WIDE IPv6 ネットワークの運用とその問題点

角川 宗近
(株) 日立製作所
sumikawa@ebina.hitachi.co.jp

山本 和彦
奈良先端科学技術大学院大学
kazu@mew.org

加藤 朗
東京大学
kato@wide.ad.jp

新 善文
(株) 日立製作所
atarashi@ebina.hitachi.co.jp

若井 宏美
慶応義塾大学
mie@sfc.wide.ad.jp

IPv4 アドレスの枯渇問題に伴い、IPv6 の開発と実装が急務となっており、また IPv6 への緩やかな移行技術の確立が望まれている。この問題に取り組むために、WIDE Project は IPv6 分科会を編成し、IPv6 実験ネットワーク (WIDE 6bone) の運用をはじめた。本稿では、活動報告として WIDE 6bone の歴史と特徴についてまとめると共に、その運用を通じて得た、静的経路制御の破綻、RIPng を用いた経路制御の限界、マルチホーム運用の難しさなどの問題点を述べる。また、point-to-point ネットワークのアドレス割り当て、リナンバリング、アドレス・スコープ、多重アプリケーションの運用などの問題点についてはその解決法も併せて述べる。

The operation of WIDE IPv6 network and its problems

Munechika Sumikawa
Hitachi, Ltd.
sumikawa@ebina.hitachi.co.jp

Kazuhiko Yamamoto
Nara Institue of Science and Technology
kazu@mew.org

Akira Kato
The Univ. of Tokyo
kato@wide.ad.jp

Yoshifumi Atarashi
Hitachi, Ltd.
atarashi@ebina.hitachi.co.jp

Hiromi Wakai
Keio University
mie@sfc.wide.ad.jp

Since the IPv4 address space is being exhausted for the rapid growth of the Internet, it is of recent interest to develop IPv6 and to migrate to IPv6 environment without great confusion. To deal with this problem, WIDE Project formed IPv6 working group and started operating our IPv6 network, called WIDE 6bone. As an activity report of IPv6 working group, WIDE Project, this paper describes the history and characteristic of WIDE 6bone. Moreover, we discuss our experienced problems including assignment network address to point-to-point link, address scope, renumbering, multi-homed organizations, management of multiple application, etc.

1 導入

IPv4 アドレスの枯渇問題に伴い、IPv6 の開発と実装が急務となっており、また IPv6 への緩やかな移行技術の確立が望まれている。この問題に取り組むために、WIDE Project は IPv6 分科会を 1995 年夏に編成した。現在 約 20 組織からの参加者があり、約 10 の実装を独立に開発し改良を続けている。同分科会は IPv6 実験ネットワークの運用を通じ、IPv6 や移行技術の問題点を発見し、解決方法の考案に努めている。

本稿では、WIDE Project IPv6 分科会のこれまでの活動を報告する。2 章では WIDE 6bone の歴史を述べ、3 章ではその特徴について触れる。また、WIDE 6bone の運用経験から得た問題点とその解決方法の一案について、4 章で解説する。最後に 5 章において、今後の課題を示す。

2 WIDE 6bone の歴史

WIDE 6bone は、1996 年 6 月 9 日に東京 NOC と奈良 NOC を専用線で結ぶことからはじまった。当初は各ノードには IPv4 互換アドレスを用い、静的な経路制御で運用した。

同年 7 月 16 日、米国 Cisco 社との接続を皮切りに 6bone に参加した。これに先だって WIDE 6bone は Internet Enginerring Task Force(IETF) で定められた実験ネットワークのためのアドレス割り当て [1] に基づいてアドレスを付け換えた (以下リナンバリングと呼ぶ)。

WIDE 6bone に接続するノードの数が増加するにつれて、設定の手間が問題になった。経路が 1 つ加わるたびに全てのルータに対して設定を追加するのは繁雑であり、また設定誤りを招きやすい。1997 年 2 月には、RIPng[2] を利用する経路制御デーモンを開発し、動的な経路制御に移行した。しかし、その半年後には WIDE 6bone 内で最も離れたノード間で 10 ホップを越え、RIPng の運用限界が近付いているのが報告された。この時期には、6bone 全体の経路数は 300 近くに増加していた。

1997 年 10 月には新しく提唱された「経路集約型アドレス」[3] に基づいて、6bone のリナンバリングが行われた。WIDE 6bone は緩やかなリナンバリングの経験を積むために、9 月下旬から約 1 ヵ月かけて移行した。

WIDE 6bone は順調な発展をとげ、1998 年 1 月 1 日現在では、参加組織数 19、NOC 数 8 と、世界でも最大規模の IPv6 ネットワークとなっている (図 1)。6bone には 3 つの海外組織へのトンネリングを通して接続されている。現在は RIPng で経路を交換しているが、BGP4+[4, 5] に移行しはじめている。アドレスを集約したため RIPng の混乱も収束し、一時期は 300 あった経路数も半分に減少している。本来ならば 6bone のコアサイト数と、WIDE 6bone 内部の経路数の和である 60 程度に経路数は収まるはずだが、一部で古い経路や適切に集約されていない経路がアナウンスされている。

3 WIDE 6bone の特徴

WIDE 6bone はその規模だけでなく、さまざまな特徴をもっている。この章では、データリンク、実装、アプリケーション、運用技術の 4 つに分けて WIDE 6bone の特徴を述べる。

3.1 データリンク

6bone は現在でもその大部分がトンネリングで構築されているのに対し、WIDE 6bone は図 1 で示したようにトンネリング以外にも多様なリンクで構成されている。トンネリングとは、IPv6 データグラムを IPv4 でカプセル化して既存の IPv4 ネットワークの上に仮想的に IPv6 ネットワークを構築することである。IPv6 のために専用線を用意したのは WIDE Project が世界初であり、いまでも希少である。

以下では特徴的なデータリンクについて述べる。

- シリアルリンク

  専用線ネットワークを構築するためのデータリンクである。WIDE 6bone の開始時

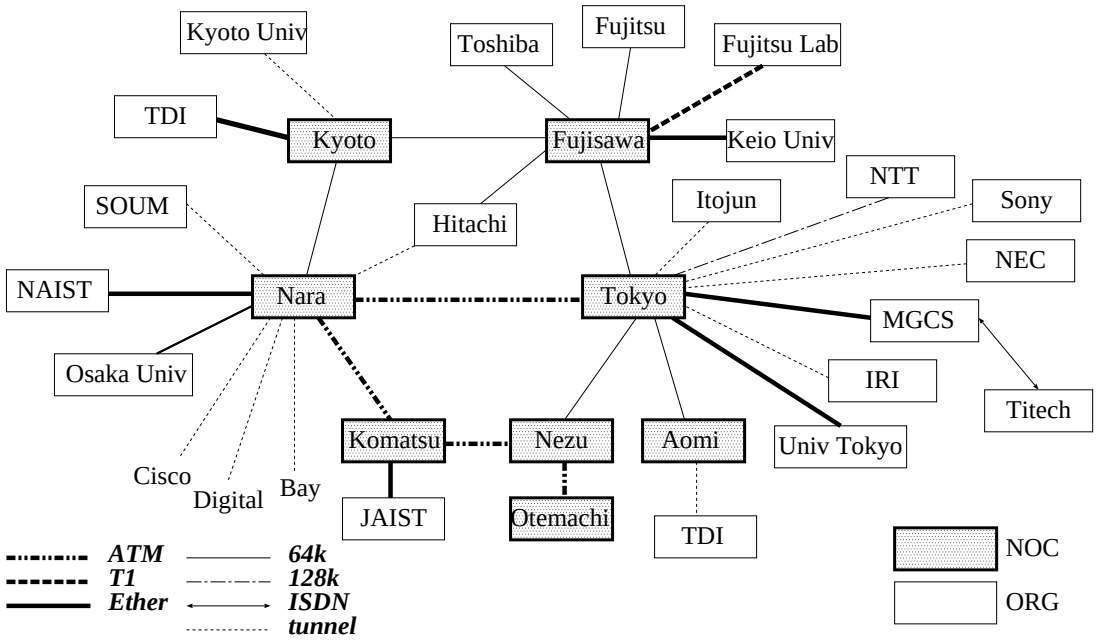


図 1: 現在の WIDE 6bone

には、多重化装置を利用して従来の IPv4 に使用している 1.5Mbps の専用線を 2 分割し、IPv4 に 1472kbps、IPv6 用に 64kbps の帯域をそれぞれ割り当てた。

このように帯域を分割することによって、専用線の工事が不要になる他、IPv4 の運用に影響を与えずに IPv6 の実験ができる。また、IPv6 の障害時には IPv4 経由の接続性を利用して、障害に対応できるという利点もある。

- ATM

  ATM の利用法には幾つかの形態がある。LAN エミュレーションに関しては、IPv6 に対する本質的な問題はないし、シェーピングが必要な遠隔地との接続には不適当であるため、WIDE 6bone では PVC を用いた Point-to-Point(以下、P2P と略記) モデルを用いることにした。実装時には PVC 上で IPv6 を配送するための仕様が存在しなかったので、IPv4 の仕様を自然に拡張することにした [6]。この形態では、ATM スイッチを柔軟な多重化装置と考えることができ、シリアルリンクで述べたのと同様の利点がある。

- その他

  ISDN を用いて IPv6 ネットワークを構築しているデータリンクや、1.5Mbps の回線をすべて IPv6 のために使用しているデータリンクも存在する。

3.2 実装

WIDE Project では、IPv6 に関する情報は共有し議論はするが、コードは共有しない方針で、複数の実装を並行に開発してきた。そのため初期の段階から WIDE 内で相互接続性を検証できた。

現在の WIDE 6bone は以下の実装から構成されている。

- Hydrangea (WIDE Project)
- globe (慶応義塾大学)
- v6 daemon (尾上氏)
- COND (東芝)

- Libra、Aries (NEC)
- LR450、LR550 (富士通)
- NR60 (日立製作所)

Hydrangea は WIDE Project の参照コードとして開発されている。WIDE Project で提案された技術はまず Hydragea に実装後検証され、優れたものは他の実装にも採り入れられることが多い。

これらの中で特徴的な実装は、v6 daemon である。これは BPF や NIT を介してユーザ空間で動作する IPv6 スタックである。そのため、例えば SunOS 4.1.4 などの IPv6 スタックが提供されていない OS 上でも IPv6 が動作する。

運用当初、実装のほとんどは PC ベースで開発が進められてきたが、現在では富士通の LR450、LR550、日立製作所の NR60 などの専用ルータも導入されはじめている。

なお、WIDE Project で開発されたものではないが、Bay Networks のルータ、Solaris および Linux も WIDE 6bone の構成要素の 1 つとして運用されている。

3.3 アプリケーション

WIDE 6bone ではさまざまなアプリケーションが開発、運用されている。

`ping`、`traceroute`、`telnet`、`ftp`、`inetd` などの基本アプリケーションは IPv6 スタックの開発初期に移植された。IPv6 スタックの相互接続試験の道具として、また WIDE 6bone の運用や保守にも役立っている。

電子メールの配送はインターネットにおいて重要なサービスの 1 つである。電子メールの配送に不可欠な DNS(`bind`、`newbie`)、MTA(`sendmail`) を IPv6 対応に修正し、運用を開始した。しかし、これらのサービスについては多重プロトコルでの運用という問題を含んでいる。この問題については 4 章で述べる。

さらに現在では、IPv4 に代わって IPv6 を通常利用できるように、WWW ブラウザ、WWW サーバ、`ssh`、`rsh` などの応用アプリケーションの移植を進めている。IPv4 アプリケーションの IPv6 への移植はそう難しい作業ではない、我々は移植の方法を提示することによって IPv6 の推進に努めている[1]。

3.4 運用技術

ネットワークを運用するには、スタックやアプリケーションの開発だけではなく、保守や記録収集といった技術も必要になる。この章では WIDE 6bone に導入している運用技術について説明する。

統計

WIDE 6bone では 1 時間に 1 回、接続性に関する情報を収集して、ホームページ上でその集計を公開している[2]。情報を視覚化することはトラブルの早期発見、修復に役立つ。 また、MRTG を IPv6 に移植、運用して 6bone を流れるパケットの特徴を解析している[3]。

WIDE 6bone 登録局

WIDE 6bone の拡大に伴って、参加組織が増大していくことが予想される。従来は人手で運営してきた接続情報等のデータベース管理、トポロジーマップの更新等をできる限り自動化すべきである。我々は WIDE 6bone 登録局ホームページを用意し、データベースを管理している[4]。また、6bone 登録局のデータ更新なども行っている。

IPv4 とは異なり、IPv6 ではネットワーク・アドレスが多段の階層構造になっているので、アドレスの割り当てやその管理が分散できるという利点がある。

[1] `http://www.itojun.org/hack/v6/v6ready.html`
[2] `http://www.v6.wide.ad.jp/Connectivity/ping/`
[3] `http://mango.itojun.org/mrtg/`
[4] `http://www.v6.sfc.wide.ad.jp/6Bone/`

4 問題点とその解決

WIDE 6boneの運用を通して、様々な問題や課題が発見された。この章ではそれらの問題を示すとともに、既に解が得られているものには解決方法を提示する。

4.1 経路制御の段階的な発展

2章で示したように、WIDE 6boneは静的な経路制御からはじまり、運用上の限界に達すると、単純な経路制御プロトコルであるRIPngに移行、さらにRIPngの限界点に達したため、より複雑なBGP4+へと移行を開始している。これはIPv4ネットワークの成長の過程と酷似している。

4.2 P2Pネットワークのアドレス割り当て

P2Pネットワークには、ノードが2つしか存在しないにも関わらずグローバル・アドレスを割り当てる必要がある。

NOC間などでは集約可能なアドレスを割り振るのが難しく、経路数が増大する原因となる。また、NOCのルータは多数のトンネルやP2Pネットワークを敷設するため、1つのルータに多くのアドレスが付き、管理が繁雑になる。また、2つのノードが異なる組織に属するときに、どのアドレスを割り振るべきかという問題もある。

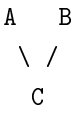
そこで、WIDE 6boneではP2Pネットワークにはリンクローカル・アドレスのみを割り当てて運用している。そのため後述の始点アドレス選択アルゴリズムに変更を加えている。

これはIPv4にも見られた問題とその解決方法であるが、さらにIPv6では経路制御プロトコルはリンクローカル・アドレスを用いるため、ルータのみが存在するセグメントでは、グロバール・アドレスを割り当てる必要がない。将来的には、ルータが数十台あるIXなどでリンクローカルのみで運用することが考えられる。

4.3 マルチホームでの運用

WIDE 6boneではマルチホームと呼ばれる複数のプロバイダに属し、複数のアドレスを持つ組織を実験的に作り、運用している。この運用を通じて、マルチホームにはIPv4と同様にさまざまな問題があることが分かった。

以下の図を例にとって説明する。図中、A、B、Cは組織、線は回線を意味する。



Cという組織がA、Bという2つのプロバイダに接続し、アドレスA:C、B:Cが与えられているとする。マルチホームの組織Cは以下の要求がある。

1. 往復パケットは同じ回線を通過させたい
   インターネット上のあるホストに、Cのホストが通信をはじめたとする。外向きのパケットがA-Cの回線を通過するとすれば、内向きのパケットもA-Cを通過させたい。そのためにはパケットの始点アドレスにはB:Cではなく、A:Cを選択する必要がある。
2. 回線を有効活用したい
   A-C、B-Cの回線を両方とも有効に利用したい。どちらか一方の帯域だけが専有されるのは避けたい。

現在のところこれらの要求を満たす一般的な解は得られていない。

運用に関しては、設定誤りによりマルチホーム組織が他組織の経路をアナウンスしてしまい、プロバイダのようにパケットを通過させてしまうといった運用上の誤りが見受けられた。上位プロバイダであるA、Bはそれぞれ A:C、B:C以外の経路を受理してはならないし、Cもその経路以外をアナウンスするべきではない。

4.4 リナンバリング

ある組織をリナンバリングするときには、できるだけ既存の通信に影響を与えないように、

緩やかに移行する必要がある。

WIDE 6bone では度重なるリナンバリングの結果、緩やかなリナンバリングの運用経験を蓄積できた。これは、IPv6 では 1 つの IF に対して複数のアドレスを登録できるという特徴を利用したものである。リナンバリングの大まかな運用手順を以下に示す。

1. ルータに対して新アドレスを追加する。旧アドレスは消去してはならないが、始点アドレスとして選択しないようにする。
2. ホストに対して新しいルータ通知メッセージを送る。古い情報は送出せずにタイムアウトして消えるのを待つ。
3. 新しい経路を組織外に対してアナウンスする。古い経路のアナウンスも継続する。
4. ネームサーバの設定を変更する。正引きに対しては旧アドレスを消去し、新アドレスだけを設定する。逆引きに関しては新旧とも登録しておく。
5. インターネットで名前情報が更新されるのを待つ。およそ数日あればよい。
6. 古い経路のアナウンスをやめる。
7. ルータに登録している旧アドレスを消す。
8. 古い逆引き情報を消す。

この手順をさらに考察して、可能な限り自動的にリナンバリングを進める方法についても提案していきたい。

4.5 アドレス・スコープ

IPv6 では複数のスコープを用意し、異なるスコープのアドレスを複数 IF に割り当てることを設計段階から考慮していた。インターネット全域での一意性が保証され、パケットの到達可能性があるのはグローバル・アドレスである。この他に、サイト内しか一意性、到達可能性がないサイトローカル・アドレスや、同一データリンク内しか一意性、到達可能性がないリンクローカル・アドレスもある。

実際に通信する際には、始点と終点のアドレス・スコープが異なっている場合には、片方向の到達可能性しか得られないという問題が発生する。我々はこれに対して、終点アドレスと同じアドレス・スコープを持つ始点アドレスを選択するべきであることを提案した [7]。終点アドレスより広いアドレス・スコープを持つ始点アドレスを選択すればよいが、現実の運用では、アドレスの安定性などを考慮すれば、無闇に広いアドレス・スコープのアドレスを選択することは問題がある。そのため、始点と終点のアドレス・スコープは同じにしておくべきである。

この提案に、前述のリンクローカルのみの P2P ネットワーク、マルチホームでの運用、リナンバリングを組みあわせると、ある終点アドレスが与えられた場合の始点アドレス選択のアルゴリズムは以下のようになる。ただし、この規則は終点アドレスがグローバルもしくはサイトローカルのときに限る。

1. 自ノードに登録されているアドレスのうち、異なるスコープのものを候補から外す。
2. 残りの候補のうち、リナンバリングのための旧アドレスは候補から外す。また、エニーキャスト・アドレスも始点アドレスには使えないので候補から外す。
3. 残った候補のうち、送出 IF に登録されたアドレスが 1 つ以上残っていれば、その中から選択する。
4. 送出 IF に候補がなければ、残りの候補から選択する。
5. 候補が 1 つも存在しなければエラーを返す。

候補が複数残ったときにどれを選択するかについては明確な結論が出ていない。現在は最長一致したものを選択している。その理由はマルチホーム環境化である程度のデータリンクの選択ができるからである。しかしながら、マルチホームの問題を完全に解決するものではない。

4.6 多重アプリケーションの運用

多重アプリケーションとは、同じサービスを IPv6 と IPv4 の両方に対して提供しているアプリケーションである。多重アプリケーションの

運用には慎重な取り扱いが必要である。

例えば、MTA の運用で同一 MX ドメイン内で複数のメール・サーバを用意したときに、最も優先順位の高いホスト、すなわち終点への到達を保証する必要がある。その実現例としては、終点のサーバをデュアル・スタックにすることが挙げられる。

また、古いメールサーバの中には IPv6 アドレスを取り扱えないだけではなく、IPv4 まで誤動作する例も報告されている。

これらの運用技術については、Internet Draft にまとめる予定である。

4.7 トランスレータ

IPv4 と IPv6 は異なるプロトコルであるため、これらのプロトコルは相互に通信できない。IPv4 から IPv6 への移行を円滑にすすめるためにはトランスレータが不可欠である。

WIDE Project では現在 3 種類のトランスレータを提案し開発している。

- SOCKS64 (富士通研究所)[8]
- FAITH (奈良先端科学技術大学院大学)
- NR60 (日立製作所)

FAITH と NR60 の基本概念については IETF に Internet Draft を提出している [9]。

5 今後の課題

今後はメーリング・リスト等の重要サービスの IPv6 への移行や、トランスレータによる円滑な移行の補助を検討している。

海外との経路交換は BGP4+に移行しはじめている。WIDE 6bone 内は現在はすべて RIPng で運用しているが、NOC 間では BGP4+の導入を検討している。

謝辞

本稿は WIDE Project IPv6 分科会のみなさんの精力的な活動の集大成です。議論に参加し、運用に携わった多くの方々に感謝します。

参考文献

[1] R. Hinden and J. Postel. IPv6 Testing Address Allocation. RFC 1897, January 1996.

[2] G. Malkin and R. Minnear. RIPng for IPv6. RFC 2080, September 1997.

[3] R. Hinden and S. Deering. IP Version 6 Addressing Architecture. Internet Draft, November 1997.

[4] T. Bates, R. Chandra, D. Katz and Y. Rekhter Multiprotocol Extensions for BGP-4 Internet Draft, September 1997.

[5] P. Marques and F. Dupont Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing Internet Draft, November 1997.

[6] K. Yamamoto, K. Cho, Y. Inoue, H. Esaki, Y. Atarashi and A. Hagiwara. IPv6 over Point-to-Point ATM Link. Internet Draft, September 1997.

[7] K. Yamamoto, A. Onoe and A. Kato The IPv6 communication model. Internet Draft(expired), September 1996.

[8] 陣崎 明、小林 伸治 SOCKS64:SOCKS プロトコルを用いた IPv4-IPv6 相互接続ゲートウェイ. インターネットコンファレンス'97, December 1997.

[9] K. Tsuchiya, M. Sumikawa, K. Watanabe, Y. Atarashi, T. Miyamoto, K. Yamamoto and J. Murai A Communication Mechanism between IPv4 and IPv6. Internet Draft, December 1997.