社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＤＳＭ－46／2007－ＱＡＩ－24（６）
2007／7／19

[招待講演] インターネットプロバイダの運用管理技術

吉田 友哉

NTT コミュニケーションズ株式会社 〒100-8019 千代田区内幸町 1-1-6
E-mail: yoshida@ocn.ad.jp

あらまし インターネットプロバイダの運用・設計に携わっている立場から、インターネットプロバイダの運用管理技術について紹介する。
キーワード Internet, Operation, Maintenance, Provider

[Invited Talk] Operation and maintenance technology of Internet Provider

Tomoya YOSHIDA

NTT Communications Corporation 1-1-6 Uchisaiwai-Cho, Chiyoda-ku, Tokyo, 100-8019 Japan
E-mail: yoshida@ocn.ad.jp

Abstract It's introduced about operation and maintenance technology of Internet Provider from the viewpoint where I concern Internet Service Provider's practical use and design.
Keyword Internet, Operation, Maintenance, Provider

インターネットプロバイダの運用管理技術

2007年7月19日
NTTコミュニケーションズ株式会社
吉田 友哉

1

内容

- はじめに
  - 自己紹介
  - OCNバックボーンの概要
- 基本的なネットワークの運用管理
- ルーティングネットワークの運用管理
- その他運用技術
- 今後

2

自己紹介（略歴）

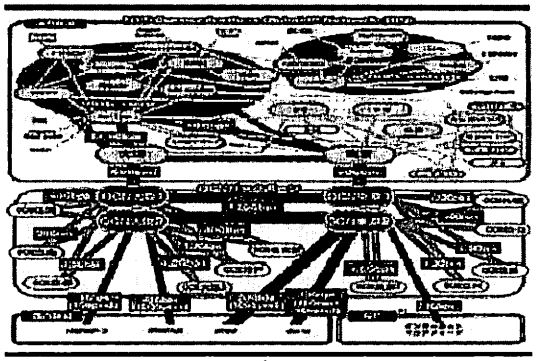
1998年　日本電信電話株式会社入社
OCNバックボーンや対外接続部の運用構築を担当

1999年　NTTコミュニケーションズ㈱へ
現在に至るまで、同ネットワークの運用・構築・設計・開発に従事

2000年～　Network+Interop NOCメンバー（現 Interop）
2002年～　JPNIC IPアドレス検討委員
2004年～　JPNIC IRR企画策定専門改委員Chair
2004年～　Network Service Provider Security JP (NSP-SEC-JP) Moderator
2005年～　番号資源利用状況調査専門家チーム委員
2006年～　JPNIC IPアドレス検討委員長
2007年～　APNIC Routing-SIG Chair
APRICOT Program Committee

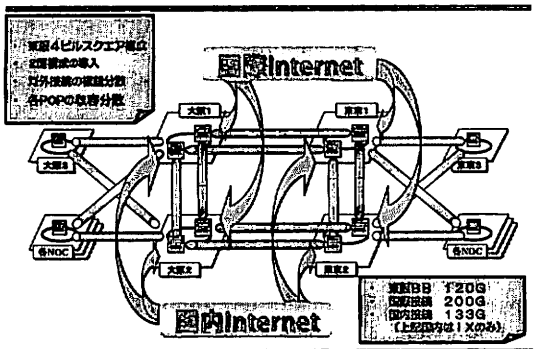
3

OCNバックボーンネットワーク



4

OCNバックボーン構成



5

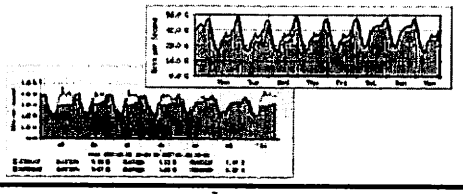
ネットワークの状態把握

- まずは、流れるパケットをモニタリング
  - きちんと帯域が確保できているか
  - 適切な通り道を通って通信できているか
  - トラフィックの分散は適切に行われているか
  - 品質の劣化はないか
- 装置のリソースに問題はないか
  - CPU、メモリ、他
- 何か起きた際には
  - 故障は交換などで迅速に対応
  - セキュリティ問題も早期対応
  - DDoS攻撃等の際には、他ISPと協調しながら対処

6

トラフィック監視

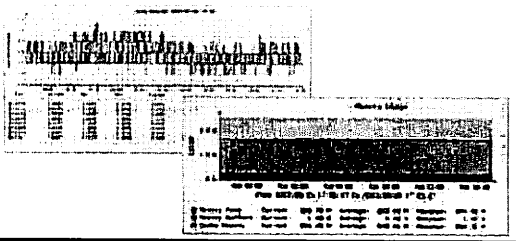
- まずは流量を把握する
  - MRTG, RRDTool, NMS, etc
- 冗長化回線のバランスを確認
- トラフィックの変動を閾値監視しアラーム通知
  - 直前のトラフィックとの比較
  - 通常時のトラフィックとの比較



7

リソース監視

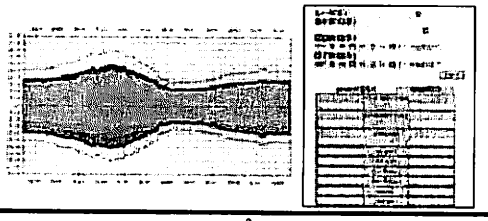
- メモリやCPU等のリソースが枯渇していないかを監視
- トラフィック状況と連動した解析も実施
  - 特定スロットのCPU高負荷
    →該当スロットをまたぐトラフィックが何らか影響か



8

フロー情報の収集

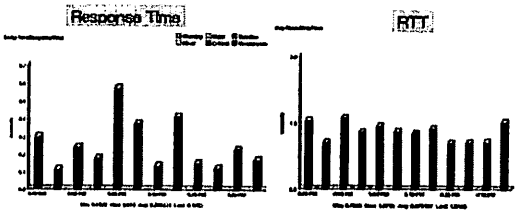
- ルータにてExport可能なフロー情報によりネットワーク全体の状況を把握する
  - NetFlow、sFlow、etc
- 可能な限りルータからフロー情報を飛ばし蓄積しておくと、トラブルシュートの際にも過去のフローを追える



9

品質の管理

- ネットワークやサーバの品質を常時モニタ
  - アプリケーションレベルの監視
- 閾値を超えた場合にはアラーム通知



10

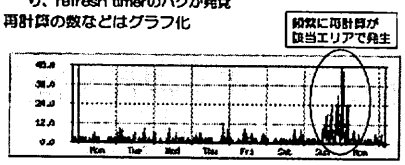
OCNバックボーンの経路制御

- トポロジー情報の管理：OSPF
  - エリア構成
  - 基本的にはECMPの冗長化運用
  - 一部経路のロードバランス適応時にOSPFを利用
  - DR/BDRの分散化
  - V4/V6ともに同様のトポロジー
- 経路情報の管理：BGP
  - iBGP/eBGP
  - RRの階層構造
  - 基本はフルルートを配送、配れる範囲で適切に配る
  - V4/V6ともに同様のトポロジー

11

OSPF運用管理－1

- LSAをモニタリングする
  - エリア毎に可能な範囲でLSAの状態を把握
    - LSDB情報をダンプ
    - LSAのType毎に状況を把握
  - 変なLSAの挙動からネットワークの不安定要因を発見することもある
    - ベンダの実装の違いで、refresh intervalが異なっていたことにより、refresh timerのバグが発覚
  - 再計算の数などはグラフ化

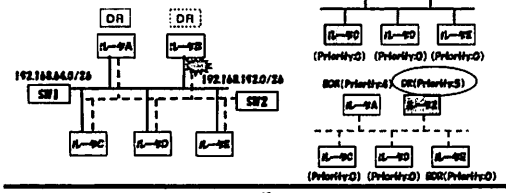


12

OSPF運用管理－2

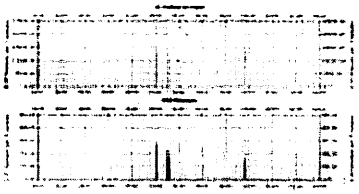
- DR/BDRの冗長化
  - SWセグメントではDR/BDRの冗長に注意しながら運用
  - IF故障などで冗長した場合には速やかに対処
  - 必要ない所はP2P-IF化

この状態でルータBが故障すると、SW1/SW2間ネットワークの再生成がルータAによって同時に行われ、通信断→SW1向けネットワークを正常後にDRをルータBからルータAに変更する



BGP運用管理－1

- 経路情報の変動を記録する
  - 過去のアップデート情報をさかのぼる
  - 異なるBGPクラスタで複数取得しておくと良い



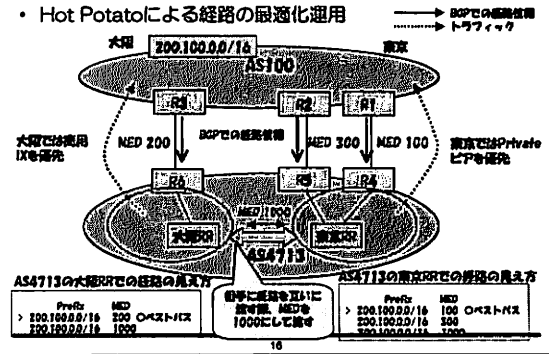
BGP運用管理－2

- 経路制御ポリシーに従い、LOCAL_PREF(LP)/MED値を付与管理しiBGP網を運用
  - 基本はLPで制御
  - 同一接続形態の回線が複数ある場合には、MEDが小さい経路か最もIGPコストが小さい経路を選択する（下記は例）

接続形態	LOCAL_PREF	MED1	MED2	MED3	優先順位
BGP顧客経路	500				1
自AS向広報経路	400				2
プライベートピア経路	300	100	110	120	3
東京IXピア経路	300	200	210	220	4
大阪IXピア経路	300	300	310	320	5
上流フルルート1	200				6
上流フルルート2	200				6

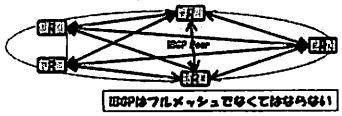
BGP運用管理－3

- Hot Potatoによる経路の最適化運用

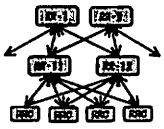


BGP運用管理－4

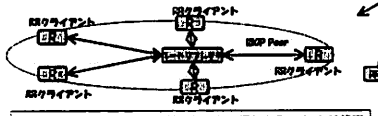
◎一般的なiBGP Peer



ルートリフレクタの階層化により大規模にも対応可

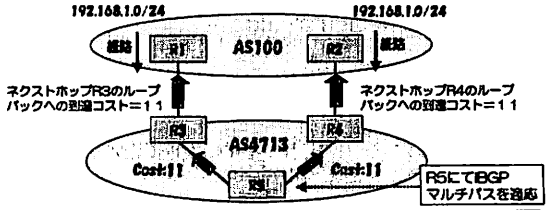


◎ルートリフレクタ(RR)を使用したiBGP Peer



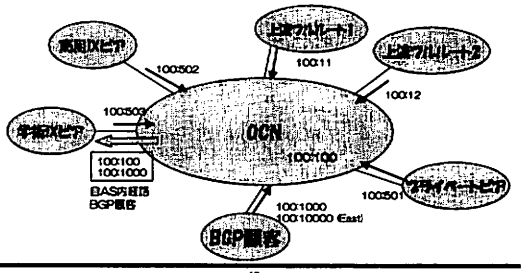
BGP運用管理－5

- BGPマルチパスによるロードバランス運用
  - 1つの宛先ネットワークに対して複数のBGP経路を生成
  - ISP間のOUTトラフィック分散等に利用
  - BGP経路比較選択プロセスで、IGPコスト比較まで同一の場合



BGP運用管理－6

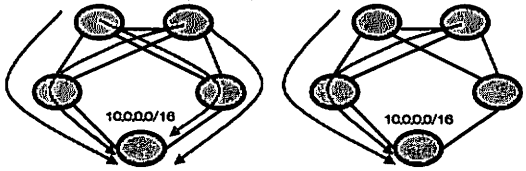
- BGPコミュニティで経路を色分けして運用管理
  - 接続形態、場所などに応じて複数に色分け
  - 色をつければポリシーに従い自動的に経路広告



19

トラフィックの分散技術

- Per Flow
  - 異なるFlow(SrcIP/DestIP/SrcPort/DestPort等が一致しているパケット）毎に分散
  - 自ルータから見た宛先が少ない場合や、特定のプリフィックスにトラフィックが偏る場合などに有効
- Per Destination
  - 異なる宛先毎に分散
  - 宛先が多数存在する場合に有効



この2つをうまく組み合わせてトラフィックコントロール

20

トラフィック増大への対応

- 10G等の回線を複数並べるのはつらい
  - 回線毎に論理的にネットワークに組み込む必要あり
  - トラフィックバランスが事実上できない

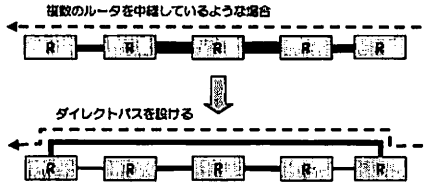
⇩

- リンクアグリゲーション技術の適応
  - 複数の物理回線を1つの論理回線にマッピング
  - Etherは802.3ad（LACPは必要に応じて適応）
  - POSは独自仕様
  - 最近はmixed aggregationにも対応
    - 回線帯域に応じたフローの割り当て

21

トラフィック交流の最適化

- BGPコミュニティとOSPF tag情報を元にトラフィックマトリックス交流を把握する
  - どの対地からどの対地へどの程度のトラフィックがあるか
  - フローの送信元/宛先情報と通過したルータ/IF情報よりわかる
- デザインの最適化へフィードバック



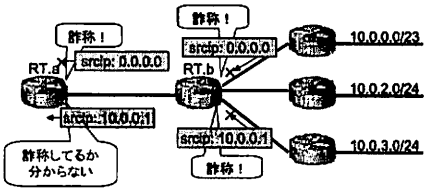
22

送信元IPアドレスの検証

- ACL
  - パケットフィルタ
  - Interface毎に個別のルールを適用する
  - ネットワークの変更時にはルールの変更が必要
- uRPF check
  - 経路情報を基にパケットの送信元IPアドレスを検証
  - 経路情報を適切に保てば、ルールは自動的に追従するところがポイント
- 上記2方式を組み合わせてネットワークに適応

23

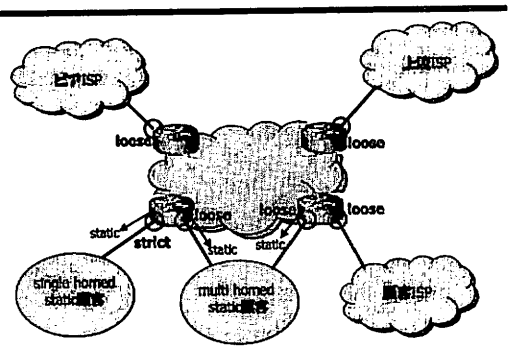
検証は送信元付近で実施



- 利用されないアドレスはどこでも判別できるが、利用されているアドレスの判別は送信元の近くでないと検証できない
- 無駄なトラヒックを網内に流さないためには、ネットワークの入り口で検証

24

プロバイダでのuRPF適応事例



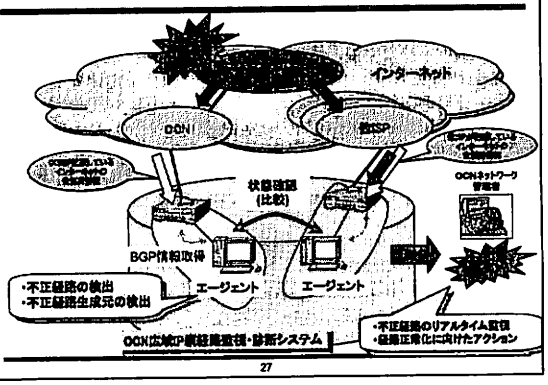
25

BGPルートハイジャック問題への対処

- 不正なBGP経路広告にるネットワークの乗っ取り
- 実際にしばし観測されている
  - 単なる数字の書き間違いがほとんどの模様
  - 意図的なハイジャックも時にはある
  - more specific な経路には勝てない
  - DNS Anycast ではより問題が深刻に
- Origin AS の検証が必要
  - 通常 Origin AS が正しいか否かに関わらず受信
  - インターネットフルルート受信時は非現実的
- 暫定対応方法
  - 被害者が more more specific(cure)経路を広告

26

OCN広域IP網経路監視・診断システム



27

IPv4/IPv6 デュアルスタック運用

- バックボーンやIPv6サービスクラスタは IPv4/IPv6デュアルスタックネットワーク運用
- 運用管理コストは実は倍近くに
  - 経路制御時や故障発生時には両プロトコルの操作確認が必要
- IPv6プロトコルの成熟度はまだ低い
  - 設計思想が違うとはいうものの
  - IPv6固有の問題は依然多い
  - これから対応するプロトコルも多数
- 管理パケットは実はIPv6
  - AAAAレコードが存在すればIPv6

28

プロバイダ同士の協調運用

- 自AS内のみでは解決出来ない問題
  - 自分は適切な経路を選択している
  - 一寸先は闇
  - 実空間を利用した偽装パケット
- 互いのboundaryを超えた協調運用の重要性
  - コミュニティの活用、連携
    - JANOG, NANOG, xNOG
    - NSP-Security-JP (NSP-SEC-JP)
    - Telecom-ISAC BGPWG
- インターネットのマナーは皆で守りましょう
  - RFC 3013
  - Recommended Internet Service Provider Security Services and Procedures

29

今後

- BGPコンバージェンス時間の更なる削減
- IX(eBGP)上での高速切断検知技術の追及
  - BFD for BGP
  - EtherOAM
- 4-octet ASNのDeployment
  - AS23456(AS_Trans)は単なる移行措置だろう
  - 3年以内ぐらいに全て4-octet ASN対応になる？
- Link-aggregation → 100G?
- IPv6 NW へのマイグレーション

30