最近のBGP ○○編

Yoshida Tomoya – yoshida@ocn.ad.jp
18 28 48 37 FE C2 D9 95 36 41 5D 7B D7 55 A2 55 63 FA E0 BF

Matsuzaki Yoshinobu - maz@iij.ad.jp
8E 9C EC 04 87 6B B5 0E 1B 6D 46 3B CB F7 72 CE

本日のアジェンダ

第一楽章「そりゃないよね編」
第二楽章「きもい編」
第三楽章「こりゃどうだ編」

本日のアジェンダ

第一楽章「そりゃないよね編」

第二楽章「きもい編」

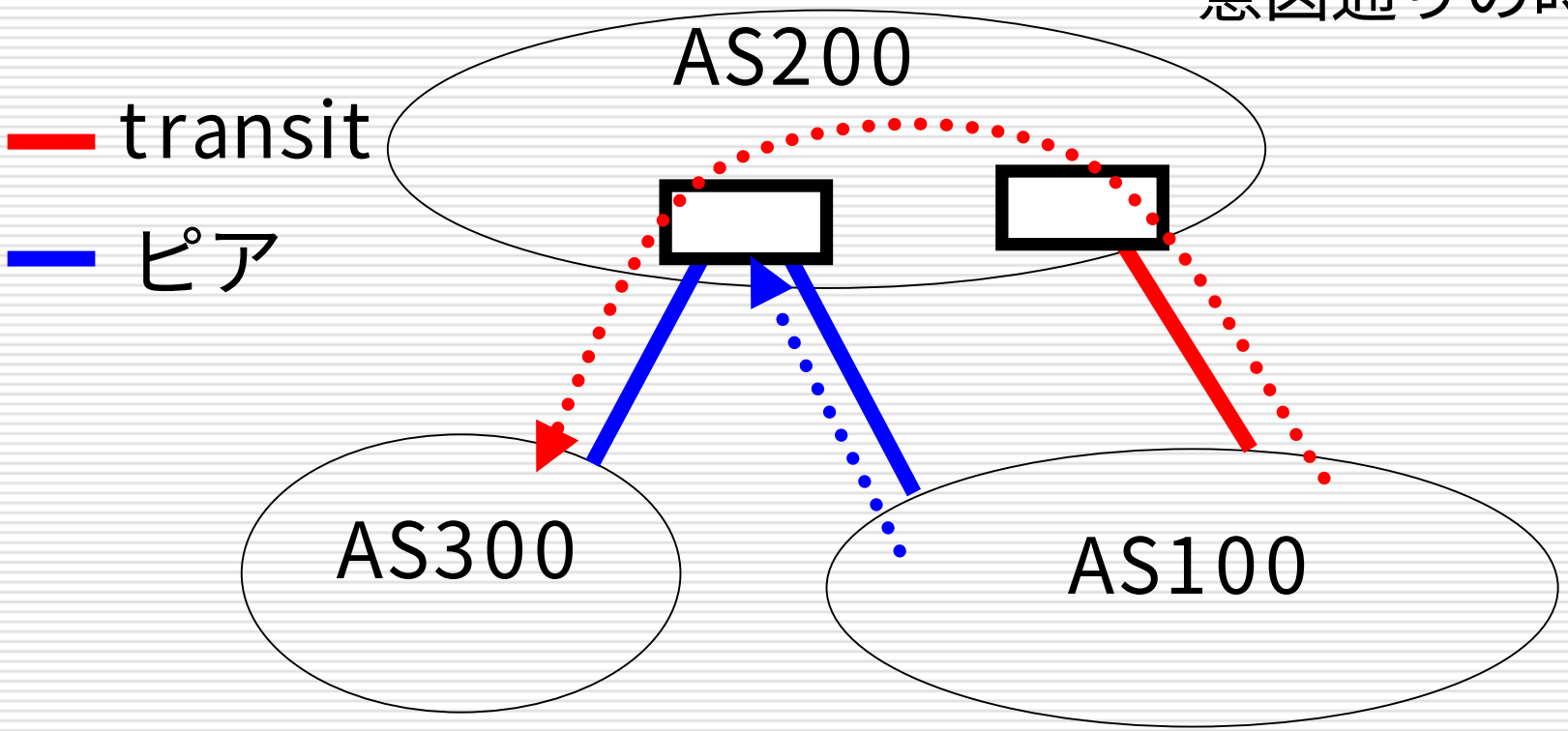
第三楽章「こりゃどうだ編」

経路が聞こえてこない。。。

- 経路の使い分け
  - 優先度は、顧客＞ピア≧アップストリーム
  - 顧客経路の優先付けを間違うと、ピア先に経路が広報されない場合がある
- 難しいポリシ実装が引き金になることも
  - 顧客∧ピア

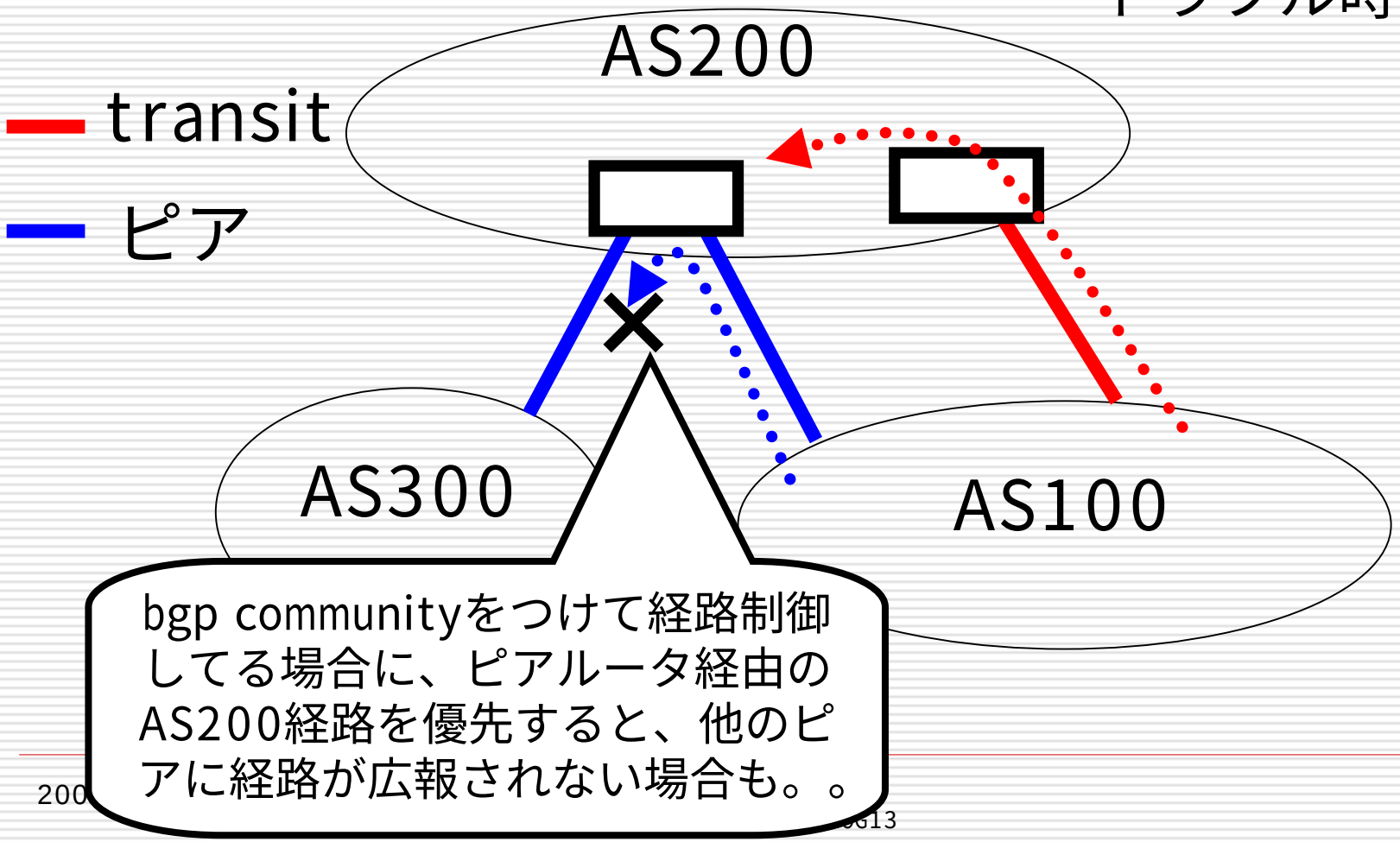
経路が聞こえてこない。。。（続き）

意図通りの時



経路が聞こえてこない。。。（続き）

トラブル時



経路が聞こえすぎる。。。

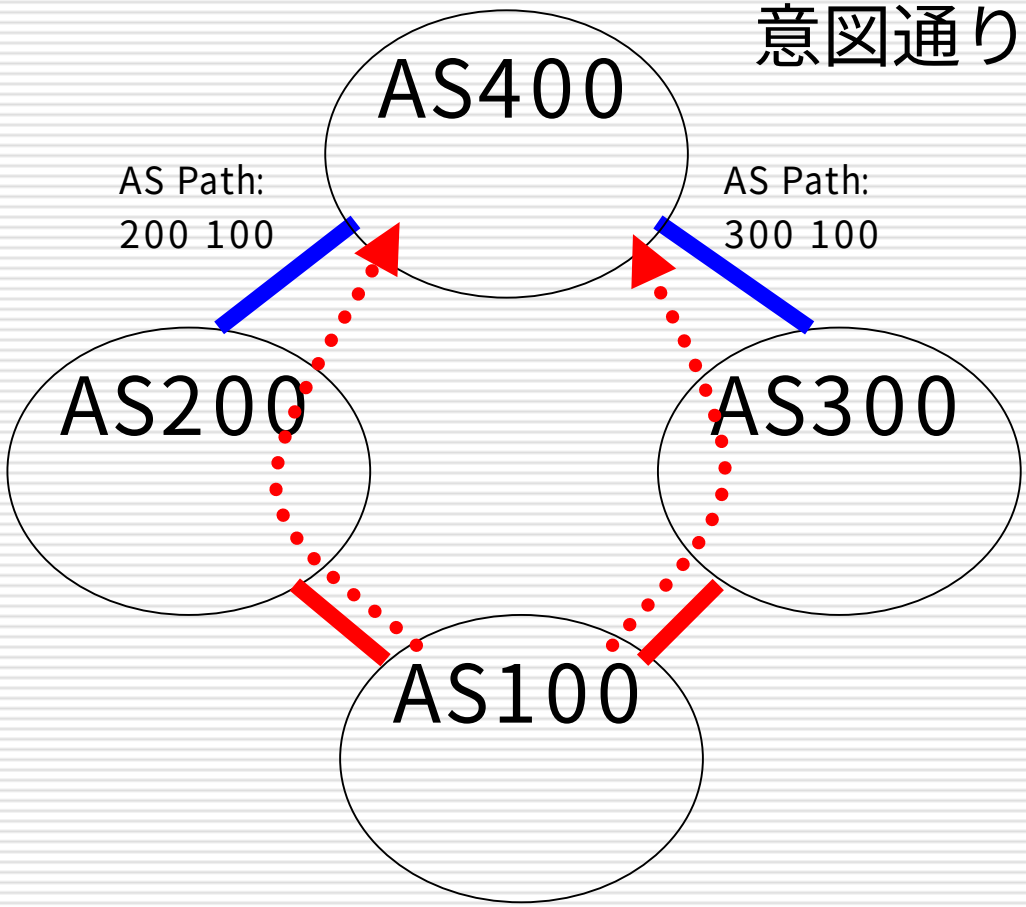
- 本当に意図してるならいいんですが
- prefix filterのみでOutbound ポリシを実装していると発生する場合が

経路が聞こえすぎる。。。（続き）

意図通りの時

transit

ピア

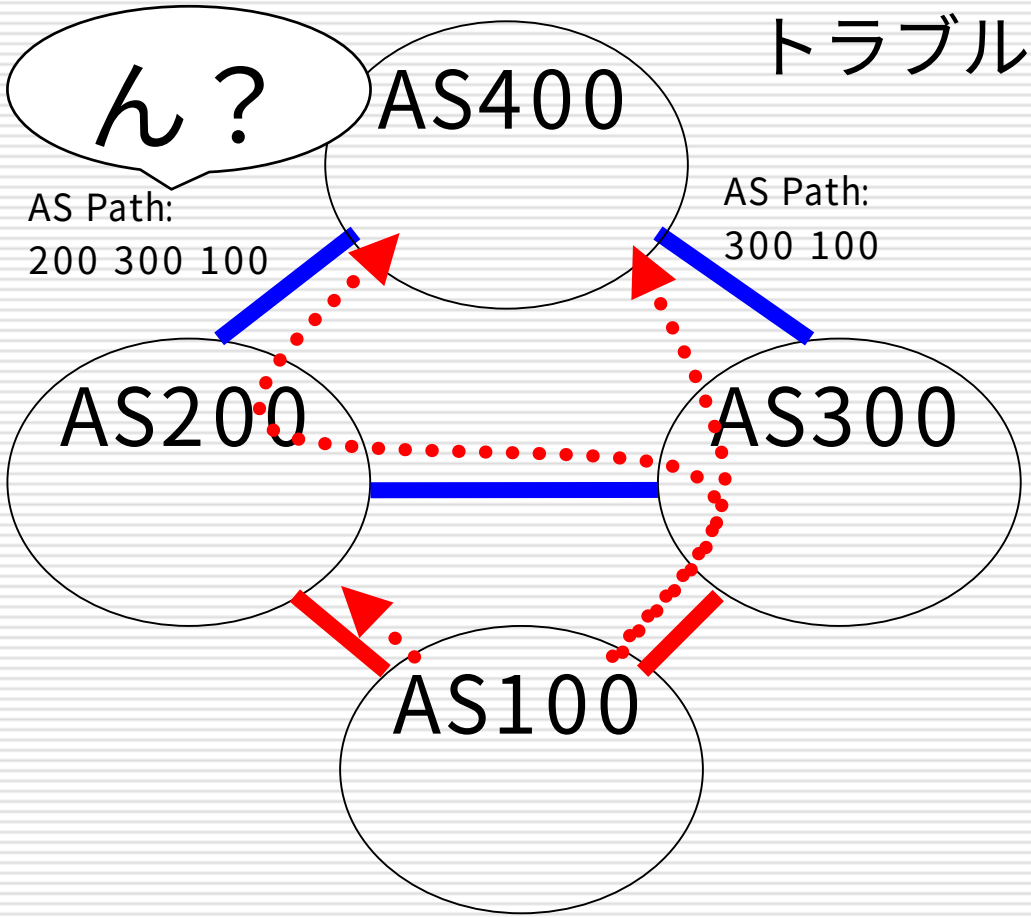


経路が聞こえすぎる。。。（続き）

トラブル？時

transit

ピア



/32はご勘弁

- やりすぎなor変なprefixの広報ががが！！
- 類似例
  - /31、/30 - Point to Point link?
- その他
  - 10.0.0.0
  - 172.16.0.0
  - 192.168.0.0
  - 0.0.0.0/0 - ほんとに？

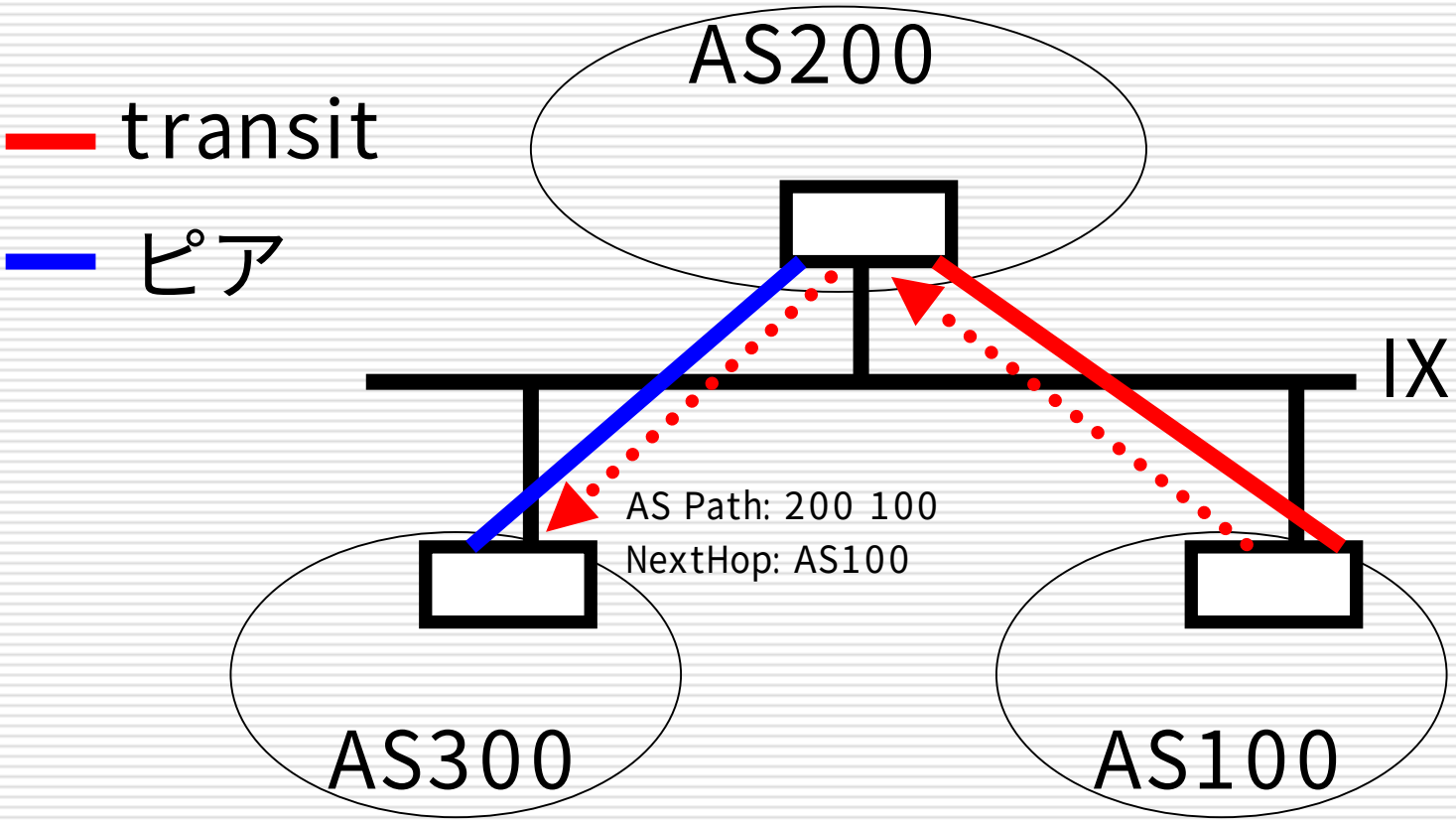
/32はご勘弁（続き）

- IngressのPrefix Filterは重要
- 考えどころは
  - default、Private、Host Loopback、etc
    - see also RFC3330
  - 細かい経路(/25 or longerとか)
  - 未割り当て経路
- 運用負荷と天秤
  - Prefix Filterの更新
    - http://www.cymru.com/Documents/secure-bgp-template.html

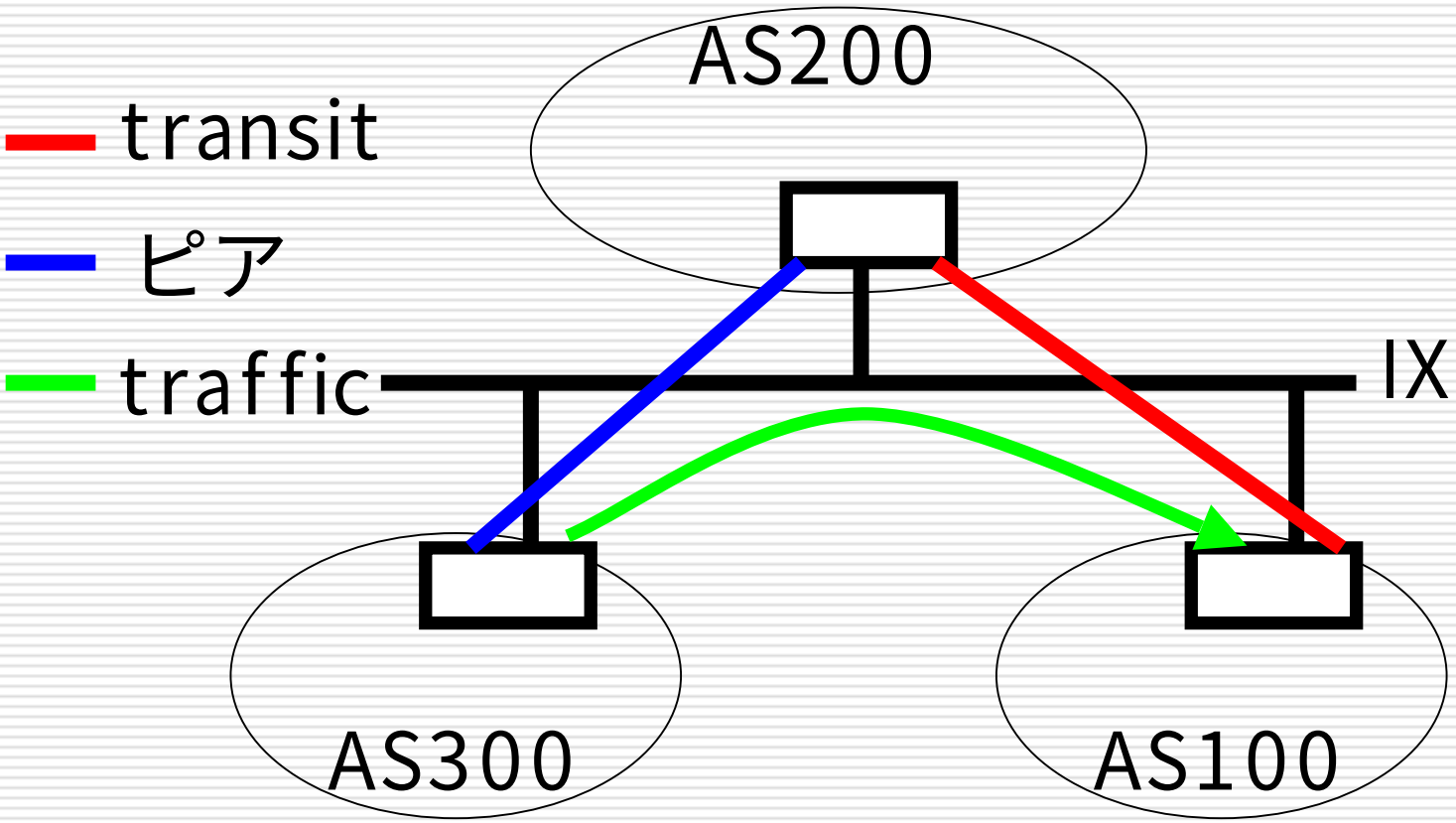
next-hop-selfを忘れずに

- ethernet上でトラヒックがピンポンするのを防いでくれますが、トラブルの引き金になることもあります
- あえて避ける理由が無いのであれば、next-hop-selfしましょう

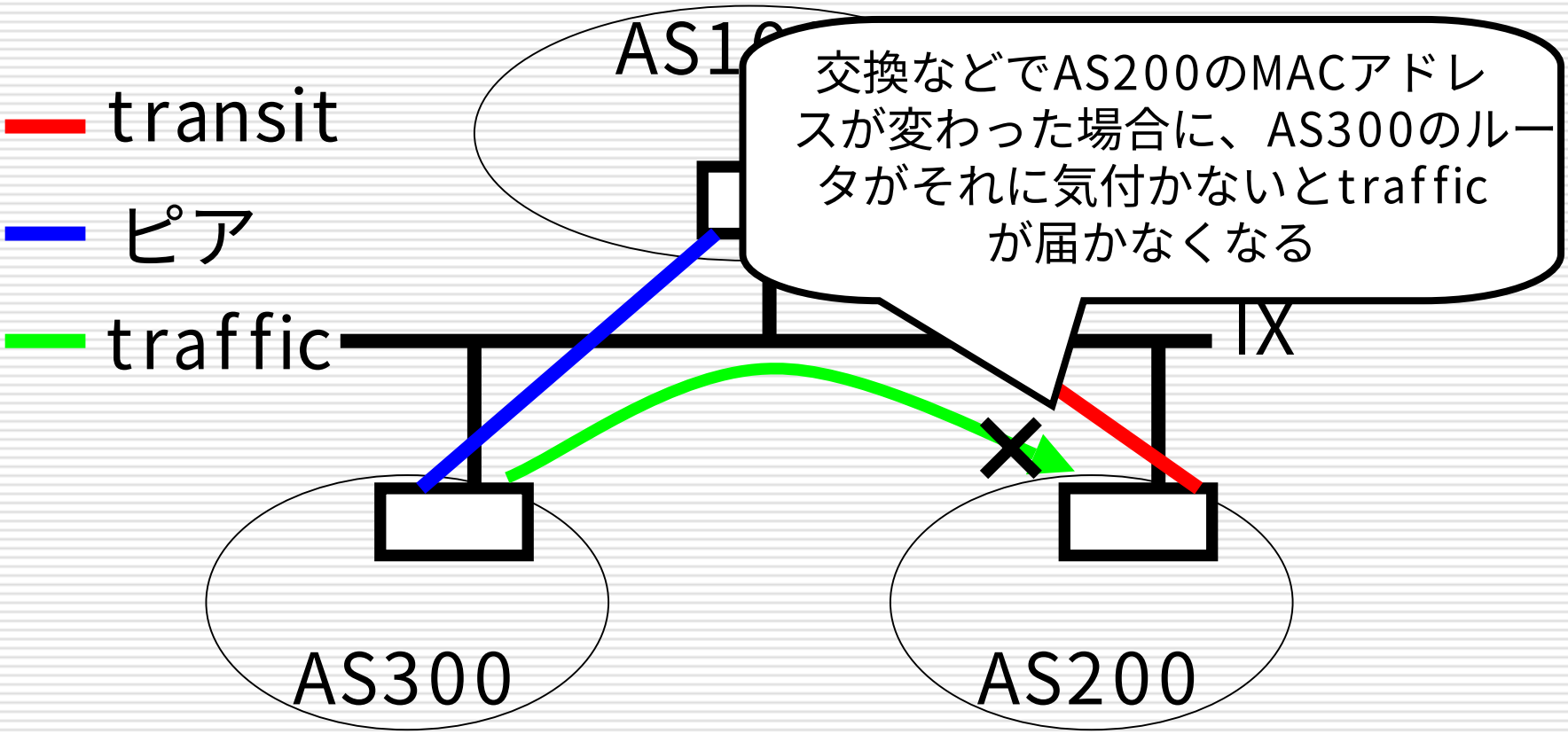
next-hop-selfを忘れずに（続き）



next-hop-selfを忘れずに（続き）



next-hop-selfを忘れずに（続き）



Ｄｙｎａｍｉｃ！

- インターネットの経路情報は日々うつろい、トラヒックの傾向も変わり続けます
- それでもＦＡＱ
  - トラヒック傾向が昨日と違う
  - 経路が変わった

Ｄｙｎａｍｉｃ！（続き）

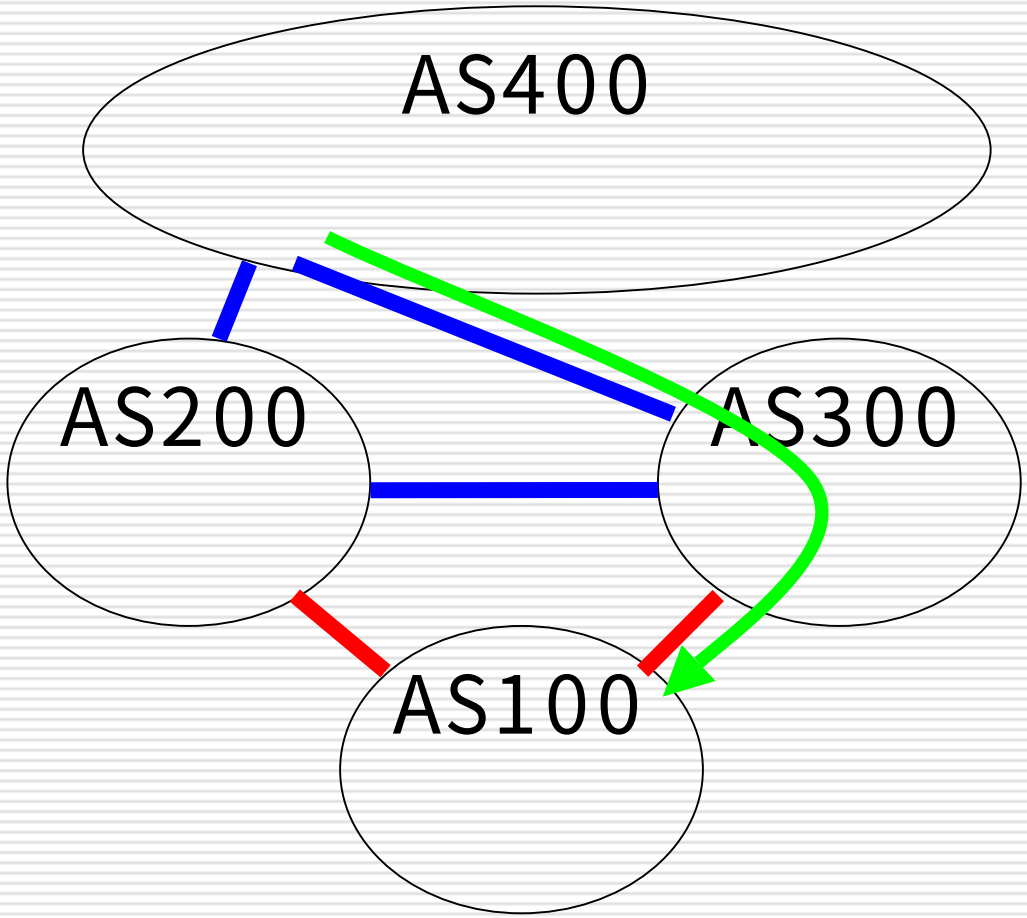
- 経路の変動
  - トラブル
    - 回線断、機器障害、設定ミス
  - メンテナンス
    - 回線、機器、構成変更
  - 顧客・ピア関係の変動
    - 開通、解約、増速減速

Ｄｙｎａｍｉｃ！（続き）

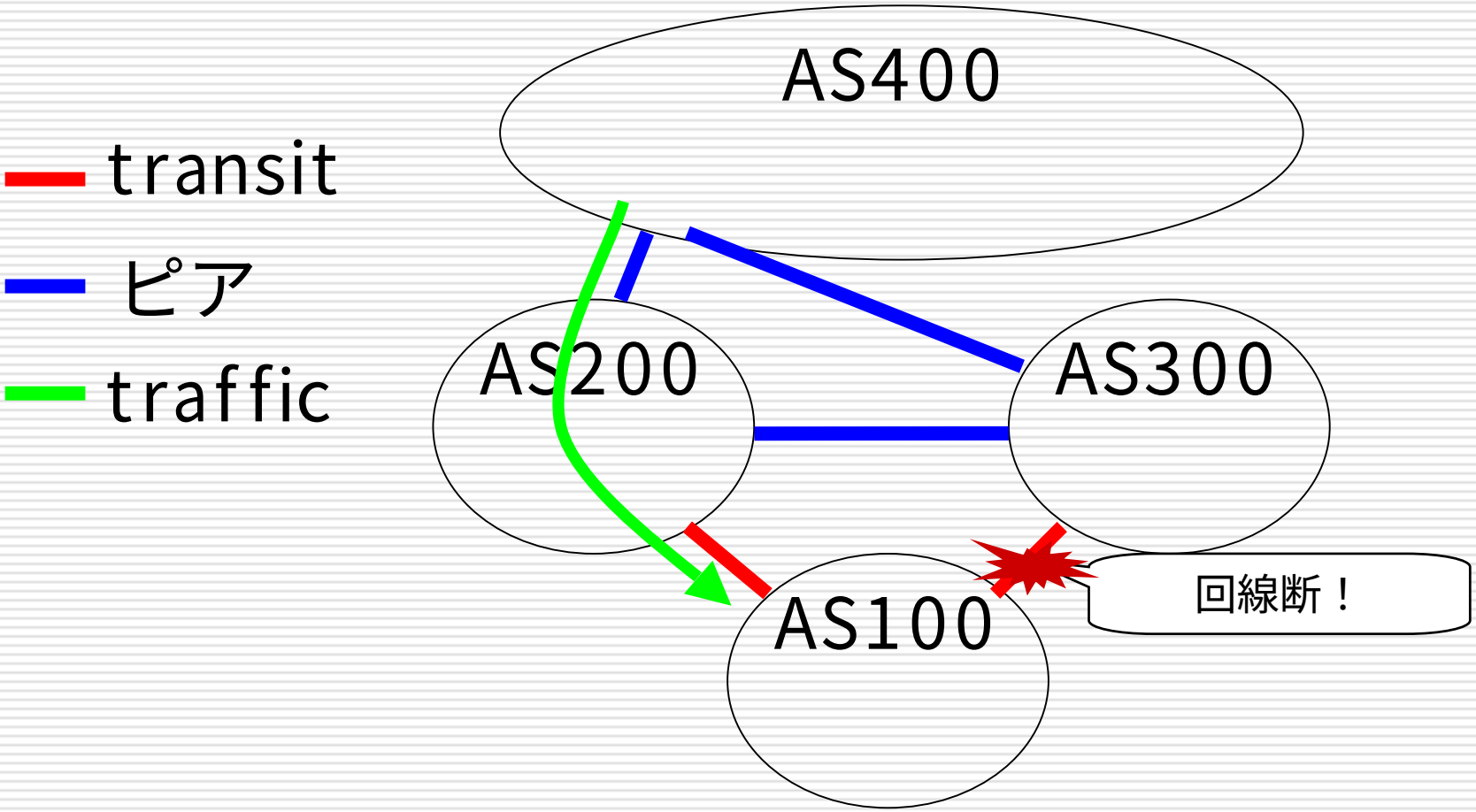
transit

ピア

traffic



Ｄｙｎａｍｉｃ！（続き）

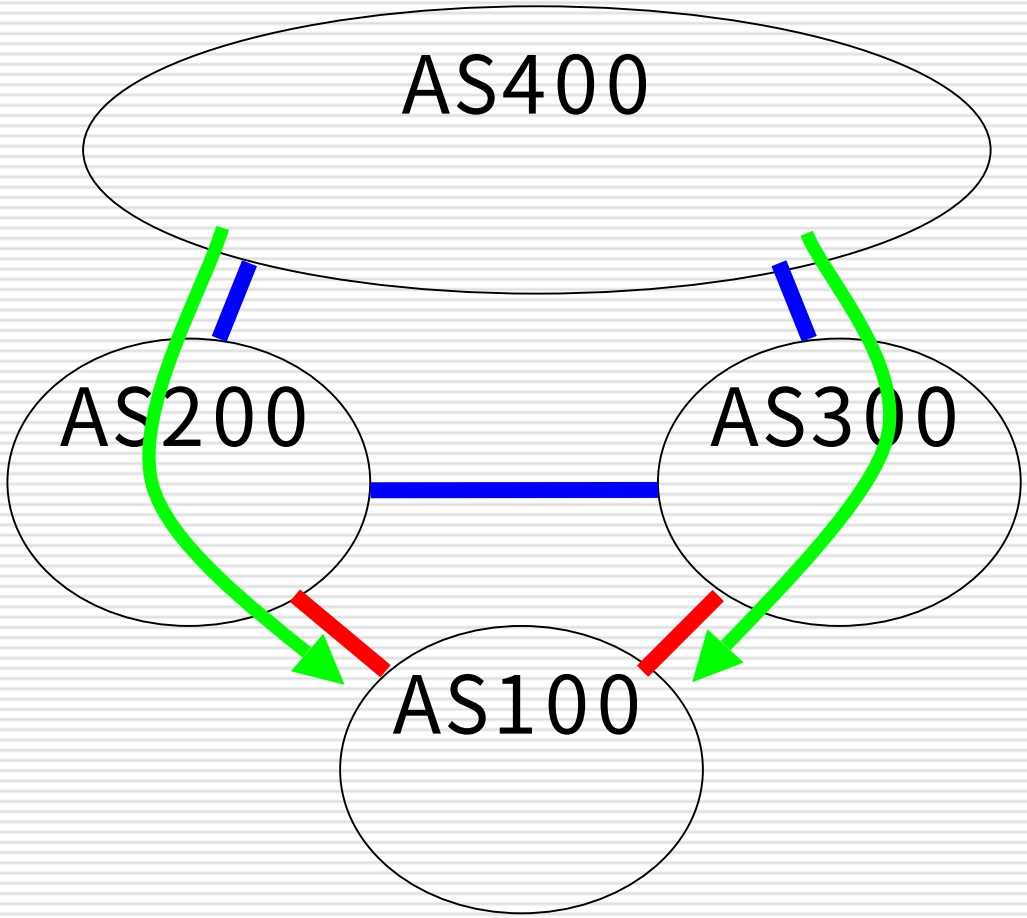


Ｄｙｎａｍｉｃ！（続き）

transit

ピア

traffic

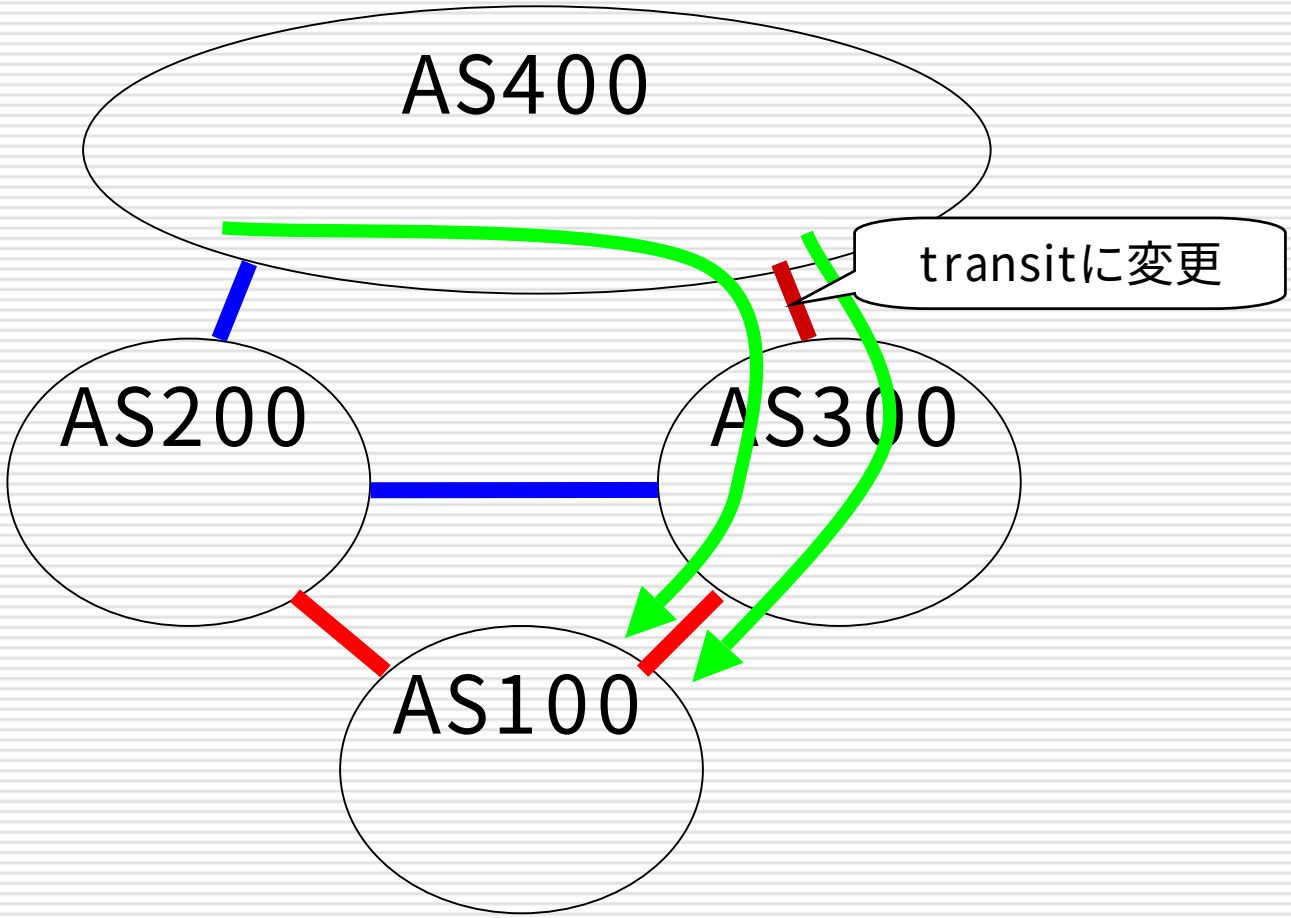


Ｄｙｎａｍｉｃ！（続き）

transit

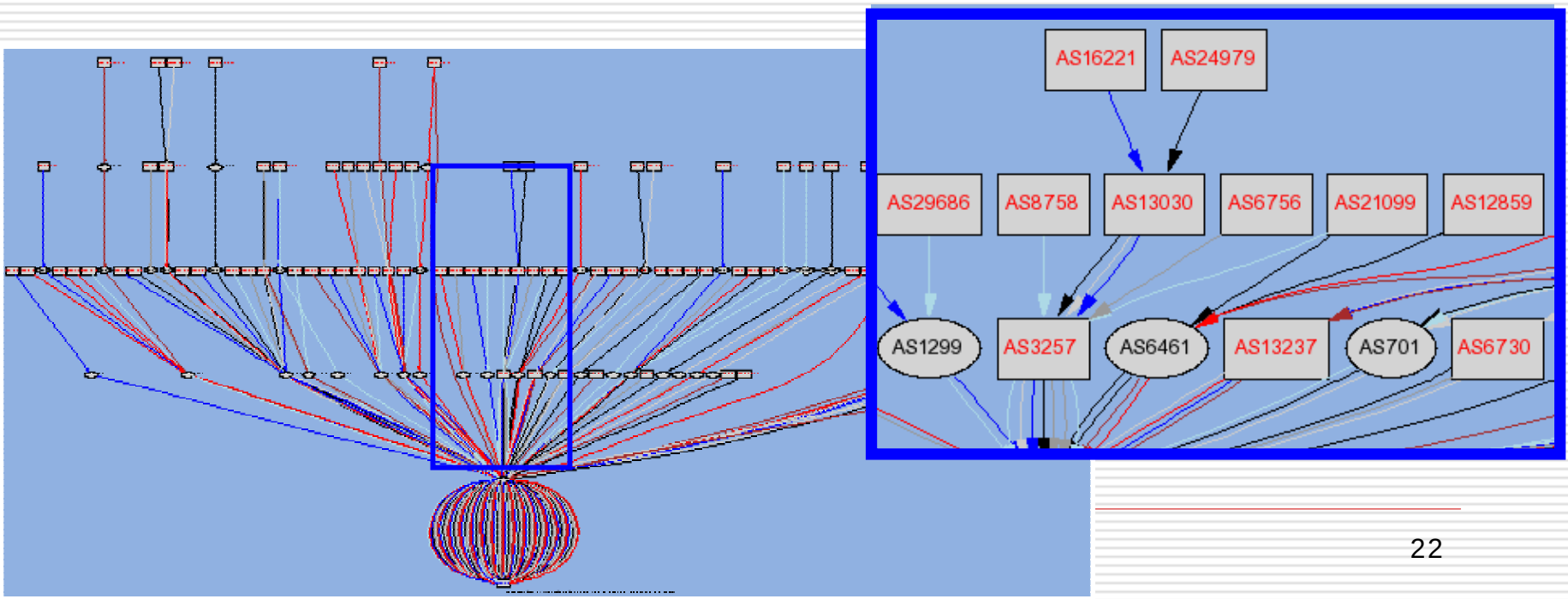
ピア

traffic



あの描画って生々しくないですか？

- Netlantis project
  - http://www.netlantis.org/



そりゃないよね編（まとめ）

- BGPの設定ミスって結構ありそうです
  - ピアにfull routeを広報
  - 不正経路の広報
    :
  - Understanding BGP misconfiguration
    - http://www.cs.washington.edu/homes/ratul/bgp/
- 経路は日々変わり続けます

そりゃないよね編（まとめの続き）

- 接続性を維持するためにも、
  - Egressで過不足なく経路を広報する
  - Ingressで不要な経路を受け取らない
  - **素直なポリシで素直に接続！**

を心がけると安心です

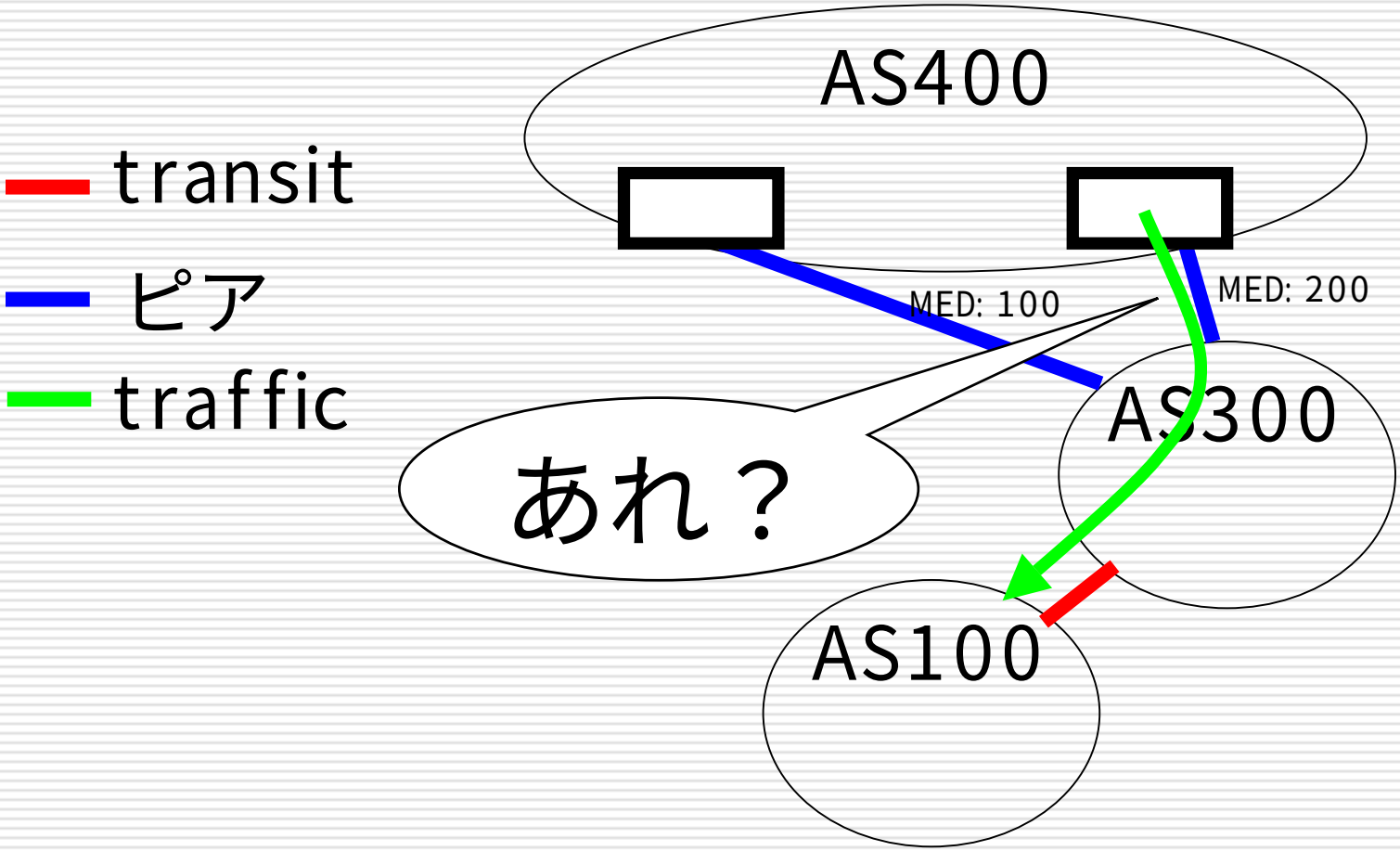
本日のアジェンダ

第一楽章「そりゃないよね編」

第二楽章「きもい編」

第三楽章「こりゃどうだ編」

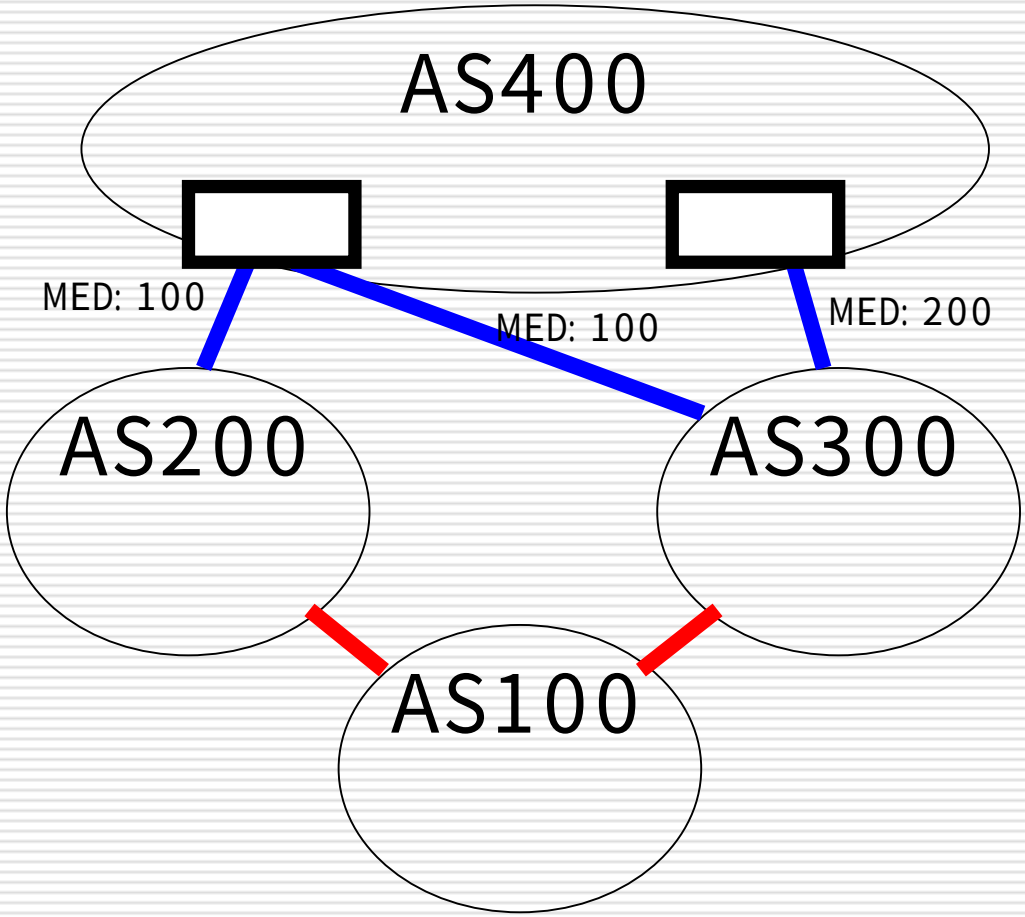
「ＭＥＤ」が無視される？



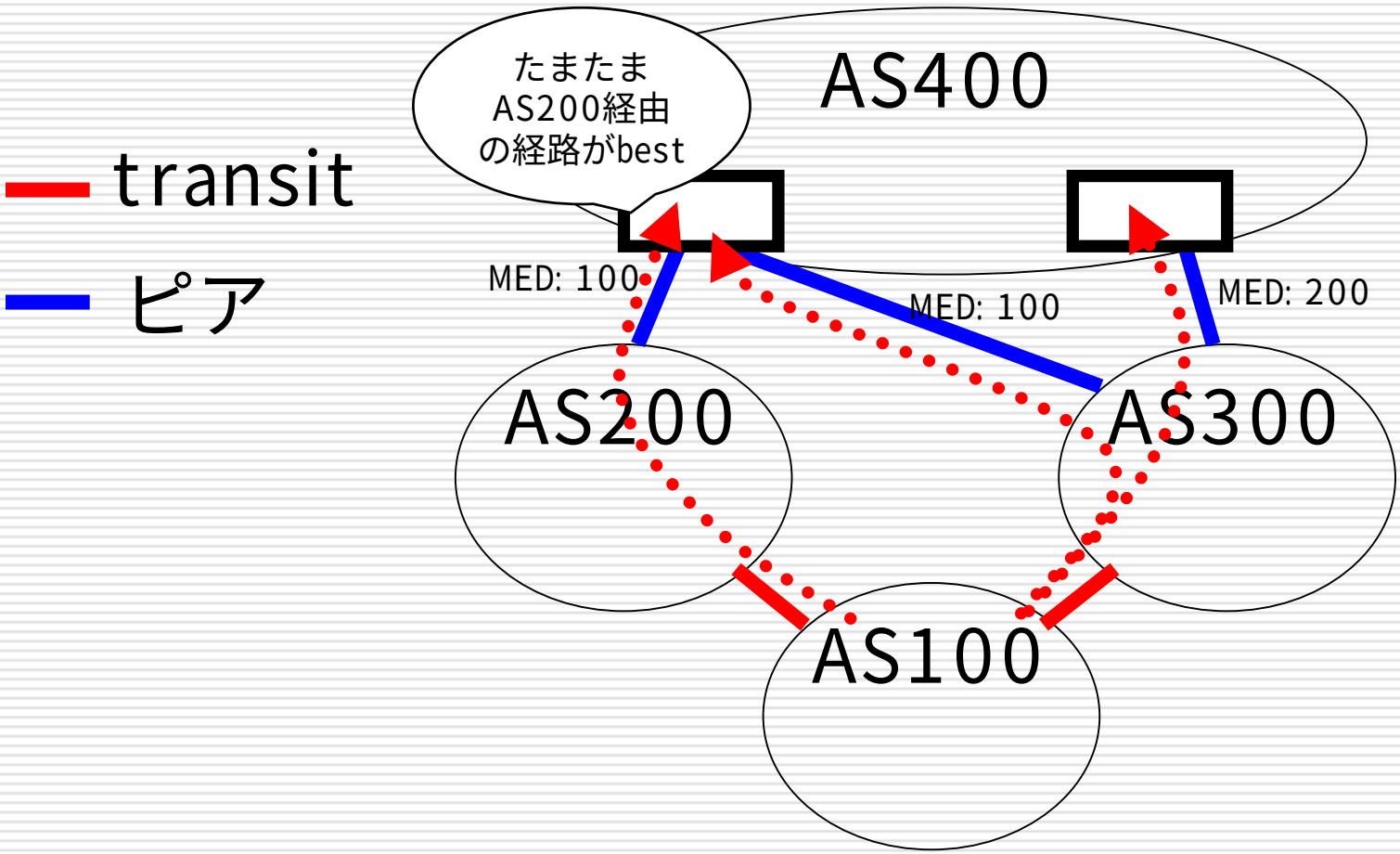
「ＭＥＤ」が無視される？（続き）

transit

ピア

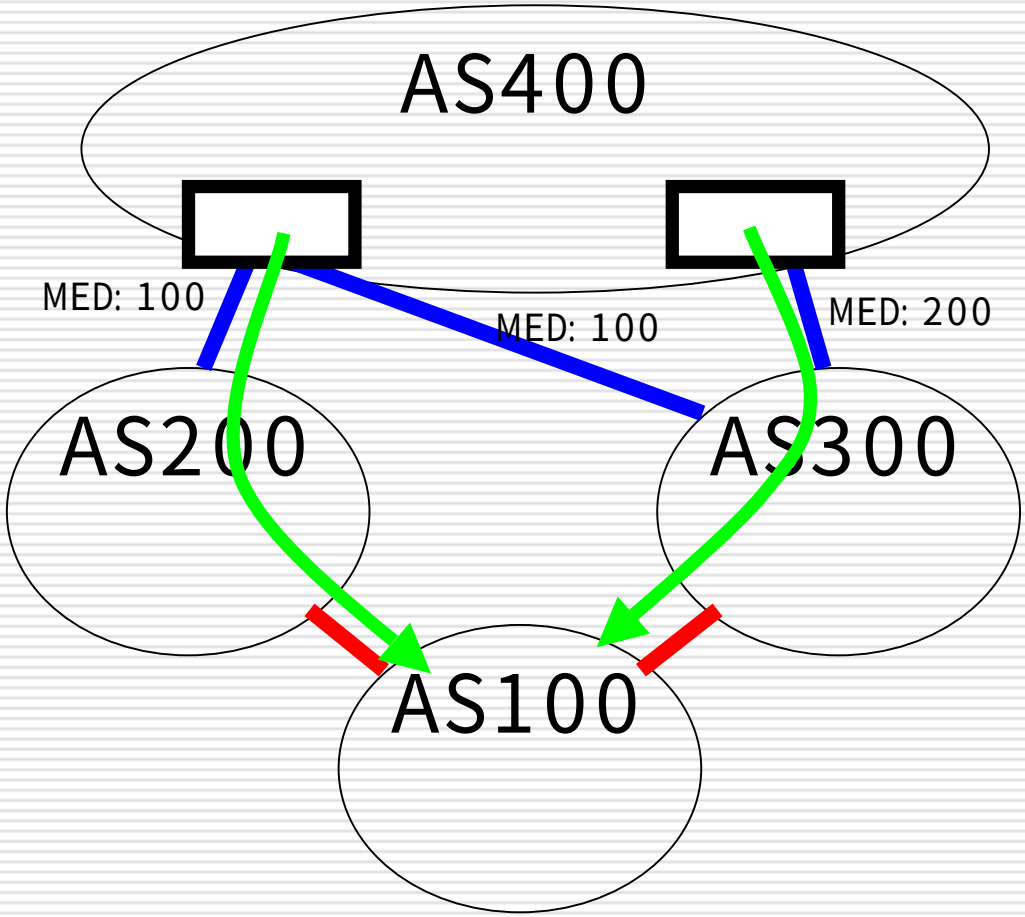


「ＭＥＤ」が無視される？（続き）



「ＭＥＤ」が無視される？（続き）

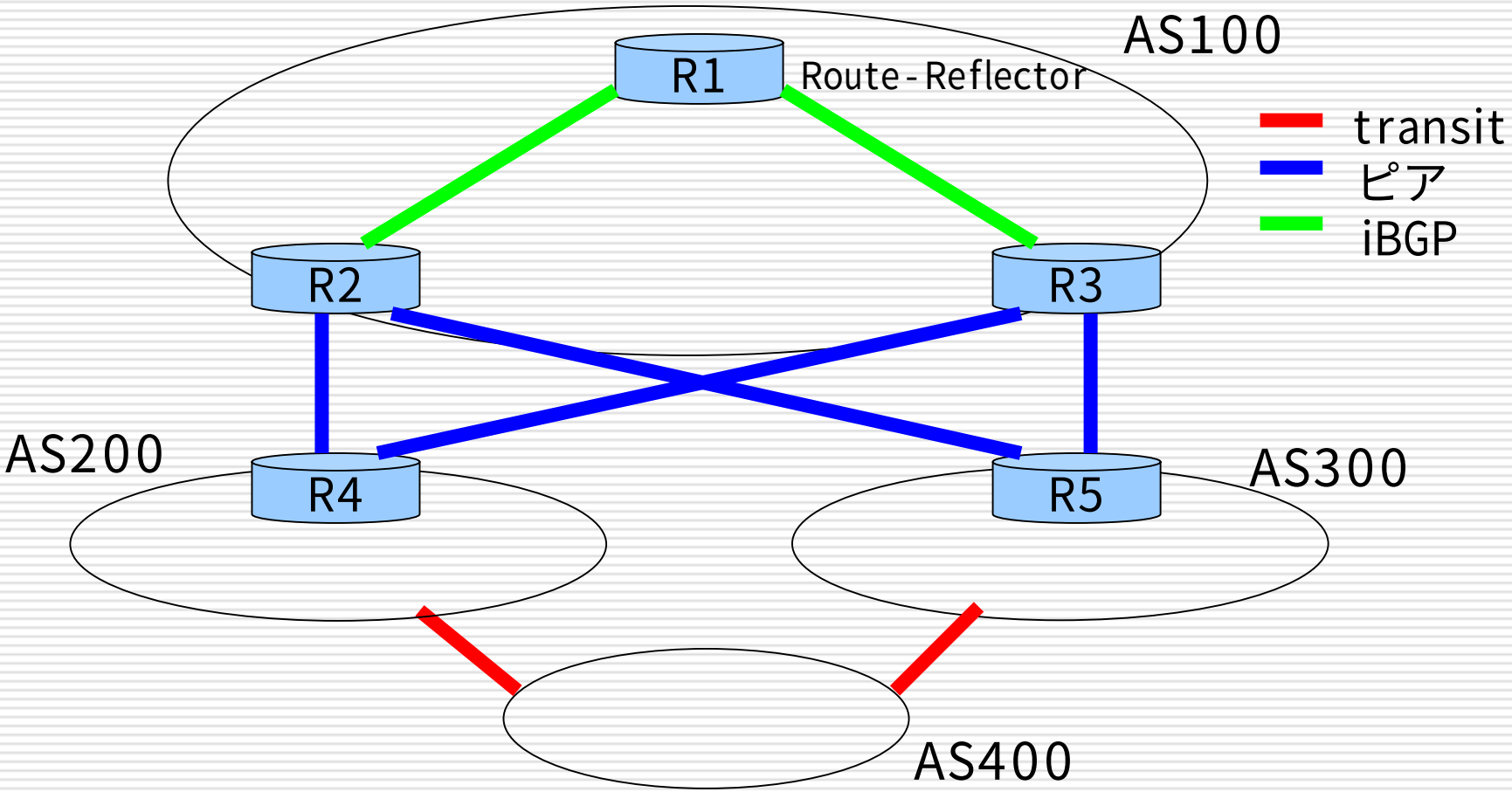
transit
ピア
traffic



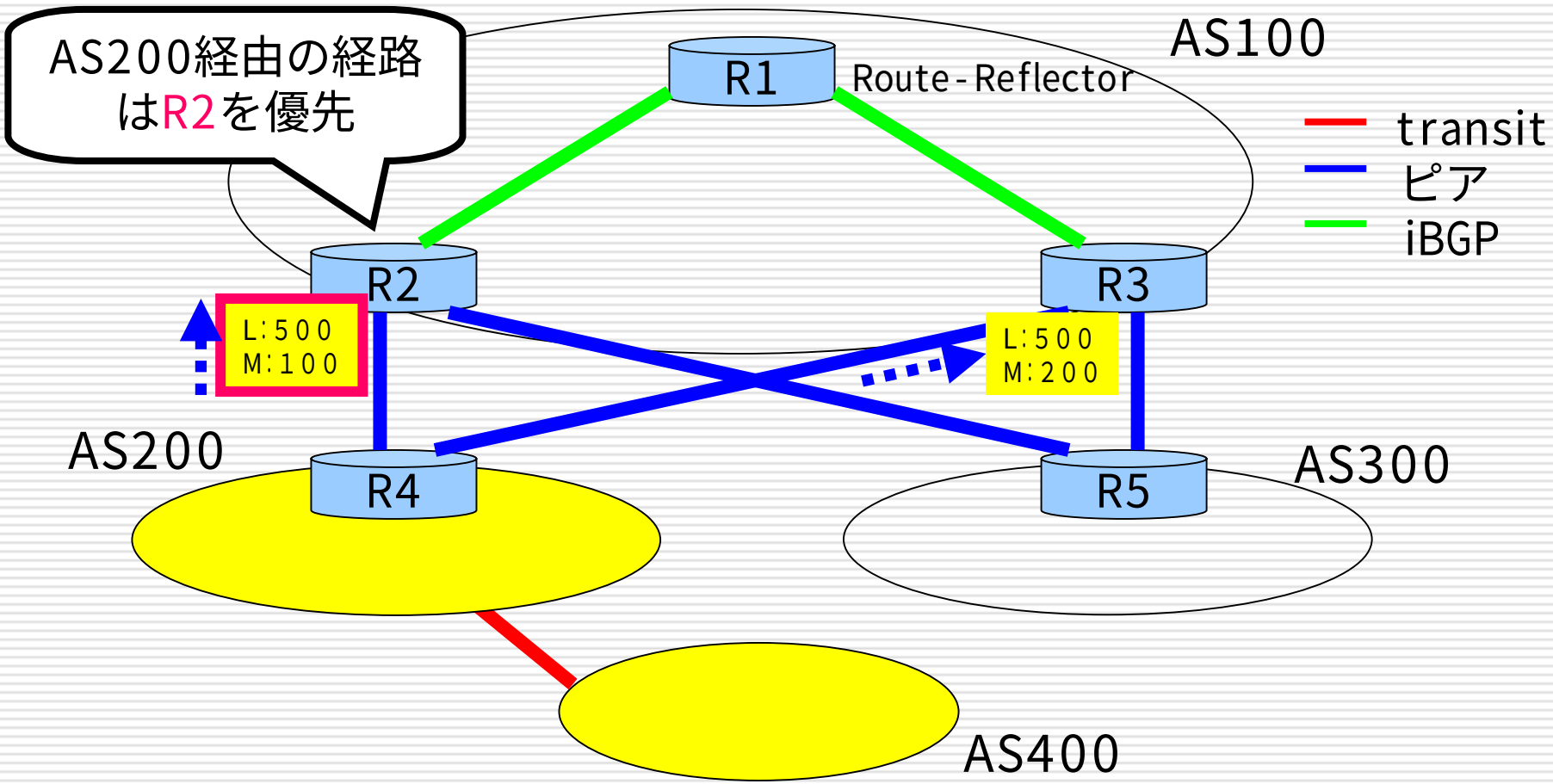
「きもい編」（その２）

- 「ＭＥＤ」がばちばち状態
  - マルチベンダ環境が引き金になることも

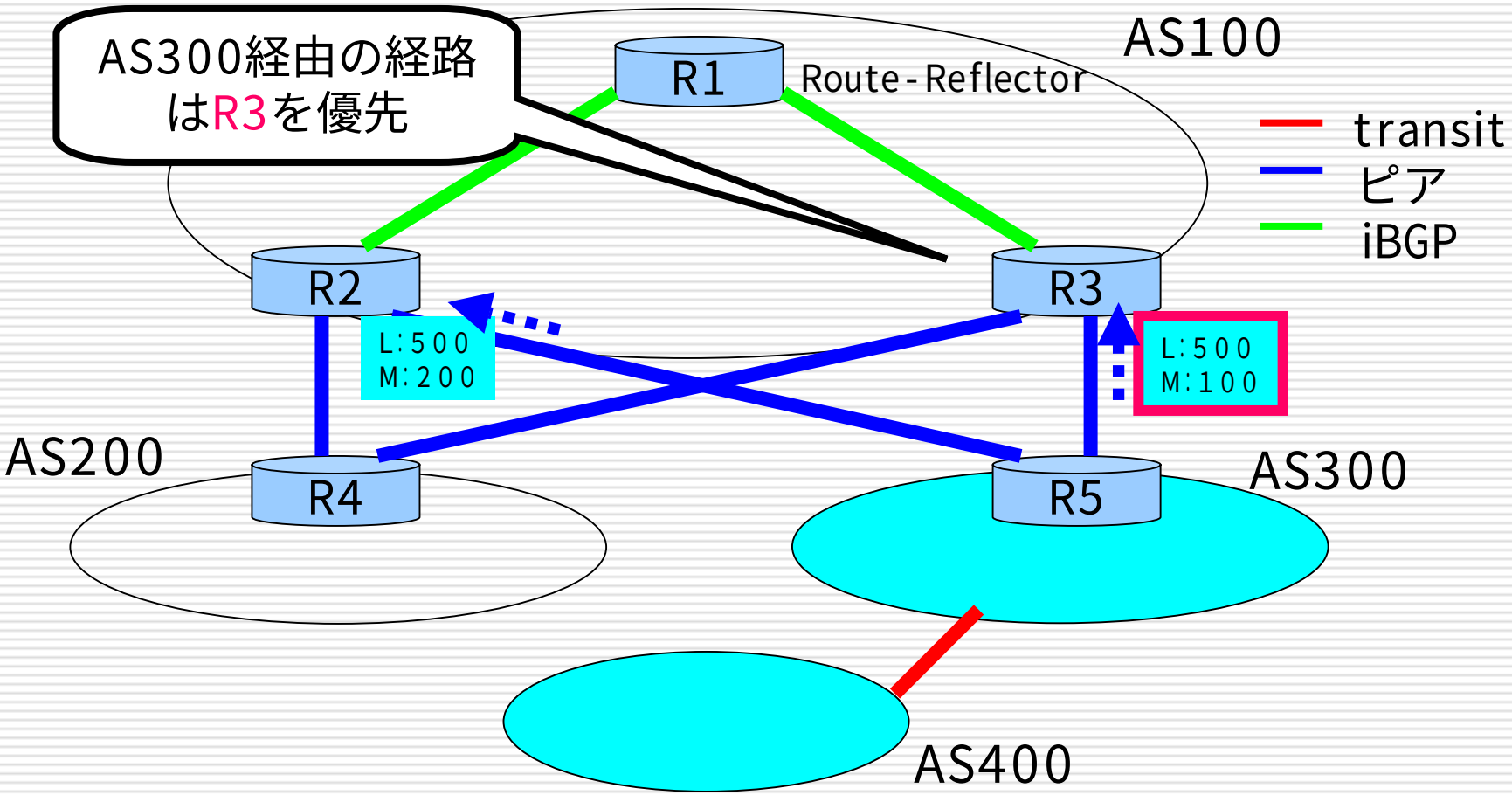
「ＭＥＤ」がばちばち状態


AS100
R1
Route-Reflector
transit
ピア
iBGP
R2
R3
AS200
R4
R5
AS300
AS400

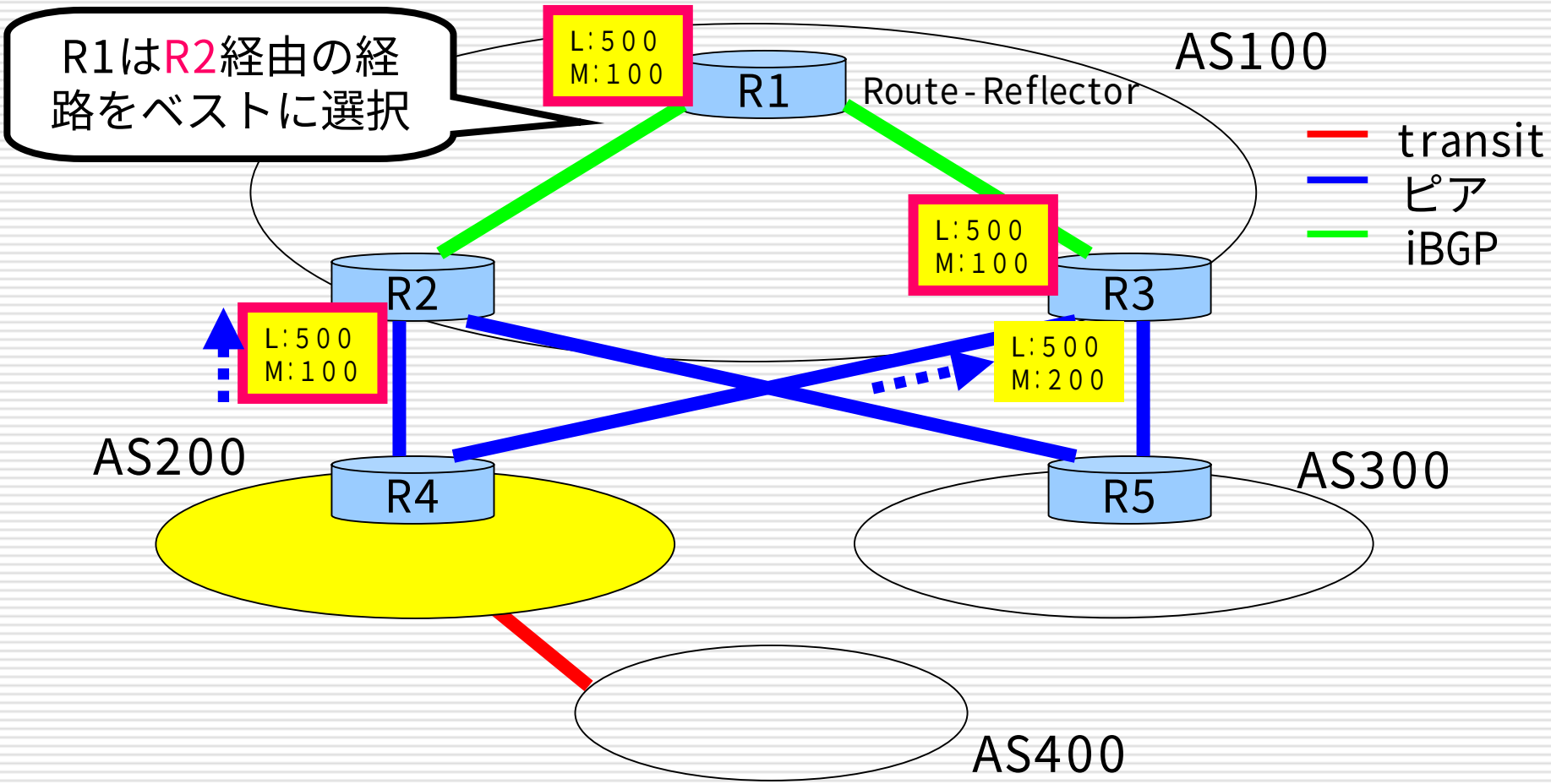
「ＭＥＤ」がばちばち状態（続き）



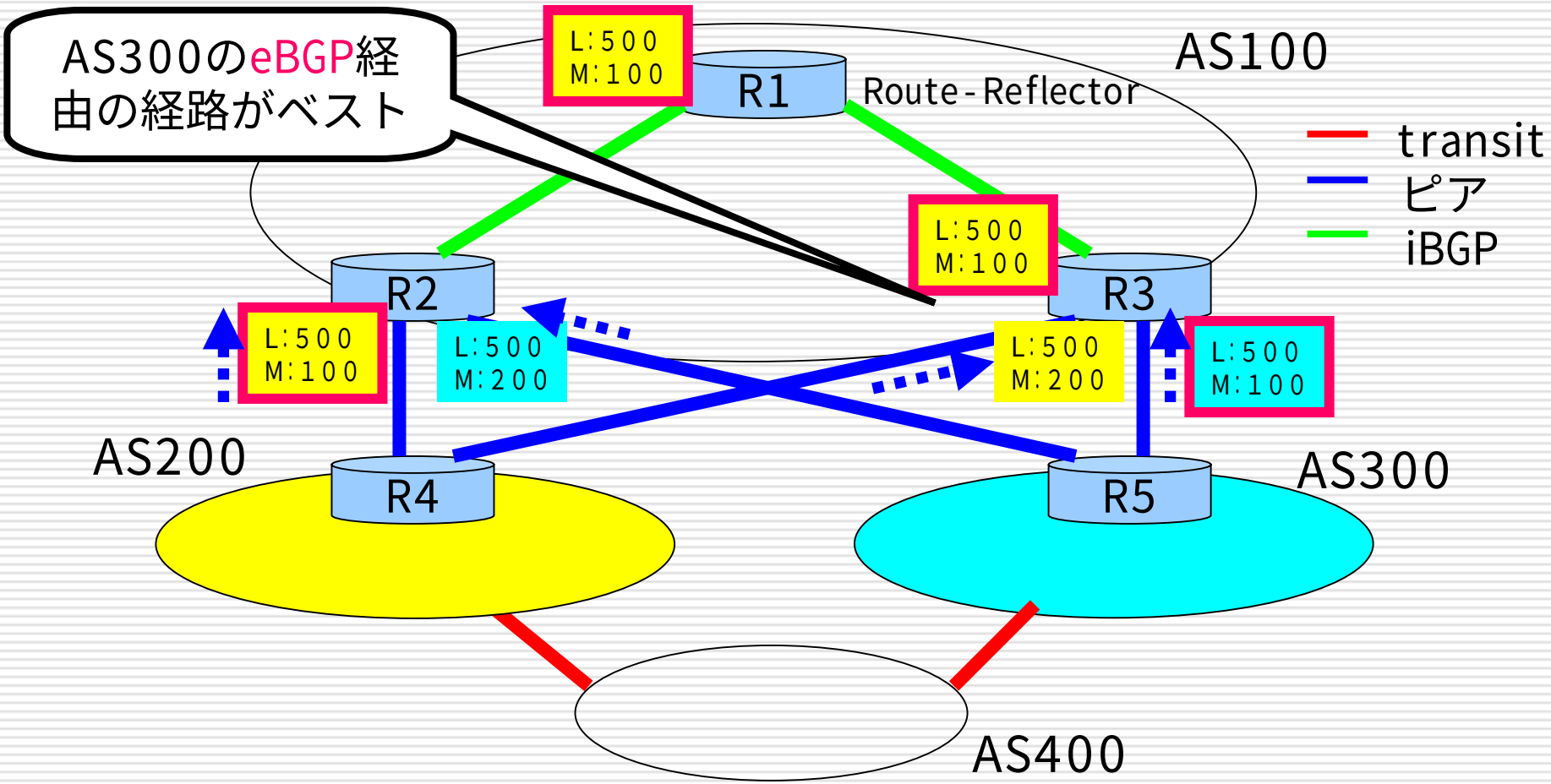
「ＭＥＤ」がばちばち状態（続き）



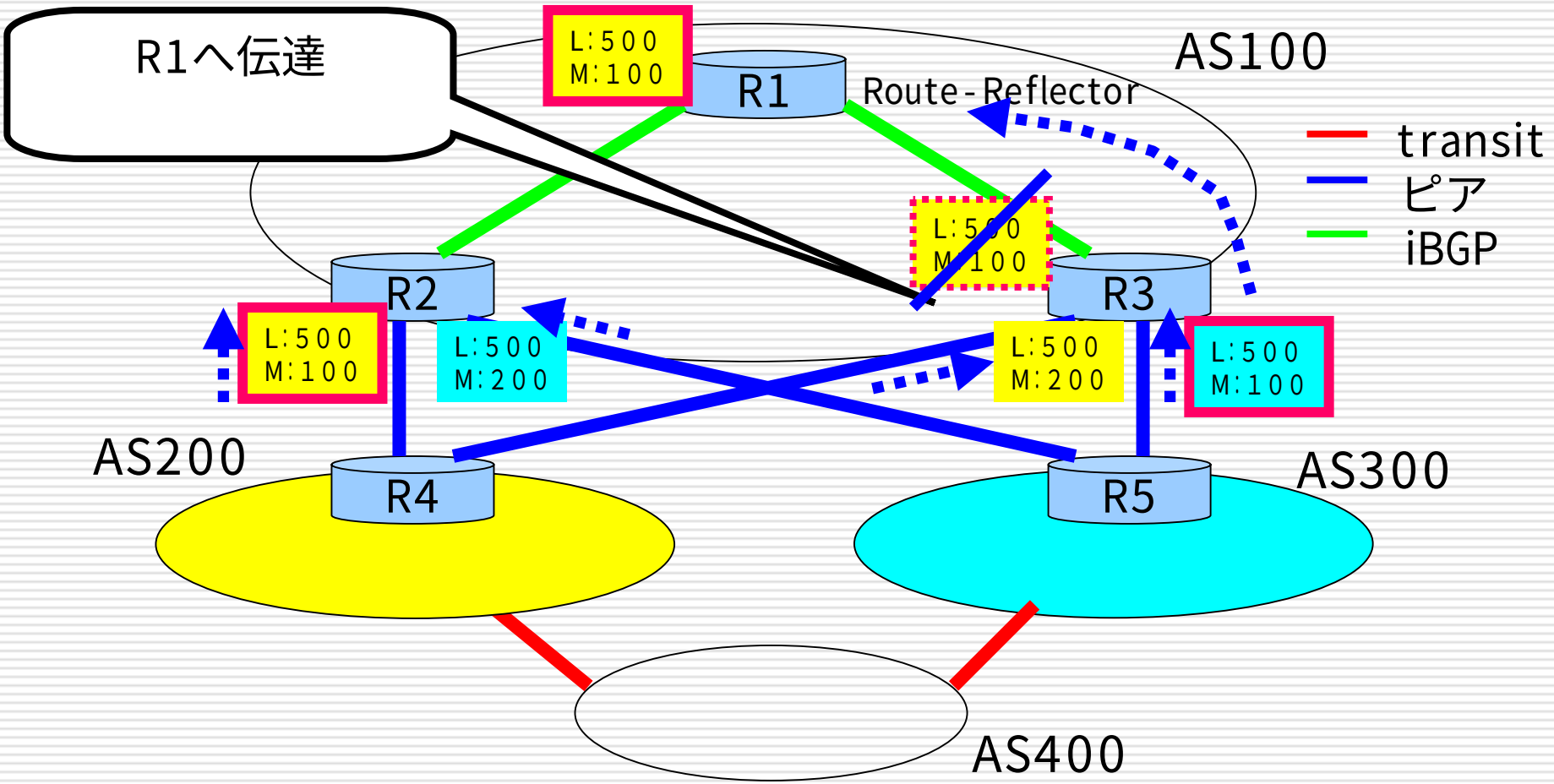
「ＭＥＤ」がばちばち状態（続き）



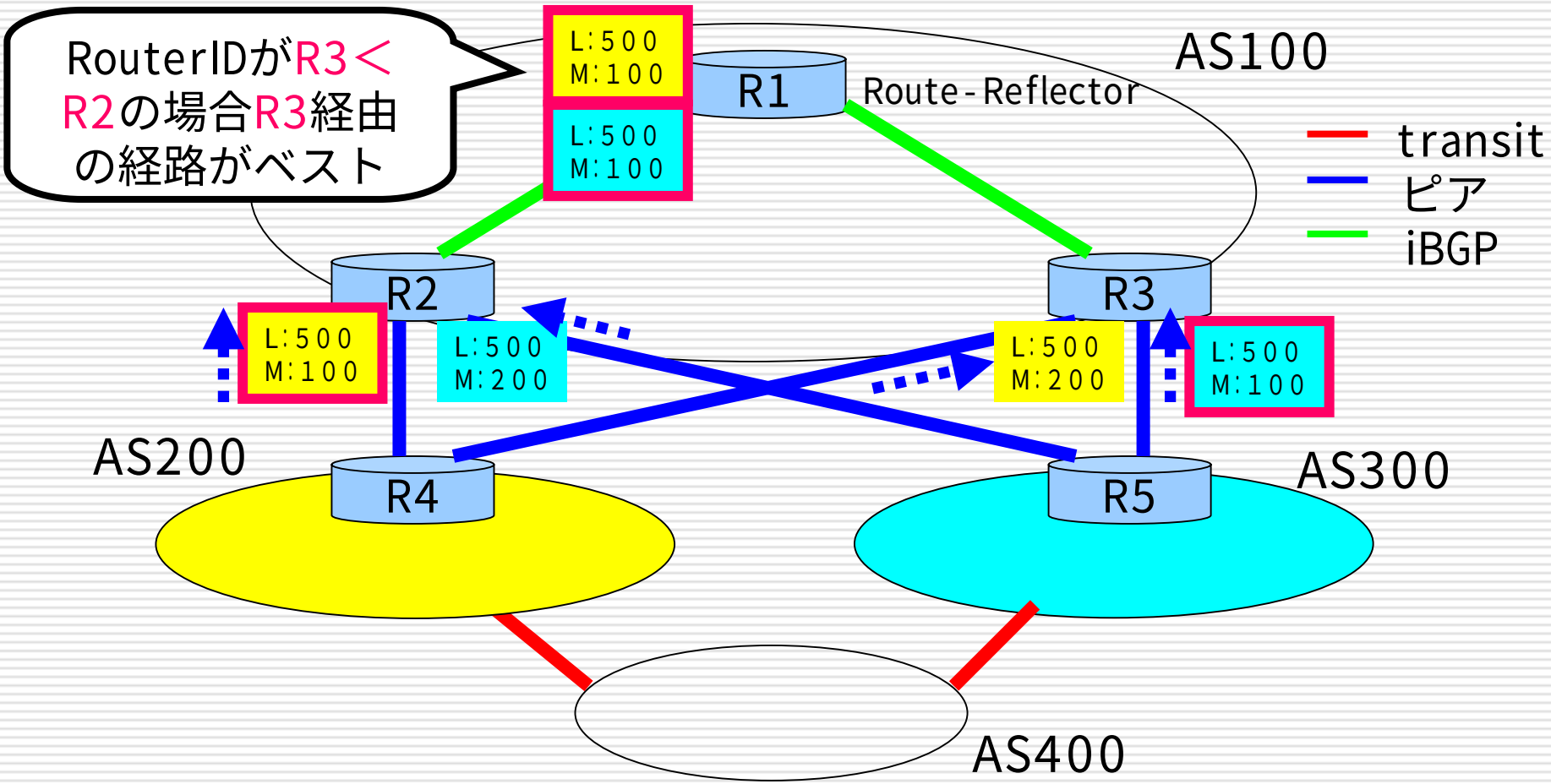
「ＭＥＤ」がばちばち状態（続き）



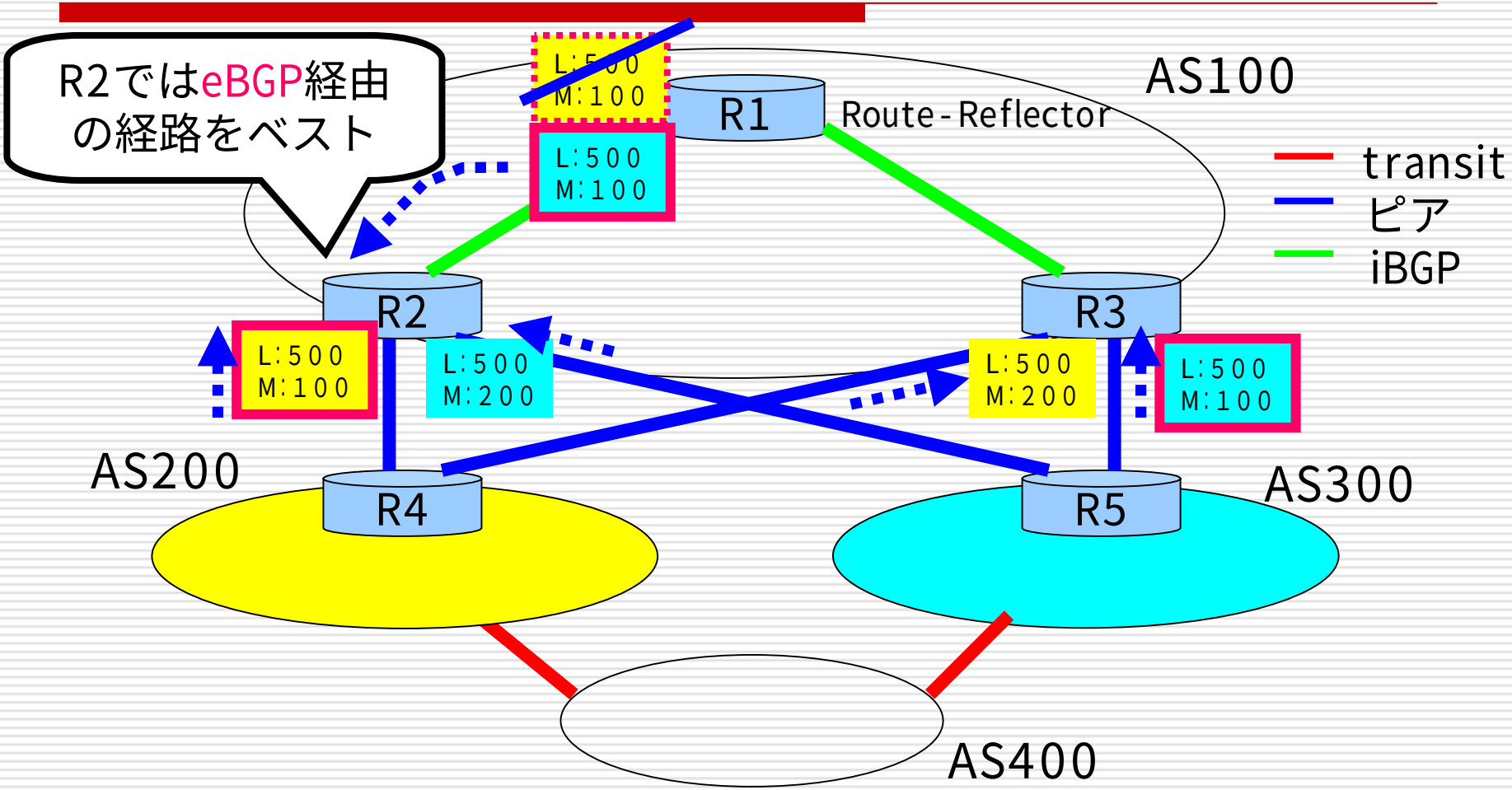
「ＭＥＤ」がばちばち状態（続き）



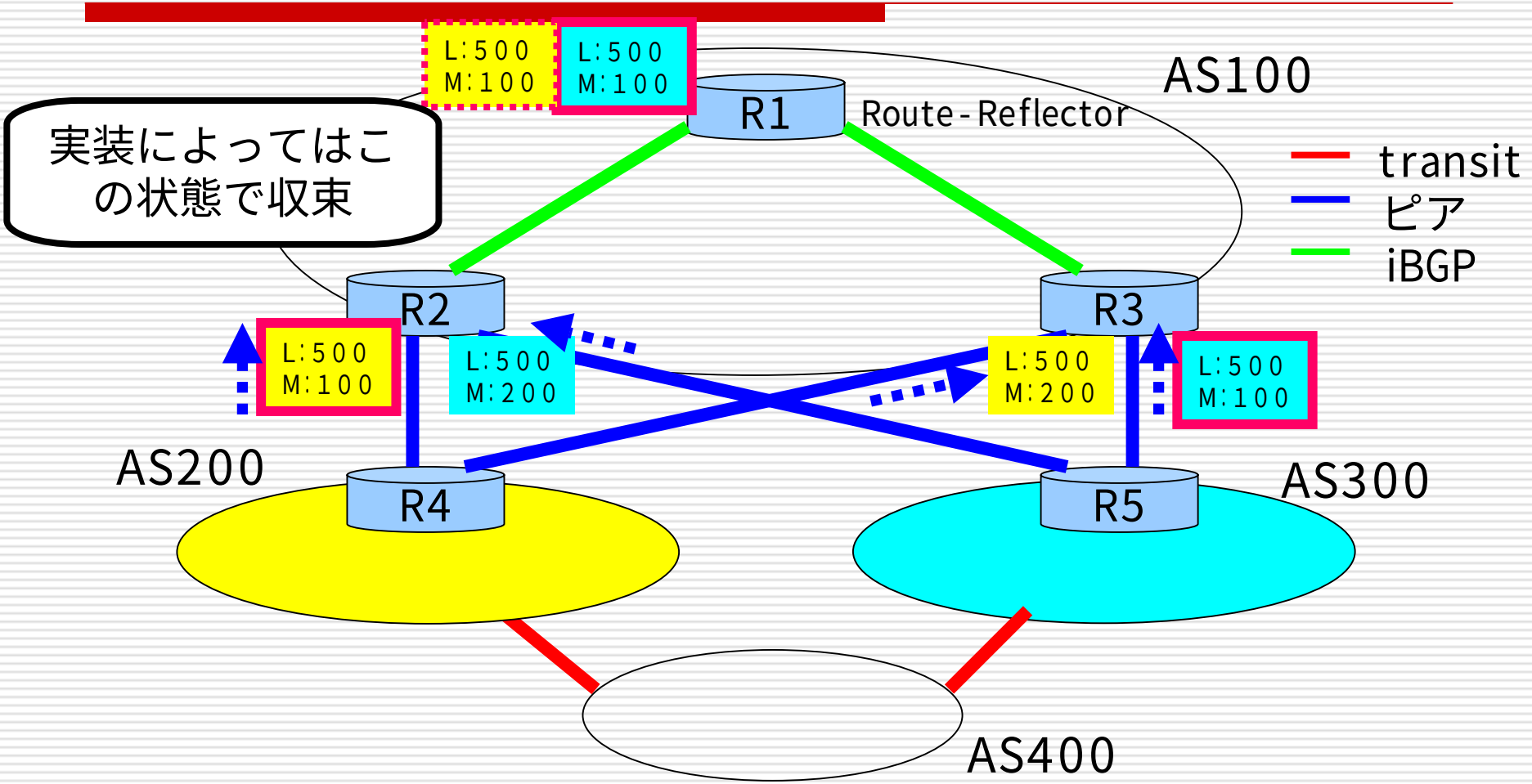
「ＭＥＤ」がばちばち状態（続き）



「ＭＥＤ」がばちばち状態（続き）



「ＭＥＤ」がばちばち状態（続き）

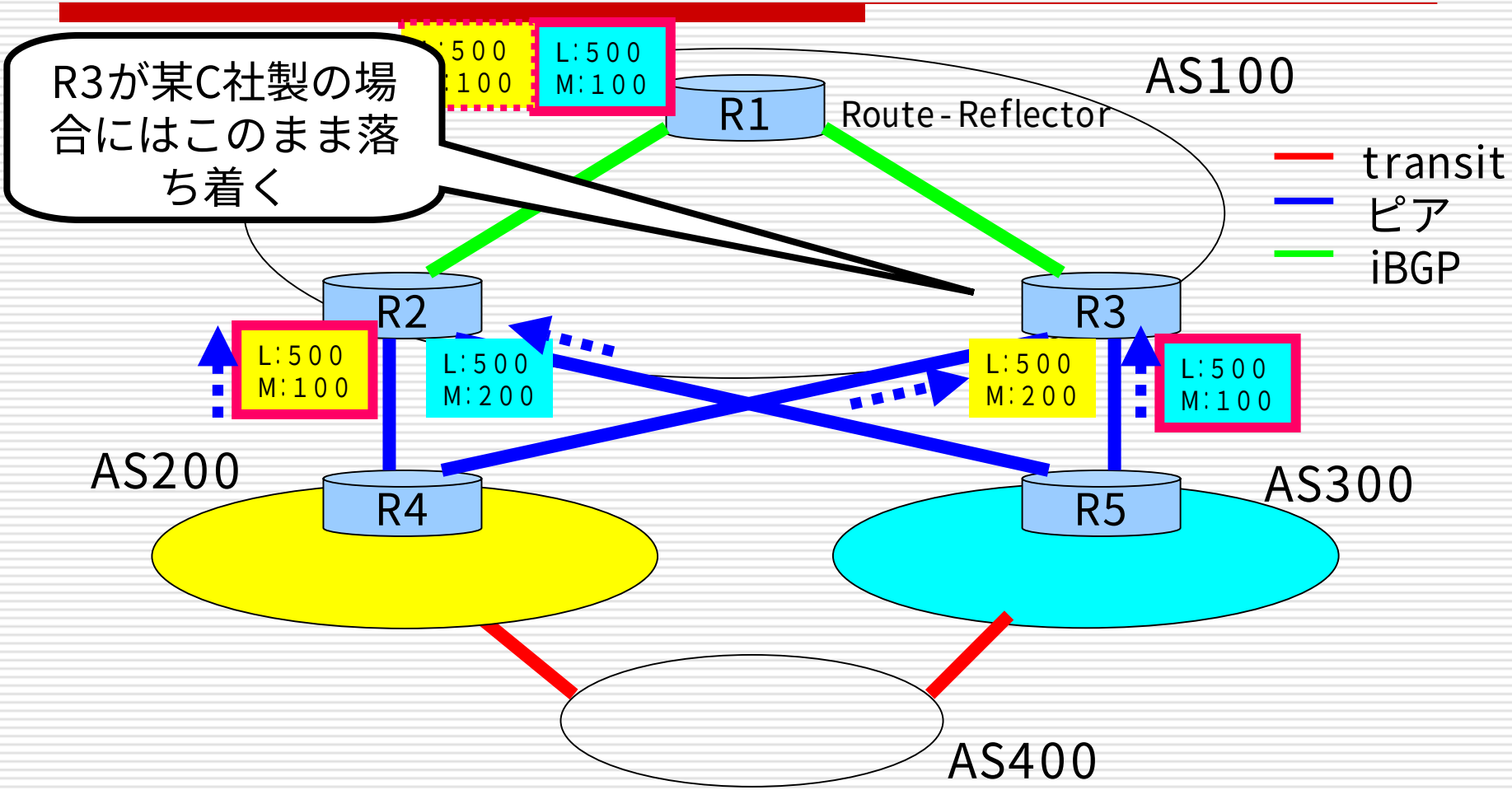


「ＭＥＤ」がばちばち状態（続き）

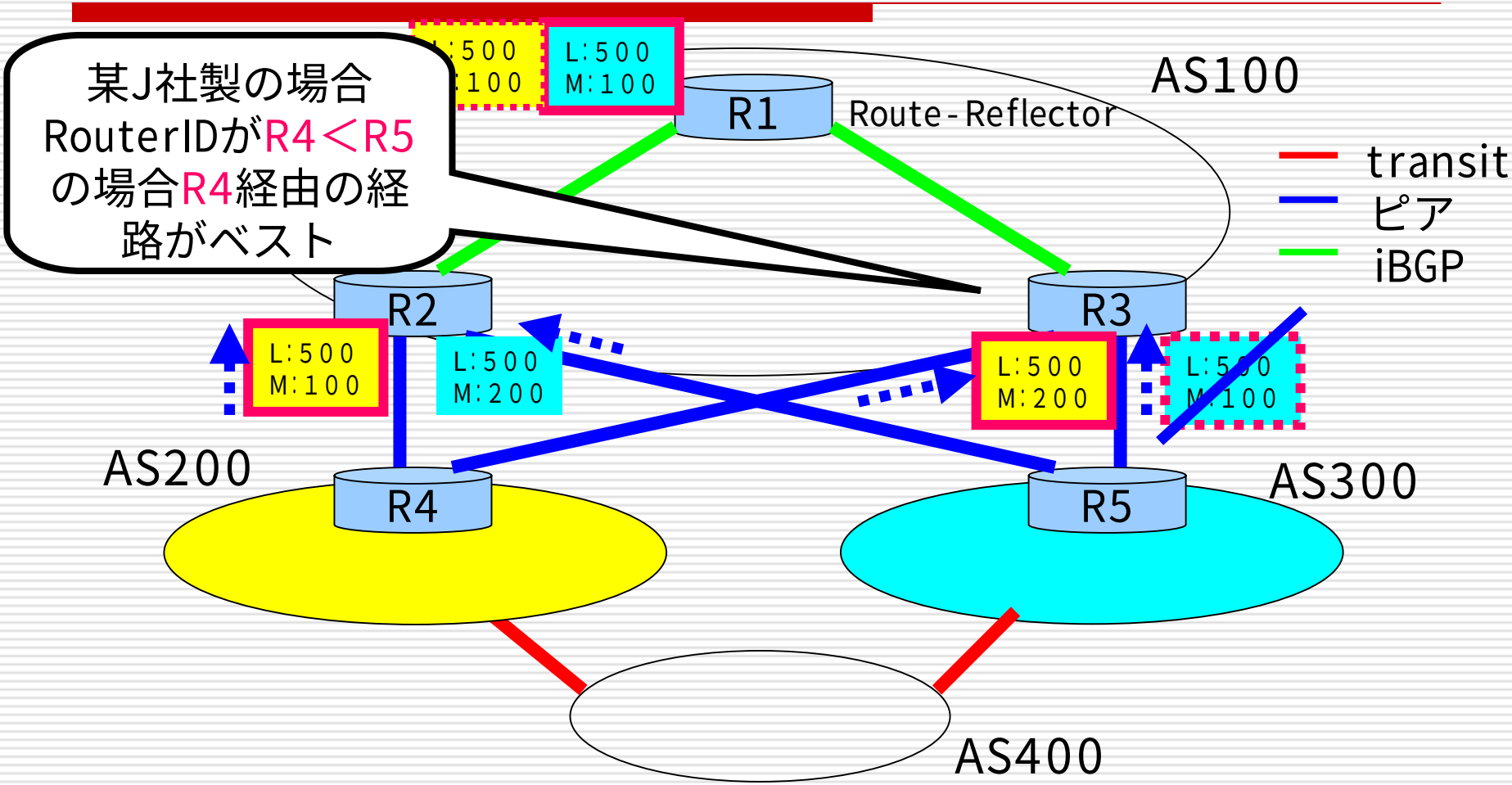
- 某C社
  - eBGPに対して，Router-IDの比較を　しない
  - ルートフラップを考慮した実装のようだ
    - 2つのeBGPピアから経路を受信している場合，同一AS_PATH長だし，安定して常に広告されている方（先に広告してきてそのままになっている）を優先的に常に選択していたほうが望ましいだろう
- 某J社
  - eBGPに対して，Router-IDの比較を　する

実装のちがい

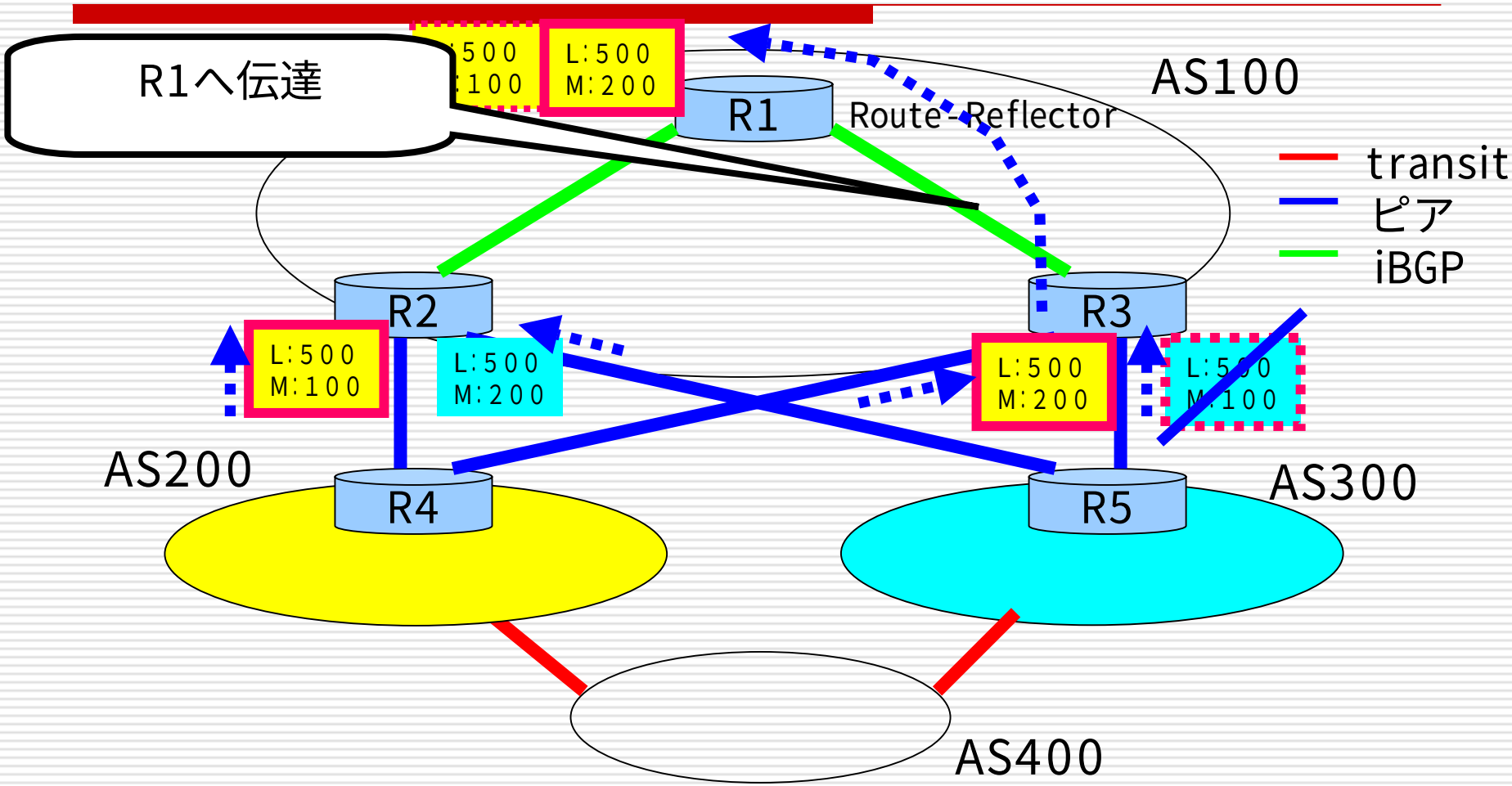
「ＭＥＤ」がばちばち状態（続き）



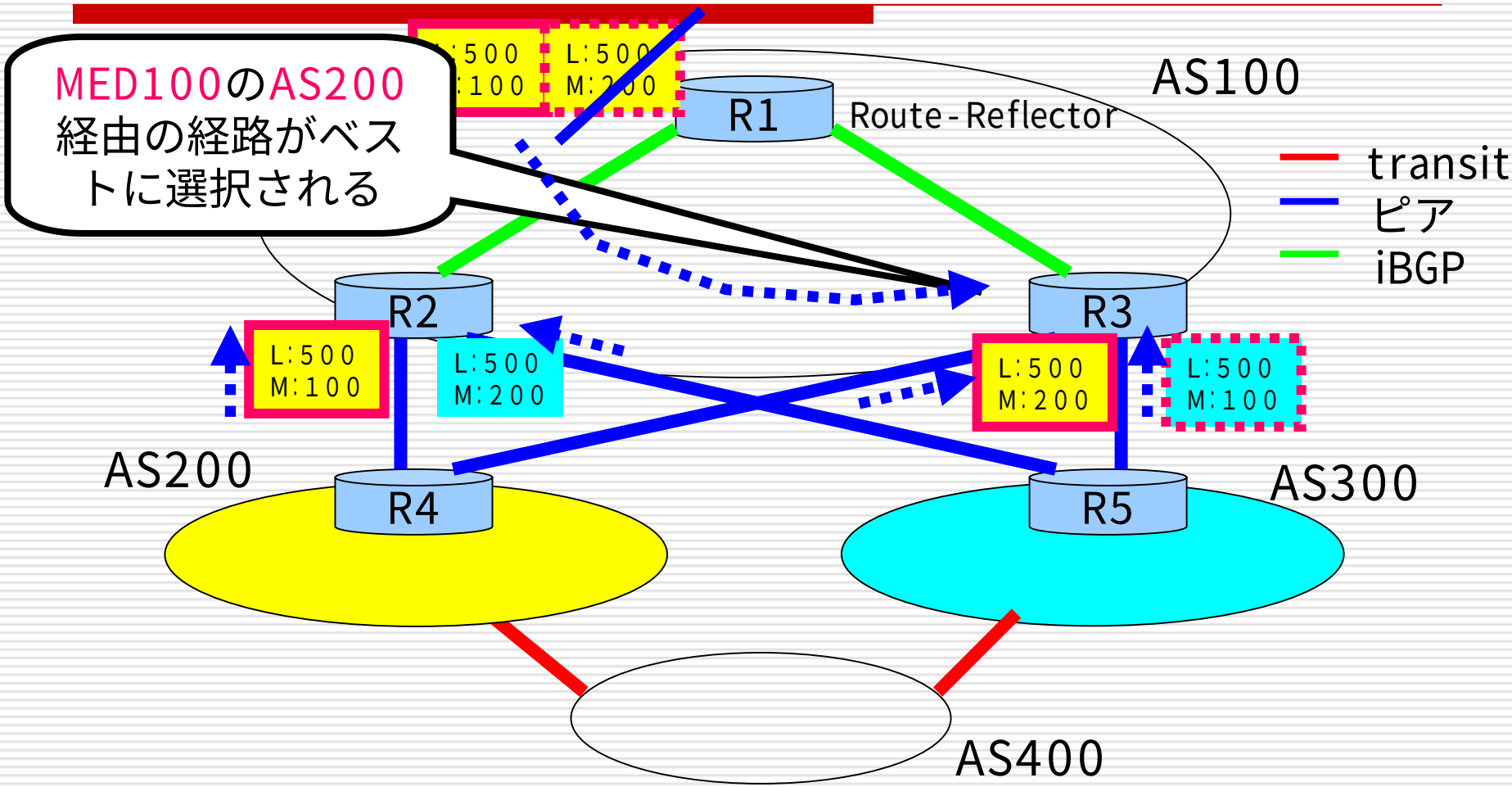
「ＭＥＤ」がばちばち状態（続き）



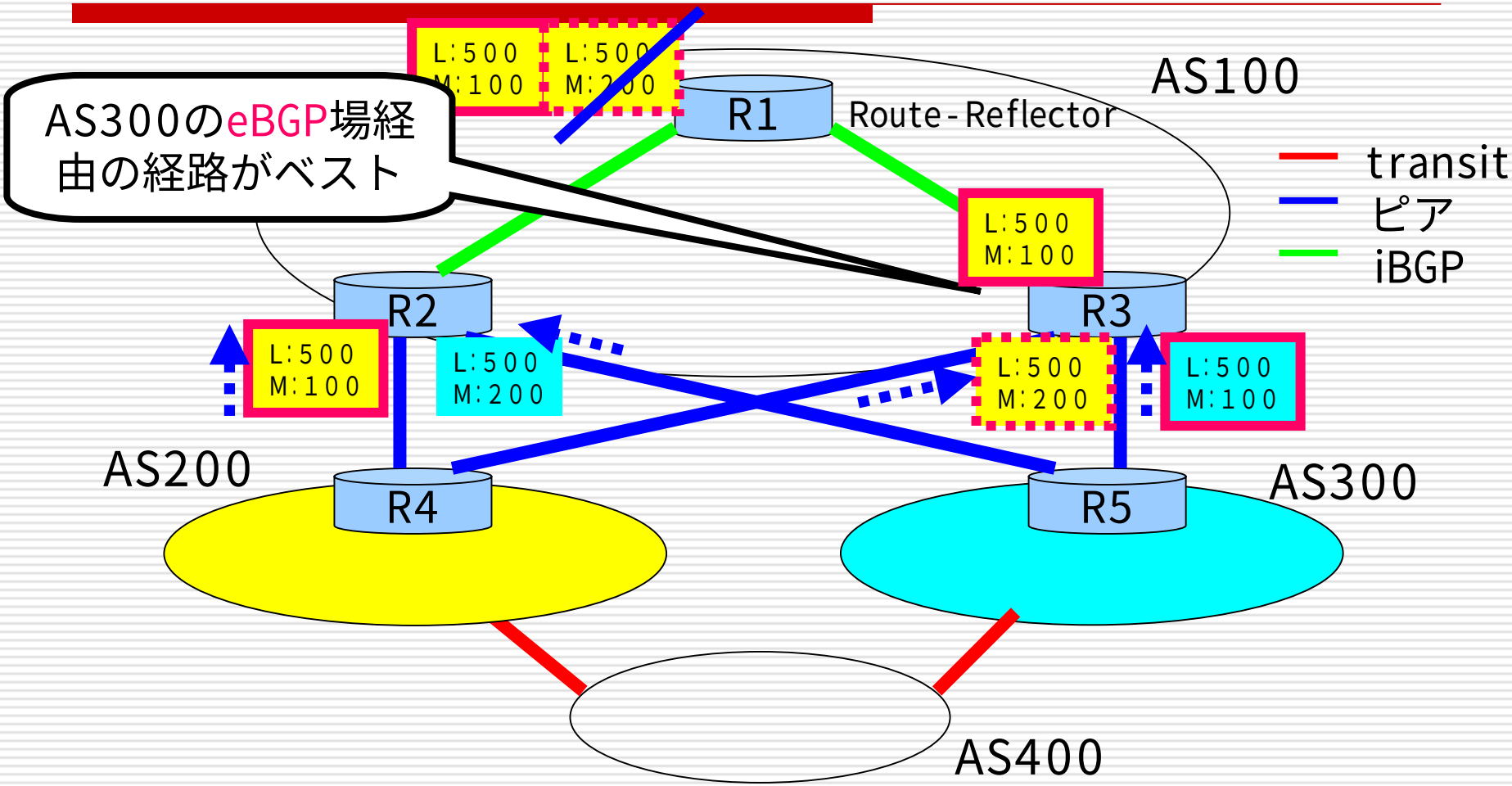
「ＭＥＤ」がばちばち状態（続き）



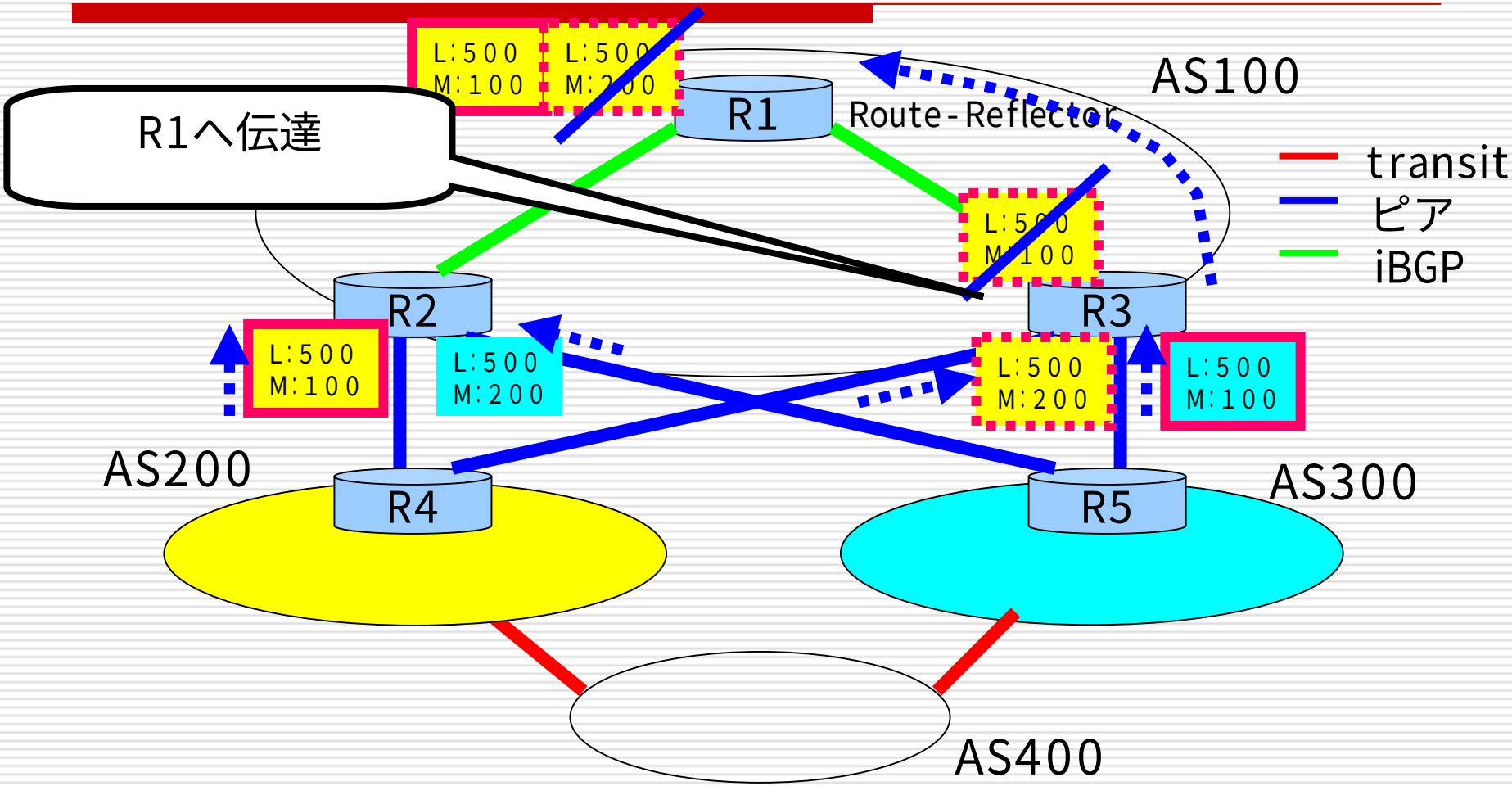
「ＭＥＤ」がばちばち状態（続き）



「ＭＥＤ」がばちばち状態（続き）



「ＭＥＤ」がばちばち状態（続き）

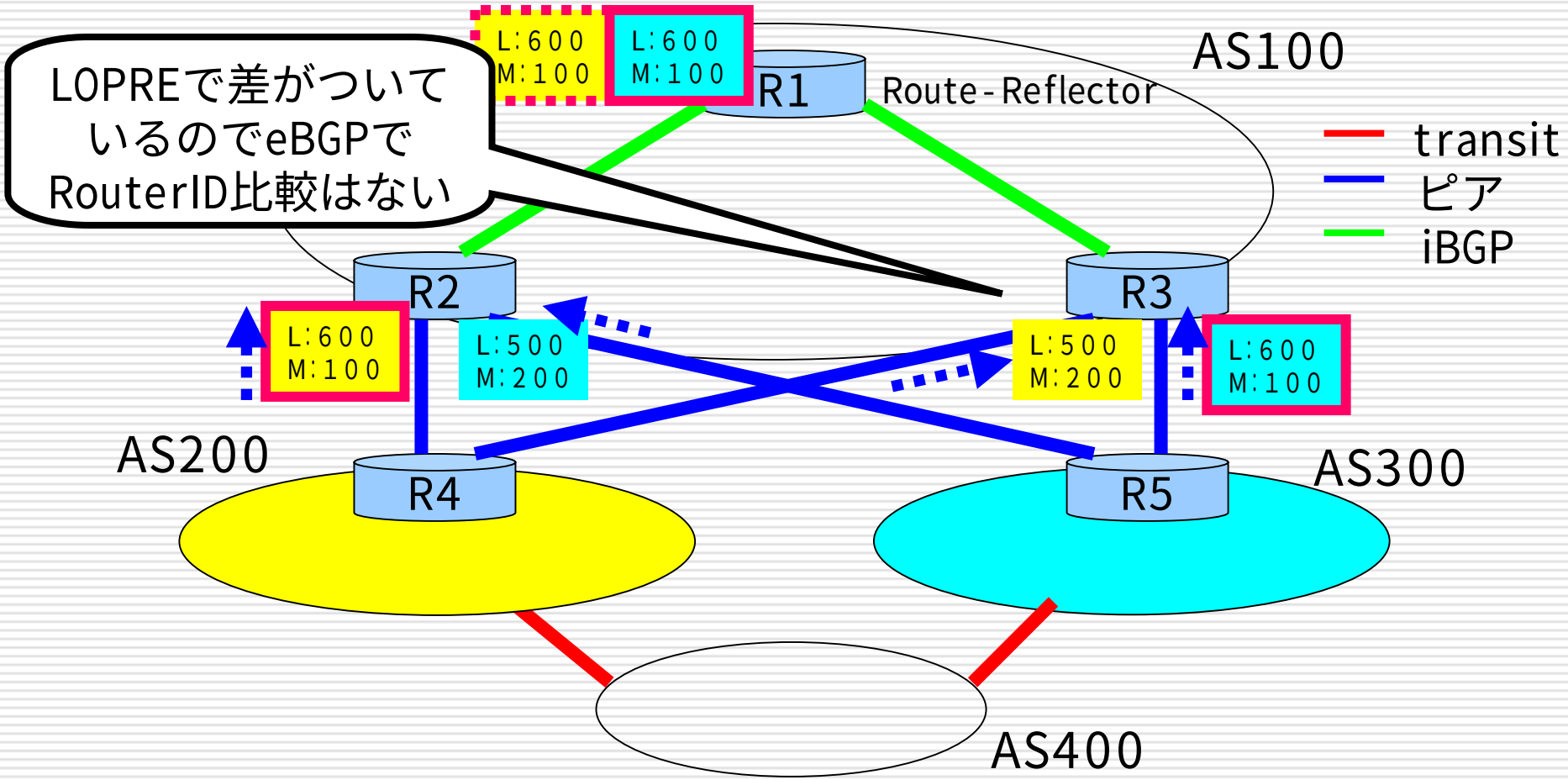


最初に舞い戻りがががが！

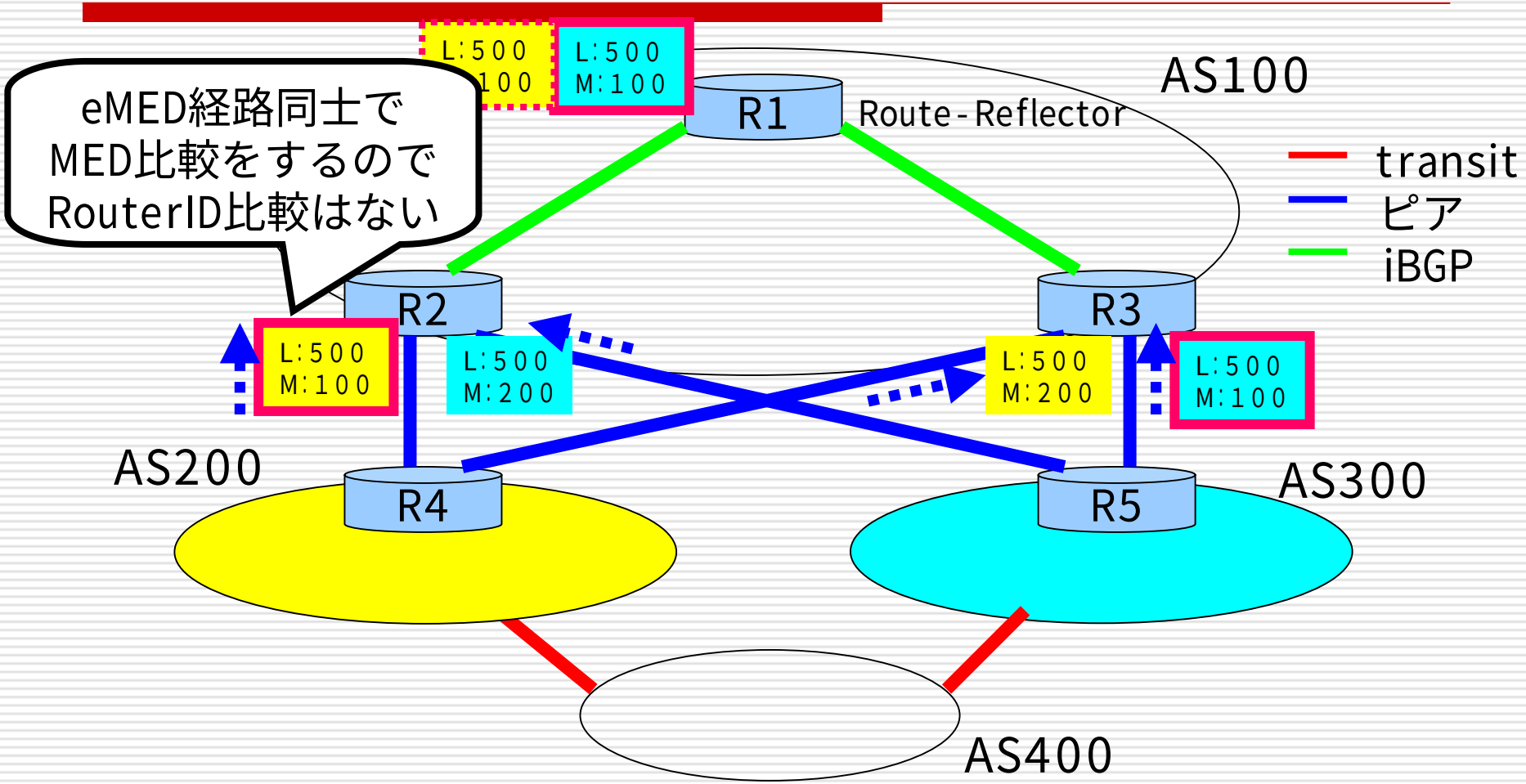
「ＭＥＤ」がばちばち状態（続き）

- 1. LOPRE
- 2. always-compare-med

「ＭＥＤ」がばちばち状態（続き）



「ＭＥＤ」がばちばち状態（続き）



2. always-compare-med

本日のアジェンダ

第一楽章「そりゃないよね編」
第二楽章「きもい編」
第三楽章「こりゃどうだ編」

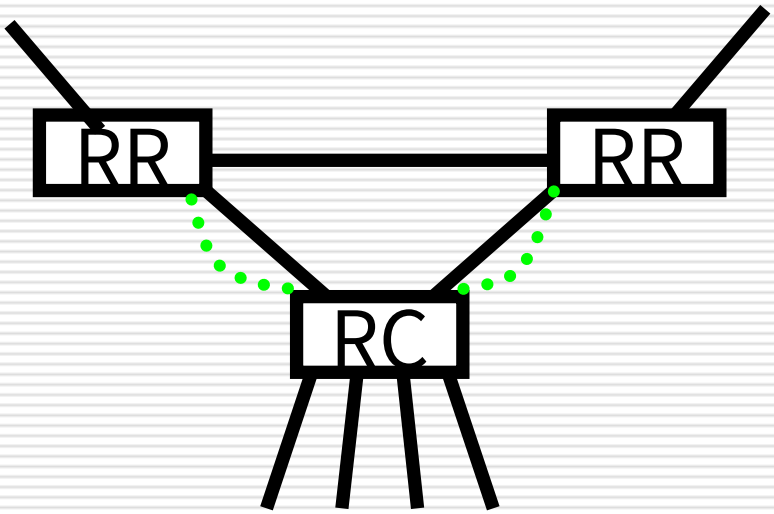
iBGPを実アドレスで張るとどう？

- 教科書通りだと、Loopbackなど落ちない仮想インタフェースでiBGPを張る
- これを、物理インタフェースのアドレスで張るとどうでしょう

iBGPを実アドレスで張るとどう？（続き）

- 良い点
  - IGPが起動しなくてもBGP sessionは上がる
- 悪い点
  - インタフェースが落ちるとsessionも落ちる
  - たくさん設定するには、少しめんどう

iBGPを実アドレスで張るとどう？（続き）



- RCの再起動時に経路収束時間が短くなるかも
  - 直結しているRRとRC間
  - iBGPの収束時間＞IGPの収束時間
- 設定は自動生成させないとつらそう
  - 手でやっても良いけど、考えるのがめんどくさい

便利なsoft reconfiguration

- Input Policyの再適用などに使われてます
- 最近はRoute refresh capabilityがサポートされている場合が多いです

```
>show ip bgp neighbors
  :
Neighbor capabilities:
   Route refresh: advertised and received(new)
```

```
> show bgp neighbor
  :
   Peer supports Refresh capability (2)
```

new：rfc　= 2
old： cisco = 128

便利なsoft reconfiguration（続き）

- soft reconfigurationだと、Input Policy適用前の経路が見られます
  - show ip bgp neighbors X.X.X.X received-route
  - show route receive-protocol bgp X.X.X.X

便利なsoft reconfiguration（続き）

- 例えばフィルタとの差分を確認するとき
- 例えばピアを上げるとき
  - Prefix Filterで deny 0.0.0.0/0 le /32
  - soft reconfiguration を有効にしてピア上げ
  - 経路確認後、普通のPrefix Filterを適用
  - Input Policy再適用

アジェンダ追加

第一楽章「そりゃないよね編」
第二楽章「きもい編」
第三楽章「こりゃどうだ編」
勝手にアンコールしながら議論「つぶやき編」

今の AS-PATH Filter はそろそろ止めよう！

- 昔は顧客経路の識別で始まったらしい
- しかし、国内のＡＳ数もピア数も増えてきて、そろそろ更新作業が大変
- 例えば一言コメントでもAS Path Filterの更新にまつわるコメントがちらほら・・・
  - 立て続けに更新のメールを送ってくるのは勘弁してほしい

今のAS-PATH Filterはそろそろ止めよう！（続き）

- IRR を使いましょう！
  - 相手に as-set オブジェクト を 事前に連絡しておいて，あとは勝手にそれを見てね，というオペレーション

```
as-set:     AS-IIJ
descr:      ASes routed by IIJ
members:    AS112, AS2497, AS2504, AS2508, AS2515,
            AS2523, AS2526, AS2527, AS4459, AS4672,
            AS4685, AS4688, AS4695, AS4718, AS4723,
            AS4777, AS4996, AS6303, AS7500, AS7502,
            AS7511, AS7516, AS7517, AS7519, AS7521,
            AS7522, AS7524, AS7529, AS7531, AS7664,
            AS7668, AS7670, AS7671, AS7672, AS7679,
            AS7682, AS7684, AS7685, AS7686, AS7687,
                    :
```

```
as-set:     AS-OCN
descr:      ASes advertised by OCN
members:    AS4713,
            AS290,  AS2504,  AS2526,  AS4249,  AS4688,
            AS4710,  AS4711,  AS4718,  AS7502,  AS7511,
            AS7521,  AS7522,  AS7524,  AS7529,  AS7668,
            AS7671,  AS7672,  AS7674,  AS7676,  AS7682,
            AS7684,  AS7686,  AS9351,  AS9353,  AS9363,
            AS9368,  AS9370,  AS9374,  AS9601,  AS9602,
            AS9605,  AS9612,  AS9614,  AS9617,  AS9618,
            AS9621,  AS9622,  AS9824,  AS9827,  AS9991,
                    :
```

- 我々も幸せになろうよ！

今のAS-PATH Filterはそろそろ止めよう！（続き）

- 現在(メールで連絡、手動更新)
  - ^(100_)+$
  - ^(100_)+(200_)+$
- 検討中(as-setから自動生成)
  - as-pathでのfilter
    - _100$
    - _200$
  - prefixでのfilter
    - 10.100.0.0/16
    - 10.200.0.0/16

最近のBGP ○○編

おしまい

質問、ご意見よろしく