ルーティング最前線と今の僕達

～力の入れどころと抜きどころ～

Matsuzaki ‘maz’ Yoshinobu
maz@iij.ad.jp
Yoshida ‘tomo’ Tomoya
<yoshida@nttv6.jp>

状況は変わり続ける

- 変化は次々にやってくる
  - プロトコル、機種、ツール、体制
  - 今風の設定、やり方
  - 社会からの要請
- これらの変化に対応しなきゃいけない

頑張って追従

- ハードルもある
  - 環境や状況の変化を知らない上司に一から説明
  - 誰が変更の判断できるか分からない
- でもって、資源は限られてる
  - 特に人
  - 既に既存作業で手いっぱい
  - トラブルでも起こると、他に何もできない

ぢゃあ、手抜きだ

- 同じことでも楽にやろうよ
  - あの時はこれが一番良いと思ったんだよ
  - でも、今なら違うやり方ができるかも
- 楽になった分、他のことに手を出せる
  - 新しいこととか、楽しいこととか ☺

ただし書き

- 各組織、企業によって状況が異なるかもしれません。営業戦略な理由とかよく分からない理由で最低限のレベルとか優先度は変化するかもしれません。何にせよ、best effortと胸を張れるぐらいには頑張りましょう。

何をやりたいかというと

- 利用可能なネットワークの維持
  - サービスの継続的な提供に繋がる
- 顧客の通信に影響が出ないように
  - 障害をいち早く取り除く
  - 障害が未然に防げるなら、なお良し
- コミュニティの要請に応える様な最善の実装
  - コンプライアンス
  - できれば楽ちんに

力の入れどころ

- 根幹にかかわる所は、ちゃんとやっとく
  - 機器の基本設定
    - 認証、アクセスコントロール
    - ログ取り、etc.
  - ルーティングポリシの策定
    - 特になければ今風のポリシでOK
    - 広報する経路は厳密に
    - 受信する経路は比較的大らかに

運用は手抜きのしどころ

- 標準化、定形化しちゃう
- やっぱり自動化
  - ちょっとしたことでも効いてくる
    - 作業ログ取り
  - scriptやツール群による支援
    - 監視とか運用管理
    - データベースから設定の生成とか
    - 様々なリソース監視とか

自動化の注意点

- 自動での状態遷移
  - 遷移しすぎ
  - 障害時に状態が追えなくなる
- 自動通知されるメッセージ
  - 情報が多すぎ
  - 必要な情報を見逃す

異常事態は起こる

- 平常時のみを想定した自動化は、時に邪魔
  - 障害時に混乱を増幅させる
  - 仕事を増やす ☹
- ちゃんと色々な状況を考えておく
  - 過負荷、遅延、分断、孤立
  - 攻撃、いたずら、設定ミス

173.0.0.0/16 and 174.128.0.0/16

- 2009/01/13にNANOGで騒ぎになった件
  - 研究目的で、ARINから割り当てられたブロック
  - それぞれのブロックを/24に分割して広報
  - AS Path属性を書き換えて、広報元ASを偽装

こんな広報

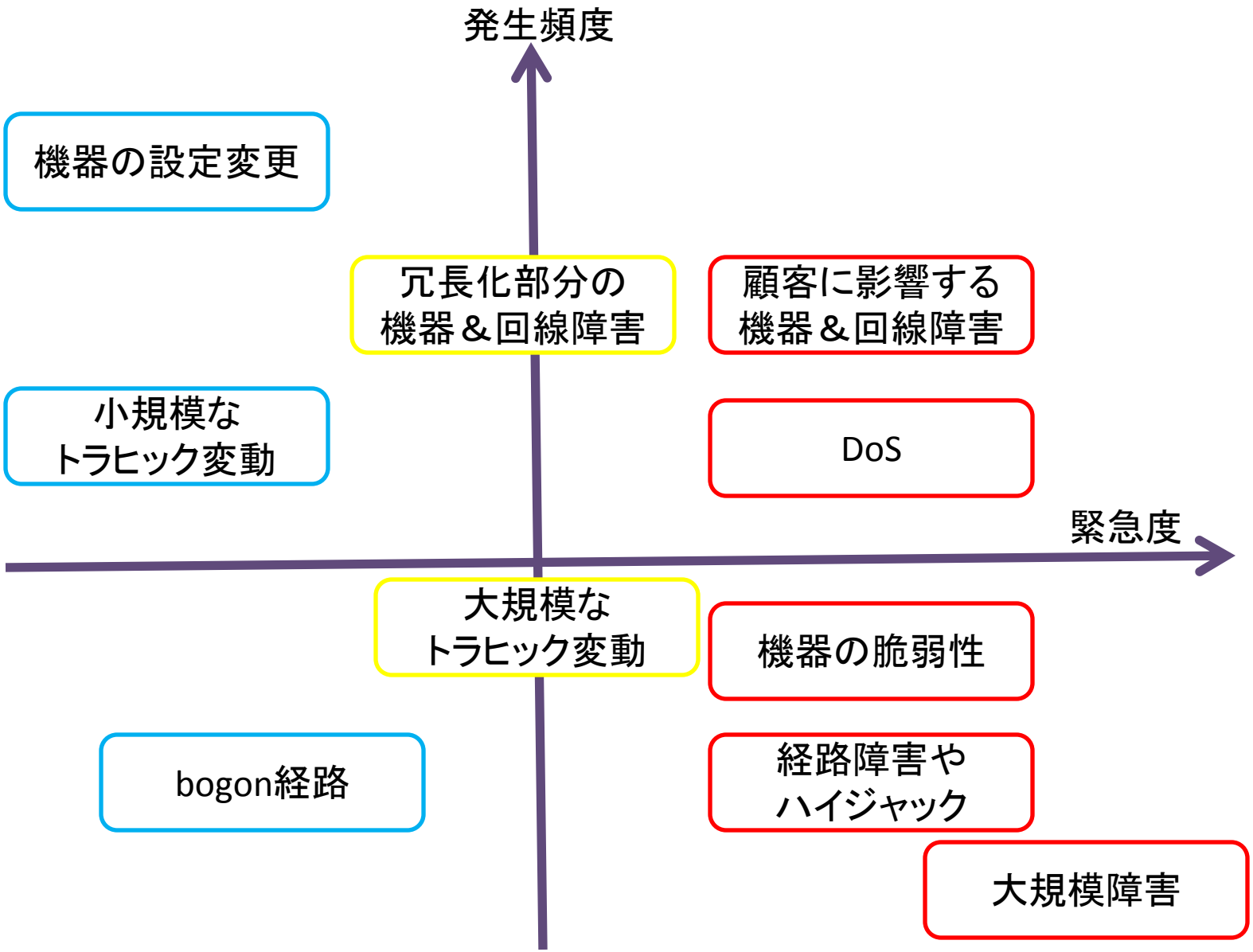
```
--prefix-------------+--as path----------
174.128.14.0/24      3130 2
173.0.14.0/24         3130 3
174.128.13.0/24      3130 4
173.0.13.0/24         3130 6
                      :
173.0.23.0/24         3130 43057
174.128.22.0/24      3130 43099
173.0.18.0/24         3130 43103
173.0.22.0/24         3130 43380
```

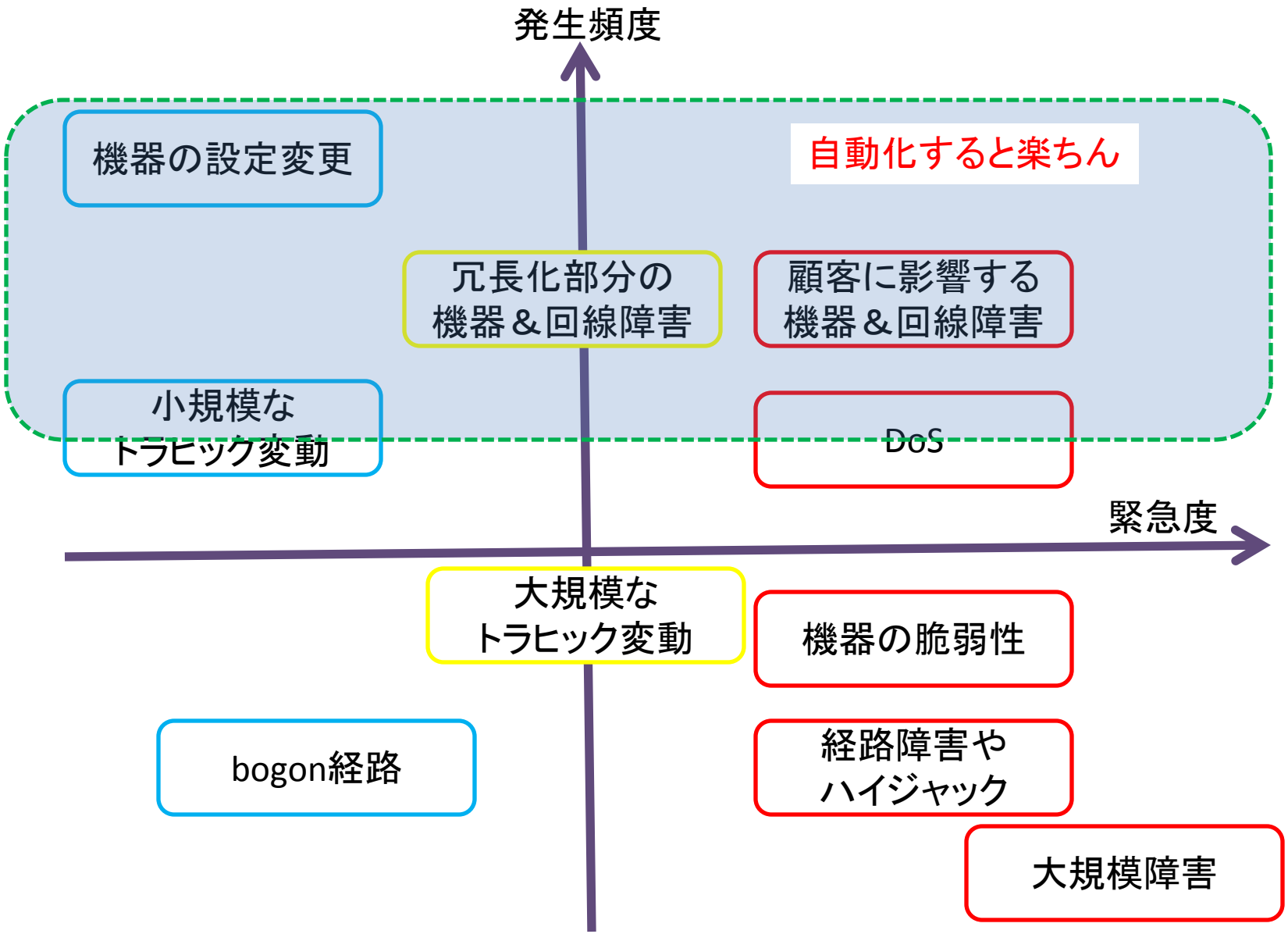
反応

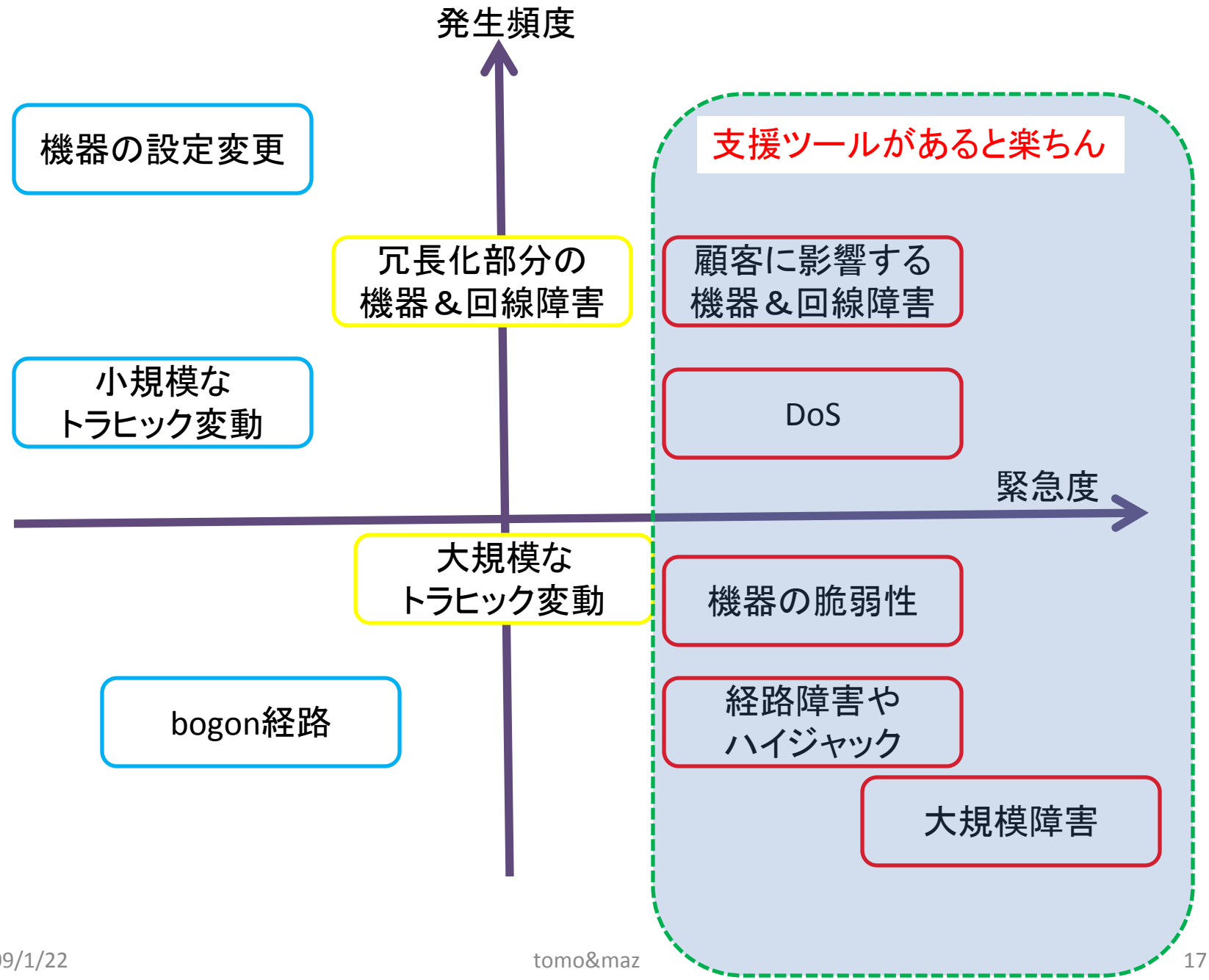
- 偽装されたASの人たちがびっくり
  - 自分のASが何か勝手に広報してるように見えた
  - 人によっては緊急対応
    - 経路広報のアラートが届いた
- 一方で、そんなに騒がなかった人もいた
  - そもそも気が付いていない
  - 顧客通信に影響がないので優先度が低かった

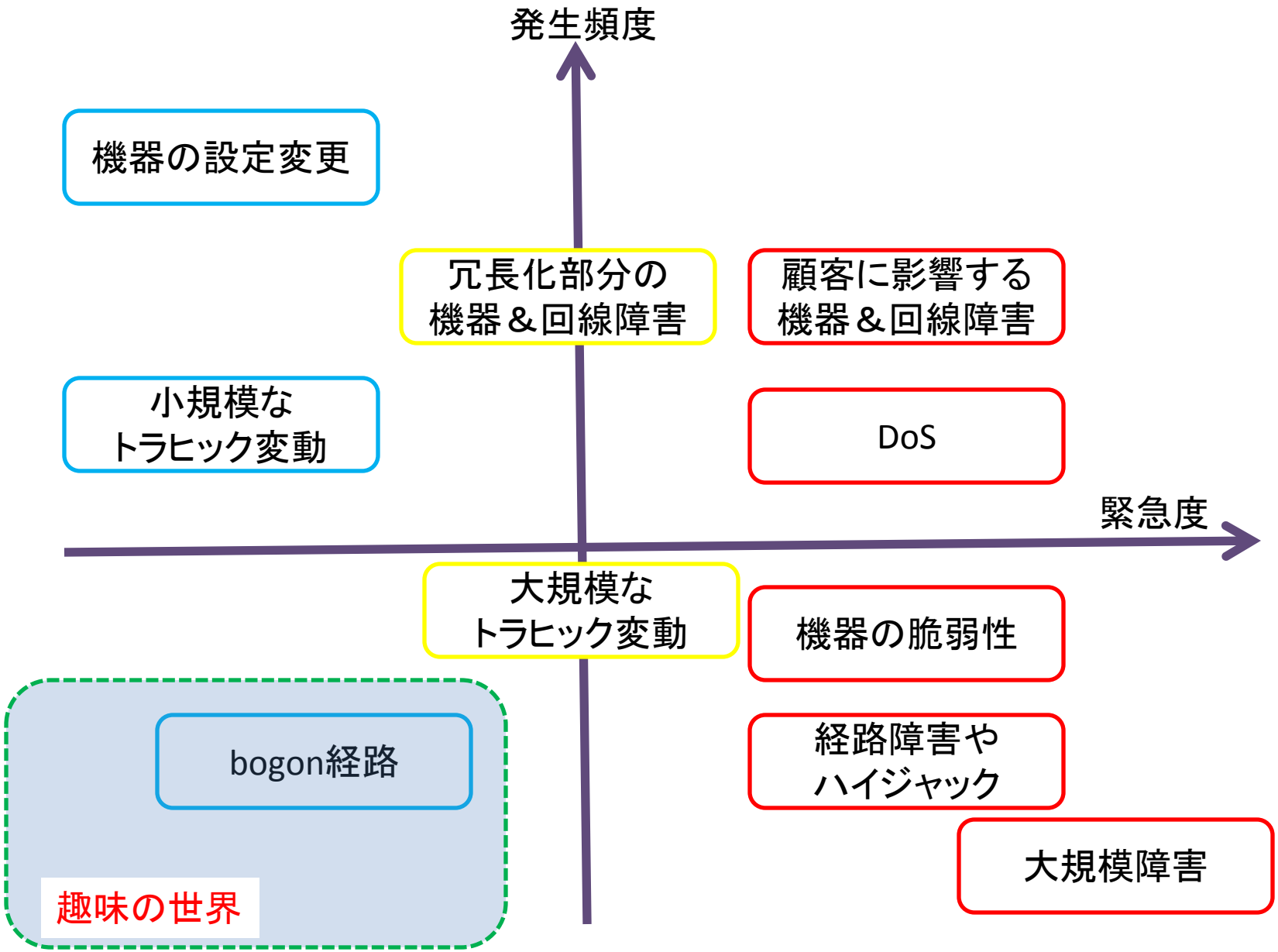
イベントと緊急度と

- 何かが起こってる事を知るのは大事
  - でも緊急度や重要度は異なる
  - 顧客の通信に影響あるアラートは重要
- 手の抜きどころ
  - 重要じゃないアラートの扱いを変える
    - 傾向を見るためにサマリとか
  - 発生頻度が高ければ、自動化が効いてくる


発生頻度
機器の設定変更
冗長化部分の
機器＆回線障害
顧客に影響する
機器＆回線障害
小規模な
トラヒック変動
DoS
緊急度
大規模な
トラヒック変動
機器の脆弱性
bogon経路
経路障害や
ハイジャック
大規模障害


発生頻度
機器の設定変更
自動化すると楽ちん
冗長化部分の
機器＆回線障害
顧客に影響する
機器＆回線障害
小規模な
トラヒック変動
DoS
緊急度
大規模な
トラヒック変動
機器の脆弱性
bogon経路
経路障害や
ハイジャック
大規模障害


発生頻度
機器の設定変更
支援ツールがあると楽ちん
冗長化部分の
機器＆回線障害
顧客に影響する
機器＆回線障害
小規模な
トラヒック変動
DoS
緊急度
大規模な
トラヒック変動
機器の脆弱性
bogon経路
経路障害や
ハイジャック
大規模障害


発生頻度
機器の設定変更
冗長化部分の
機器＆回線障害
顧客に影響する
機器＆回線障害
小規模な
トラヒック変動
DoS
緊急度
大規模な
トラヒック変動
機器の脆弱性
bogon経路
経路障害や
ハイジャック
大規模障害
趣味の世界

運用のデフラグ

- みんな、思い出した頃にデフラグやるよね？

- 運用も時々は見直ししなきゃ
  - ツール群を揃えると運用が固定化するので注意
    - 誰かが書いたぐっとくるツールがあれば移行を検討
  - 手順の見直しとか
    - 効率化できないか考えてみる

最近の実装と課題

1. トラヒック制御

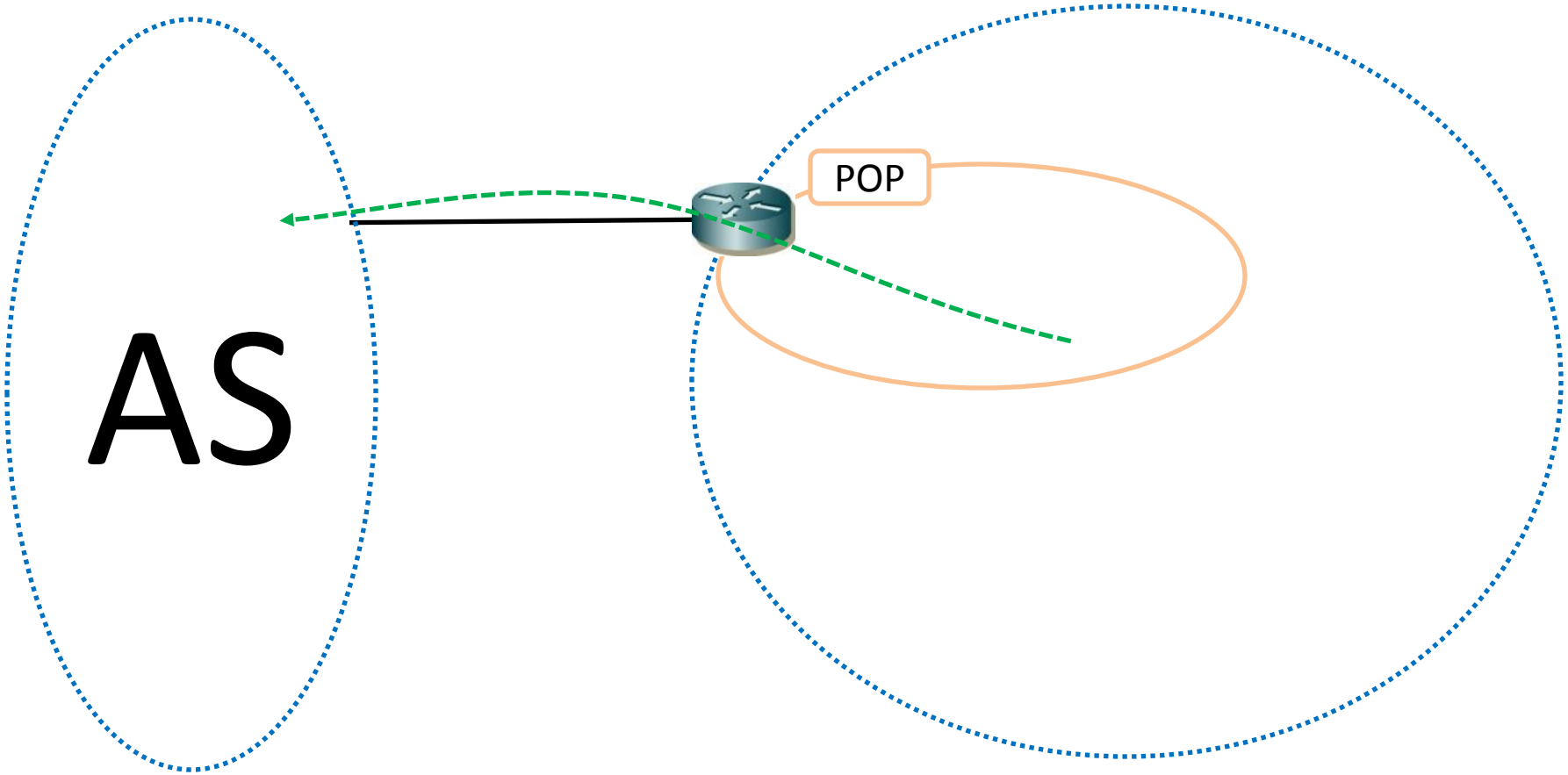
2. 4octet AS
   - DOTとPLAINの混在環境への対応

3. 不正な経路情報への対策
   - IRR
   - RPKI

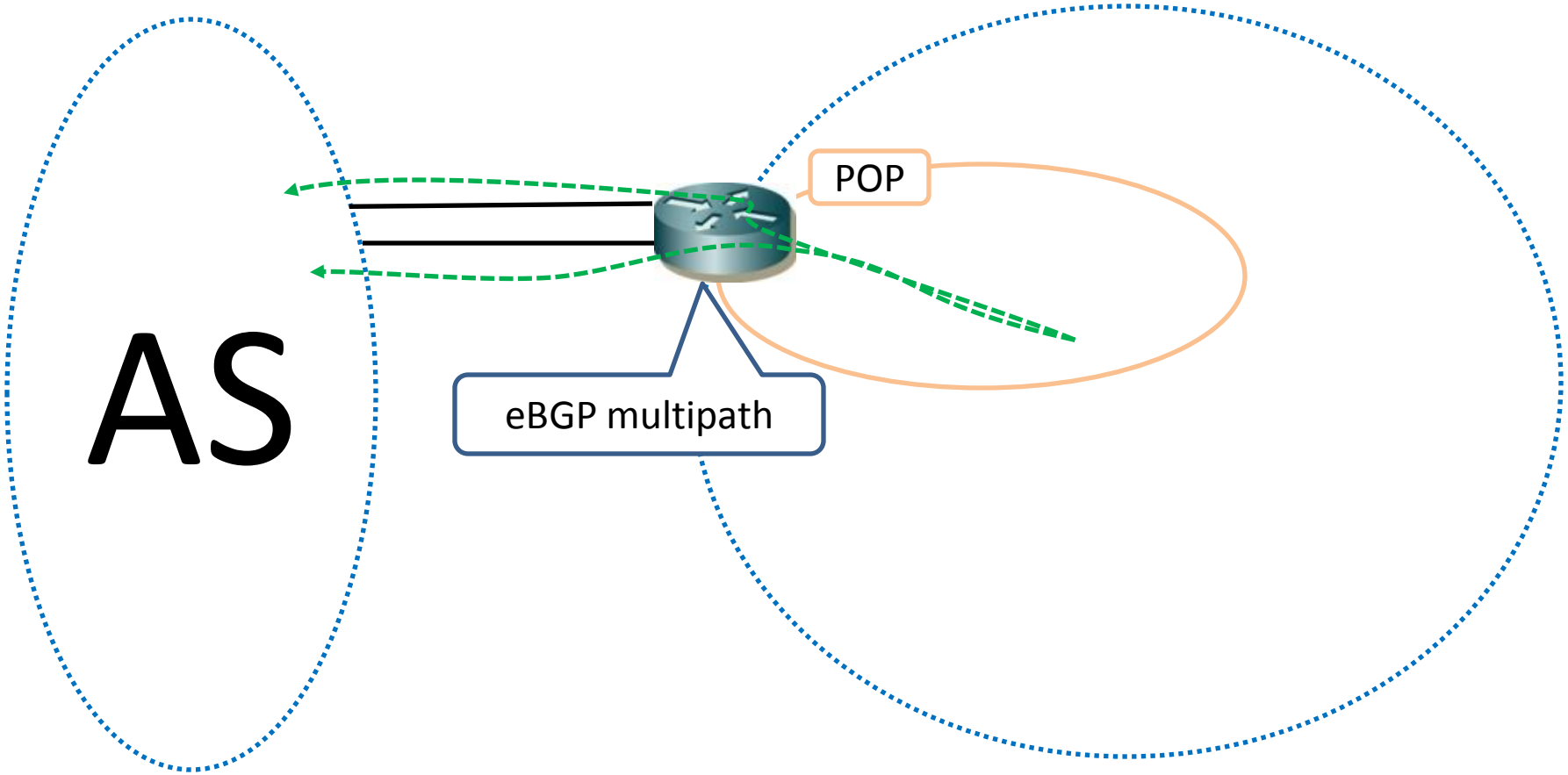
1. トラヒック制御は悩みどころ

- 入りのトラヒックを制御するのは至難の業
  - 調整すれば、一時的には何とかなる
  - でも他のASの制御や事情で激変しちゃう
- 出のトラヒックを制御するのは比較的簡単
  - しかし比較的でしかない
- なんにせよ、帯域が限られていると辛い
  - 100G早くほしい

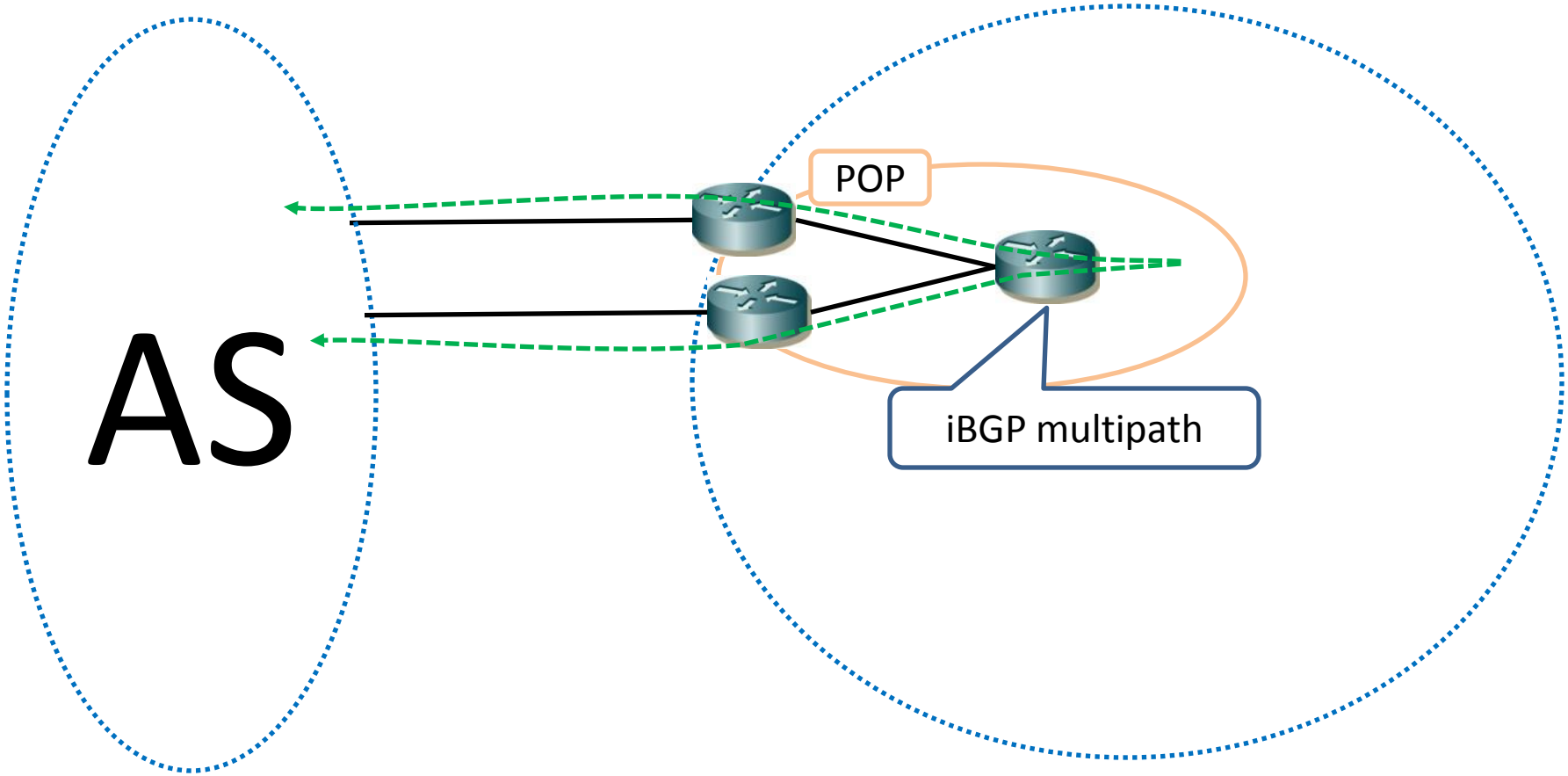
相互接続



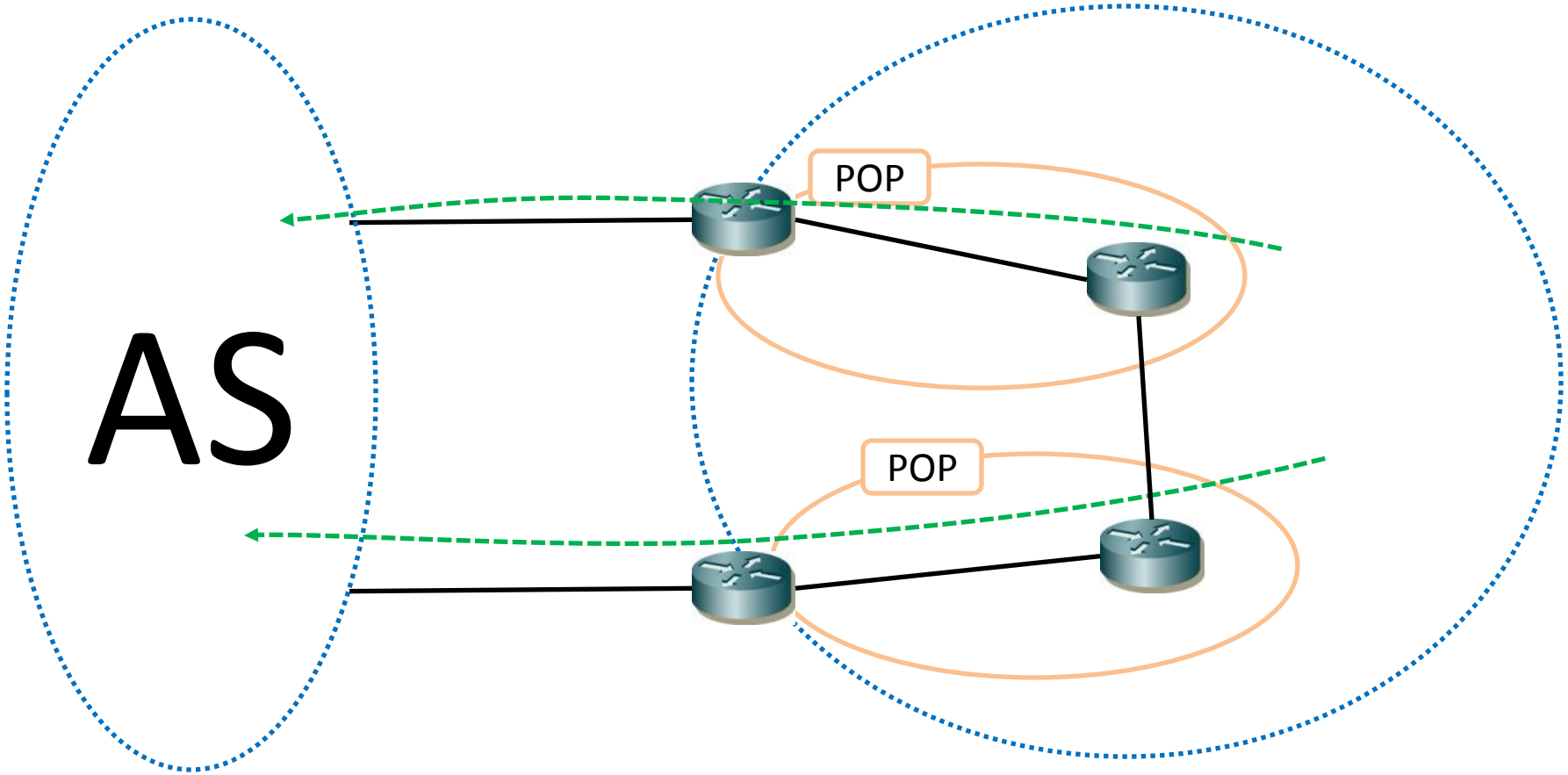
eBGP multi path



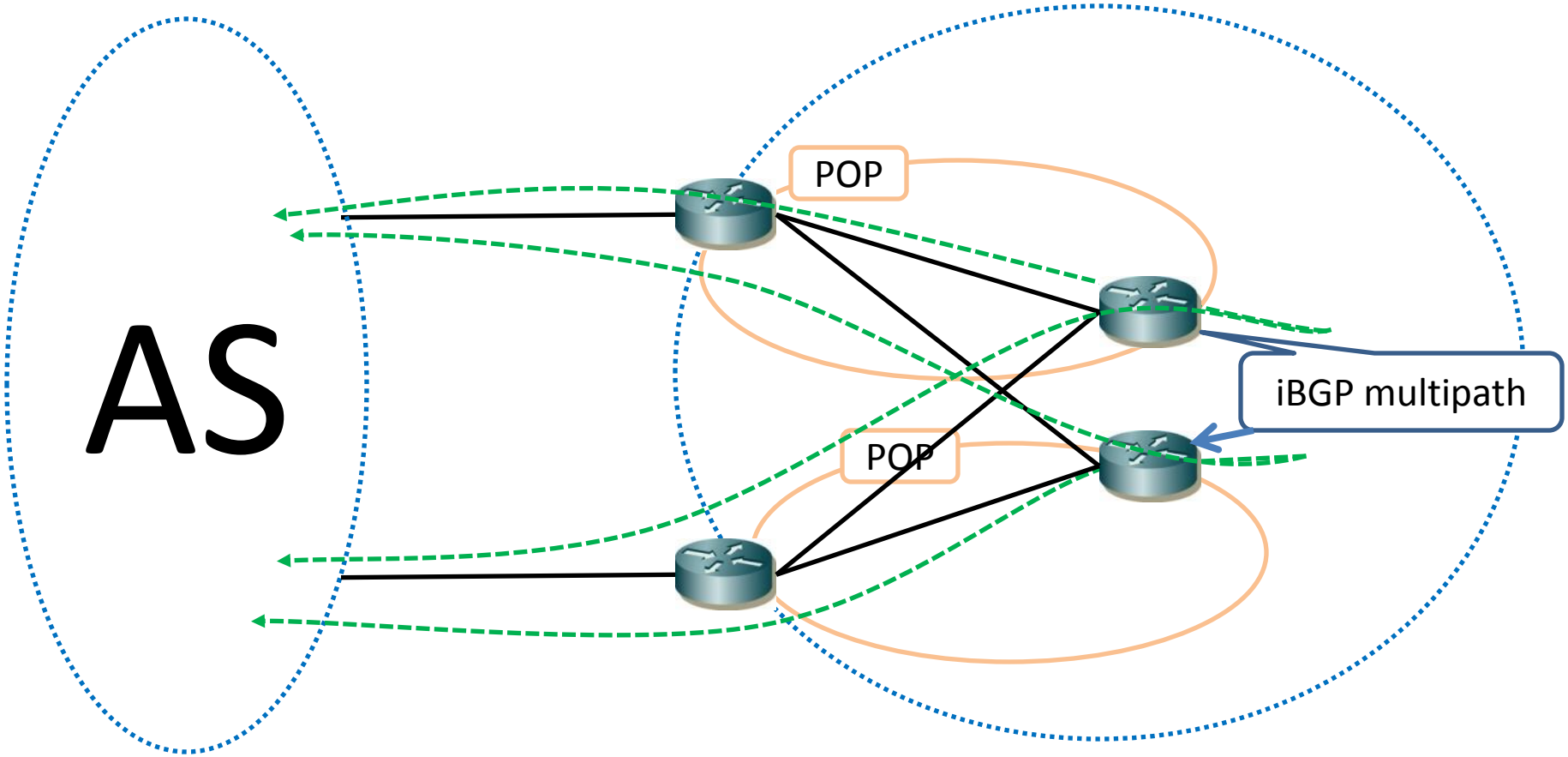
iBGP multipath



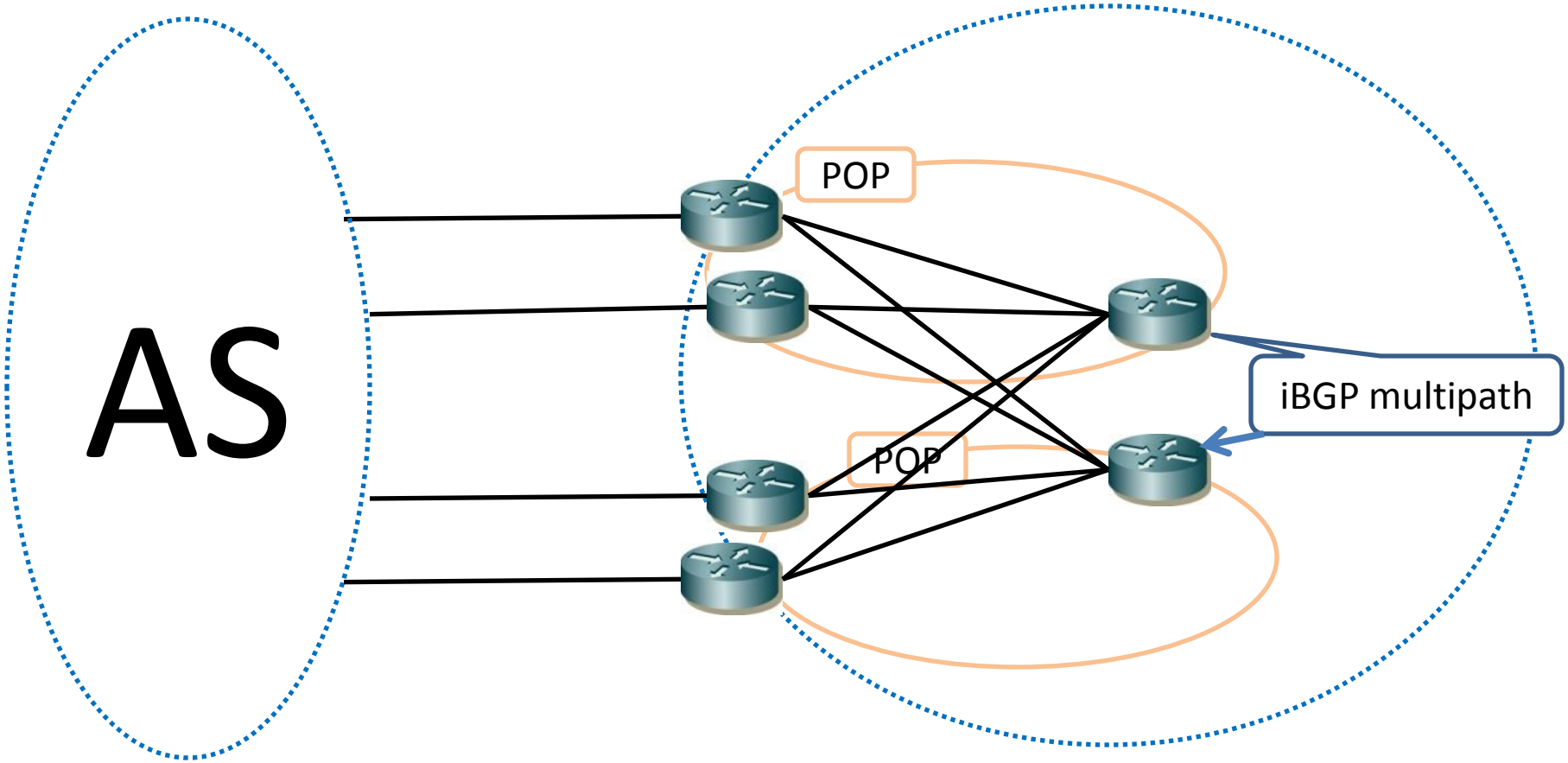
closest exitが一番簡単



POP越えのiBGP multipath



POP越えのiBGP multipath



2. 4byteへの移行

- まだ未対応 or ごく一部のみ対応済みの人が多い
（と聞いています。）
- AS4_PATHをみるとある程度わかる
  - Jan 21 17:08:15.075747 BGP RECV flags 0x40 code ASPath(2) length 16: XXXX YYYY ZZZZ 13249 6886 23456 35748
  - Jan 21 17:08:15.075769 BGP RECV flags 0x40 code NextHop(3): 122.28.179.205
  - Jan 21 17:08:15.075778 BGP RECV flags 0x80 code MultiExitDisc(4): 0
  - Jan 21 17:08:15.075790 BGP RECV flags 0xe0 code AS4Path(17) length 10: 196629 35748

- まだまだASDOTとASPLAINが今後も混在していくことが予想される
- 関連システムや諸々の実装もこれから

2byte cloud view (19routes)

IOS(2byte)>show ip bgp regexp _23456_

Network	Path
* i64.127.137.0/24	4713 2914 18508 23456 i
* i84.205.64.0/24	4713 2914 3561 1273 1103 1125 23456 12654 i
* i84.205.80.0/24	4713 2914 3561 1273 1103 1125 23456 12654 i
* i91.207.218.0/23	4713 2914 35320 23456 23456 ?
* i91.208.44.0/24	4713 2914 3257 1213 23456 i
* i169.222.0.0/24	4713 2914 701 7091 715 23456 i
* i192.26.93.0/24	4713 2914 4697 23456 i
* i193.5.68.0/23	4713 2914 6830 8758 23456 i
* i193.31.7.0/24	4713 2914 3549 5539 23456 i
* i195.47.195.0/24	4713 2914 3257 8495 23456 i
* i195.128.230.0/24	4713 2914 35320 23456 23456 35748 i
* i195.128.231.0/24	4713 2914 35320 23456 23456 35748 i
* i196.1.15.0/24	4713 2914 7018 3741 23456 i
* i197.255.248.0/22	4713 2914 174 3741 23456 i
* i202.255.47.0/24	4713 2516 7667 23456 i
* i205.233.128.0/24	4713 2914 10026 7657 23754 23754 9439 23456 i
* i2001:df0:2::/48	4713 2914 4697 23456 I
* i2001:4810:2000::/35	4713 2914 33437 23456 I
* i2403:2000::/32	4713 2914 4635 23911 24489 24490 9270 7660 2500 18146 23456 I

今朝この2経路が削除

4byte cloud view (19routes)

JUNOS(4byte)> show route aspath-regex ".* [65536-4294967295] .*"

inet.0: 270995 destinations, 270999 routes (270974 active, 21 holddown, 0 hidden)

64.127.137.0/24	AS path: 4713 2914 18508 393222 I
84.205.64.0/24	AS path: 4713 2914 3561 1273 1103 1125 196613 12654 I
84.205.80.0/24	AS path: 4713 2914 3561 1273 1103 1125 196613 12654 I
91.207.218.0/23	AS path: 4713 2914 35320 196629 AS_TRANS ?
91.208.44.0/24	AS path: 4713 2914 3257 1213 196623 I
169.222.0.0/24	AS path: 4713 2914 701 7091 715 131076 I
192.26.93.0/24	AS path: 4713 2914 4697 131075 I
193.5.68.0/23	AS path: 4713 2914 6830 8758 196621 I
193.31.7.0/24	AS path: 4713 2914 3549 5539 196611 I
195.47.195.0/24	AS path: 4713 2914 3257 8495 196624 I
195.128.230.0/24	AS path: 4713 2914 35320 196629 AS_TRANS 35748 I
195.128.231.0/24	AS path: 4713 2914 35320 196629 AS_TRANS 35748 I
196.1.15.0/24	AS path: 4713 2914 7018 3741 327681 I
197.255.248.0/22	AS path: 4713 2914 174 3741 327681 I
202.255.47.0/24	AS path: 4713 2516 7667 131078 I
205.233.128.0/24	AS path: 4713 2914 10026 7657 23754 23754 9439 131110 I

今朝この2経路が削除

inet6.0: 1381 destinations, 1391 routes (1381 active, 0 holddown, 0 hidden)

2001:df0:2::/48	AS path: 4713 2914 4697 131075 I
2001:4810:2000::/35	AS path: 4713 2914 33437 393219 I
2403:2000::/32	AS path: 4713 2914 4635 23911 24489 24490 9270 7660 2500 18146 131081 I

勝手にAS_TRANSを利用？

■4byte対応ルータ

>show ip bgp 195.128.231.0/24

BGP routing table entry for 195.128.231.0/24

Paths: (13 available, best #9, table Default-IP-Routing-Table)

Advertised to non peer-group peers:

4713 2914 35320 3.21 23456 35748

通常では出現しないはず?
現在広告元に問い合わせ中

■4byte未対応ルータ

>show ip bgp 195.128.231.0/24

BGP routing table entry for 195.128.231.0/24, version 49541885

Paths: (2 available, best #2, table Default-IP-Routing-Table)

Not advertised to any peer

4713 2914 35320 23456 23456 35748

各ベンダの対応状況

- 対応PLATFORM等詳しい状況は各社ベンダに要問合せ

	Cisco	Force10	Juniper	Foundry	Alaxala	Redback
IMPLEMENTATION	ASDOT	ASPLAIN ASDOT	ASPLAIN ASDOT	ASPLAIN ASDOT ASDOT+	ASPLAIN? (planning)	ASPLAIN? ASDOT?
CURRENT OS	IOS-XR3.4- 12.2SXI 12.4T	FTOS -ASPLAIN(7.7.1-) -ASDOT(7.8.1-)	JUNOS -ASPLAIN(9.1-) -ASDOT(9.2-) JUNOSe -ASPLAIN(4.1. -)	NetIronXMR MLX Version04.0.00-		SEOS All support version
FUTURE OS	CY09Q2～Q3 頃 12.0S (32SY/33S)- 12.2SXI- 12.4T-	CY09Q1~Q2 FTOS8.2 Additional ASPLAIN Functions			CY09Q1~Q2 （planning）	

変更や追記などありましたら yoshida@nttv6.jp まで

各種4byte対応

項目	4byte対応状況
BGP Community	draft-rekhter-as4octet-ext-community-03.txt を実装している所が多い
IRR	RIPE, APNIC, RADB等はASDOT対応 JPIRRはASPLAIN表記可（但し要注意）
RPSL	draft-uijterwaal-rpsl-4byteas-ext-03.txt :expired
Looking glass	RISなど徐々に対応されてきている （但しASDOT表記）
xFLOW	Netflow v9で対応（対応は機種による） sFLOWも仕様上対応済み コレクタ側の対応はこれから
MRT	対応済みにみえる draft-ietf-grow-mrt-08.txt
MIB	RFC4273 : Definitions of Managed Objects for BGP-4 SYNTAX　inteter32(0..65535) draft-ietf-idr-bgp4-mibv2 : expired
Multicast	eGLOPの拡張提案が過去でたが、pending

IRR based Filter

- AS_PATHフィルタの置き換え
  - AS-SETオブジェクトのmember ASより自動生成
- 4byteASはまだAS-SETに書けない
  - IRRToolsetも未対応
- スクリプトで変換
  - AS-SETのmember ASに AS23456が含まれている場合
  - 本来の4byteASをdesc:などに記載の上参照する

従来のAS_PATH UPDATE

- 手動の場合には、以下の例だと非常に親切
  1. **^(AS131077_)+$** **4octet/ASPLAIN**
  2. **^(AS2¥.5)_+$** **4octet/ASDOT**
  3. **^(AS23456_)+$** **2octet/AS_TRANS**
  - 受信側は4octet ASで受信するなら1.か2.で対応、2octetで受信するなら3.で対応する必要がある

- さらに他に広告する場合は以下イメージ
  1. **^(AS65400_)+(AS131077_)+$** **4octet/ASPLAIN**
  2. **^(AS65400_)+(AS2¥.5_)+$** **4octet/ASDOT**
  3. **^(AS65400_)+(AS23456_)+$** **2octet/AS_TRANS**

4byte運用対処

- 複雑には管理したくない
- けどきちんと管理する必要がある
- ツールなどを用いてきちんとコンバートできるような対応が当面移行時には必要
  - ASDOT(+)->ASPLAIN
  - ASPLAIN->ASDOT(+)
- Script等のtoolが公開できるといいかも
- IRRへの標準的な記述方法を作りましょう

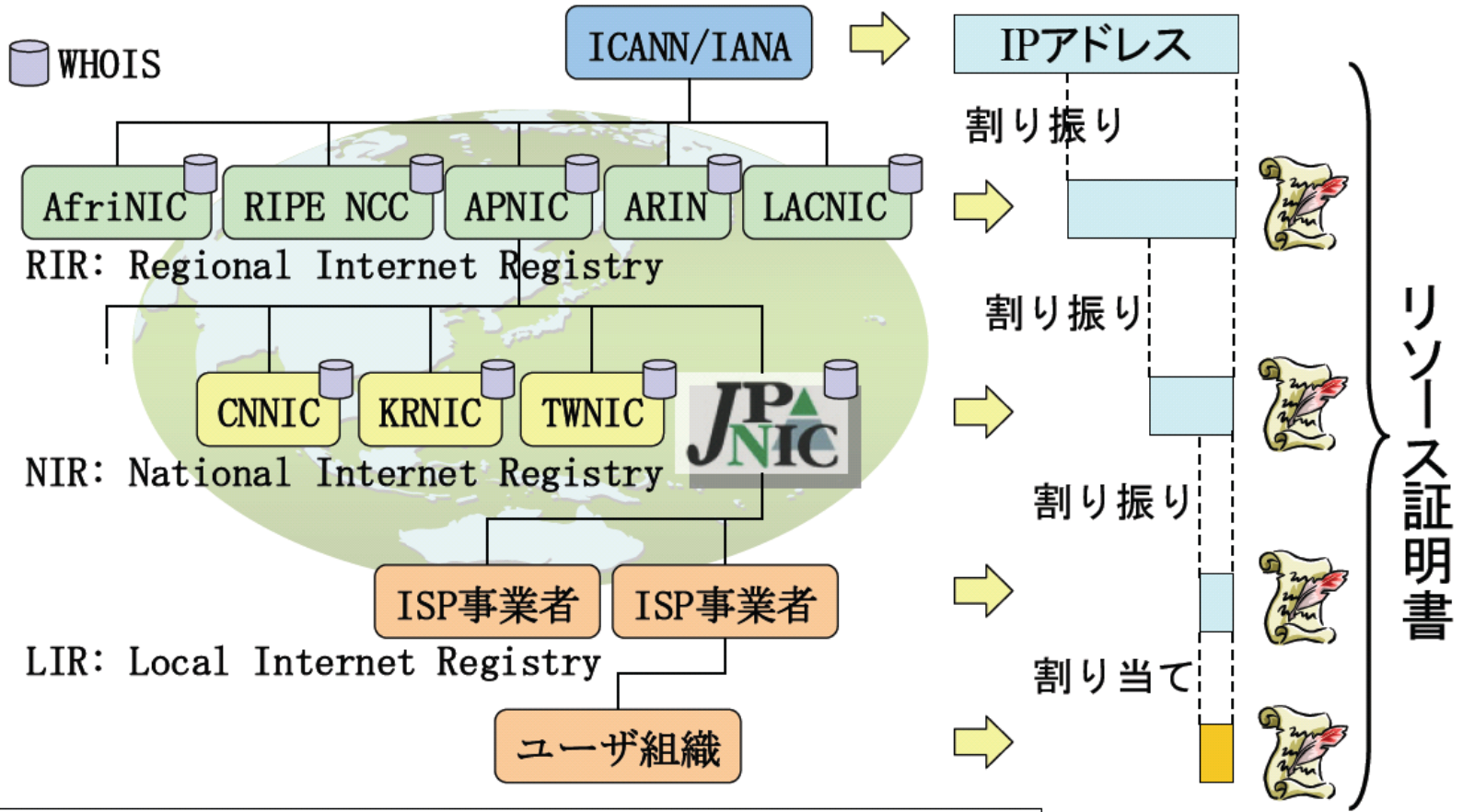
3. 不正経路混入防止

- 不正経路が流入することを未然に防ぐ
- BGPアップデートの受信時に経路の正当性をルータが確認
  - 不正な場合にはフィルタ
  - 怪しいなどの場合にもオペレータのポリシー次第ではフィルタが可
  - ポリシーの整合性の問題はある

SIDR WG

- Secure Inter-domain Routing WG
  - Inter-domain Routingにおけるセキュリティフレームワークアーキテクチャの検討
    - 拡張可能なインタードメインにおけるセキュアルーティングアーキテクチャの文書化
    - セキュアルーティングアーキテクチャに含まれる証明書の利用方法に関する文書化
    - RPSEC WGにより決定された安全なルーティングへの要求に焦点をおいた、このアーキテクチャが提供するルーティング機能の構成要素の文書化
  - 2006年末よりWG化
  - http://www.ietf.org/html.charters/sidr-charter.html

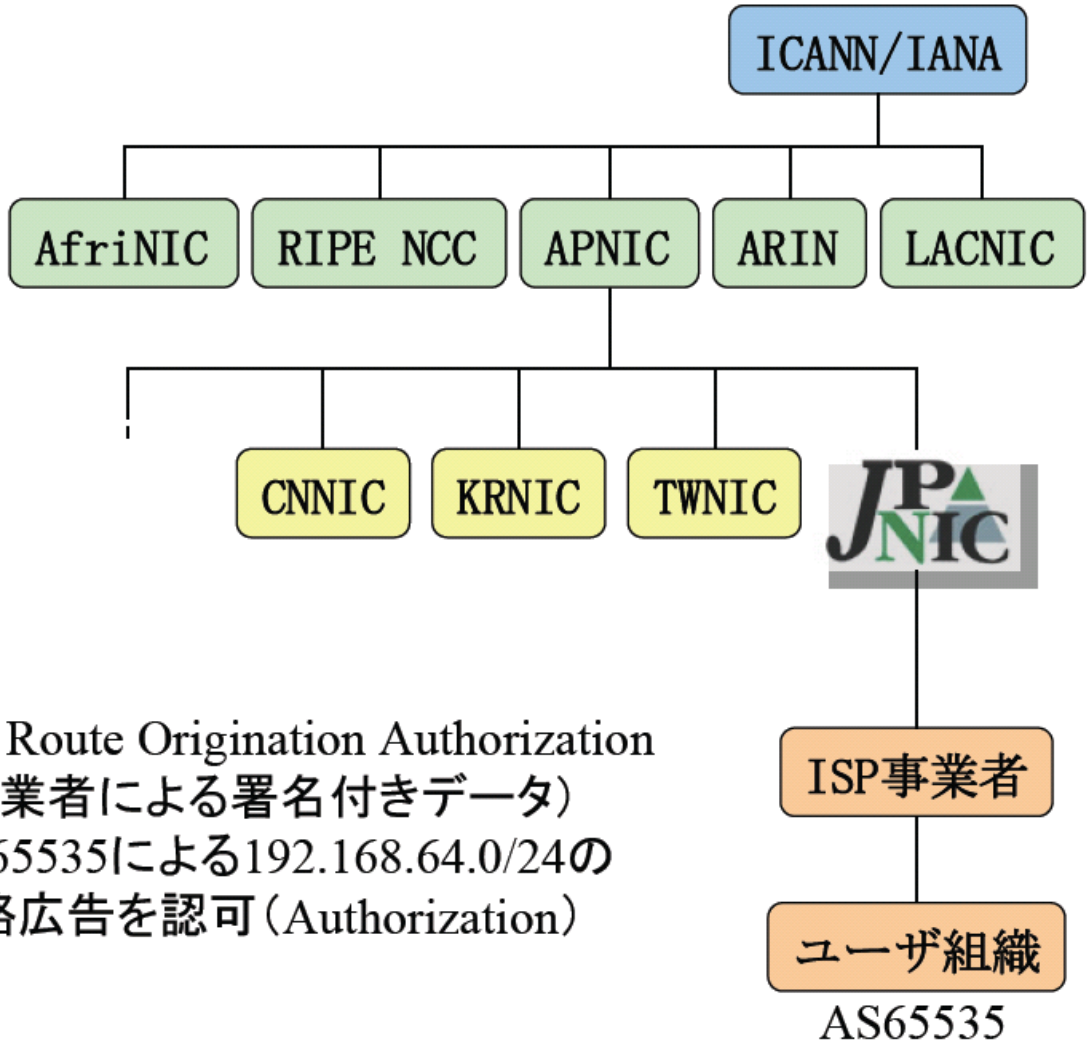
RPKI



IPアドレスの一意性を保証し経路制御の適応性を向上させる仕組み

出典：InternetWeek2007 JPNIC木村さんより

ROA



発行元：APNIC
対象：JPNIC
アドレスブロック：
192.0.0.0/8

発行元：JPNIC
対象：ISP事業者
アドレスブロック：
192.168.0.0/16

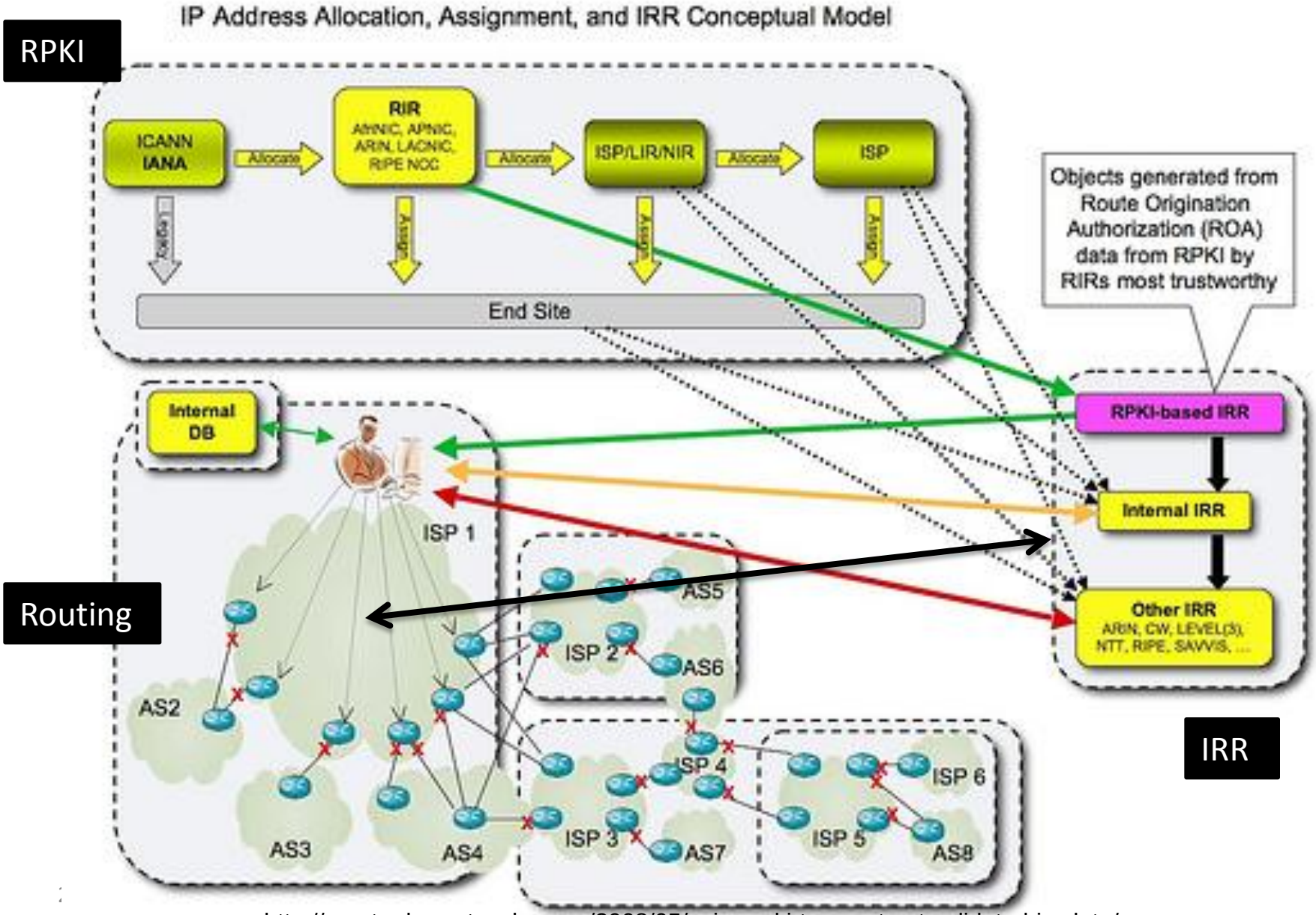
発行元：ISP事業者
対象：ユーザ組織
アドレスブロック：
192.168.64.0/22

ROA – Route Origination Authorization
（ISP事業者による署名付きデータ）

- AS65535による192.168.64.0/24の経路広告を認可（Authorization）

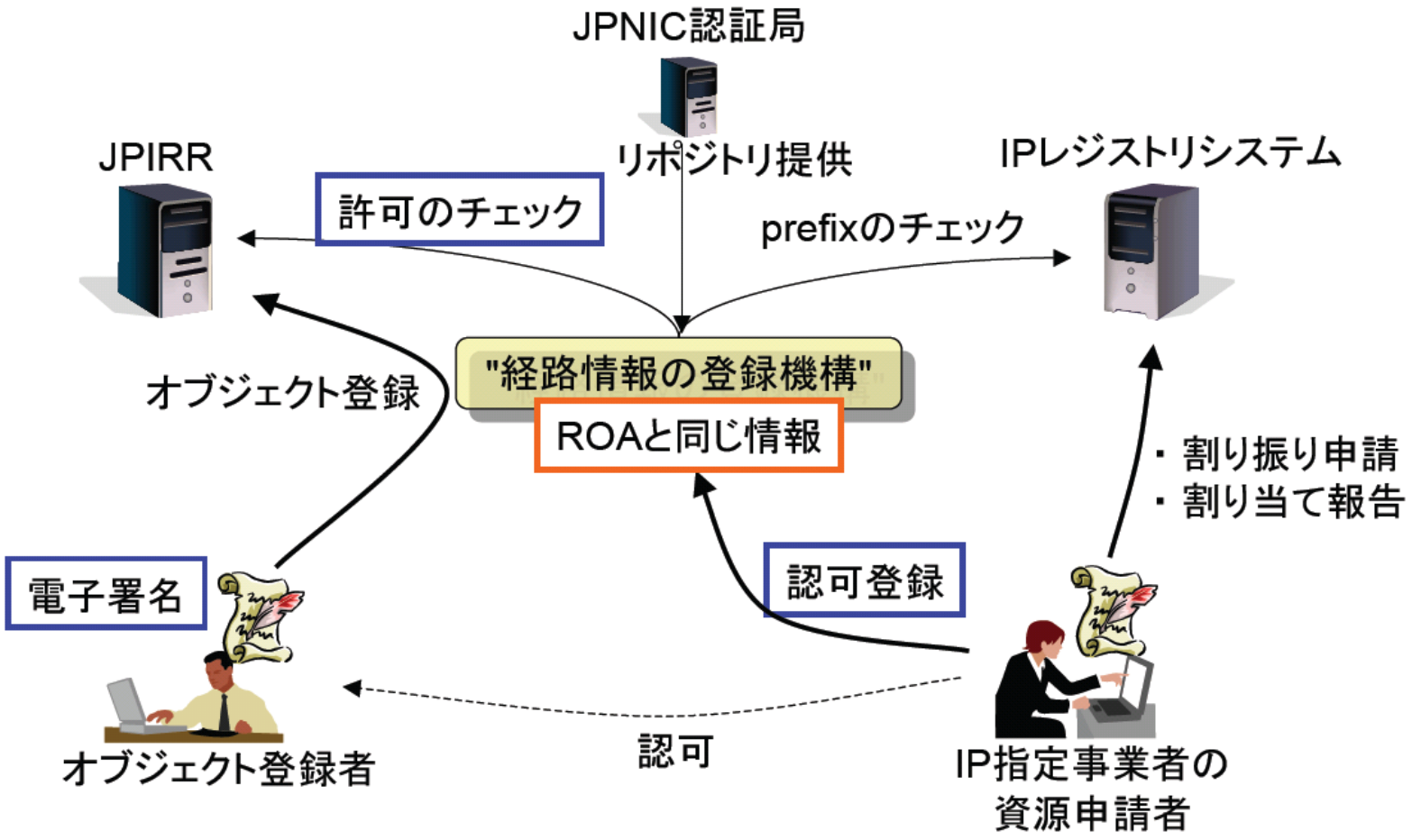
出典：InternetWeek2007 JPNIC木村さんより

RPKI+IRR+Routing



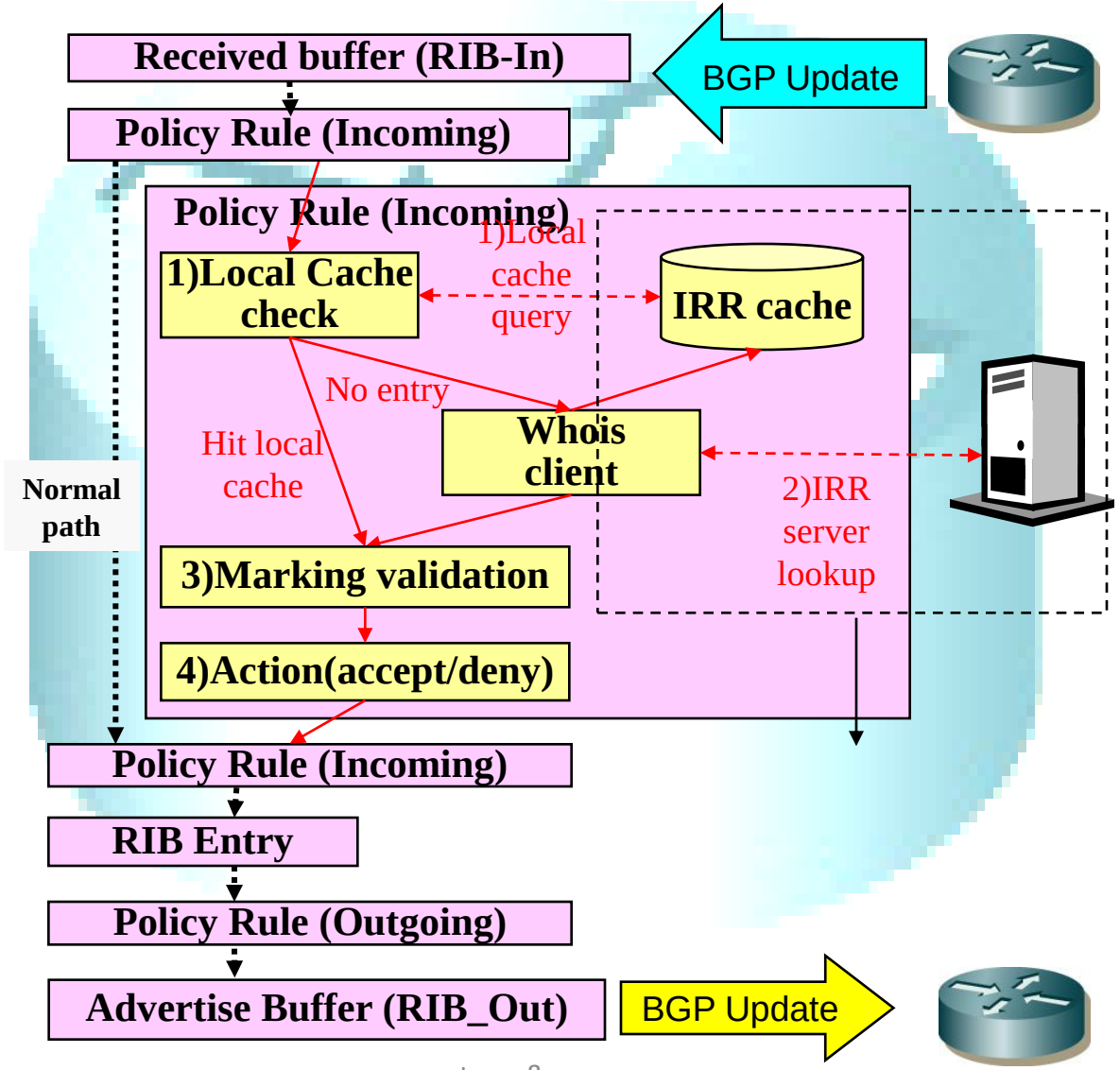
http://asert.arbornetworks.com/2008/05/using-rpki-to-construct-validated-irr-data/

JPNIC

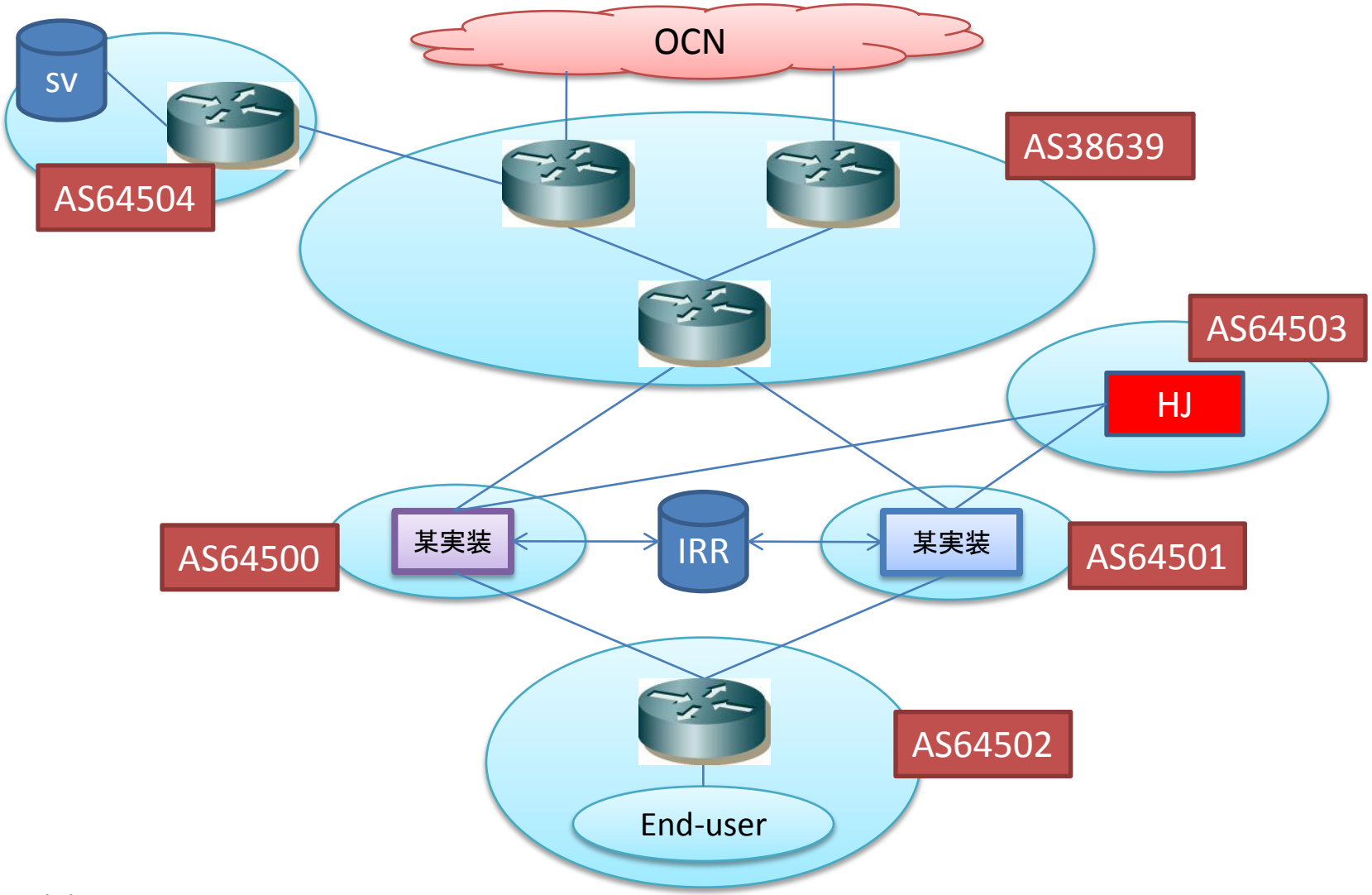


出典：InternetWeek2007 JPNIC木村さんより

Router Extention



Experimentation



御清聴有難うございました。

おしまい