



ISPバックボーンネットワークにおける 経路制御設計

- その理論と実践 -

Intec NetCore Inc. 近藤 邦昭
France Telecom Long Distance 前村 昌紀



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



経路制御基礎 ～復習～

経路とは何か
そして
それを制御する意味とは



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



経路とは

- 経路
 - パケットが送信元から送信先に向けて転送されてゆく道筋である。
- 経路表とは
 - 受信したパケットのDestinationアドレスに向けてそのパケットを転送する次のルータへのアドレスを示したデータベースである。



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



ルータとは

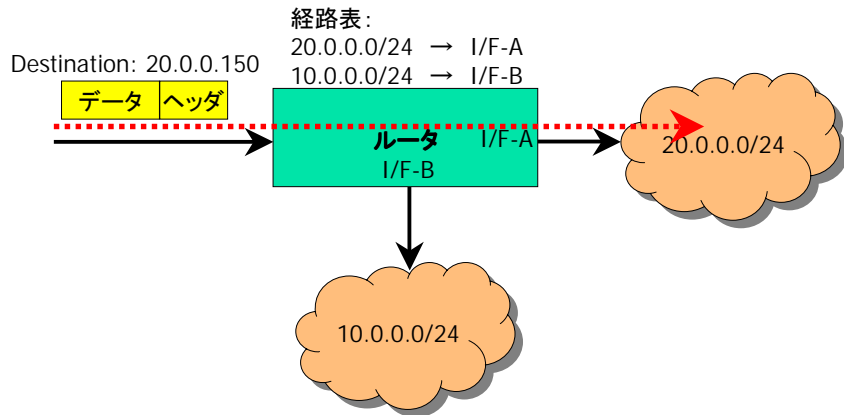
- IPパケットのヘッダに書かれているあて先情報を元に、ルータ内の経路表にしたがってパケットの送出先を決める機器をいう。
 - ってなことをあらわす図 & 文字



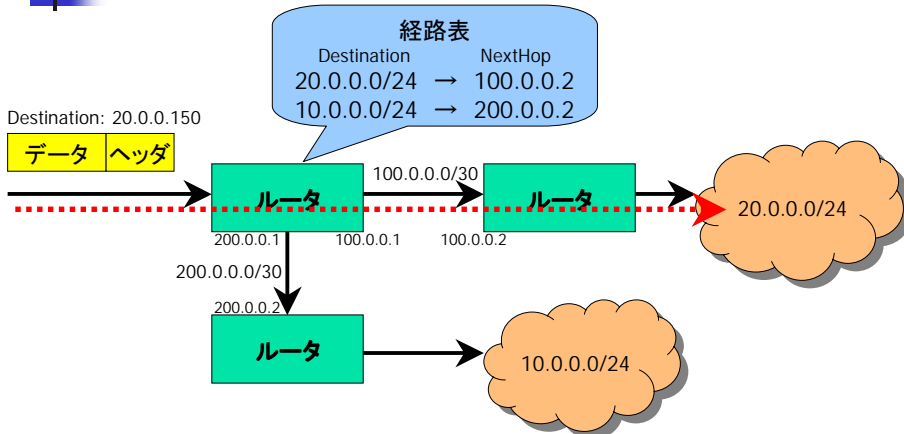
Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



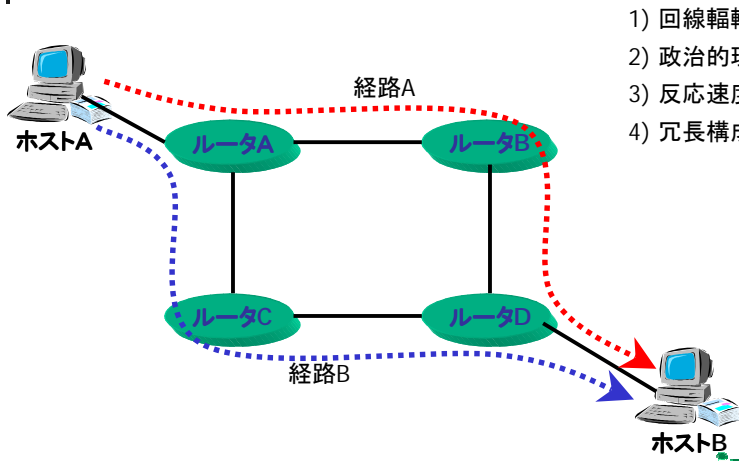
経路を決める



経路を決める - 実際には

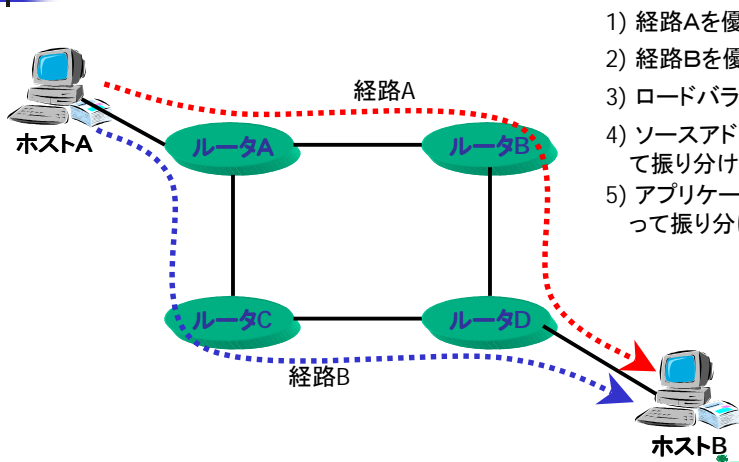


経路を制御する - その理由



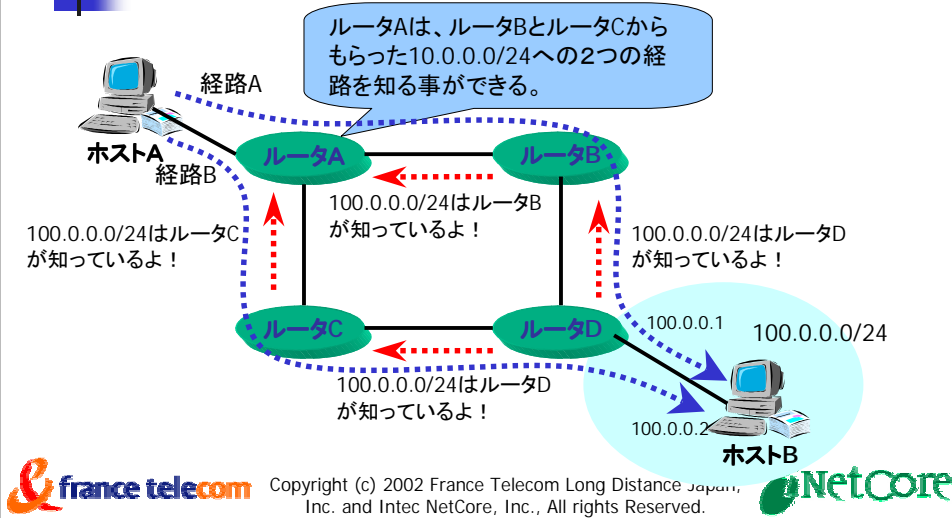
- 1) 回線輻輳
- 2) 政治的理由
- 3) 反応速度(RTT)
- 4) 冗長構成

経路を制御する - 技術的方法

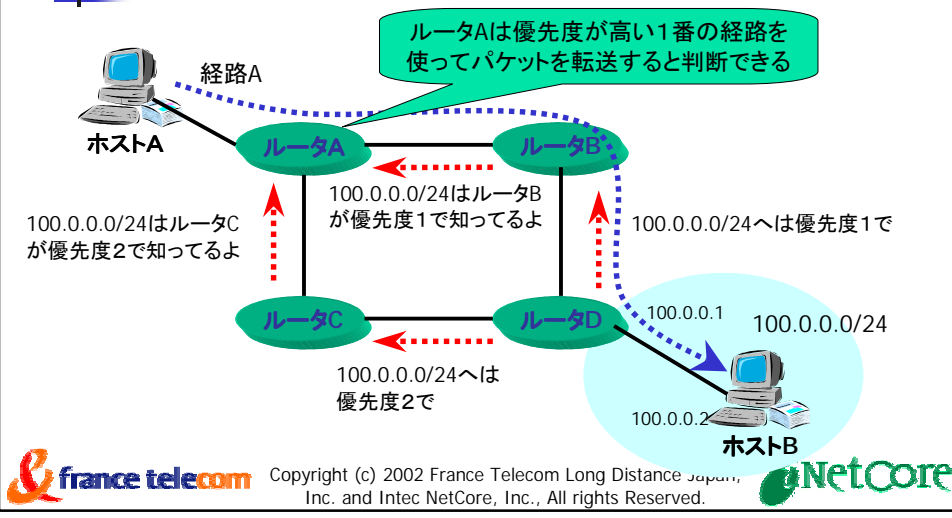


- 1) 経路Aを優先させる
- 2) 経路Bを優先させる
- 3) ロードバランスさせる
- 4) ソースアドレスによって振り分ける
- 5) アプリケーションによって振り分ける

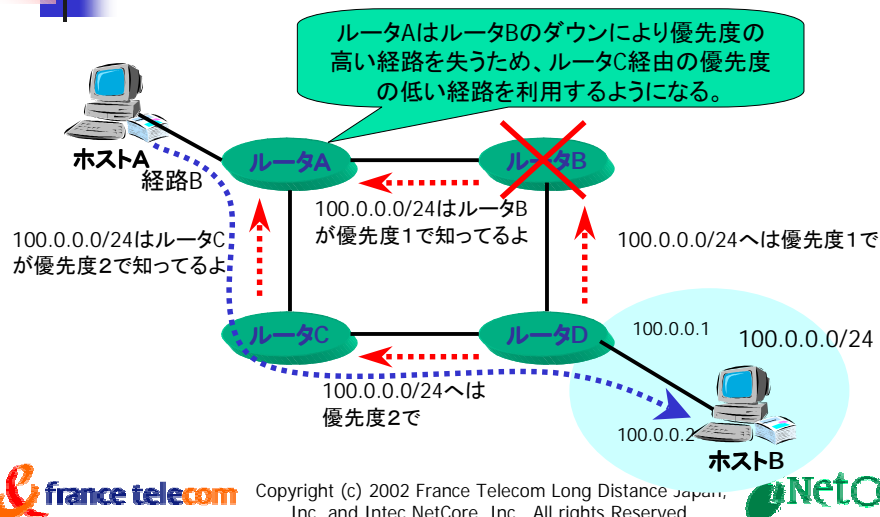
自動的に経路を制御する ～経路の伝播～



自動的に経路を制御する ～経路の選択～



自動的に経路を制御する ～経路の選択～



経路制御基礎 - おさらい

- 経路制御とは
 - 静的経路制御と動的経路制御がある
 - 目的のネットワークまでの経路を定義する
- 動的経路制御の必要性
 - 同一ネットワークまでに複数の経路がある場合に、そこまでの経路を自動的に構築する
 - 動的経路制御を行うことで、ネットワークの耐障害性を向上させる
- 動的制御経路の特徴
 - 経路の伝播方向とトラヒックの伝播方向は逆

大規模制御経路用語

～大規模には醍醐味がある～



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



14

捨てなくてはいけない概念

- 歴史的アドレス利用法
 - アドレスクラス : Class A,B,C
 - IPアドレスには、歴史的にネットワークの規模に応じたクラスという概念があった。たとえば、Class Aであれば、先頭ビットが0でネットマスクが255.0.0.0であるアドレスをいう。
- アドレスクラスを基本としたアドレス利用は既に過去のものである。
 - アドレスマスク長を"/"で記述する形式が標準
 - 192.168.100.128/25 ⇒
192.168.0.100.128/255.255.255.128



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.





VLSM

- Variable Length Subnet Mask
 - VLSMとは1つのネットワークをサブネットに分割する場合に複数の長さのサブネットマスクを利用する方法
 - 例えば、Class C相当のネットワークを分割するときに/26と/27を同時に利用したりすること
 - 例えば、同じアドレスクラスでは種類のアドレスマスク長しか使えないというものは、VLSMに対応していない。
 - Class Cで/24しか使えないというものはサブネットにも対応していないということである。
- クラスの概念があるときに利用されたもので、二つのClass Cをあわせて/23とかすることはできない。
 - 一般的にはSuperNetと呼ぶ。



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.




CIDR

- Classless Inter Domain Routing
- 経路情報を扱う際に完全に“アドレスクラス”という概念を取り払ったもの。
- ドメイン間(ISP間など)において経路情報を交換する場合に、アドレスPrefixというネットワーク情報とそのマスク長を交換することでクラス概念をなくした。
 - SuperNetをそのまま経路情報として扱うと思えばよい。



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



Route Aggregation

- 複数の経路を束ね、1つの経路に集約してしまうこと。
 - 192.168.100.0/23と192.168.102.0/23は、192.168.100.0/22という経路にまとめることができる。
 - 経路を集約することで、経路テーブル内の経路情報のエントリ数を抑えることができる。
 - 大規模ネットワークで経路数を抑えることは、安定運用のために必須といえる。
 - ネットワークの構成によっては集約できないケースも発生する。設計者の腕が試される部分でもある。
 - 詳しくは、後半にて...



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



経路制御プロトコル(IGP) ～基礎編～

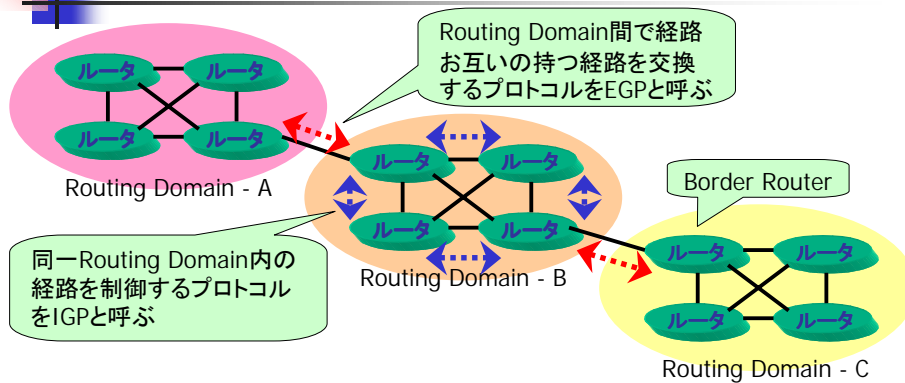
IGPとは何か
RIP



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



IGPとEGP



- IGP = Interior Gateway Protocol
- EGP = Exterior Gateway Protocol



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



RIP – Routing Information Protocol

- ルーティングプロトコルの中でもっとも単純なプロトコル
 - 距離ベクトル型 (Distance Vector)
 - RFC1058
- RIPのバージョン
 - RIPv2 – VLSM (Variable Length Subnet Mask)に対応
 - RIPv6 – IPv6版RIP
- 制御パラメータ
 - Metric ただひとつ(ただし最大は16)

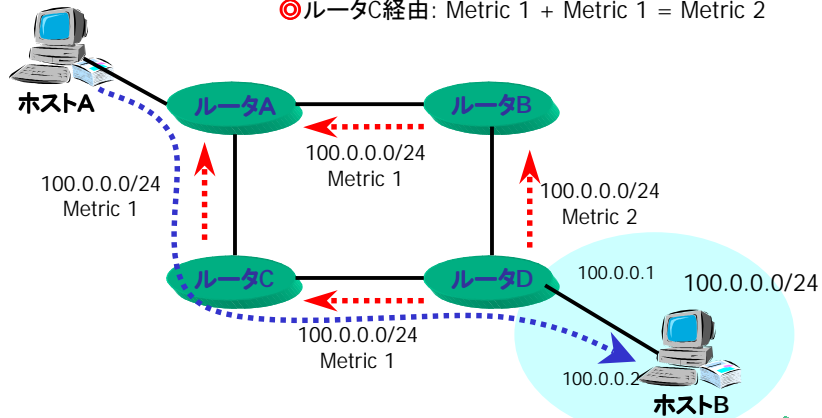


Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



RIP – 経路を制御する

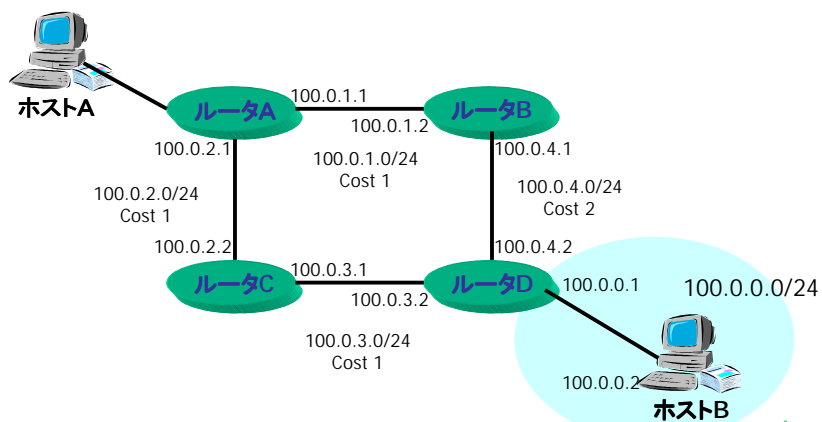
- × ルータB経由: Metric 2 + Metric 1 = Metric 3
- ◎ ルータC経由: Metric 1 + Metric 1 = Metric 2



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



RIP – Sample Network



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



Sample Configurations - RIP

ルータDの設定

```

hostname routerD
!
ip subnet-zero
!
interface Ethernet0
ip address 100.0.4.2 255.255.255.0
!
interface Ethernet1
ip address 100.0.3.2 255.255.255.0
!
interface Ethernet3
ip address 100.0.0.1 255.255.255.0
!
router rip
version 2
offset-list 1 out 1 Ethernet0
network 100.0.0.0
!
ip classless
!
access-list 1 permit any

```

ルータD-B間のMetricを1加算し、
Metricを2に設定している



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



OSPF

- 概論と特長
- とにかく動かして見る
- プロトコル上の仕組み
- 設計上の留意点



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



OSPF – 概論と特長

- どんなプロトコルか
- 精密な経路制御
- 良好な規模対応性



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



26

OSPFとはどんなプロトコルか

- Open Shortest Path Fast
- RFC2328 – 1998年4月
- IPの上に直接乗る。プロトコル番号89
- マルチキャストで隣接ルータとやり取りを行う
- リンクステートアルゴリズム



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



OSPFの特長

- 精密な経路制御(リンクステートアルゴリズム)
 - 高信頼化設計に適したトポロジ管理を実現
 - 素早い収束
- 良好な規模対応性(スケーラビリティ)
 - インクリメンタルアップデート(incremental update)
 - エリア分割機能
 - エリア毎のトポロジ管理
 - エリア毎の経路集約



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



コスト

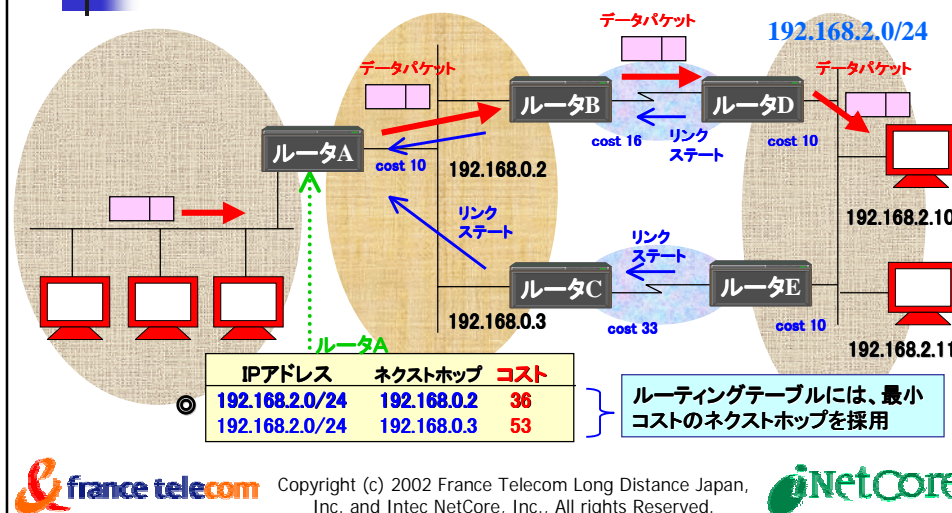
- 距離指標。1から65535までを取る
- リンクの回線容量に応じた自動的なコスト割付も可能
 - Ciscoの例: $\text{コスト} = 100\text{Mbps} \div \text{回線容量}$
- 対地までの全リンクのコスト総和を比較して経路を選択
- トポロジとして管理されるので、代替経路、ロードバランシングなどの高信頼化を実現



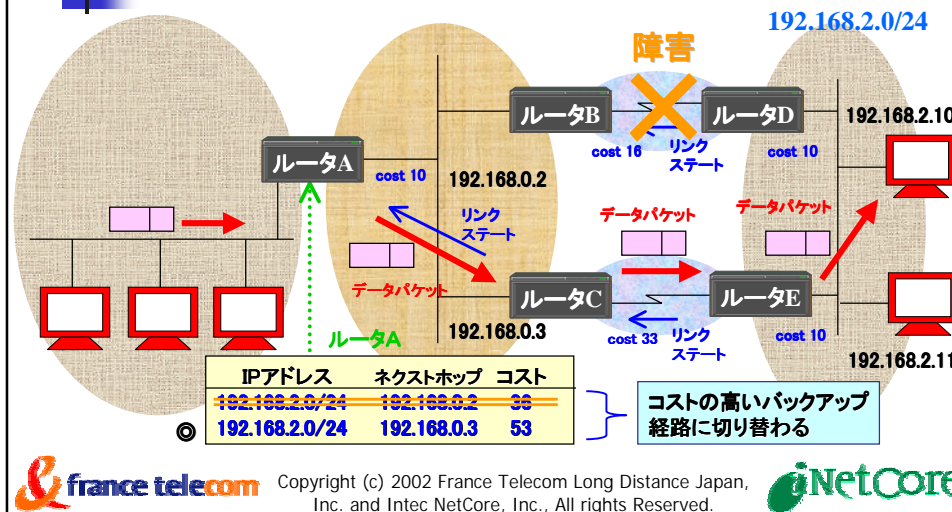
Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



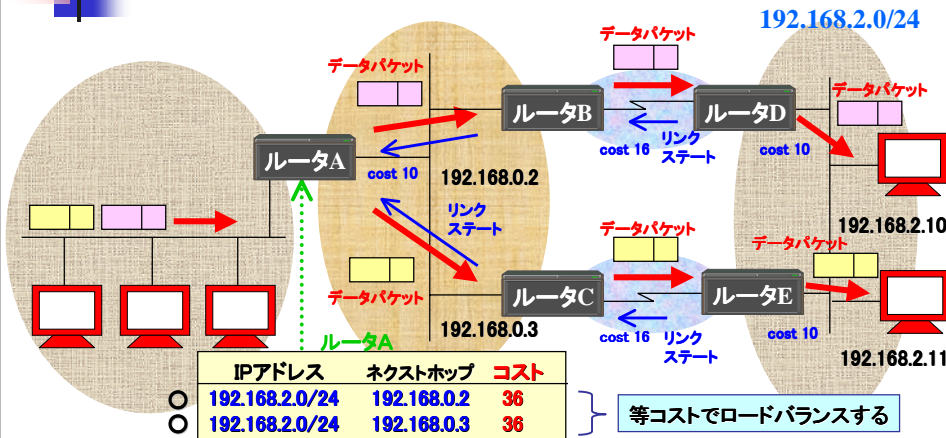
コスト - 1対地2経路の例



最小コスト経路障害時



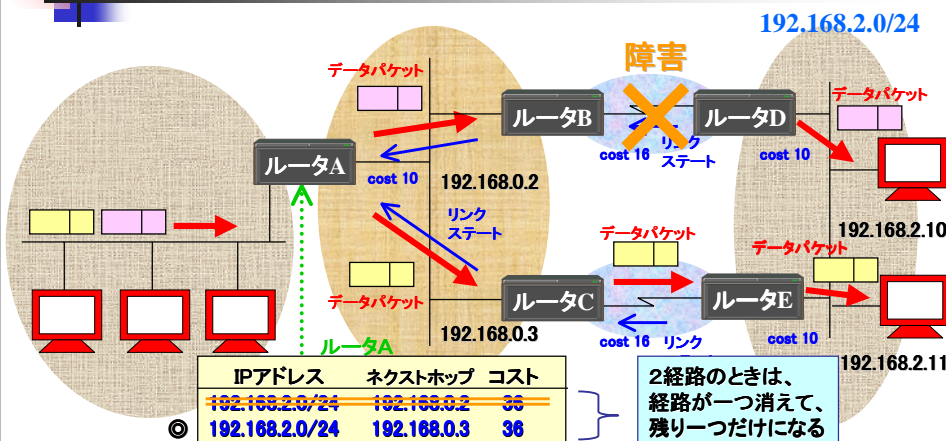
等コスト複数経路の例 ロードバランシング (load balancing)



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



等コスト複数経路の例 障害時の反応



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



エリア

- OSPF管理領域を「エリア」で分割
 - エリアIDは32ビット空間。IPアドレスの形式も可能。
 - バックボーンエリア(area0)が全てのエリアに接する
 - 同一エリア内の全ルータでトポロジデータベースは共通
 - 他のエリアはトポロジの管理は行わず、対地経路として扱う
 - 経路の集約が可能

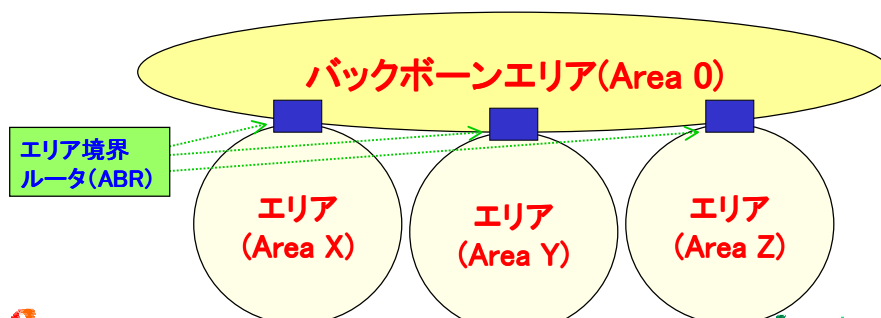


Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



エリア

- 各エリアとバックボーンエリアを結ぶルータをエリア境界ルータと呼ぶ
- 集約経路、他エリアに対するサマリの生成



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



OSPF – とにかく動かして見る

- 基本コンフィグレーション
- ループバックインターフェースの勧め
- 基本的な適用テクニック



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



36

CiscoにおけるOSPF設定

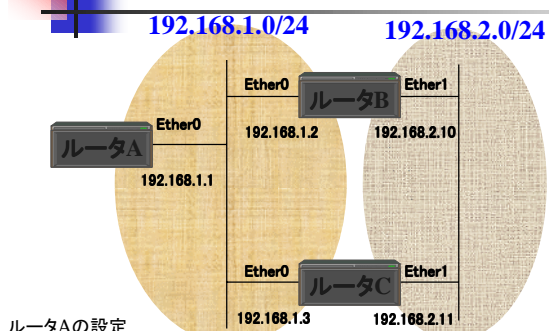
- router ospf <process ID>
 - 自分のASと同じ番号にすることが多い
 - 一つのAS内で一つしかOSPF processを走らせない場合
 - process IDは1~65535の何番にしてみてもいい
- network 192.168.0.0 0.0.0.15 area 0
 - 0.0.0.15はワイルドカードマスク
 - アドレスのうち無視する部分をマスクする
 - 192.168.0.0~192.168.0.15の範囲にあるアドレスのインターフェースで
 - OSPFを area 0 で話す
 - そのインターフェースのネットワークをOSPFに広告する



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



基本設定



ルータAの設定

```
interface Ethernet0
ip address 192.168.1.1 255.255.255.0
!
router ospf 1
network 192.168.1.0 0.0.0.255 area 0
```

ルータBの設定

```
interface Ethernet0
ip address 192.168.1.2 255.255.255.0
!
interface Ethernet1
ip address 192.168.2.10 255.255.255.0
!
router ospf 1
network 192.168.1.0 0.0.0.255 area 0
network 192.168.2.0 0.0.0.255 area 0
```

ルータCの設定

```
interface Ethernet0
ip address 192.168.1.3 255.255.255.0
!
interface Ethernet1
ip address 192.168.2.11 255.255.255.0
!
router ospf 1
network 192.168.1.0 0.0.0.255 area 0
network 192.168.2.0 0.0.0.255 area 0
```



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



ループバックインターフェースの勧め

■ ルータID

- OSPFのやり取りの際にルータ同定のために利用
- ループバックインターフェースを定義するのが定石
- Ciscoの場合、ループバックがない場合物理インターフェース中最大のIPアドレスが選ばれる
 - ダウンしたら他を利用。全てのOSPFのやり取りをやり直し

```
Interface loopback 0
description *for routerID and iBGP*
ip address 192.0.2.1 255.255.255.255
!
Router ospf 5511
network 192.0.2.1 0.0.0.0 area 0
passive-interface loopback 0
```



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



適用テクニック

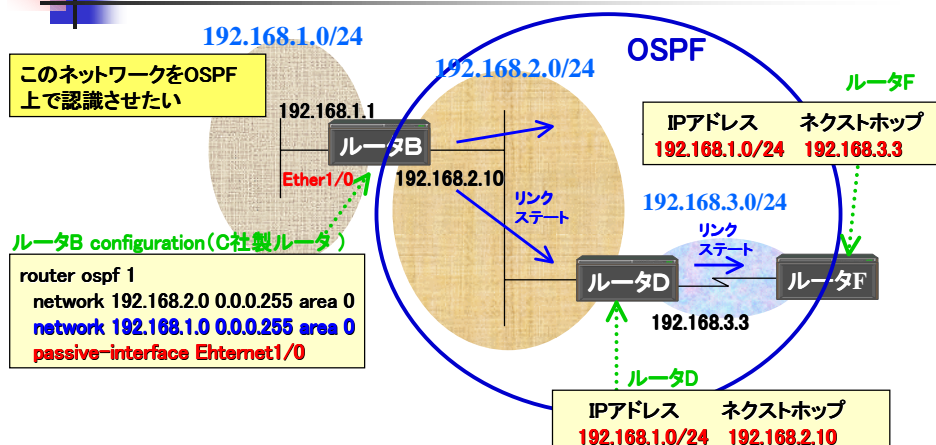
- Passive-Interface
 - ネイバがないセグメントをOSPFで認識させる
- Redistribute
 - Static などOSPF以外の経路情報をOSPF上で流通する
- Default-information originate
 - 特にデフォルトルートをOSPF上で流通する



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



passive-interface



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



外部経路(External Routes)

- OSPFに対してredistributeされた経路外部経路を生成するOSPFルータをAS境界ルータ(ASBR: AS Border Router)と呼ぶ

- ここでのASはOSPF領域のみを指す

- 扱い方に2つのタイプ

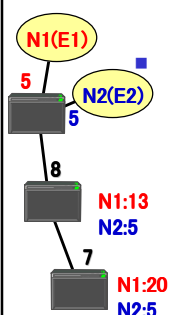
- Type1: redistributeされたメトリック+OSPFコスト

- Redistribute時のメトリックを一定にすれば、最も近いASBRに経路制御される

- Type2: redistributeされたコストのまま

- Redistribute時のメトリックで優先制御を決定付ける

- 同一プリフィクスに対して以下のように優先制御
IntraArea>InterArea>External 1>External 2



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



OSPF – プロトコル上の仕組み

- リンクステートデータベース
- DR, BDR, ネイバ, Adjacency
- パケットタイプ, LSAタイプ, スタブエリア



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.

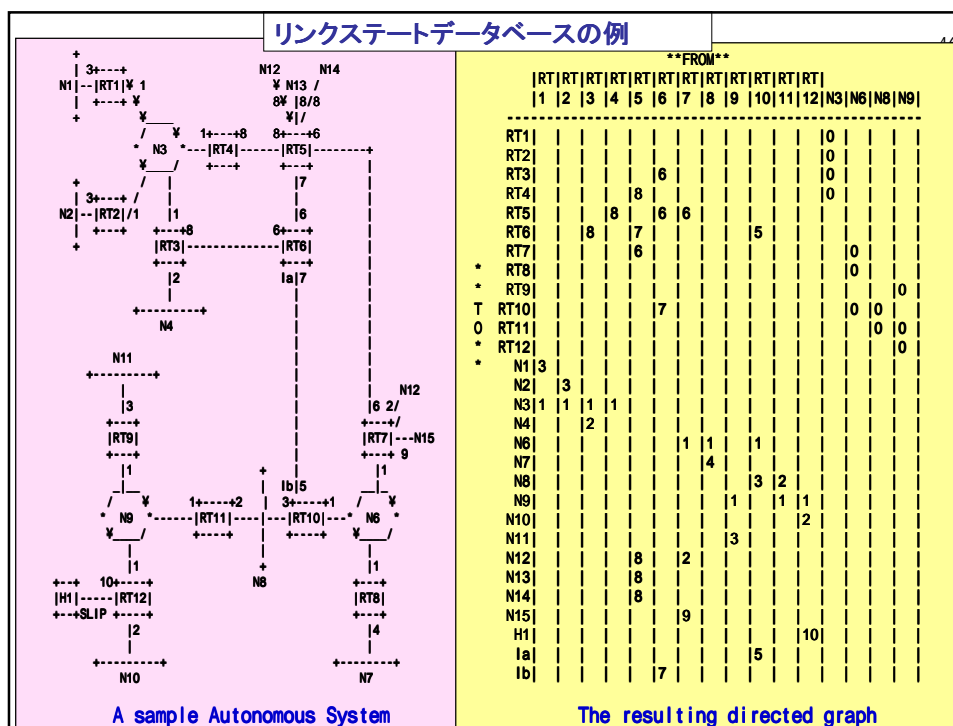


リンクステートアルゴリズム

- 管理領域(エリア)内の全てのリンクに関して、全ルータでリンクステートをデータベースとして管理
 - リンクステートデータベース(LSDB)
 - エリア内のトポロジを管理している
- SPF(shortest Path First)アルゴリズムで最短パスツリーを作成して、
- ルーティングテーブルを生成する



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



マルチプルアクセスネットワーク におけるOSPFのやり取り

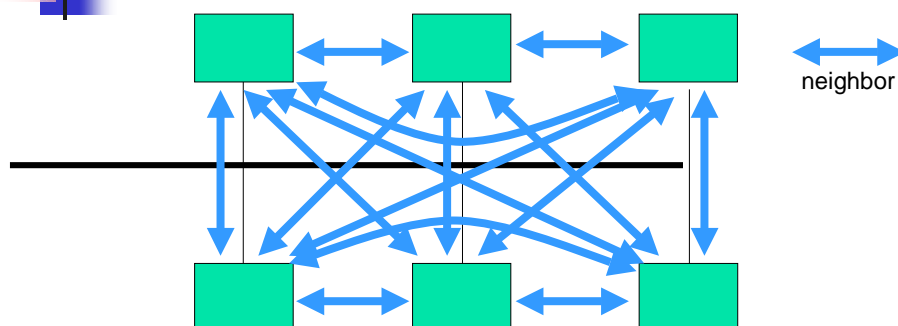
- 隣接ルータ=ネイバ(neighbor)
- フルメッシュでやり取りすると非効率なので
- DR(Designated Router: 指名ルータ)が代表して全ネイバとのリンクステートのやり取りを行う
- DRのバックアップとしてBDR(backup DR)も、全ネイバとリンクステートをやり取りする
- リンクステートをやり取りする関係: Adjacency(アジャセンシー)



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



ネイバ



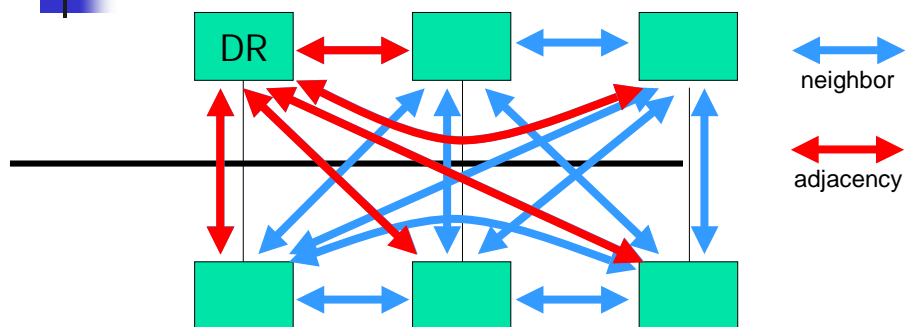
ネイバーはHelloパケットのやりとりを行う。



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



DR(指名ルータ)とAdjacency



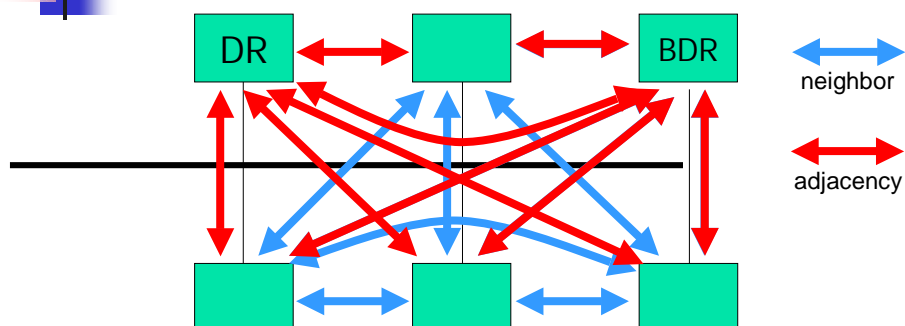
DRはネットワークを代表して
全ネイバとリンクステートのやり取りを行う



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



BDR



DRのバックアップとして
BDRも全ネイバとAdjacentになる



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



DR, BDRの選出方法

- Priority値を比較して、高いほうが優先される
- しかしネイバの新規追加ごとに再選出はしないので、結局最初に立ち上がった2つがDR,BDRを担当
 - DR,BDR以外: DROTHER
- ospf priority 0はDR/BDRに選ばれない
 - 負荷が大きくなると困るルータなどは0にする
 - DR,BDRの負荷は高くなるので、2つのセグメントをまたぐルータなどは選ばれないようにする



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



OSPFのパケットタイプ

パケットタイプ	内容
Hello	ネイバー確立と維持 10秒ごとにマルチキャスト
Database Description	アジャセンシー確立時のDB内容の伝達
Link-State Request	リンクステート情報の要求
Link-State Update	リンクステート情報の伝達・更新
Link-State Acknowledgement	Link-State Updateの受領確認



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



リンクステート広告(LSA)

- Link-State Update パケットの中に複数含まれる。
- リンクステートに関する情報を伝達

LSA名	内容
ルータLSA	ルータと接するリンクのリンクステート情報
ネットワークLSA	リンクに接するルータのリスト
サマリLSA	エリア外対地の経路情報
AS外部LSA	外部経路の経路情報



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



スタブエリアとNSSA

- スタブエリア(Stub Area)
 - 外部経路を持たないエリア
 - エリア外経路をデフォルトルートで代用し、処理を軽くする目的
 - 但し、ユーザのスタティックルートをredistributeすることを考えると、スタブエリアが適用できる局面が少ない
- NSSA(Not So Stubby Area : 準スタブエリア)
 - スタブ的なエリアでredistributeされる外部経路に対して、別のLSAタイプを適用、他の外部経路、エリア外経路との区別を可能としたもの



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



OSPF – 設計上の留意点

- 大原則・諸元
- Keep it simple
- 規模対応



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



56

大原則

- 安定動作の保証が欲しければ、ルータベンダなどのコンサルテーションを受け、当たり前に見える事項も全て確認を取る
 - あらゆるシステム設計の大原則ですが
 - メモリ実装量や諸元・設計限界などに関して
- 日常運用の局面で性能指標をよく監視し、性能が著しく低下したらアップグレードを考える
 - ネットワーク変更の収束時間の急増
 - CPU負荷の急増



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



Keep it simple

- 必要以上に複雑な構造にしない
 - むやみにエリア分けしない
 - 数十台規模ならarea0だけでも大丈夫？
 - (但し大原則に従ってくださいね)
 - エリア境界ルータは3エリア(area0含む)以上受け持たない
 - むやみにリンク数を増やさない
 - 冗長性確保はできるだけ二重化に留める
 - Point-to-pointリンクのメッシュより、マルチプルアクセスLANの1セグメントのほうがいい
 - 小さいルータ多数より、大きなルータ一台



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.



負荷配分と規模対応

- DR,BDRを集中させない、非力なルータにやらせない
- エリアごとに経路集約ができるようにアドレッシングする
- デフォルトルートで済むところはデフォルトルートを活用する



Copyright (c) 2002 France Telecom Long Distance Japan, Inc. and Intec NetCore, Inc., All rights Reserved.

