



## IPv6 ～ ISP的な運用 ルーティング編

---

向井 将  
株式会社パワードコム  
IPv6運用研究会 Co-Chair  
InternetWeek 2001



## 内容

---

- ルーティングアーキテクチャ
  - IGP
  - EGP
  - マルチホーム/トラフィックコントロール/IX
- IPv4からIPv6へ
  - 運用上の課題
- そしてIPv6へ



## はじめに

---

- このセッションでは、ルーティングにテーマを絞ってIPv4とIPv6の違いについてお話します
- また、IPv6オペレーション研究会での検討結果を一部引用しています



## ルーティングアーキテクチャ

---

## ルーティングプロトコル

---

- IGP
  - 自組織のネットワークの内部で利用されるルーティングプロトコル
- EGP
  - 他組織のネットワークとの間で利用されるルーティングプロトコル

## ルーティングプロトコル

---

- IGP
  - RIP
    - 小規模ネットワークで利用されることが多いルーティングプロトコル
  - OSPF
    - 中規模から大規模までのネットワークで利用されるルーティングプロトコル
- EGP
  - BGP
    - AS(独自のポリシーをもったネットワーク)でルーティングを行う

## RIP

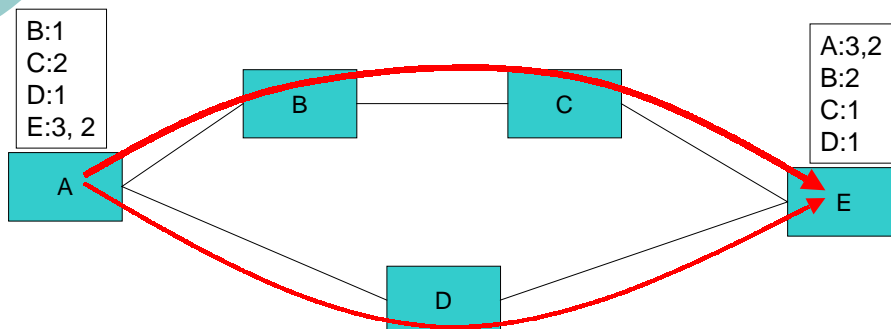
- 小規模なネットワークで利用されるプロトコル
- 距離ベクトル型
- destinationまでの距離(hop数)をRIPルータ間でやり取りをする
  - hop数が小さい方を選択する
  - hop数は15hopsが限界
- RIPv1
  - RFC1058
  - UDP(port 520)をつかう
- RIPv2
  - RFC1388
  - UDPに加え、マルチキャストを使ってルータ間でやり取りが可能になった

InternetWeek2001

7

## RIP

- AからEの通信を考えた場合
  - 通常は、A-D-Eを利用
  - D-E間のリンクが落ちた場合、A-B-C-Eを使う



InternetWeek2001

8

## OSPF

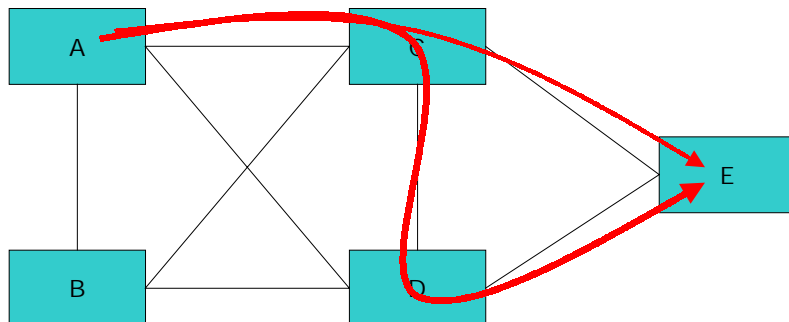
- 中規模から大規模ネットワークまで利用できるプロトコル
- リンクステート型
- リンクに重み(cost)を与え、destinationまでのコストの和が一番小さいものを選ぶ
  - 自ノードを中心としたツリーをつくりSPFアルゴリズムを用いて最短パスを見つける
- RIPに比べ障害時の経路収束が早い
- OSPFルータ間はIP(port 89)を利用する
- OSPFv2
  - RFC2328

InternetWeek2001

9

## OSPF

- AからEの通信を考えた場合
  - 通常はA-C-Eを利用
  - C-E間のリンクが落ちた場合、A-C-Eを利用



InternetWeek2001

10

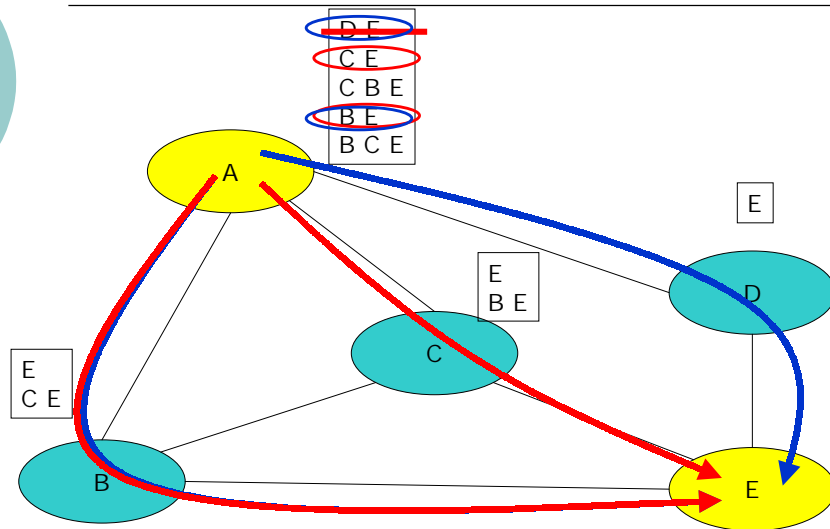
## BGP

- 今のインターネットを構成するプロトコル
  - AS(独自のポリシーをもったネットワーク)単位でルーティングを行う
  - このASの中でIGP(RIP、OSPF)が利用される
- パスベクトル型
- RFC1711で定義
- BGPルータ間ではTCP(port 179)を利用
  - TCPを利用することで、フロー制御などを任せることができ、ルーティングプロトコルの設計が簡単になる
- BGPは2種類
  - iBGP: AS内部のBGPルータ間で利用
  - eBGP: AS間で利用されるBGPルータ間で利用

InternetWeek2001

11

## BGP



InternetWeek2001

12



## ルーティングに関連する技術

---



## ルーティングに関連する技術

---

- マルチホーム
  - インターネットに対する冗長性を目的とした技術
- トラフィックコントロール
  - 経路情報をコントロールすることにより、流入するトラフィックをコントロールする手法
- IX (Internet eXchange)
  - 経路・トラフィックを交換する目的で作られた共用ポイント

## マルチホーム

### ○ 目的

- インターネットに対して複数の到達性をもたせるため

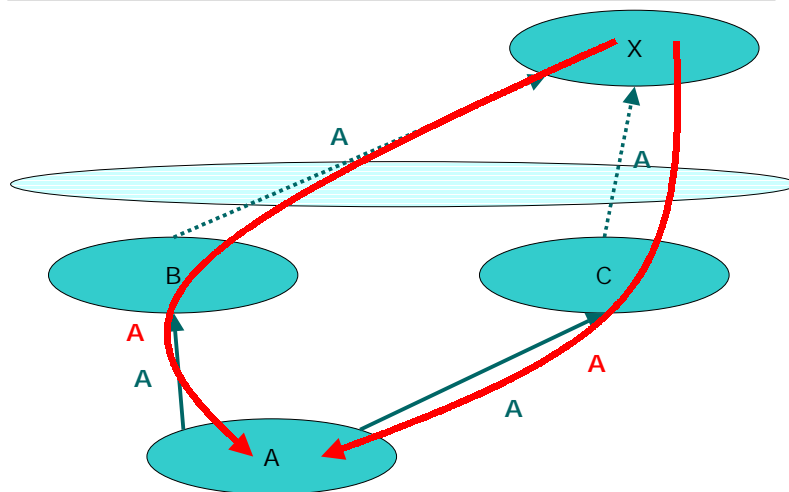
### ○ 方法

- static、BGP、NATなどさまざまなモデルがある
- 複数の到達性を得るということは、冗長性が得られることを示す
  - 複数の到達性 = 複数の回線

InternetWeek2001

15

## マルチホーム(BGP利用)



InternetWeek2001

16



## トラフィックコントロール

- 目的
  - アナウンスする経路情報を変えることにより、流入するトラフィック量をかえること
- 必要なケース
  - upstreamにつながるルータの負荷が高くなってきた場合
  - upstream向けの回線帯域が逼迫した場合
- 複数の回線、複数のupstreamを必要とし、マルチホームを行う

InternetWeek2001

17

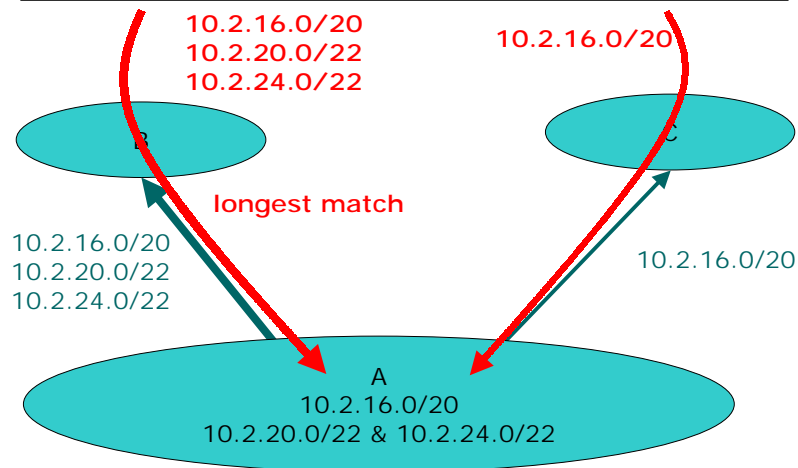
## トラフィックコントロールの手法

- 経路情報の選択手順によるもの
  - 最長一致優先規則を利用
  - 経路選択する際に、ネットマスク長/prefixの長さが長いものを優先することを利用
- ルーティングプロトコルによるもの
  - オプションを利用し、経路選択手順を意図的に変え、特定の経路に対する優先度を変更する
  - 例: BGPの経路選択アルゴリズム

InternetWeek2001

18

## 経路情報の最長一致優先規則を用いる 場合のトラフィックコントロール



InternetWeek2001

19

## punching hole

### ○ 定義

- ISPは経路集約を目的として、マスクでまとめられる形で割り振りをうけている(CIDR: RFC1519, Classless Inter-Domain Routing)
- Internet Registryから割り振りをうけたアドレス空間の一部を他のASからアナウンスをすること

### ○ 目的

- AS番号を取得するほど大規模ではないネットワークがマルチホームを行うために割り当てをうけたISP以外からもそのアドレスの経路情報を流す
  - /24の場合が多く観測されている

InternetWeek2001

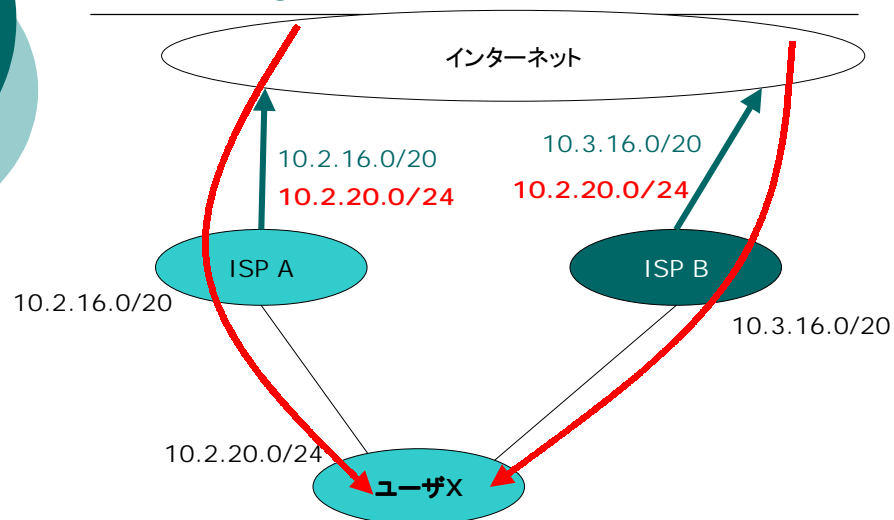
20

## punching hole

### ○ 問題点

- IPv4インターネットの経路情報数(フルルート)が増大する
- 経路の収束により多くの時間を必要とする

## punching hole



## IPv4インターネットの経路情報数(フルルート)

- punching holeの影響でIPv4フルルートは増大している

- 全体の6割を占める

Number of prefixes announced per prefix length (Global)

/1:0	/2:0	/3:0	/4:0	/5:0	/6:0
/7:0	/8:18	/9:5	/10:7	/11:12	/12:34
/13:89	/14:218	/15:382	/16:7034	/17:1321	/18:2402
/19:7040	/20:5951	/21:4759	/22:7104	/23:8967	/24:60644
/25:304	/26:292	/27:282	/28:277	/29:158	/30:191
/31:1	/32:75				

APNIC bgp-stats MLより

InternetWeek2001

23

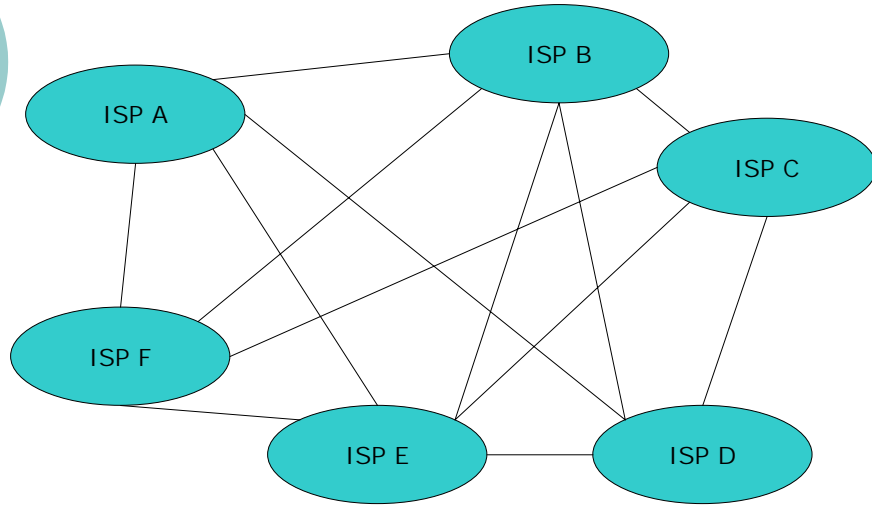
## IX(Internet eXchange)

- 目的
  - ISPが複数の他ISPとの接続のため、設置する共用接続ポイント
  - コスト、運用負荷の低減を図れる
- 一般的にはLayer2(LAN形態)での接続性がある
- 通常BGPを使用して経路交換を行う
- 有名なIX
  - 日本: NSPIX、JPIX、JPNAP
  - 海外: PAIX、STARTAP、AMS-IX

InternetWeek2001

24

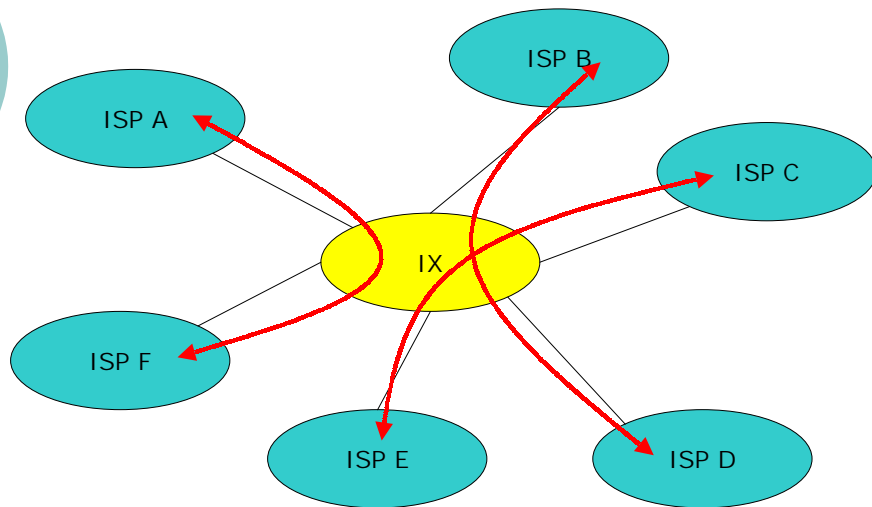
## IXがない場合



InternetWeek2001

25

## IXがある場合



InternetWeek2001

26



## IPv4からIPv6へ

---



## アドレスアーキテクチャ

---

- IPv6ではアドレス長が128bits
  - IPv4は32bits
- RFC2373、RFC2374にて定義
  - RFC2373  
IP Version 6 Addressing Architecture
  - RFC2374  
An IPv6 Aggregatable Global Unicast Address Format

## アドレス体系

---

- リンクローカルアドレス
  - 同一サブネットの中でのみ有効なアドレス
- グローバルユニキャストアドレス
  - インターネットにて一意性が確保されたアドレス
- サイトローカルアドレス
  - 同一サイト内でのみ有効なアドレス
- マルチキャストアドレス
  - ネットワークインターフェースのグループに到達するアドレス
- エニキャストアドレス
  - ネットワークインターフェースのグループに到達するアドレス
  - ただし、最初のインターフェースに到達をした後は、他のインターフェースにパケットを転送しない

## 現在のアドレスの割り振り

---

- IPv4
  - 日本であれば、JPNICより割り振りをうける
- IPv6
  - RIR(Regional Internet Registry)より割り振りをうける

## Internet Registry

---

- 定義
  - IPアドレス空間をメンバーまたは顧客に分配し、その分配を登録する責任を持つ。
    - 「JPNICにおけるアドレス空間管理ポリシー」より引用
- Internet Registryは階層構造をもち、その構造の下でIPアドレスは割り振られる
  - RFC2050で定義

InternetWeek2001

31

## Internet Registry/RIR/LIR

---

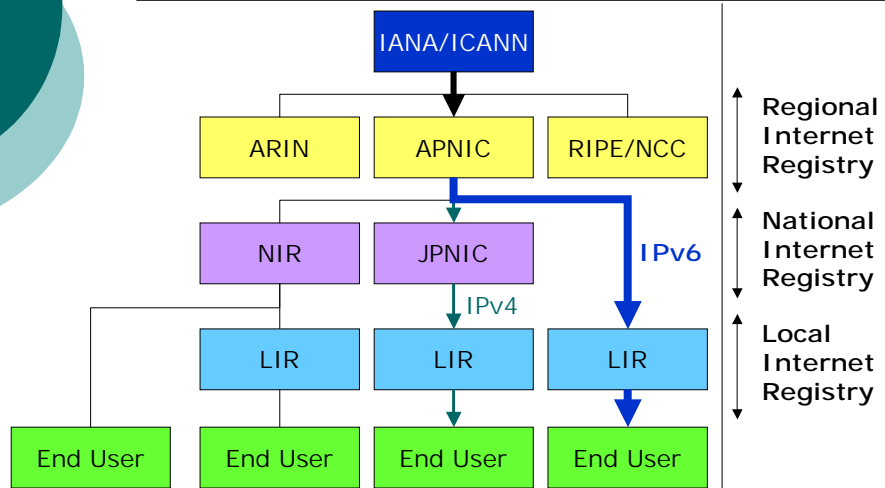
- RIR(Regional Internet Registry)は大きな地域にたいしてアドレスの割り振りを行う組織
  - APNIC: アジア・太平洋地域
  - ARIN: アメリカ大陸・アフリカ南部
  - RIPE/NCC: ヨーロッパ・アフリカ北部
- RIRは、NIR(National Internet Registry)へアドレスを割り振り、LIR(Local Internet Registry)へ割り振る
  - NIR: 各国に存在するInternet RegistryでLIRへアドレスを割り振る(日本ではJPNICがこれにあたる)
  - LIR: NIRより委任をうけエンドユーザにアドレスを割り当てる(IPアドレス管理指定事業者がこれにあたる)

InternetWeek2001

32



## Internet Registryの階層構造と アドレス割り振り



InternetWeek2001

33

## IPv6でのルーティングプロトコル

## IPv6でのルーティングプロトコル

---

- 通常、ルーティングプロトコルはリンクローカルアドレスで通信を行う
  - IPv6マルチキャストアドレスを利用可能な場合もある
- 認証はIPSecを用いるモデルが提示されている
  - RIPng、OSPFv3

## IPv6でのルーティングプロトコル

---

- IGP
  - RIPng
  - OSPFv3
- EGP
  - BGP4+

## RIPng

---

- RIPv2をIPv6対応
- RFC2080で定義
  - IPv4版からの変更は最小限にとどまる
- IPv6対応のため、ルーティングエントリをIPv6アドレス分(128ビット)に変更
- IPv6マルチキャストアドレスをつかって隣接RIPngルータと通信が可能
  - ff02::9に対してアナウンス
  - リンクローカルアドレスを使って通信も可能

InternetWeek2001

37

## RIPng Configuration

---

- C社での設定例
  - RIPngを有効
  - default routeを生成
  - RIPngを利用するインタフェースでRIPngを有効にする

```
ipv6 router ripng foo
  default-information originate
```

```
interface ethernet 0
  ipv6 ripng foo enable
```

InternetWeek2001

38

## OSPFv3

---

- OSPFをIPv6に対応
- RFC2740で定義
  - RIPngと同様、変更は最小限にとどまる
  - OSPFルータを識別するroute-id
    - IPv4アドレスを利用していたものから32bitsの識別子にかわる
- IPv6マルチキャストアドレスで通信が可能
  - ff02::5 : OSPFルータ間
  - ff02::6 : DR、BDRとOSPFルータ

## OSPFv3 Configuration

---

- ルーティングソフトウェアzebraでの設定例
  - OSPFv3を有効にし、router-idを設定
  - インタフェースにてareaの設定を行う

```
router ospf6
router-id x.x.x.x
interface fxp0 area 0.0.0.0
```

## BGP4+

---

- BGP4をIPv6用に拡張
  - 複数のプロトコルに対応するためにマルチプロトコル対応 (mBGP: Multi protocol BGP)
- RFC2545、RFC2858で定義
  - RFC2545  
Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing
  - RFC2858  
Multiprotocol Extensions for BGP-4
  - BGP4+はIPv4が動作することが前提
  - OSPFと同様にrouter-idは識別子

## BGP4+

---

- neighbor address
  - グローバルアドレス、リンクローカルアドレスのどちらでもよい
- next-hop address
  - グローバルアドレス、リンクローカルアドレスの両方



## BGP4+ Configuration

---

### ○ C社での設定例

- AS100とAS200でのpeering
- BGP4+のみのpeer

```
router bgp 100
  no bgp default ipv4-unicast
  neighbor 3FFE:1100:0:CC00::1 remote-as 200
```

```
address-family ipv6 unicast
  neighbor 3FFE:1100:0:CC00::1 activate
exit address-family
```



## IPv6でIGPの運用はどう変わるのか？

---

## IGP

---

- IPv6アドレスは経路集約を目的としたアドレス構造をとっている
- どの大きさprefixをふればよいのか？
  - loopback address
  - ルータ間のアドレス
  - ユーザへ割り振るアドレス空間

## IGPでの経路集約

---

- IGPの経路数は、IPv6になってもIPv4と同じくらいになるだろう
- しかし、IPv6なので出来るだけ経路数は減らしたい
  - IPv6のネットワークをはじめからつくるので、経路集約を念頭においたネットワーク設計ができる

## ISP内部におけるアドレス割り当て

- 手法
  - ネットワークをいくつかのグループに分割
  - ある程度まとまった大きさをPOPにアドレスを割り振る
    - 例えば、回線数、顧客の需要予想数など
    - /40、/42など
    - その中からloopback addressや、ルータ間のアドレスを割り当てる
  - まとまった単位で経路集約を行っていけば、ある程度まで集約は可能になる

InternetWeek2001

47

## ISP内部におけるアドレス割り当て

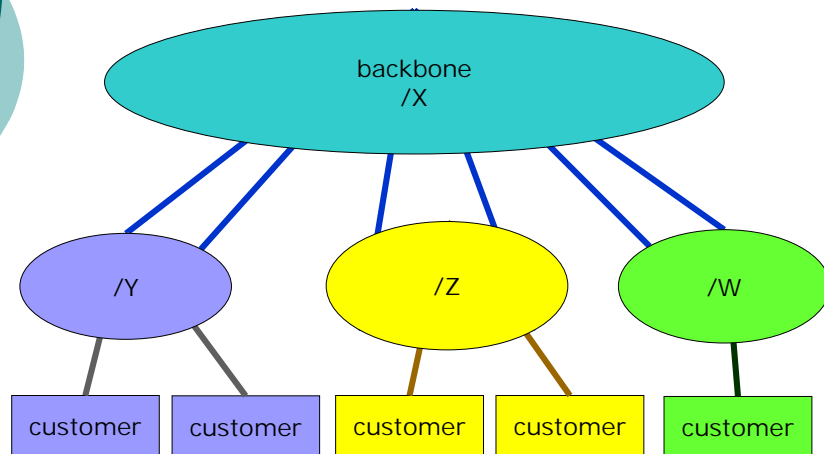
- POPに割り当てるprefixとユーザ数(/48を割り当てるとする)
  - /39 per POP : 512ユーザ
  - /40 per POP : 256ユーザ
  - /41 per POP : 128ユーザ
  - /42 per POP : 64ユーザ
- POPに割り当てたprefixからユーザに割り当てる

InternetWeek2001

48



## ISP内部におけるアドレス割り当て



InternetWeek2001

49

## ISP内部のアドレス割り当て

- loopback address
  - /128
- ルータ間のアドレス
  - /64
  - IPv4で/30で割り当てていた感覚だと/126という議論もあるが、/64で十分
  - /48から/64は65535個割り当てられる
- ユーザに割り当てるアドレス
  - /48, /64, /128
  - サービスにもよる

InternetWeek2001

50



## IPv6でEGPの運用はどう変わるのか？

---



## IPv6でEGPの運用はどう変わるのか？

---

- EGPの運用
- マルチホーム
- トラフィックコントロール
- IX(Internet eXchange)

## EGPの運用

- BGP4+で運用をする場合、他のISPと隣接するルータ(site exit router)では、受ける経路があらかじめわかっているのでフィルタリング可能
  - prefixの長さでフィルタリングが出来る
  - C社であれば、prefix-listを利用

## EGPの運用

- IPv6フルルートにアナウンスされる経路は2種類
  - 6bone(IPv6の実験網)の経路
  - RIRが割り振った経路
- 6boneの経路
  - 3ffe::/16の空間がIANAより割り振られている
  - RFC2772をベースでfilterを書くことができる
    - 6boneでのガイドラインが書かれている
- RIRが割り振った経路
  - 現在は、2001::/16の空間が割り振られている
  - 6boneとは違い、商用利用可能なアドレス空間

## 6bone

---

### ○ 目的

- IPv6の運用・研究を目的として、世界規模で構築されたネットワーク
- IPv6の実験用として、3ffe::/16のアドレス空間が割り振られている

## IPv6インターネットの経路情報数 (IPv6フルルート)

---

- 現在、190経路前後流れている
  - 6ヶ月で2倍弱に増加している
- 6boneの経路とRIRが割り振った経路の割合
  - 6boneの経路: 90経路前後
  - RIRが割り振った経路: 90経路前後
  - ほぼ5:5の割合になっている

## EGPの課題

---

- マルチホーム
  - 経路集約が進むため、マルチホームが困難になる可能性がある
  - punching holeをすると、経路集約は破綻する
    - IPv4の繰り返し
- トラフィックコントロール
  - 1ASにつき、そのASがもつprefixしかアナウンスできないため、ASの出口となるルータの一部にトラフィックが集中する可能性がある

## IPv6マルチホーム

---

- 概要
  - マルチホームはIPv4で、ISPLレベル、および特殊な接続で行われている
  - インターネットに対する冗長性確保のためにIPv6であってもマルチホームを行いたい
- IPv6でマルチホームは可能なのか？

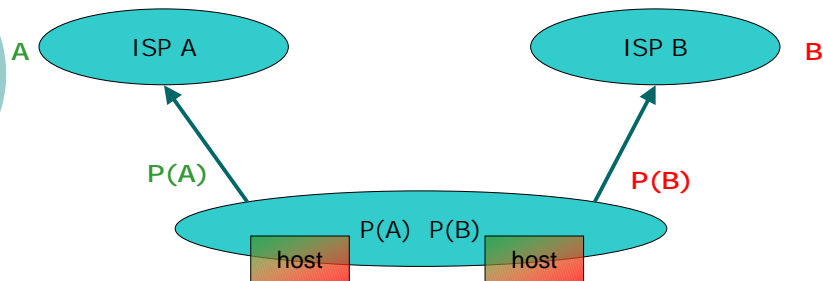
## マルチホームの分類

- BGPを利用する
  - AS及びプロバイダ集約型アドレスの場合
    - PA: Provider Aggregatable address)
    - これはIPv6でも変わらない
- プロバイダ独立型アドレス(PI)を流す
  - PI: Provider Independent address
  - 現在、IPv6でのPIはない
- punching hole
  - IPv6でpunching holeの影響はわからない面がある
- Multi prefix
  - 複数のupstreamからアドレスを割り当てられたケース
  - source address selectionモデルが利用できる
  - しかし、実装依存である
- RFC3178モデル
  - 実現可能だが、コスト・運用面で問題がある

## Multi Prefix型のマルチホーム

- 概要
  - upstreamからそれぞれアドレスが割り当てられる
  - 組織の各ホストにupstreamから割り当てられた複数のアドレスを割り当てる
  - ホストは、source address selectionを利用して、パケットを送出し通信を行う

## Multi Prefix型のマルチホーム



- hostには複数のアドレスを割り当てることが可能
- hostがP(A)、P(B)のどちらを利用するかは実装依存

InternetWeek2001

61

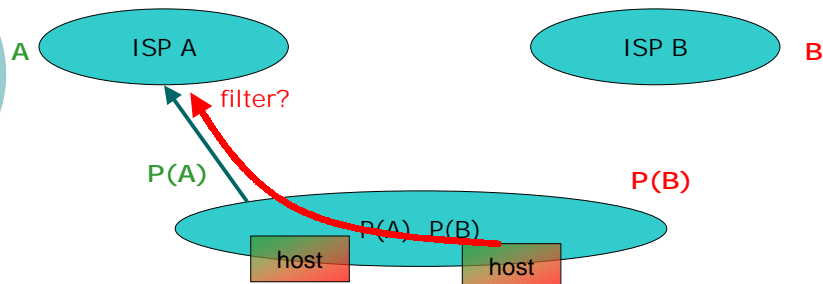
## Multi Prefix型のマルチホームの課題

- 実装は存在しているが、実装依存な部分が多い
  - source addressの選択に関しては各host任せになる
- マルチホームという点で、バックアップの回線の利用、ロードバランスを利用することについては不完全な一面がある

InternetWeek2001

62

## Multi Prefix型のマルチホームの課題



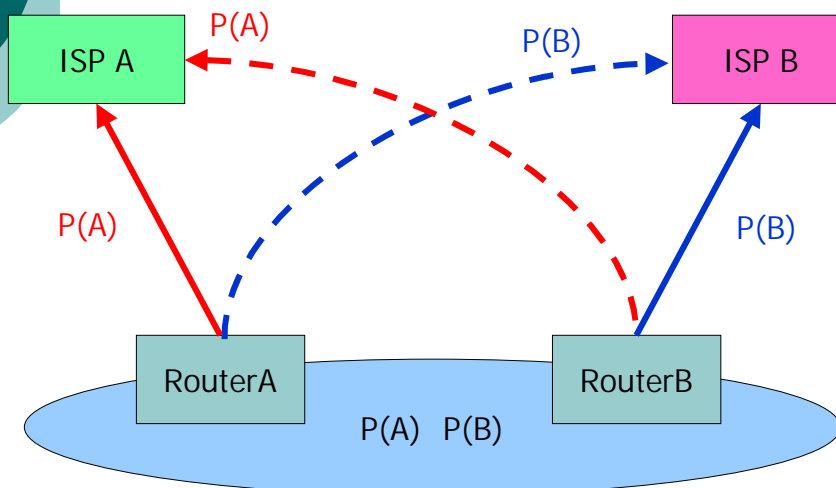
- ポリシールーティングできるルータが一般的にならなくてはならない
  - source address baseのルーティングが必要となる
  - upstreamとのリンクが切れた状態で、そのupstreamから割り当てられたアドレスをsource addressとして選択した場合、別なupstream側でfilterがかけられていると通信ができない可能性がある

## RFC3178モデル

- RFC3178
  - IPv6 Multihoming Support at Site Exit Routers
- 複数の回線をupstreamとの間に用意する
  - 回線
  - IPv6 over IPv4 tunnel
  - ATM PVC



## RFC3178モデル



65

## RFC3178モデルの課題

- トンネルを使用した場合
  - 障害切り分けが難しい
  - 責任分解点を明確にできない
  - トラフィックが他のサイトを通るのでセキュリティ状の問題がある
  - トンネルに対する帯域が制御できない
- トンネルではなく、回線を用意するとすれば、複数の回線コスト増になる

InternetWeek2001

66

## IPv6トラフィックコントロール

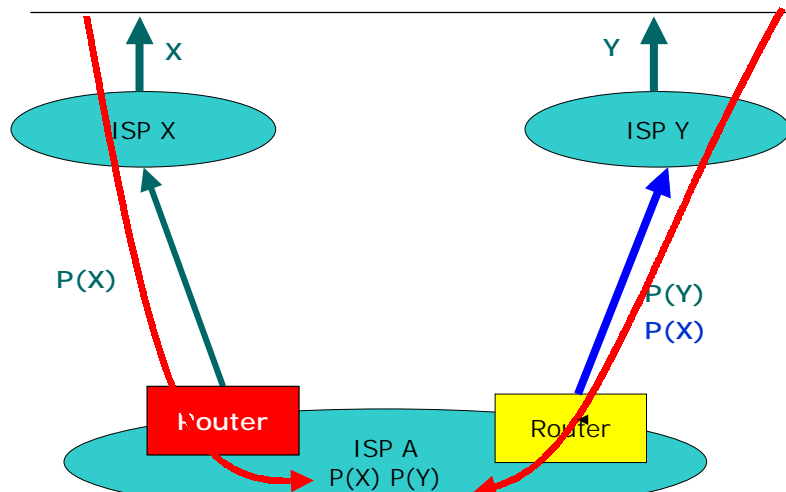
### ○ 概要

- IPv6でもIPv4と同様にトラフィックコントロールを必要とする場面は想定できる
  - ASがアナウンスできるのは、そのASのもつアドレス空間(prefix)のみ
  - ASの出口となるルータの一部にトラフィックが集中する可能性がある
- IPv6でトラフィックコントロールは可能なのか？

InternetWeek2001

67

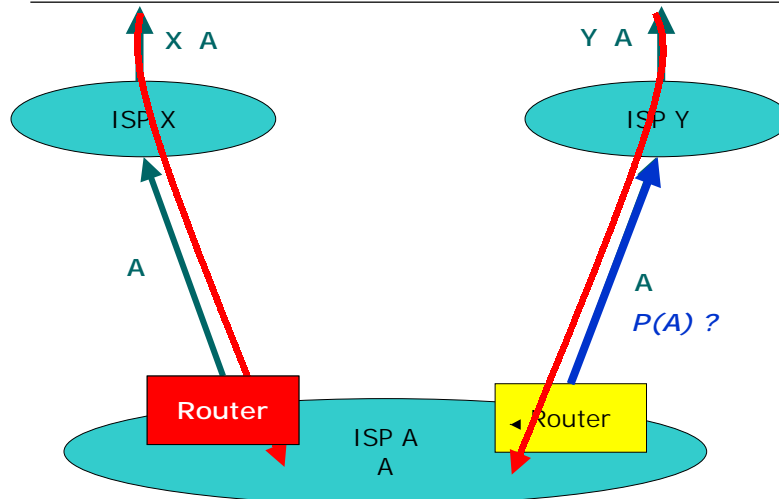
## IPv6トラフィックコントロール upstreamからアドレスを割り当てられた場合



InternetWeek2001

68

## IPv6トラフィックコントロール Internet Registryよりアドレスを割り振られた場合



InternetWeek2001

69

## IPv6トラフィックコントロールの課題

- upstreamからアドレスを割り当てられた場合
  - 割り当てられたアドレスを他のupstreamからアナウンスすることはできないため、コントロールができない
    - 仮に行うとすると経路爆発を招く
- Internet Registryからアドレスを割り振られた場合
  - BGPを利用するのであれば、ある程度までBGPの経路選択アルゴリズムを利用してコントロールは可能
  - しかし、アナウンスできる経路はIRから割り振られた経路に限定されるため、最長一致優先の規則を利用したコントロールはできない

InternetWeek2001

70

## IPv6のIX

---

- 経路交換、トラフィック交換目的であることは同じ
- peeringのアドレス
  - BGPでのpeeringは、グローバルアドレスでもリンクローカルアドレスの両方で可能
  - これは、peeringを行うAS同士のポリシー次第
- IXは、インターネット上で独立したアドレス空間を持つ必要がある

## IPv6のIX

---

- 有名なIX
  - 日本: NSPIXP-6、JPIX
  - 海外: PAIX-6、6TAP、AMS-IX
- IPv6 Exchange Point Web
  - <http://www.v6nap.net/>



## そして、IPv6へ

---



## IPv6時代がくる

---

- IPv6は運用されている
- IPv6対応ルータもある
  - ローエンドからハイエンドまで
  - L3SWも対応しつつある
- 商用サービス・実験サービスなど、IPv6の接続性は手元に届く
- 課題はいくつか残っているが、IPv6へ向けて準備をする必要はある

## IPv6関連の情報

---

- UNIX
  - FreeBSD
  - NetBSD
  - Linux
  - Solaris
- ルーティングソフトウェア
  - zebra
- Cisco IPv6 web
- IETF
  - Site Multihoming in IPv6 WG
  - IPNG WG
  - Next Generation Transition WG

InternetWeek2001

75

## IPv6 ~ ISP的な運用 ルーティング編

---

向井 将  
株式会社パワードコム  
IPv6運用研究会 Co-Chair  
InternetWeek 2001 YOKOHAMA