

## 第 10 部

# マルチキャスト通信



# 第 1 章

## はじめに

MC WG では、JP Mbone コミュニティーと係し、日本国内の MBONE の運用に携わり、トラフィック監視やツールの開発を行い、安定運用に貢献している。ここでは、JP Mbone の現状を述べ、MC WG が提供している統計収集技術とその情報を報告する。また、WIDE 大手町 NOC で行なっている multicast exchange project について報告する。

## 第 2 章

### 国内の MBONE の現状

まず、日本国内の MBONE アクティビティを統計情報で示す。表 2.1 は、1998 年 4 月 27 日現在の、JP MBONE に接続されているルータと組織の数である。

表 2.1: JP MBONE ルータと組織数

	合計	ac.jp	ad.jp	co.jp	go.jp	gr.jp	or.jp	ne.jp	etcJP	notJP	NA
組織数	82	37	16	14	7	0	4	0	1	2	1
(active)	58	25	13	9	5	0	4	0	0	1	1
ルータ数	213	116	38	16	7	0	4	0	1	6	25
(active)	121	67	25	9	5	0	4	0	0	5	6

(etcJP = その他の jp、notJP = jp 以外、NA = DNS 未登録)

この調査は、DVMRP\_ASK\_NEIGHBORS2 の IGMP を用いて `mbone.otemachi.wide.ad.jp` から順に `threshold` が 64 未満の範囲をたどっていくことで得られるデータを元に算出した。次に、JP Mbone で使われているマルチキャストルータの種類を表 2.2 に示す。

表 2.2: JP MBONE ルータの種類 (1998 年 4 月 27 日)

ルータの種類	数
11.1PM	9
11.2PM	9
11.3PM	2
3.255PGM	50
3.8PGM	51
not active	92
合計	213

現在 JP MBone において使われているマルチキャストルータの種類は、大半が mrouted3.8 および mrouted3.9 であり、一部の組織で cisco IOS version 11 が使われていると推測される。表 2.2 中の version 3.255 は、そのルータが DVMRP version3 に従っていることを示し、実際には、それらはほとんどが mrouted3.9 であると思われる。

JP MBone における経路問題の解決および現状の把握をするため、JP MBone 以外とのリンクや、ループ構造を持つ部分に関係する国内主要ルータについては接続トポロジ図を常に最新の状態にあわせ維持している。ここ 1 年では、新たに SINET とタイとの国際リンクが設けられたほか、国内でも一部主要ルータ間の接続が変更され経路が変わるなどの変化があった。このようすを表 2.3 に示す。

表 2.3: JP MBONE のトポロジ (1998 年 4 月 27 日)

```

Loopback0.MB1.SJC1.Alter.Net --- Loopback0.MB1.SF01.Alter.Net
      | 5,1/64      pim      | 5/64
      maew-mbone.nsn.nasa.gov      dec3800-2-fddi-0.SanFrancisco.mci.net
mr.ai3.net      |      |      mbone.nectec.or.th
| 1/64      | 5,1/64      | 1/64      | 1/64 |
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% boundary 239.133/16 %%%%%%%%%
|      |      |      |
|      mroute00.iij.ad.jp --- sw01.tokyonet.ad.jp      |
|      | 6,3/32 / 1/32      |
| mbone.otemachi.wide.ad.jp ---- mbone.nc-u-tokyo.ac.jp ---- totoro.sinet.ad.jp
|      | 1/32 \ 1/32      | 1/32      |
|      sun1.tokyo.wide.ad.jp --- mbone.imnet.ad.jp -----|
|      | 1/32      | \ 1/32      | 1/32      | 1/32      |
mr.nara.wide.ad.jp | 1/32 ---- ftp.cfi.waseda.ac.jp      |
|      |      | 4/32      |
1/32 | jp-gate.wide.ad.jp --- handshake.gw.kyoto-u.ac.jp -----|
|      | 1/32 / 3/32      | 1/32      |
sun15.kyoto.wide.ad.jp ---- aki.csi.ad.jp ---- saijo.csi.ad.jp ----|
|      | 9/32 8/32      | 1/32      | 1/32      |
sun1.fukuoka.wide.ad.jp ---- nic.karrn.ad.jp -----|
|      | 1/32      | 1/32      |
    
```

また、MBONE 全体で調査したルータ数とルータの種類を表 2.4 と表 2.5 に示す。

表 2.4: MBONE ルータと組織数

ドメイン	ルータ数	ルータ数 (active)	組織数	組織数 (active)
arpa	1	0	1	0
com	122	54	99	38
edu	460	344	113	80
gov	152	113	21	17
mil	81	56	8	7
net	391	283	104	79
org	26	14	21	12
br	17	10	9	4
ca	35	29	24	19
hk	1	0	1	0
id	2	2	1	1
in	1	0	1	0
jp	183	110	79	55
kr	90	37	6	4
ru	3	1	3	1
sg	7	6	4	4
th	2	1	2	1
tn	1	0	1	0
us	9	3	6	3
NA	302	92	-	-
合計	1886	1155	504	325

表 2.5: MBONE ルータの種類

ルータの種類	数
1.0	6
2.2	2
3.3	1
3.6PGM	7
3.8PGM	311
3.38PGM	2
3.255PGM	136
10.2	13
10.3	14
11.0M	1
11.0PM	53
11.1PM	312
11.2PM	268
11.3PM	29
NG	731
TOTAL	1886

## 第 3 章

# MBONE 上のトラフィック収集と監視技術

### 3.1 統計情報収集システムの必要性

マルチキャストネットワークを運営していくにあたって、特に注意を払わなければならないのは既存の運用ネットワークに対して過大な負荷をかけないということである。そのためには、マルチキャストトラフィックを常に監視する必要がある。

WIDE インターネット内のマルチキャストトラフィックの配送では、マルチキャストルータに mrouterd を使用しているため、各拠点において mrouterd に対してシグナル SIGUSR1 を送って得られるダンプファイルを解析し、仮想インタフェース (以下、VIF) 毎に送受信されたパケットの量、バイト数を計測できる。

ある地点のトラフィック量を計測できるが、ネットワーク全体のトラフィックの流れを計測するには、各ルータでこのようなファイルを記録する必要がある、作業が繁雑である。同時に、管理ポリシ的に受け入れられない場合も多い。

このような問題を認識し、リモート監視を容易にするために、マルチキャストルーティングソフトウェアとして、SNMP によって統計データを取得できる mrouterd を利用して統計情報の収集実験を行なった。WIDE バックボーンの mrouterd のすべてで SNMP 対応がなされている。

mrouterd の最新バージョンである mrouterd3.9beta3 では、SNMP 対応が行われていないため、Merit によるバージョン 3.8 の SNMP 化を参考にし、WIDE 内で SNMP 化を行った。(ftp://ftp.kyoto.wide.ad.jp/multicast/mrouterd/snmp-mrouterd/)

当初は perl で作成したスクリプトによって、トラフィック監視を行っていたが、次のような三つの要求が出てきた。

- トラフィックの履歴を残したい
- 問題点がより明瞭になる形で表現したい
- 情報を公開することで、分散トラブルシューティングを促したい

これらの要求に対応するため、World Wide Web によって、統計情報の履歴をリアルタイムで公開するとともに、JP-MBone におけるトラフィックの流れの概略のフロー図を作成、提示することとした。



## 3.2 システム構成

各拠点における統計情報のグラフ化はスイス工科大学 Tobias Oetiker 氏によって製作された MRTG という可視化ソフトウェアを使用した。(http://ee-staff.ethz.ch/~oetiker/webtools/mrtg/mrtg.html)

JP-MBone におけるトラフィックのフロー図の作成にあたっては、小野秀貴氏(京都大学工学部電子工学科)によって製作された makefeedmap を使用した。(http://www.tamaru.kuee.kyoto-u.ac.jp/ono/tools/makefeedmap/)

リンクの生成、消滅に関しては、定期的ポーリングによって自動的に対応が行われるようになっている。

## 3.3 統計情報の収集例

### 3.3.1 大手町 NOC におけるトラフィック

mbone.otemachi.wide.ad.jp における 1998 年 4 月 27 日の統計情報を図 3.1 から示す。

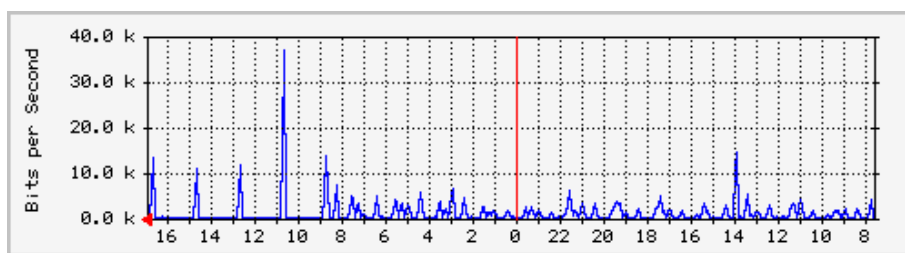


図 3.1: 大手町 MEX 間

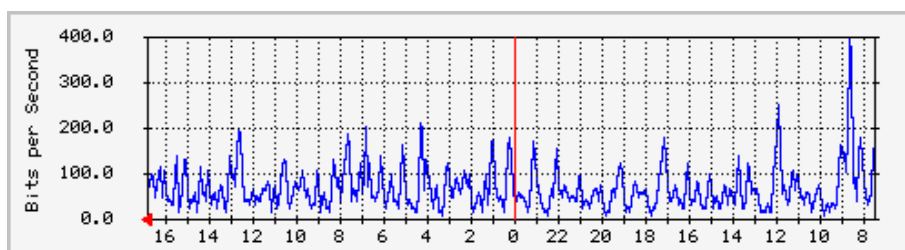


図 3.2: 大手町 IBMNET 間

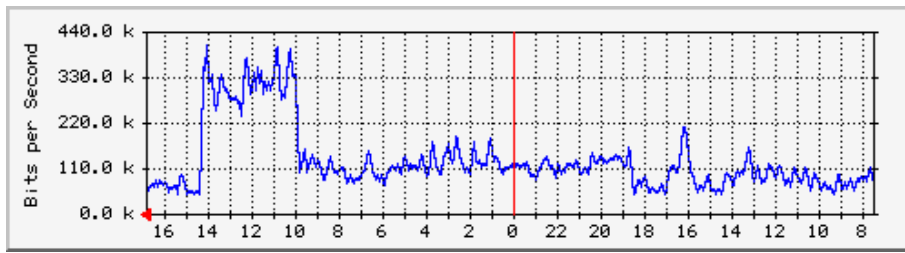


図 3.3: 大手町 IMNET 間

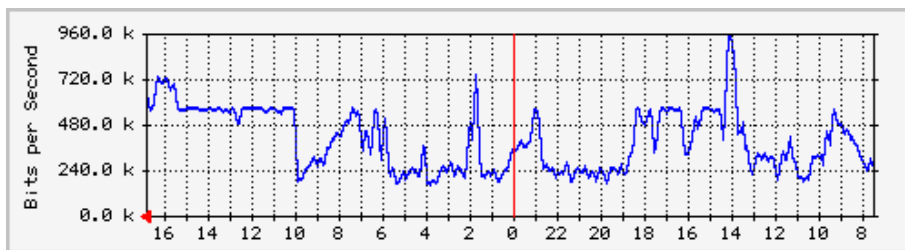


図 3.4: 大手町 東京大学間

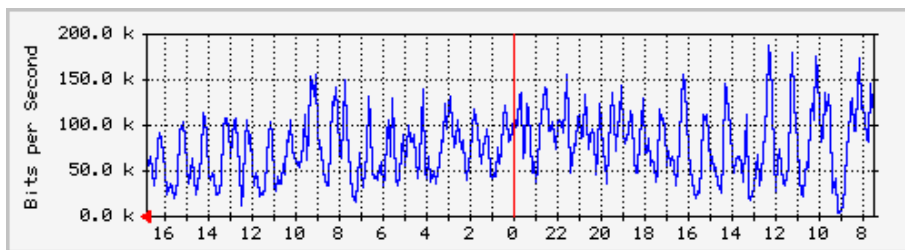


図 3.5: 大手町 MESH 間

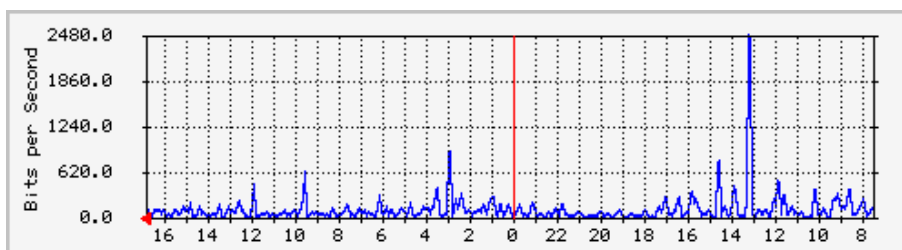


図 3.6: 大手町 INFOSPHERE 間

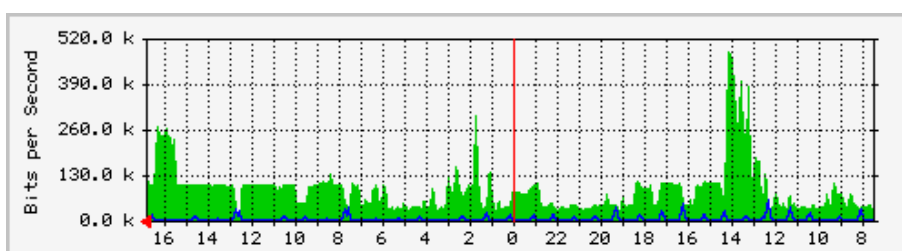


図 3.7: 大手町 IJ 間

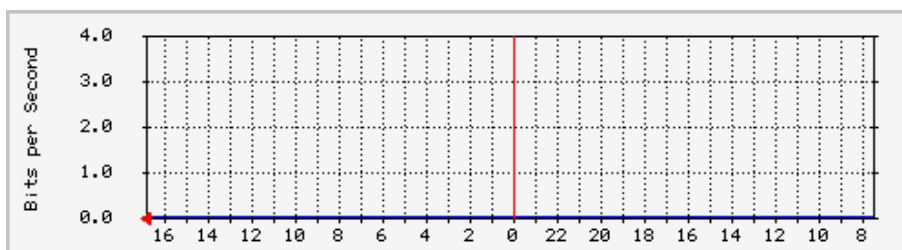


図 3.8: 大手町 青梅間

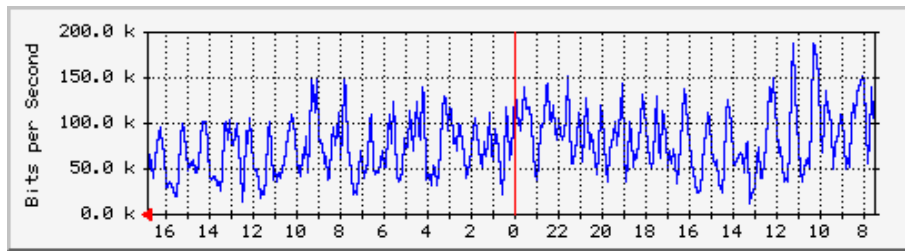


図 3.9: 大手町 SINFONY 間

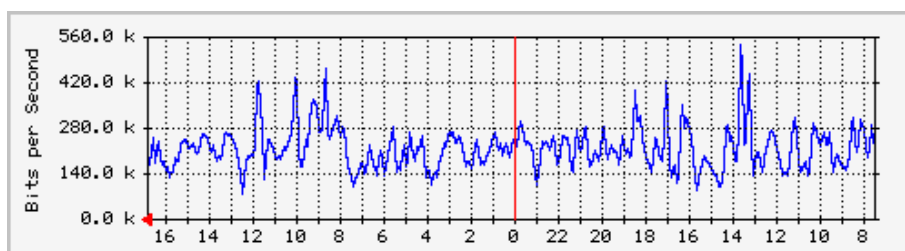


図 3.10: 大手町東京 NOC 間

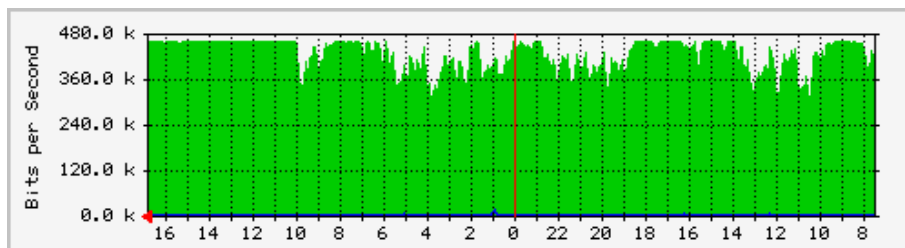


図 3.11: 大手町 TOKYONET 間

### 3.3.2 海外リンクのトラフィック

東京ネットと MCI の間の JP Mbone のトラフィックを図 3.12 から図 3.13 に示す。

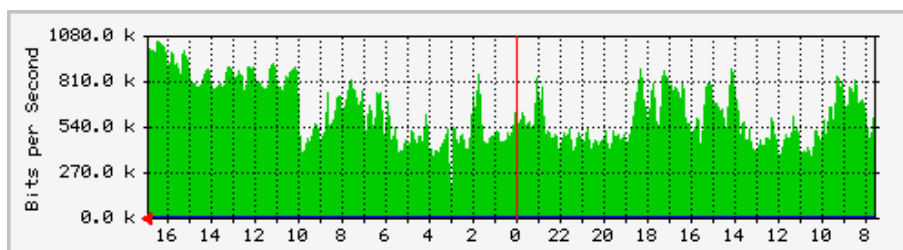


図 3.12: 海外トラフィック (1 日の変化)

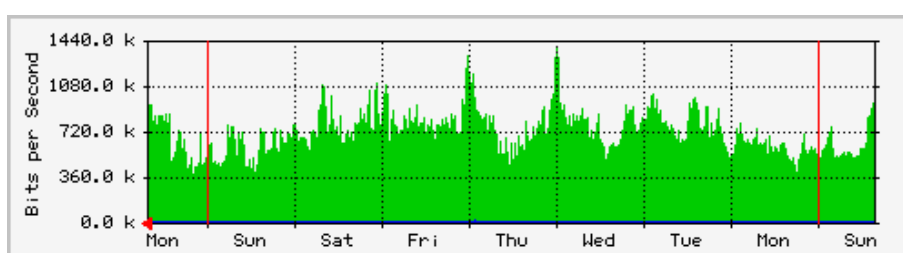


図 3.13: 海外トラフィック (1 週間の変化)

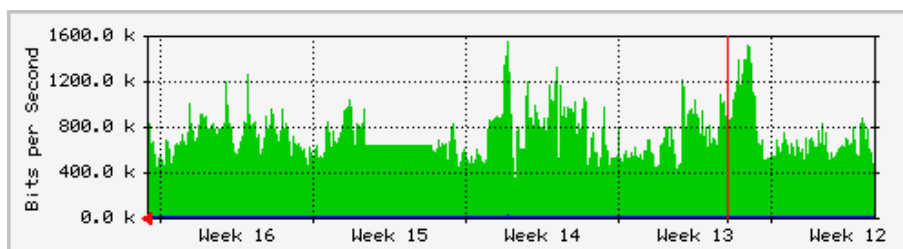


図 3.14: 海外トラフィック (1 か月の変化)

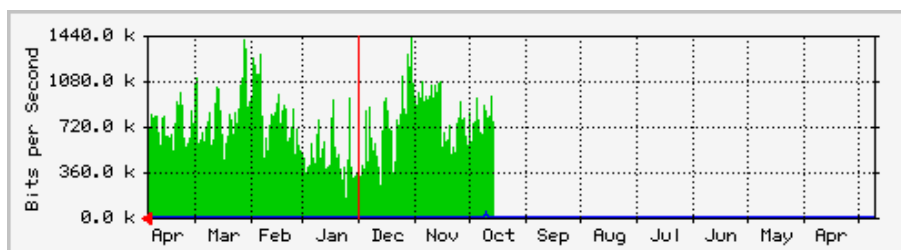


図 3.15: 海外トラフィック (1 年間の変化)

### 3.3.3 流量図

MBONE のトラフィックのフロー解析を行ったものを図 3.16 および図 3.17に示す。

## 3.4 考察

本システムによって、non pruning router の早期発見と、World Wide Web を用いた分散監視を行えるようになった。

また、統計情報を継続的に記録することによって、正しく運用がなされたマルチキャストネットワークのサービスクオリティはロスレートなどの点において、非常に高いことが改めて確認された。

しかしながら、現システムには次のような問題点がある。

- 統計情報の記録単位が VIF であるということ- グループ毎の情報が有効である場合が多い
- ○ 統計情報データベース形式- 現状では MRTG 固有のデータベースであるが、再利用性を考慮した形式を用意すべきである
- SNMP 対応の mouted のみしか対応していない- SNMP 対応がなされていない mouted や、専用ルータの扱いが考慮されていない

これらに関しては、今後の課題としたい。

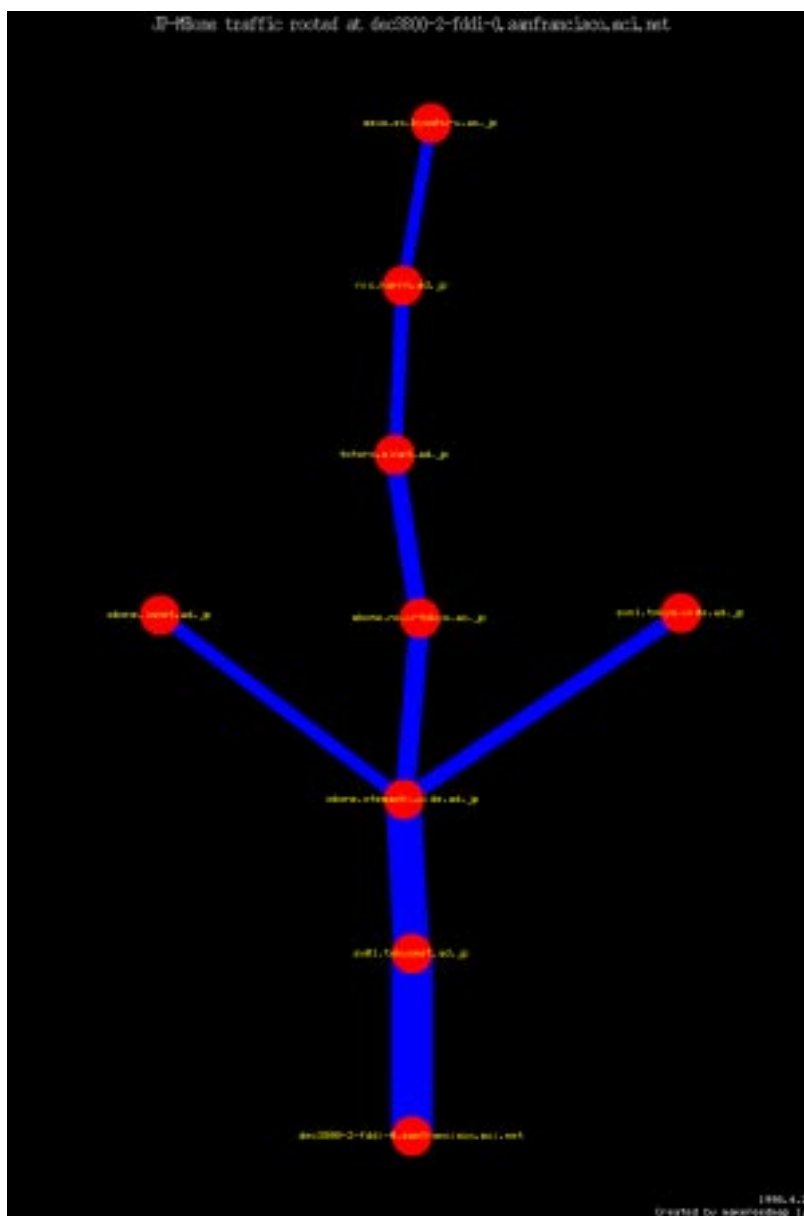


図 3.16: 1998 年 4 月 27 日 03:50 における流量図

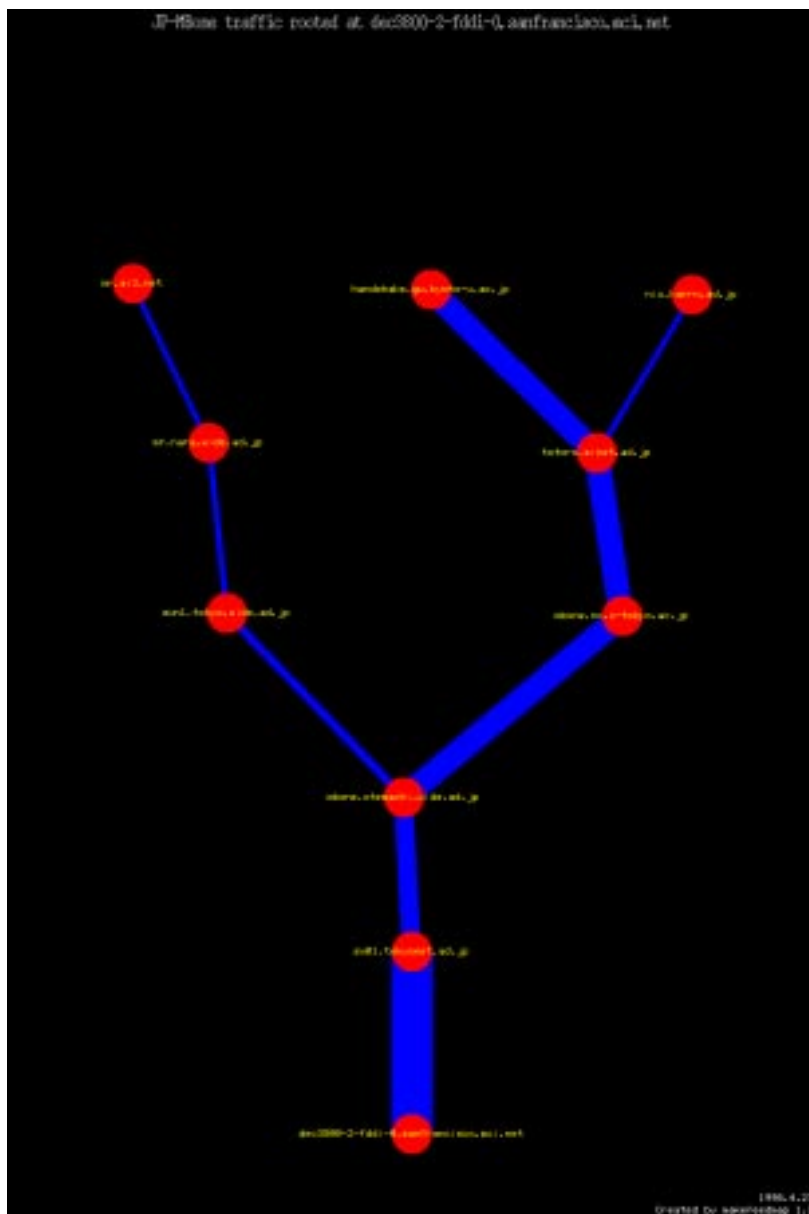


図 3.17: 1998 年 4 月 27 日 16:52 における流量図



## 第 4 章

# multicast exchange project

ISP 間の接続の整理および商用 ISP の新規接続を促すため、1997 年 3 月に WIDE 大手町 NOC に mbone.otemachi.wide.ad.jp を設置し、主に NSPIX2 を経由して各 ISP との間を順次トンネルにより接続を行なってきた。これを、multicast exchange project phase I と呼ぶ。

しかし、この方法は、mbone.otemachi.wide.ad.jp を中心とした放射型の接続トポロジとなっているため、次のような構造上の基本的な問題があった。

- 各 ISP 同士のやりとりが必ず WIDE 経由の間接やりとりになっている。
- 常に mbone.otemachi.wide.ad.jp を通過するため、そこが過負荷になる。
- もしも mbone.otemachi.wide.ad.jp が落ちると、すべて切断されてしまう。
- すべての接続がトンネル利用であり、また、mbone.otemachi.wide.ad.jp はユニキャスト的には各 ISP に対してルータでないため、すべてのトンネリングのパケットが出入り口である 1 つの物理インタフェースに集中し、トンネル数が増えるとともに比例する形で通過するトラフィックが増えてしまう。

そこで、これらの問題を解決するため multicast exchange project phase II として 100BaseTX のセグメントを新規に設置し、そこへ各 ISP のマルチキャストルータを接続することでそれらが直接マルチキャストパケットをやりとりできるようにする新たな研究実験計画を打ち出した。JP Mbone コミュニティへ以下の文書を提案し、現時点で 12 の ISP より実験参加の希望が出ている。

『multicast exchange project phase II について』

目的:

国内における複数 ISP 間での IP マルチキャスト接続交換の研究のため  
接続実験運用を行なう。

期間:

特に厳密な期間は定めず臨機応変に対応するが、およそ一年後をめどに大きく見直すとともに、必要であれば更に継続するものとする。

#### 経緯:

国内の Mbone トポロジの整理を目的に、1997 年 3 月、おもに当時増えてきた商用 ISP のマルチキャストルータを接続するための基幹ルータとして `mbone.otemachi.wide.ad.jp` を NSPIXP2 近傍に設置し、これと複数の ISP のルータ間をそれぞれトンネリングにて接続してきた。これを、multicast exchange project の phase I とする。今回のものはそれを改善発展させて行なう phase II として位置づける。

#### phase I での問題点:

接続のトポロジ的に `mbone.otemachi.wide.ad.jp` を中心とした放射型となっているため、

- ・各 ISP のルータが直接パケットをやりとりする形になっていない。
- ・必ず `mbone.otemachi.wide.ad.jp` を通るため、そこが過負荷になる。
- ・万が一 `mbone.otemachi.wide.ad.jp` が落ちると、すべて切れてしまう。
- ・すべてがトンネリングであり、また、`mbone.otemachi.wide.ad.jp` はユニキャスト的にはルータでないため、すべてのトンネリングのパケットが 1 つの物理インタフェースに集中し、トンネル数が増えるとともに通過するトラフィックが増えてしまう。

などといった、構造上の基本的な問題があった。

#### その他の問題点:

phase I での問題を解決するには、たとえば、NSPIXP2 などにおいて各接続ルータが同時にマルチキャストルータとしてもふるまい、NSPIXP2 上においてマルチキャストの交換もしてしまえばよい。

しかし、現在の状況では NSPIXP2 に各 ISP が接続している専用ルータ、たとえば、cisco などにおいて同時にマルチキャストのルーティングをやらせるには負荷的に問題ありと判断している管理者が多い。また、各 ISP が実際に対外接続マルチキャストルータとして用意している物は PC やワークステーションでの `mrouted` ベースの物が多い。

#### 今回の接続構成:

マルチアクセスのメディアのセグメントを 1 つ新たに用意し、各 ISP のマルチキャストルータをそのセグメントにすべて接続する。そして、そのセグメントにおいてそれら複数のマルチキャストルータが

直接マルチキャストパケットをやりとりするようにする。ただし、このセグメントを越えてのユニキャストの交換は一切行なわない。

今回の構成は、現在国内の MBone に接続している ISP の多くがたまたま NSPIXP2 などがある KDD 大手町ビル内にハウジングしており、同時に對外マルチキャストルータもそこに設置している例が多かったため、その特殊な環境を利用して上記のセグメントを NSPIXP2 近傍に設置する。

セグメントを構成するマルチアクセスのメディアとして、今回は 100BaseTX を採用する。これは、多くの ISP のマルチキャストルータを接続して本格的利用が始まると 10Base ではすぐに飽和が予想されることと、現在において PC などで 100Mbps 級を構成するのにもっとも安価なものは 100BaseTX と思われるためである。また、これに伴い、ここに接続される各 ISP のマルチキャストルータ間については rate limit を撤廃する。

#### NSPIXP プロジェクトとの関係:

今回必要となるセグメントの設置場所および機材などについては NSPIXP プロジェクトが提供し、今回の NSPIXP 関連プロジェクトとしても位置づけられる。(現在調整中)

ただし、この multicast exchange project への参加についてはオープンとし、NSPIXP プロジェクトに参加しているかどうかに関わらず、自由に参加できる。

#### JP MBone コミュニティとの関係:

今回の実験運用において出てくる諸問題および議論や成果などは JP MBone コミュニティ内(実際には mbone-jp メーリングリスト)においてオープンに行なう。つまり、今までと同様に JP MBone の一部としてみんなで運用管理の調整を協調的に行なっていくものとする。

#### 物理的接続について:

今回の接続参加にあたっては、その物理的制約により、設置する 100BaseTX セグメントの HUB の所までカテゴリ 5 ケーブルなどを延ばしてこれることが事実上参加に必要な条件となる。ただし、光ファイバなどで HUB の近くまで延ばしてきて小さな変換器などで変換するなどの方法は交渉次第で可能。

各参加 ISP のマルチキャストルータから HUB へのケーブル敷設など、現場での作業についてはみんなで調整して行なう方が効率がよいことも多いため、物理的な接続などについてのみ調整や話し合いを行なうメーリングリストを新たに設置する。それ以外の、マルチキャストルータ

間のやりとりや設定、その他 Mbone 諸問題についての議論は、  
mbone-jp メーリングリスト上においてオープンに行なう。

なお、今回の物理的接続による制限のために参加できない ISP については  
今まで通り `mbone.otemachi.wide.ad.jp` にてトンネリングによる接続を  
行なうため、JP Mbone への接続は引き続き保たれる。新規接続も受け付ける。

参加費用について:

今回の phase II において必要となる費用は分担を必要とするほど高額では  
ないため、NSPIXP および WIDE プロジェクトが場所や HUB などを用意する。  
そのため、参加をする各 ISP は HUB までの接続だけ用意すれだけでよい。  
ただし、今後の運用実験を進めていく上で、費用のかかる、より高度な  
実験環境を構築することになった場合については、別途決めることとする。