

# T4 ルーティングチュートリアル

## AS内制御 IGP – OSPF –

IIJ 津辻 文亮

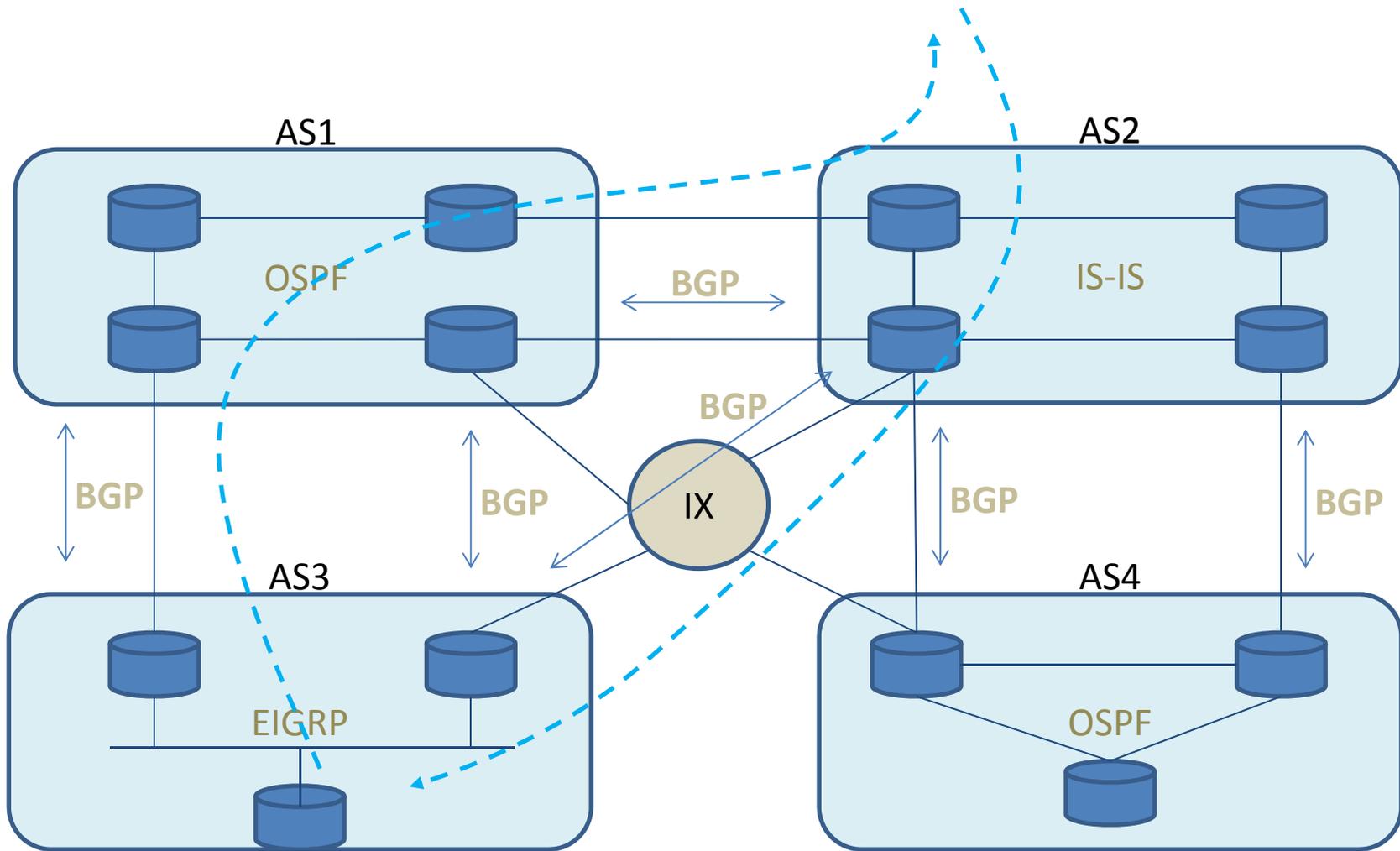
<tsutsuji@iij.ad.jp>

# IGP/EGP

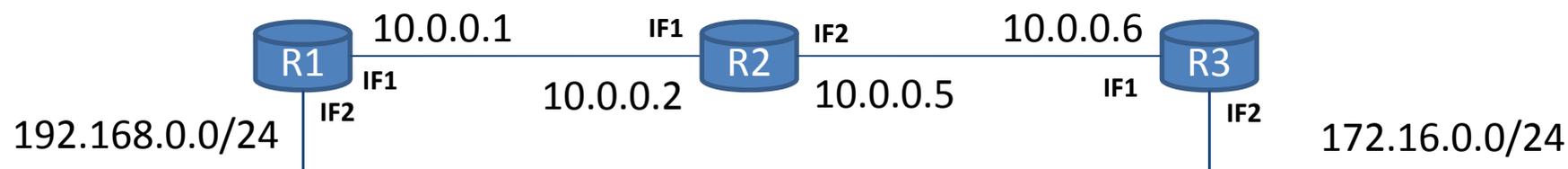
- EGP(Exterior Gateway Protocol)
  - AS間とのやり取りに利用する経路制御プロトコル
  - 例: BGP
- IGP(Interior Gateway Protocol)
  - AS内で利用する経路制御プロトコル
  - 例: OSPF/IS-IS/EIGRP/RIP

IGP		EGP
ディスタンスベクタ	リンクステート	パスベクタ
RIP	OSPF / IS-IS	BGP

# インターネット



# 経路表(Routing Table)



R1の経路表

宛先	ネクストホップ	学習元
192.168.0.0/24	IF2	direct/connected
10.0.0.0/30	IF1	direct/connected
10.0.0.4/30	10.0.0.2	Dynamic routing
172.16.0.0/24	10.0.0.2	Dynamic routing

- 経路表は宛先をどこに送るかの集合
- 各々のルータが経路表を作成
- 直接つながっていないものを色々なルーティングプロトコルから学習する
- 各々が経路表を作成し、HOP-by-HOPでパケットを送り届ける

# IGP Routing

- 静的ルーティング
  - Connected / Direct: 直接接続
  - Static: 静的な安定したルーティング
- 動的ルーティング
  - ディスタンスベクタ型 (RIP)
    - 自分のルーティングテーブルを交換
    - 定期的に情報を交換
  - リンクステート型 ( OSPF / IS-IS )
    - 自分のインターフェース情報を交換
    - 変更があった場合に差分を即座に更新

# 宛先の検索

- Longest Match
  - 最長一致。
  - 192.168.0.0/27 > 192.168.0.0/24 >> 0.0.0.0/0
  - 2001:db8::/64 > 2001:db8::/48 >> ::/0
- Protocol による優先度(ベンダ特有)
  - 同じprefixの場合はprotocol優先度で勝負
  - ex. connected > ospf > bgp

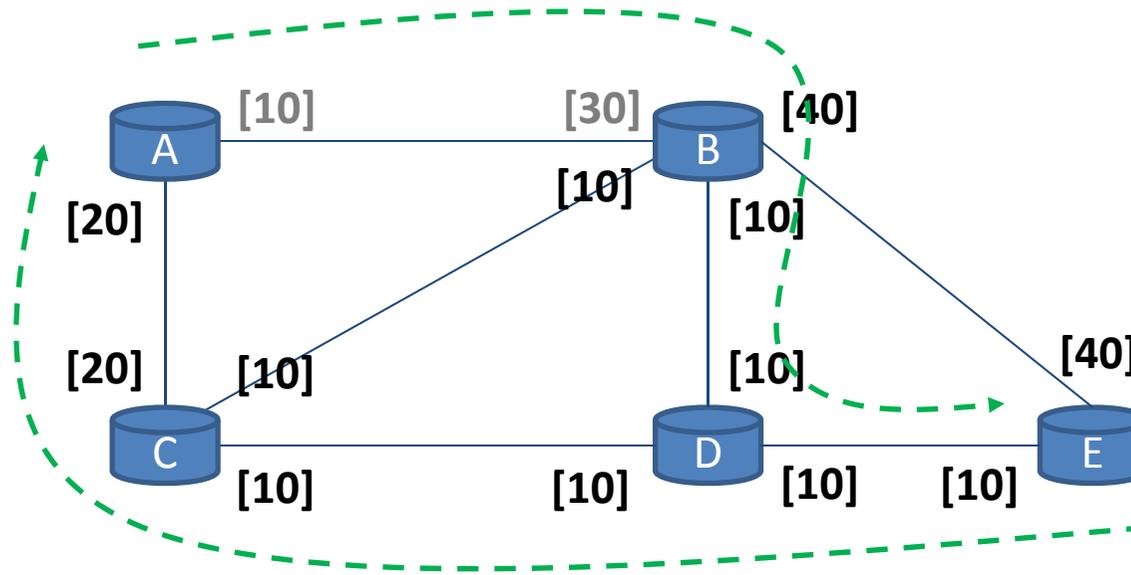
OSPF

# OSPF 概要

- リンクステート型
  - 各ルータが持つリンクの情報を交換
  - 変更があった場合に即座に通知
- SPFを用いて最短経路を探索
  - リンクステート集合から各々がトポロジーを作成
- 規模に合わせたエリア構成
  - エリア毎の計算による負荷低減
  - 2階層のエリア構成



# リンクコストとトラフィック



■ リンクコストには向きがあり、コストによっては、行きと帰りが異なる

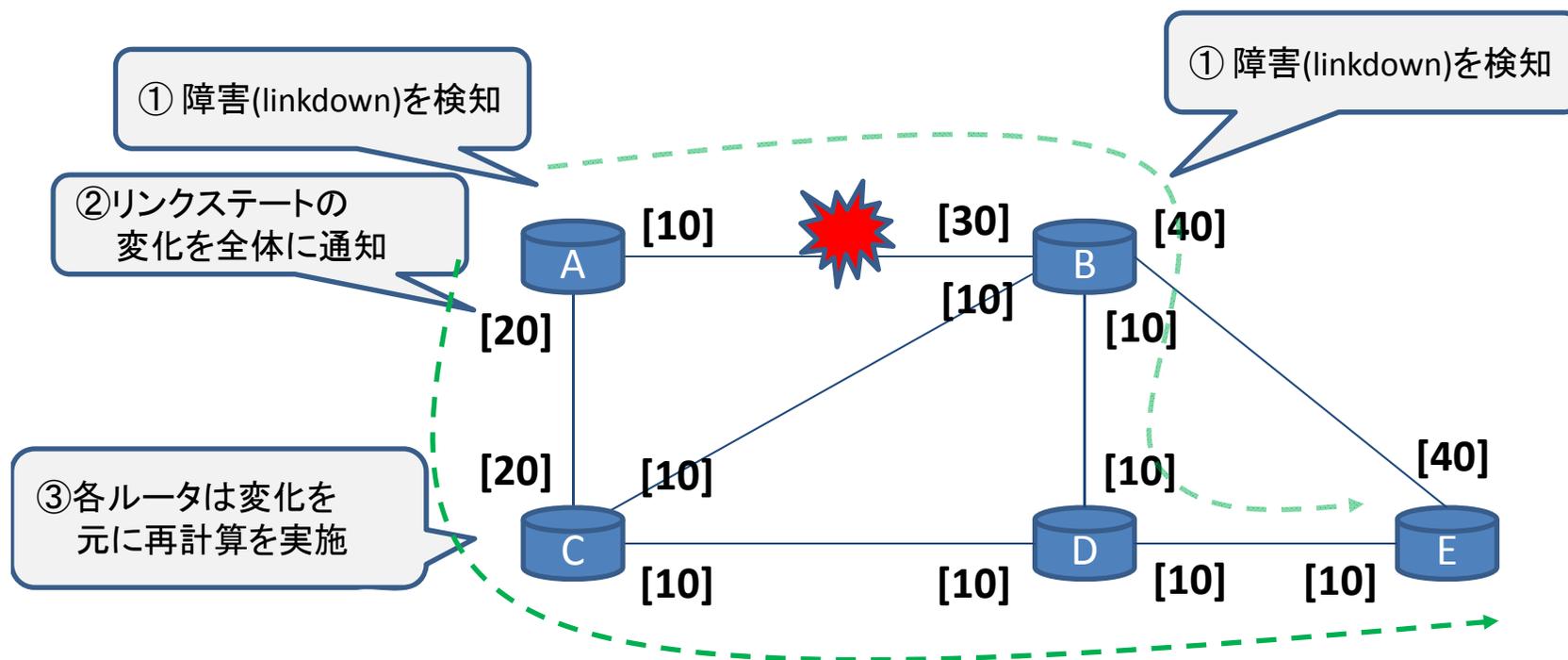
A → B → D → E      Cost = 30

E → D → C → A      Cost = 40

# LSA/LSDB

- LSA(Link state Advertisement)
  - OSPFで言うところのリンクステート情報
  - リンク情報やネットワーク情報
  - 情報に応じていくつかのタイプが存在する
  - 各ルータが皆に広告する
- LSDB(Link state DataBase)
  - 各々のルータが広告したLSAを使って構築するDB
  - LSDBを用いて最短パスを決定

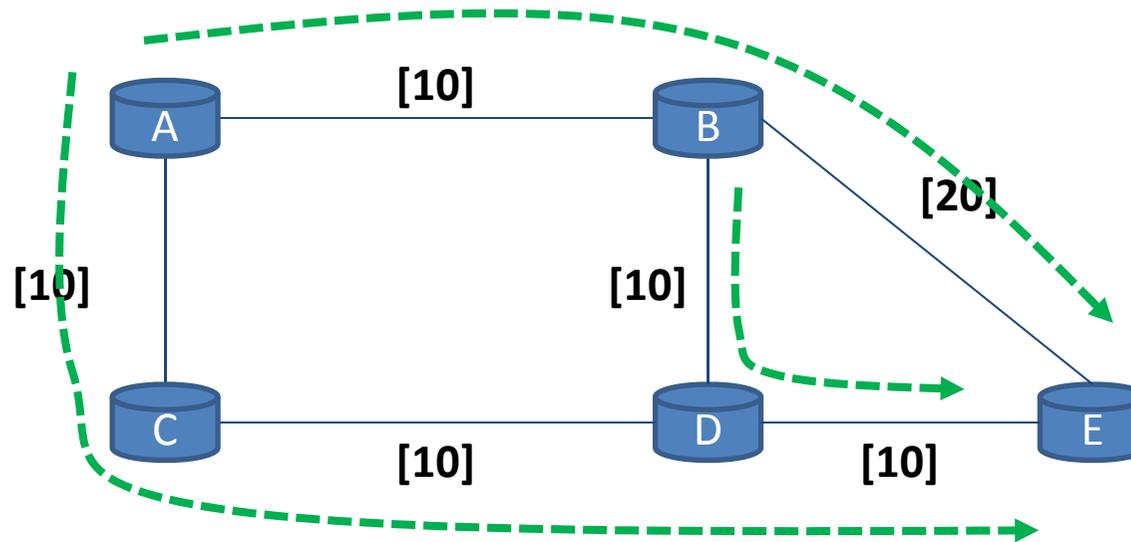
# 障害時の動作



- リンク状態を配布するルータが変更を通知
- 障害時には再計算して自動で迂回がかかる

# ロードバランス

## ECMP(Equal-Cost Multi-Path)



■ 最短コストが複数ある場合には、トラフィックを分散させることができる

A -> B -> E = Cost 30 nexthop B

A -> B -> D = Cost 30 nexthop B

A -> C -> D -> E = Cost 30 nexthop C

# OSPFパケットタイプ

- Type1: Hello パケット
  - neighborの発見、DR/BDR選出、生存確認で利用
- Type2: DBD(データベース記述)パケット
  - データベースのサマリを交換するのに利用
- Type3: LSR(リンク状態要求)パケット
  - Link Stateを要求する場合に利用
- Type4: LSU(リンク状態更新)パケット
  - Link Stateの更新時に利用
- Type5: LSAck(リンク状態確認応答)パケット
  - LSAの受信確認に利用

定常時  
でも利用

初期の  
DB同期に利用

状態変化時  
に利用

受信確認  
に利用

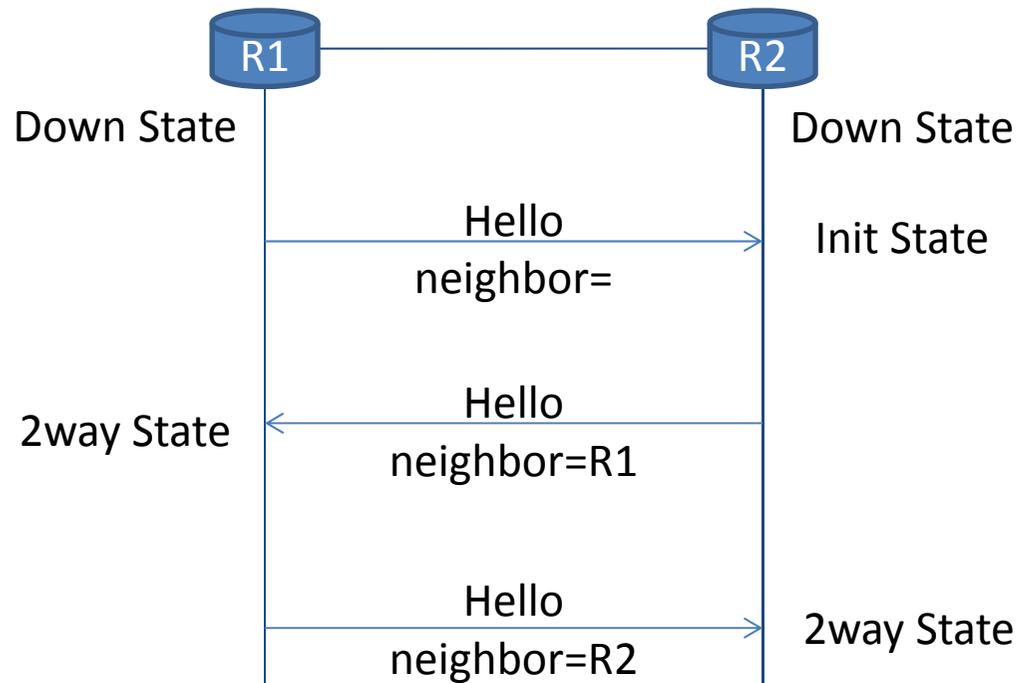
# OSPFのながれ

- ネイバー(Neighbor)の確立
- DR/BDRの選出( 必要な場合 )
- アジャセンシ(Adjacency)の確立
- LSDBの構築と最短パスツリーの構築
- ルーティングテーブル反映

# NeighborとAdjacency

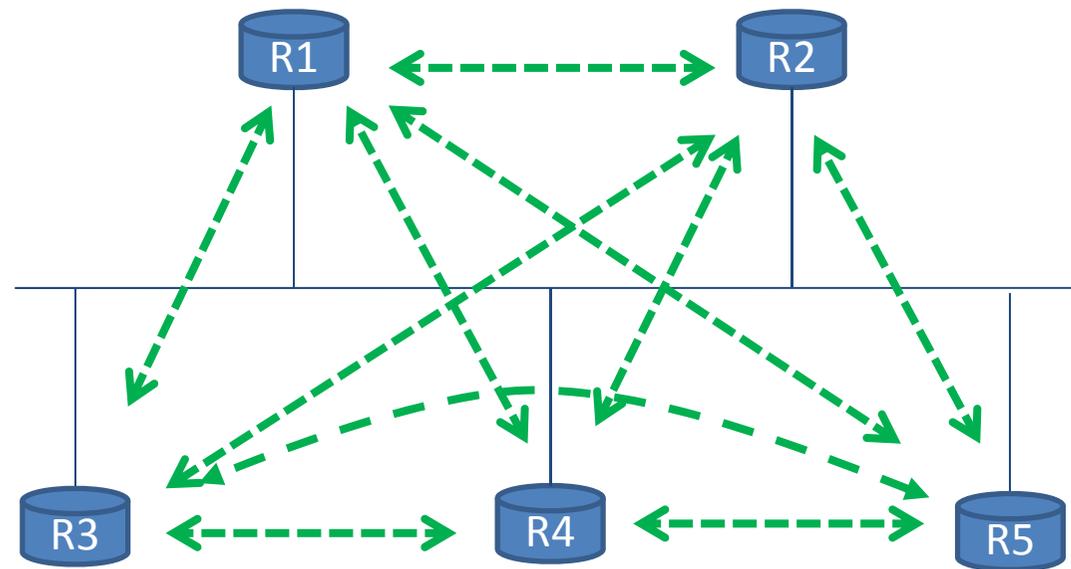
- Neighbor
  - OSPF リンク上に存在する隣接した関係
  - OSPF Helloを送り合い隣接を発見
- Adjacency
  - LSA/リンクステートを交換し合う関係
  - Adjacencyを確立した同士でLSDBの同期を実施

# Neighborの確立





# DR/BDR とは



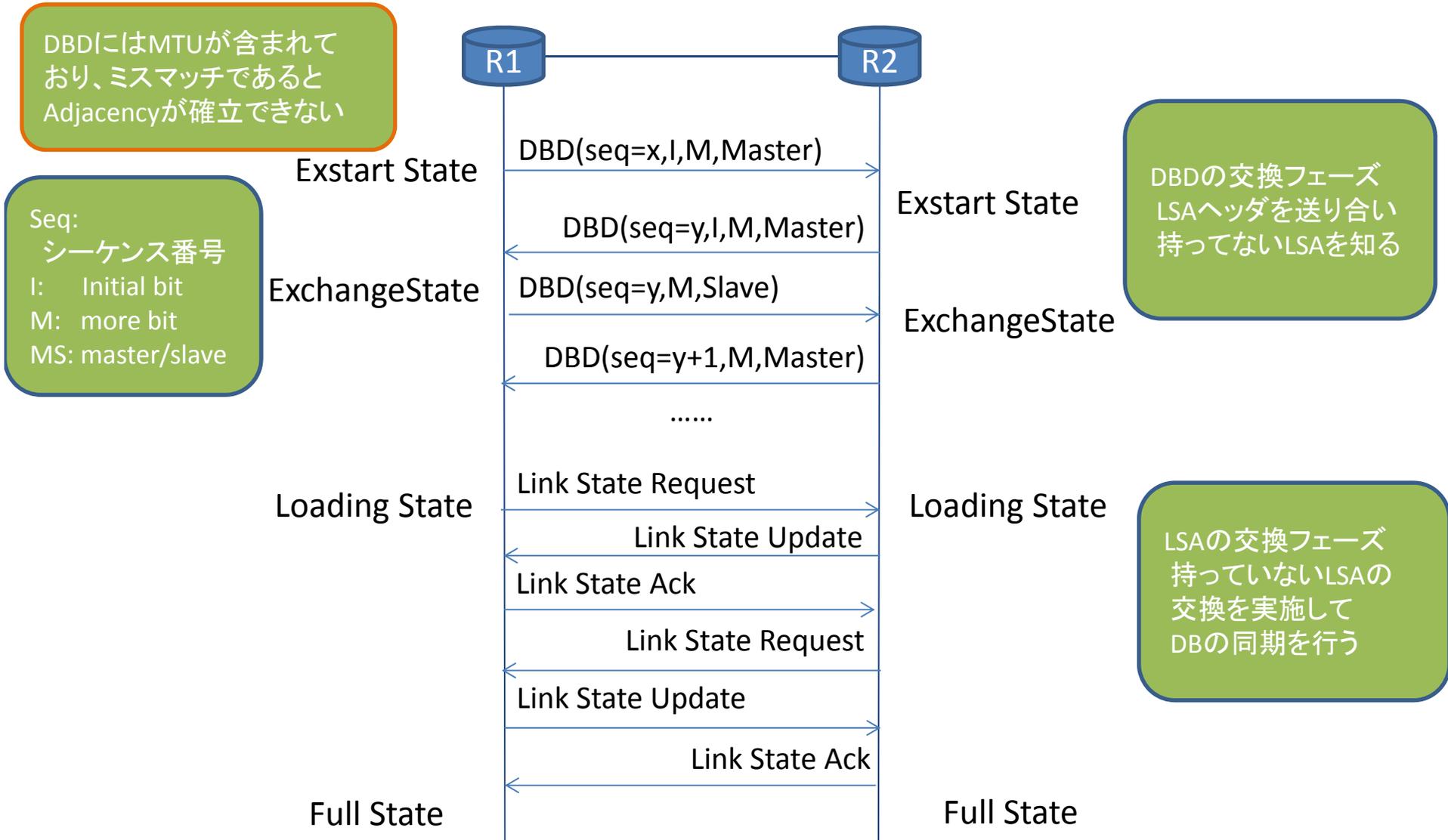
- DR/BDRはそれぞれ指名ルータ、バックアップ指名ルータと呼ばれる  
DR/BDRがなく、メッシュ状に隣接関係を結ぶと隣接関係が大きくなり  
やりとりが増大するため、DRという代表者を決めて隣接関係を形づくる

# DR/BDRの選出



- Hello packetにPriority値が含まれて、HelloでDR/BDRを選出
- Priorityが同じ場合には、RouterIDの大きいものが選出される
- DR/BDRがその他(DROther)とadjacencyを確立する
- DRがNetwork LSAを生成する

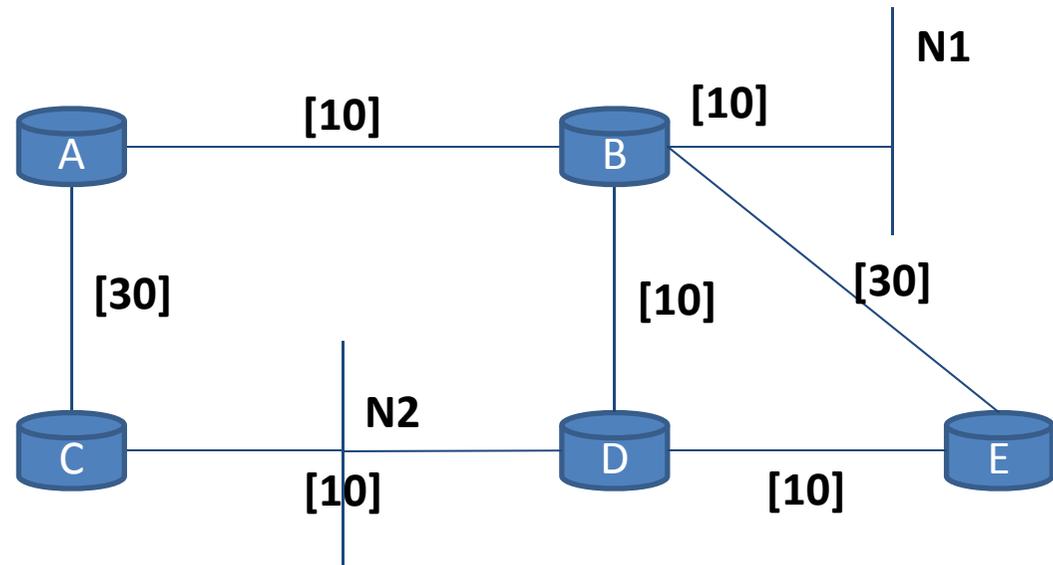
# Adjacencyの確立



# LSDBの構築

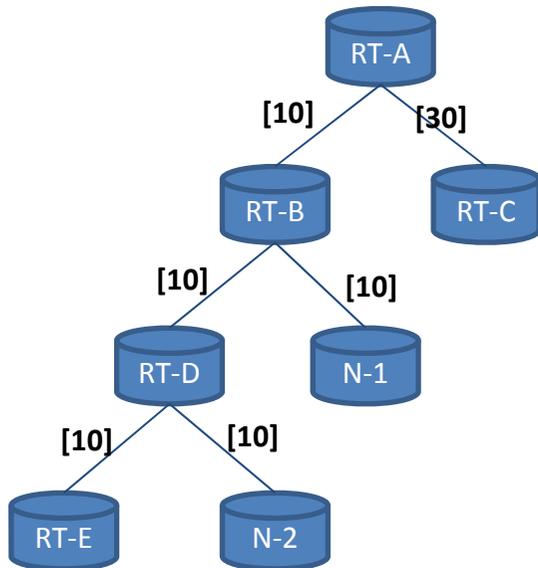
- リンクステート(Link State)からLSDBを構築

To/From	RT-A	RT-B	RT-C	RT-D	RT-E	N-2
RT-A		10				
RT-B	10			10	30	
RT-C	30			10		0
RT-D		10	10		10	0
RT-E		30		10		
N1		10				
N2			10	10		



# 最短パスツリーの構築とルーティングテーブル作成

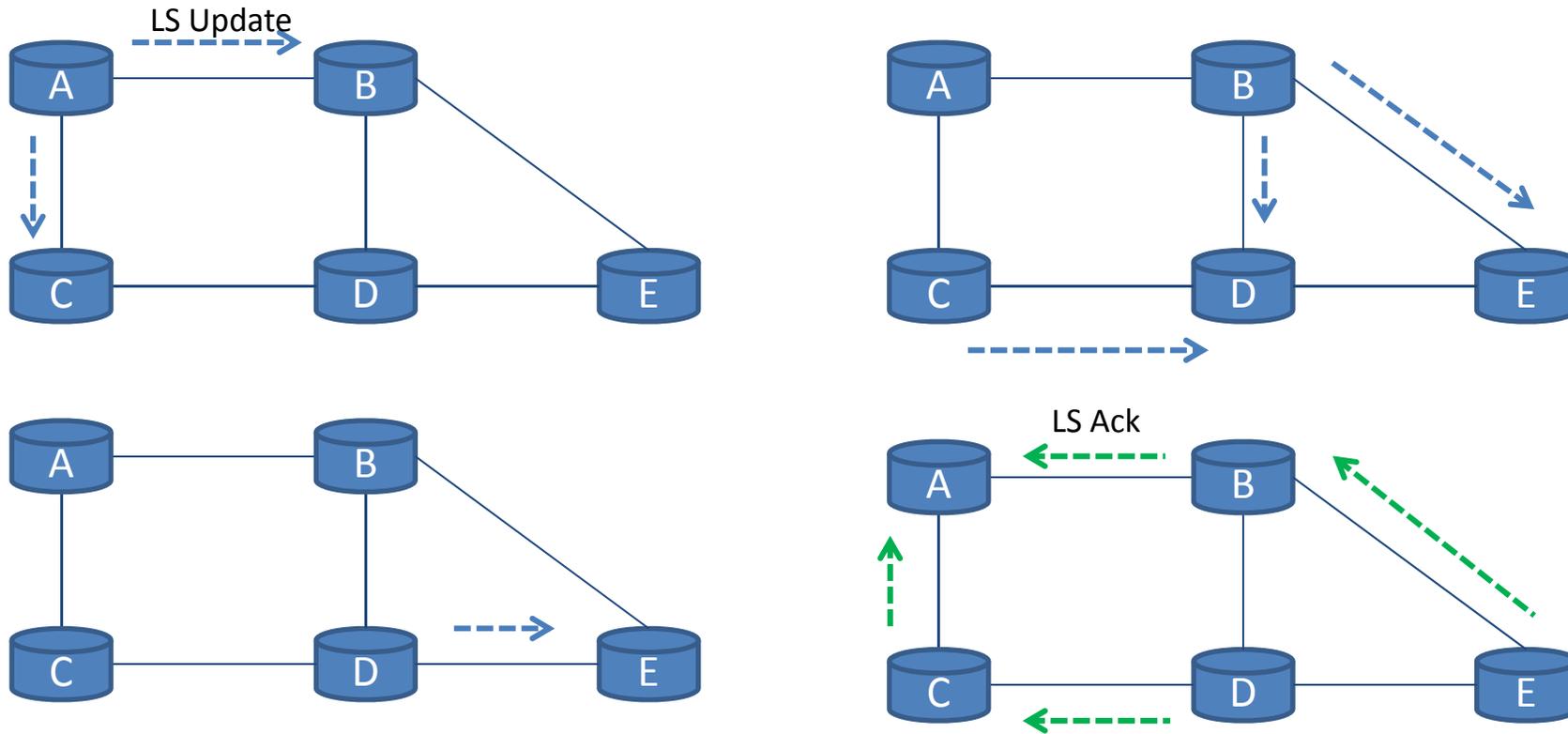
- 自分を始点(root)としてSPTを作成



Destination	nexthop	Cost
N-1	Link A-B	20
N-2	Link A-B	30
.....		
.....		

ルータに設定したIPアドレスはRT-Xに接続されているものとしてLSDBが形成され、SPT(最短パスツリー)が作成される。

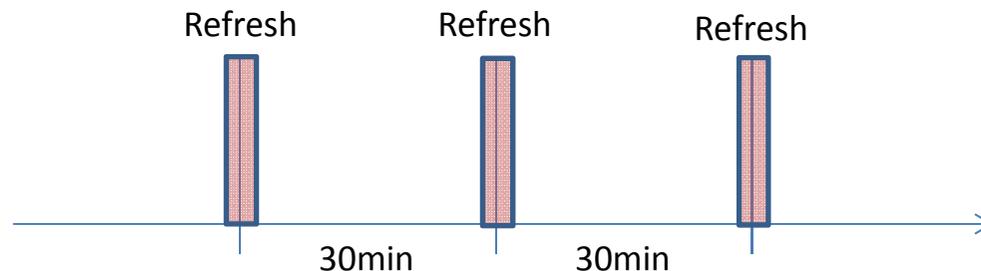
# Flooding



- LSAの変化がある場合には、LSUpdateを用いてadjacencyへ伝達される
- Checksumとより新しいかを確認し、他の隣接したインターフェースへ伝達する
- LS Ackで応答。応答があるまで定期的に再送することで信頼性を確保

# aging timeとrefresh

- Aging timeは、生成されてからの経過時間
- Max age(=60min)でLSAは捨てられる
- LSAは、30min程度でsequenceを上げてリフレッシュされることで保持され続ける
- OSによってdefaultのrefresh timeが異なる
  - 実際は負荷分散のために、分散してUpdate



# Router ID

- 32bitでOSPFルータをuniqueに識別
- Router IDの選択
  1. Router-idで指定したIPアドレス
  2. Loopback Interfaceの最大のIPアドレス
  3. ルータ上の最大のIPアドレス

RouterIDを変更したり、障害で選択されるIDが変化するとLSDBが崩壊する恐れがあるため、Router-IDを設定するのがよいでしょう



# LSAの種類

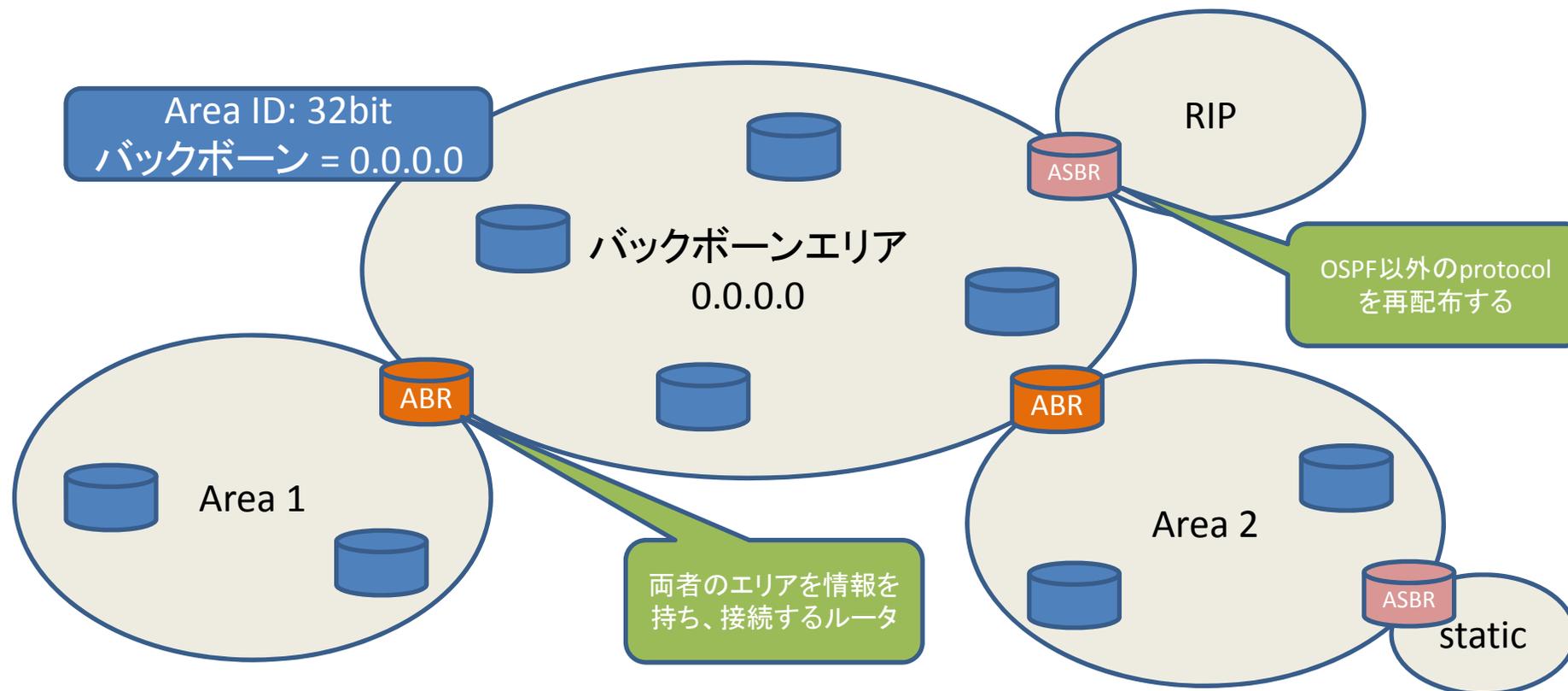
1. Router LSA
  - すべてのルータが生成
  - ospfに参加するすべてのリンクの情報
2. Network LSA
  - DRが生成。Networkに含まれるルータの情報
3. Summary LSA(Type 3)
  - ABR(エリア境界ルータ)が生成
  - Area外のNetwork情報
4. Summary LSA(Type 4)
  - ABRが生成。ASBR(AS境界ルータ)の情報
5. AS-External LSA
  - ASBRが生成。外部経路(AS境界ルータの向こう側)のNetwork情報
  - 実際には、staticからの注入だったり。別protocolからの注入

# 拡張LSA

6. Group-membership LSA
  - MOSPF用に拡張
7. NSSA-external LSA
  - NSSA エリア用に拡張
8. External-attributes LSA
  - BGPのattributesを運ぶために拡張
9. Opaque LSA(link-local scope)
10. Opaque LSA(area-local scope)
11. Opaque LSA(as-wide scope)
  - Traffic-engineering-lsa, grace-lsa...

エリア

# エリアの概念

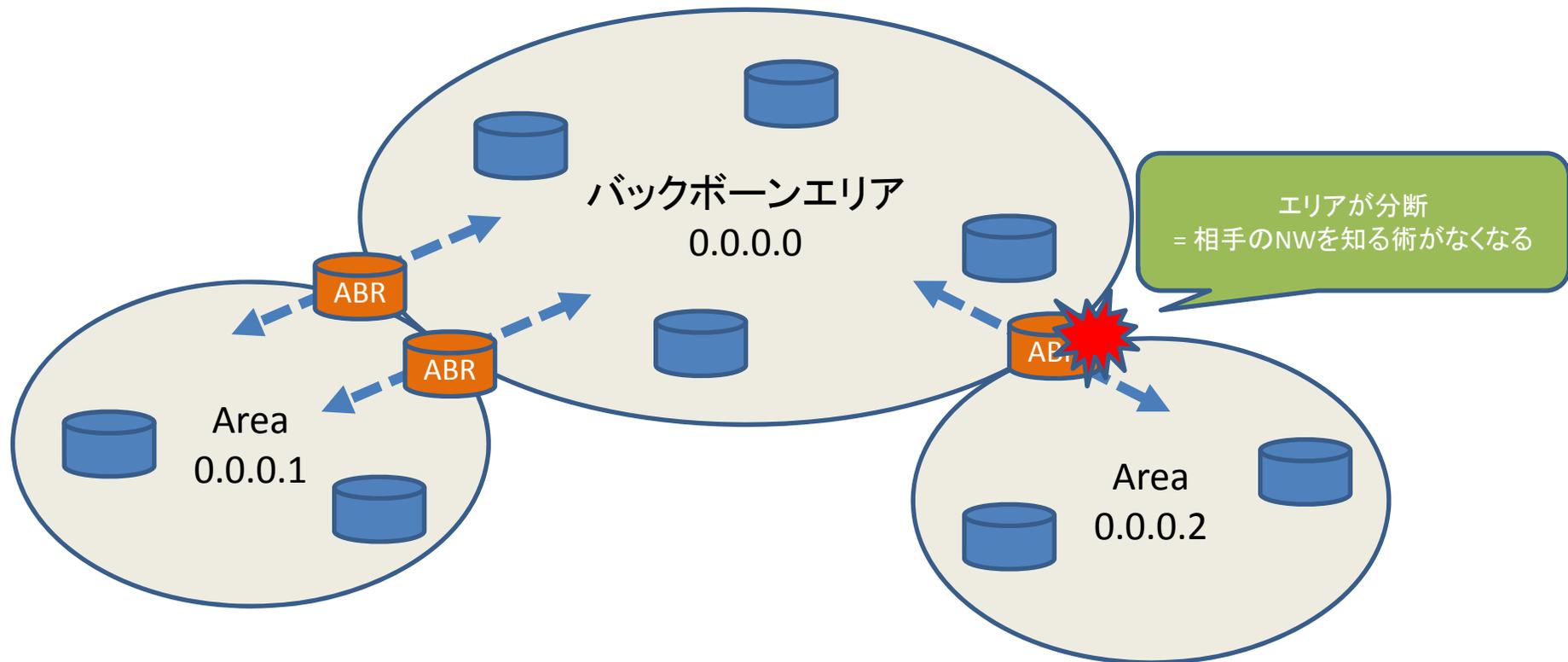


- エリアを分割することでSPF計算の範囲をエリアに限定
- ABR(エリア境界ルータ)は複数にエリアに属しエリア毎のトポロジーを知る
- 各エリアは必ずバックボーンエリアに隣接する必要がある

# エリアの種類

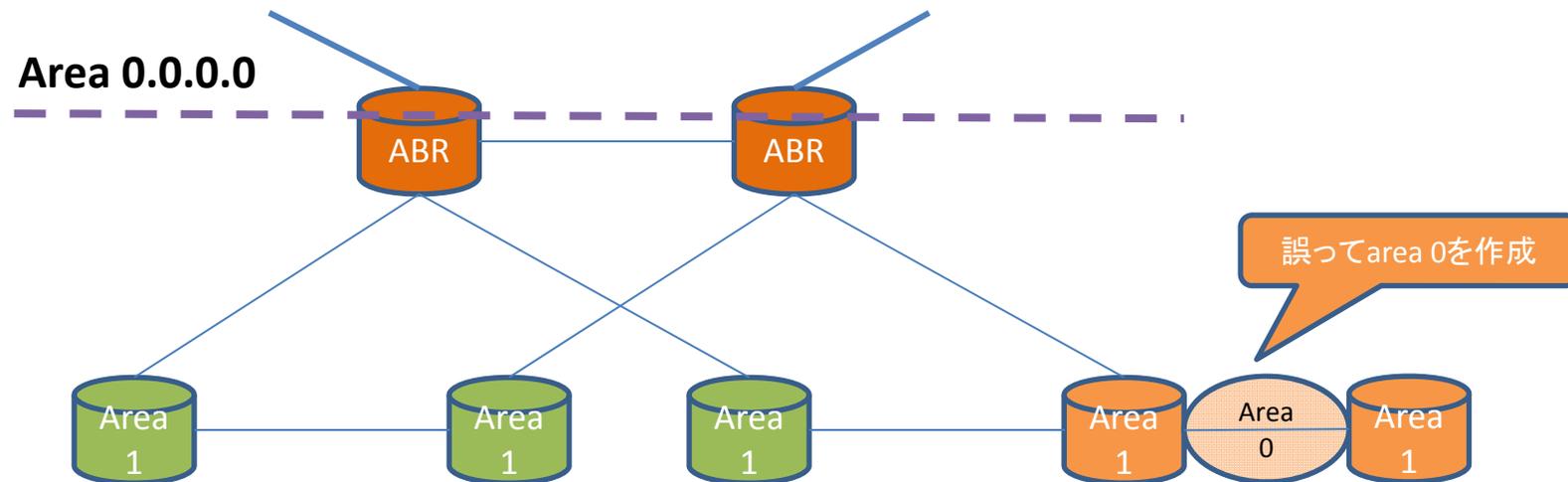
- バックボーンエリア(Backbone Area)
  - すべてのエリアがこのエリアに隣接する
- 標準エリア(Area)
  - バックボーンエリアへは、ABR(Area Border Router)が仲介
- スタブエリア(Stub Area)
  - 外部経路を受け取らない → default routeで代用
  - ASBRは配置できない
- トータルリースタブエリア(Totally Stubby Area)
  - Summary経路も受け取らない → default routeで代用
- NSSA(Not so Stubby Area)
  - スタブエリアにASRRを配置できるようにしたもの
  - NSSA内はType-7をAdvertiseし、Area 0.0.0.0へはType-5に変換

# エリア分断



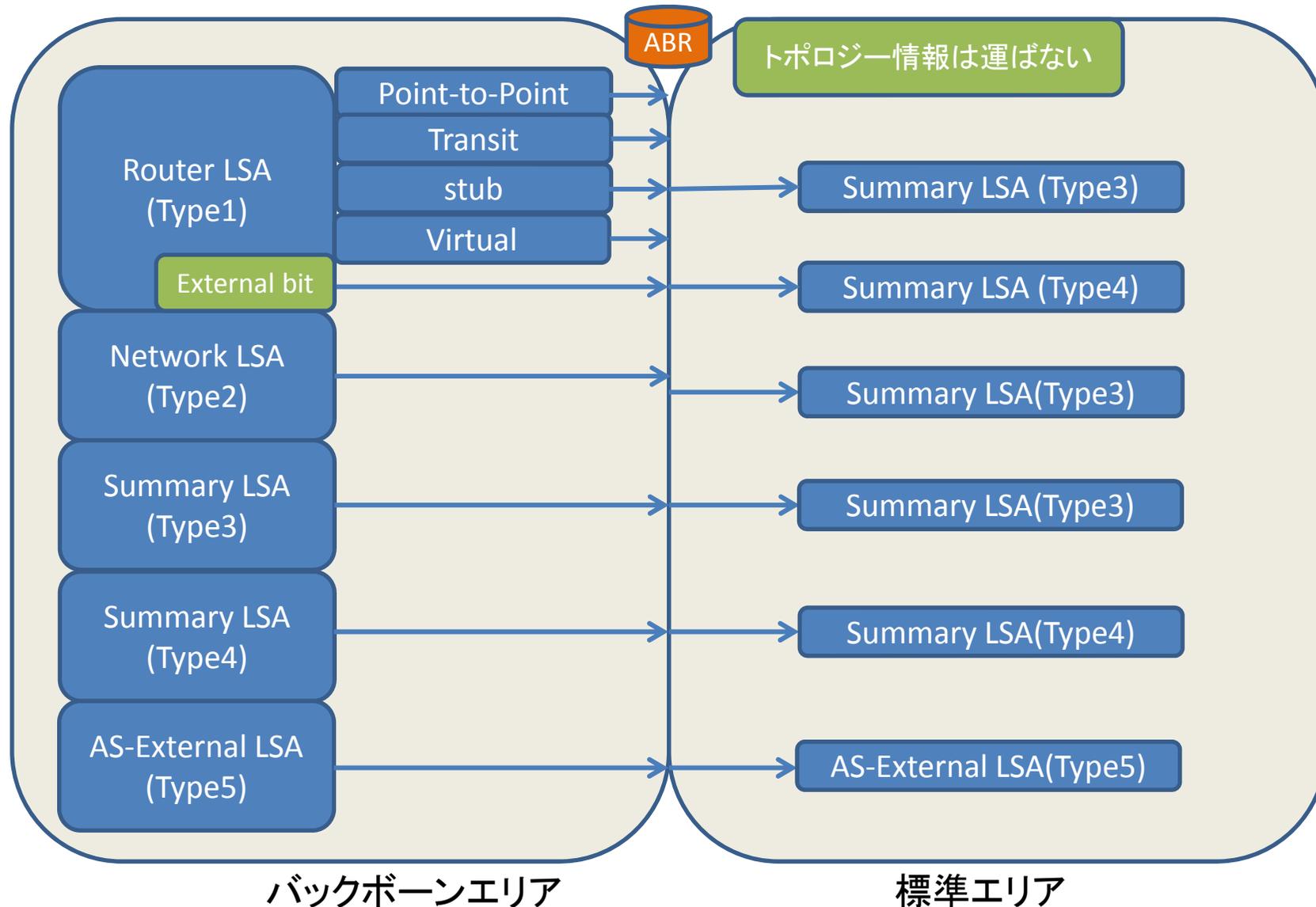
- エリアが分断することを絶対避けるように設計が必要
- ABRは両者のトポロジーを計算する
- 分断を避けるために2台以上のABRは必要

# エリア分断 Cont.



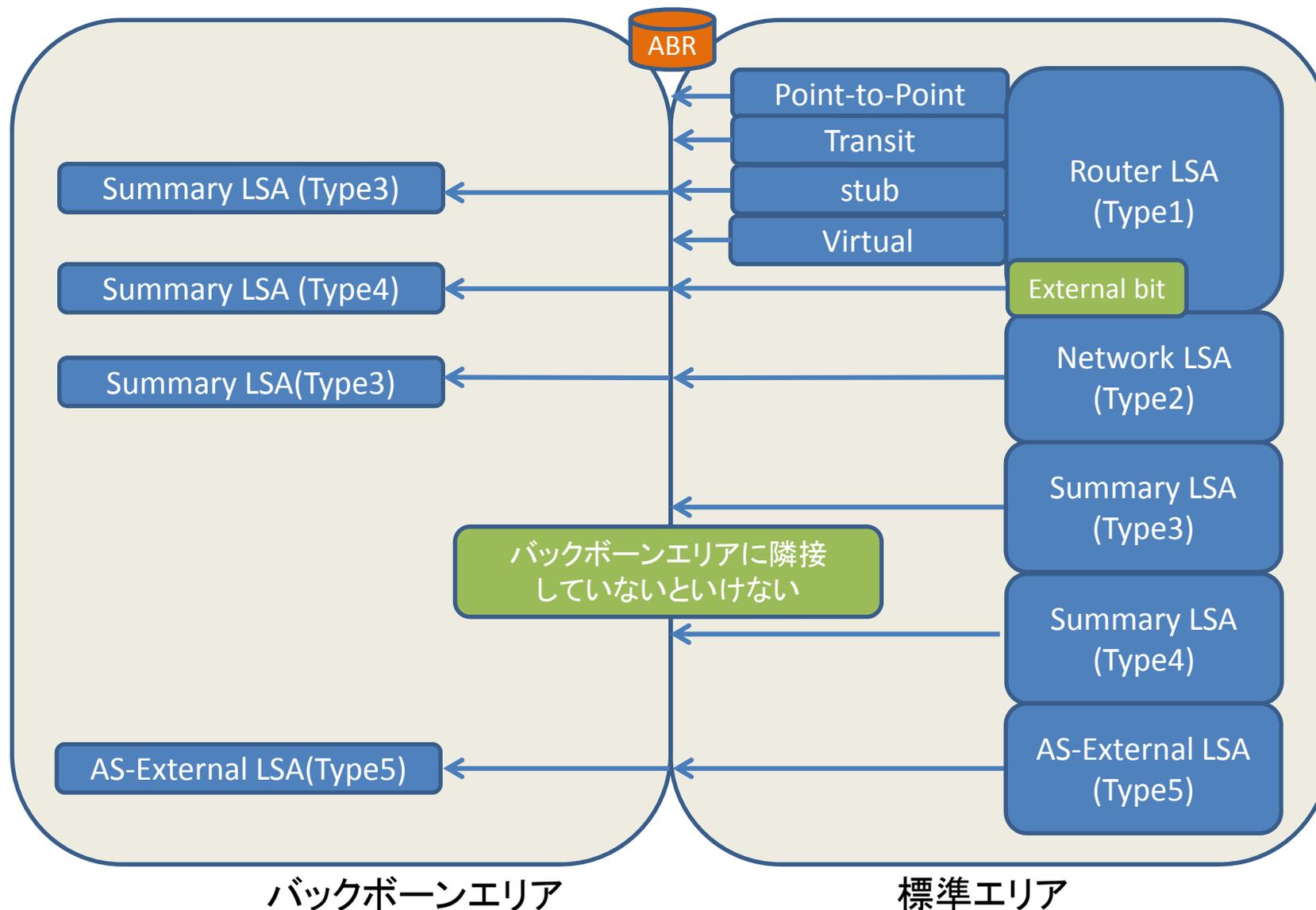
- ABRはバックボーンエリアからのsummary-LSAしか採用しない
- 孤立したarea 0.0.0.0をつくった場合には、孤立したバックボーンエリアからのsummary-LSAしか計算されないため、結局は孤立した結果となるので要注意

# エリア0から他エリアへ

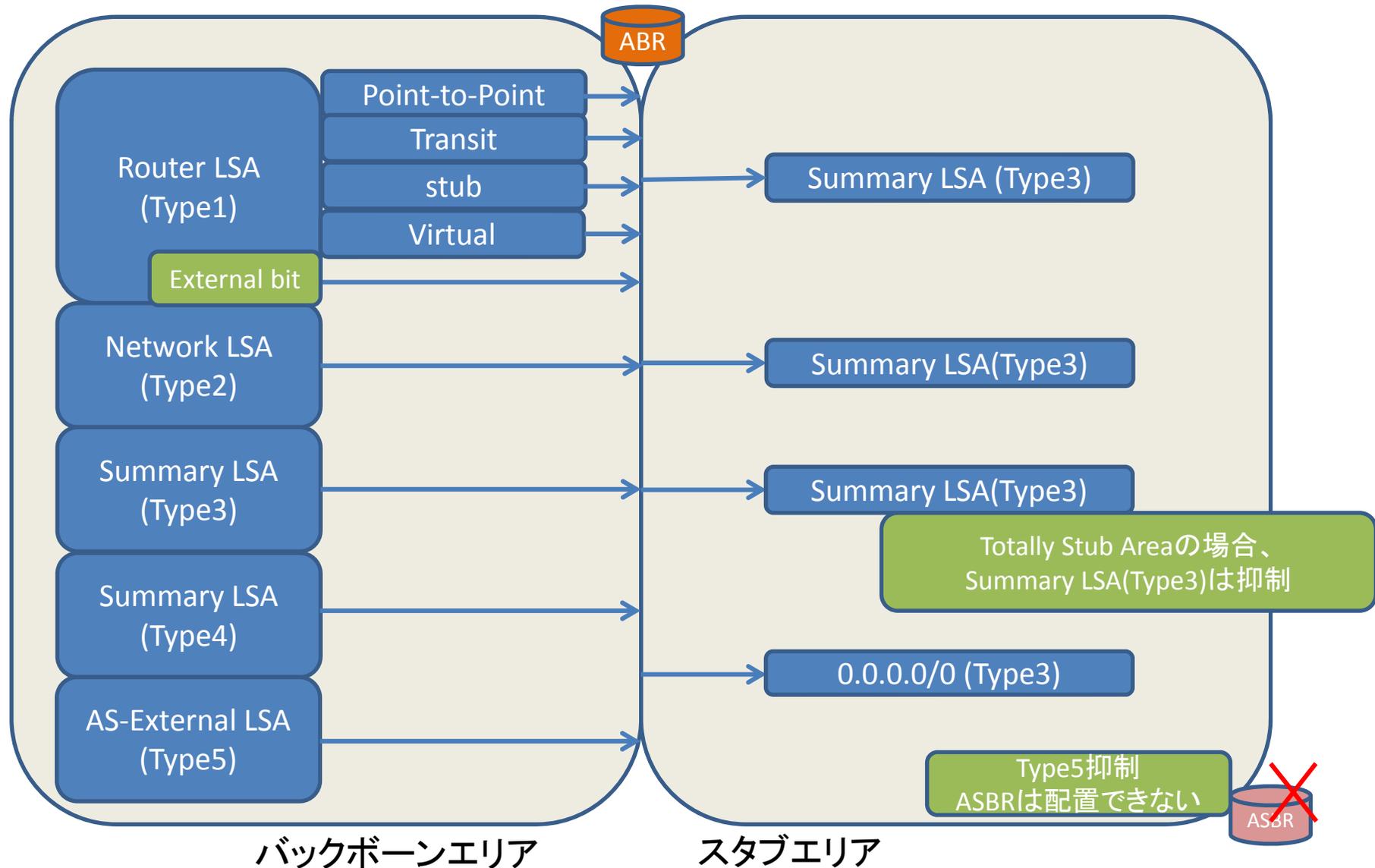




# 他エリアからエリア0へ



# Stub Area



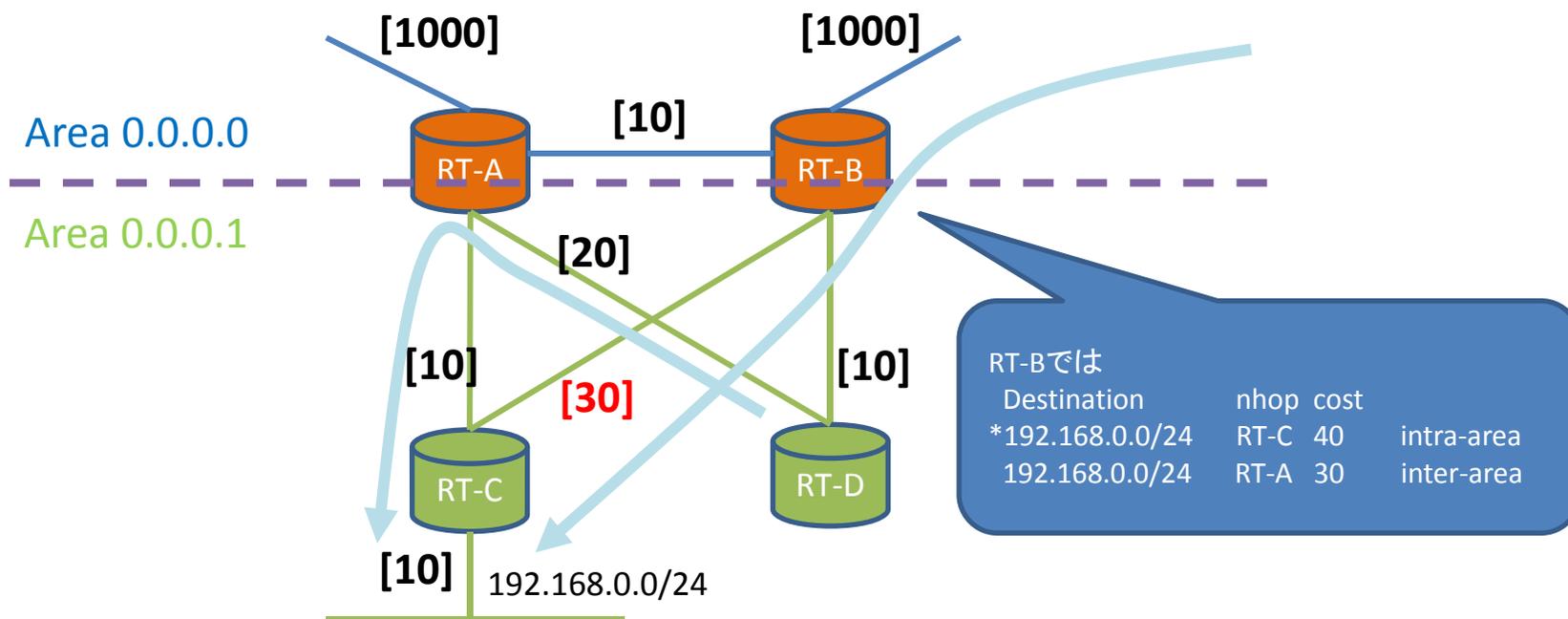
# エリア経路の優先

AS内・AS外経路	経路の種類名前	説明
内部ルート	Intra-area	LSA type1/type2の経路
	Inter-area	LSA type3/4の経路
外部ルート	external type1	外部と内部コストの合計
	external type2	外部コストのみ

- 優先順序

1. Intra-area(エリア内経路)
2. Inter-area(エリア外経路)
3. type1 as-external(外部経路 type1)
4. type2 as-external(外部経路 type2)

# エリア経路の優先 Cont.



- 経路の優先度により、costに関係なくintra-area routeが優先される
- エリアを構成する時は、よく考えて設計が必要

# エリアの利用

- メリット
  - トポロジー(LSDB)を小さくできる
  - SPFの計算の負荷を下げる  $O(l * \log(n))$
- デメリット
  - 設定、構成が複雑になり、メンテナンス性が悪い
  - ABRの設計と冗長化で機器、運用共にコスト高

デメリット >> メリット

最初からエリアをきることは考えない方がシンプル

# エリアの利用 Cont.

- ルータの高性能化
  - メモリ、CPUの大容量化
  - ルーティングとフォワーディングの分離
  - ルーティングに専念できるパフォーマンス
- 高速迂回のトレンド
  - 経路集約は他エリアにイベントを通知しない
  - エリア毎のSPF時間

バックボーンエリアで十分  
LSA(type 1/2)のみ考えることがシンプル

OSPFv3

# Protocol番号と利用Address

- IP Protocol=89を利用
  - ospfv3も同様(next header = 89)
- Multicastを利用して情報を交換
  - ALLSPFRouters(224.0.0.5/ff02::5)
  - ALLDRRouters(224.0.0.6/ff02::6)
  - ospfv3では、link-local Addressを利用



# OSPFv3 OSPFv2の違い

- LSAが再定義。トポロジーが分離
- OSPF headerはアドレス非依存に
- Flooding Scopeが定義
- IPv6 link-Local Addressを利用
- multiple instance per link
- Authentication がヘッダから削除

# LSAパケットフォーマット (ospfv2, ospfv3)

# ospfv2 / ospfv3 header

ospfv2 Header

Version	Type	Packet length
Router ID		
Area ID		
Checksum	AuthType	
Authentication		
Authentication		

ospfv3 Header

Version	Type	Packet length
Router ID		
Area ID		
Checksum	Instance ID	0

- 24 bytes → 16bytes
- RouterID/AreaIDは32bitのまま維持
- multiple instance対応(defaultは0)
- Authenticationヘッダ削除 → RFC4552

# ospfv2 / ospfv3 LSA header

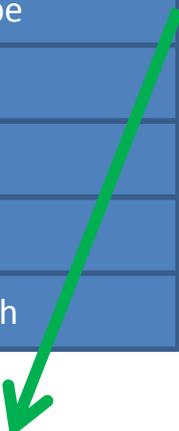
ospfv2 LSA Header

LS Age	Options	LS type
Link State ID		
Advertising Router		
LS Sequence Number		
LS Checksum	Length	

Options fieldは各種LSAの中で定義

ospfv3 LSA Header

LS Age	LS Type
Link State ID	
Advertising Router	
LS Sequence Number	
LS Checksum	Length



U	S	S	LSA function code
1	2		

U: function codeが理解できない  
場合の取り扱いについて  
S1/S0: flooding scope

# LSA type

Type	LSA	Function code	LSA	LS type code	Flooding scope
1	Router LSA	1	Router LSA	0x2001	Area Scope
2	Network LSA	2	Network LSA	0x2002	Area Scope
3	Network Summary LSA	3	Inter-Area-Prefix LSA	0x2003	Area Scope
4	ASBR Summary LSA	4	Inter-Area-Router LSA	0x2004	Area Scope
5	AS-external LSA	5	AS-External LSA	0x4005	AS Scope
6	Group Membership LSA	6	Group-Membership LSA	0x2006	Area Scope
7	NSSA External LSA	7	Type-7 LSA	0x2007	Area Scope
8	External Attributes LSA	8	Link LSA	0x0008	Link-local Scope
9	Opaque LSA(link-local scope)	9	Intra-Area-Prefix LSA	0x2009	Area Scope
10	Opaque LSA(area-local scope)	10	Intra-Area-TE LSA	0xa00a	Area Scope
11	Opaque LSA(AS scope scope)				

拡張

# Router LSA(0x2001)

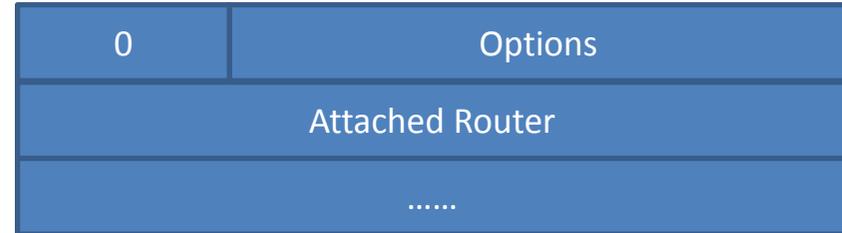
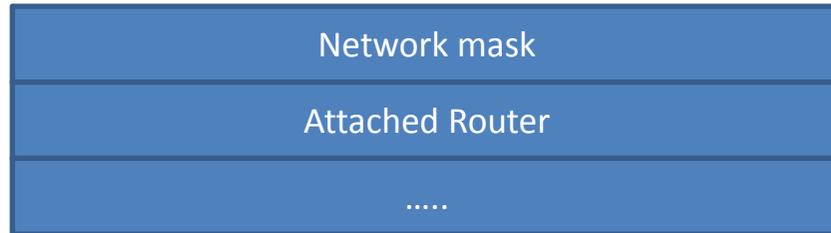
00000VEB	0	#links
Link ID		
Link Data		
Type	# TOS	Metric
.....		
TOS	0	TOS Metric

000NtxVEB	Options	
Type	0	Metric
Interface ID		
Neighbor interface ID		
Neighbor Router ID		

- Type: {point-to-point, transit, **reserved**, Virtual link}
- interface ID = MIB II ifindex (ではない場合も)

Link state ID	Link Type	Link id	Link data
Router ID	1	Neighbor Router ID	Interface IPaddress MIB II ifindex
	2	DRのIP address	Interface IP Address
	3	Network address	IP subnet mask
	4	Neighbor Router ID	Interface IP Address

# Network LSA(0x2002)



- link state idは、DRのIP Address → DRのinterface ID
- attached Routerは32bitのままなので、Router IDが入る

# inter-area-prefix LSA(0x2003)

Network mask	
0	Metric
TOS	TOS Metric

0	Metric	
Prefixlength	PrefixOptions	0
Address Prefix(128bit)		

- ospfv2のtype 3(summary LSA)はarea外のprefixを運ぶ(役割は同じ)
- ospfv2では、Link state IDにnetwork addressがセット  
LSAの中のNetmask maskのセットでprefixを運ぶ



# inter-area-Router LSA(0x2004)

Network mask	
0	Metric
TOS	TOS Metric

0	Options
0	Metric
Destination Router ID	

- ospfv2のtype 4(summary LSA)はASBRの情報を運ぶ(役割は同じ)
- ospfv2の場合は、Link State IDにASBRのRouter IDがセットされ、Netmask mask=0がセットされる(実際のexternal経路はtype5で運ぶ)
- ospfv3では、ASBRのRouter IDだけがセットされる

# AS-external LSA(0x4005)

Network mask		
E0000000	Metric	
Forwarding Address		
External Route Tag		
E	TOS	TOS Metric

0	E	F	T	Metric	
Prefixlength		PrefixOptions		Referenced LS Type	
Address Prefix(128bit)					
Forwarding Address(optional)					
External Route Tag(Optional)					
Referenced Link State ID(Optional)					

- ospfv2/ospfv3の両者ともexternal route(外部経路)の情報を運ぶ
- E-bitは1=type2, 0=type1, F-bit = Forwarding Prefix, T=External Route Tag

# link LSA(0x0008) ospfv3 only

Options=24bit( 下位8bitのみ記載 )

```
+---+---+---+---+---+---+---+---+---+---+
| | |DC| R| N| x| E|V6|
+---+---+---+---+---+---+---+---+---+---+
```

V6-bit: clearされるとrouting calculationから除外

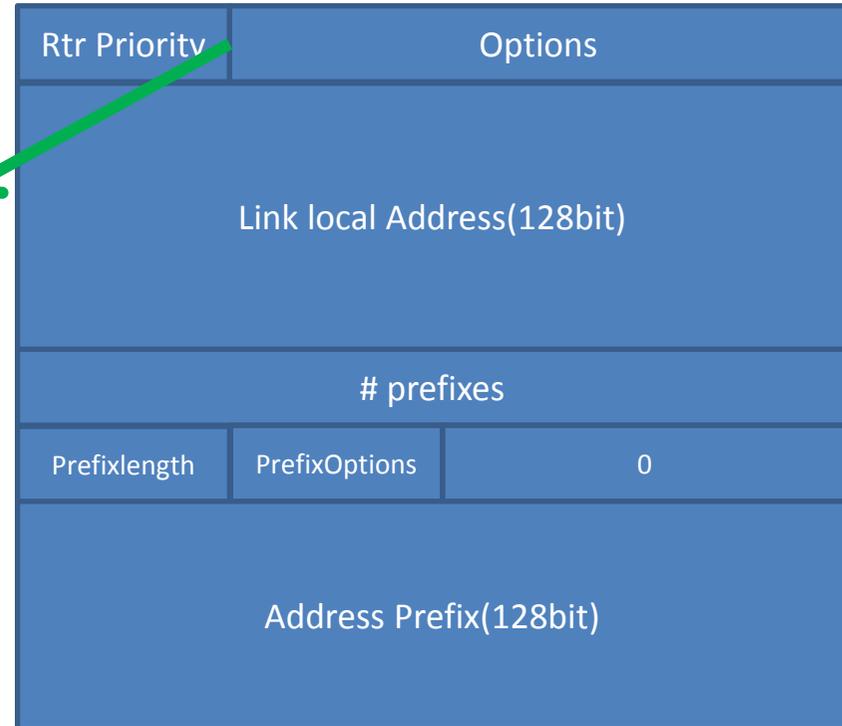
E-bit : External bit

X-bit: MOSPFで利用

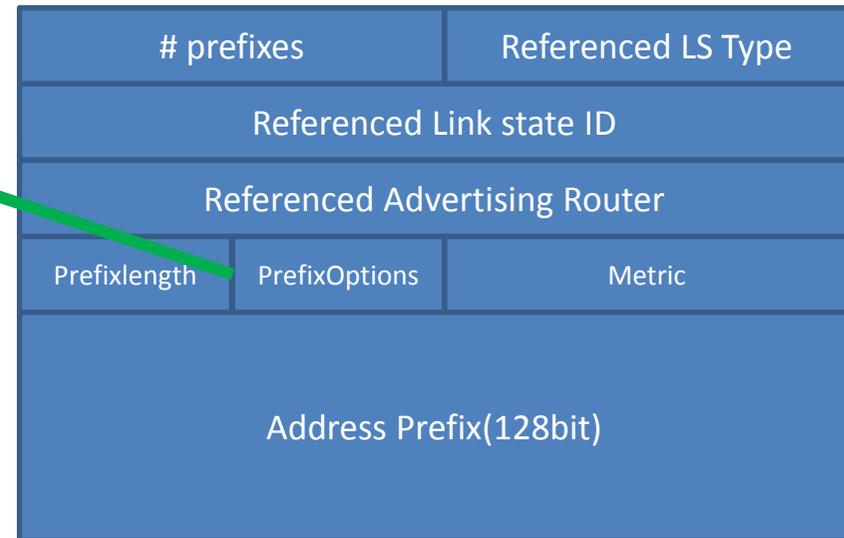
N-bit: NSSAで利用

R-bit: Router bit(active)

DC-bit: Demand circuit bit



# intra-area-prefix LSA(0x2009) ospfv3 only



- 参照LS Type , Link State ID, Advertising Routerがセット
- Prefix情報を伝搬

# RFC

- rfc2328 OSPF Verion 2
- rfc3101 The OSPF Not-So-Stubby Area(NSSA) Option
- rfc3137 OSPF Stub Router Advertisement.
- rfc3630 Traffic Engineering (TE) extensions to OSPF Version 2
- rfc4552 Authentication/Confidentiality for OSPFv3
- rfc5250 The OSPF opaque lsa Option
- rfc5340 ospf for ipv6
- rfc5329 Traffic Engineering Extensions to OSPF Version 3
- rfc5838 Support of Address Families in OSPFv3

# IGP 設計と運用

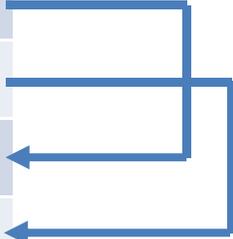
# 設計のポイント

- 信頼性(安定稼働)
  - 物理リンクの不安定さが余計な再計算を生む
  - 信頼のおけるメディアを選択
  - 日々のOSPFのリソースやイベントを知ることが必要(SNMP/debug)
- 拡張性
  - prefixをOSPFに注入のし過ぎはLSDBの増大を招く
  - トラフィックコントロールは拡張性のために、ある程度の指針を決めておく
- 運用性
  - NW設計とオペレーションはシンプルに保つ
  - 切り分けや障害時間の短縮
- 品質
  - N+1設計や十分な帯域確保
  - 障害時の迂回速度を考えたネットワーク設計
  - セキュリティ(OSPF AUTH)で自組織を守る

# BGPとIGPの関係

- BGPはprefixを伝搬
  - BGPはprefix情報とNEXT-HOP属性を運ぶ
  - NEXT-HOP属性を解決するのにIGPで再帰解決
  - IGPの収束・変更がBGPの経路に大きく影響する

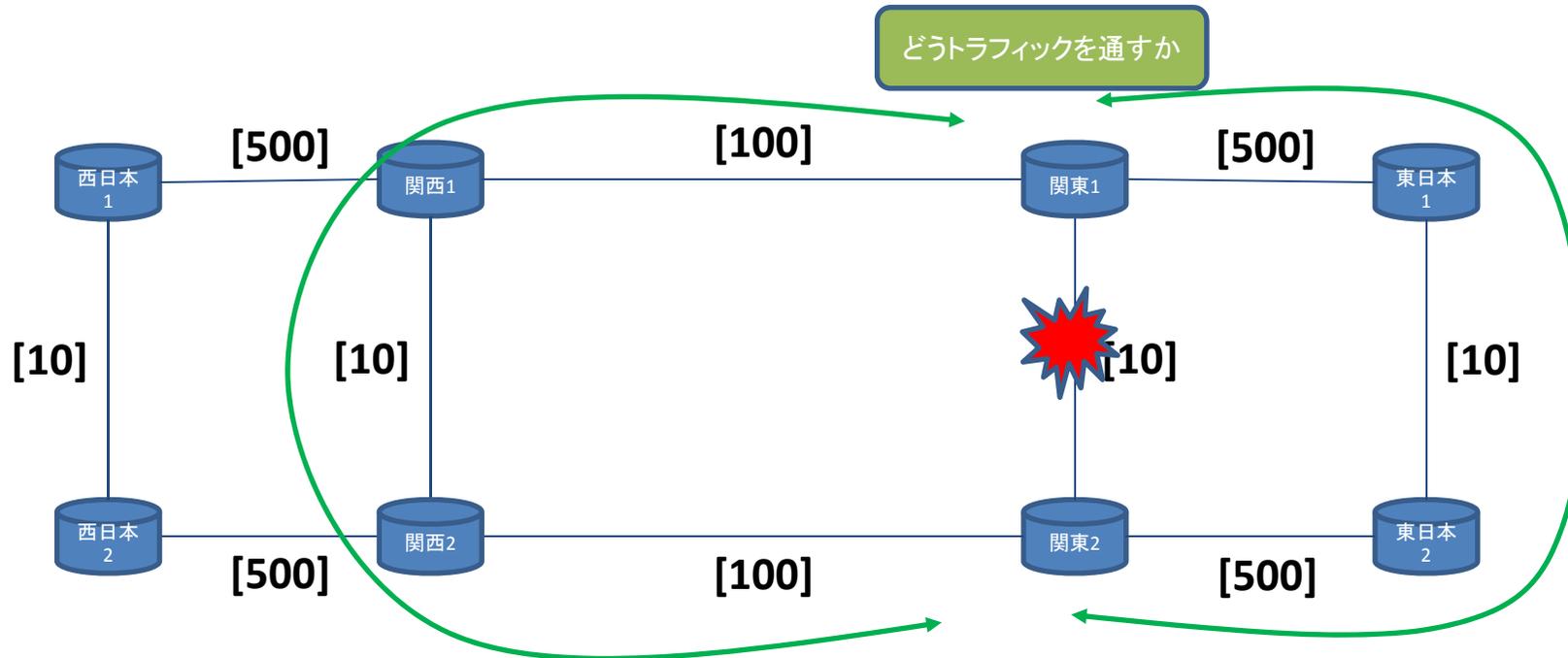
宛先	ネクストホップ	学習元
10.0.0.0/16	192.168.0.1	BGP
10.1.0.0/16	192.168.0.2	BGP
192.168.0.1	IF_X	IGP
192.168.0.2	IF_Y	IGP



The diagram shows four blue arrows pointing from the '学習元' (Learning Source) column to the 'ネクストホップ' (Next Hop) column. The first arrow points from 'BGP' to '192.168.0.1'. The second arrow points from 'BGP' to '192.168.0.2'. The third arrow points from 'IGP' to 'IF\_X'. The fourth arrow points from 'IGP' to 'IF\_Y'. This illustrates that BGP learns prefixes from BGP and uses IGP to resolve the next hop IP addresses to interface names.

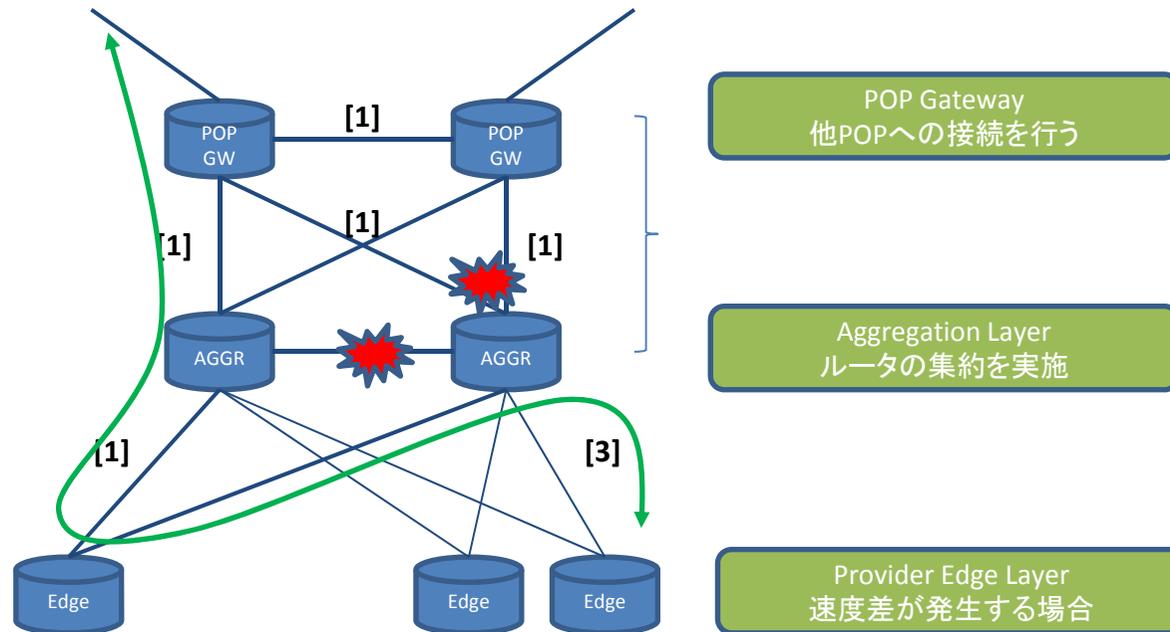


# コスト設計



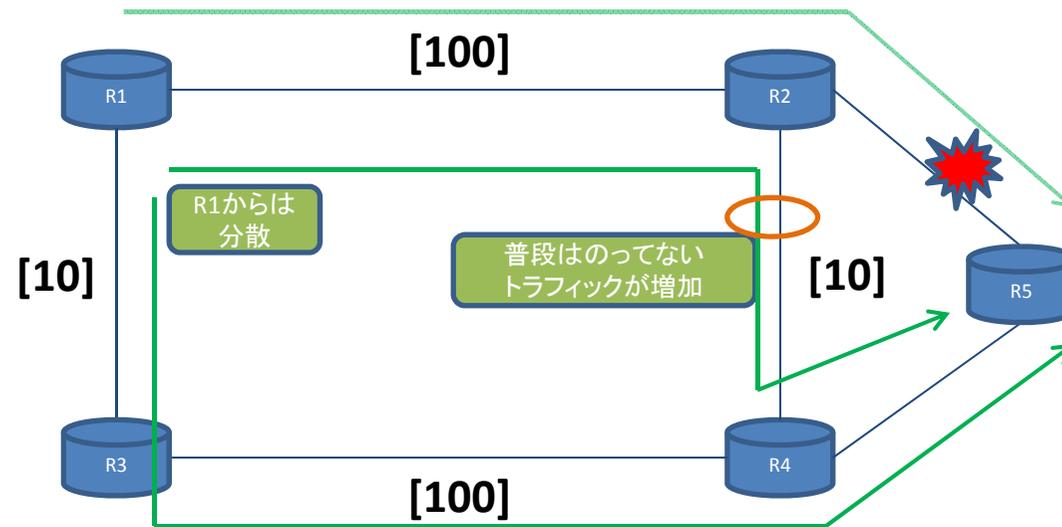
- コア拠点を中心に考え、距離(遅延)を考慮したポリシーを決めておく
- コスト設計 = ポリシーの設計なので、どこを利用するのがいいかをあらかじめ決めておく
  - 関東コア同士が切れた際に東日本POPを経由したくない  
関東1 ~ 東日本1 ~ 東日本2 ~ 関東2 > 関東1 ~ 関西1 ~ 関西2 ~ 関東2  
そもそも関東コア同士は切れないように冗長回線が必要
  - 西日本1 ⇔ 西日本2同士の通信は普段はコアを回したくない  
西日本1 ~ 西日本2 < 西日本1 ~ 関西1 ~ 関西2 ~ 西日本2

# コスト設計 Cont.



- POP内の設計は遅延差がないので速度差を考えてリンクコストを決めておく
  - シャーシ型ルータではカード単位で死ぬことも考え、カード分散も考慮に入れる
  - いざトラフィックが流れてもいいようにコスト差をつけておき障害に備える
  - エリアを切るとintra-area優先や分断を考慮して設計する必要がある

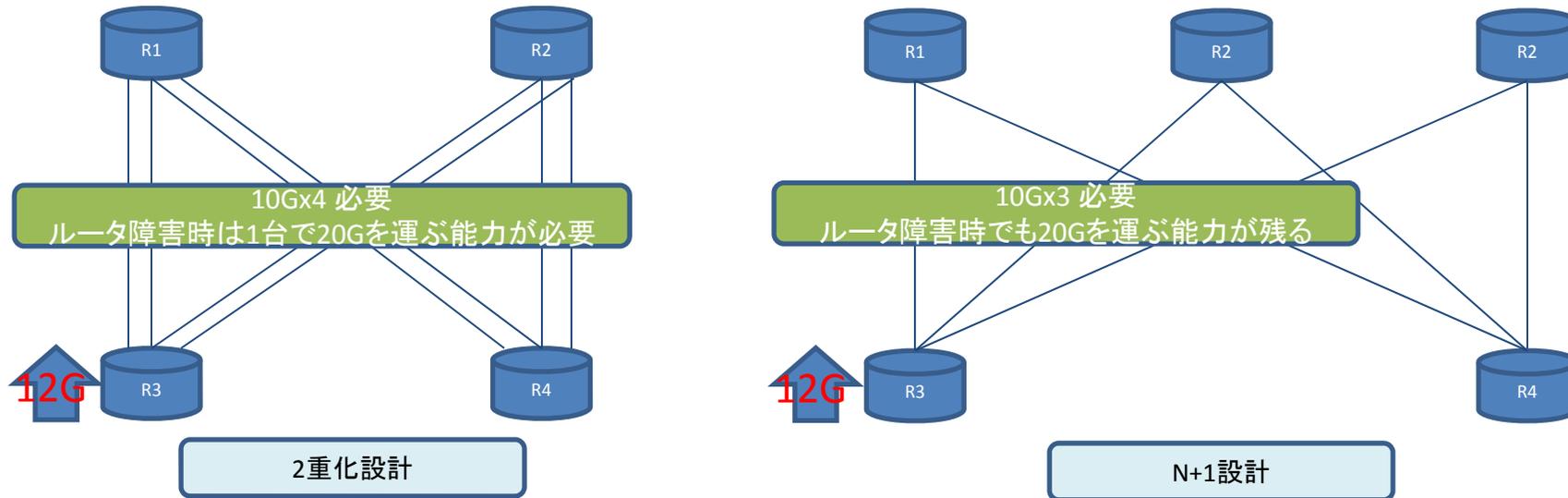
# ネットワーク設計 1



■ 普段利用されないが、障害時を想定して帯域の確保が必要

- R2 ⇔ R4のリンクに新たに増えるトラフィック
- R2 ⇔ R5のトラフィックの全部
- R1 ⇔ R5のトラフィックの半分くらい (ECMP)
- R3 ⇔ R4のリンクに新たに増えるトラフィック
- R1 ⇔ R5のトラフィックの半分くらい

# ネットワーク設計



- ルータ間の接続リンクが回線である場合には、N+1設計を実施するとリンク数が抑えられる
- |        |                               |
|--------|-------------------------------|
| 2重化設計: | 12Gのトラフィックをさばくためには、40Gのリンクが必要 |
| N+1設計: | 12Gのトラフィックをさばくためには、30Gのリンクで十分 |

# アドレッシング

- インフラアドレス
  - Loopback BLOCK(/32, /128)
  - Point-to-Point(/30, /126, /127)
  - POP単位
- ボーダアドレス
  - 顧客ボーダアドレス(/30, /2X)
  - AS Border(IX, Private Peering)
- 顧客割り当て
  - 割り当て用
  - サービス用途
- 特殊用途
  - Multihop用途のアドレスブロック

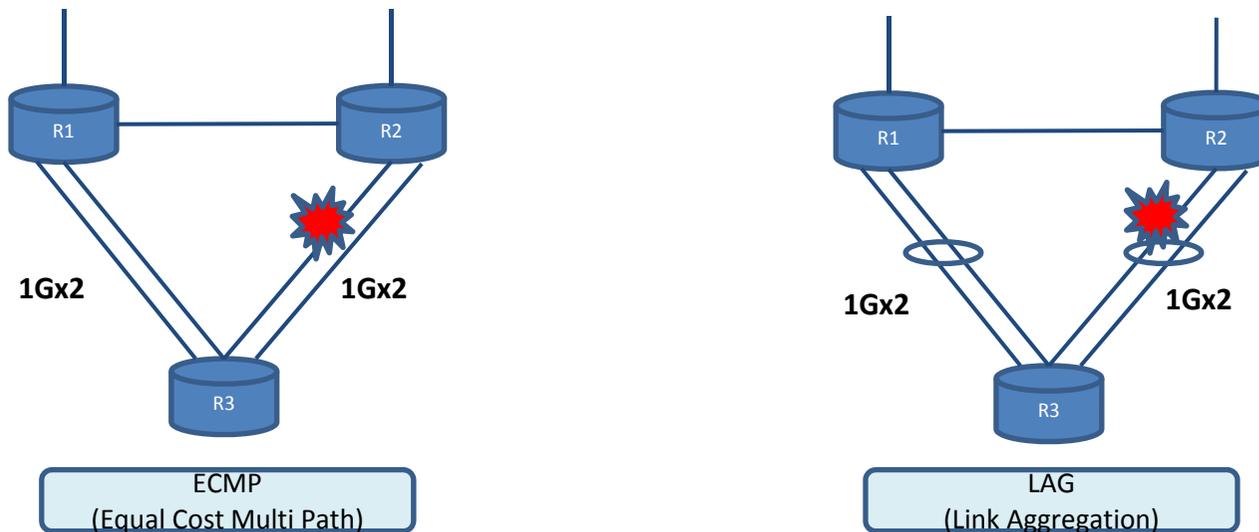
IGPから孤立した場合に隣接したルータからログインできるようにしておくのかそれとも完全に裏線を用意しておくのか予め対策を決めておくことが大事

AS外から到達する必要のないインフラブロックをシンプルにACLが書けるように心がける

# DR/BDR設計

- DR/BDRの選択
  - DR/BDRは不安定なルータを選択しない
  - セグメントの障害を避けるために二重化
  - 二重化したセグメントのDRを同一にしない
  - 障害でDR/BDRが重なった場合は正しい姿に
- DR/BDRの操作
  - “ip ospf priority <0-255>”でDRを指名
  - “priority 0”でDRを放棄することが可能

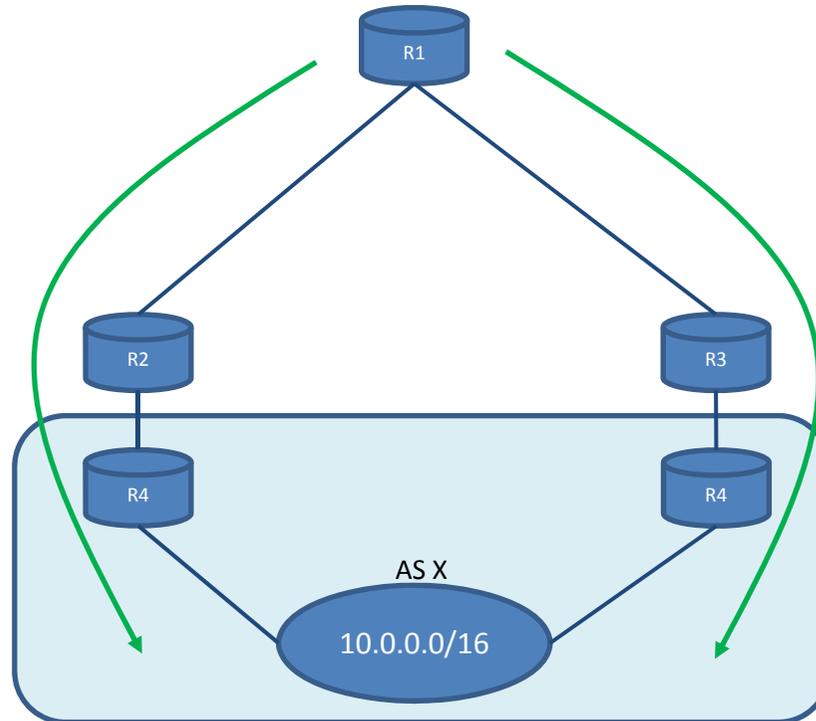
# ECMP/LAG



L3情報(SRC/DST Address)やL4情報(SRC/DST Port)や  
その他keyによってトラフィックが分散される(ベンダ依存)

- ルータ間に2本以上の接続がある場合は、障害で速度差が発生する

# ibgp multipath



R1:

\*10.0.0.0/16 nexthop R2

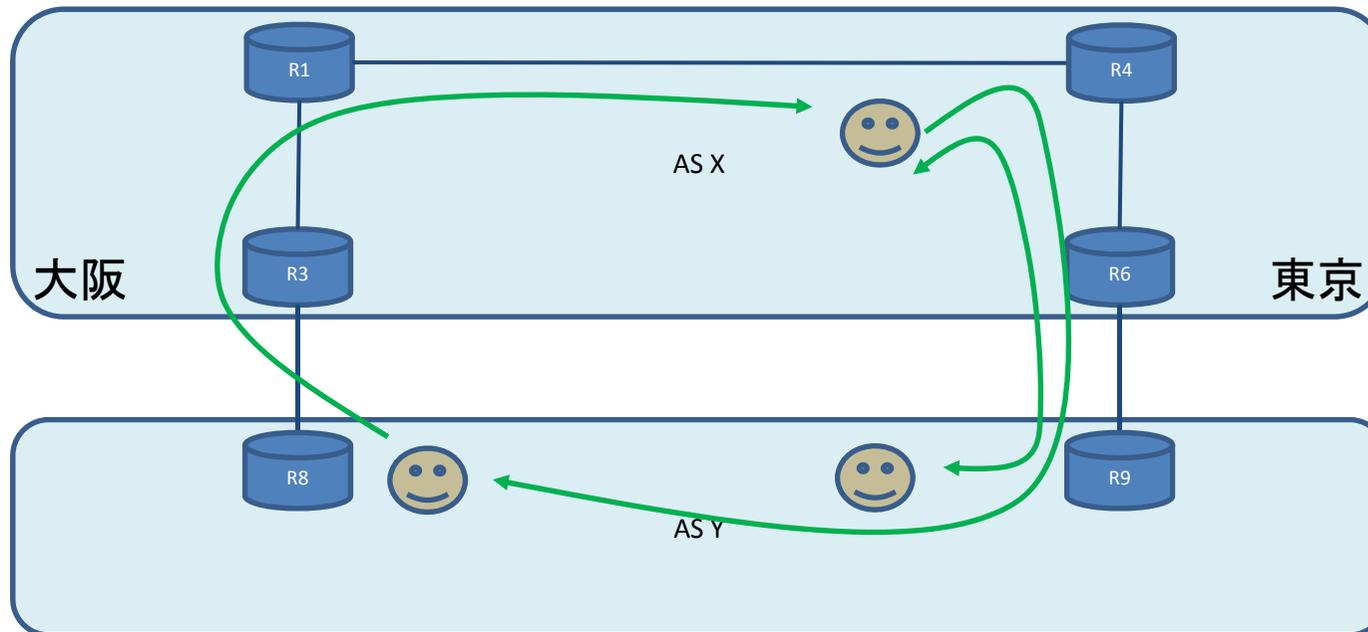
10.0.0.0/16 nexthop R3

Best経路は、1つしか選択されないので  
通常はR2にしかトラフィックは流れない

- BGP経路を複数の出口からバランスさせたい場合にibgp multipathを利用
- BGPのベストセレクションでnexthopまでのigp costが同一であれば、マルチパスが実現可能

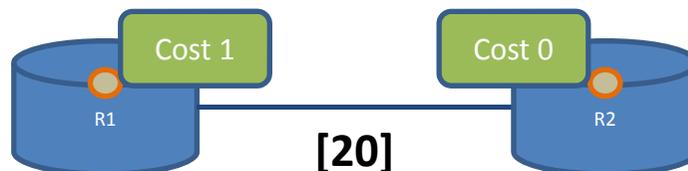


# AS間トラフィック交換



- 対等な関係のAS間のトラフィックの交換は、hot-potatoルーティングが基本  
自ASのIGPのメトリックに応じて出口が決定 (自分のリソースが最小限になる)

# loopbackまでの距離 1



IOS と JUNOS

R1 Loopback: 10.0.0.1/32 Cisco

R2 Loopback: 10.0.0.2/32 Juniper

R1 - R2: 192.168.0.1/30 - 192.168.0.2/30

```
R1> sh ip route 10.0.0.2
Routing entry for 10.0.0.2/32
  Known via "ospf 1", distance 110, metric 20, type intra area
  Last update from 192.168.0.2 on GigabitEthernet1/1 6d21h
ago
Routing Descriptor Blocks:
  * 10.0.0.2, from 10.0.0.2, 6d21h ago, via GigabitEthernet1/1
    Route metric is 20, traffic share count is 1
```

```
R2> > show route 10.0.0.1

inet.0: 36 destinations, 37 routes (36 active, 0 holddown, 0
hidden)
+ = Active Route, - = Last Active, * = Both

10.0.0.1/32    *[OSPF/10] 6d 21:54:09, metric 21
              > to 192.168.0.1 via ge-0/0/1.0
```

Loopbackはbgpのnexthopになり得るので、機器を変更するとコスト差が発生して意図しないバランスになるので要注意

# loopbackまでの距離 2



LoopbackはLA bitがセット  
Prefix length 128、Cost 0で運ぶ

ospfv2 と ospfv3  
R1 Loopback: 2001:db8:10::1/128 Cisco IOS  
R2 Loopback: 2001:db8:10::2/128 Juniepr Junos  
R1 - R2: 2001:db8::1/64 - 2001:db8::2/64

```
R1> show ipv6 route 2001:db8:10::2
Routing entry for 10.0.0.2/32
  Known via "ospf 1", distance 110, metric 20, type intra area
  Route count is 1/1, share count 0
Routing paths:
  FE80::217:CBFF:FEDA:625, GigabitEthernet1/1
  Last updated 00:05:19 ago
```

```
R2> > show route 2001:db8::10::1
inet6.0: 24 destinations, 25 routes (24 active, 0 holddown, 0
hidden)
+ = Active Route, - = Last Active, * = Both

2001:db8:10::1/128
  *[OSPF3/10] 00:04:28, metric 20
  > to fe80::20a:41ff:fe43:d080 via ge-0/0/1.0
```

ospfv2 と ospfv3でCost差が出てしまうので  
Dualstack環境では要注意

# STUB Router Advertisement

- Router LSAのtype stub以外のCostを上げることでどのどこから見ても遠くみえるのでパケットは通らない
  - Cisco: max-metric router-lsa
  - Junos: set protocols {ospf|ospfv3} overload
  - シンプル イズ ベスト . SPFも一度に！
- 一定時間指定を入れておくと再起動時も安心
  - Cisco: on-startup
  - Junos: timeout
  - 経路学習する時間以上を設定

# OSPF Hello Timer

- Neighbor同士で死活監視に利用
  - Hello interval = 10sec
  - Dead timer = 40sec
  - 40秒相手からHelloがなかったらDownと判断
- Helloタイマの変更
  - 検知高速化に使うには負荷がかかる(last resort)
  - 物理レイヤで落ちるリンクを利用する
  - BFD(Bidirectional Failure Detection)利用
  - iBGP hold-timer > IGP Down timer
    - BGPが落ちると、経路を持たないblackholeルータの原因に

# ルーティングの収束

- IGP Convergence =

検知時間 + 伝搬時間 + SPF計算時間 + RIB反映時間 + FIB更新時間

## 検知短縮のアプローチ

警報転送可能なメディア、P2Pトポロジー、BFDを使って検知時間

link-down検知するまでの遅延時間(carrier-delay, upが早いと不安定に)

## 伝搬時間/SPF計算時間のアプローチ

伝搬時間は様々な要因があるのでチューニングベンダとよく相談

余計に不安定な要素を生む可能性あり

伝搬時間に頼らないmicro-loopの起こりにくいトポロジーも効果的

## RIB/FIB更新時間のアプローチ

インターネット経路の増大に伴うFIB更新時間のインパクトは大きい

FIB階層化も効果的

# Timer Tuning

```
R1> sh ip ospf
Routing Process "ospf 1" with ID 10.0.0.1
Start time: 00:02:14.428, Time elapsed: 3w0d
Supports only single TOS(TOS0) routes
Supports opaque LSA
Supports Link-local Signaling (LLS)
Supports area transit capability
Supports NSSA (compatible with RFC 3101)
Event-log enabled, Maximum number of events: 1000, Mode:
cyclic
It is an area border router
Router is not originating router-LSAs with maximum metric
Initial SPF schedule delay 5000 msec
Minimum hold time between two consecutive SPF 10000 msec
Maximum wait time between two consecutive SPF 10000 msec
Incremental-SPF disabled
Minimum LSA interval 5 sec
Minimum LSA arrival 1000 msec
...
```

## IF設定

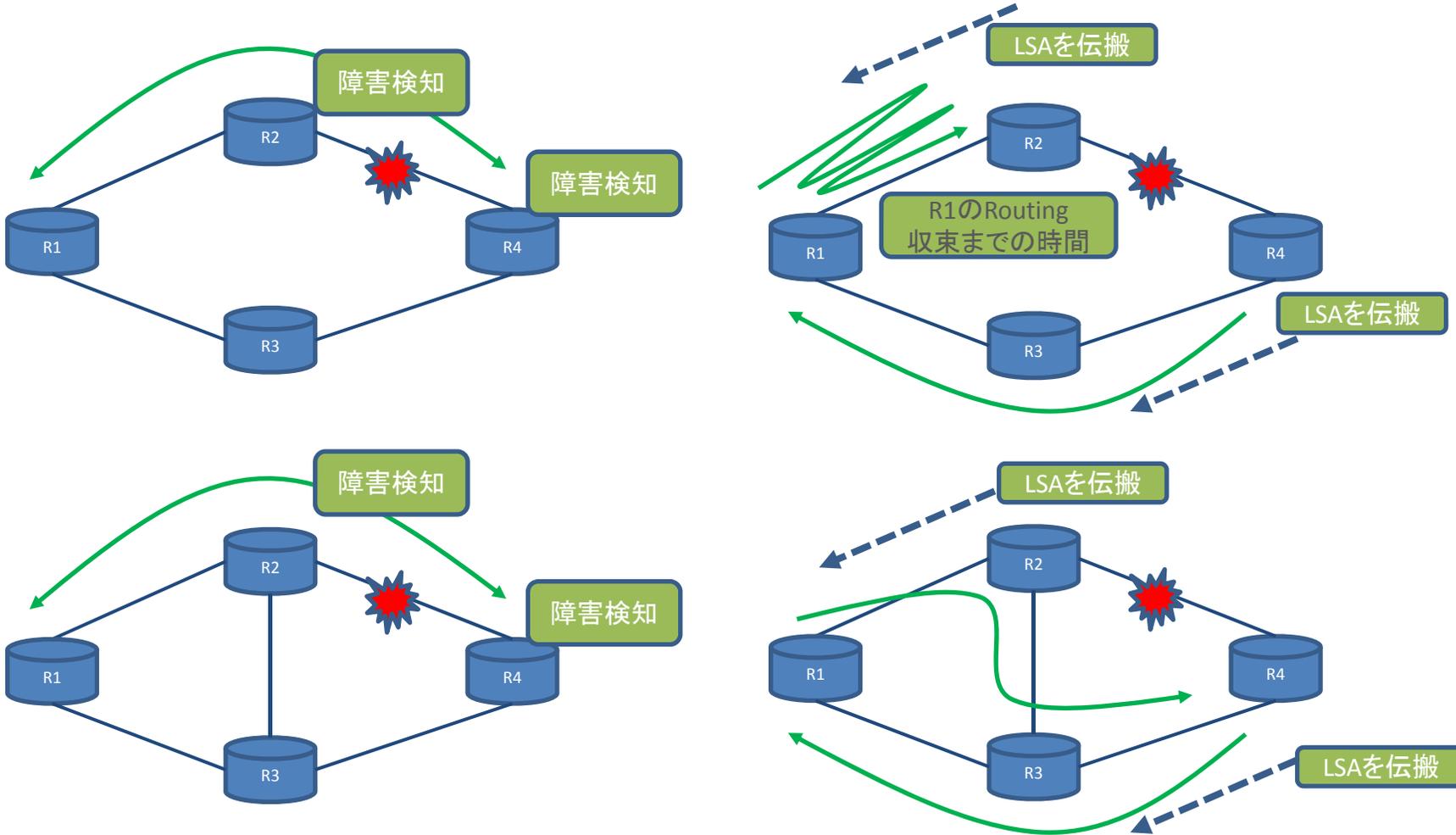
```
carrier-delay down <ms> up <ms>
upは早い必要はない
```

## OSPF timer

```
timers throttle spf <start> <hold> <max>
LSAを受け取ってSPFを開始する時間
timers throttle lsa all <start> <hold> <max>
LSAを生成する時間
timers lsa arrival <hold>
LSAを受信する時間
ベンダによく相談してから入れましょう
```

IOSのdefaultは決して早くはないが、設定変更する場合は他機器と値が異なれば、全体の収束がバラバラになるので要注意

# micro-loop



物理的なトポロジーでDown時間も変わってくる



# Dualstack 対応

- 性能
  - 機器性能の確認、IPv6転送処理
  - FIB容量(IPv4/IPv6/MPLS)
    - FIB階層化検討
- IGP選択
  - ospfv2/ospfv3のmulti-process
  - IS-ISのsingle-process
  - Support of Address families in ospfv3(RFC5838)
- 構成
  - v4/v6が異なるトポロジ―は超混乱の元！

# セキュリティ

- OSPF Authentication
  - 信頼できるリンクと信頼できないリンクはMD5は分けておく(つながっちゃーまずい)
  - ospfv3はISPECで守る
- Infrastructure ACL
  - ルータでの制御
    - 信頼されたブロックからのACL
    - RP防御のrate-limitなど
  - ASボーダでの制御
    - AS外部からインフラ宛への到達性は不要

お わ り