



ISPバックボーンネットワーク における経路制御設計 ～ 理論編 ～

Matsuzaki Yoshinobu
<maz@iij.ad.jp>

Copyright (C) 2005 Internet Initiative Japan Inc.

1

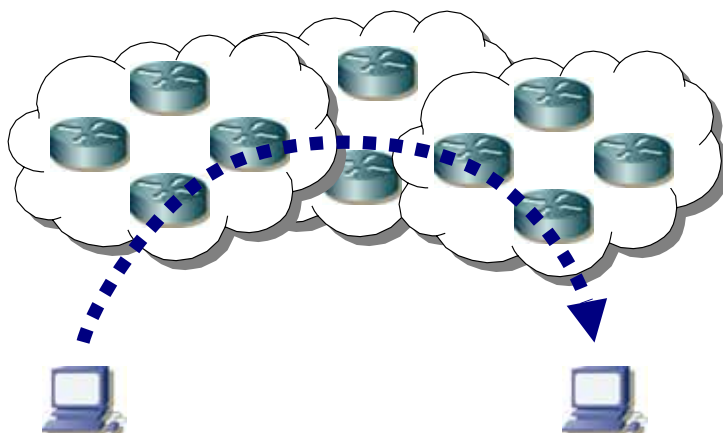
ルーティング整理

ルーティングの基礎事項を整理する

Copyright (C) 2005 Internet Initiative Japan Inc.

2

ルーティング



Copyright (C) 2005 Internet Initiative Japan Inc.

3

ネットワークのプレフィックス表記

- 192.168.0.0/24
 - = 192.168.0.0 ~ 192.168.0.255
 - = 192.168.0.0 mask 255.255.255.0
- 連続ネットマスクが前提
 - 非連続ネットマスクは表現できない
 - 192.168.0.10 mask 255.255.0.255
 - でも、昨今こんなの使わない
 - 複数行での表記になる場合
 - 192.168.0.0 ~ 192.168.2.255
 - 192.168.0.0/23, 192.168.2.0/24

Copyright (C) 2005 Internet Initiative Japan Inc.

4

クラスレス(Classless)

- クラスの概念は過去の遺物なので忘れよう
- 昔はネットワークアドレスの認識に利用
 - IPアドレスを見れば、ネットマスクが分かった
 - RIPなどで利用
 - 最近ではプロトコルでプレフィックス長を伝播する
 - 今やクラスレスが標準

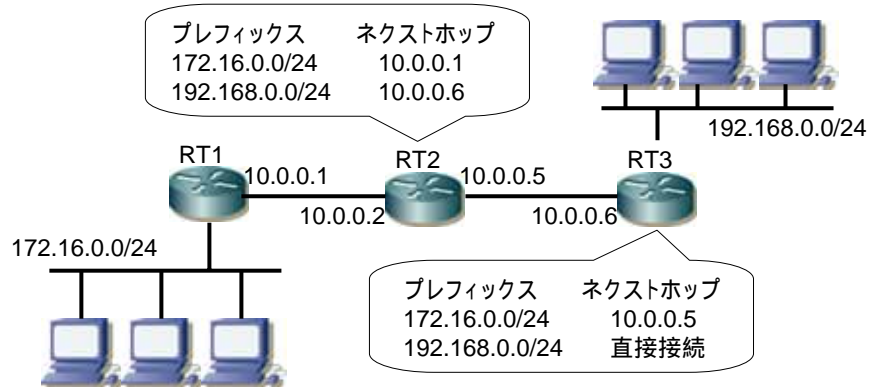
クラスA	0.0.0.0 ~ 127.255.255.255	/8
クラスB	128.0.0.0 ~ 191.255.255.255	/16
クラスC	192.0.0.0 ~ 223.255.255.255	/24

ルーティングとは

- どこを経由して、パケットを宛先に届けるか
- ルータは、パケットの宛先アドレスをみて次の送り先を判断する

経路情報

- 宛先プレフィックス + ネクストホップ の集合

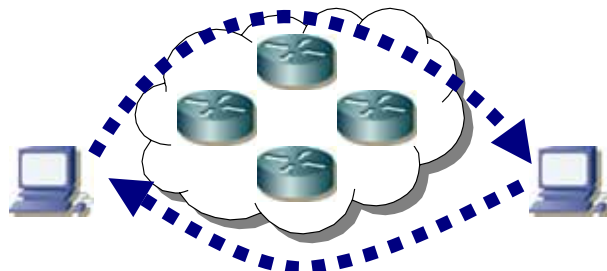


Copyright (C) 2005 Internet Initiative Japan Inc.

7

パケットと経路

- 送信元から宛先まで経路に矛盾が無ければ、パケットが届く
- 双方向で問題が無ければ、相互に通信できる
 - 行きと帰りの経路は違うかもしれない



Copyright (C) 2005 Internet Initiative Japan Inc.

8

経路の種類

- 静的経路
 - connected経路
 - ルータが直接接続して知っている経路
 - static経路
 - ルータに静的に設定された経路
- 動的経路
 - ルーティングプロトコルで動的に学習した経路
 - OSPFやIS-IS、BGPなどで学習した経路

Copyright (C) 2005 Internet Initiative Japan Inc.

9

経路の優先順位

1. prefix長が長い(経路が細かい)ほど優先
 - ホスト経路(/32) \leftrightarrow default経路(0.0.0.0/0)
2. 経路種別で優先
 - connected経路
 - static経路
 - 動的経路(ospf, bgp, etc...)
 - 内訳はベンダ依存

Copyright (C) 2005 Internet Initiative Japan Inc.

10

動的経路制御

インターネットと動的経路制御

Copyright (C) 2005 Internet Initiative Japan Inc.

11

動的経路制御の必要性

- ネットワーク変化を経路情報に反映
 - もちろん事前の設計は必要
- ISPのバックボーン運用では必須
 - インターネットは変化し続けてる
 - プロトコルごとの得手不得手を把握しておく
 - 何を設定しているのか理解しておく

Copyright (C) 2005 Internet Initiative Japan Inc.

12

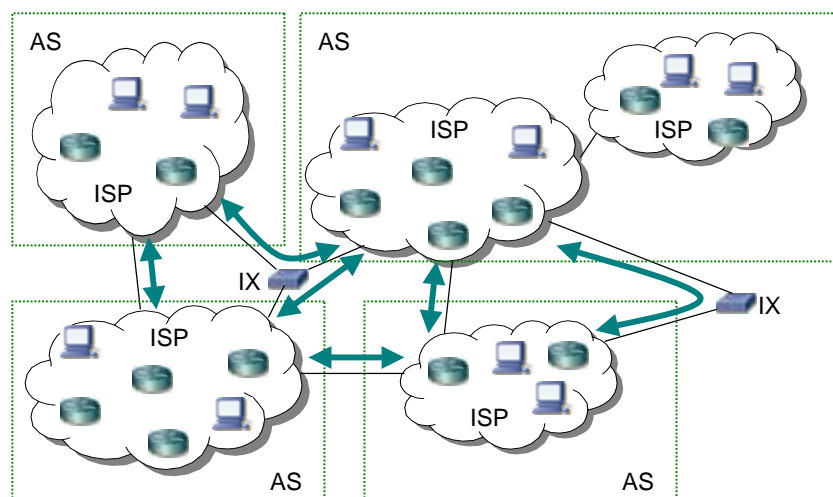
動的経路制御の種類

- ディスタンスベクタ(distance vector)
 - RIPなど、距離と方向を扱うプロトコル
- リンクステート(link state)
 - OSPFやIS-ISなど、リンクの状態を収集して管理するプロトコル
- パスベクタ(path vector)
 - BGPなど、パス属性と方向を扱うプロトコル

Copyright (C) 2005 Internet Initiative Japan Inc.

13

インターネットの構成

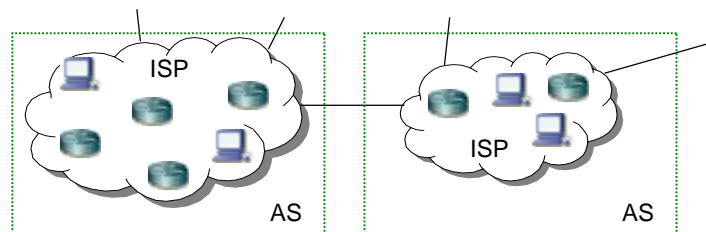


Copyright (C) 2005 Internet Initiative Japan Inc.

14

AS

- Autonomous System
- 統一のルーティングポリシーのもとで運用されているIPプレフィックスの集まり
- インターネットではASの識別子として、IRから一意に割り当てられたAS番号を利用する

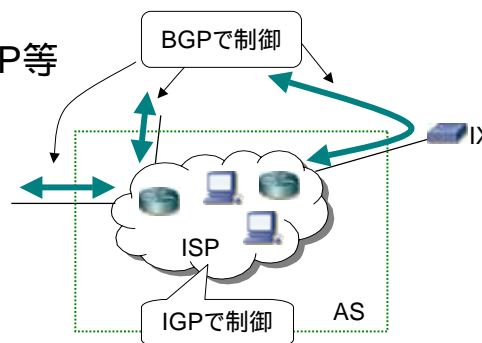


Copyright (C) 2005 Internet Initiative Japan Inc.

15

IGPとEGP

- IGP
 - OSPF、IS-IS、BGP等
 - AS内
- EGP
 - 事実上BGPのみ
 - AS間



Copyright (C) 2005 Internet Initiative Japan Inc.

16

プロトコルの利用法

- OSPF or IS-IS
 - ネットワークのトポロジ情報
 - 必要最小限の経路で動かす
 - 切断などの障害をいち早く通知、迂回
- BGP
 - その他全ての経路
 - 顧客の経路や他ASからの経路
 - 大規模になっても安心

Copyright (C) 2005 Internet Initiative Japan Inc.

17

OSPF

Copyright (C) 2005 Internet Initiative Japan Inc.

18

OSPF概要

- リンクステート型
 - 全ルータがネットワークのトポロジ情報を持つ
 - ネットワークに変更があれば通知
- SPFアルゴリズムによる最適経路の選択
 - リンクのコストによる優先付け
 - 同一コストの複数パスによる負荷分散
- エリアによる階層化
 - エリア境界はルータ
- 隣接のルータと情報を交換

Copyright (C) 2005 Internet Initiative Japan Inc.

19

OSPF RFCs

- 必読
 - [RFC2328] OSPF Version 2
 - この他にもいっぱい
 - [RFC2370] The OSPF Opaque LSA Option
 - [RFC2740] OSPF for IPv6
 - [RFC3101] The OSPF NSSA Option
 - [RFC3137] OSPF Stub Router Advertisement
- :

Copyright (C) 2005 Internet Initiative Japan Inc.

20

OSPF用語

- AS - Autonomous System
 - 共通のプロトコル(OSPF等)で経路情報を交換するルータのグループ
 - BGPなどでいうASとは概念が異なる
 - インターネットでのASはBGPのASを指す場合が多い
- ASBR – AS boundary router
 - AS境界ルータ
 - 外部の経路(static等)をAS内に広報するルータ

Copyright (C) 2005 Internet Initiative Japan Inc.

21

OSPF用語

- ルータID
 - OSPFのAS内でルータを識別する32bitの数値
 - 特に指定が無い場合、ルータのインタフェースのIPアドレスから選ぶ場合が多い

 - ルータIDを変更する場合は、OSPFプロセスの再起動が必要なため、実運用では変更が発生しないようにloopbackインタフェースに付与したIPアドレスを利用する

Copyright (C) 2005 Internet Initiative Japan Inc.

22

OSPFの基礎

OSPFの基礎を整理する

Copyright (C) 2005 Internet Initiative Japan Inc.

23

リンクステートデータベース(LSDB)

- ルータとネットワークの接続状態を表す
 - 各ルータが同じ情報を持つ
- 各ルータは、このデータベースをもとに、同じアルゴリズム (SPF) で経路情報を生成する

Copyright (C) 2005 Internet Initiative Japan Inc.

24

リンクコスト(link cost)

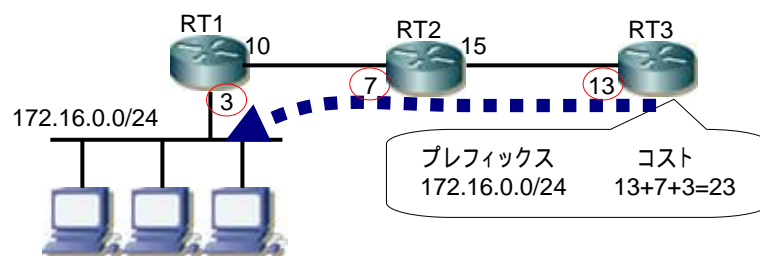


- ルータが、そのインタフェースからパケットを送出するときのコスト(負担)
- 1 ~ 65535の整数を管理者が設定する

Copyright (C) 2005 Internet Initiative Japan Inc.

25

コスト(cost)

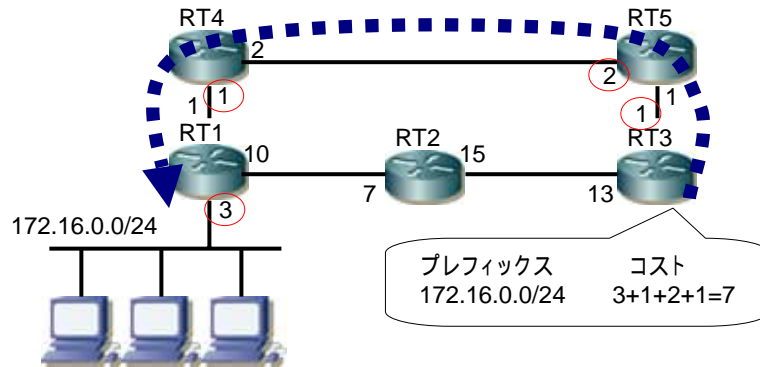


- 宛先までのパスで、パケットが出力されるインタフェースのリンクコストを合計した値

Copyright (C) 2005 Internet Initiative Japan Inc.

26

小さいコストのパスが優先

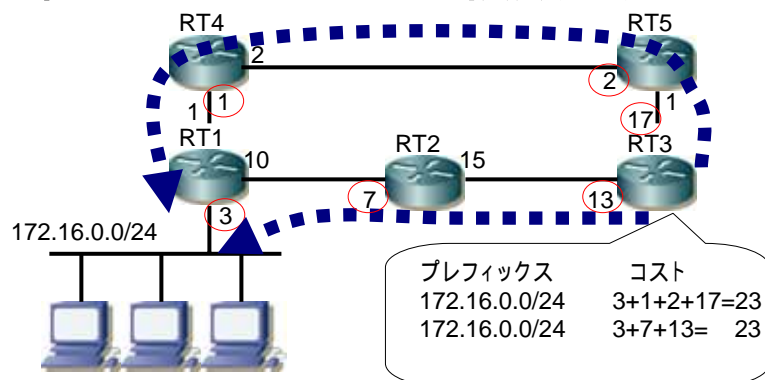


- 最も小さいコストの経路を探索するのが、SPFアルゴリズム

Copyright (C) 2005 Internet Initiative Japan Inc.

27

同じコストのパスを複数利用できる

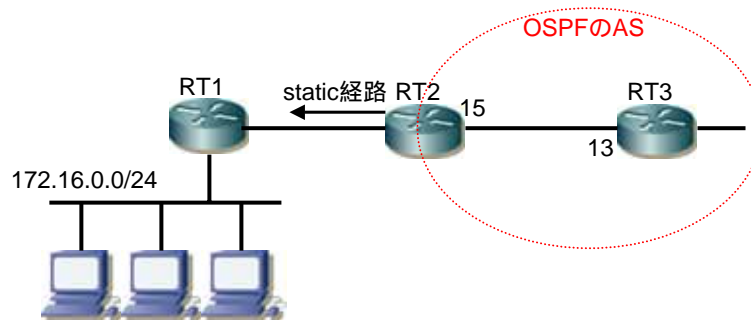


- 同じコストの経路を同時に利用できる
- Equal Cost Multi Path(ECMP)と呼ばれる

Copyright (C) 2005 Internet Initiative Japan Inc.

28

外部経路(external route)



- RT2はstatic経路をOSPF AS内に広報できる
- 外部経路を広報する際にメトリックを付加できる

Copyright (C) 2005 Internet Initiative Japan Inc.

29

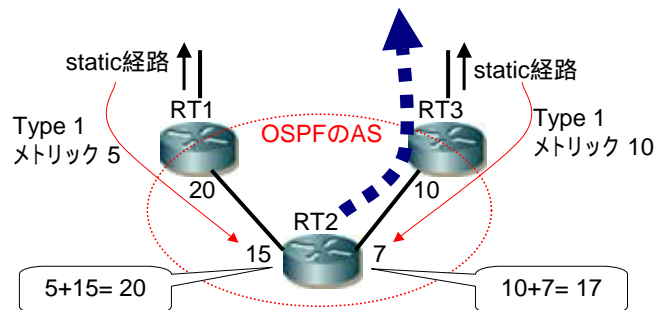
外部経路メトリック

- Type 1 - リンクコストと同様に加算される
 - 同じ宛先のType 1外部経路があった場合、途中リンクのコストも加算して、もっとも小さなコストの経路が選ばれる
- Type 2 - とにかく小さな値が選ばれる
 - 同じ宛先のType 2外部経路があった場合、もっとも小さなType 2メトリックの経路が選ばれる
 - 同じType 2メトリックの場合、転送先アドレスまでのコストがもっとも小さな経路が選ばれる
- 同じ宛先のType 1とType 2の外部経路があった場合、Type 1の経路が選ばれる

Copyright (C) 2005 Internet Initiative Japan Inc.

30

Type 1 外部経路

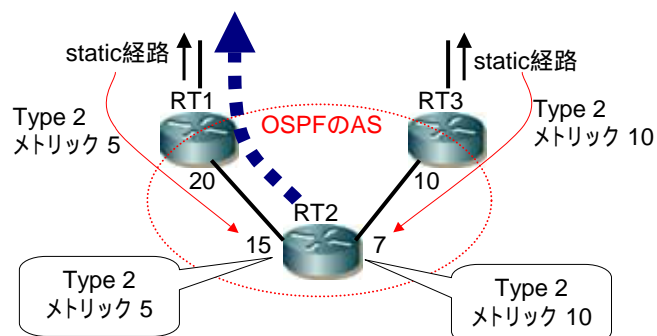


- リンクコストの加算の結果、RT3が広報する外部経路が優先される

Copyright (C) 2005 Internet Initiative Japan Inc.

31

Type 2 外部経路

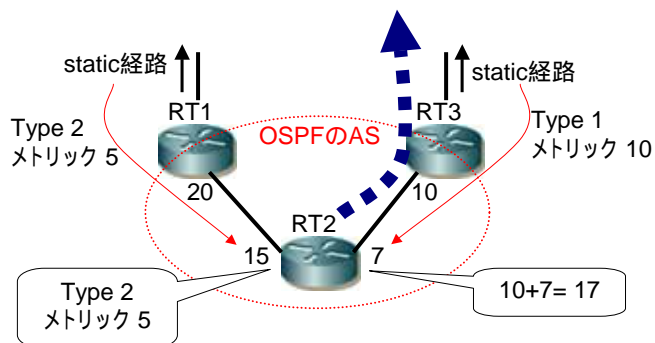


- RT1が広報している、小さなType 2メトリックを持つ外部経路が優先される

Copyright (C) 2005 Internet Initiative Japan Inc.

32

外部経路の混在

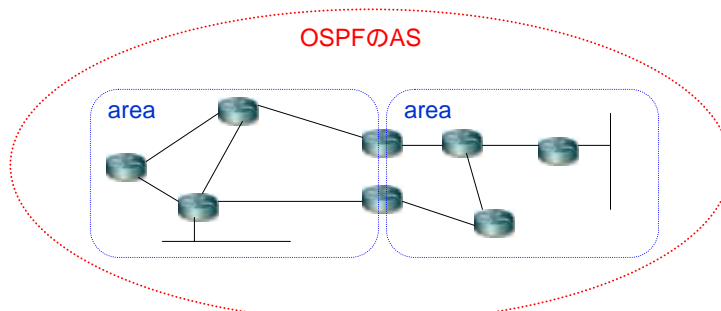


- 常にType 1経路が優先される

Copyright (C) 2005 Internet Initiative Japan Inc.

33

エリア(area)



- OSPFでは連続したネットワークのグループを作成できる。これに接続するルータを含めて、エリアと呼ぶ。

Copyright (C) 2005 Internet Initiative Japan Inc.

34

エリアの概要

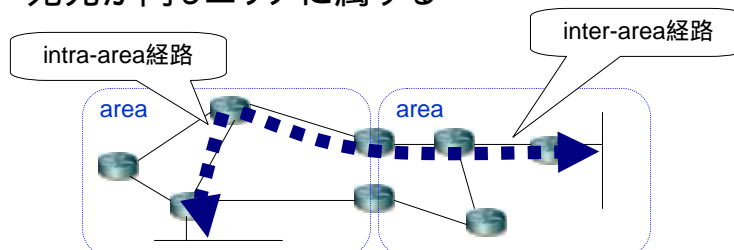
- それぞれのエリアで独立にLSDBが管理され、経路情報が計算される
- あるエリアのトポロジは、他のエリアからは見えない
 - 必要な経路情報のみが伝播する
 - 計算負荷の軽減

Copyright (C) 2005 Internet Initiative Japan Inc.

35

エリア間とエリア内

- エリア間経路(inter-area経路)
 - 宛先が異なるエリアに属する
- エリア内経路(intra-area経路)
 - 宛先が同じエリアに属する



Copyright (C) 2005 Internet Initiative Japan Inc.

36

エリアID(area ID)

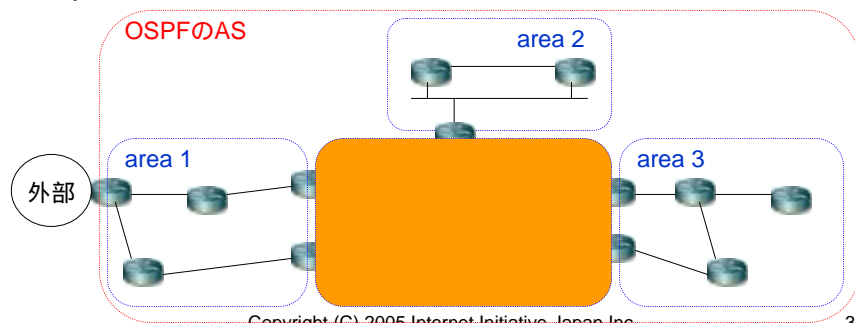
- 各エリアを識別する32bitの数値
 - 各エリアに管理者がIDを設定する
- そのまま数字で表記する書式
 - area 0
- IPアドレスの様に8bit毎に区切った書式
 - area 0.0.0.0

Copyright (C) 2005 Internet Initiative Japan Inc.

37

バックボーン エリア(area 0)

- エリアIDが0
- 各エリアの経路情報を交換できる特別なエリア
- OSPFのマルチエリア構成は、バックボーンエリアを中心としたスター型



Copyright (C) 2005 Internet Initiative Japan Inc.

38

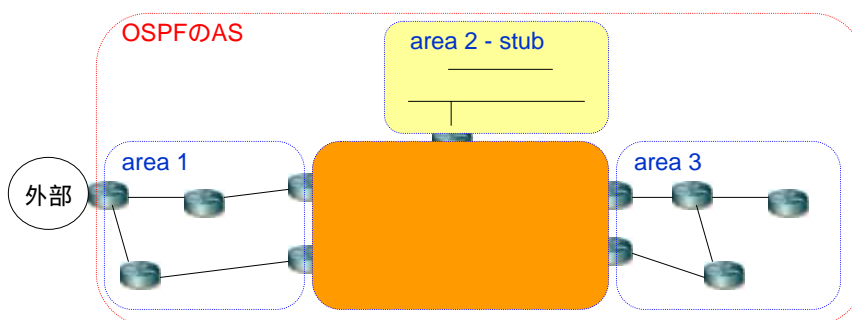
stubエリア

- 幾つかのOSPFの構成ではほとんどが外部経路だった
 - 外部経路は基本的に全エリアに広報される
 - これを軽減する仕組みを考えた
- 外部経路を伝播しない代わりに、default経路をエリア境界ルータが広報する
 - ASBRが無いエリアに適用できる
 - エリア内の全てのルータでstubエリアと設定

Copyright (C) 2005 Internet Initiative Japan Inc.

39

area 2をstubエリアにしている例



- area 2にはエリア間経路(inter area経路)とdefault経路のみが広報される
- この場合、area 3もstubエリアにすることができる

Copyright (C) 2005 Internet Initiative Japan Inc.

40

OSPFプロトコルパケット

IPヘッダ	IPヘッダ	IPヘッダ	IPヘッダ	IPヘッダ
OSPFヘッダ	OSPFヘッダ	OSPFヘッダ	OSPFヘッダ	OSPFヘッダ
HELLO	DD	LS Req	LS Update	LS Ack
ネイバs	LSAヘッダs	LSAヘッダs	LSAs	LSAヘッダs

Copyright (C) 2005 Internet Initiative Japan Inc.

41

LSA

LSA(link state advertisement)の
パケットフォーマットを解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

42

リンクステート広告 (LSA)

- link state advertisement
- 各ルータが広告する情報のこと
- LSAの集合がLSDBになる
- 目的に応じた幾つかのタイプがある

Copyright (C) 2005 Internet Initiative Japan Inc.

43

LSAの種類 (基本)

LSタイプ

- 1. ルータLSA**
ルータに接続するリンク、ネットワーク情報を運ぶ
リンク種別に応じてp2p, transit, stub, virtualの4種類
- 2. ネットワークLSA**
ネットワークに接続するルータ情報を運ぶ
- 3. サマリLSA (ネットワーク)**
エリア外のネットワークへの経路を運ぶ
- 4. サマリLSA (ASBR)**
エリア外のASBRへの経路を運ぶ
- 5. AS-external-LSA**
外部経路を運ぶ
Type 1とType 2のメトリックタイプが存在する

Copyright (C) 2005 Internet Initiative Japan Inc.

44

LSAの種類(拡張)

LSタイプ

6. group-membership-LSAs
MOSPFで利用
7. NSSA-external-LSA
NSSAエリアで、外部経路を広告する
8. (External-attributes-LSA)
RFCにならずに消失
9. Opaque LSA(link-local scope)
特殊な情報を運ぶLSAで、リンク内のみに伝播する
10. Opaque LSA(area-local scope)
特殊な情報を運ぶLSAで、エリア内のみに伝播する
11. Opaque LSA(as-wide scope)
特殊な情報を運ぶLSAで、OSPFのAS内に全て伝播する

Copyright (C) 2005 Internet Initiative Japan Inc.

45

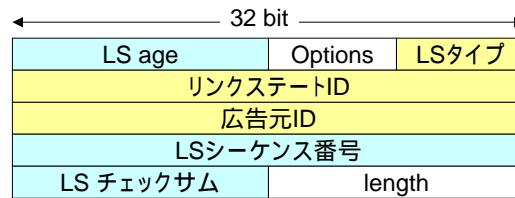
LSAに共通に含まれる情報

- LS age
 - LSAが生成されてからの秒数
 - 生成時は0で1時間を越えたものは利用されない
 - LSタイプ
 - LSAの種類
 - 広告元ID
 - LSA生成ルータのルータID
 - LSシーケンス番号
 - 32bitの符号付整数で、大きいほど新しいLSAを示す
 - 最大値に達した場合は、LSAを一旦消してから、新たに広告しなおす
 - リンクステートID
 - LSAで示される情報のID
 - LSタイプに応じた値をとる
- | LSタイプ | リンクステートIDの意味 |
|-------|-----------------|
| 1 | 広告ルータのルータID |
| 2 | DRのリンクのIPアドレス |
| 3 | 宛先ネットワークのIPアドレス |
| 4 | ASBRのルータID |
| 5 | 宛先ネットワークのIPアドレス |

Copyright (C) 2005 Internet Initiative Japan Inc.

46

LSA header



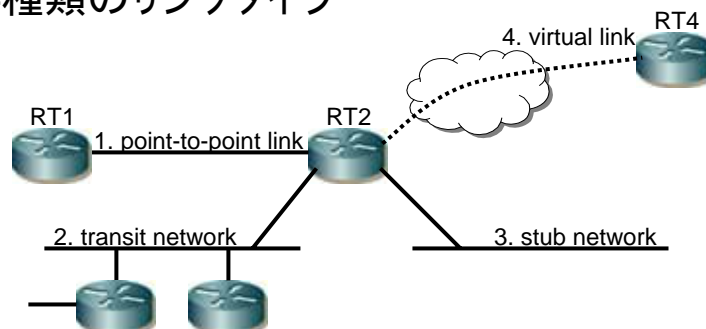
- 20-octetの固定長
- LSAを個別に識別できる
 - LSタイプ, リンクステートID, 広告元ID
- 新しいLSAを識別できる
 - LS age, LSシーケンス番号, LSチェックサム

Copyright (C) 2005 Internet Initiative Japan Inc.

47

LSタイプ1 - ルータLSA

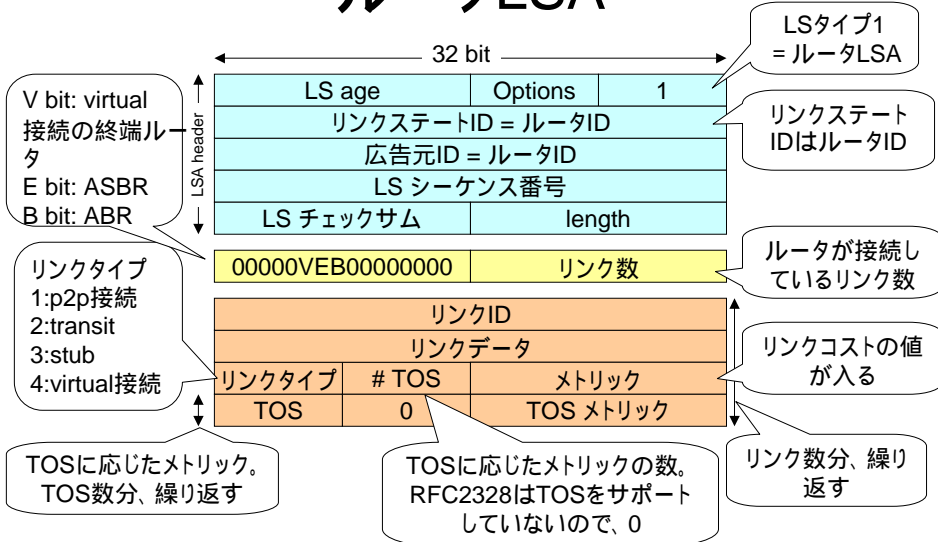
- 全ルータが一つずつ広告する自己紹介
- 4種類のリンクタイプ



Copyright (C) 2005 Internet Initiative Japan Inc.

48

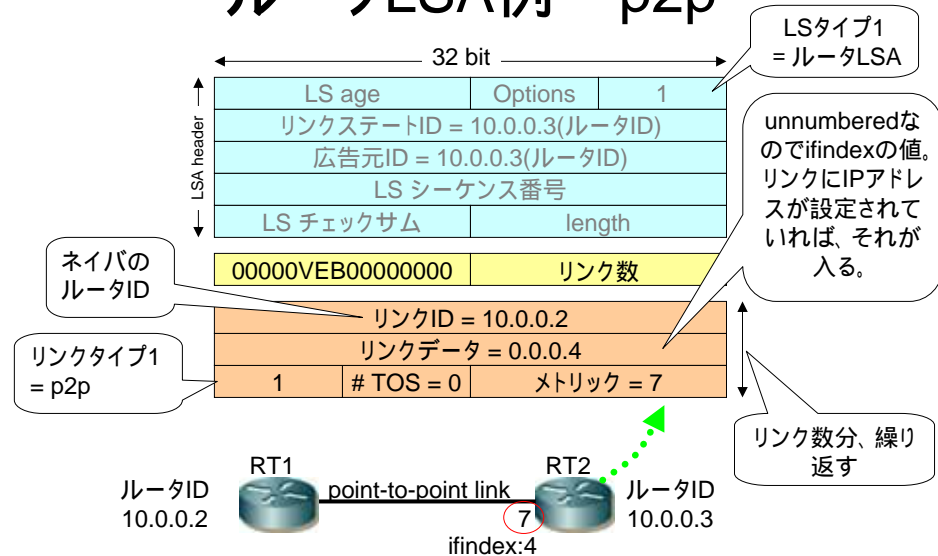
ルータLSA



Copyright (C) 2005 Internet Initiative Japan Inc.

49

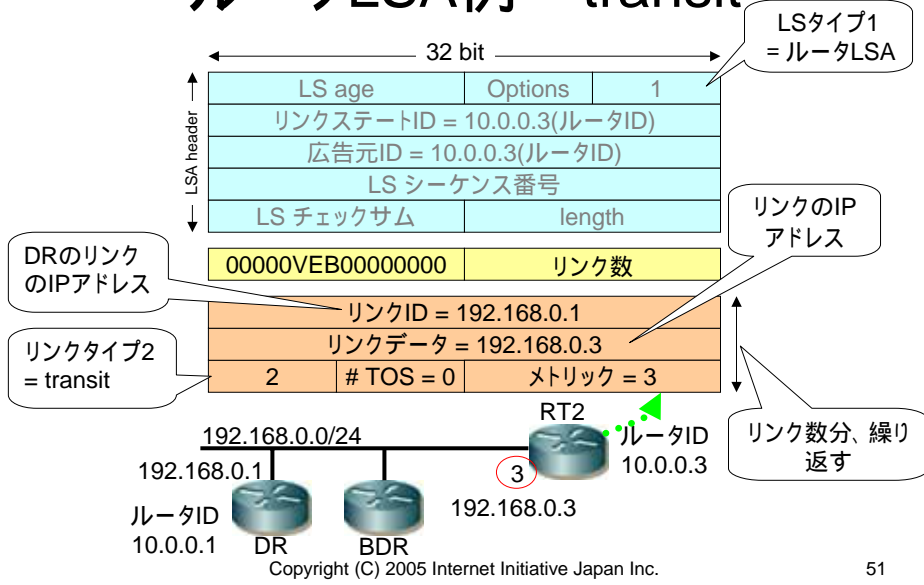
ルータLSA例 - p2p



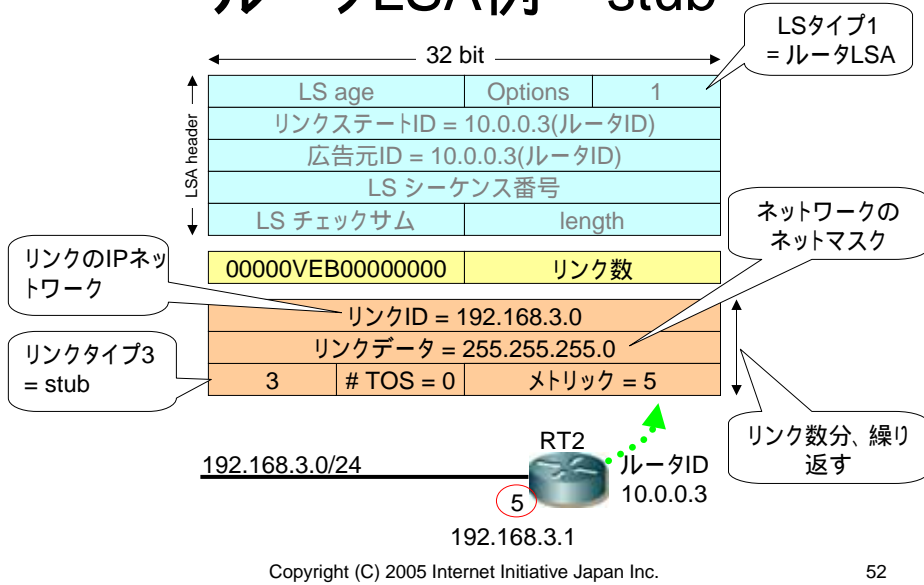
Copyright (C) 2005 Internet Initiative Japan Inc.

50

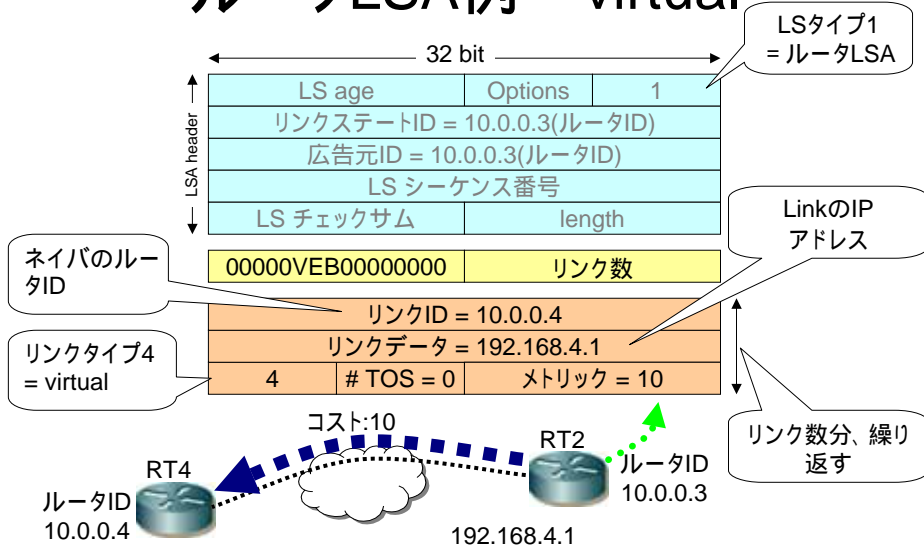
ルータLSA例 – transit



ルータLSA例 – stub



ルータLSA例 – virtual



53

ルータLSAで運ぶ情報

- stubネットワークのみが経路情報を運び、その他はトポロジ情報を運ぶ

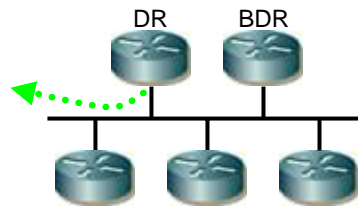
リンクタイプ	リンクID	リンクデータ	データ種類
1 p2p	ネイバのルータID	MIB-II ifindex値 (あれば、リンクのIPアドレス)	トポロジ
2 transit	DRのリンクのIPアドレス	リンクのIPアドレス	トポロジ
3 stub	リンクのIPネットワーク	ネットワークマスク	経路
4 virtual	ネイバのルータID	リンクのIPアドレス	トポロジ

Copyright (C) 2005 Internet Initiative Japan Inc.

54

LSタイプ2 - ネットワークLSA

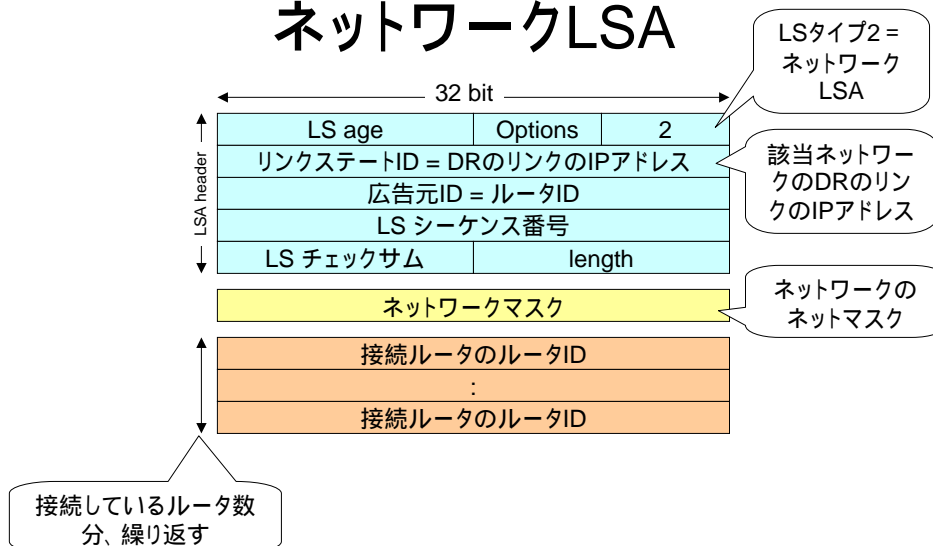
- transitネットワークに接続するルータのリスト
- ネットワークにつき1台の代表ルータ(DR)のみが広告する



Copyright (C) 2005 Internet Initiative Japan Inc.

55

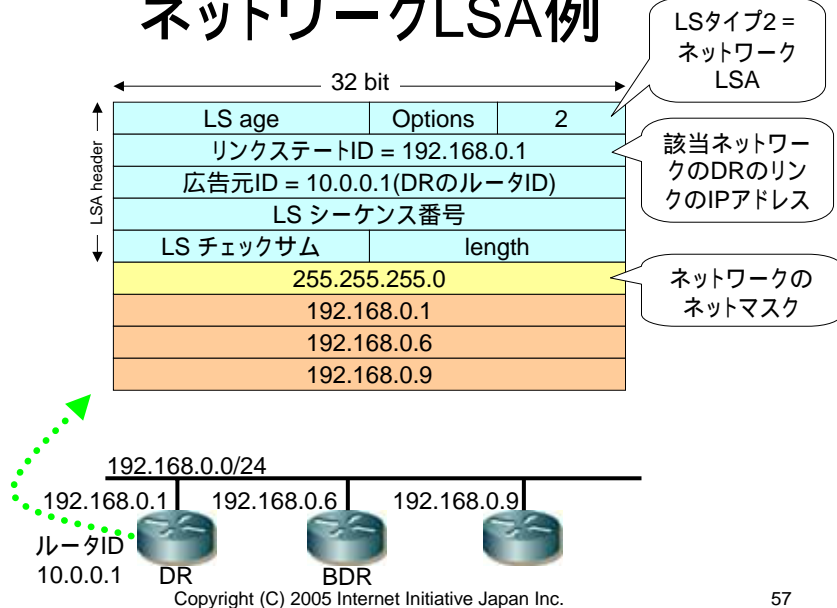
ネットワークLSA



Copyright (C) 2005 Internet Initiative Japan Inc.

56

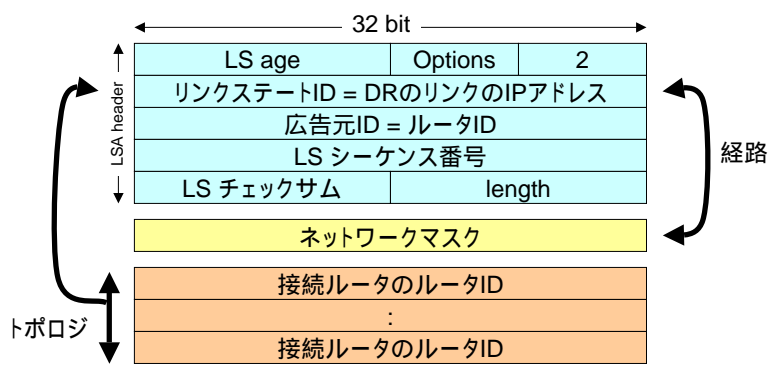
ネットワークLSA例



57

ネットワークLSAで運ぶ情報

- トポロジと経路を同時に運ぶ

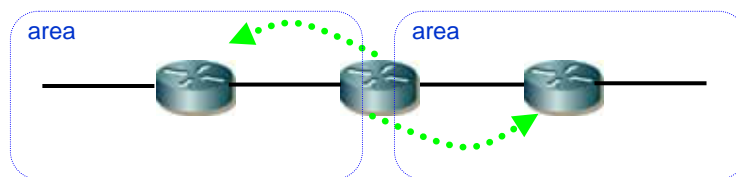


Copyright (C) 2005 Internet Initiative Japan Inc.

58

LSタイプ3 - サマリLSA

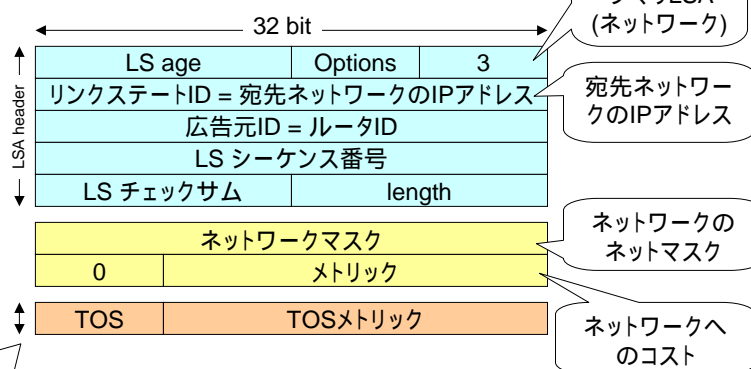
- エリア外のネットワークへの経路情報を運ぶ
- エリア境界で、エリア境界ルータが生成する



Copyright (C) 2005 Internet Initiative Japan Inc.

59

LSタイプ3 サマリLSA

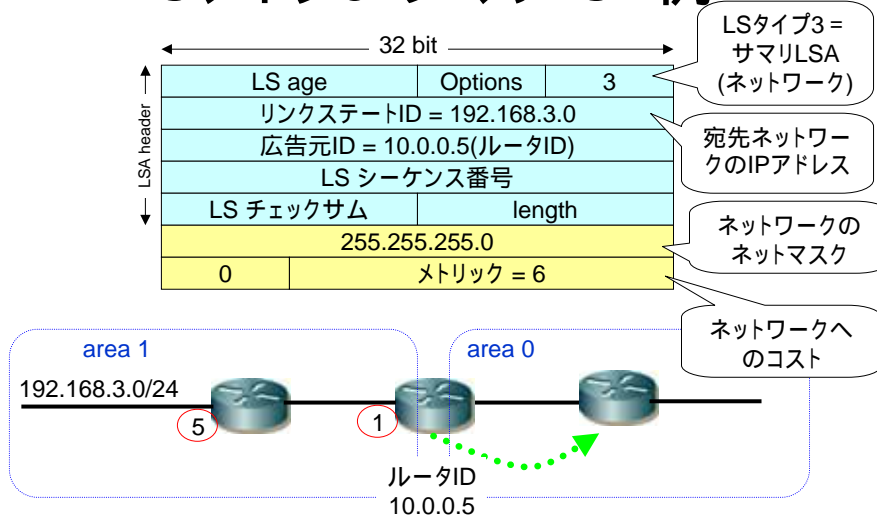


TOSに応じたメトリック。TOS数分、繰り返す。RFC2328はTOSをサポートしていないので付加されない

Copyright (C) 2005 Internet Initiative Japan Inc.

60

LSタイプ3 サマリLSA例

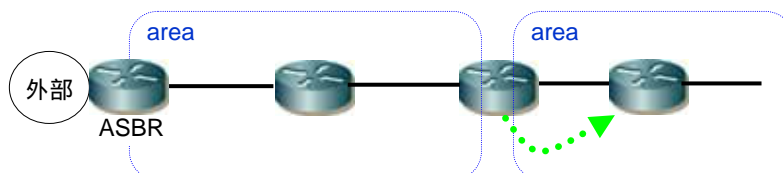


Copyright (C) 2005 Internet Initiative Japan Inc.

61

LSタイプ4 - サマリLSA

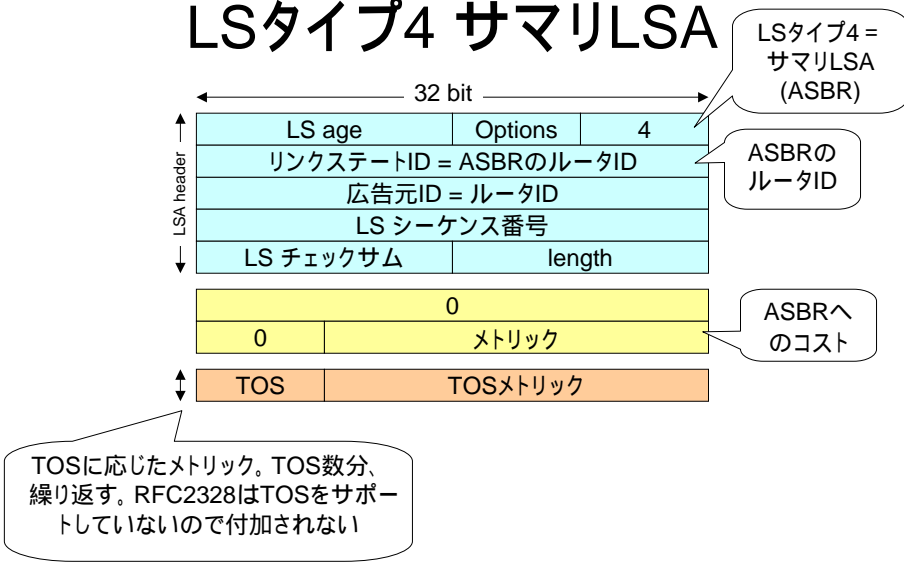
- エリア外のASBRへの経路情報を運ぶ
- エリア境界で、エリア境界ルータが生成する
- LSタイプ3とほぼ一緒
 - リンクステートIDがルータID、ネットマスクが0



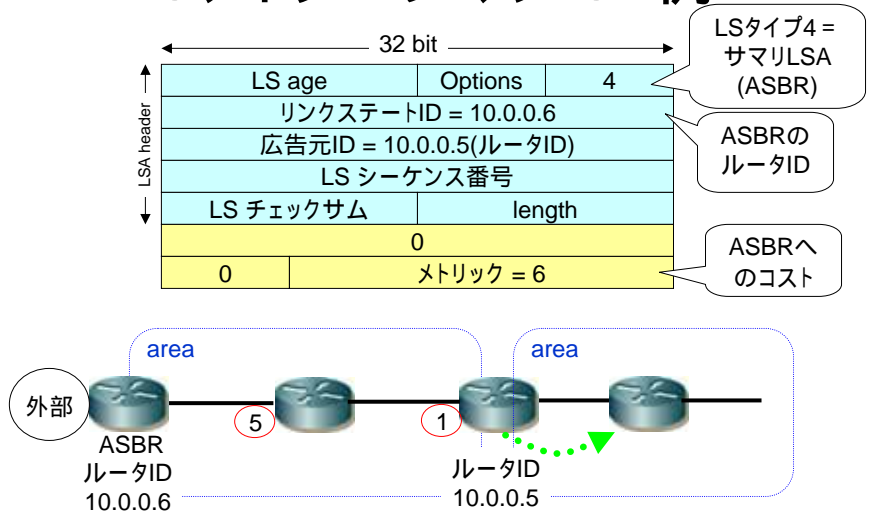
Copyright (C) 2005 Internet Initiative Japan Inc.

62

LSタイプ4 サマリLSA

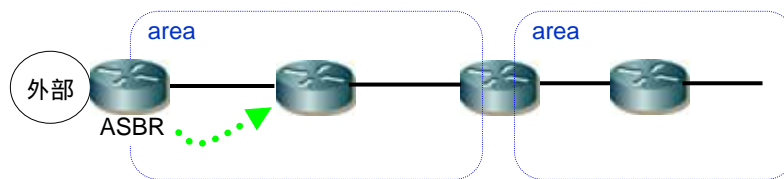


LSタイプ4 サマリLSA例



LSタイプ5 - AS-external-LSA

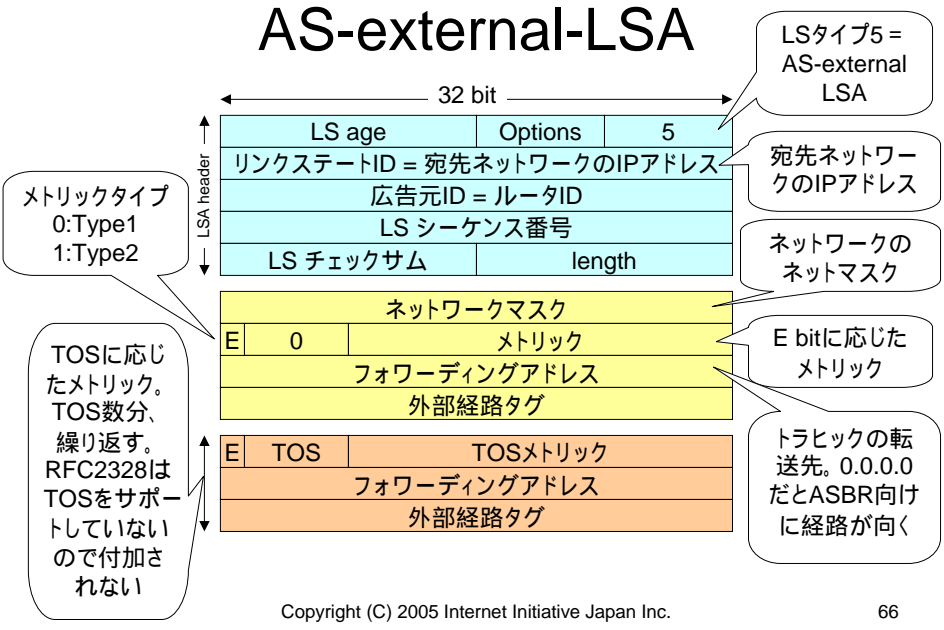
- 外部経路の経路情報を運ぶ
- AS境界ルータ(ASBR)が生成する
- Type1またはType2のメトリック



Copyright (C) 2005 Internet Initiative Japan Inc.

65

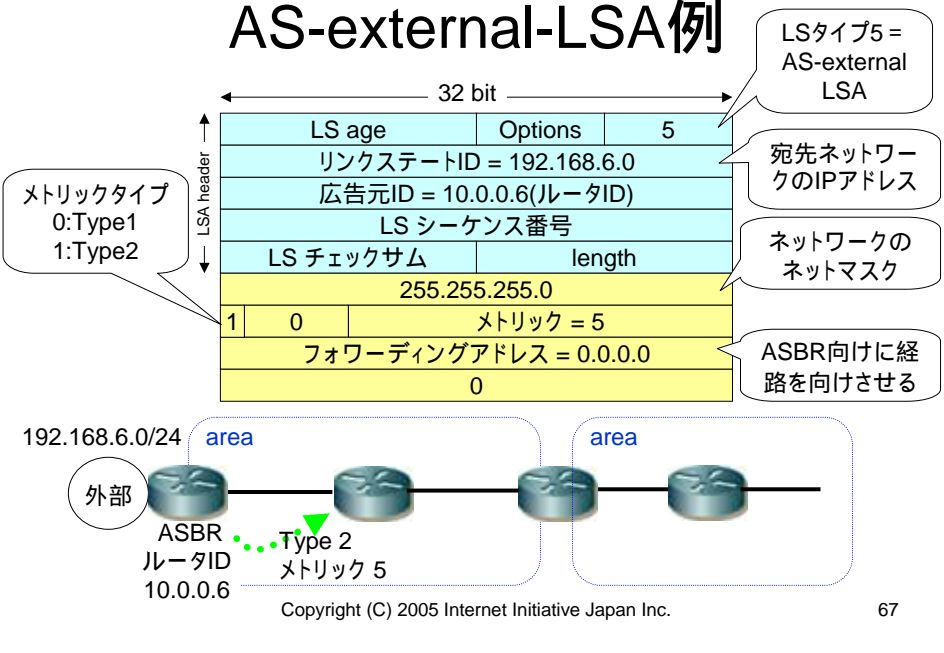
AS-external-LSA



Copyright (C) 2005 Internet Initiative Japan Inc.

66

AS-external-LSA例

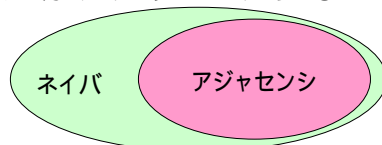


隣接関係

隣接関係とDR、BDRの役割

ネイバとアジャセンシ

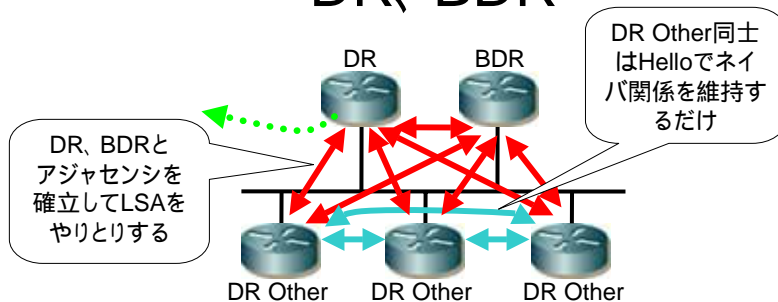
- ネイバ(neighbor)
 - 同じネットワークに接続する2台のルータの関係
 - 多くの場合、Helloで自動的に探索、維持される
- アジャセンシ(adjacency)
 - 経路情報を交換するネイバの関係
 - 全ネイバがアジャセンシになるわけではない



Copyright (C) 2005 Internet Initiative Japan Inc.

69

DR、BDR



- Designated Router (代表ルータ) とそのBackup
- DR Other (その他のルータ) は、DR と BDR とだけアジャセンシを確立する
- DR がネットワーク LSA を広報する責任を負う

Copyright (C) 2005 Internet Initiative Japan Inc.

70

DR、BDRの選出

1. 現時点でDRではないルータからBDRを選出する
 1. 複数のルータがBDRだと主張した場合は以下の順序で選ぶ
 1. ルータ優先度(Router Priority)が最も高い
 2. 大きなルータIDを持つ
 2. BDRがない場合は以下の順序で選ぶ
 1. ルータ優先度(Router Priority)が最も高い
 2. 大きなルータIDを持つ
2. DRを選出する
 1. 複数のルータがDRだと主張した場合は、以下の順序で選らぶ
 1. ルータ優先度(Router Priority)が最も高い
 2. 大きなルータIDを持つ
 2. DRがない場合は、BDRがDRになる

Copyright (C) 2005 Internet Initiative Japan Inc.

71

DR、BDRの選出要約

- ルータ優先度の高いルータが選出
 - 同じ場合は、ルータIDの大きな方
- 既に選出が終わっていても、基本的に置き換わらない
- DRが故障したときはBDRがDRへと移行する
 - BDRがDRになるまで、新たなBDRは選ばれない
 - このルールを記述するために複雑になっている

Copyright (C) 2005 Internet Initiative Japan Inc.

72

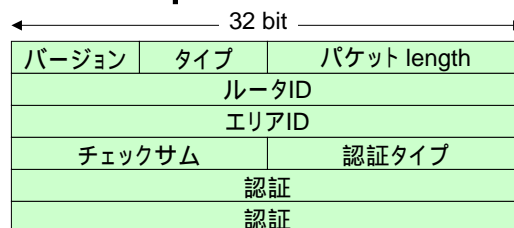
OSPFパケット

OSPFのプロトコルパケットの フォーマットを解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

73

OSPF packet header



• 24-octetの固定長

タイプ:

- 1 Hello
- 2 Database Description
- 3 Link State Request
- 4 Link State Update
- 5 Link State Acknowledgment

認証タイプ:

- 1 認証なし
- 2 シンプルパスワード認証
- 3 暗号認証

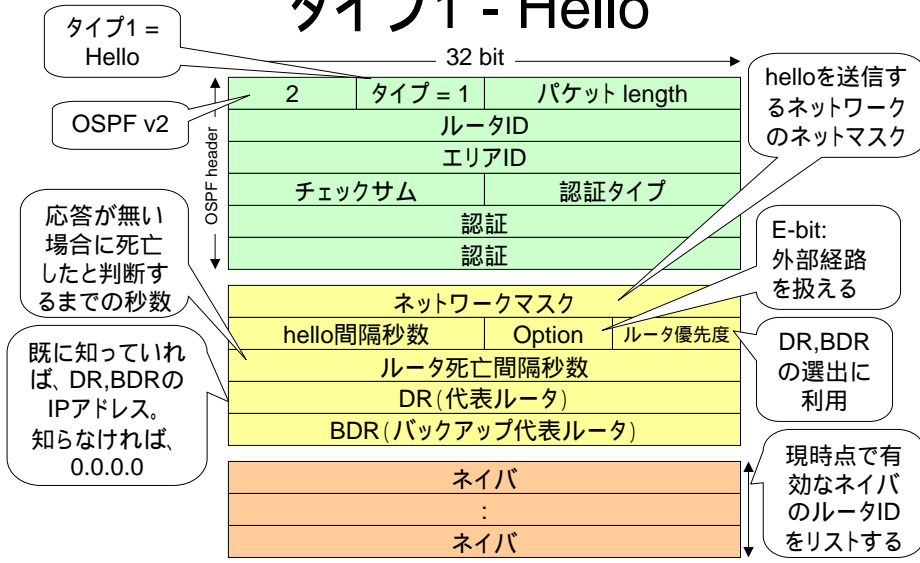
Copyright (C) 2005 Internet Initiative Japan Inc.

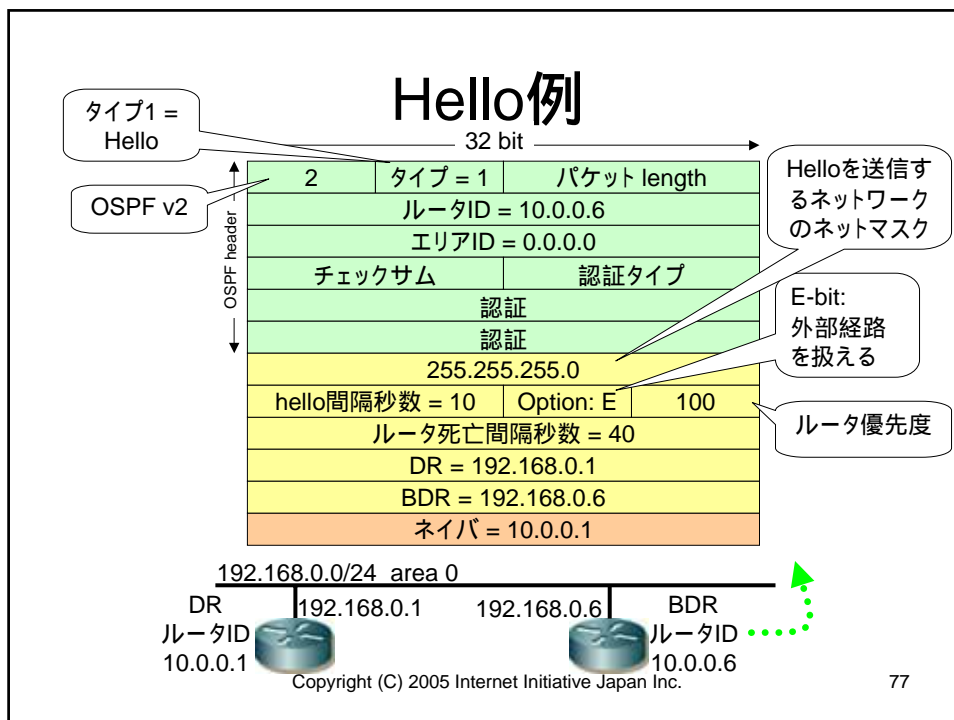
74

タイプ1 - Hello

- ネイバの管理
 - ルータID、エリアIDの通知
 - ネイバの検出、維持、生死確認
- ルータ優先度の通知
 - DR、BDRの選出
- Optionフィールドでルータの機能の通知
 - E-bit: 外部経路が扱えるかどうか
Stubエリアでは0、それ以外は1

タイプ1 - Hello

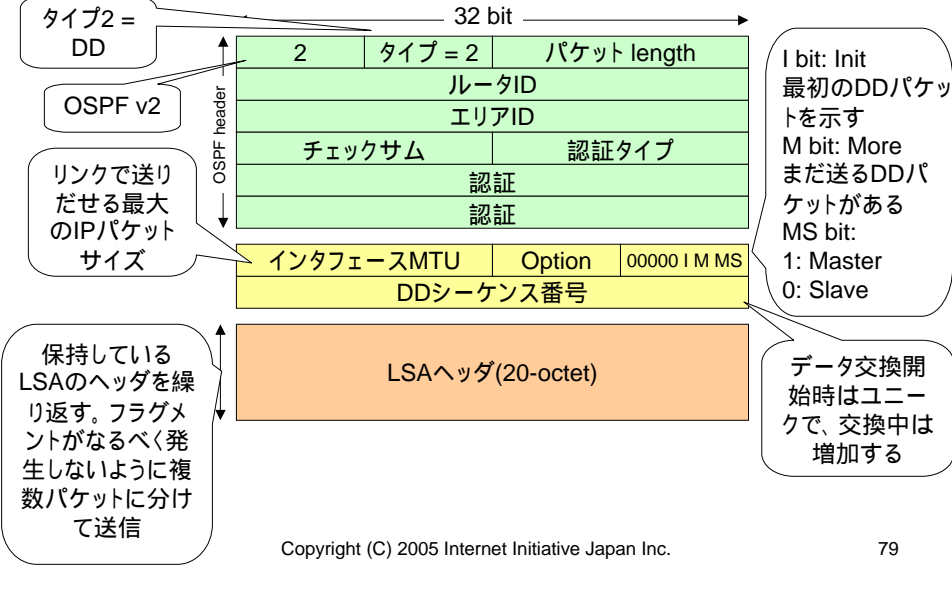




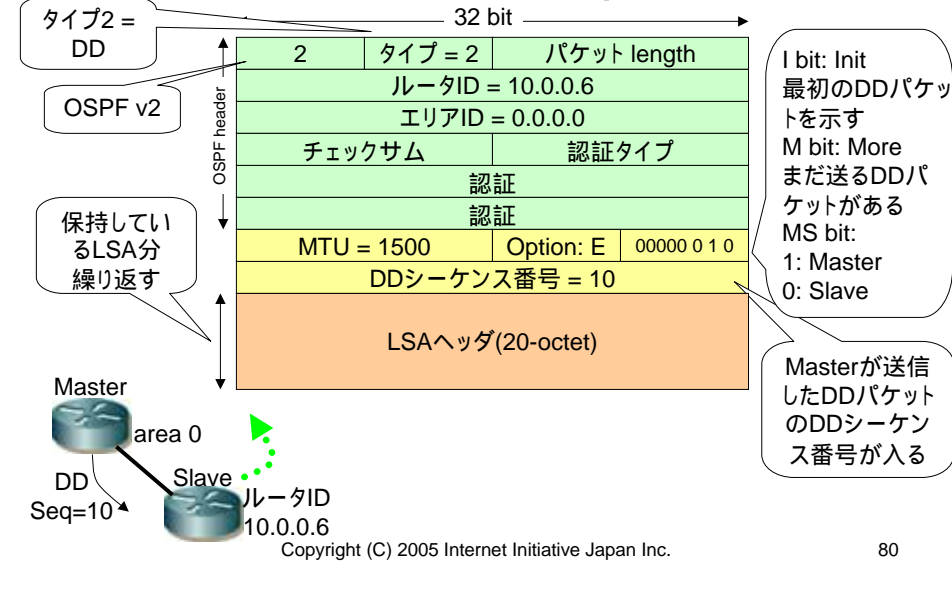
タイプ2 - Database Description

- アジャセンシ確立時に、保持するLSAを通知
 - LSDBの同期をとる
 - 全LSAのヘッダのみを伝える
 - マスタとスレーブになって情報を交換
 - Router IDの大きい方がマスタ
 - スレーブはマスタのDDシーケンス番号に同期する
- インタフェースのMTUを伝える
- OptionフィールドでRouterの機能の通知
 - E-bit: 外部経路が扱えるかどうか(Stubエリアでは0)

タイプ2 - Database Description



Database Description例



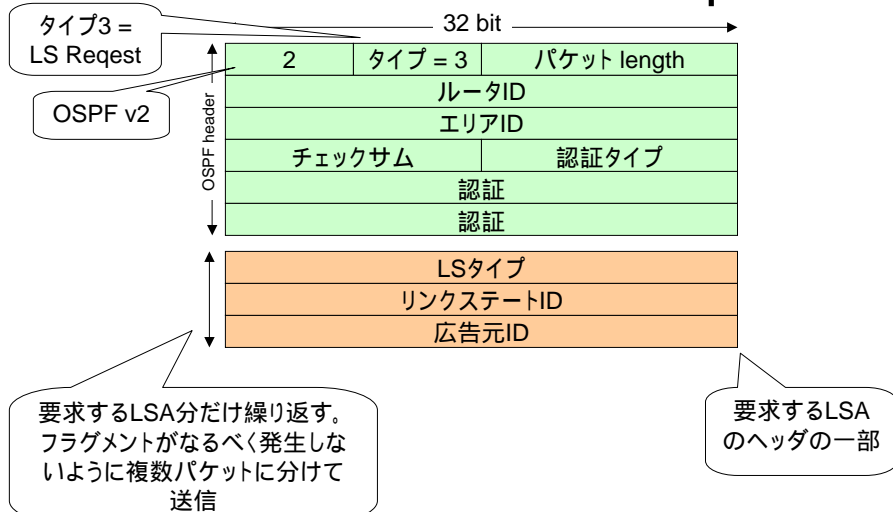
タイプ3 - Link State Request

- DDパケットでLSA情報を交換後、差分を埋めるためにLSAを要求する
 - 最新のLSAを要求
 - 保持していないLSAを要求
- LSAが識別できる情報をリストして送信する
 - LSタイプ、リンクステートID、広告元ID

Copyright (C) 2005 Internet Initiative Japan Inc.

81

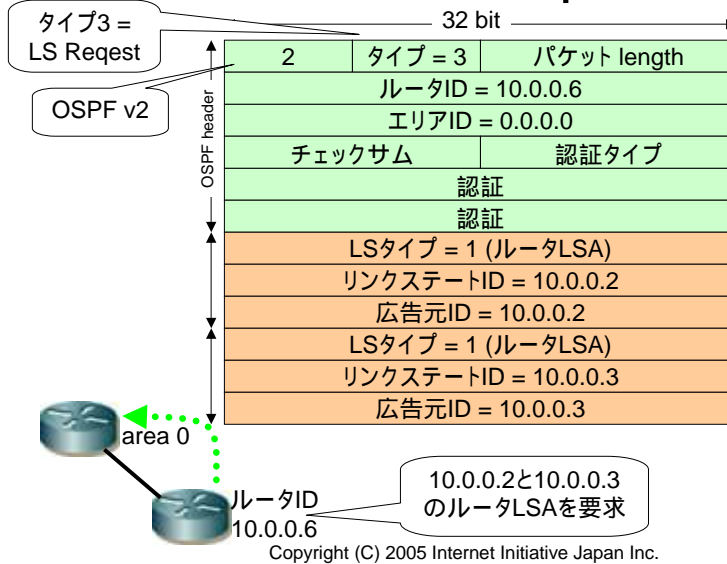
タイプ3 - Link State Request



Copyright (C) 2005 Internet Initiative Japan Inc.

82

Link State Request例



83

タイプ4 - Link State Update

- 一つ以上のLSAを運ぶ
- 隣接のルータまで通知される
 - LSAの転送はHop by Hop
 - アジャセンシ確立後の受信確認にはLink State Acknowledgmentを利用する

Copyright (C) 2005 Internet Initiative Japan Inc.

84

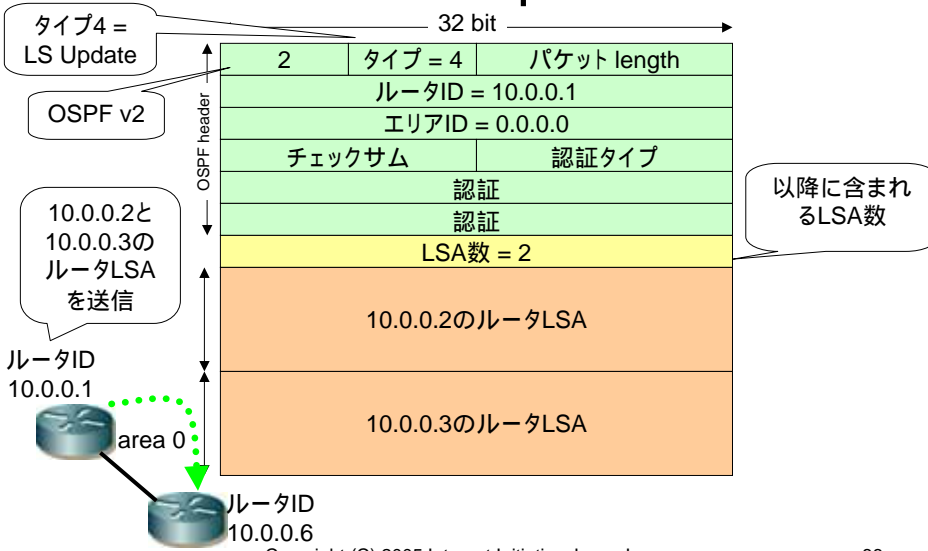
タイプ4 - Link State Update



Copyright (C) 2005 Internet Initiative Japan Inc.

85

Link State Update例



Copyright (C) 2005 Internet Initiative Japan Inc.

86

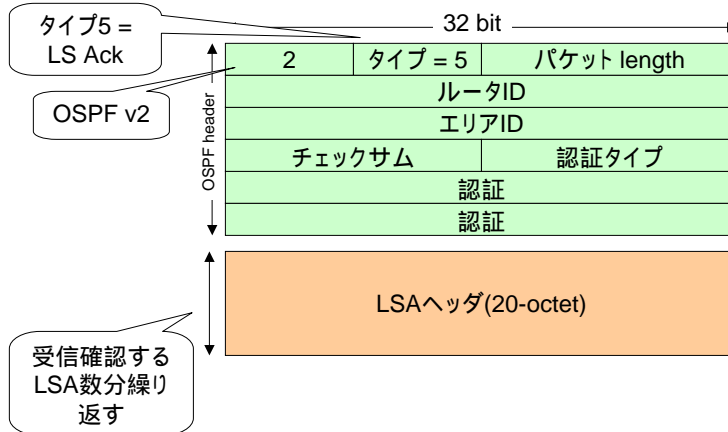
タイプ5 - Link State Acknowledgment

- 受信確認を通知する
 - LSAのヘッダを通知する
 - これで確実にLSAが伝わったことを保証する
- 一つ以上の受信確認を運ぶ

Copyright (C) 2005 Internet Initiative Japan Inc.

87

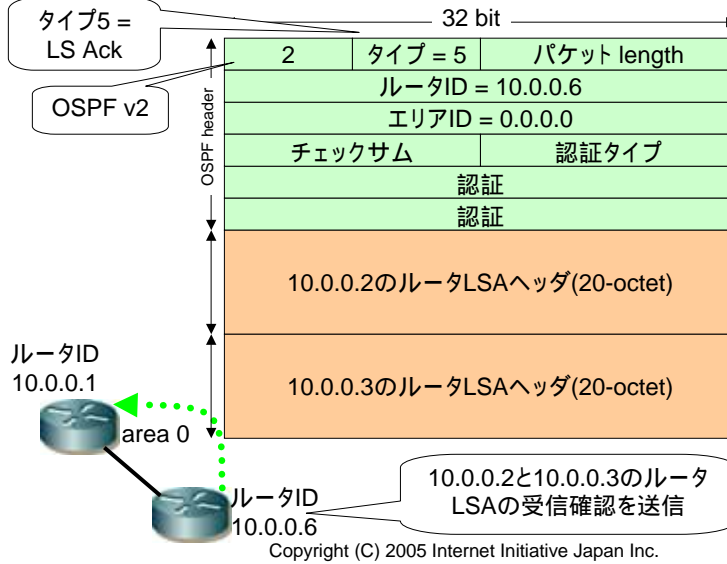
タイプ5 - Link State Acknowledgment



Copyright (C) 2005 Internet Initiative Japan Inc.

88

Link State Acknowledgment例



89

OSPFパケットの送信先

- ALLSPFRouter[224.0.0.5]
 - 全てのOSPFルーターが受信する
- ALLDRRouter[224.0.0.6]
 - DR, BDRのみが受信する
- p2p接続では [224.0.0.5]宛
- ブロードキャストネットワークで、
 - Hello及びDRとBDRからのLS update、LS Ackは[224.0.0.5]宛
 - DROtherからのLS update、LS Ackは[224.0.0.6]宛
- その他
 - ネイバへのunicast宛

Copyright (C) 2005 Internet Initiative Japan Inc.

90

トポロジの把握のために

隣接関係の確立と、LSAによるトポ
ロジ情報の伝播

Copyright (C) 2005 Internet Initiative Japan Inc.

91

アジャセンシの確立まで

1. Helloでネイバを確立(2way)
 - お互いのHelloパケット中に自ルータIDがある
 - 必要であればDR、BDRの選出
2. DDでデータベースの情報交換(Exchange)
3. 差分があれば
 - LS Requestで必要なLSAを要求(loading)
 - LS UpdateでLSAを送信
4. 同期完了、アジャセンシ確立(Full)

Copyright (C) 2005 Internet Initiative Japan Inc.

92

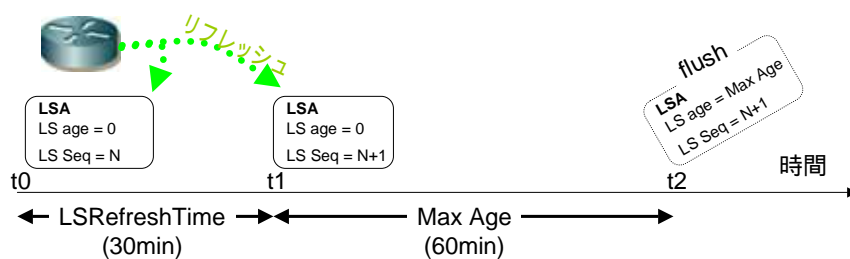
アジャセンシ確立の条件

- Hello段階
 - 認証が同一
 - エリアIDが同一
 - 同一ネットワークに所属(p2pとvirtual接続を除く)
 - OptionのE-bitが同一(stubエリアかどうか)
 - hello送信間隔が同一
 - ルータ死亡間隔秒数が同一
- DD段階
 - インタフェースのMTUが同一
 - OptionのE-bitが同一(stubエリアかどうか)

Copyright (C) 2005 Internet Initiative Japan Inc.

93

LSAのリフレッシュ



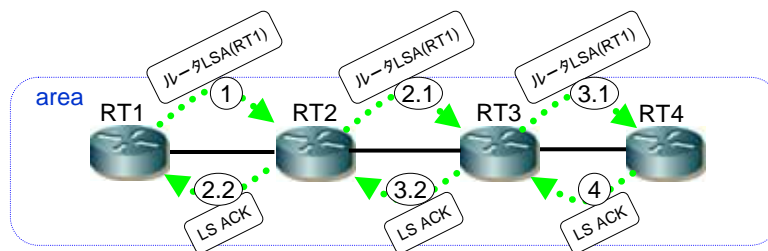
- 各ルータはそれぞれのLSAの経過秒数を保持
- Max Ageに達するまでにLSAのリフレッシュor更新が無ければ、該当のLSAは消される
- ベンダごとのタイマーの非互換は大問題

Copyright (C) 2005 Internet Initiative Japan Inc.

94

LSAの伝播

- LSAの伝播は1ホップずつ
- 隣接に広報してから、ACKを返す



Copyright (C) 2005 Internet Initiative Japan Inc.

95

OSPFの再計算

- トポロジの変化などがLSAで通知されると、SPFの再計算を行う
- 検知
 - ルータが変更気づく
- 通知
 - 周囲のルータにLSAを送付して通知
- 再構成
 - 最適経路を計算して経路情報を更新する

Copyright (C) 2005 Internet Initiative Japan Inc.

96

LSAとエリア

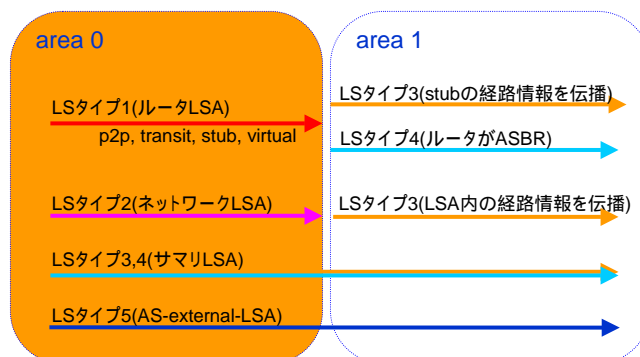
エリア間のLSAの伝播を整理する

Copyright (C) 2005 Internet Initiative Japan Inc.

97

バックボーンから他エリアへのLSA

- ルータLSA,ネットワークLSAの経路部分が伝播する
- その他のLSAはそのまま伝播する

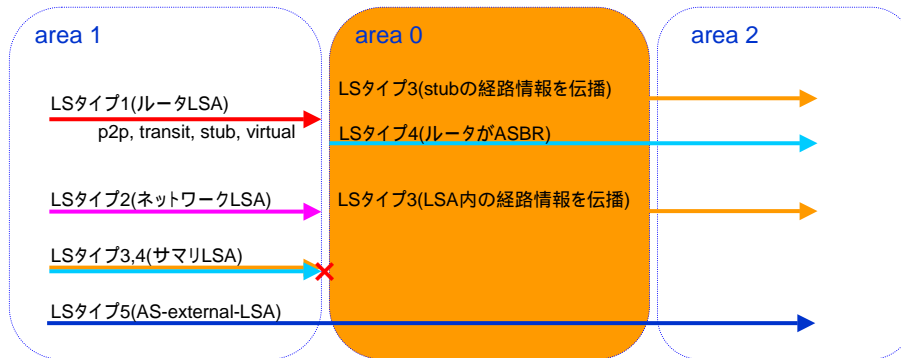


Copyright (C) 2005 Internet Initiative Japan Inc.

98

エリア間でのLSA

- バックボーンエリアのみがサマリLSAを他エリアに中継できる

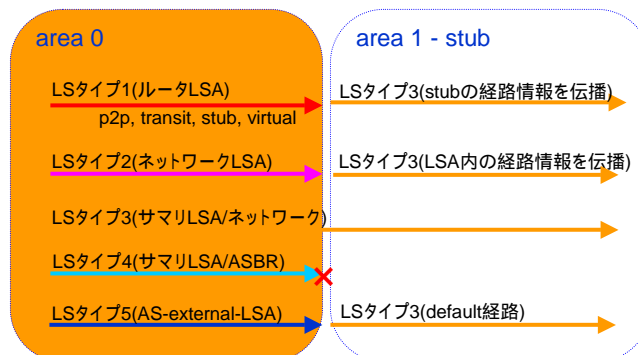


Copyright (C) 2005 Internet Initiative Japan Inc.

99

stubエリアへのLSA

- 外部経路が伝播せず、default経路が広告される
- ASBRへの経路も必要ないので、伝播しない

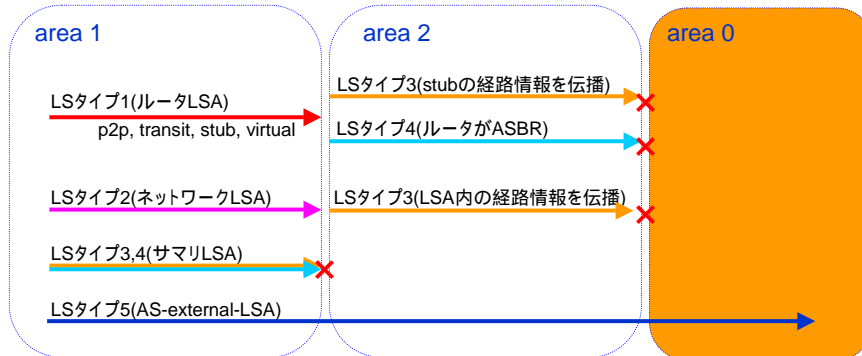


Copyright (C) 2005 Internet Initiative Japan Inc.

100

エリア構成を誤った場合

- area 1がバックボーンエリアに接していない
- ほとんどの経路がバックボーンに届かない
- 外部経路も転送先アドレスが到達できないと考えられるので、ほぼ全ての経路が利用できない

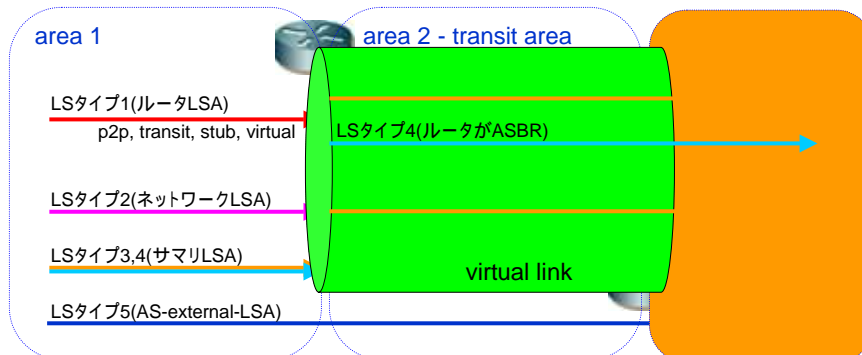


Copyright (C) 2005 Internet Initiative Japan Inc.

101

virtual linkの利用

- area 0が張り出している様にLSAのやり取りを行う
- 構成が複雑になるので、緊急時以外お勧めしない

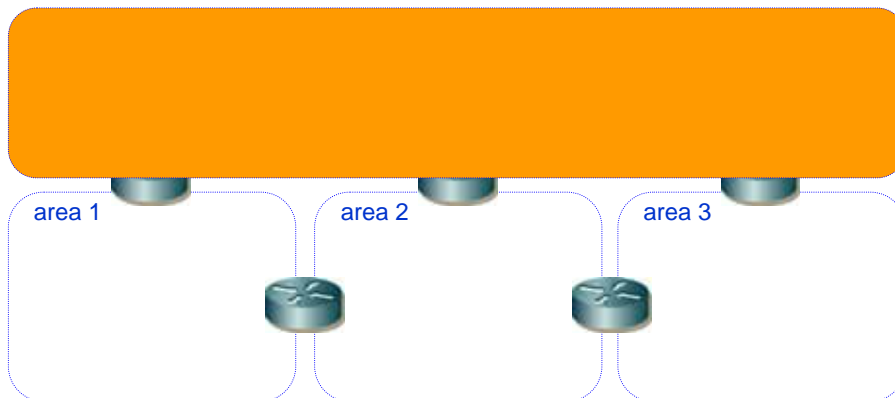


Copyright (C) 2005 Internet Initiative Japan Inc.

102

練習問題

- エリア間の通信を考えてみましょう



Copyright (C) 2005 Internet Initiative Japan Inc.

103

OSPF経路の優先順序

1. エリア内経路(intra area経路)
ルータLSA(LSタイプ1)
ネットワークLSA(LSタイプ2)
2. エリア間経路(inter area経路)
サマリLSA(LSタイプ3, 4)
3. 外部経路タイプ1
AS-external-LSA(LSタイプ5) メトリックタイプ1
4. 外部経路タイプ2
AS-external-LSA(LSタイプ5) メトリックタイプ2

Copyright (C) 2005 Internet Initiative Japan Inc.

104

OSPFv3

- OSPF for IPv6のこと
 - 詳しくは[RFC2740]
 - IPv6に対応するために、変更が加えられた
- トポロジ情報と経路情報の分離
 - ルータLSA、ネットワークLSAから経路を削除
 - ルータLSAのstubネットワーク
 - ネットワークLSAのネットマスク
 - 代わりにIntra-Area-Prefix-LSAを用意
- LSAにFlooding Scopeの要素が追加

Copyright (C) 2005 Internet Initiative Japan Inc.

105

OSPFv3

- LSAを分かりやすく改名
 - サマリLSA(LSタイプ3) Inter-Area-Prefix-LSA
 - サマリLSA(LSタイプ4) Inter-Area-Router-LSA
- リンクの識別手法の変更
 - OSPFv2 – 3種類
 - リンクタイプ
 - リンクID、リンクデータ
 - OSPFv3 – 4種類
 - リンクタイプ、
 - インタフェースID、ネイバインタフェースID、ネイバルータID

Copyright (C) 2005 Internet Initiative Japan Inc.

106

他のプロトコルとの比較

- OSPFとIS-IS
 - 1980年代の後半に開発が始まる
 - IS-ISのdraftが原型となり、OSPFが開発される
 - その後、各ベンダが実装
 - 現在もIETFのそれぞれのworking groupで議論が続く
- どちらもSPFアルゴリズムを利用してリンクステートデータベースから経路情報を計算する

Copyright (C) 2005 Internet Initiative Japan Inc.

107

IS-ISとOSPFv2

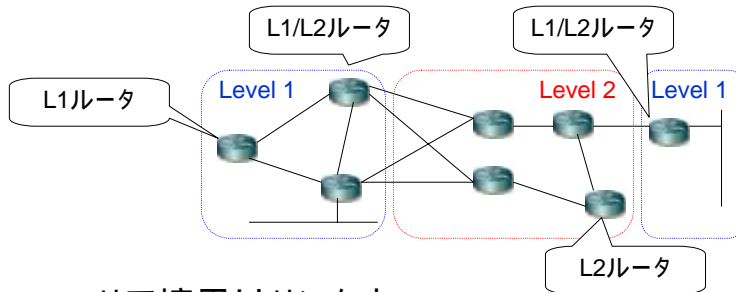
- IS-ISでIPの経路制御を出来るように拡張したものが、Integrated IS-IS。または Dual IS-IS

	Integrated IS-IS	OSPFv2
プロトコルパケットの転送	CLNS	IPv4
扱う経路情報	CLNS and/or IPv4	IPv4
階層化	Level1(エリア内) Level2(エリア間)	中継を担うarea0と その他のエリア
エリア境界	リンク	ルータ

Copyright (C) 2005 Internet Initiative Japan Inc.

108

IS-IS



- エリア境界はリンク上
- Level2がエリア間の通信を担う
- Level1からエリア外への通信は近隣のL1/L2ルータに頼る
 - L2の経路をL1内に伝播させることもできる[RFC2966]

Copyright (C) 2005 Internet Initiative Japan Inc.

109

BGP

Copyright (C) 2005 Internet Initiative Japan Inc.

110

BGP概要

- パスベクタ型プロトコル
 - プレフィックスに付加されたパス属性で経路制御
- AS番号によって、組織間、組織内を認識する
- 経路交換にTCPを利用
 - BGP接続のために事前に経路情報が必要
- 変更があった場合にのみ通知
 - ベスト経路のみを通知する
- 現在のバージョンは4 (BGP4)

Copyright (C) 2005 Internet Initiative Japan Inc.

111

BGP RFCs

- 基本
 - [RFC1771] A Border Gateway Protocol 4
- この他にもいっぱい
 - [RFC1997] BGP Communities Attribute
 - [RFC2796] BGP Route Reflection
 - [RFC3065] AS Confederations for BGP
- 伏兵
 - draft-ietf-idr-bgp4-26.txt
 - This document obsoletes RFC1771.

Copyright (C) 2005 Internet Initiative Japan Inc.

112

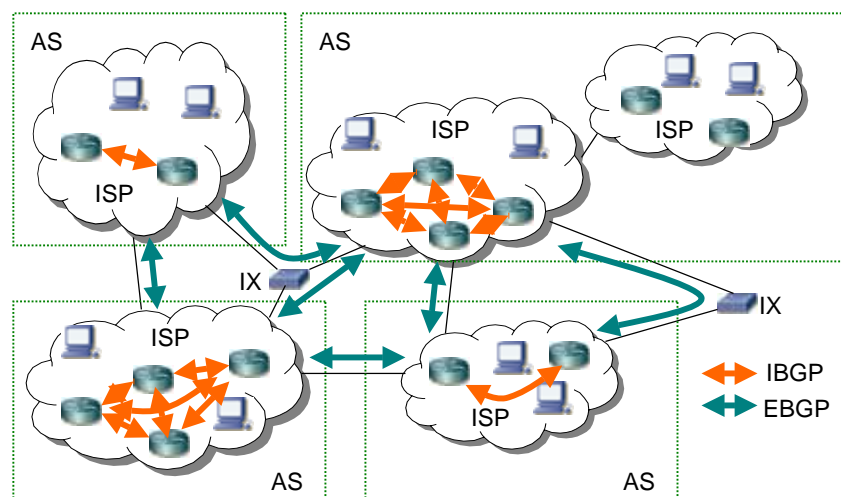
BGP用語

- BGP ID
 - ルータを識別する32bitの数値
 - インタフェースのIPアドレスから選ばれる
 - 実運用では変更が発生しないようにloopbackインタフェースに付与したIPアドレスを利用する
- NLRI
 - Network Layer Reachability Information
 - ネットワーク層到達可能性情報
 - prefixで示される

Copyright (C) 2005 Internet Initiative Japan Inc.

113

IBGPとEBGP

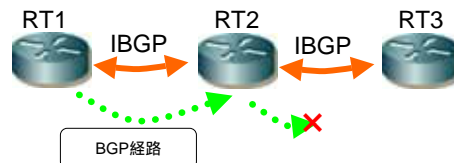


Copyright (C) 2005 Internet Initiative Japan Inc.

114

IBGP(Internal BGP)

- 同じAS内でのBGP接続
- IBGPで受信した経路は他のIBGPルータに
広報されない
 - 全ての経路を伝えるには、AS内の全BGPルータ
がfull-meshでIBGPを張る必要がある

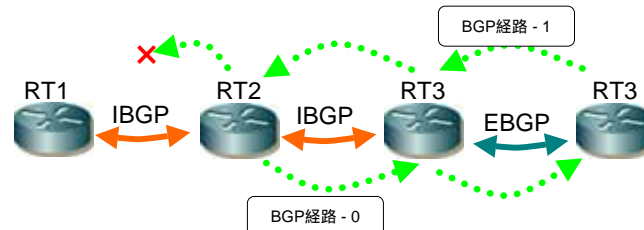


Copyright (C) 2005 Internet Initiative Japan Inc.

115

EBGP(External BGP)

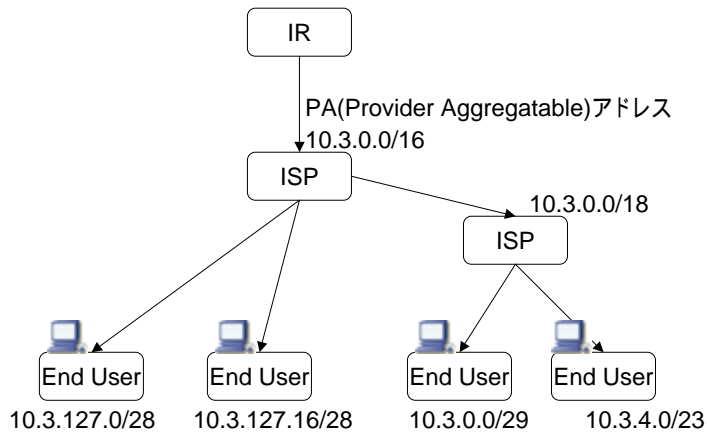
- 異なるASとのBGP接続
- EBGPから受信した経路は、他のBGPルータ
に広報する
 - IBGPから受信した経路もEBGPには広報する



Copyright (C) 2005 Internet Initiative Japan Inc.

116

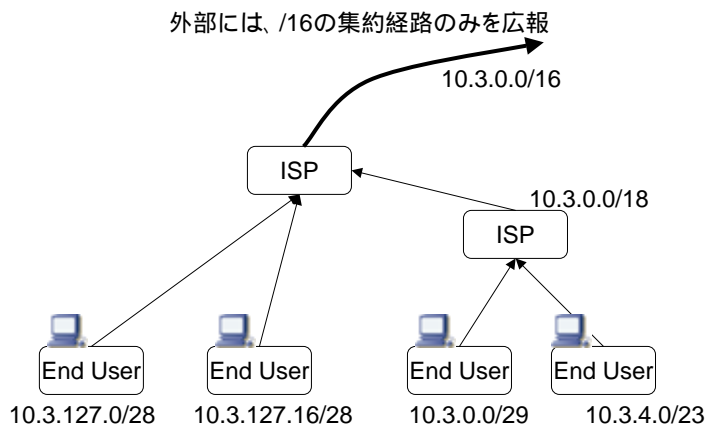
アドレス割り当ての階層化



Copyright (C) 2005 Internet Initiative Japan Inc.

117

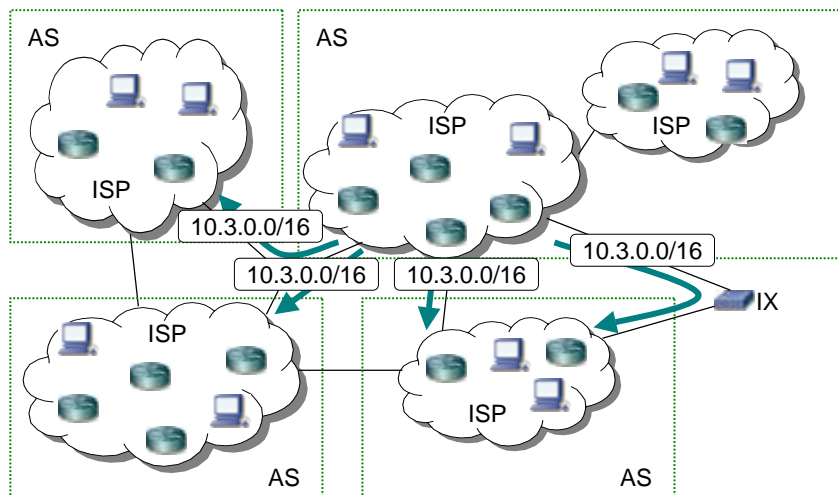
経路の集約



Copyright (C) 2005 Internet Initiative Japan Inc.

118

集約経路の広報



Copyright (C) 2005 Internet Initiative Japan Inc.

119

BGP4プロトコルパケット

IPヘッダ	IPヘッダ	IPヘッダ	IPヘッダ	IPヘッダ
TCPヘッダ	TCPヘッダ	TCPヘッダ	TCPヘッダ	TCPヘッダ
BGP OPEN	BGP UPDATE	BGP NOTIFICATION	BGP KEEPALIVE	BGP ROUTE-REFRESH
OPEN Message	WITHDRAWN Path Attr.+NLRI	Error Code		AFI+SAFI

Copyright (C) 2005 Internet Initiative Japan Inc.

120

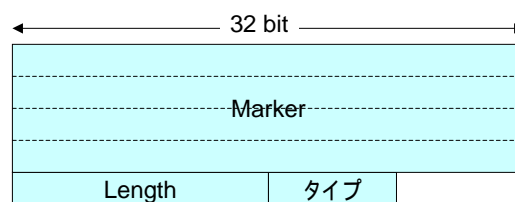
BGPパケット

BGPのプロトコルパケットの フォーマットを解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

121

BGP Message header



- Marker(マーカ)
 - 16-octetの全bitが1
 - 過去との互換性のため
- Length
 - 2-octetのメッセージ長
 - 19 ~ 4096
- タイプ (1-octet)
 1. OPEN
 2. UPDATE
 3. NOTIFICATION
 4. KEEPALIVE
 5. ROUTE_REFRESH

Copyright (C) 2005 Internet Initiative Japan Inc.

122

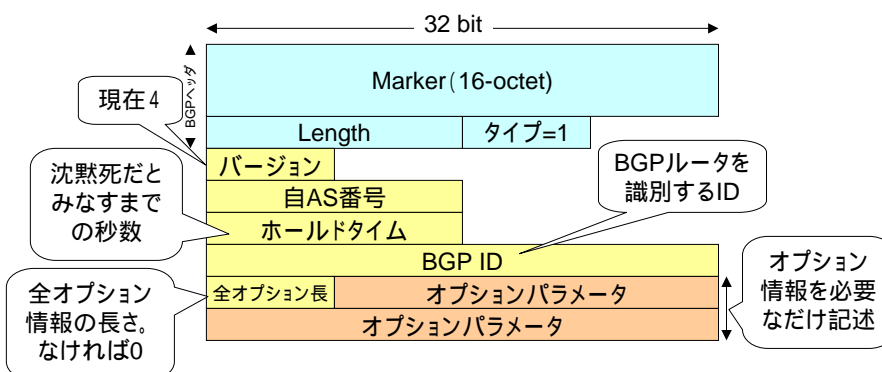
タイプ1 OPENメッセージ

- TCP接続が確立後、最初にやりとりされる
- パラメタの交換
 - バージョン、AS番号やBGP ID、ホールドタイム
 - オプションパラメータで各種機能の交換
- タイプ4 KEEPALIVEで接続確立

Copyright (C) 2005 Internet Initiative Japan Inc.

123

タイプ1 OPENメッセージ

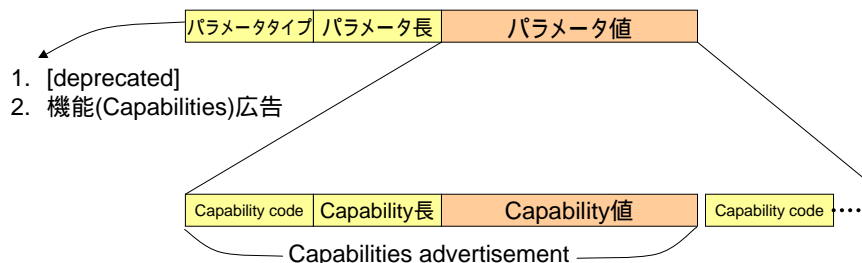


- ホールドタイムは0もしくは3以上
 - 小さな値が採用される
 - 0の場合、セッション維持にKEEPALIVEを利用しない

Copyright (C) 2005 Internet Initiative Japan Inc.

124

オプションパラメータフォーマット



- オプションパラメータはCapability広告で利用される
- Capability広告で自身の機能をピア先へ広報する

Copyright (C) 2005 Internet Initiative Japan Inc.

125

Capabilityコード

1 Multiprotocol Extension	サポートする<AFI, SAFI>の広告
2 Route Refresh	rfc版のRoute Refresh機能広告
3 Cooperative Route Filtering	
4 Multiple routes to a destination	
64 Graceful Restart	
65 Support for 4-octet AS number	
67 Support for Dynamic Capability	
128 Route Refresh(cisco)	Cisco独自のRoute Refresh機能広告

Copyright (C) 2005 Internet Initiative Japan Inc.

126

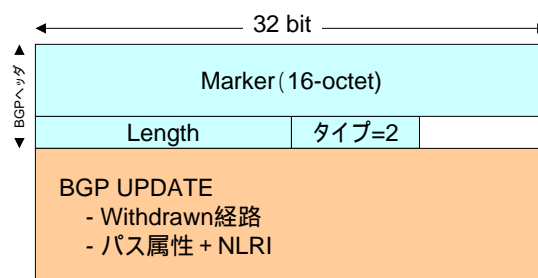
タイプ2 UPDATEメッセージ

- 経路情報を運ぶ
- 一つのメッセージで以下の情報を運べる
 - 複数のWithdrawn(取り消された)経路
 - 同じパス属性を持つ複数のNLRI
- 情報の伝播保証はTCPまかせ

Copyright (C) 2005 Internet Initiative Japan Inc.

127

タイプ2 UPDATEメッセージ



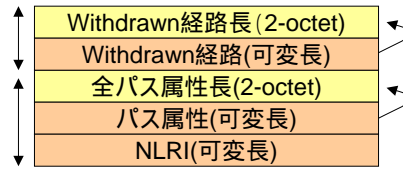
- Withdrawn経路に含まれる経路は、同じメッセージ中でNLRIに含まれてはならない
- パス属性が異なる経路は、異なるUPDATEメッセージで運ばれる

Copyright (C) 2005 Internet Initiative Japan Inc.

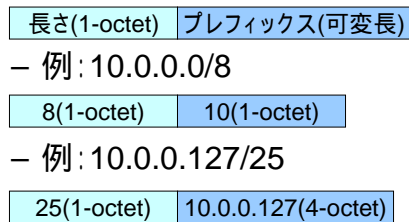
128

BGP UPDATEフォーマット

- Withdrawn経路
 - Withdrawnの長さ(2-octet)
 - Withdrawn経路の列挙
- 到達可能経路
 - 全パス属性の長さ(2-octet)
 - パス属性の列挙
 - NLRIの列挙



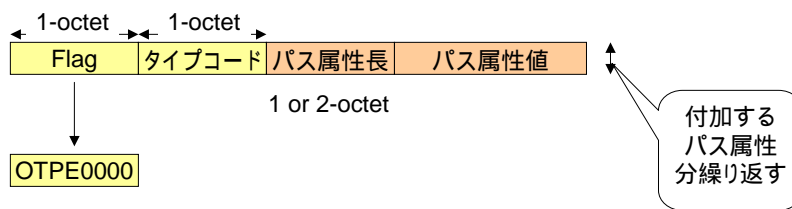
プレフィックスの格納形式



Copyright (C) 2005 Internet Initiative Japan Inc.

129

パス属性フォーマット



O bit: Optional(パス属性の種類別)

0=Wellknown, 1=optional

T bit: Transitive(パス属性の転送)

0=non-transitive, 1=transitive

P bit: Partial(パス属性の処理)

0=complete, 1=partial

E bit: Extended length

0=パス属性長は1-octet

1=パス属性長は2-octet

Partial bit

- オプション属性が、経路が広報されてから経由した全てのルータで解釈されたかどうかを示す
- 0:全てのルータで解釈された
- 1:解釈されなかったルータあり

Copyright (C) 2005 Internet Initiative Japan Inc.

130

パス属性の4つのカテゴリ

- 周知必須 - well-known mandatory [T]
 - 全てのBGPルータで解釈可能
 - NLRI情報があれば必ずパス属性に含まれる
- 周知任意 - well-known discretionary [T]
 - 全てのBGPルータで解釈可能
 - 必ずしも含まれない
- オプション通知 - Optional transitive [OT]
 - 一部のBGPルータでは解釈できないかもしれない
 - 解釈できなくても、そのまま他のルータに広報する
 - この際、Partial bitを1にセットする
- オプション非通知 - Optional non-transitive [O]
 - 一部のBGPルータでは解釈できないかもしれない
 - 解釈できない場合は、他のルータに広報するとき属性を削除する

Copyright (C) 2005 Internet Initiative Japan Inc.

131

パス属性値コードタイプ

1	ORIGIN	周知必須	経路の生成情報
2	AS_PATH	周知必須	経路が通過したASの情報
3	NEXT_HOP	周知必須	経路の宛先IPアドレス
4	MULTI_EXIT_DISC	オプション非通知	複数出口から経路選定する際の優先度
5	LOCAL_PREF	周知任意	経路の優先度
6	ATOMIC_AGGREGATE	周知任意	経路が途中で集約された情報
7	AGGREGATOR	オプション通知	経路集約を行ったルータ
8	COMMUNITIES	オプション通知	処理を行うための情報

Copyright (C) 2005 Internet Initiative Japan Inc.

132

パス属性値コードタイプ 続き

9	ORIGINATOR	オプション非通知	クラスタ内での経路生成ルータ
10	CLUSTER_LIST	オプション非通知	経路を反射したクラスタIDのリスト
14	MP_REACH_NLRI	オプション非通知	マルチプロトコルの到達可能経路
15	MP_UNREACH_NLRI	オプション非通知	マルチプロトコルの到達不可能経路

Copyright (C) 2005 Internet Initiative Japan Inc.

133

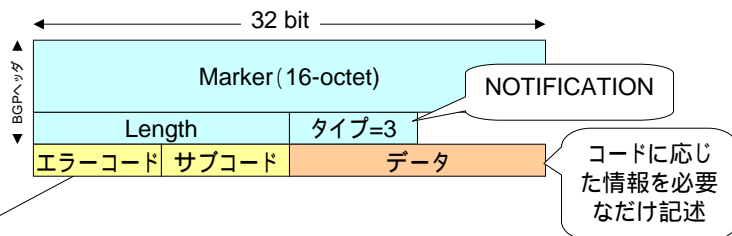
タイプ3 NOTIFICATIONメッセージ

- エラーを検出すると送信する
 - 送信後、すぐにBGP接続を切断する
- エラー内容がエラーコードとエラーサブコードで示される
 - 必要であれば、追加のデータも通知される

Copyright (C) 2005 Internet Initiative Japan Inc.

134

タイプ3 NOTIFICATIONメッセージ



1. メッセージヘッダエラー
2. OPENメッセージエラー
3. UPDATEメッセージエラー
4. HoldTime超過
5. 状態遷移エラー
6. Cease

Copyright (C) 2005 Internet Initiative Japan Inc.

135

コード1 メッセージヘッダエラー

- メッセージヘッダの処理中にエラーを検出



サブコード	エラー内容	データに含まれる値
1.	Markerの値が不正	
2.	Lengthの値が不正	不正だと判断したLengthの値
3.	解釈できないタイプ	解釈できなかったタイプの値

Copyright (C) 2005 Internet Initiative Japan Inc.

136

コード2 OPENメッセージエラー

- OPENメッセージの処理中にエラーを検出

2	サブコード	データ
---	-------	-----

サブコード	エラー内容	データに含まれる値
0.	特定なし	
1.	バージョン不一致	サポートする最も近いバージョン
2.	AS番号でエラー	
3.	BGP IDが不正	
4.	解釈できないオプションパラメータがある	
5.	[Deprecated]	
6.	ホールドタイム値が受け入れられない	
7.	未サポートのCapability	サポートしていないCapabilityコード

Copyright (C) 2005 Internet Initiative Japan Inc.

137

コード3 UPDATEメッセージエラー

- UPDATEメッセージの処理中にエラーを検出

3	サブコード	データ
---	-------	-----

サブコード	エラー内容	データに含まれる値
1.	アトリビュートが不正	
2.	周知必須属性が解釈できなかった	エラーを検出した属性(TLV)
3.	あるべき周知必須属性が無かった	無かった周知必須属性のタイプコード
4.	フラグが不正	エラーを検出した属性(TLV)
5.	パス属性長が不正	エラーを検出した属性(TLV)
6.	ORIGIN属性値が未規定の値	エラーを検出した属性(TLV)
7.	[Deprecated]	
8.	NEXT_HOP属性値の書式が不正	エラーを検出した属性(TLV)
9.	オプション属性値でエラー	エラーを検出した属性(TLV)
10.	NLRIの書式が不正	
11.	AS_PATH属性の書式が不正	

Copyright (C) 2005 Internet Initiative Japan Inc.

138

コード4 HoldTimer超過

- HoldTimer期間中に、UPDATEもKEEPALIVEも受信しなかった

4	サブコード	データ
---	-------	-----

Copyright (C) 2005 Internet Initiative Japan Inc.

139

コード5 状態遷移エラー

- 予期せぬイベントが発生

5	サブコード	データ
---	-------	-----

Copyright (C) 2005 Internet Initiative Japan Inc.

140

コード6 Cease

- その他のエラーを検出

6	サブコード	データ
---	-------	-----

サブコード	エラー内容	データに含まれる値
1.	最大受信経路数に到達	<AFI(2), SAFI(1), prefix上限値(4)>
2.	Administrative Shutdown	
3.	設定削除	
4.	Administrative Reset	
5.	接続拒否	
6.	その他の設定変更	
7.	接続競合の解決	
8.	リソース不足	

Copyright (C) 2005 Internet Initiative Japan Inc.

141

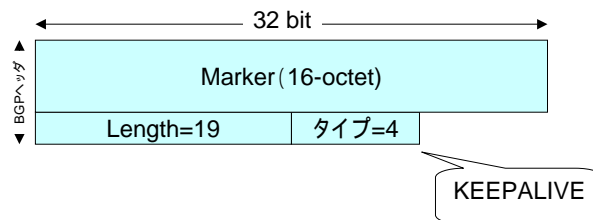
タイプ4 KEEPALIVEメッセージ

- BGP接続を確立させる
- BGP接続を維持する
 - 送信間隔内にUPDATEが無ければ送信
 - 送信間隔はホールドタイムの1/3程度
 - 最小で1秒
 - ホールドタイムが0の場合は、送信されない

Copyright (C) 2005 Internet Initiative Japan Inc.

142

タイプ4 KEEPALIVEメッセージ



- KEEPALIVEであること以外、何も運ばない
- 最小のBGPメッセージ

Copyright (C) 2005 Internet Initiative Japan Inc.

143

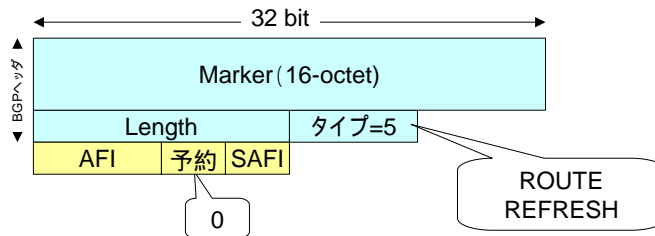
タイプ5 ROUTE-REFRESHメッセージ

- 全経路の再広報を依頼する
 - <AFI, SAFI>を指定 (IPv4 unicastなど)
- 受信時、知らない<AFI, SAFI>であれば無視
- メッセージを送信するには、OPENメッセージのCapability広告でROUTE_REFRESH機能が通知されている必要がある

Copyright (C) 2005 Internet Initiative Japan Inc.

144

タイプ5 ROUTE-REFRESHメッセージ

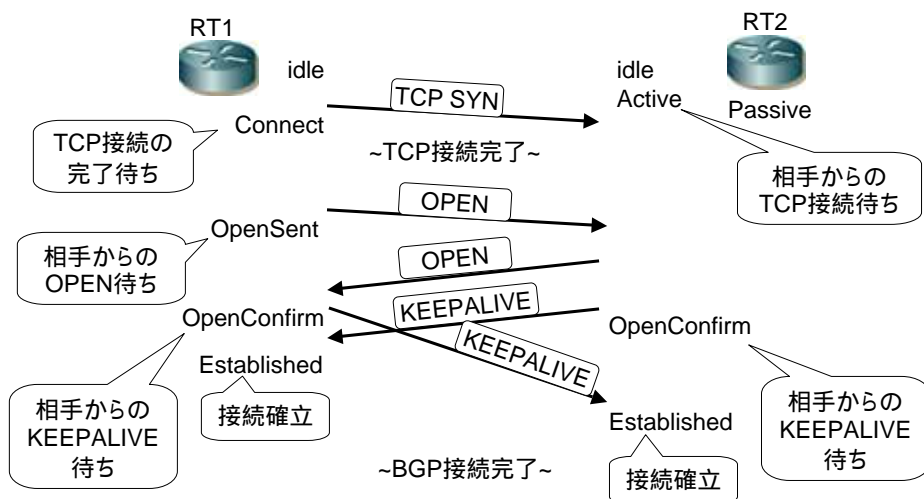


- AFI = Address Family Identifier
 - IPv4やIPv6など
- SAFI = Subsequent Address Family Identifier
 - UnicastやMulticastなど

Copyright (C) 2005 Internet Initiative Japan Inc.

145

BGP接続の確立までの例



Copyright (C) 2005 Internet Initiative Japan Inc.

146

パス属性の処理

経路の送受信時のパス属性の処理
を解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

147

EBGP & IBGPとパス属性

パス属性	EBGP	IBGP
ORIGIN	必須	必須
AS_PATH	必須	必須
NEXT_HOP	必須	必須
MULTI_EXIT_DISC	任意	任意
LOCAL_PREF	不許可	付加すべき
COMMUNITIES	任意	任意

Copyright (C) 2005 Internet Initiative Japan Inc.

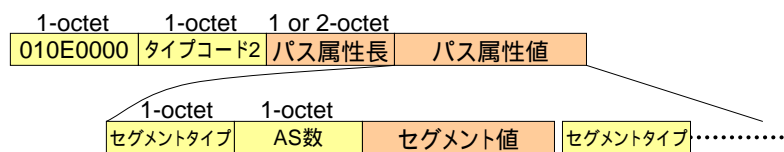
148

ORIGIN属性値

- NLRIの起源を示す3つのタイプ
- 経路生成元で付加され、その後変更されない

0 – IGP … AS内部にある
 1 – EGP … EGP[RFC904]で学習した
 2 – INCOMPLETE … その他の方法で学習した

AS_PATH属性フォーマット



セグメントタイプ

- 1: AS_SET
 UPDATEが経由したAS番号。順序は意味を持たない
 異なるAS Pathの経路を集約したときに生成される。
 Internetではあまり見かけない

- 2: AS_SEQUENCE
 UPDATEが経由したAS番号。順序に意味がある
 経由した最新のAS番号はセグメント値の一番左

AS数

octet数ではなく、AS数
 つまり、255個のASまで

セグメント値

2-octetのAS番号のリスト

- 新しい情報は先頭(左)に付加される

AS_PATH属性の処理

- 経路を転送する場合

広報先	
IBGP	変更しない
EBGP	自AS番号をAS_SEQUENCEタイプでAS_PATH属性の先頭に付加する

- 経路を生成する場合

広報先	
IBGP	空のAS_PATH属性を生成する
EBGP	AS_SEQUENCEタイプで自AS番号のみのAS_PATH属性を生成する

Copyright (C) 2005 Internet Initiative Japan Inc.

151

NEXT_HOP属性の処理

- IBGPに経路を転送するときは
 - 変更しない
 - ただし、設定で自身のIPアドレスに変更することも可能
- IBGPに生成した経路を広報するときは
 - その宛先に到達するためのネクストホップを設定する
 - ただし、自身のIPアドレスを設定することも可能
- EBGPに経路を広報するときは
 - BGP接続に利用している自身のIPアドレスを設定する
 - ただし、宛先のネクストホップがEBGPルータと共通のサブネットに属する場合は、他のルータのIPアドレスや自身の別なインタフェースのIPアドレスを設定することも可能

Copyright (C) 2005 Internet Initiative Japan Inc.

152

COMMUNITIES属性

- 32bitの数値で情報を伝える
 - この数値を元に、予め実装した処理を適用する
 - 経路生成時や経路の送受信時に付加、変更、削除される
- 4つのWell-Known-community
 - NO_EXPORT(0xFFFFFFFF01)
 - 他ASに広報しない
 - コンフェデレーション内のメンバASには広報する
 - NO_ADVERTISE(0xFFFFFFFF02)
 - 他BGPルータに広報しない
 - NO_EXPORT_SUBCONFED(0xFFFFFFFF03)
 - 他ASに広報しない
 - コンフェデレーション内でメンバASにも広報しない
 - NOPEER(0xFFFFFFFF04) [RFC3765]
 - 対等ピアには広報しない
 - まだ実装は無さそう

Copyright (C) 2005 Internet Initiative Japan Inc.

153

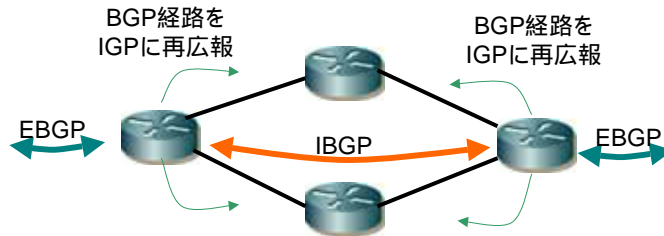
IBGPの構成

IBGPの構成を解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

154

BGPのいにしえのモデル



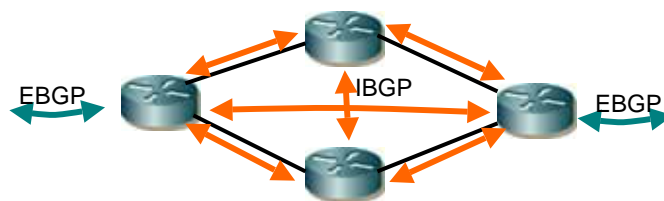
- EBGPルータのみがBGPルータとなる
- BGP経路をIGP(OSPFやIS-IS)に再広報してAS内部はIGPで経路制御

…経路数が増大すると破綻

Copyright (C) 2005 Internet Initiative Japan Inc.

155

経路数の増大に対応したBGPモデル



- 主要なルータは全てBGPルータ
- IGPはトポロジを運び、BGPで経路を運ぶ

…IBGP接続の増大

Copyright (C) 2005 Internet Initiative Japan Inc.

156

IBGP full-mesh $n*(n-1)/2$

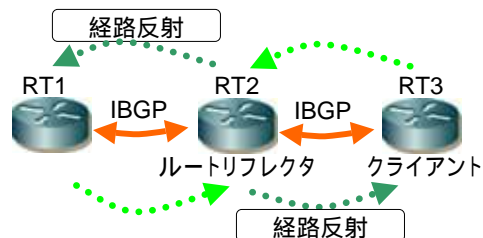
- AS内にBGPルーターが増える毎にIBGP接続が増大していく
 - 20台目のBGPルーターが接続すると19接続追加
 - ルータリソースの問題、設定負荷の問題
- 解決策の模索
 - [RFC2796] ルートリフレクタ
 - [RFC3065] コンフェデレーション
 - 気にせずリソースを強大にする
 - ルーターを減らす

Copyright (C) 2005 Internet Initiative Japan Inc.

157

ルートリフレクタ

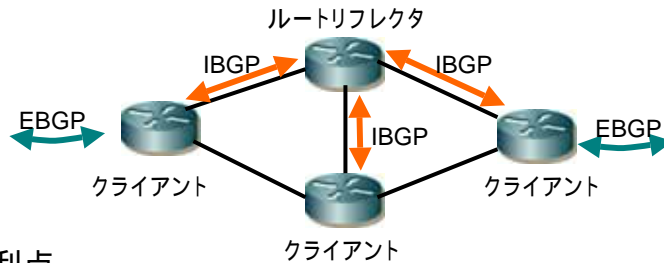
- IBGPで受信した経路の転送ルールを変更
- ルートリフレクタの機能
 - BGP接続ごとに設定される
 - クライアント以外のIBGPで受信した経路をクライアントに送信
 - クライアントから受信した経路を他のIBGPルーターに送信
- ベスト経路のみを広報するルールは変わらない



Copyright (C) 2005 Internet Initiative Japan Inc.

158

ルータリフレクタの利点と欠点



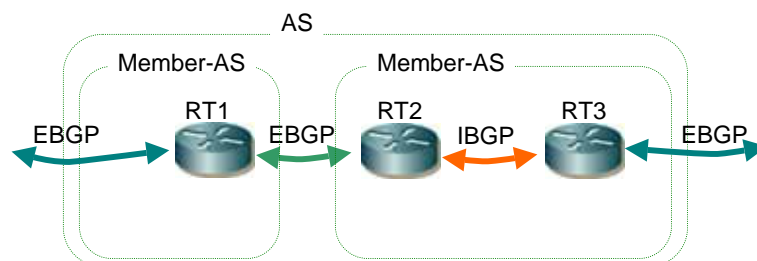
- 利点
 - IBGP接続数が削減できる
 - 比較的容易に導入できる
- 欠点
 - 経路削除時に、UPDATEが増える可能性がある
 - 経路情報が隠蔽されるため最適ではない経路を選ぶ可能性がある
 - リフレクタの階層はできるだけ物理トポロジに合わせるべし!

Copyright (C) 2005 Internet Initiative Japan Inc.

159

コンフェデレーション

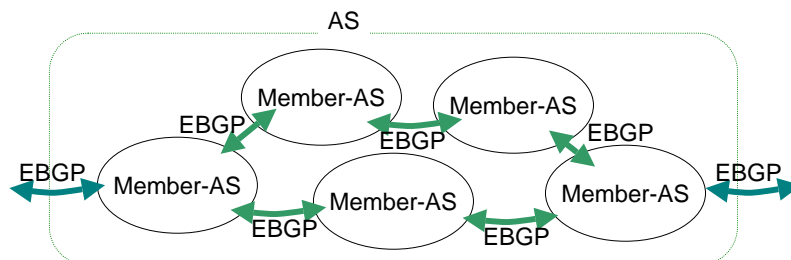
- 外部からは一つのASのままだが、内部を複数のメンバASで構成する
- メンバAS間のBGP接続はEBGPに似た挙動をする
- メンバASにはプライベートASを使うのが一般的



Copyright (C) 2005 Internet Initiative Japan Inc.

160

コンフェデレーションの利点と欠点



- 利点
 - IBGP接続数が削減できる
 - 管理区分を分けられる
- 欠点
 - 経路削除時にUPDATEが増える可能性がある
 - 経路情報が隠蔽されるため最適ではない経路を選ぶかもしれない

Copyright (C) 2005 Internet Initiative Japan Inc.

161

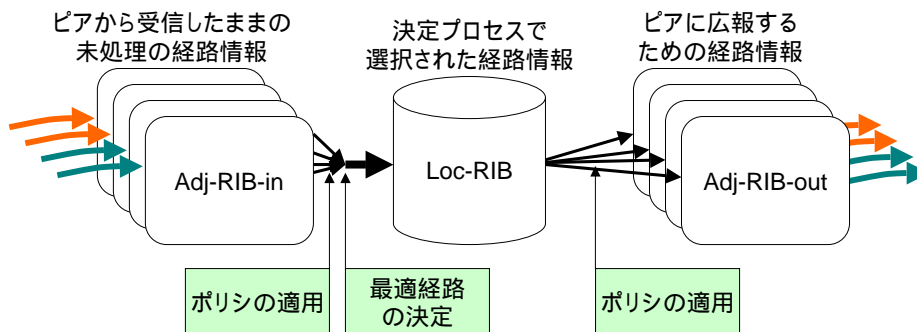
BGPの経路選択

経路処理方法や、経路選択ルール
を解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

162

BGPの経路処理



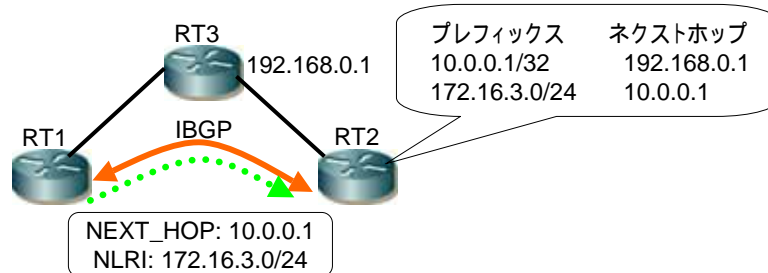
- ポリシは設定/実装依存
- 無理なポリシーを適用すると、経路Loopを引き起こす可能性があるので注意

Copyright (C) 2005 Internet Initiative Japan Inc.

163

NEXT_HOP解決

- NEXT_HOPのIPアドレスは(IGPで)到達できる必要がある
 - IGPで解決するという制限は、現在のdraftからは削除されている



Copyright (C) 2005 Internet Initiative Japan Inc.

164

経路優先度

1	NEXT_HOP RESOLVABLE	NEXT_HOPが到達可能
2	LOCAL_PREF	LOCAL_PREF属性値が大きい経路を優先
3	AS_PATH	AS_PATH属性に含まれるAS数が少ない経路を優先 (AS_SETタイプは幾つASを含んでも1として数える)
4	ORIGIN	ORIGIN属性の小さい経路を優先 (IGP < EGP < INCOMPLETE)
5	MULTI_EXIT_DISC	同じASからの経路はMED属性値が小さな経路を優先 (MED属性が付加されていない場合は、最小(=0)として扱う)
6	PEER_TYPE	IBGPよりもEBGPで受信した経路が優先
7	NEXT_HOP METRIC	NEXT_HOPへの内部経路コストが小さい経路が優先 (コストが算出できない経路がある場合は、この項目をスキップ)
8	BGP_ID	BGP IDの小さなBGPルータからの経路が優先 (ORIGINATOR_IDがある場合は、これをBGP IDとして扱う)
9	CLUSTER_LIST	CLUSTER_LISTの短い経路が優先
10	PEER_ADDRESS	ピアアドレスの小さなBGPルータからの経路を優先

draft-ietf-idr-bgp4-26.txt + draft-ietf-idr-rfc2796bis-02.txt の簡略版

Copyright (C) 2005 Internet Initiative Japan Inc.

165

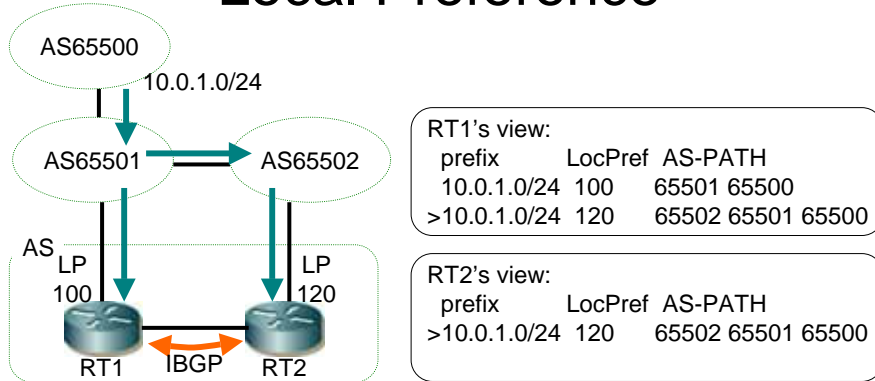
属性値の評価

属性値がどう評価されるかを
解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

166

Local Preference

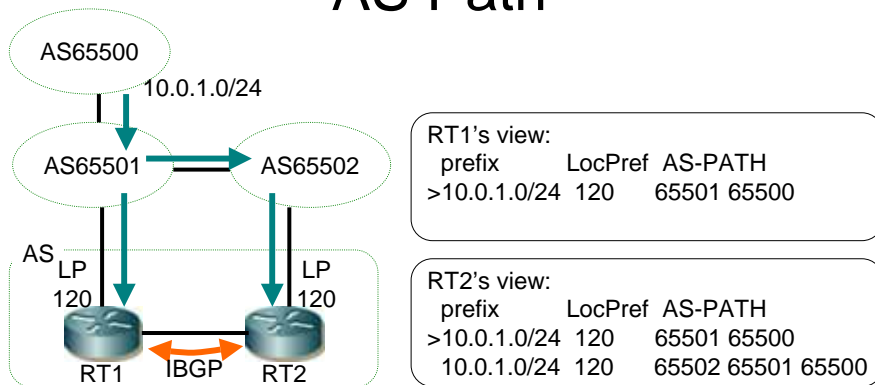


- Local Preferenceの大きな値が優先
– EBGPからの受信時 or 経路の生成時に設定

Copyright (C) 2005 Internet Initiative Japan Inc.

167

AS Path

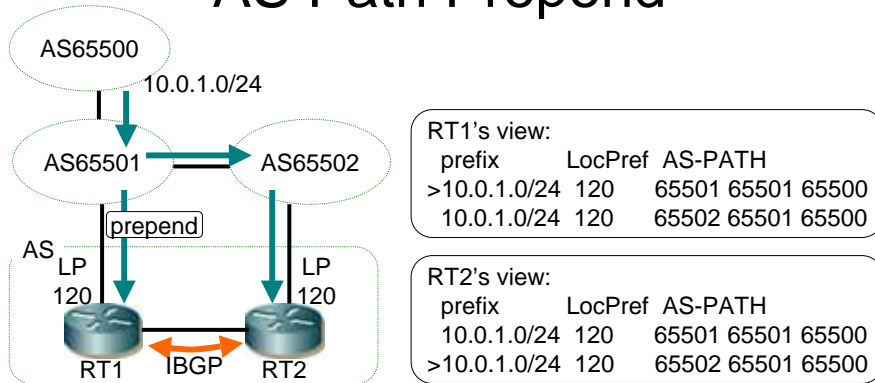


- Local Preferenceの大きさが同じなら、
AS_Path長の短い経路が優先

Copyright (C) 2005 Internet Initiative Japan Inc.

168

AS Path Prepend

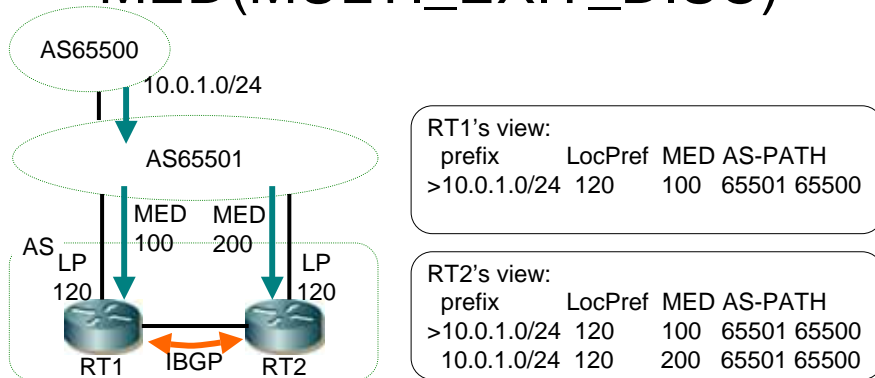


- 優先度を変化させるために、EBGPに広報する際にAS Pathに自ASを追加で付加する

Copyright (C) 2005 Internet Initiative Japan Inc.

169

MED(MULTI_EXIT_DISC)

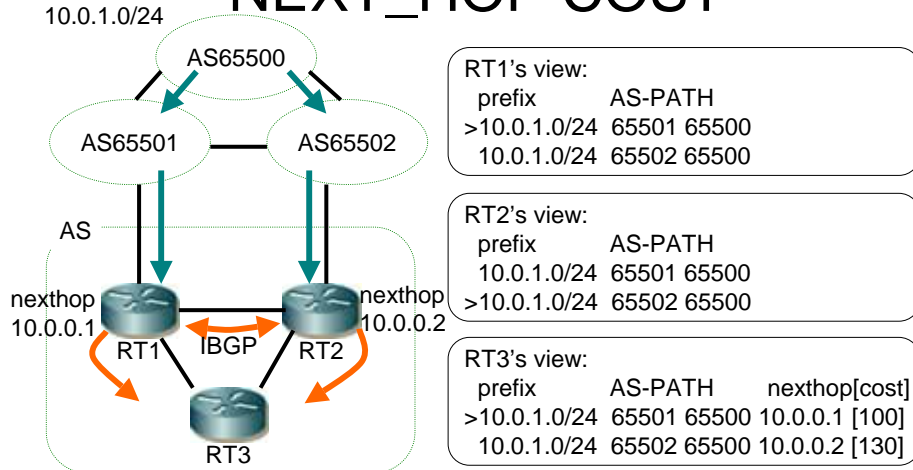


- MEDの値が小さな経路が優先
 - MEDはEBGPに広報時に設定
 - 経路受信時に設定することも可能

Copyright (C) 2005 Internet Initiative Japan Inc.

170

NEXT_HOP COST



- NEXT_HOPへのコストが小さい経路が優先

Copyright (C) 2005 Internet Initiative Japan Inc.

171

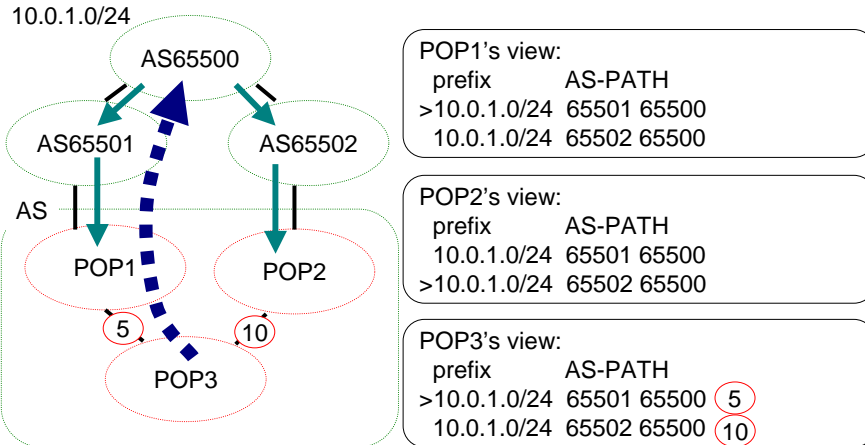
BGPのパス選択

OSPFとBGPの関わりを
解説する

Copyright (C) 2005 Internet Initiative Japan Inc.

172

OSPFとBGP

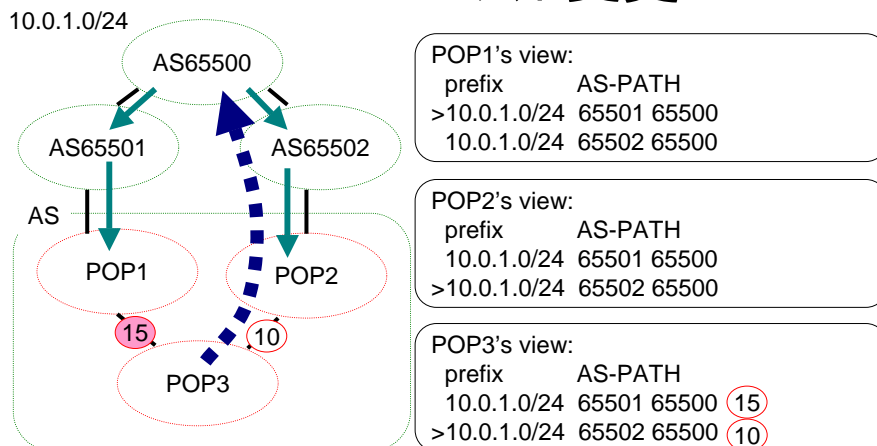


- POP3では、65501 (POP1経由) を選択中

Copyright (C) 2005 Internet Initiative Japan Inc.

173

OSPFのコスト変更

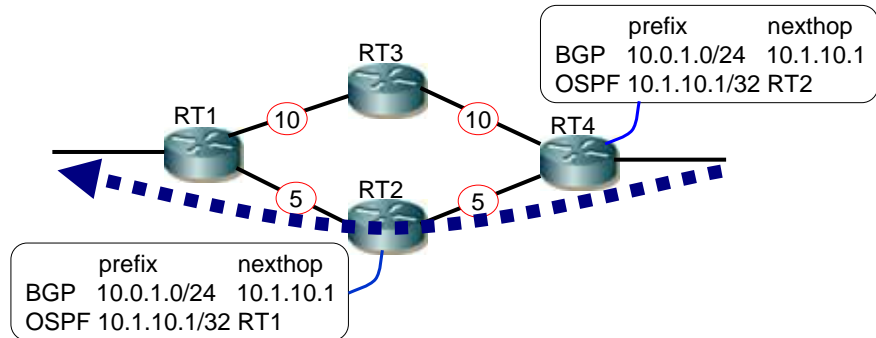


- POP3は65502 (POP2経由) に変更

Copyright (C) 2005 Internet Initiative Japan Inc.

174

OSPFコストとBGP

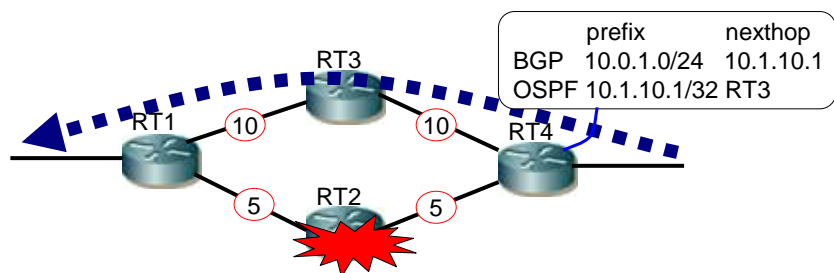


- BGPネクストホップへのOSPFコストが一番小さな経路が選択される

Copyright (C) 2005 Internet Initiative Japan Inc.

175

RT2が突然reload

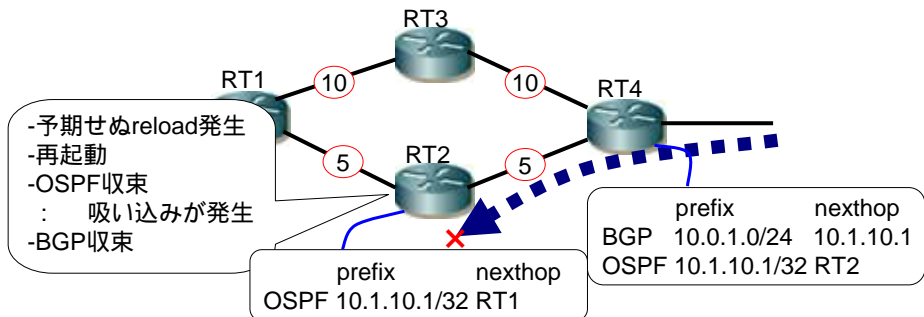


- RT2が再起動中・・・
- 他のルータが障害を検出し、OSPF再計算
- トラヒックはRT3を迂回している

Copyright (C) 2005 Internet Initiative Japan Inc.

176

OSPFとBGPの収束時間が違う

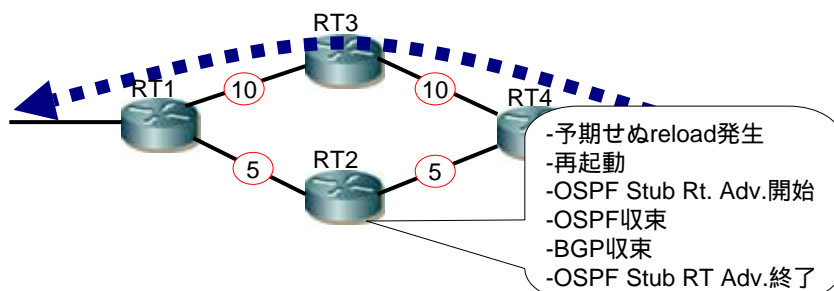


- OSPFはOK/完了 しかーし BGPは未だ
- BGPの経路が無いよー---
- その間、RT2がトラフィックを吸い込んでしまう

Copyright (C) 2005 Internet Initiative Japan Inc.

177

OSPF Stub Router Advertisement [RFC3137]



- ルータを経由するトラフィックを迂回させる機能
- OSPF起動後に実施して、BGP収束までトラフィックを迂回させる等の利用が考えられる

Copyright (C) 2005 Internet Initiative Japan Inc.

178

BGP4+

- BGP4のマルチプロトコル(IPv6)対応
 - [RFC2545] [RFC2858]
- OPENメッセージでマルチプロトコル対応を通知
- BGPセッションはIPv4 or IPv6どちらでも可
 - IPv6だと global unicast or link-localが選べる
 - IPv6の到達性を保証するには、IPv6でセッションを確立するのがお勧め
- NEXT_HOPは global unicast (+ link-local)
 - プレフィックスと共にMP_REACH_NLRIで運ばれる

Copyright (C) 2005 Internet Initiative Japan Inc.

179

BGPの転用

- BGPは、ルータにTCPで情報を通知できる
 - ルーティング以外の用途に転用されまくり
- パス属性で情報を運ぶ
 - IPv6経路等もパス属性で運ばれる
 - パス属性のみでNLRIが無いUPDATEも有効
- 経路を運ぶ以外の目的にも利用されるようになってきた。

Copyright (C) 2005 Internet Initiative Japan Inc.

180

おわり



Copyright (C) 2005 Internet Initiative Japan Inc.

181