

M Root DNS サーバの運用

2011 年報告書

加藤 朗 (kato@wide.ad.jp) / 関谷 勇司 (sekiya@wide.ad.jp)

1 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、名前空間は 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い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。

2 M Root DNS サーバの構成

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 Project 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。

M Root DNS は、図 1 に示す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、問い合わせは他の 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。

2009 年はインターネットにおける経路表の増大に伴い、一部のルータでメモリ不足が発生した。2009 年 12 月の時点で、インターネット全体の経路数は 30 万経路を突破し、M Root DNS サーバと IX を接続する一部のルータにメモリ不足によるパフォーマンスの劣化が発生した。そのため、急遽これらのルータを異なる機種に交換、もしくはマネージメントカードの交換を行うことで、経路表の増大に対応した。

2010 年も引き続きこの傾向が見られ、それに対応するために一部拠点にてルータの変更を行った。また、Root zone の DNSSEC 署名に対応するための性能向上と、機材の老朽化による障害を未然に防ぐため、Paris 拠点にて機器更新を行った。サーバとルータ、スイッチを含む全ての機材を更新し、海外保守体制を充実させた。この機材更新は 3 月 3 日から 6 日にかけて行われた。この機材更新にあわせて、サーバの OS バージョンも更新した。その際、TCP による DNS クエリ処理に一部の不具合が見つかり、DNS サーバソフトウェアを bind から nsd に一時的に変更する

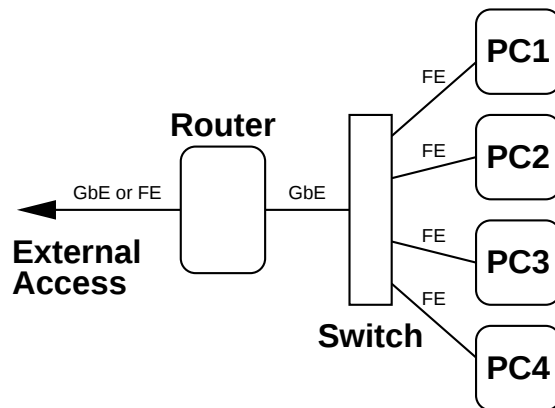


図 1: Anycast 用基本構成

ことによって対処を行った。その後、この不具合の原因は特定され、DNS サーバソフトウェアは nsd から bind に戻された。

また、2010 年は Root zone の DNSSEC による署名が開始された。これにともない、M Root DNS サーバが提供する zone も DNSSEC によって署名された root zone に変更された。また、一部拠点において、M Root DNS サーバ機器の更新が行われた。これは、DNSSEC にて署名された zone を提供するための性能向上と、故障を未然に防ぐための予防交換を目的として行われた。

2011 年の主な出来事は、2011 年 6 月に D Root DNS サーバが IPv6 での運用を開始し、d.root-servers.net に IPv6 アドレスが登録された。また、Root DNS サーバへの、正常な問い合わせとは考えにくい名前問い合わせクエリの突発的な増加が、ほぼ全ての Root DNS サーバにて数回観測されたため、その分析結果を RIPE が [1] に公開した。

M Root DNS サーバとしては、2010 年に引き続き機材の更新を進めた。東京近郊に存在する 3 つのクラスタのうち、2 つのクラスタは機材更新が完了し、新機材にて運用を開始している。また、JPRS との共同運用体制をより推し進め、Root DNS サーバの運用者会議や、機材更新、日々の運用管理に関しても、WIDE と JPRS が合同にて行なっていく体制づくりを進めた。

3 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 の経路選択ポリシーに依存するものの、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” [2] 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 (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。

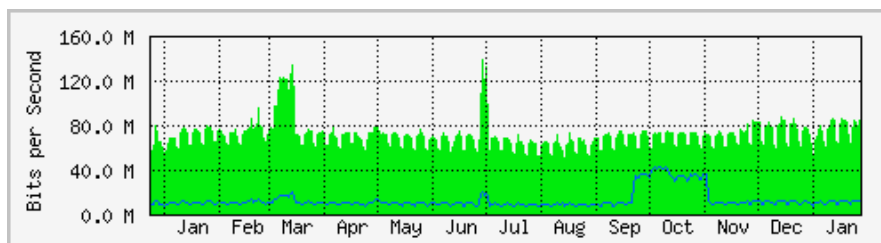
経路広告に BGP の NO_EXPORT 属性を添付するいわゆる local anycast として運用を行なっているが、学術系のネットワークの収容を目的として NCA — National Computerization Agency — が運用している Layer-3 IX である KIX では、NO_EXPORT を外して学術系ネットワークに対して経路の広報を行なっている。しかし、韓国での主要二大 ISP である KT および Daemon への接続性がないため、現在、Seoul で処理されている問い合わせは毎秒 100~200 クエリ程度である。

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 から独立の回線として transit 提供を受けている。また、2010 年 10 月に Equinix Paris に加入し、複数の ISP と経路交換を開始した。

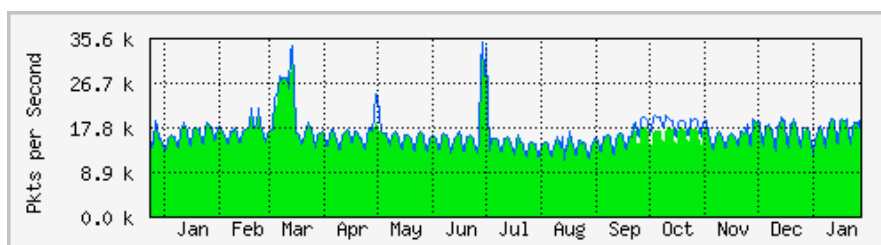
現在は多くの ISP に対して NO_EXPORT をつけて経路広告を行なっているが、幾つかの ISP に対しては NO_EXPORT なしに経路広告をしている。ヨーロッパ全域にサービスを提供している transit ISP とも多く peer しているため、そのサービスエリアはフランスに留まっていない。このため、毎秒 3000~7000 クエリ程度の問い合わせがある。

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 Los Angeles NOC 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 で処理されていた。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 サーバへの問い合わせは 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。

しかし、2010 年 10 月の WIDE Los Angeles NOC 撤退にともない、WIDE San Francisco NOC が日本との直接の回線を持たない独立 NOC として存在するようになったため、M Root DNS San Francisco も AS2500 と直接の接続性を持たない独立拠点として運用を開始した。



(a) トラフィックの推移



(b) パケット数の推移

図 2: 2011 年における M-Root DNS 全体の問い合わせ数の推移

図 2 に M-Root 全体に対するトラフィックの 2011 年における推移を示す。2010 年には、DNSSEC

導入に伴うトラフィック量の変化が見て取れたが、2011 年は大きな変動も無く、トラフィック量ならびにクエリ量とも安定した推移を示している。

4 他の Root DNS サーバ

2011 年 12 月時点での、Root DNS サーバの設置状況を表 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 Root DNS サーバでは 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 Root DNS サーバでは、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 Root DNS サーバは原則として、経路情報に NO_EXPORT BGP 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。その一方で J Root DNS サーバや L Root DNS サーバは、ほぼ全てのノードでグローバルな経路広告を行なっている。このあたりに、各 Root DNS サーバ管理組織毎の運用ポリシーの違いが見て取れる。

2009 年、2010 年から比較すると、2011 年は Anycast ノードの増加もそれほどなく、各拠点とも DNSSEC のさらなる普及に向けてのノードの増強に注力したものである。M Root DNS サーバも、拠点数を増やすことよりも、現存の拠点の機材増強を中心とした運用を行った。

5 DNSSEC 導入後の状況

2010 年に Root zone に DNSSEC が導入され、A Root DNS サーバから M Root DNS サーバまでの全ての Root DNS サーバに、署名済みの Root zone が導入された。その後、2011 年は特に DNSSEC に関して大きな動きは無く、各 Root DNS サーバとも安定運用を行った。

Root zone の Trust Anchor は <https://www.iana.org/dnssec/> から取得できるようになっており、Root zone を署名する KSK (Key Signing Key) を更新するための KSK Ceremony も 2011 年中に 4 回開催された。

6 まとめ

M Root DNS サーバは、14 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、ならびに JPRS との共同運用開始によって、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。

M Root DNS サーバでは、WIDE Project の監督責任のもと、JPRS と共同で管理運用を行い、より安定した運用と保守体制、ならびに機材更新を行っていく所存である。

参考文献

- [1] Wolfgang Nagele and RIPE NCC. Analysis of increased query load on root name servers, 2011. <https://labs.ripe.net/Members/wnagele/analysis-of-increased-query-load-on-root-name-servers>.
- [2] T. Hardie. RFC 3258: Distributing Authoritative Name Servers via Shared Unicast Addresses, April 2002. Status: INFORMATIONAL.

表 1: Root DNS サーバの設置状況

サーバ	設置都市		
A	Los Angeles, CA	New York, NY	Frankfurt (DE)
	Hong Kong (HK)	Palo Alto, CA	Ashburn, VA
B	Marina Del Rey, CA		
C	Herndon, VA	Los Angeles, CA	New York, NY
	Chicago, IL	Frankfurt (DE)	Madrid (ES)
D	College Park, MD		
E	Mountain View, CA		
F	Ottawa (CA)	Palo Alto, CA	San Jose, CA
	New York, NY	San Francisco, CA	Madrid (ES)
	Hong Kong (HK)	Los Angeles, CA	Rome (IT)
	Auckland (NZ)	Sao Paulo (BR)	Beijing (CN)
	Seoul (KR)	Moscow (RU)	Taipei (TW)
	Dubai (AE)	Paris (FR)	Singapore (SG)
	Brisbane (AU)	Toronto (CA)	Monterrey (MX)
	Lisbon (PT)	Johannesburg (ZA)	Tel Aviv (IL)
	Jakarta (ID)	Munich (DE)	Osaka (JP)
	Prague (CZ)	Amsterdam (NL)	Barcelona (ES)
	Nairobi (KE)	Chennai (IN)	London (UK)
	Santiago de Chile (CL)	Dhaka (BD)	Karachi (PK)
	Torino (IT)	Chicago, IL	Buenos Aires (AR)
	Caracas (VE)	Oslo (NO)	Panama (PA)
	Quito (EC)	Kuala Lumpur (MY)	Suva (Fiji)
	Cairo (EG)	Atlanta, GA	Podgorica (ME)
	Maarten (AN)		
G	Colombus, OH	San Antonio, TX	Honolulu, HI
	Fussa (JP)	Stuttgart-Vaihingen (DE)	Naples (IT)
H	Aberdeen, MD		
I	Stockholm (SE)	Helsinki (FI)	Milan (IT)
	London (UK)	Geneva (CH)	Amsterdam (NL)
	Oslo (NO)	Bangkok (TH)	Hong Kong (HK)
	Brussels (BE)	Frankfurt (DE)	Ankara (TR)
	Bucharest (RO)	Chicago, IL	Washington, DC
	Tokyo (JP)	Kuala Lumpur (MY)	Palo Alto, CA
	Jakarta (ID)	Wellington (NZ)	Johannesburg (ZA)
	Perth (AU)	San Francisco, CA	Singapore (SG)
	Miami, FL	Ashburn, VA	Mumbai (IN)
	Beijing (CN)	Manila (PH)	Doha (QA)
	Colombo (LK)	Vienna (AT)	Paris (FR)
	Taipei (TW)	Porto Alegre (BR)	Yerevan (AM)
	Thimphu (BT)		
J	Dulles, VA (3 sites)	Ashburn, VA	Miami, FL
	Atlanta, GA	Seattle, WA	Chicago, IL
	New York, NY	Honolulu, HI	Mountain View, CA (3 sites)
	San Francisco, CA (2 sites)	Dallas, TX	Amsterdam (NL)
	London (UK)	Stockholm (SE) (2 sites)	Tokyo (JP)
	Seoul (KR)	Beijing (CN)	Singapore (SG)
	Dublin (IE)	Kaunas (LT)	Nairobi (KE)
	Montreal (CA)	Perth (AU)	Sydney (AU)
	Cairo (EG)	Cairo (EG)	Warsaw (PL) (2 sites)
	Brasilia (BR)	Sao Paulo (BR)	Sofia (BG)
	Prague (CZ)	Johannesburg (ZA)	Toronto (CA)
	Buenos Aires (AR)	Madrid (ES)	Fribourg (CH)
	Hong Kong (HK) (2 sites)	Turin (IT)	Mumbai (IN)
	Oslo (NO)	Brussels (BE)	Paris (FR) (2 sites)
	Helsinki (FI)	Frankfurt (DE)	Riga (LV)
	Milan (IT)	Rome (IT)	Lisbon (PT)
	San Juan (PR)	Edinburgh (UK)	Tallin (EE)
	Taipei (TW)	New York, NY	Palo Alto, CA
	Anchorage, US	Moscow (RU)	Manila (PH)
	Kuala Lumpur (MY)	Luxembourg City (LU)	Guam, US
	Vancouver (CA)	Wellington (NZ)	
K	London (UK)	Amsterdam (NL)	Frankfurt (DE)
	Athens (GR)	Doha (QA)	Milan (IT)
	Reykjavik (IS)	Helsinki (FI)	Geneva (CH)
	Poznan (PL)	Budapest (HU)	Abu Dhabi (AE)
	Tokyo (JP)	Brisbane (AU)	Miami, FL
	Delhi (IN)	Novosibirsk (RU)	Dar es Salaam (TZ)
L	Amsterdam (NL)	Atlanta, GA	Auckland (NZ)
	Bogota (CO)	Brisbane (AU)	Brussels (BE)
	Buenos Aires (AR)	Cairo (EG)	Cape Town (ZA)
	Chicago, IL	Crete (GR)	Copenhagen (DK)
	Culpeper (VA)	Dammam (SA)	Denver, CO
	Dhaka (BD)	Dakar (SN)	El Segundo, CA
	Guam, GU	Istanbul (TR)	Jeddah (SA)
	Johannesburg (ZA)	Kathmandu (NP)	Kharkov (UA)
	Kiev (UA)	Lima (PE)	Los Angeles, CA
	Lyon (FR)	Luxembourg City (LU)	Manama (BH)
	Manila (PH)	Malmö (SE)	Maputo (MZ)
	Marseille (FR)	Melbourne (AU)	Miami, FL
	Nairobi (KE)	Paris (FR)	Perth (AU)
	Philadelphia, PA	Prague (CZ)	Reston, VA
	Riyadh (SA)	Saint-Denis (RE)	San Jose, CA
	Oslo (NO)	Sao Paulo (BR)	Santiago de Chile (CL)
	Stockholm (SE)	Sydney (AU)	Toronto (CA)
	Warsaw (PL)		
M	Tokyo (JP)	Seoul (KR)	Paris (FR)
	San Francisco, CA		