事例から学ぶIPv6 トラブルシューティング ～ISP編～

Internet Week 2011

NEC BIGLOBE, Ltd.
Seiichi Kawamura
kawamucho at mesh.ad.jp

IPv4脳

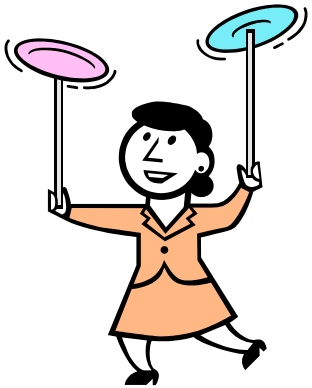
IPv6脳

今までのISPネットワーク設計

- 今まではIPv4のことを考えればよかった
  - IPv4のトラブルシューティング
  - IPv4のルーティング
  - ユーザの通信はすべてIPv4
- 原則1つのIPアドレス
  - Peering
  - DNS
  - コアルータのインタフェース
  - etc…

Welcome to the New World!!!

- デュアルスタックにすると全システムアドレスは少なくとも2個（link local入れると3個）
- ユーザの通信はIPv4とIPv6が混ざってやってくる
  - 例：DNSはIPv4、HTTPはIPv6



IPアドレス設計は？

フィルタポリシーは同じ？

ルーティング設計は？

私のパートの主題

- IPv4とIPv6は同じではありません
- ISPのネットワークにIPv6の通信機能を追加する設計時および、その後運用する際にどういう差分に気を付けないといけないのか
- 重要かつ多くの人に関連する「注意ポイント」をピックアップして説明
- コアネットワークのお話が中心です

自己紹介

- NECビッグローブ株式会社勤務
  - 2001年 NEC入社
    - 2004年まで営業SE（インターネット関係なし）
  - 2004年からBIGLOBEに移籍
    - 2008年まではIPv6/VPN担当
    - 2008年ごろからIPv6/コアネットワーク運用担当
    - 2011年から運用現場を離れIPv6、NW設計、Peering担当
- JANOG（日本ネットワークオペレーターズグループ　運営委員）
- 標準化（IETF）活動、APNICなどに出没します

目次

- IPアドレス設計：20分
- コアネットワーク：10分
- インターネット接続：5分
- サービス提供：5分
  - ISPで最低限必要なサーバ
  - その他

IPアドレス設計：主な登場人物

- Global Unicastアドレス
- Link localアドレス
- Multicastアドレス
- Unspecifiedアドレス
- 場合によっては：Unique Localアドレス

※各アドレスの詳細はRFC4291参照

Globalアドレスの概要

- APNICのポリシーでは、Global IPアドレスの最少割り振りサイズは/32 (JPNICも同様)
  - http://www.apnic.net/policy/ipv6-address-policy#4.3
- 多くのプロバイダにとって、1度の割り振りで十分なサイズ
  - IPv4では複数回割り振りを受ける

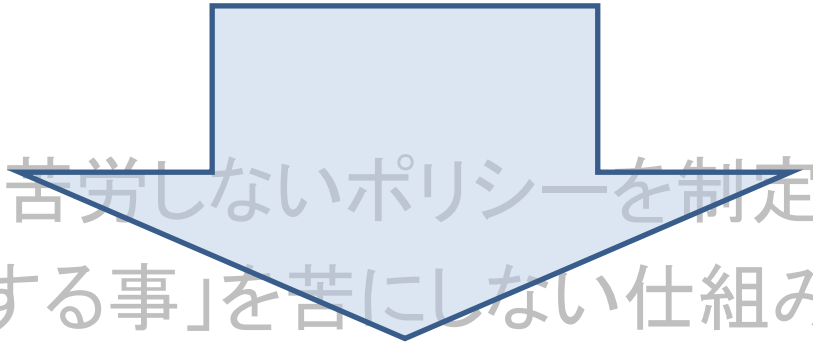
/32の適切な管理は成功する運用のキー

適切な管理のポイントとは？

- ゴール
  - 運用で苦労しないポリシーを制定
  - 「管理する事」を苦にしない仕組みの検討

適切な管理のポイントとは？

- ゴール
  - 運用で苦労しないポリシーを制定
  - 「管理する事」を苦にしない仕組みの検討



IPv6の適切な知識を基に検討する事が必要

自分の事業形態に合った手法を採用

時にはIPv4の常識を一度忘れてみる

失敗例

- ルーティングの集約を目的としてしまう
  - 例：データセンターのフロア単位で集約する

- データセンターを運営している場合例外発生確率が高い
- 集約を目的とすると、運用がそれに依存してしまう
- 依存が高まると例外の扱いが難しくなり、結局運用を苦しめる

3.4 Aggregation

Wherever possible, address space should be distributed in a hierarchical manner, according to the topology of network infrastructure. This is necessary to permit the aggregation of routing information by ISPs, and to limit the expansion of Internet routing tables.

-----http://www.apnic.net/policy/ipv6-address-policy#3.4

内部経路には当てはまりません！！！

IPv6の適切な知識

- 経路集約性とIPv6は無関係
- 「IPv6は階層型なので集約しやすい」と書いてある本を見かけますが無視しましょう
- 昔sTLAなどIPアドレスの割り振りに階層の概念が存在していた時代にそういう事は言われていましたが、実際の運用とあまりにかけ離れているため現在では無くなっています

- 集約しなくてよい、というわけではありません。集約するほどルーティングの効率は高くなりますが、サービス提供や運用を苦しめる設計は回避する必要がある、という意味です

ましな例

- /32を/40単位に分割して、機能毎にわける
  - 2001:db8:1000::/40はルータに充てるアドレス
  - 2001:db8:1100::/40はプライベートピア用
  - 2001:db8:1200::/40は顧客用
  - 2001:db8:1300::/40は裏ネットワーク用
- 覚えやすい（裏なのか、ルータのアドレスなのか、顧客のアドレスなのか）
- ロケーションに依存しないので、アドレスを持ち運びしてもOK
- クラウドネットワークで、L2を広域に伸ばしても気にならない

割り当てポリシー

- IPv4の一般的な概念
  - /24毎の割り当て管理
  - ルータ間は/30
  - 節約のため細かくサブネット分け
- IPv6では？？？
  - /48毎の割り当て管理
  - ルータ間は/64, /112, /126, /127など
  - ネットワークサイズは管理しやすく、運用しやすい単位で設計
    - わかりやすい境界（/48, /56, /64）が一般的

IPv6の正しい知識

- IPv6では/64が「default network size」と言われています
  - 実際ほとんどの実装はこれを前提にしています
- しかし/64以上のPrefixが使えないわけではない
- /64より長いPrefixにすると、SLAAC(RFC4862)が利用できなくなる事を意識すればよい

色々なセグメントサイズ

- Point to point link(SONETやトンネルなど)は/127が利用できるルータなら/127がおすすめ
  - それ以外を選択する場合は、Neighbor Discoveryの扱いについてベンダーに確認してみよう！
- Broadcastが飛ぶようなメディアでも、IPアドレスが少ない方がセキュリティ的に守りやすいため/126でも問題はない
- ルータ間リンクは、管理性を重視して/64を適用する場合もある
  - ただしRFC4443のICMPv6実装である事が重要！メーカーに確認してみよう！
- サーバセグメントにはシンプルに/64がおすすめ
  - SLAACは基本的には使わないと思うが、IPv4みたいにケチる必要は無い部分

ルータ間アドレス設計のトラブル

Point to point link の場合

2001:db8::100:1/64　　2001:db8::100:2/64

2001:db8::100:3宛

お！2001:db8::/64だから
隣のやつかな

お！隣のやつかな

Loop!

Broadcast Mediaの場合

2001:db8::100:3宛

ICMP unreachable

2001:db8::100:4宛

ICMP unreachable

ルータ負荷
上昇!

Link localアドレス

- IPv4に無い概念
  - 一応定義はされてはいるが・・・
- どのノードにも必ず付与されるアドレス
  - fe80::/10

はたしてこれは管理すべきアドレスなのか？？？

Link localアドレスの登場シーン

- BGP/OSPF/static routeのnext hop
- OSPFのネイバーアドレス
- SLAACが有効なゾーンでのデフォルトルート

かなり重要なところで登場する
1）管理せずMACアドレス生成のEUI-64を使う
2）管理するためにStaticで定義する
どっちにしますか？

こたえ

- 自らの事業形態、運用内で重要視したいポイント、置諸元を考慮して選択する事が望ましい
- Link localを固定で割り当てられない装置も存在する
  - 例外が出てくる
- 固定で割り当てておくとトラブルシュートが格段に楽
- 固定での割り当ては「管理」目的よりも「運用性向上」が目的なのでポリシーを決めておけば管理表などは不要
  - Globalの下64bitと同じに設定する、など
- EUI-64を使う場合、デフォルトルートがどこを向いているのか、OSPFのこのネイバーは誰なのか、BGPのnext hopがどこを向いているのか、判別の仕方を考えれば良いだけ

ポイント

- Link localは設計段階で「意識」しておかないと後でポリシーを変更する事は難しい

- 運用で必ず登場する。「見えないアドレス」では無い

その他アドレス

- MulticastとUnspecified
  - 注意点一つだけ：フィルタしてはいけません
  - 普段はほとんど気にすることはない
- Unique Local Address
  - 扱いが難しいアドレス
  - IPv4のPrivate(RFC1918)スペースとは若干意味合いが異なる
  - Privateのように扱っても良いが、現状お勧めの利用方法は定義されておらず、標準として曖昧さがある
    - IPv6ではNATは「あまり」推奨されていない
    - 逆引きの扱いが曖昧
  - 現状、ISPネットワークでは評価環境内でしか使わない方がよいかも

コアネットワークのトピック

- ルーティング
  - ルーターとプロトコルの注意ポイント
- 運用
  - データ取得の注意ポイント
  - ツールの注意ポイント

IPv4/IPv6の実装差分

- IPv4とIPv6で機能に差分がある場合があります
  - IPv6ではOSPF stub router announcementが実装されていない場合がある(draft-retana-ospf-rfc3137bis-01)
  - コンフィグの階層が異なる場合がある
    - レガシーIOSやquagga
    - OSPFをInterface階層に設定するかrouter階層に設定するか
  - 保持できる経路数が異なります
    - IPv4では100万経路持てるのにIPv6は1万程度、など

どのメーカーもfeature parityは必ずあります
主要な機能は差分を明確にしてもらう方が良いです

プロトコル差分：BGP

- BGP4+（BGPのマルチプロトコル対応拡張)で経路交換
  - 特に珍しい実装ではないが、VPN接続用途で実装されているBGPでは使えない場合が稀にある
- ほぼBGP4と同等の動き
  - RFC4271、RFC4760、RFC2545

プロトコルの差分：BGP

- IPv4のセッションの上でIPv6経路情報を交換する事もできますが、おすすめしません
- Peer設定のSource addressをlink-localにする事もできますがおすすめしません
  - 一般的にGlobalを想定しているため、バグが稀にある
  - Next-hopはGlobalアドレスでないと中に伝達しても意味が無いため、基本すべての設定はGlobalで実施する事が望ましい

プロトコルの差分：OSPFv3

- 基本的な動作はOSPFv2と同じ
- リンク(インタフェース)単位で実行される
- LSAが異なる
- Link-local中心の設計に慣れる

OSPFv3~linkの考え方~

■ “link”単位での動作。”network”という概念がありません。

Ciscoに慣れてる人は
router ospf 10
　network 10.10.10.0 0.0.0.255 area 0

interface [interface]
　ipv6 ospf *process-id* area *area-id* [instance *instance-id*]

※juniperとかは元々interface指定なので違和感はない

OSPFv3~neighborの考え方~

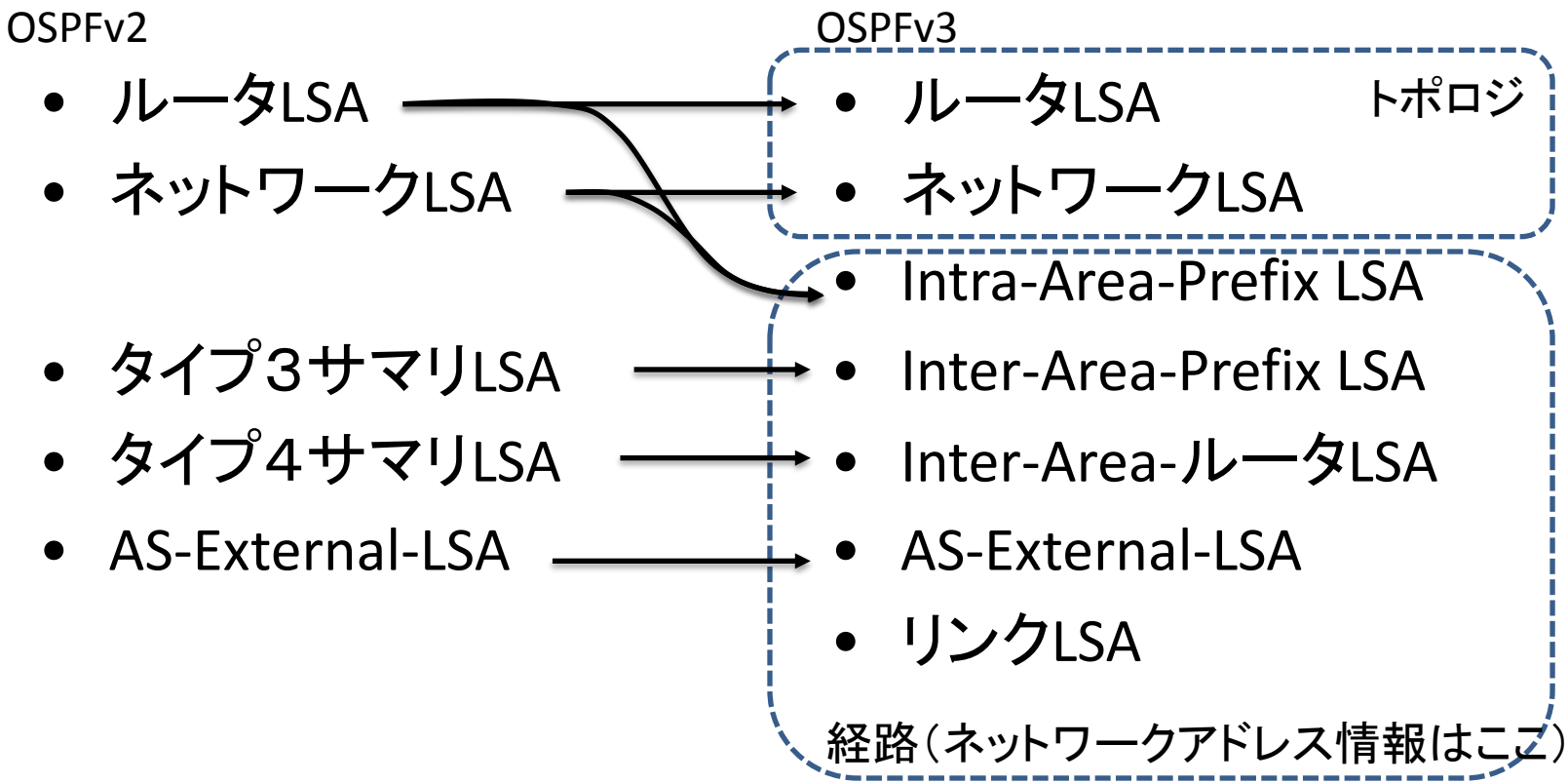
- router IDはIPv4アドレスサイズと同じ32bitのまま

  →loopbackアドレスを設定するポリシーだった人は、
  ポリシー差分がでます

- show neighbor系で見えるのは、routerID単位で見える。
  next-hopアドレスがキーとして見えるのではない
  router-IDの設計は重要

- 認証はMD5でなくIPsecで

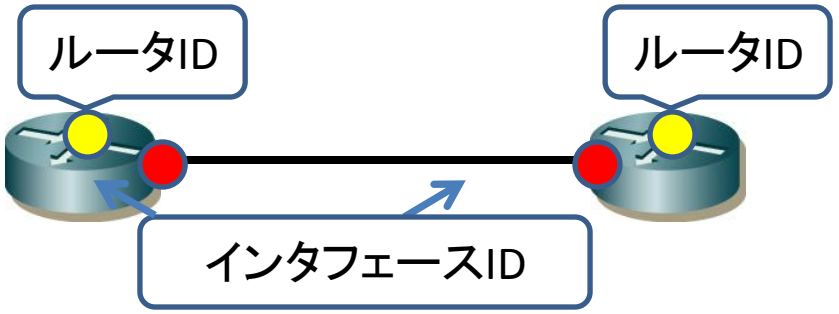
＜参考＞OSPFv3のLSA



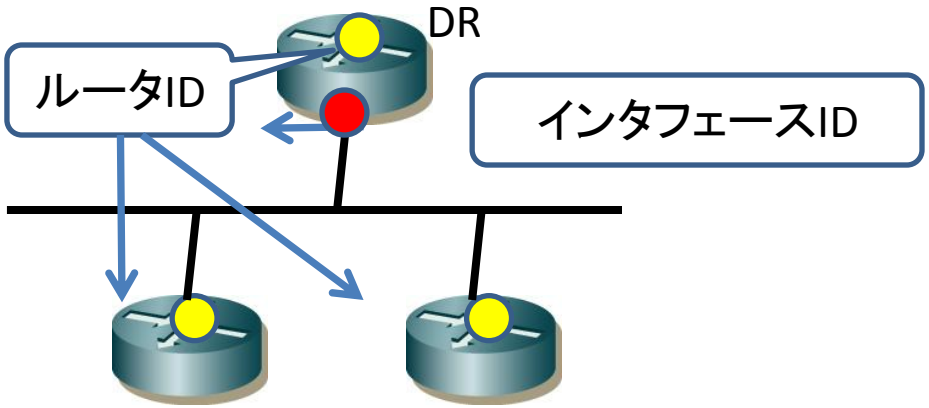
IPv4ではルータLSA、ネットワークLSA内にネットワークプレフィックス情報が含まれていた。

＜参考＞OSPFv3のルータLSAとネットワークLSA

- IPv6非依存でトポロジ情報だけを運ぶ
- ルータLSA
  - ルータの接続情報

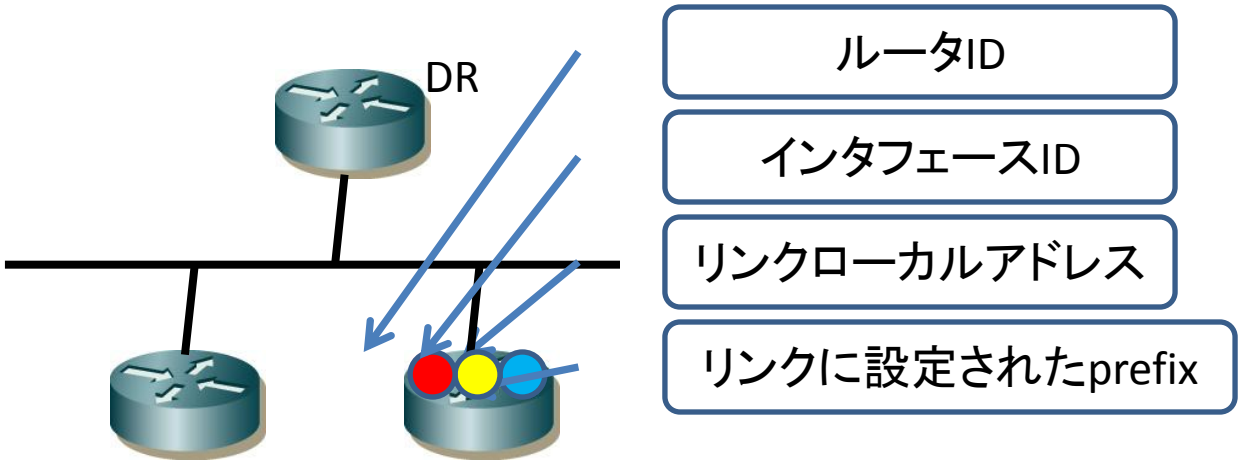


- ネットワークLSA
  - リンクに接続しているルータのリスト



＜参考＞リンクLSA

- リンク内のみで交換されるLSA
  - リンクローカルアドレスの通知
  - リンク上で有効なprefixの通知



＜参考＞Intra-Area-Prefix LSA

- OSPFv2の時にルータLSAやネットワークLSAが運んでいた経路情報を運ぶ
  - Stubネットワークやtransitネットワークの経路
  - loopbackの経路もこのLSAで運ばれる
- リンクLSAをDRが収集して、経路情報を代表してIntra-Area-Prefix LSAとして広報する
  - リンク上の一部ルータにだけ設定されているprefixでもリンクのDRが代表して広報する

OSPFv3~link localに慣れる~

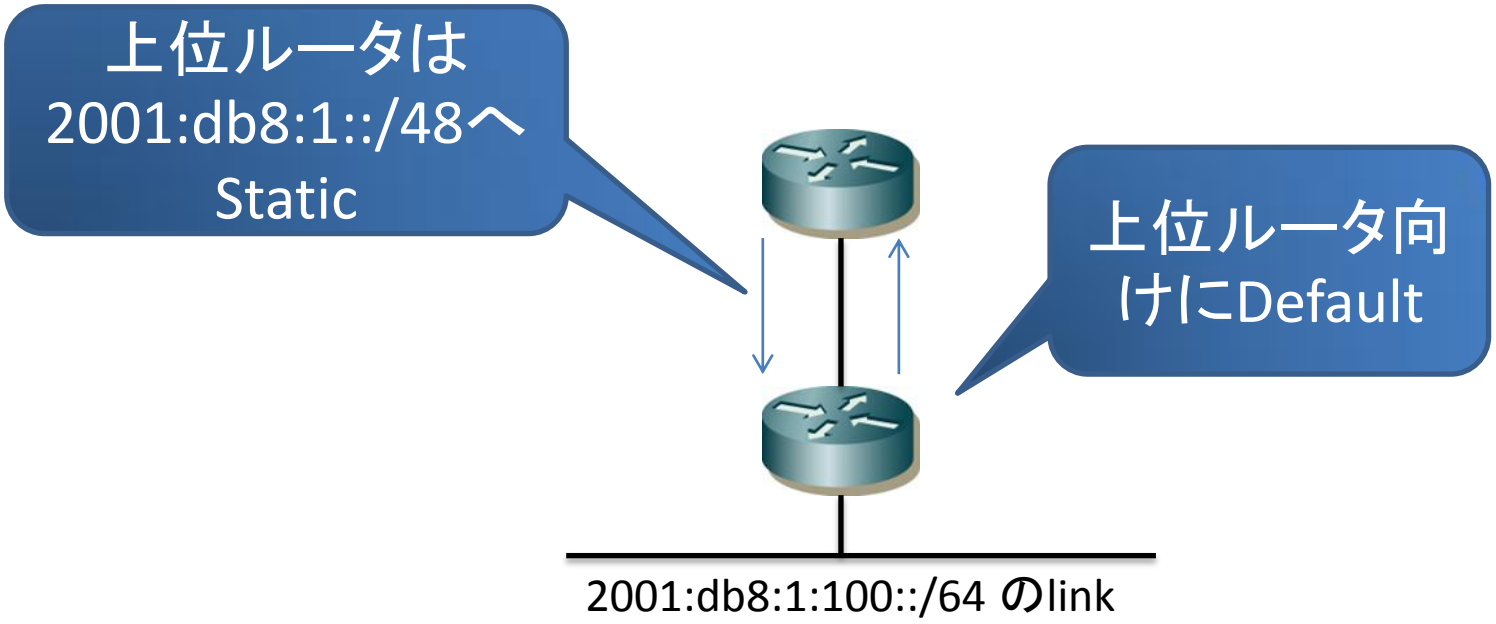
■ OSPFのパケットはすべてlink-localがsource-ipとなる

→ next hopはlink local!

（例）tracerouteで出てくるアドレス＝グローバル
show routeのnext-hop＝リンクローカル
show ospf neighbor ＝router-ID
（追加で実装によりneighborのlink localだったり IF-IDだったり）

慣れるまで違和感はあるが
実際の運用はそんなに難しい事はない

Static/defaultのヒヤリ



Loopします！！！

- 2001:db8:1:100::/64以外の65k個のPrefix宛がloop
- 下のルータは、2001:db8:1::/48をnullに向けましょう

運用差分

- SNMP/xflowのデータ取得
  - Netflowv9対応が必要
    - ルータによってはハードウェアアップグレードが必要
  - IF-indexを使ったデータ取得ではIPv4とIPv6が混ざったデータしか取れない
- ツール
  - ExPINGなどIPv6に対応していない運用ツールは多数
    - 対応ツールは3年前と比較すると格段に増えた
  - 今まで自作したPerlツールなどは使えなくなるケースが多発
    - コードを綺麗に作り直すに良いチャンス？

インターネット接続のトピック

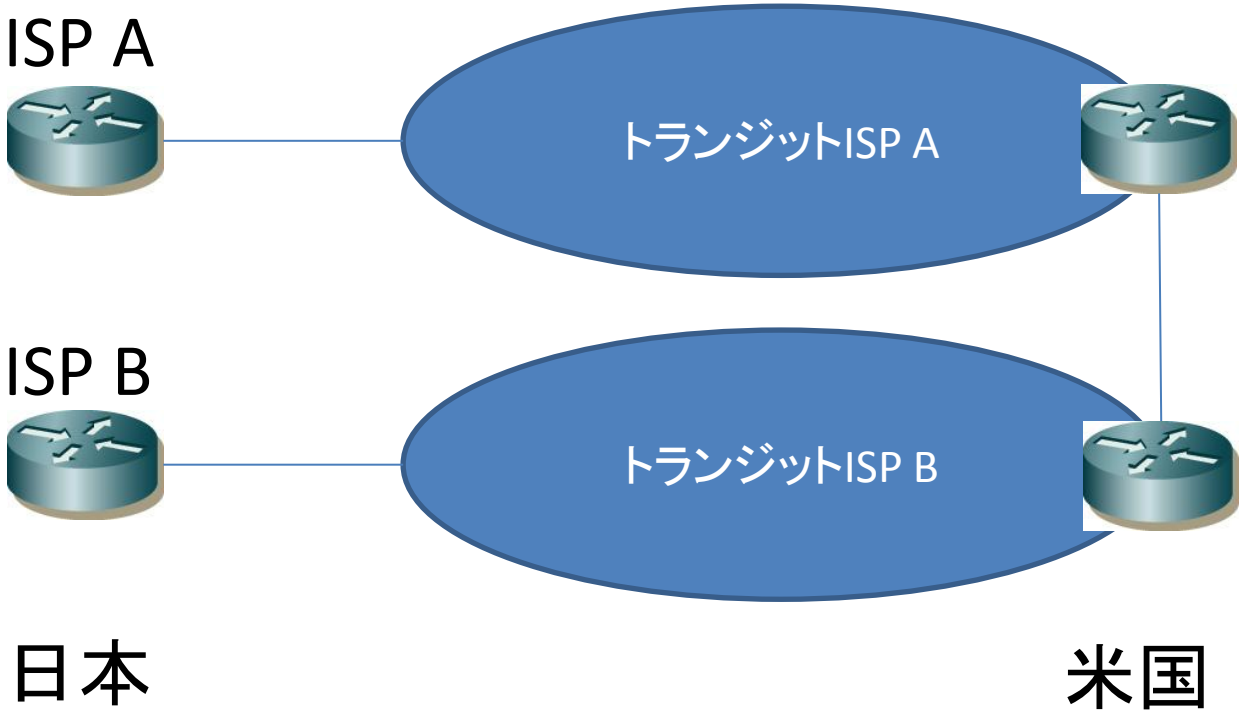
- トランジットとフルルートについて
- BGPフィルタポリシーの注意ポイント

トランジットとフルルート

- IPv4と異なり、まだIPv6のPeeringは安定していない
  - シングルホームが多かったり
  - フィルタリングポリシーがバラバラだったり
- Route Viewsを見ると、経路数差分は大きい
  - 日本でTransitを販売している事業者はそんなにひどい差分は無い
  - 何経路もっているか、フィルタリングポリシーは何なのかは、確認が必要

経路トラブル

- Peeringが少ない状況での注意点
  - 遠回りしてしまう通信



BGPフィルタポリシー

- 受信経路のフィルタポリシーと広告ポリシーの「慣例」がまだ浸透していない
  - /48でフィルタする人が多いものの、/64で広告しているケースがみられる
  - 厳密な割り振りサイズでフィルタしている人も散見するが、経路を分割して広告しているASへ到達性が無い
- 現状は、受信経路は/48～/64の間でフィルタし、経路広告する場合は分割したとしても、割り振りサイズのものを広告する方が望ましい

ISPで最低限必要なサーバ

- DNS（キャッシュサーバ）
  - AAAAの応答は黙ってても行われている
  - IPv6のDNSサーバをユーザに提供する必要があればインタフェースにIPv6アドレスを付ける
- メール
  - MXへAAAAを付けると、他ISPのメールサーバとIPv6で通信できるようになる（今のところ大きな問題にはならない）
  - ただし、ユーザ向けのPOP/SMTPは慎重になる必要がある
    - AAAAレコードの応答で挙動がおかしくなるクライアントが存在する（企業が独自で開発したものなど）

ユーザサポートとデュアルスタック

- 顧客にIPv4/IPv6 デュアルスタック環境を提供する場合、IPv4でうまく接続できていないのか、IPv6でうまく接続できていないのか確認が必要
  - 確認用のツールがあった方が良い
- WindowsXP、MacOSX Lion以前、多くのスマートフォンはDHCPv6に対応していないため、DNSはIPv4でクエリーし、通信はIPv6で行う。
  - IPv4でのDNSトラブルがIPv6通信に影響を与える

まとめ

- アドレス設計、ルーティングともにIPv4とは差分があり、独立したポリシーが必要な場合もある
- 経路運用には細心の注意を払う必要があるが、IPv4と比べてすごく大変になるわけでもない
- ユーザ向けに提供するサーバやサポートは慎重な対応が必要

ありがとうございました