第 XXXI 部

M Root DNS サーバの運用

第 31 部
M Root DNS サーバの運用

第 1 章　はじめに

インターネット上の資源は、木構造の名前空間であるドメイン名によって指定される。ドメイン名から、IP アドレスなどの名前に対応した種々の情報を得る操作は名前の解決と呼ばれるが、この名前解決を担当するシステムが DNS — Domain Name System —である。

DNS では、名前空間を Zone と呼ばれる連続した部分空間に分割して管理が行われており、分散的なアルゴリズムによって名前の解決が行われる。木構造の頂点である Root ゾーンの解決を行う DNS サーバは、特に Root DNS サーバと呼ばれており、DNS の名前解決にとって非常に重要である。特に DNS での UDP を用いた場合のメッセージ長の制約から、多数の Root DNS サーバを設定することはできない。DNS ではキャッシュを多用することによって効率を改善するとともに、Root DNS サーバ等の上位ドメインに対応するゾーンを担当するサーバへの問合わせを減らすような努力がなされているが、Root DNS サーバが重要な存在であることには変わりはない。

Root DNS サーバは現在 A.ROOT-SERVERS.NET〜M.ROOT-SERVERS.NET という 13 システムで運用が行われている。このうち、M.ROOT-SERVERS.NET は、1997 年 8 月に WIDE プロジェクトによって運用が始まった。Root DNS サーバはインターネットにおける分散が制限されている資源の一つであるため、障害等によるサービス中断を最低限に押さえる必要がある。そのため、M Root DNS サーバは、1997 年の運用開始時から、サーバの冗長構成を導入し、主サーバの障害時には副サーバが自動的にサーバ機能を提供するような運用を行っている。

現在は、以下に示すような基本構成をユニットとし、後述の Anycast を用いてサービスの提供を行っている。各ユニットは 4 台のサーバから構成されており、サーバの OS や DNS ソフトウェアの更新時にもサービスを停止する必要はない。ルータなどの更新時にはサービスを停止せざるを得ないが、サービス停止に先だって経路広告を停止することにより、問合わせは他の active な Anycast サーバによって処理されるため、事実上のサービス停止は発生しない。

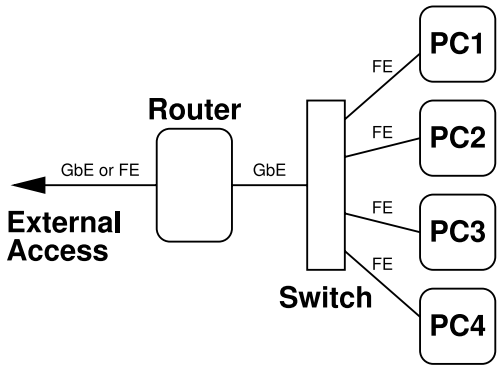


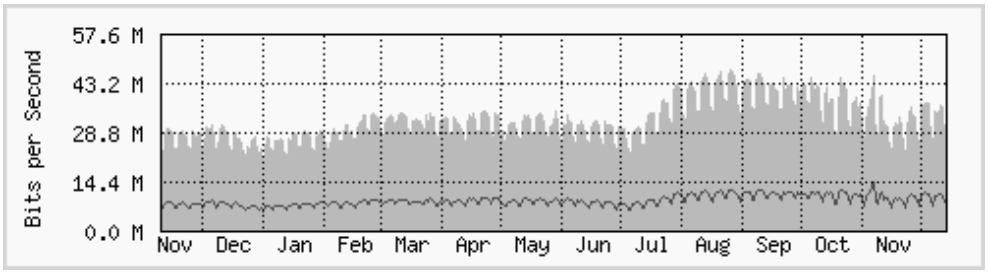
図 1.1. Anycast 用基本構成

第 2 章　Anycast

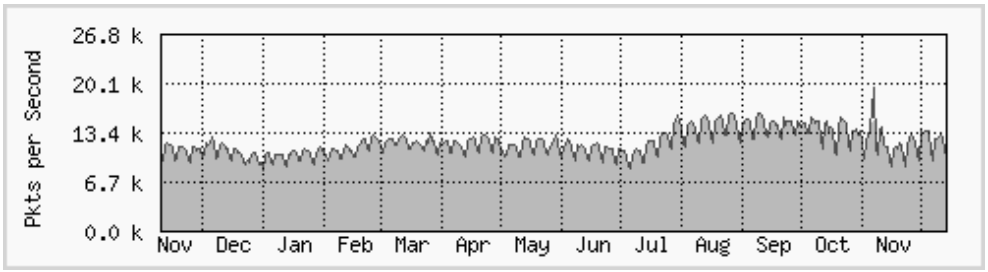
Root DNS サーバは 13 台と限られた存在であるため、インターネット上に普く分布させることはできない。そこで、同じデータを供給するサーバを複数インターネット上に設置し、それぞれのサーバは同一サービスアドレスでサービスを提供する様にする。このサービスアドレスを含む経路情報を BGP でアナウンスすることにより、BGP の経路選択ポリシに依存するものの、一つのアドレスで複数台のサーバを運用することができる。この運用方法は RFC3258 “Distributing Authoritative Name Servers via Shared Unicast Addresses”[60] で定義されており、一般的には BGP Anycast と呼ばれている。

この Anycast に関しては、RFC が出版されたのは 2002 年 4 月であるが、最初の Internet Draft が IETF の DNSOP WG に提案されたのは 1999 年 10 月であり、その間議論が続けられてきた。

M Root DNS サーバでは、2004 年に入り、Seoul



(a) トラフィックの推移



(b) パケット数の推移

図 2.1.　2008 年における M-Root DNS 全体の問合わせ数の推移

(KR) および Paris (FR) での設置を行ない、運用準備を進めてきた。このうち、Seoul に関しては、韓国で唯一の Layer-2 IX である KINX — Korea Internet Neutral Exchange —のご協力を得て、2004 年 7 月 21 日より運用を開始した。経路広告に BGP の NO_EXPORT 属性を添付するいわゆる local anycast として運用を行なっているが、学術系のネットワークの収容を目的として NCA — National Computerization Agency —が運用している Layer-3 IX である KIX では、NO_EXPORT を外して学術系ネットワークに対して経路の広報を行なっている。しかし、韓国での主要二大 ISP である KT および Daemon への接続性がないため、現在、Seoul で処理されている問合わせは毎秒 50〜100 クエリ程度と多くはない。

一方、Paris は Telehouse Europe、Renater、France Telecom、および Open Transit の協力を得て、Telehouse Voltaire にて 2004 年 9 月 1 日より運用を開始した。ここでは二つの独立な IX である、Renater が運用する SFINX と France Telecom が運用する PARIX に接続している他、2004 年 10 月からは TISCALI が独立に transit を提供して頂いている。現在は多くの ISP に対して NO_EXPORT をつけて経路広告を行なっているが、幾つかの ISP に対しては NO_EXPORT なしに経路広告をしている。ヨーロッパ全域にサービスを提供している transit ISP とも多く peer しているため、そのサービスエリアはフランスに留まっていない。このため、毎秒 4000 クエリ程度の問合わせがある。

San Francisco は WIDE San Francisco NOC に設置されており、WIDE とは別な FastEthernet で PAIX/Palo Alto に接続されている。WIDE の Los Angeles での upstream である AS701 からのトラフィックは東京に送るのではなく San Francisco で処理されている。また、アメリカ合衆国の研究教育ネットワークである Internet2 Network とは IPv6 による PAIX 上の peer をしているが、2006 年夏に IPv4 での peer を追加した。これによって、アメリカ合衆国の主な大学からの M-Root DNS サーバへの問合わせは TransPAC 等を経由して東京で処理されるのではなく、San Francisco で処理されるようになり、RTT の改善に貢献している。

図 2.1 に M-Root 全体に対するトラフィックの 2008 年における推移を示す。2008 年 7 月からトラフィック増加が見て取れる。これは主に DIX-IE における M-Root DNS への問い合わせが増加したためである。本年 7 月頃に話題となった Kaminsky Attack によって、各所の DNS サーバ実装がアップデートされたことに起因すると思われるが、正確な理由は不明である。

第 3 章 他の Root DNS サーバ

2002 年 10 月 22 日早朝（日本時間）に発生した 13 台の Root DNS サーバをターゲットにした DDoS 攻撃をきっかけに、幾つかの Root DNS サーバでは、Anycast サーバの設置を図っている。特に、ISC が運用している F Root DNS サーバでは、APNIC 等との協調により、精力的に Anycast サーバの設置を行っている。

2008 年 12 月時点での Root DNS サーバの設置状

表 3.1. Root DNS サーバの設置状況

サーバ	設置都市		
A	Dulles, VA	Ashburn, VA	
B	Marina Del Rey, CA		
C	Herndon, VA Chicago, IL	Los Angeles, CA Frankfurt (DE)	New York, NY Madrid (ES)
D	College Park, MD		
E	Mountain View, CA		
F	Ottawa (CA) New York, NY Hong Kong (HK) Auckland (NZ) Seoul (KR) Dubai (AE) Brisbane (AU) Lisbon (PT) Jakarta (ID) Prague (CZ) Nairobi (KE) Santiago de Chile (CL) Torino (IT) Caracas (VE) Quito (EC) Cairo (EG)	Palo Alto, CA San Francisco, CA Los Angeles, CA Sao Paulo (BR) Moscow (RU) Paris (FR) Toronto (CA) Johannesburg (ZA) Munich (DE) Amsterdam (NL) Chennai (IN) Dhaka (BD) Chicago, IL Oslo (NO) Kuala Lumpur (MY)	San Jose, CA Madrid (ES) Rome (IT) Beijing (CN) Taipei (TW) Singapore (SG) Monterrey (MX) Tel Aviv (IL) Osaka (JP) Barcelona (ES) London (UK) Karachi (PK) Buenos Aires (AR) Panama (PA) Suva (Fiji)
G	Colombus, OH		
H	Aberdeen, MD		
I	Stockholm (SE) London (UK) Oslo (NO) Brussels (BE) Bucharest (RO) Tokyo (JP) Jakarta (ID) Perth (AU) Miami, FL Beijing (CN) Colombo (LK)	Helsinki (FI) Geneva (CH) Bangkok (TH) Frankfurt (DE) Chicago, IL Kuala Lumpur (MY) Wellington (NZ) San Francisco, CA Ashburn, VA Manila (PH)	Milan (IT) Amsterdam (NL) Hong Kong (HK) Ankara (TR) Washington, DC Palo Alto, CA Johannesburg (ZA) Singapore (SG) Mumbai (IN) Doha (QA)
J	Dulles, VA (3 sites) Miami, FL Chicago, IL Honolulu, HI Dallas, TX Stockholm (SE) (2 sites) Beijing (CN) Nairobi (KE) Dublin (IE) Warsaw (PL) Sofia (BG) Toronto (CA) Vienna (AT) Turin(IT) Brussels (BE) Frankfurt (DE)	Ashburn, VA Atlanta, GA New York, NY Mountain View, CA, (2 sites) Amsterdam (NL) Tokyo (JP) Singapore (SG) Montreal (CA) Sydney (AU) Brasilia (BR) Prague (CZ) Buenos Aires (AR) Fribourg (CH) Mumbai (IN) Paris (FR) Riga (LV)	Vienna, VA Seattle, WA Los Angeles, CA San Francisco, CA (2 sites) London (UK) Seoul (KR) Kaunas (LT) Quebec (CA) Cairo (EG) Sao Paulo (BR) Johannesburg (ZA) Madrid (ES) Hong Kong (HK) Oslo (NO) Helsinki (FI)
K	London (UK) Athens (GR) Reykjavik (IS) Poznan (PL) Tokyo (JP) Delhi (IN)	Amsterdam (NL) Doha (QA) Helsinki (FI) Budapest (HU) Brisbane (AU) Novosibirsk (RU)	Frankfurt (DE) Milan (IT) Geneva (CH) Abu Dhabi (AE) Miami, FL
L	Los Angeles, CA	Miami, FL	
M	Tokyo (JP) San Francisco, CA	Seoul (KR)	Paris (FR)

況を表 3.1 に示す。各サーバの最初の都市が元々運用されていた都市であり、それ以降は Anycast によるものである。Anycast の運用形式も各サーバで異なっており、例えば、C では Cogent Communications のバックボーンにおける IGP による Anycast を実施している他、F では、Palo Alto, CA と San Francisco, CA のサーバはグローバルな経路広告を行っているのに対し、その他の F サーバは原則として、経路情報に NO_EXPORTBGP Community を添付することによるローカルな Anycast サービスを提供している。

第 4 章　IPv6 サービス

Root DNS サーバは、IANA が生成する Root ゾーンを無編集でそのまま提供することになっている。Root ゾーンの生成は技術的には Verisign で行われているが、全ての変更は IANA の指示に基づくものであり、U.S. Department of Commerce の承認を経たものである。Root ゾーンに含まれる TLD に関して、IPv6 のサービスを提供するために必要な TLD サーバに関する AAAA レコードは 2004 年 7 月 21 日に JP および KR のサーバに対して最初に追加された。これによって、Root への問合わせは依然として IPv4 による必要があるものの、TLD 以下に対しては IPv6 のみでも解決が可能な名前が存在したことになる。2008 年 11 月現在、280 ある TLD のうち 153 の TLD が、少なくとも一台は IPv6 でアクセスできるサーバで提供されている。むろん、殆んどの名前の解決を IPv6 で行うことは現在もできないが、IPv6 で解決可能な範囲は少しづつ広がっている。

Root DNS サーバのアドレスは root.cache などのファイルによって与えられるが、Root DNS サーバのアドレスはあまり頻繁な変更はないものの、未来永劫に一定というわけでもない。そのため、bind などの DNS Software は、起動時に root.cache ファイルで与えられたサーバに対して、QTYPE = NS、QCLASS = IN、QNAME = "." なる問合わせを発行し、最新の Root DNS サーバの一覧およびそのアドレスを知り、以降はこの得られた情報を元に問合わせを行う。この機構は *priming* と呼ばれ、この機構によって Root DNS サーバまでの到達性を確認し、最新の Root DNS サーバの一覧を取得することができる。すなわち、ある DNS サーバが起動時に保有している Root DNS サーバの一覧が古いものであったとしても、少なくとも一台の Root DNS サーバから応答を得ることができれば、Root DNS サーバの最新一覧を取得することができるため、動作に影響を及ぼすことはない。

Root DNS サーバにおける IPv6 サービスは長い間の懸案事項であった。一つの要因は、特に *priming* 時におけるパケット長の問題があった。DNS では、UDP を用いた問合わせの場合、応答パケット長は UDP や IP のヘッダを除いた（但し 12 byte の DNS 固定ヘッダは含む）メッセージ長の上限は 512 byte と定められている。この制約は RFC2671[182] に規定される EDNS0 によって緩和することができるが、依然として多くの DNS の問合わせは EDNS0 の pseudo-RR を伴っておらず、これを前提とした運用を始めるには時期尚早である。

IPv4 のみの時代での priming の応答は 13 の NS レコードおよび 13 の A レコードを含む 436 byte であった。この場合、76 byte 余裕があることになり、AAAA レコードが一つ 28 byte であることを考えると、二つまでの AAAA レコードは 512 byte の制約に抵触せずに追加することができる。しかし実用的に考えた場合、特に IPv6 の接続性が IPv4 のそれに比べてまだ十分な密度ではないため、二つというのは不十分であることが指摘されていた。

このことから、2 つ以上の Root DNS サーバの AAAA レコードを追加する方向で ICANN RSSAC/SSAC を中心に議論がまとまり、2007 年 1 月に SSAC の答申である

> Accomodating IP Version 6 Address Resource Records for the Root of the Domain Name System (SAC018)

が公表された。一つの懸念は、EDNS0 を併用し UDP によって 512 byte より大きな応答が返った場合、それを廃棄するような firewall 装置が存在していたことであるが、これも新しい firmware に更新することにより解決できるため、大きな障害にはならないと判断された。その結果、2007 年秋までに正式に IANA に対して AAAA レコードの追加を申請していた F/H/K/M の 4 つの Root DNS サーバに関して実施することが ICANN Board で承認された。これは 2007 年 12 月 31 日付で、2008 年 2 月 4 日に実施する旨のアナウンスが幾つかの Mailing List に投

表 4.1. Root DNS サーバのアドレス

Root DNS サーバ	IPv4 アドレス	IPv6 アドレス
A.ROOT-SERVERS.NET	198.41.0.4	2001:503:ba3e::2:30
B.ROOT-SERVERS.NET	192.228.79.201	
C.ROOT-SERVERS.NET	192.33.4.12	
D.ROOT-SERVERS.NET	128.8.10.90	
E.ROOT-SERVERS.NET	192.203.230.10	
F.ROOT-SERVERS.NET	192.5.5.241	2001:500:2f::f
G.ROOT-SERVERS.NET	192.112.36.4	
H.ROOT-SERVERS.NET	128.63.2.53	2001:500:1::803f:235
I.ROOT-SERVERS.NET	192.36.148.17	
J.ROOT-SERVERS.NET	192.58.128.30	2001:503:c27::2:30
K.ROOT-SERVERS.NET	193.0.14.129	2001:7fd::1
L.ROOT-SERVERS.NET	199.7.83.42	2001:500:3::42
M.ROOT-SERVERS.NET	202.12.27.33	2001:dc3::35

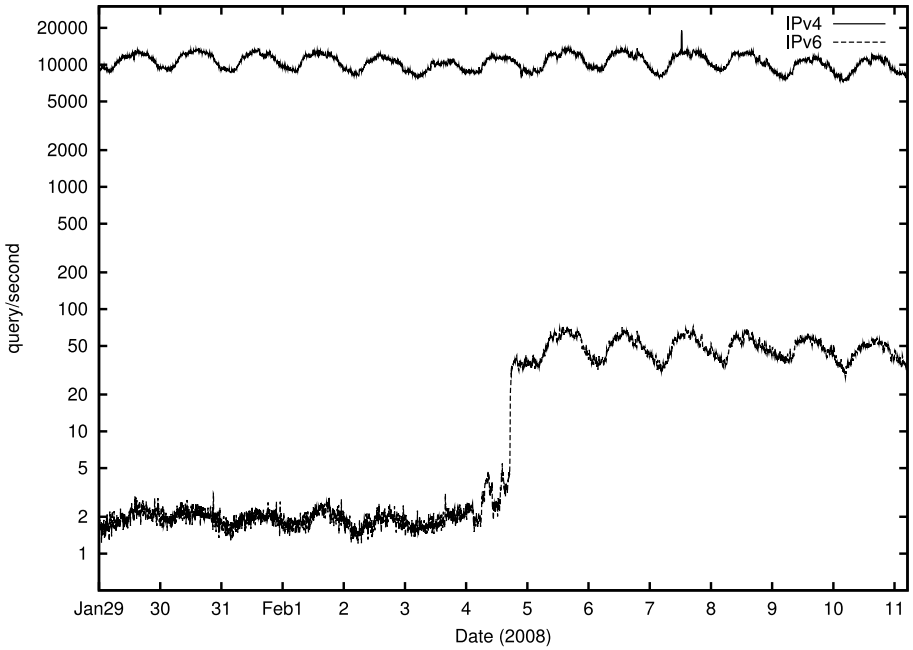


図 4.1. AAAA 追加時の IPv4 と IPv6 の問合わせ数の推移

稿された。実際にはその後 A/J が追加され、さらに 2008 年 12 月に L に対しても追加された。2008 年 12 月現在、表 4.1 に示すように、7 つの AAAA レコードが存在している。

M-Root DNS では、この決定のアナウンスを受け、それまでの実験的なサービスを運用していたサーバ群を引退させ、IPv4 を提供しているサーバで IPv6 のサービスも行うことにした。ただし、NSPIXP-6 に対応するクラスタがないため、現在まだ IPv6 サービスが始まっていない JPNAP 用のクラスタを流用することにし、また実験用クラスタのルータには十分なメモリが搭載されていなかったため、2004 年まで運用に用いられていたルータを使用することにした。さらに、San Francisco および Paris における設定変更を行い、Seoul 以外の Anycast ノードでの IPv6 サービスも開始できるように準備を進めてきた。

その結果、日本時間で 2008 年 2 月 5 日午前 2 時を数分回った頃、Root ゾーンおよび root-servers.net ゾーンに AAAA レコードが追加された。その前後一週間の M-Root DNS サーバが全体として受信している問合わせ密度の推移を図 4.1 に示す。図の日付は日本時間ではなく UTC である。2 月 5 日以前にもある

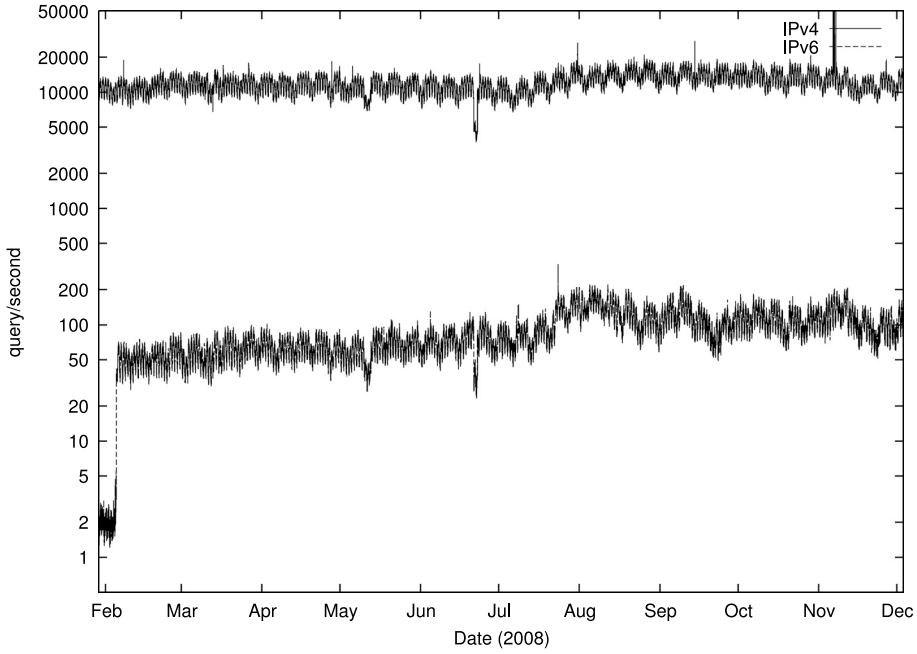


図 4.2.　IPv4 と IPv6 の問合わせ数の推移

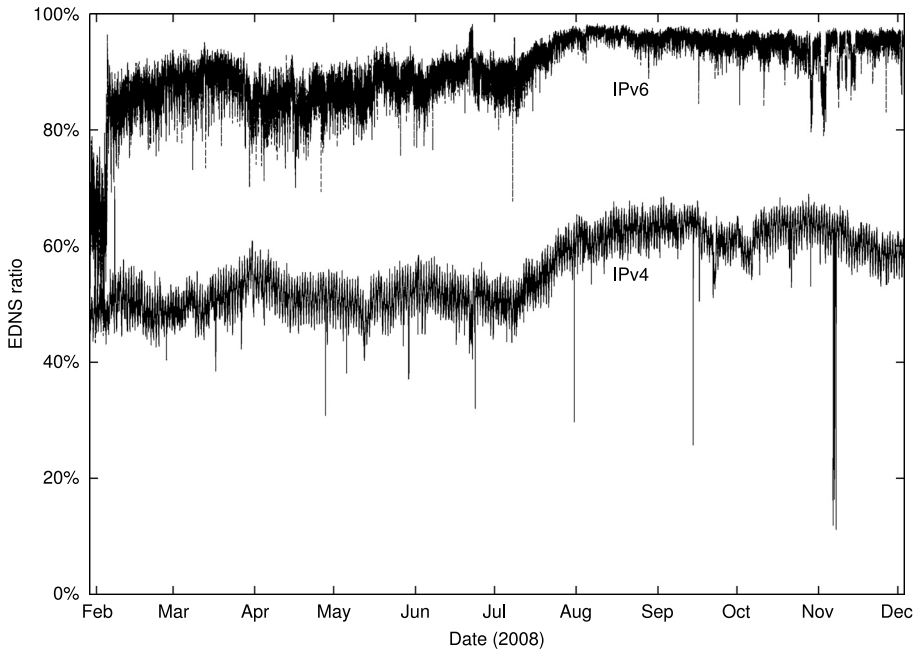


図 4.3.　EDNS0 対応問合わせ数の推移

程度の量の IPv6 による問合わせを受信していたが、これらのうちの幾分かは RIPE/NCC の DNSMON プロジェクトによる active measuremenet によるものであることは判明しているが、残りは不明である。いずれにせよ、2 月 5 日になる直前から IPv6 による問合わせが急増している。しかしながら、ピークで 50 qps を少々越える程度であり、IPv4 と比較すると 0.5%程度になっている。この数字に関しては、現在の IPv6 の展開の状況を示唆しているものと考えられ、M 以外の Root DNS サーバについても、多くても 1%程度に留まっている。

その後 IPv6 による問い合わせは徐々に増加する傾向にある。2008 年 2 月から 2008 年 12 月初旬までの M-Root DNS 全体における IPv4 による DNS

問い合わせとIPv6によるDNS問い合わせの割合を図4.2に示す。グラフ上部の曲線がIPv4による問い合わせ、グラフ下部の曲線がIPv6による問い合わせを示す。

7月から8月にかけて、IPv6による問い合わせが階段状に増加している点が見られる。これは、Kaminsky Attackによる脆弱性注意喚起によって、多くのDNS管理者がソフトウェアの更新を行ったため、IPv6対応になったと考えられる。

また、同様に2008年2月から2008年12月初旬におけるIPv4とIPv6による問い合わせのうち、EDNS0に対応した問い合わせの割合を図4.3に示す。

2008年2月の時点で、Root DNSサーバに対してAAAAレコードが追加されたことにより、IPv6による問い合わせが増加し、IPv6問い合わせにおけるEDNS0対応問い合わせの割合が一気に増加した。これは、AAAA追加前によせられていたIPv6問い合わせはほとんどが計測によるものであり、そのためEDNS0非対応な問い合わせがよせられていたと考えられる。AAAAレコードの追加により、通常のDNSサーバからの問い合わせが増えたため、EDNS0対応比率も増加したと考えられる。

また、8月頃にIPv4、IPv6問い合わせともにEDNS0対応率が増加している。これもやはりKaminsky Attackの影響でDNSソフトウェアが更新され、EDNS0対応になったためと考えられる。

第5章　まとめ

M Root DNSサーバは、11年以上に渡り安定的にサービスを提供してきた。特に多階層の冗長構成の導入により、サービスの停止を伴わずにサーバやサーバソフトウェアの保守作業が可能になったことは、サービス停止を伴う保守作業は72時間前に他のRoot DNSサーバオペレータに連絡することが要請されていることを考えると、運用面で大きなメリットがある。また、数多くのISPやIXの協力により、サーバそのものの安定運用に留まらず、インターネットの広い範囲に対して安定なサービスを提供できたことも特筆すべきである。また、Root DNSサーバのIPv6によるサービスも始まり、IPv6の展開に新たなtriggerになったと言うことができる。

現在、M-Root DNSサーバはWIDEプロジェクトの監督責任のもと、JPRSと共同で管理運用が行われているが、今後とも各方面と協力の上、より安定なサービス提供に努めていきたい。