

4-byte AS の世界へようこそ

JANOG20 (2007.7.12)

外山勝保^(*1)、藤崎智宏、新延史郎、松本存史

NTT情報流通プラットフォーム研究所

(^{*1}) 現所属: インターネットマルチフィード株式会社

本日の論点

□ インターネットサービスを継続して提供する「プロバイダの目線で」、4バイトASへの対応を議論したい

■ 4バイトASへの対応(移行)は容易なのか

□ 一気に変える？それとも徐々に変えていける？

■ 4バイト対応(移行)するとしたらいつ頃がよい？

□ 設備計画・工事計画、お客様への連絡、などもあるし。

■ そもそもいま、何か問題は生じていないか？ 対策は？

などなど。。

□ みなさんと情報共有・議論したいと思います。

アウトライン

(1) 4バイトASについて

- (1.1) 背景
- (1.2) 4バイトASの基本的な仕組み
- (1.3) 4バイトAS割当スケジュール

(2) 4バイトASの現状

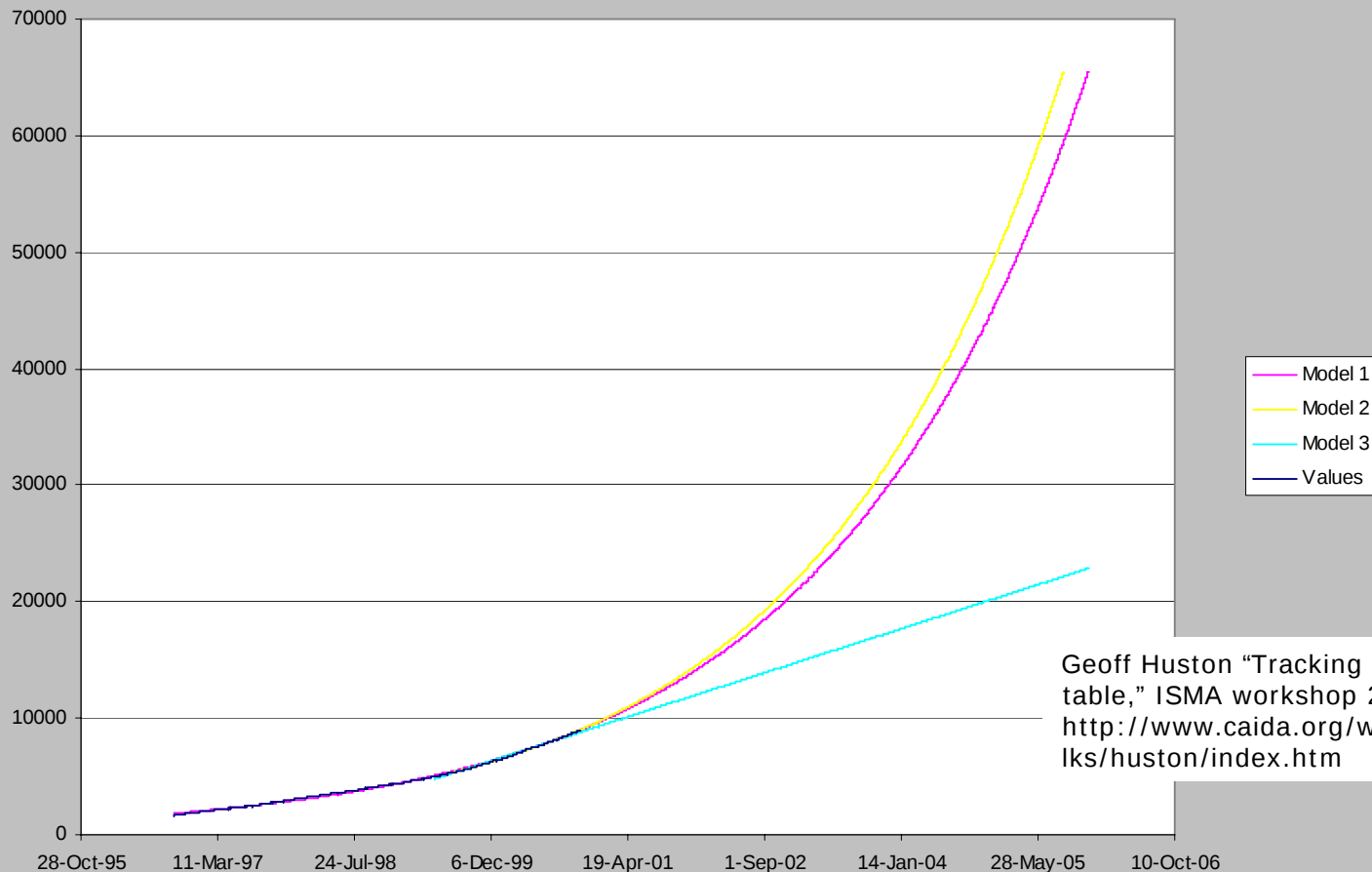
(3) プロバイダの対応

(4) 課題と議論

(1.1) 背景: AS番号の枯渇(1)

□ AS番号の消費トレンド議論

- 2000年頃、すでにGeoff Huston氏、Tony Bates氏、J.Scott氏などがAS番号についても割当トレンドと将来予測を実施。
- Geoff Huston氏は、当時、指数関数的に伸びると2005年にはAS番号は足りなくなるペースと予想した。



Geoff Huston "Tracking the Internet's BGP table," ISMA workshop 2000, CAIDA
<http://www.caida.org/workshops/isma/0012/talks/huston/index.htm>

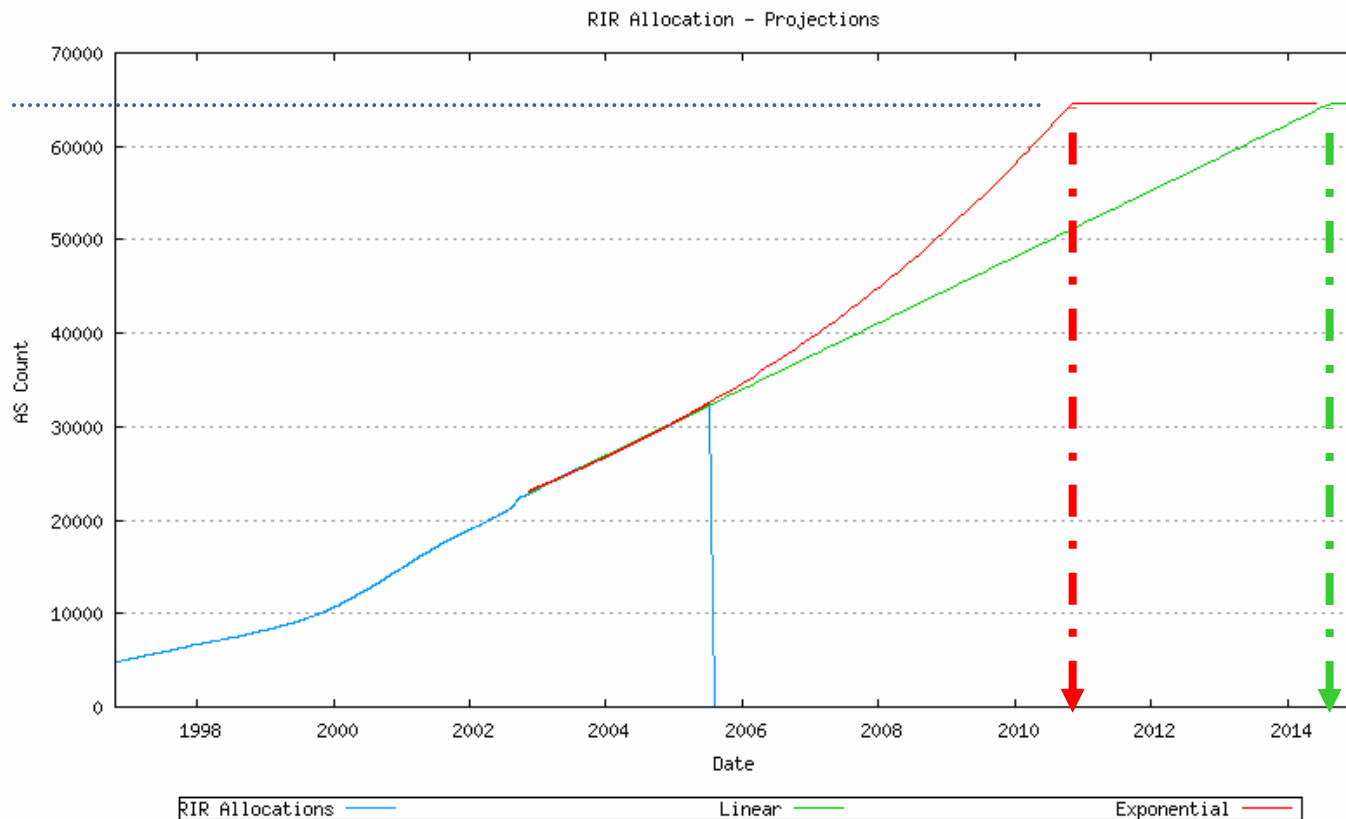
(1.1) 背景 AS番号の枯渇(2)

□ 近年の予測

■ 2005年8月当時の予測では、2010年後半頃に枯渇

□ AS番号は16ビット空間 → 上限64511個(0とプライベートASを除く)

64511



Goeff Huston, "Exploring Autonomous System Numbers," ISP Column, Aug 2005
<http://www.potaroo.net/ispcol/2005-08/as.html>

(1.2) AS番号の枯渇対策技術 -- 4バイト化

□ IETF idr working group で長一く議論

- 最初のドラフト draft-chen-as4bytes-00 は、2000年11月に submit されている

- この時点で“New AS PATH” (AS4_PATH) は盛り込まれていない

- 2000年12月のIETF49で議論

- idr working group

- 2001年1月に draft-ietf-idr-as4bytes-00

- RFC直前のI-D: draft-ietf-idr-as4bytes-**13**

- 2007年5月にRFC4893に。

(1.2) AS番号の4バイト化

□ 設計目標

- 既存の2バイト対応BGPスピーカには影響を与えないこと

□ 4バイト化の要点

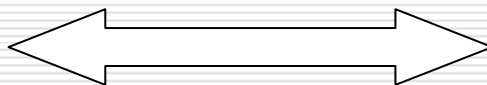
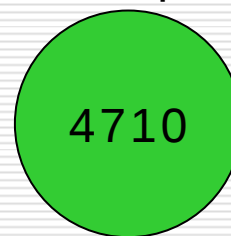
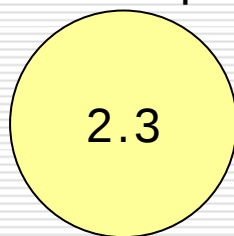
- BGPでピアを張る相手が4バイトかどうかを確認する
 - OPEN message の capability チェックを使う
- 4バイト対応ルータは、AS番号をすべて4バイトとして取り扱う
 - AS_PATH属性、およびAGGREGATOR属性
- 4バイト対応ルータが2バイト対応ルータと通信する際に、4バイト側がすべてをケアする
 - 2バイト側へ経路広告する際に、AS_PATH属性やAGGREGATOR属性の中を2バイトへ書き換える。
 - AS4_PATH属性を新たに定義(Optional Transitive)し、4バイトで表現したAS_PATHを、2バイト対応ルータが「スルー」して流せるようにする。
 - 2バイト対応ルータから受け取ったAS_PATH(2バイト表現)と、「スルー」されたAS4_PATH属性から、4バイト対応ルータは「4バイトのAS_PATH」を復元する。

(1.2) 4バイトASのしくみ

□最初のネゴシエーション(4バイト→2バイト)

New BGP Speaker

Old BGP Speaker

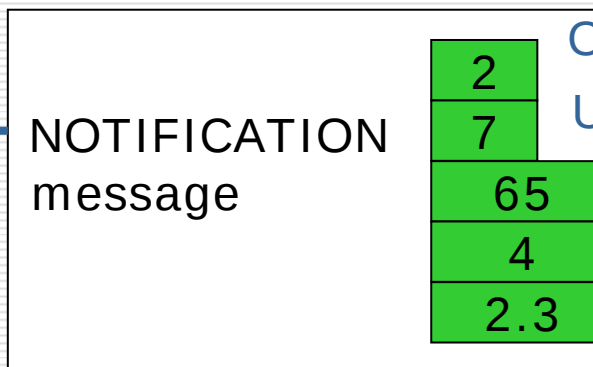
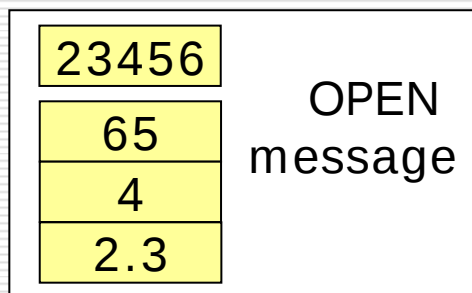


My AS

Capability
Code

Length

ASN



OPEN msg err
Unsupported

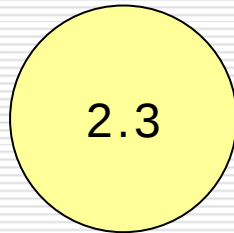
OPENと同じ
情報を返す

Capability (65)
を知らないから、
相手は4バイト対
応していない

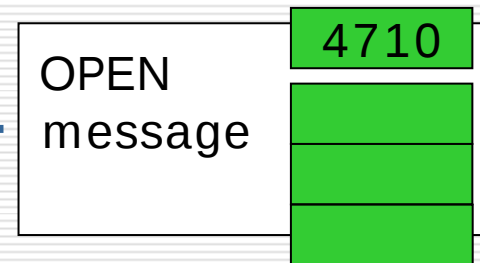
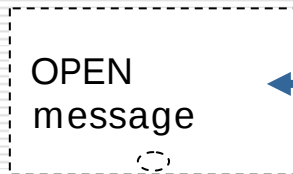
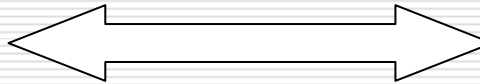
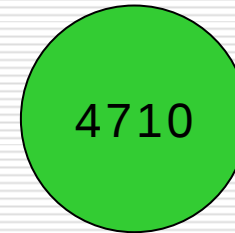
(1.2) 4バイトASのしくみ

□最初のネゴシエーション(2バイト→4バイト)

New BGP Speaker

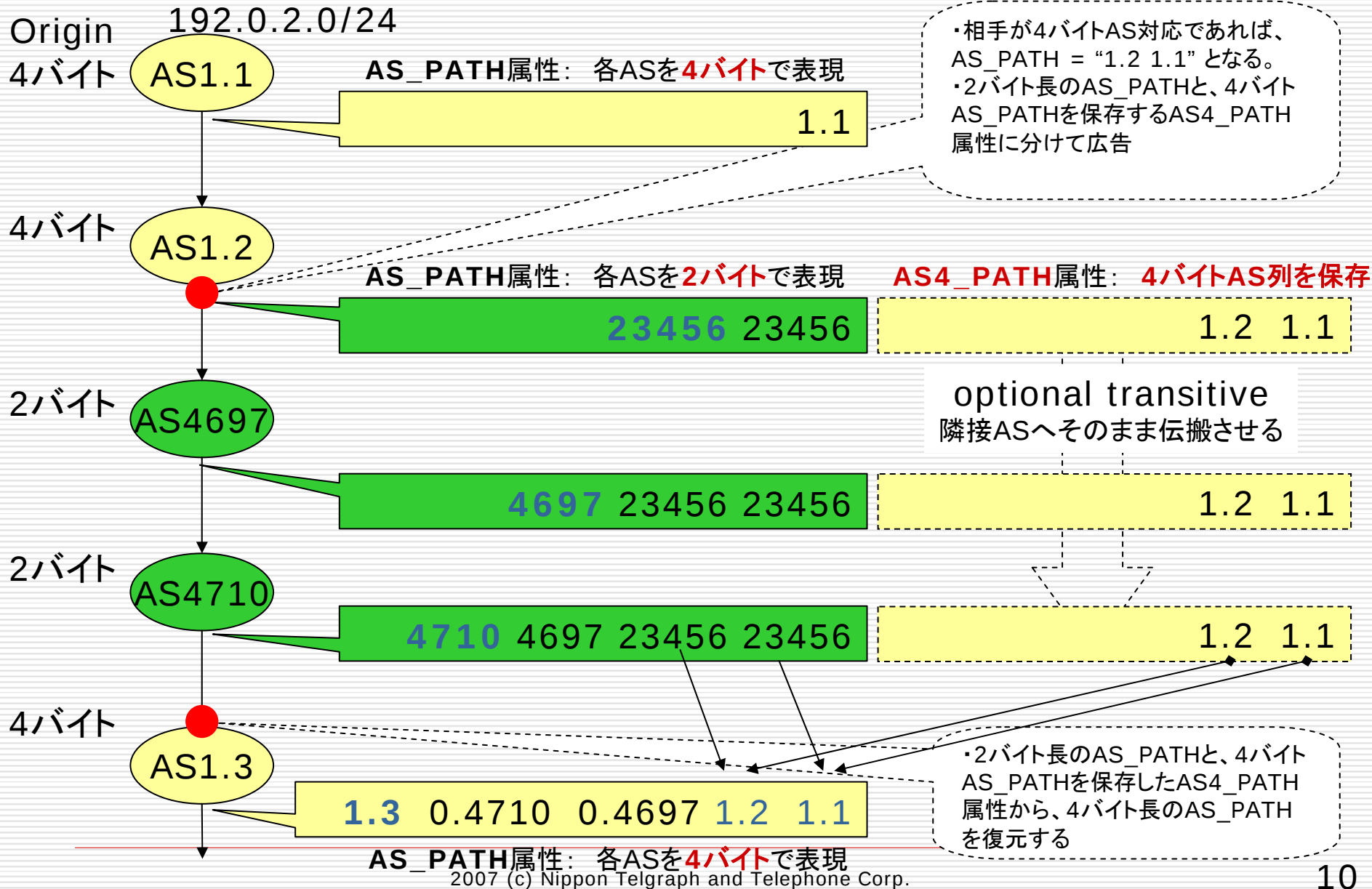


Old BGP Speaker



Capability (65)
が入っていないから、
相手は4バイト
対応していない

(1.2) 4バイトASのしくみ (AS_PATHの扱い)



(1.2) 4バイトASのしくみ(ループ検出)

- AS-PATHのなかに自ASが含まれていることで、ループを検出する。
 - 2バイトASの世界でも、2<->4バイト変換がきちんと行われていれば、自ASをas-pathに発見してループ検出できる。
 - ただし、4バイトASが含まれている経路(AS23456が含まれている)を2バイトASにてaggregateすると、その後as-pathが復元できずループ検出できない場合がある。

 - as-path中にAS23456が飛び飛びで複数含まれていても、ループとはみなさない。
 - 例:
[4byte] 4697 4710 2.3 2.7 4713 2.2 2914
[2byte] 4697 4710 **23456** **23456** 4713 **23456** 2914
- ※ループではない

(1.3) 4バイトASの配布スケジュール



4バイトAS番号の分配について

7

2006年5月

ポリシー施行決定

APNIC理事の承認待ち

2
バイト

- ✓ベンダー・ISPへの周知
- ✓ベンダーによる機器対応
- ✓ISPルータの4バイトAS対応
- ✓2バイト・4バイト間通信試験

2007年1月

希望者に4バイトASを分配

デフォルトは2バイトASを分配

2
バイト

4
バイト

早期4バイトAS利用者を救う
にはこの時点で対応が必要

2009年1月

デフォルト4バイトASを分配

希望者には2バイトAS番号を分配

2
バイト

4
バイト

基本的にこの時期を目指して
対応完了(2.5年後！)

1.5年後

2010年1月

2バイト、4バイト区別なく分配

どちらかを明示的に希望することは
できなくなる

区
別
な
し

ここで対応できていないと
新たにASを取得するNWと
通信できる方法がない

You are here!

Copyright © 2006 JPNIC All Rights Reserved.

奥谷泉, “IPアドレスのルール作りとオペレーションコミュニティ”, JANOG18, Jul. 2006
http://www.janog.gr.jp/meeting/janog18/files/IPadr_Okutani.pdf

アウトライン

(1) 4バイトASについて

(2) 4バイトASの現状

- (2.1) 4バイトAS割当状況
- (2.2) 4バイトASに係わるアドレス広告状況
- (2.3) 4バイトAS対応ソフトウェア

(3) プロバイダの対応

(4) 課題と議論

4バイトASの割り当て

□ IANAからRIRに割り当て委譲済

■ Specially designated Autonomous System Numbers:

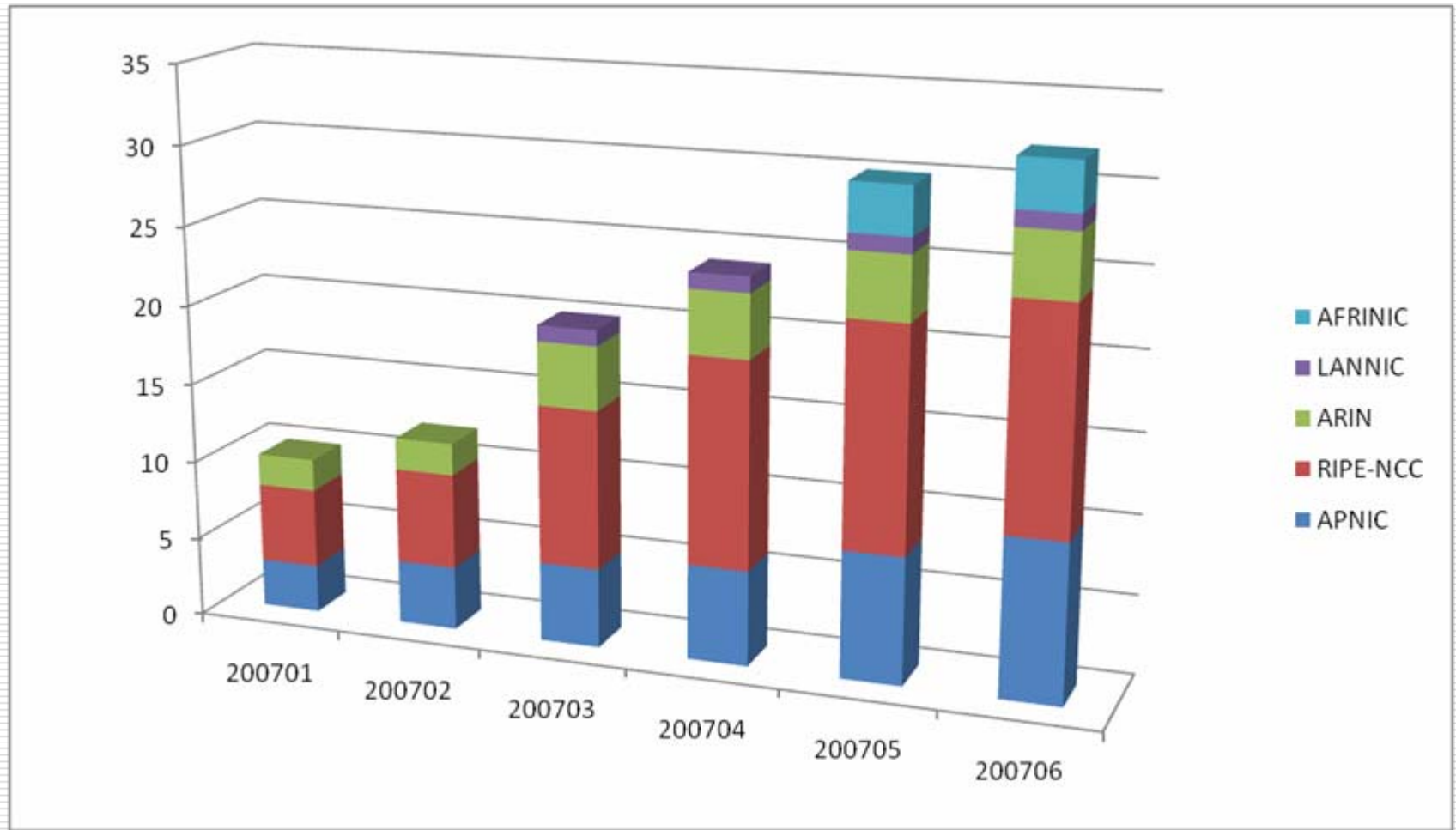
□ 23456 AS_TRANS [RFC4893]

□ 割り当て(2007年7月4日現在)

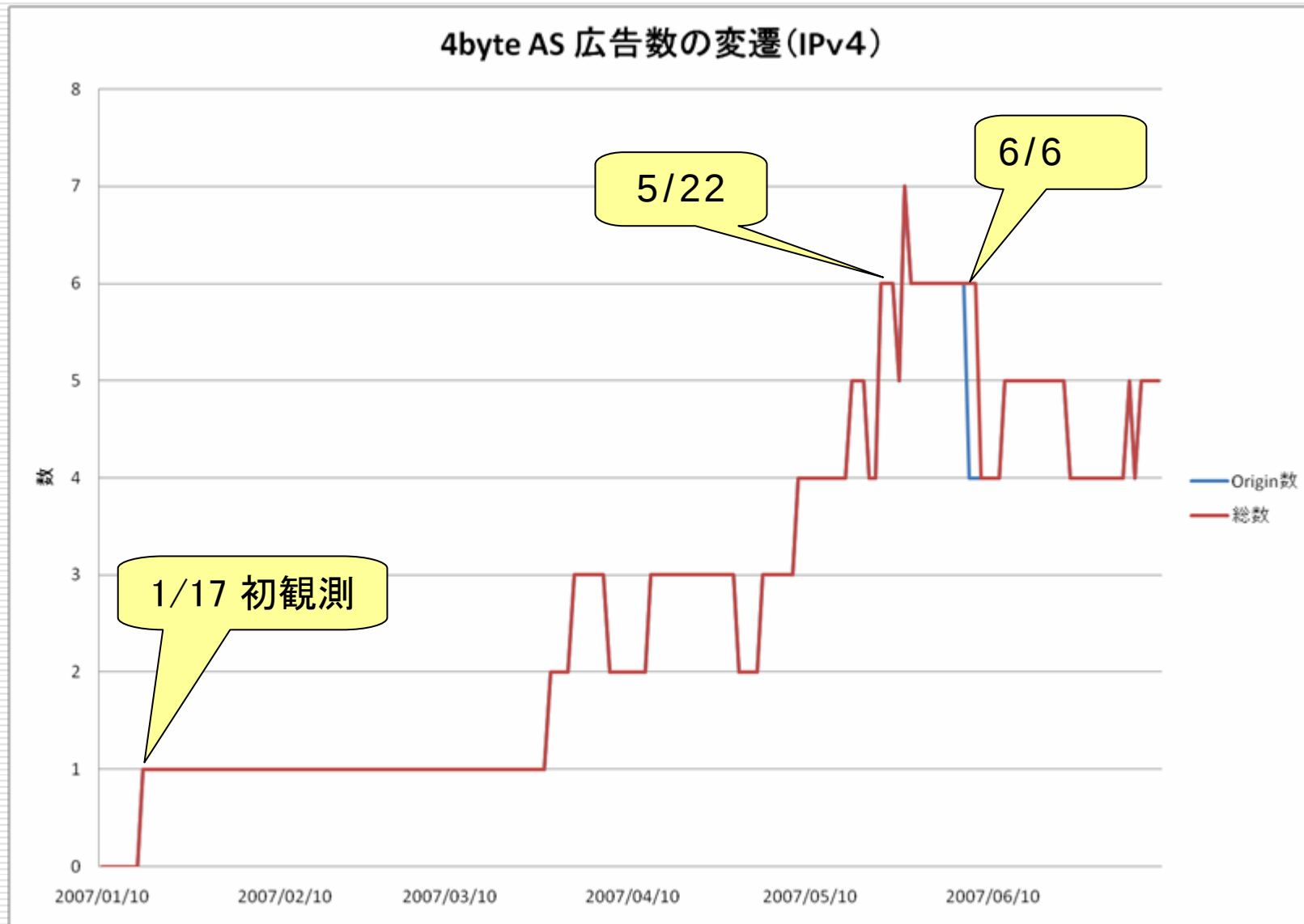
範囲	割り当て
1.0-1.65535	Reserved
2.0-2.1023	APNIC
3.0-3.1023	RIPE NCC
4.0-4.1023	LACNIC
5.0-5.1023	AfriNIC
6.0-6.0123	ARIN
65535.65535	Reserved

AS2.5 - AS2.14 は, JPNIC-4Byte-ASBLOCK-AP として委譲されている

4バイトAS取得状況



4バイトASの状況 ～IPv4広告～



4バイトASの状況 ～IPv4広告～

2007年7月10日

Network	Path
> 84.205.88.0/24	4697 2914 12654 3.7 i
> 145.125.0.0/20	4697 2914 3549 1103 1125 3.5 i
> 192.26.93.0	4697 2.3 i
> 193.31.7.0	4697 2914 3561 1273 5539 3.3 i
> 202.255.47.0	4697 2914 4713 2516 7667 2.6 i

2007年5月22日

連続した4バイトAS

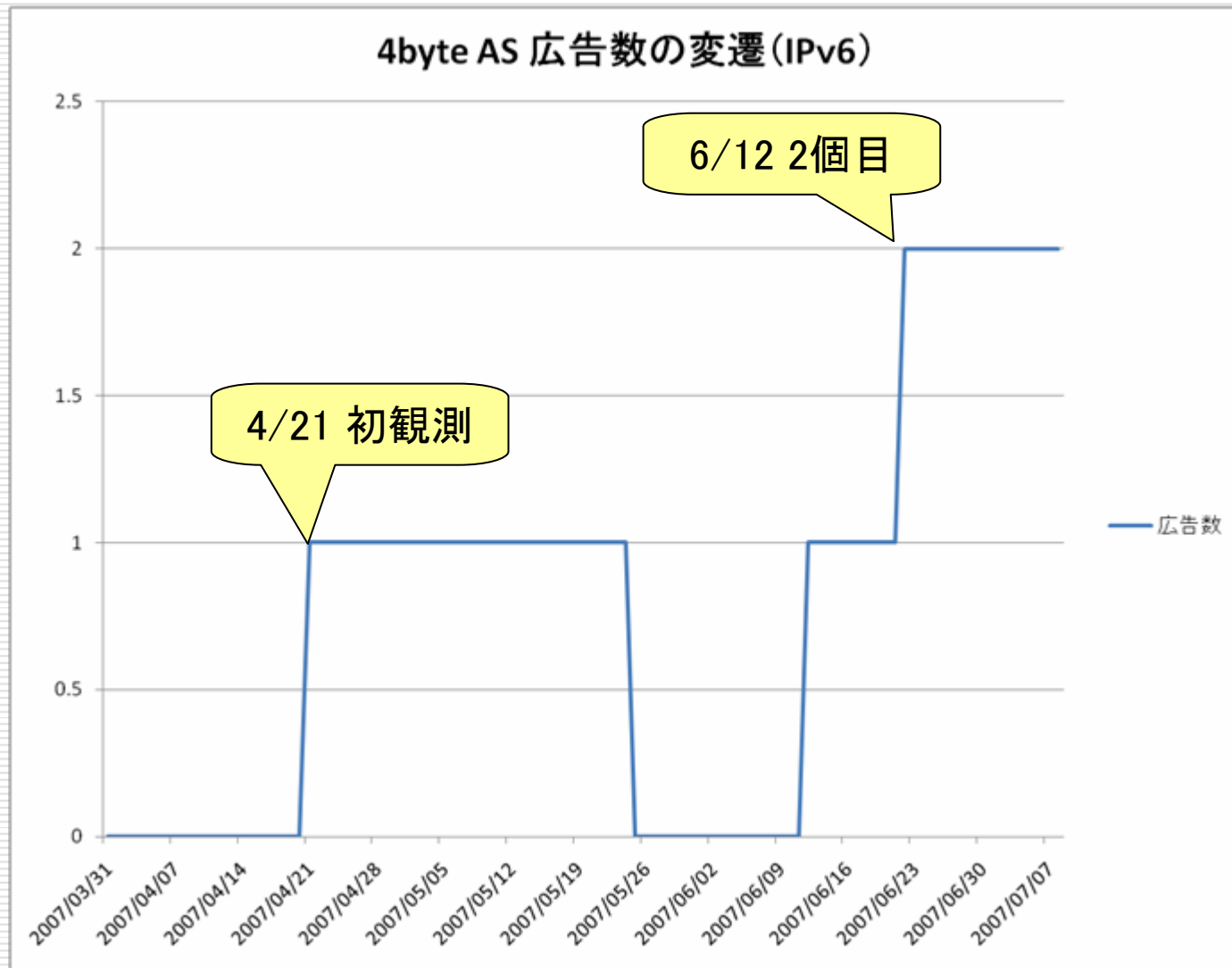
203.10.63.0	4710 4713 2914 3561 4637 4637 4637 4637
	4637 4637 4637 4637 4637 1221 2.2 2.0 i
> 203.10.63.0	4697 2914 3561 4637 4637 4637 4637 4637
	4637 4637 4637 4637 1221 2.2 2.0 i

2007年6月6日

連続した4バイトAS／2バイトASをトランジット

> 57.87.2.0/24	4697 2914 5511 2874 41712 i
57.87.2.0/24	4710 4713 2914 3561 4637 4637 4637 4637
	4637 4637 4637 4637 4637 1221 2.2 2.0
	12654 16186 2874 41712 i

4バイトASの状況 ～IPv6広告～



4バイトASの状況 ～IPv6広告～

2007年7月10日

```
> 2001:df0:2::/48      4697  2.3  i
> 2001:4810:2000::/35  4697 2914 33437 6.3  i
```

(2.3) 現在使える4バイトASルータ

- ❑ IOS XR3.4.0以降 (CISCO CRS-1, GSR)
- ❑ JUNOSe (Juniper E-series)
- ❑ OpenBGPD (3.9, 4.0) + 4-byte AS patch
- ❑ Quagga (0.99.6, 0.99.7) + 4-byte AS patch

各社ルータの4バイトAS対応状況

ベンダ	ルータ等	4バイトAS対応版	出荷予定時期
AlaxalA	-	-	リリース時期検討中
CISCO	CRS-1, GSR*	IOS-XR 3.4以降	対応済み
Force10	E1200, E600, E600i, E300	FTOS 7.6.1	2007年11月予定
Foundry	NetIron XMR		2008年Q2予定
	NetIron MLX		2008年Q2予定
	BigIron RX		2008年Q3予定
Juniper	E-Series	JUNOSe	対応済み
	T,M-Series	JUNOS9.1	2008年Q2予定

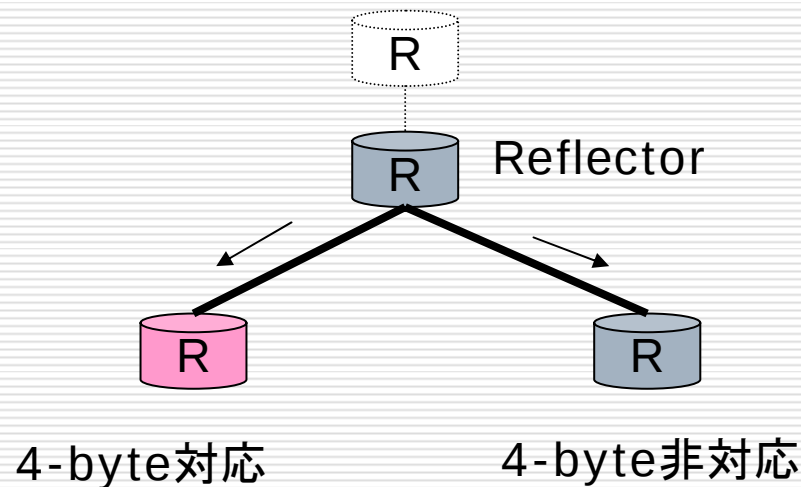
*GSR: Processor=PRP/PRP2, Interface Card=Engine3

☆2007年7月6日問い合わせ時点での情報

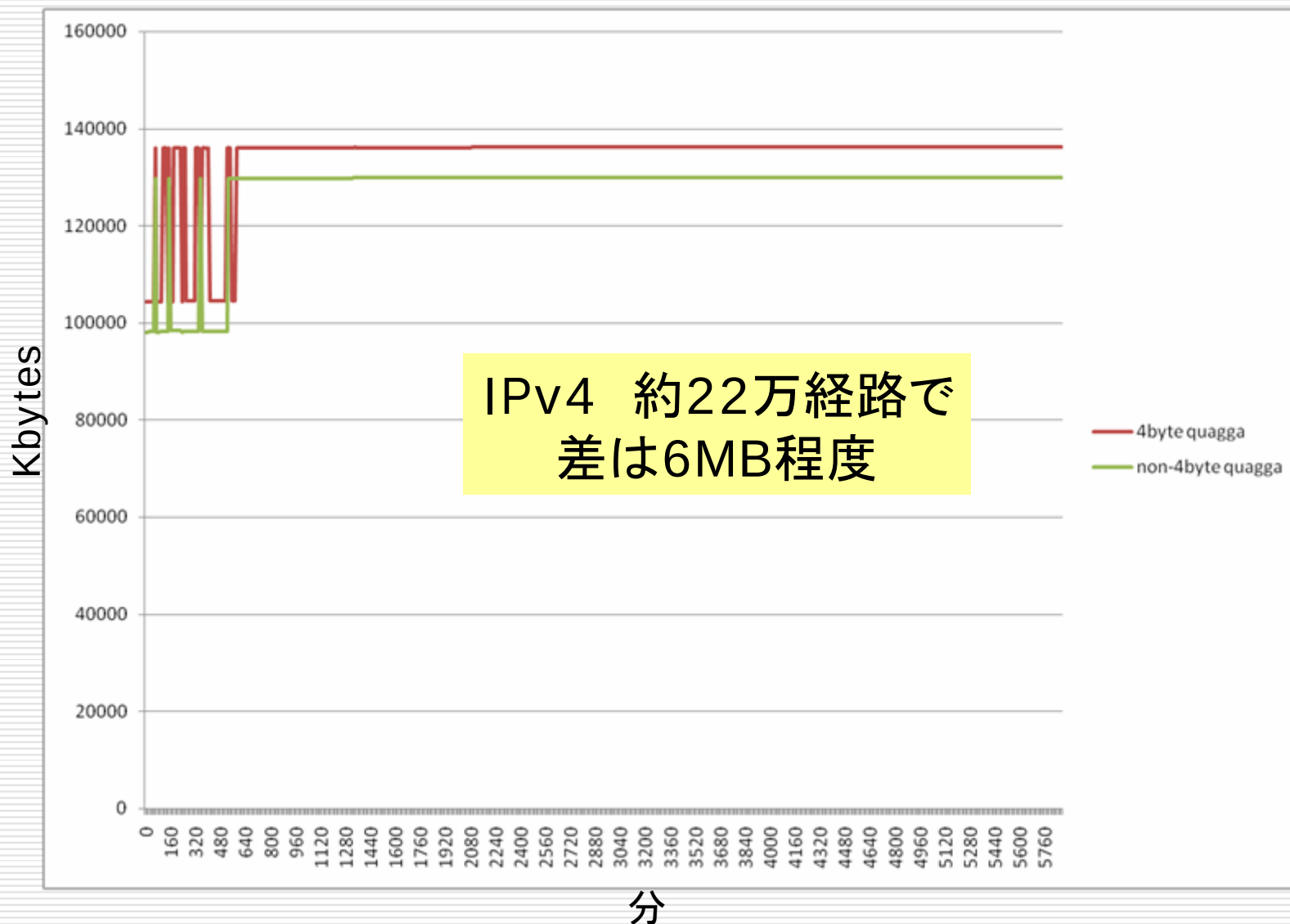
実装比較: quagga のメモリ使用量

□ 実験環境

- リフレクタに, 同スペックのPCを二台接続し, 4byte AS対応／非対応のquagga を同時にスタートさせた.
- bgpd プロセスのメモリ使用量を10分おきに観測



実装比較: quagga のメモリ使用量



アウトライン

(1) 4バイトASについて

(2) 4バイトASの現状

(3) プロバイダの対応

- (3.1) 2バイトASの世界に居続けるとどうなるか？

- 経路フィルタをかける点で問題はあるか？

- (3.2) プロバイダは2バイトASの世界から、4バイトASの世界に簡単に移行可能なのか？

- 具体的には、「ネットワークを止めずに2バイト対応から4バイト対応化できるのか？」

(4) 課題と議論

(3.1) 2バイトASの世界に居続けると

□ 4バイトASがすべてAS23456に見えてしまう

□ 問題点

■ MEDによる経路選択に影響が出る可能性

□ RFC4893にも指摘あり

□ 通常、異なる隣接ASからの経路のMED値は比較されない。
しかし、隣接するASに4バイトASが複数あると、同じASとして扱ってしまう。

■ AS PATHフィルタで区別できない

□ AS PATHフィルタを用いる際に、4バイトASを明示的に記述できない

■ 例： AS2.3とAS2.7が同じ上流プロバイダの場合、AS2.3のOriginの経路をAS2.7から流しても、2バイトの世界では区別できない

(3.1)

□いつかは2バイトASの世界から、4バイトASの世界へ移行する必要がある

■「4バイトAS移行」の意味

□例えばAS4697→AS2.3へ変えるのではなく

□AS4697内のルータを、4バイトAS対応のソフトウェアに更新し、AS"0.4697"化するという意味。

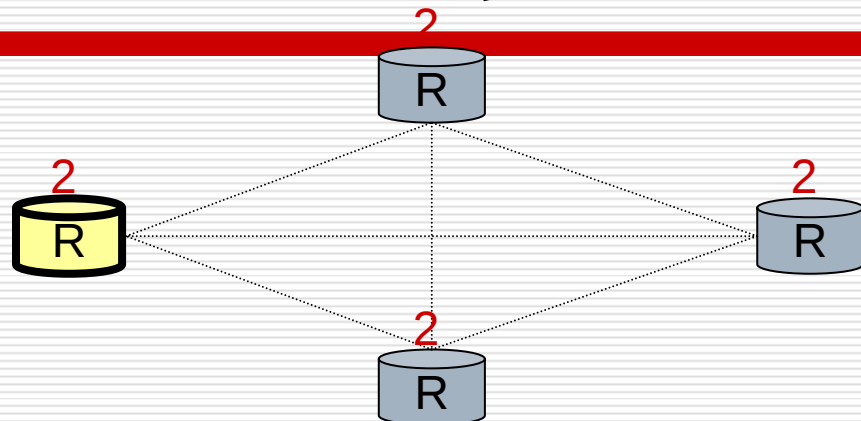
(3.2) プロバイダの4バイトAS移行は容易か？

- eBGPでの接続はうまくいくらしい。
しかしプロバイダに取っては内部の移行も重要。

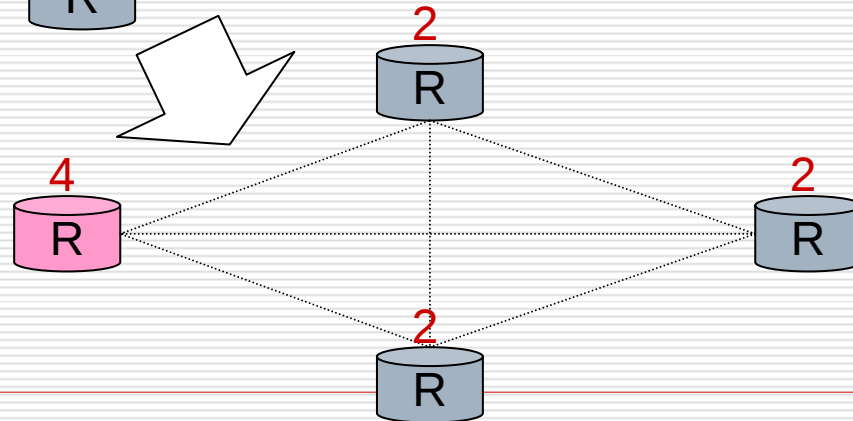
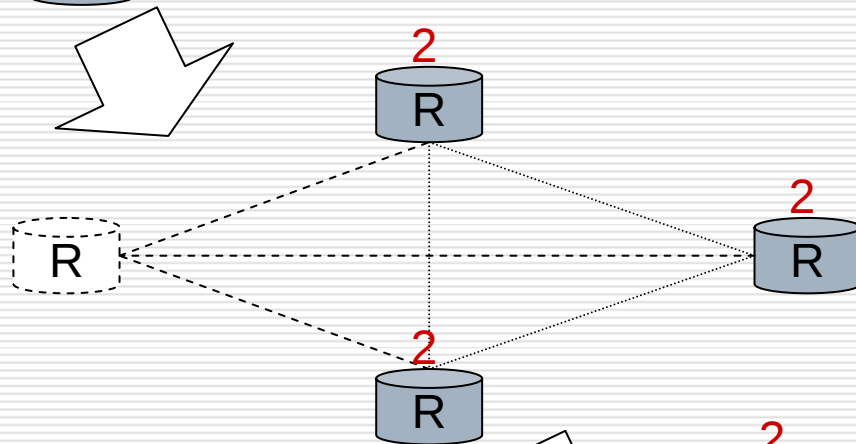
- プロバイダの内部のBGP構成には
 - フルメッシュ (full-mesh)
 - リフレクタ (reflector)
 - コンフェデレーション (confederation)などがある。

- うまく移行可能か実験してみた。
 - フルメッシュ、リフレクタ(クライアント側・リフレクタ側)

フルメッシュの場合

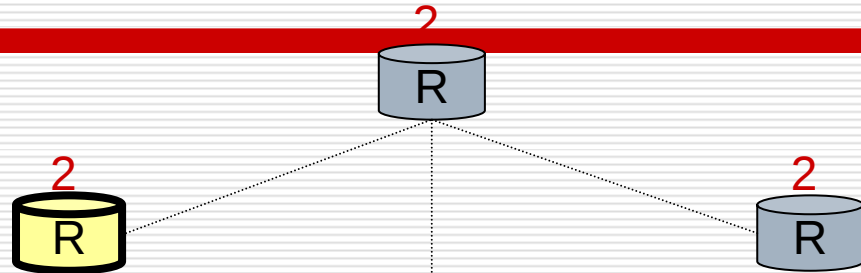


フルメッシュ参加の
BGPスピーカ1台を。。。。

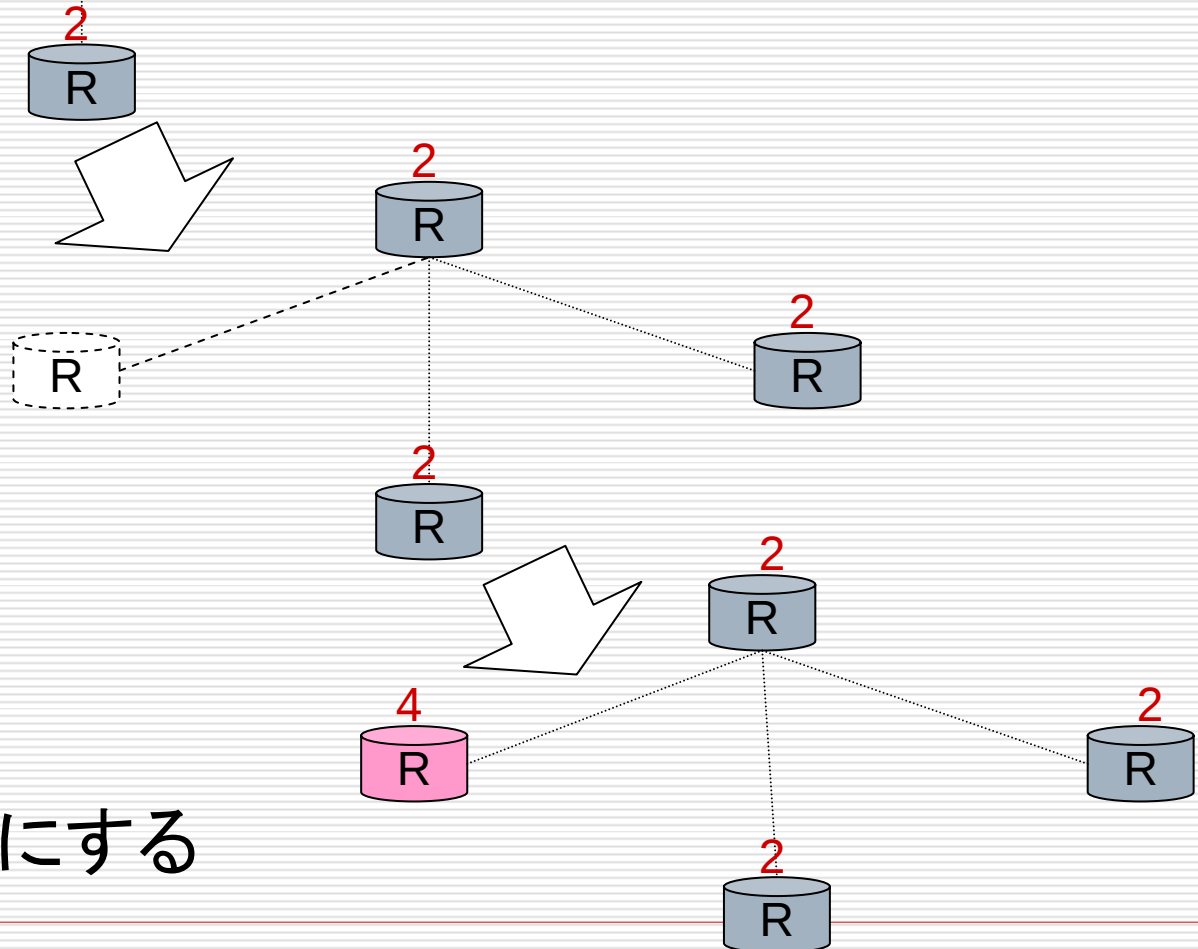


4バイトAS対応にする

リフレクタのクライアント側の場合



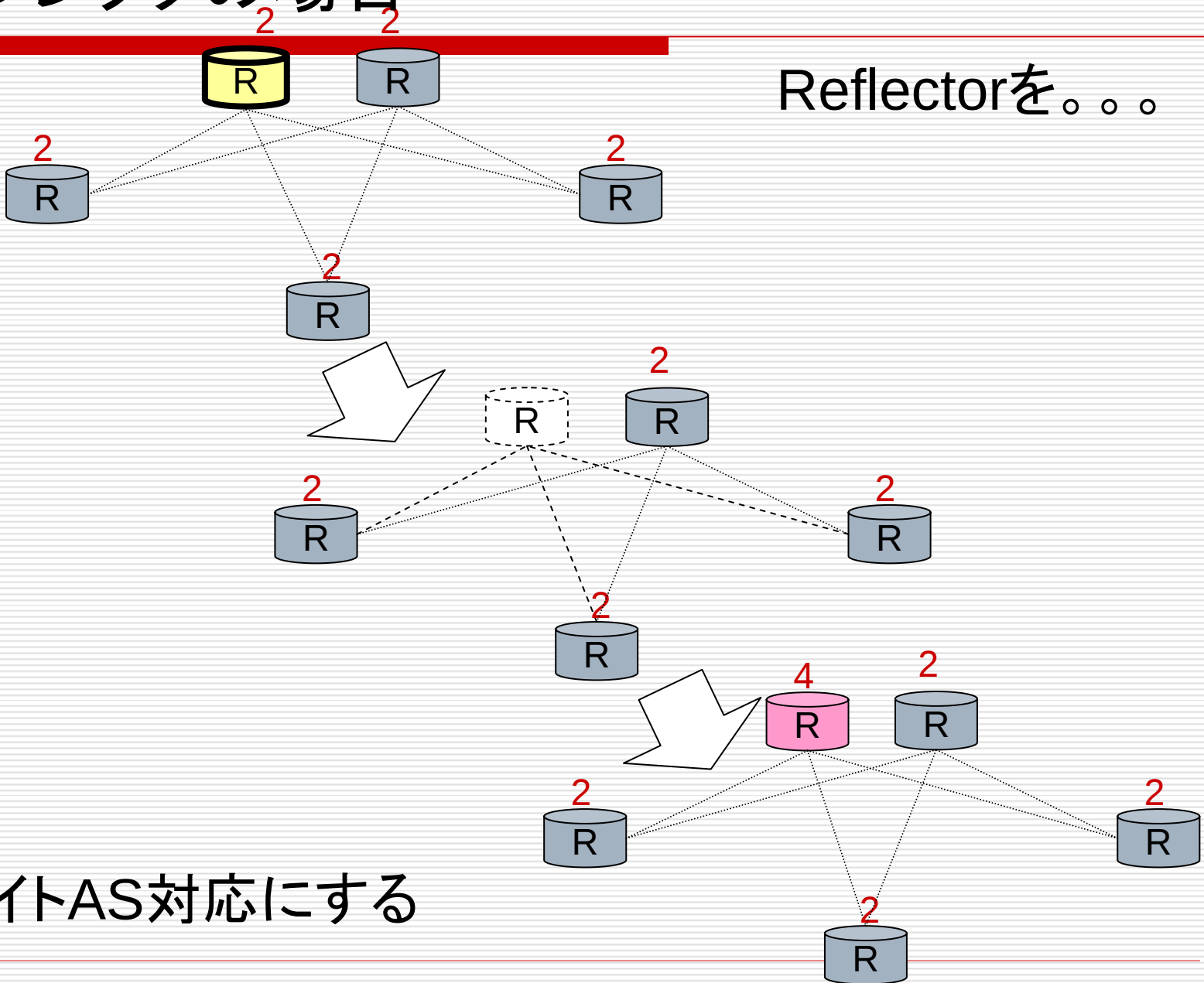
Reflector構成の
BGPスピーカ1台を。。。



4バイトAS対応にする

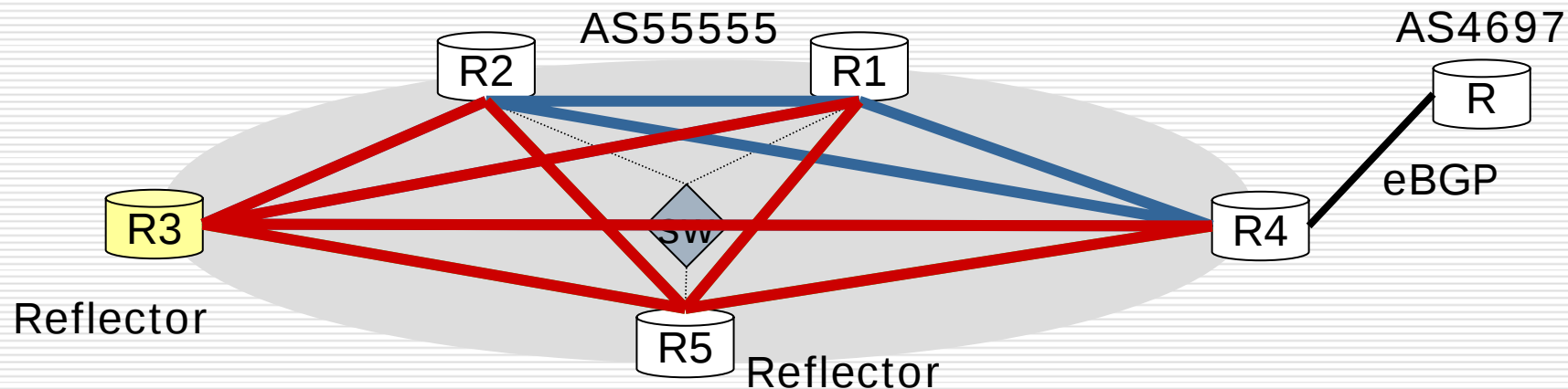
リフレクタの場合

Reflectorを。。。



4バイトAS対応にする

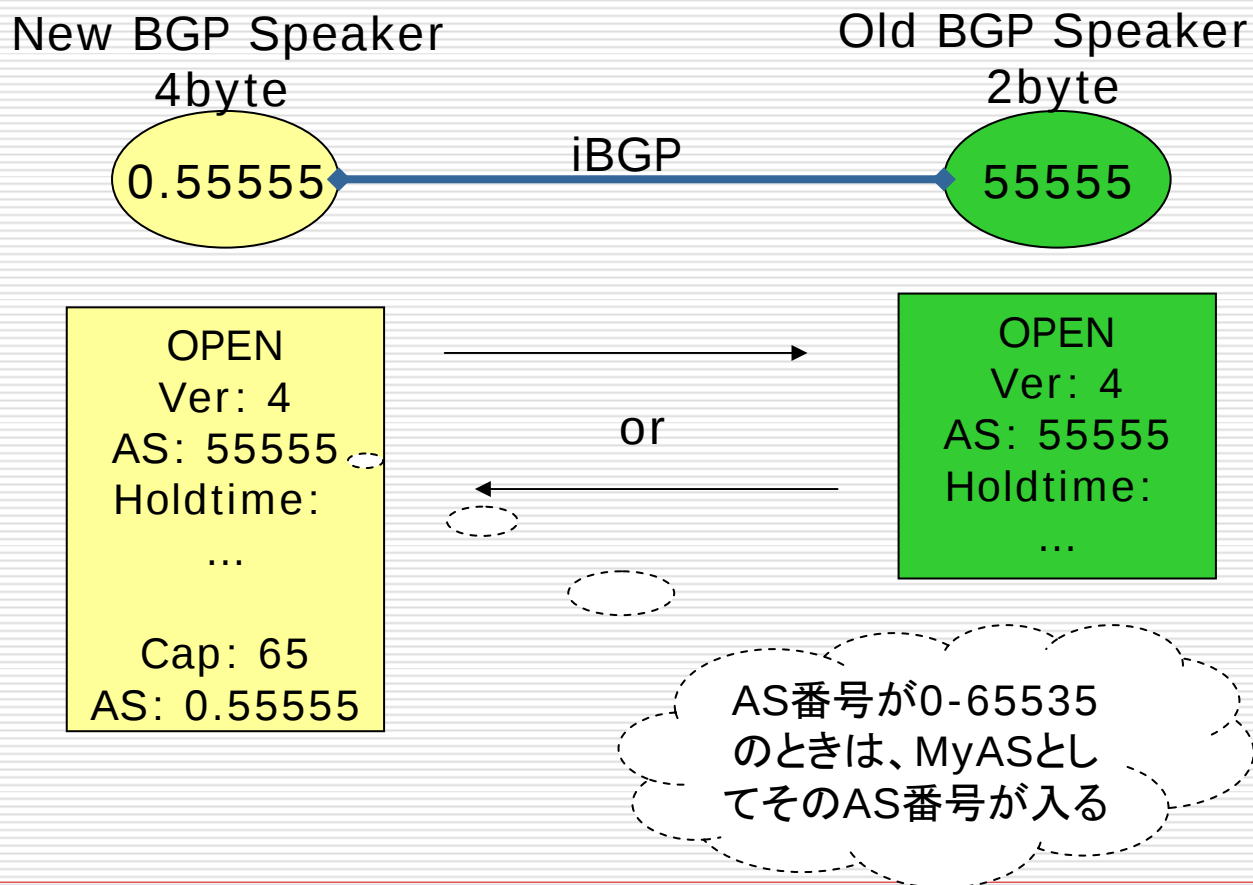
実験

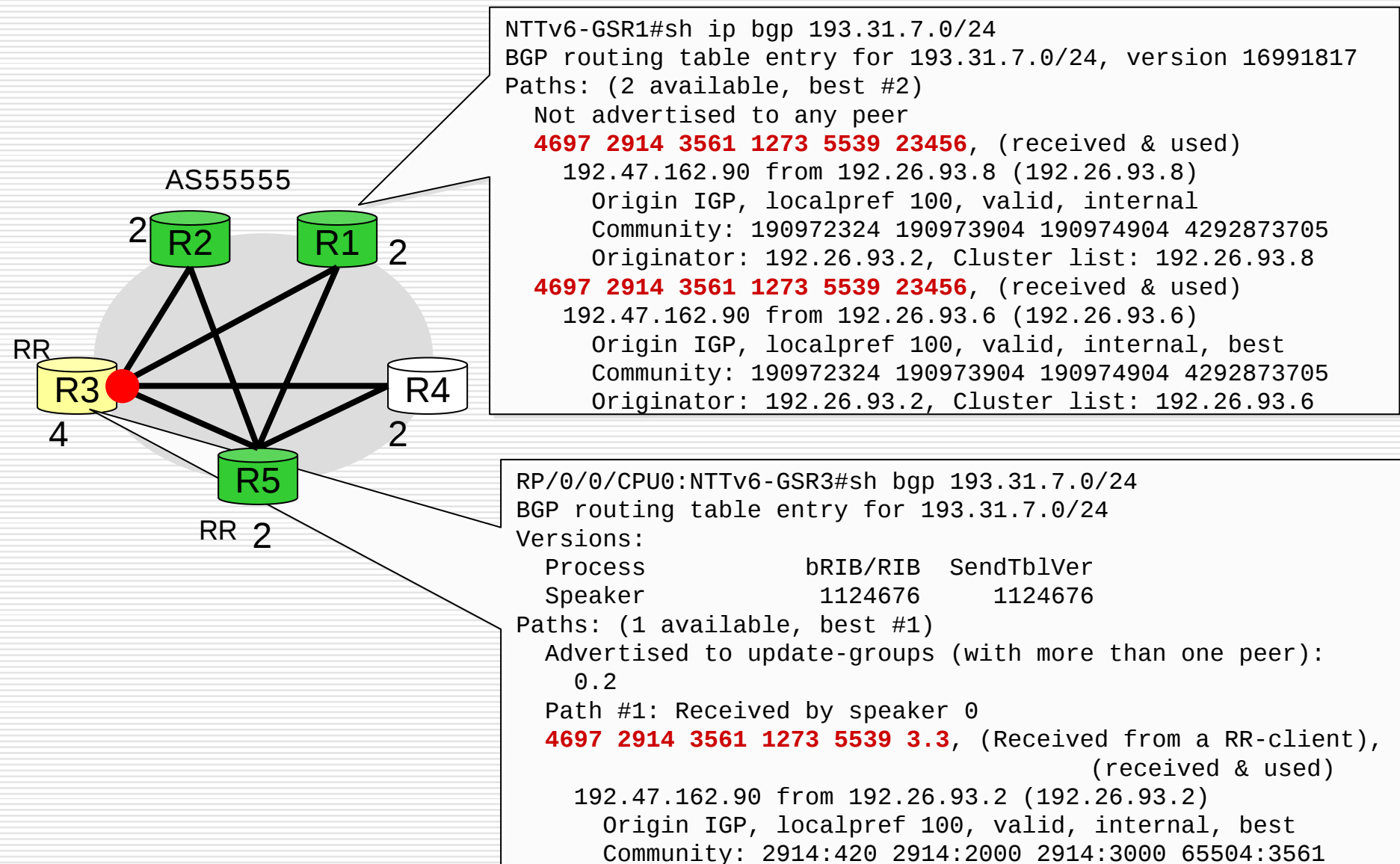


	ルータ	ソフト (2-byteのみ)	ソフト (4-byte対応)
R1	CISCO GSR12404 (1GB)	IOS 12.0(31)S	
R2	CISCO GSR12404 (1GB)	IOS 12.0(31)S	
R3	CISCO GSR12404 (2GB)	IOS 12.0(32)S6	IOS XR3.4.1
R4	PC + Quagga	0.99.7_2	0.99.7_2 w/ 4byte-AS patch (v.7)
R5	PC + Quagga	0.99.7_2	

結果： 良好

- どのケースも、実験の範囲内では大きな問題は見られなかった





実験結果より

□プロバイダ内部のルータを、**一つ一つ4バイト対応に移行することは可能**と考えられる。

- つまり、ネットワーク全体を止めなければならない事態は回避できそう。

□もちろん

- 各社のネットワーク機器(性能やメモリ)、トポロジー(特にiBGPやeBGPのピアリング)、顧客収容状況、フィルタなど様々な要因があるため、容易に移行できるかどうかは各社ごとに異なる。

プロバイダとしての移行順序

- ピアリング先(プロバイダや顧客)が4バイトASで接続してくる場合には、当面収容ルータを分けておく

- リフレクタ構成の場合
 - リフレクタのクライアントを4バイトAS対応OSにしてから、最後にリフレクタを4バイトAS対応とする順序が良さそう。

- フルメッシュの場合
 - はやめにすべて4バイト対応化しておくほうが良い？

アウトライン

(1) 4バイトASについて

(2) 4バイトASの現状

(3) プロバイダの対応

(4) 課題と議論

- (4.1) 実際には問題は生じていないのか？

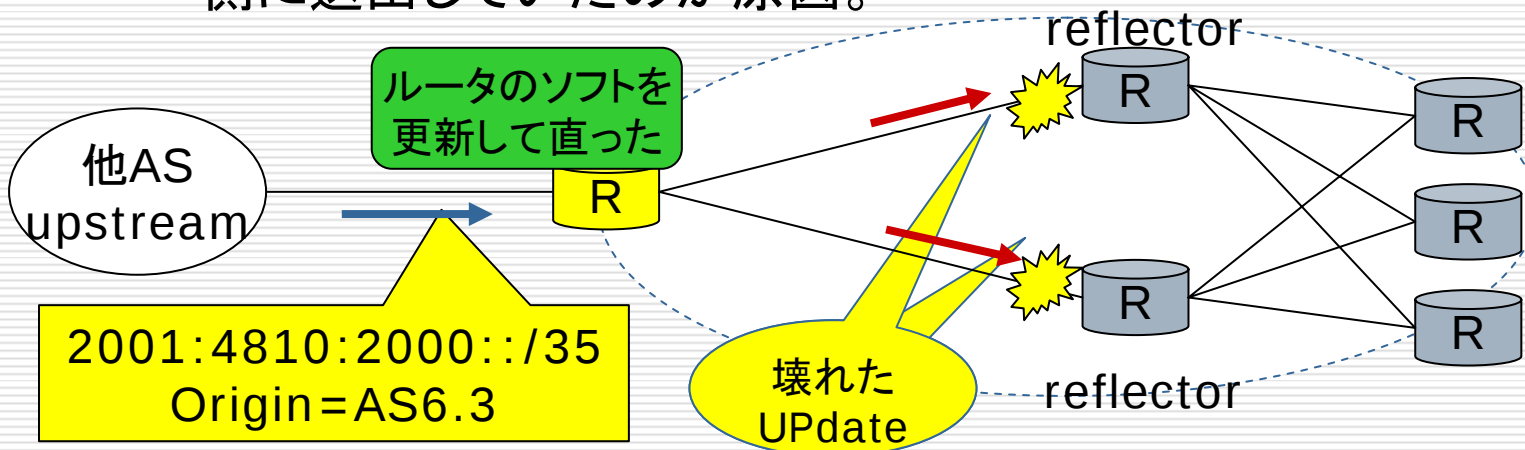
- (4.2) 4バイトAS化に伴って変更すべき点は？

- (4.3) みなさんの活動紹介

- ... 議論

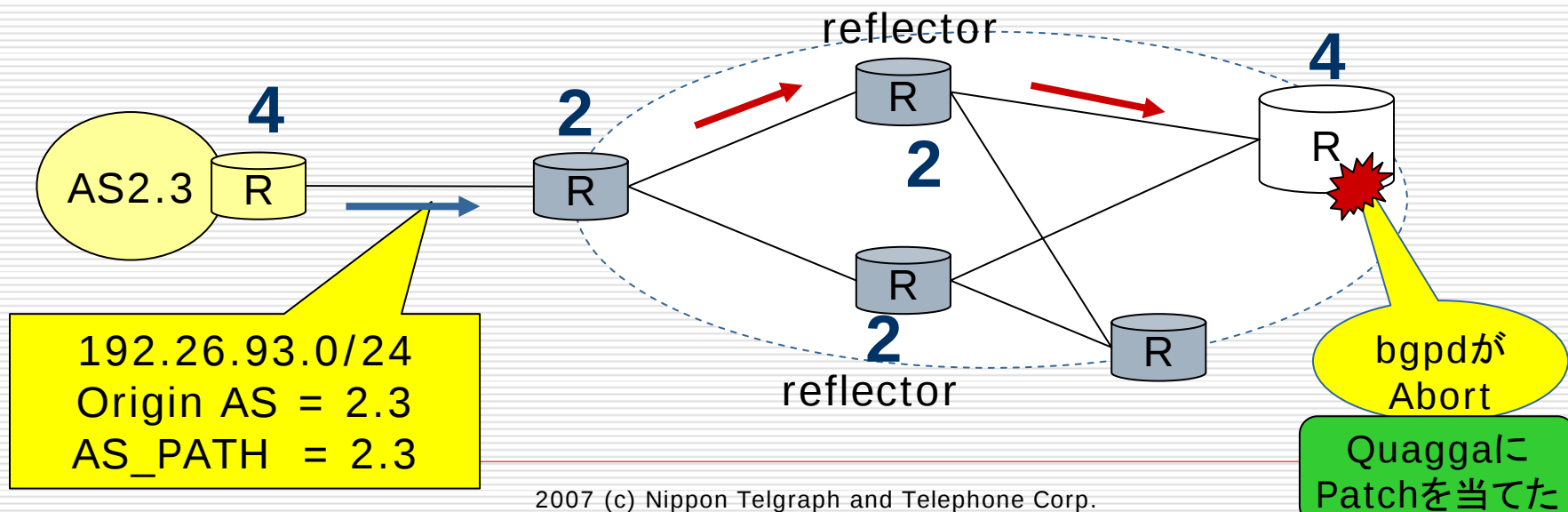
(事象1)

- ある日突然iBGPピアがダウン。
- ReflectorがUpdateメッセージを処理できずピアを切断していた。
- しばらくしてピアが再接続しても、経路数がある程度増えたところで再度ダウン。
- 処理できないUpdateメッセージを調べてみると、ある特定の経路のところで事象が発生していた。
 - 2001:4810:2000::/35 AS_PATH ... 6.3
- その経路をフィルタするとiBGPピアは上がる。
- あるルータが、その経路に対して壊れたUpdateメッセージをiBGP側に送出していたのが原因。



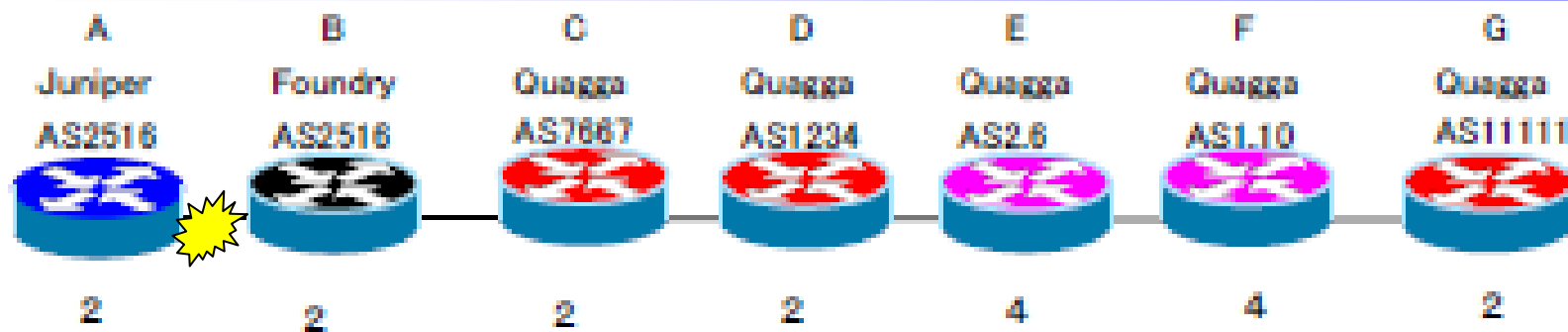
(事象2)

- 隣接ASに4バイトASを接続した。
- 4バイト対応のQuaggaをリフレクタクライアントとして接続していたが、bgpdがAbortして接続できなくなった。
- bgpdがAbortする原因を調べてみると、Quaggaが4バイトのAS_PATHを再構築するときに、隣接ASが4バイトASの場合、解放済みのメモリを参照するバグを発見。(by 松本存史)
- パッチを当てると、問題なく動くようになった。
 - パッチは、Quaggaを4バイトAS化した作者へ送付済み



(事象3)

- JUNOS8.3R1.5を下記ルータAにインストール。
 - ルータG(AS11111)から、ルータF(AS1.10)経由で経路を送信。
 - as-path: 2516 7667 1234 2.6 1.10 11111
 - このas-pathを受けると、A-B間のピアがダウンする。
 - 右端のルータG(AS11111)を止めると、ピアは回復する。
-
- 4バイトASがトランジットした as-path で、Juniperのピアがダウンする。



(4.2) 4バイト化に伴い影響を受ける点

- Community 属性
- xFlow (NetFlow, sFlow)
- MRTフォーマット
- MIB
- IPv4マルチキャストアドレス (GLOP)
- IRR
- RPSL

Community属性

- Community 32ビットのうち上位16ビットはAS番号を入れて使う
- “Four-octet AS Specific BGP Extended Community,” draft-rekhter-as4octet-ext-community-02.txt で議論中

xFlow関係

□ NetFlow

■ Version 9 (RFC3954)

□ AS番号は4バイト長で表現可能

■ デフォルトは2バイト

■ Version 5

□ 2バイトのみ。

■ 4バイト対応のルータでNetflow v.5をexportするとき、4バイトASのAS番号はやはりAS_TRANS(23456)でexportするのか？

□ sFlow

NetFlow v.9 のパケット

No. .	Time	Source	Destination	Protocol	Info
7	66.136438	192.47.163.20	192.47.163.53	CFLOW	total: 8 (v9)
8	70.143438	192.47.163.20	192.47.163.53	CFLOW	total: 1 (v9)
9	91.928441	192.47.163.20	192.47.163.53	CFLOW	total: 1 (v9)
10	102.207645	192.47.163.20	192.47.163.53	CFLOW	total: 1 (v9)

+ User Datagram Protocol, Src Port: 26623 (26623), Dst Port: 9999 (9999)	
- Cisco NetFlow/IPFIX	
Version: 9	
Count: 1	
Sysuptime: 12967269	
- Timestamp: Jul 2, 2007 08:15:25.000000000	
CurrentSecs: 1183331725	
FlowSequence: 489	
SourceId: 16	
- FlowSet 1	
Data FlowSet (Template Id): 260	
FlowSet Length: 60	
- Flow 1	
Packets: 4	
Octets: 240	
SrcAddr: 192.47.163.53 (192.47.163.53)	
DstAddr: 84.205.88.1 (84.205.88.1)	
InputInt: 1	
OutputInt: 2	
EndTime: 12950.785000000 seconds	
StartTime: 12934.692000000 seconds	
SrcPort: 0	
DstPort: 0	
SrcAS: 0	
DstAS: 0	
BGPNextHop: 0.0.0.0 (0.0.0.0)	
SrcMask: 32	
DstMask: 24	
Protocol: 1	
TCP Flags: 0x00	
IP ToS: 0x00	
Type 61	
Type 89	
Padding (1 byte)	

0000	00 0b 97 b9 e6 d6 00 19 55 e3 dc fc 08 00 45 00	.	.	.	.	.	.	U	.	.	.	.	E
0010	00 6c 01 e9 00 00 ff 11 f2 ee c0 2f a3 14 c0 2f	.	.	.	.	.	.	.	.	.	.	.	/
0020	a3 35 67 ff 27 0f 00 58 00 00 00 09 00 01 00 c5	.	5g	.	.	.	.	.	.	.	.	.	x
0030	dd 65 46 88 35 8d 00 00 01 e9 00 00 00 10 01 0a	.	eF	.	5	.	.	.	.	.	.	.	.
0040	00 3c 00 00 00 04 00 00 00 f0 c0 2f a3 35 54 2d	.	.	.	.	.	.	.	.	.	.	.	/
0050	58 01 00 00 00 01 00 00 00 02 00 c5 9d 01 00 c5	.	.	.	.	.	.	.	.	.	.	.	5T
0060	5e 24 00 00 00 00 00 00 00 00 00 00 00 00 c5	.	.	.	.	.	.	.	.	.	.	.	.
0070	00 00 20 18 01 00 00 00 40 00	.	.	.	.	.	.	.	.	.	.	.	.

Destination AS (cfldstas), 4 bytes

P: 26 D: 26 M: 0

4バイト幅に
なっている

MRTフォーマット

- MRT routing information export format (draft-ietf-grow-mrt) にて規定
- -02 (March 6, 2006) より, 4バイトAS対応
 - BGP4MP_MESSAGE_32BIT_AS (-03まで)
 - BGP4MP_MESSAGE_AS4(-04より)
 - Type名が変わった他, BGP4+へのフル対応などの変更が入っている.
- Quagga の4バイト対応パッチがMRTフォーマットの情報をはける(ように見える)

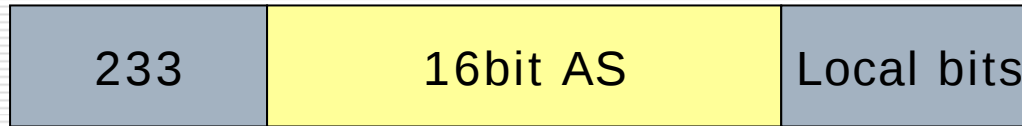
MIB関連

- 現状, 最新のMIBは, RFC4273: Definitions of Managed Objects for BGP-4)
 - AS関係は, Integer32 として規定されているが,
bgpLocalAs OBJECT-TYPE
SYNTAX Integer32 (0..65535)
なっており, 改訂が必要.
 - draft-ietf-idr-bgp4-mibv2 に取り込まれる方向で議論が進行中
 - が, 現在, このドラフトはExpire状態

IPv4マルチキャストアドレス

□ GLOPアドレス(RFC3180)の対応

■ GLOP アドレス



□ 一応, RFCに, 16bit のASでしか使えないよ, とは書いてはある.

■ eGLOPアドレス(RFC3138)

□ GLOPアドレスの, プライベートAS番号部分(64512-65534)部分を配布する

■ 配り方を規定(RFC2365, RFC2770(RFC3180), SSMじゃだめなことを説明する必要あり)

■ eGLOPアドレスの配布について, RIRで提案中

□ APNICでは, 次回のミーティングに提案があがっている

(4.3) 他のみなさんの活動

□ 次世代IX研究会 (Distix)

- ルータ相互接続WG（主査：高知工科大学：菊池先生）

□ <http://www.distix.net/router-wg/exp12/index.html>

□ George Michaelson氏@APNIC

- “dix-ie” に4バイトASスピーカを設置。

- **ピアリング募集中。**

□ 連絡先：東大 加藤朗先生

まとめ

□プロバイダとしての今後の対応

■ すぐに起きうる問題点への対処

□ 当面、eBGPおよびiBGPに注意を払う

■ 秘孔を突く経路が突然来るかもしれない

□ ルータOSのバグ情報に注意する

■ 4バイト対応OSへの移行

□ 基本動作としては問題なし。ネットワークを止めずに移行できる。

□ 自社で使っているルータのOSがいつ4バイトAS対応となるかをベンダに確認しておく

■ 場合によっては設備更改にあわせて4バイト化する。

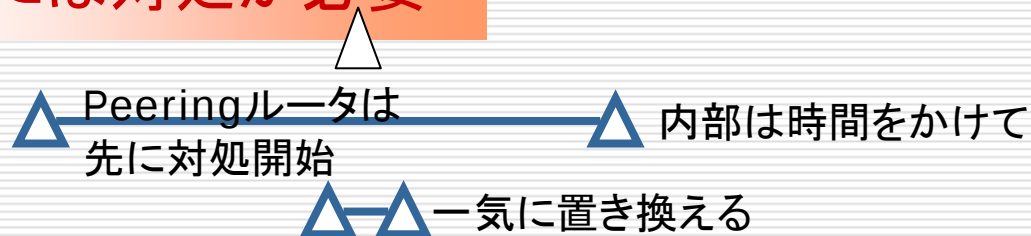
■ 4バイト対応には、メモリを増やす・CPU性能を上げる必要が出てくるかもしれない。

□ 事前に類似環境で検証し問題がないか確認しておくが良い。

対応のスケジュール感



2009年ぐらいまでには対応が必要



- 少なくとも、ピアリングするルータは2009年には4バイトAS対応しないとまずそう

ベンダさんへのお願い

- いつ頃対応できるのかの情報を早めに教えて下さいね。
- 予定通りに「安定した」バージョンを提供して下さいね。
- ルータを買い換えなくても済むように配慮して下さいね。

ご静聴ありがとうございました

□ 質問、コメント、議論など