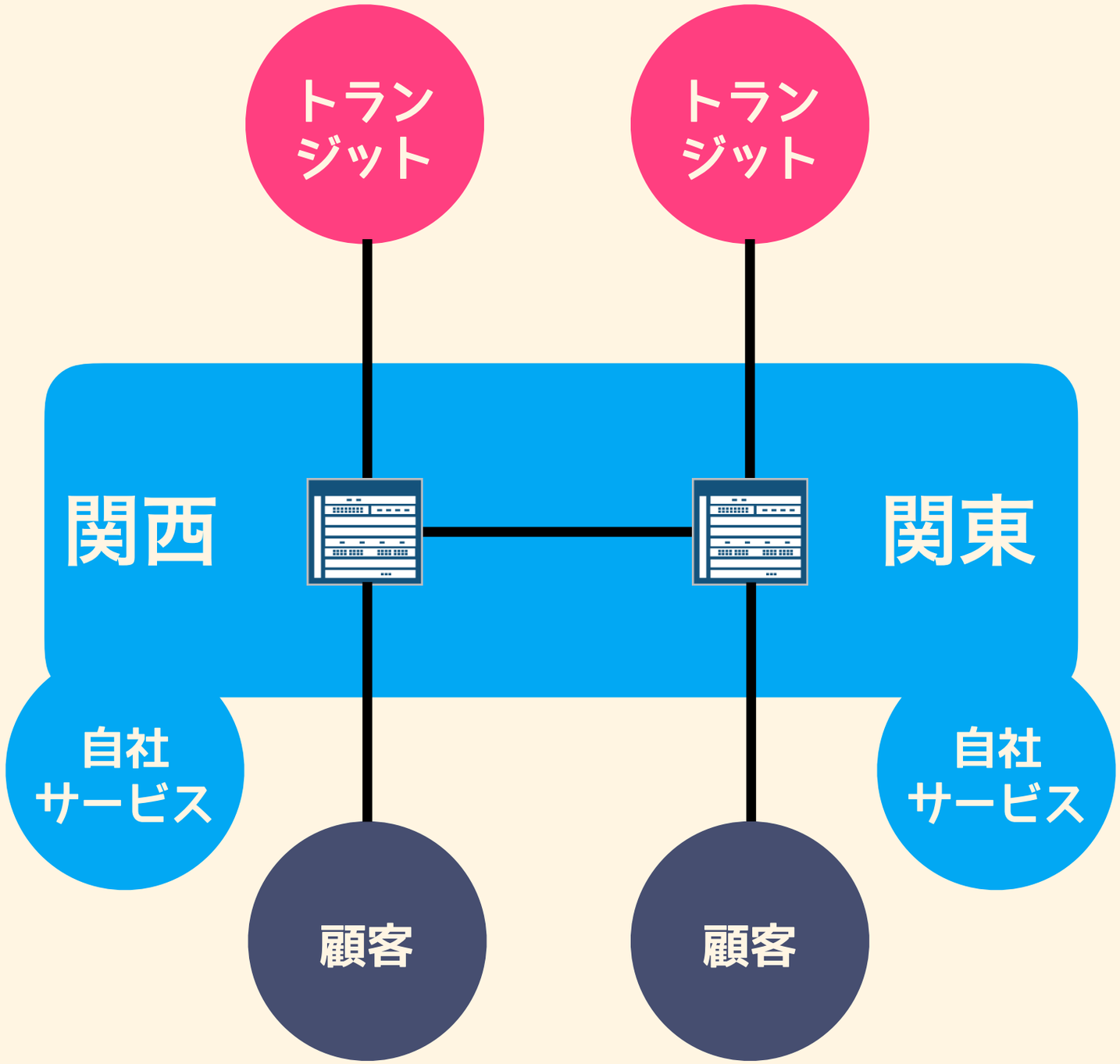
わかりやすさ重要。
自分たちに便利なように

それはそうなんだけど、

- わかりやすい？
- 他のCommunity とのコンボがちゃんと動く？

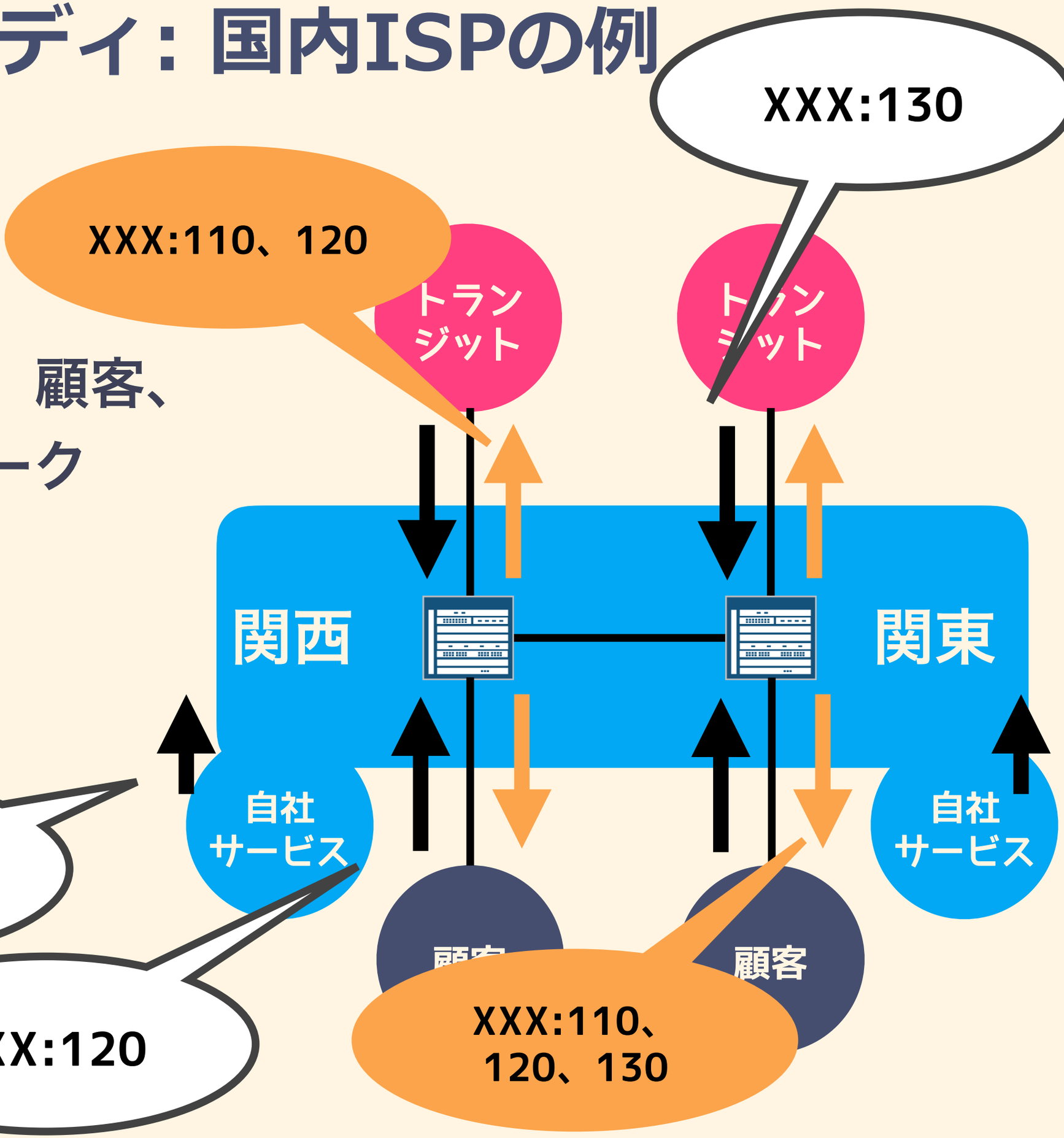
ケーススタディ：国内ISPの例

- 関東、関西に分かれてる
- 自社サービスの経路は
  関西 / 関東に閉じる
- インターネットの経路は
  アジア / US / EU くらいの
  粒度で見てる
- 「デフォルト + モバイル
  っぽいASの経路だけ欲しい」
  という顧客が複数いる



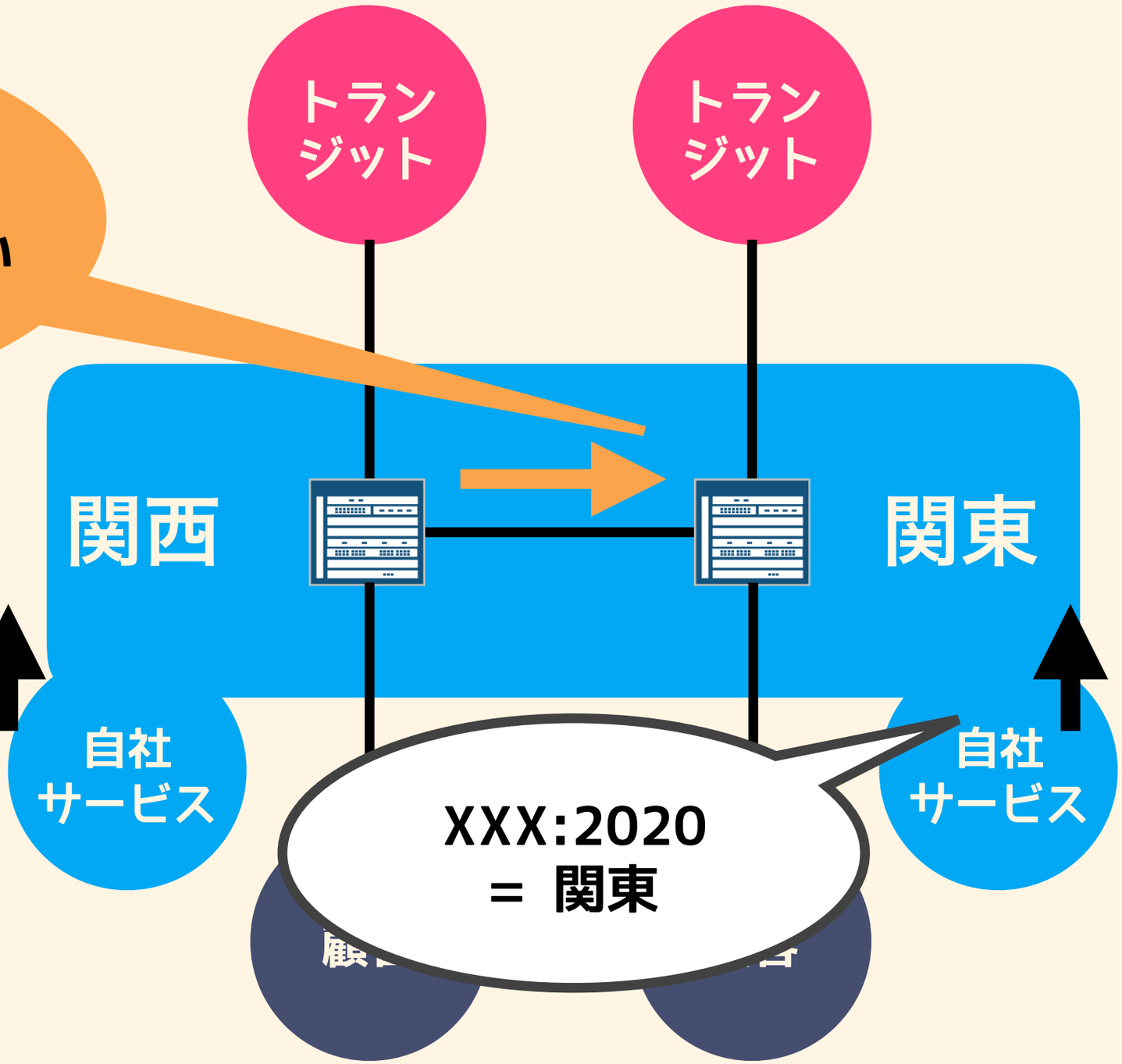
ケーススタディ：国内ISPの例 (1) 基本

- 受信
  - トランジット、顧客、自社経路をマーク
  - LP をセット
- 送信
  - フィルター



ケーススタディ：国内ISPの例
(2) 自社サービスの経路は関西 / 関東に閉じる

受信時
XXX:205 が付いていて、
XXX:2020 が付いてない
経路を捨てる



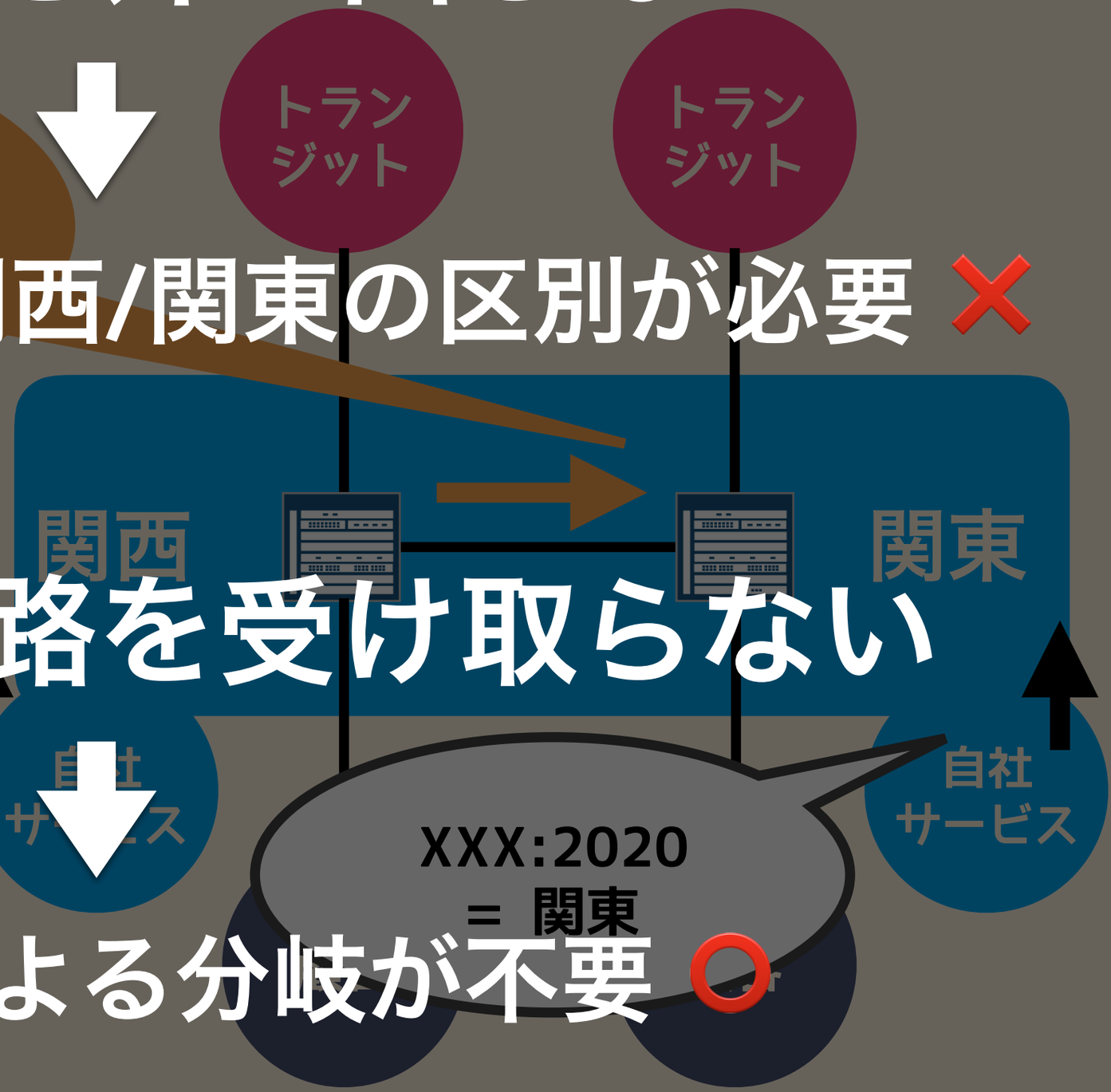
XXX:205 = リージョン
に閉じるフラグ
XXX:2010 = 関西

ケーススタディ：国内ISPの例
(2) 自社サービスの経路は関西 / 関東に閉じる

関西の経路を外に出さない

受信時
XXX:205 が付いていて、
XXX:2020 が付いてない、
経路を捨てる

Neighborごとに関西/関東の区別が必要 ✕



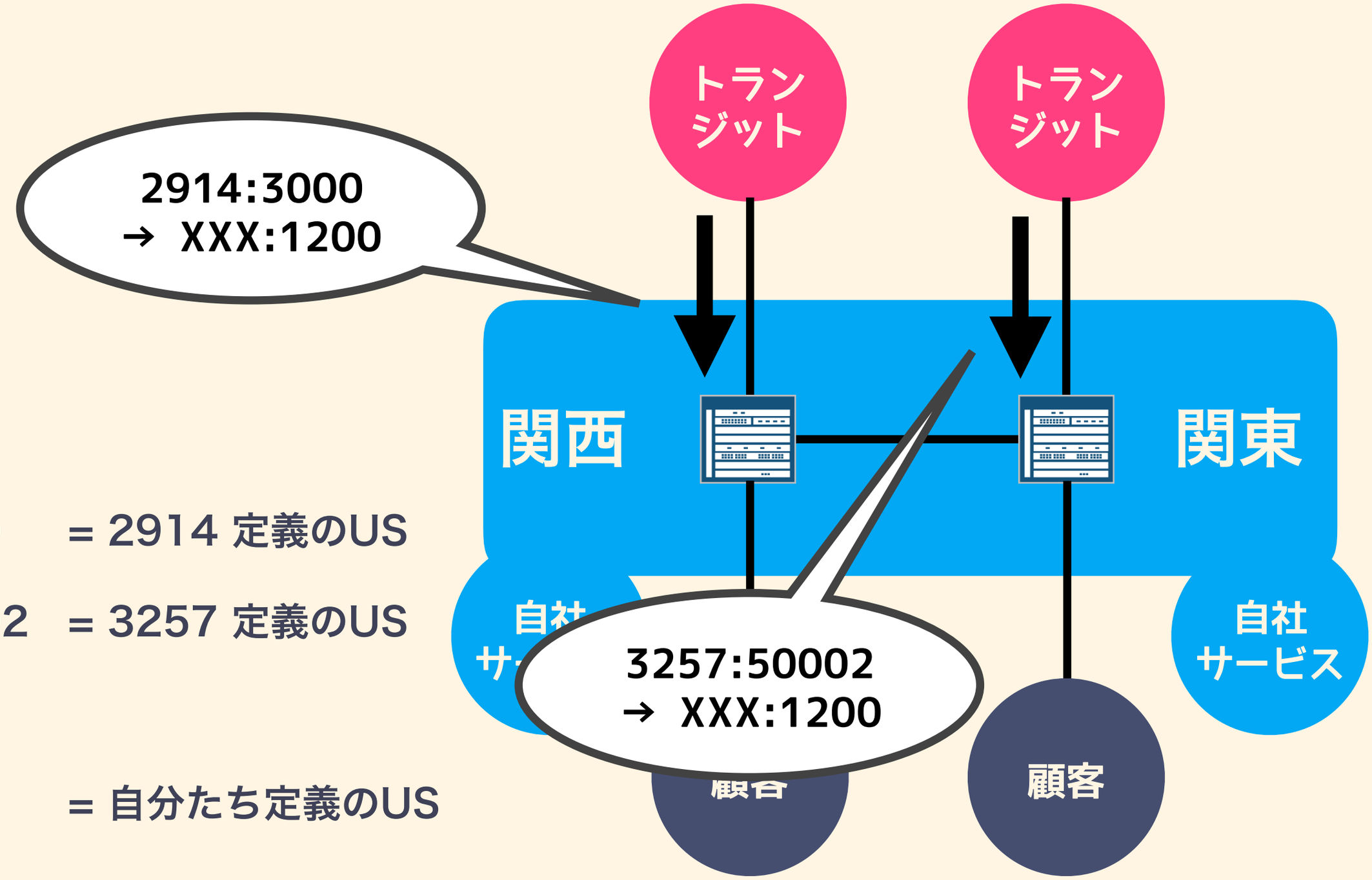
関東以外の経路を受け取らない

XXX:205 = リージョンに閉じるフラグ
XXX:2020 = 関東

XXX:2020
= 関東

Neighbor による分岐が不要 ⭕

ケーススタディ：国内ISPの例
(3) インターネットの経路はアジア / US / EU くらいの粒度で見てる



2914:3000　　= 2914 定義のUS

3257:50002　= 3257 定義のUS

→

XXX:1200　　= 自分たち定義のUS

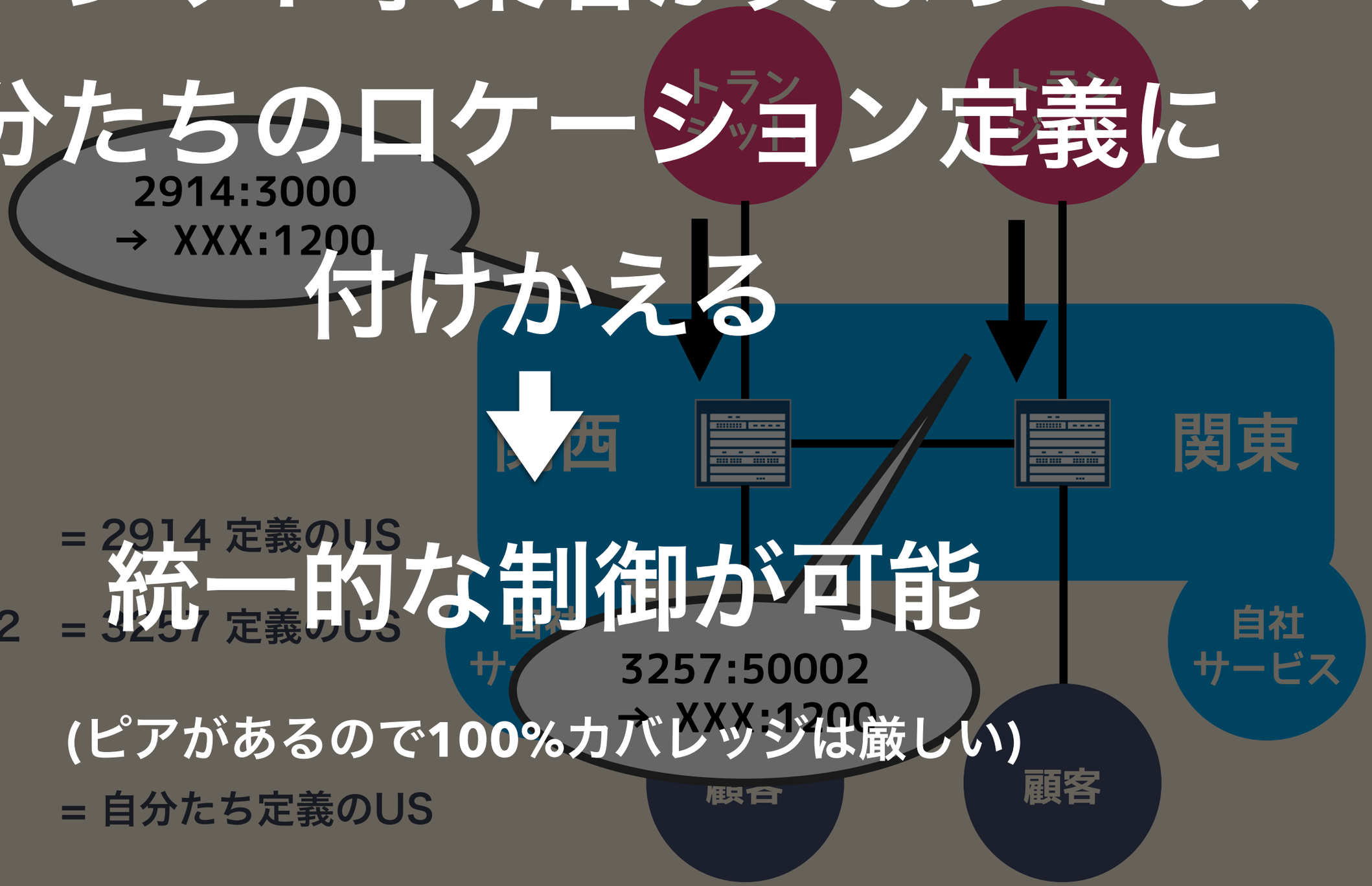
に付けかえる

ケーススタディ：国内ISPの例
(3) インターネットの経路はアジア / US / EU くらいの粒度で見てる

トランジット事業者が異なっても、

自分たちのロケーション定義に

付けかえる



統一的な制御が可能

(ピアがあるので100%カバレッジは厳しい)

2914:3000　= 2914 定義のUS

3257:50002　= 3257 定義のUS

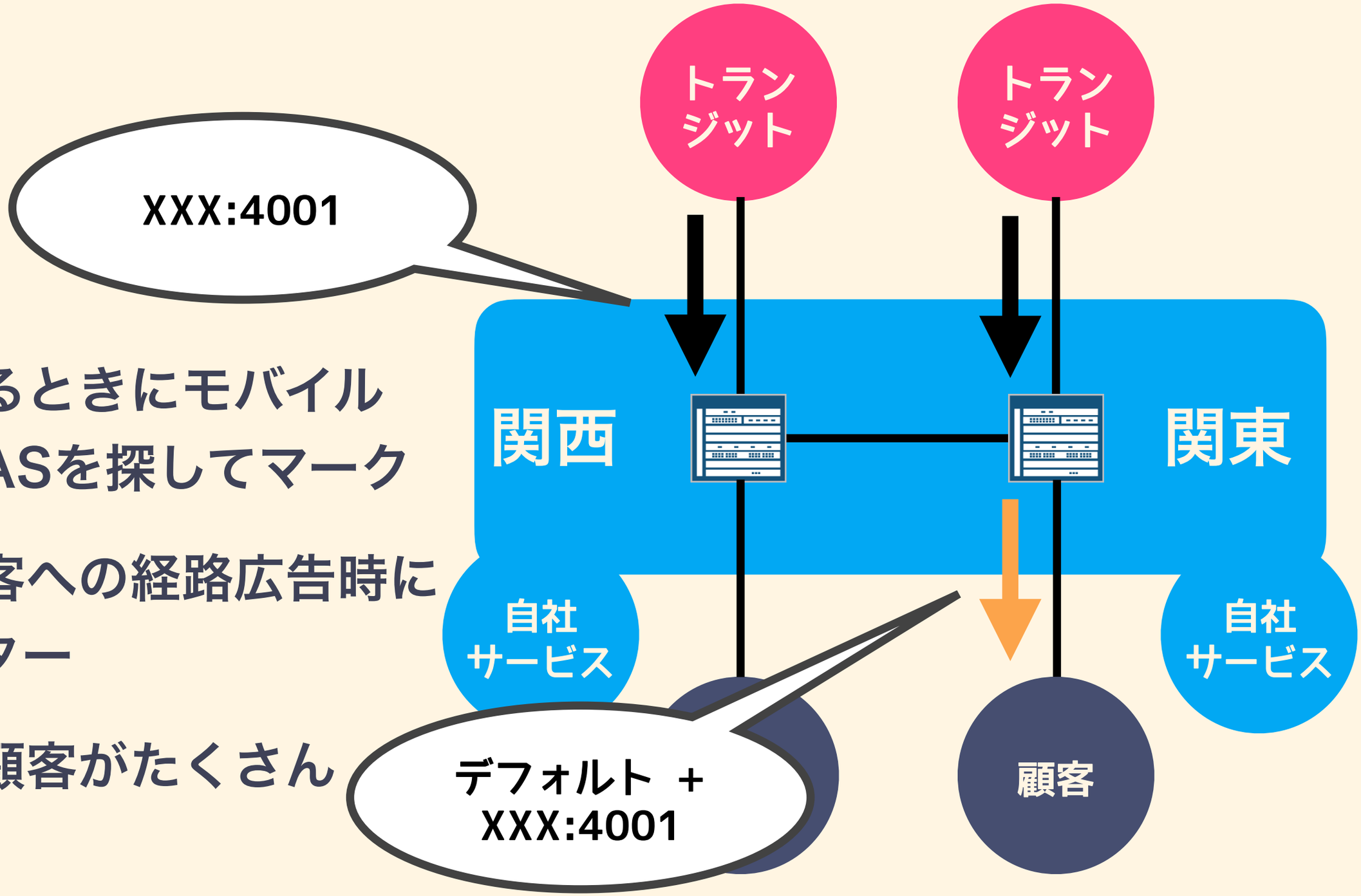
→

XXX:1200　= 自分たち定義のUS

に付けかえる

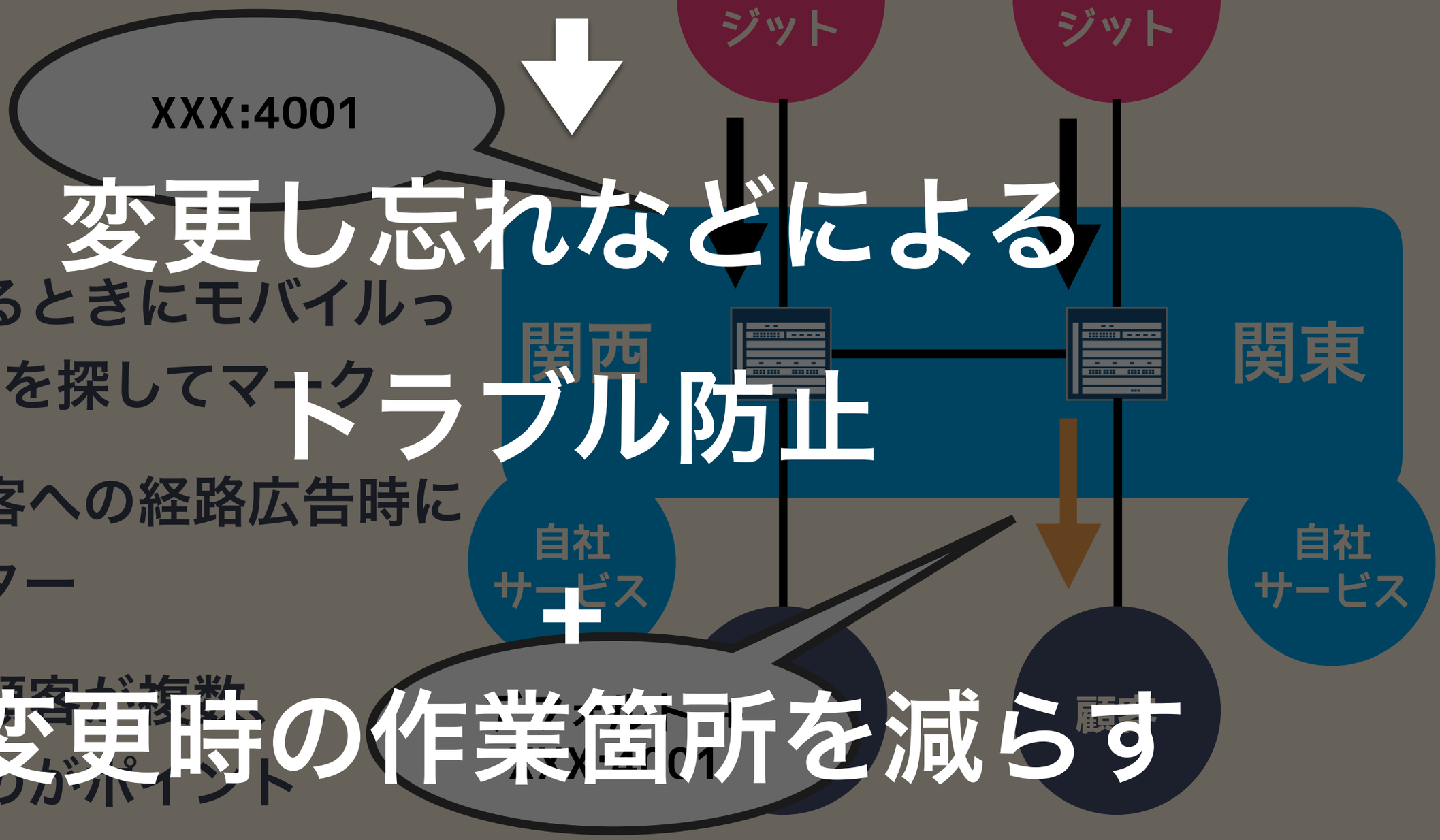
ケーススタディ：国内ISPの例
(4)「デフォルト + モバイルっぽいASの経路だけ欲しい」という顧客が複数いる

- 受け取るときにモバイルっぽいASを探してマーク
- 特定顧客への経路広告時にフィルター
- そんな顧客がたくさんいる



ケーススタディ：国内ISPの例

- 「ゴールデンタイムだけ帯域が欲しい」という顧客が複数いる
- 受け取るときにモバイルっぽいASを探してマーク
- 特定顧客への経路広告時にフィルター
- そんな顧客が複数というのがポイント



複雑な判定ロジックを複数箇所に書かない工夫

↓

変更し忘れなどによるトラブル防止

＋

変更時の作業箇所を減らす

TIPS

「自由になる2バイトの部分、広すぎて手にあまる 🤕」

XXX:　0〜　99　← 狭いので使わない。例えば:5と:50は違う種別に見える

...

XXX: 200〜 249　← 似た機能はケタ数と頭1~2ケタを揃える

XXX: 250〜 299　(予約)

...

XXX:1000〜1099

XXX:1100〜1199　(予約)

XXX:1200〜1299　← パターンが多い機能ほど広く使う

XXX:1300〜1399　(予約)

...

まとまり感

隣をあけておくと、見直しで
リナンバーしやすい

すべてにハマる設計はない

- 解決したい問題 / 規模によって、ベストな BGP Community 設計は違う
- まずは小さくやってみましょう
  - やってみないと分からない
  - 「似たパターンでこれも欲しい」ってどんどん出てくる

- たまに「これ顧客が喜ぶんでは」というのが見つかる

使っているBGP Community、顧客に公開するべき？

- 顧客の問題を解決しそうなら公開するべき
  - 自分たちが便利 = 自分たちの問題を解決した
  - であれば、ついでに顧客の問題を解決できるかも
    - インターネットだから。パケットは似たところを通る

- 制御権を一部渡すことで、顧客のスピードが上がったり、自分たちのOPEXが下がりそうなら公開するべき
  - 「メールでオーダーください」は遅い
  - BGP Community = API

10年前の、この状況に似てる

Be the first to clip this slide

AWSクラウドの起源は、

Amazon社内の

ビジネス課題を解決するために

生まれた

7 of 53

AWSが誕生するまでの秘話

3,418 views

http://www.slideshare.net/horiyasu/20140703-geechs

10年前の、この状況に似てる

自分たちでは解決しづらいが、

トランジット事業者ならなんとか

できそうな問題がある

ビジネス課題を解決するために

生まれた

「お願いすればやってくれる。遅いけど」

より、「API を使って自分たちで

さっと解決できる」ほうがいいでしょう？

AWSが誕生するまでの秘話

3,418 views

http://www.slideshare.net/horiyasu/20140703-geechs

対する懸念

・いちど BGP Community を公開したら、
変えられないのでは？

・いまの設計が失敗だったら？

そこで Large BGP Community ですよ

Large BGP Communities

/ Spec / Implementations / About / News / FAQ / Talks / Subscribe

Large BGP Communities are a novel way to signal information between networks. Large BGP Communities are easy to use, implement and deploy.

An example of a **Large BGP Communities** is: `2914:65400:38016` . **Large BGP Communities** are composed of three 4-byte integers, separated by a colon. This is easy to remember and accomodates advanced routing policies in relation to 4-Byte ASNs.

This website brings together implementors and end users of the **Large BGP Communities**.

Recent News articles:

INEX Deploys Large BGP Communities in Production

Nov 7, 2016

The Internet Neutral Exchange Association (INEX) is the first netw

http://largebgpcommunities.net/

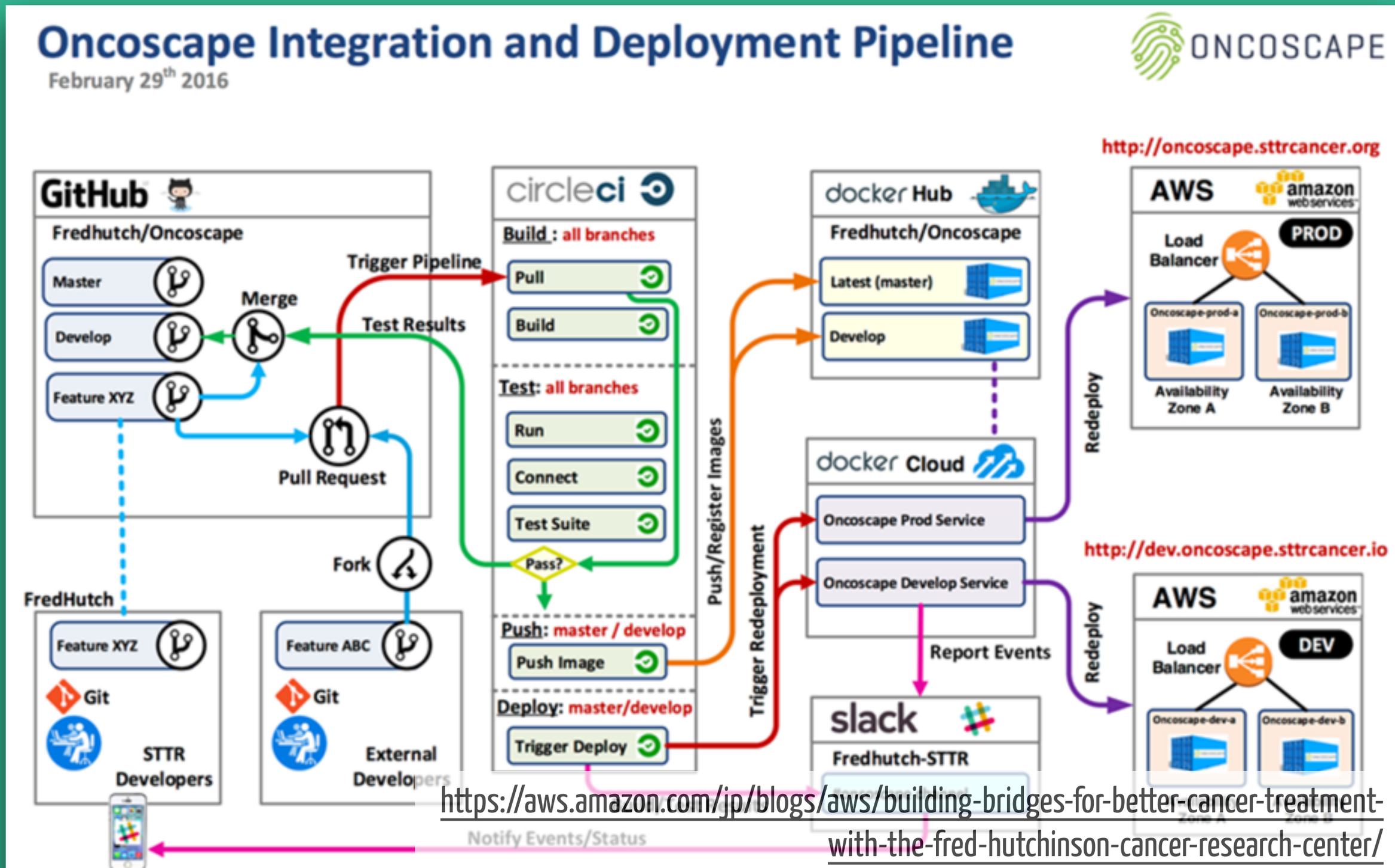
何を言ってるのか

- 近い将来、Large BGP Community なるものが来る
- 2バイト:2バイト → 4バイト:4バイト:4バイト
- リナンバーのチャンス！ (顧客への言い訳的に)
- それまでに知見をためて、設計を見直すことができる

繰り返しだけど、
BGP Community = API

- ネットワーク機能を外から発動する
- 自分たちがうれしいものは、顧客にとってもうれしい
- 他社の API をマッシュアップして自社サービスにつなげることもできる
- BGP Community は Transitive なので、本質的に連携しやすい
  - どこかから伝われば発動する

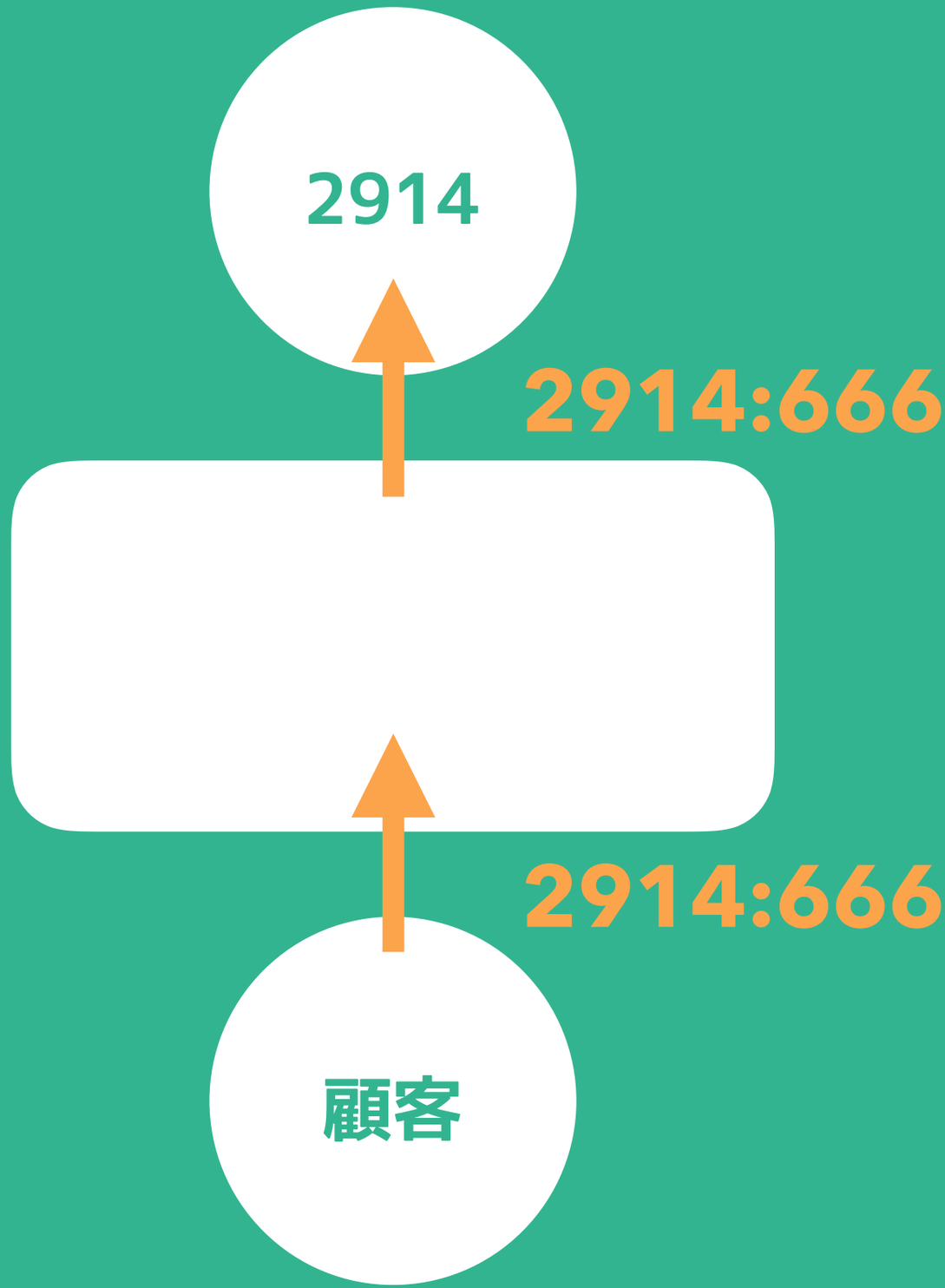
お隣さんは、
API 連携で発展している…⚡⚡⚡



https://aws.amazon.com/jp/blogs/aws/building-bridges-for-better-cancer-treatment-with-the-fred-hutchinson-cancer-research-center/

BGP Community を消さないことも価値

- たとえば2914のRTBH、
  顧客が使えるとしたらどう？
- AUP的な問題はあるかも…?
- 経路フィルターは開いてる？



まとめ

- BGP Community の設計方法
  - まずは自分たちが便利に使えることが重要
- 顧客に公開することで、
  顧客の問題を解決することを考えよう
- BGP Community は API
  - 呼んだり呼ばれたりしよう
- BGP Community を消さない選択肢もご検討を

Questions ?