第21部

M Root DNS サーバの運用

加藤 朗、関谷 勇司

第１章　M-Root DNSサーバ基盤の概要

M-Root DNS Serverは1997年8月に運用を開始してから20年間、大きなトラブルなく運用を継続している。2017年はM-Root DNS Server運用開始から20周年であり、記念のステッカー等を作成した。

運用開始当初は、Cisco 4500をルータとし、PentiumPro 200MHzをCPUとして搭載したPCを２台用いて、プライマリ/バックアップ体制にて運用を開始した。2002年に東京拠点で\Anycast in a Rack"として、Anycastの予備的な運用を開始した。これはラック内においてAnycastを行うことでサービス拠点内における耐障害性を向上させる技術である。2004年までに、Seoul拠点、San Francisco拠点、Paris拠点にて"Anycast in a Rack"の運用を開始した。

運用開始当初は202.12.27.33というアドレスを使用し、IPv4のみのサービスを提供していたが、2008年2月からは2001:dc3::35というアドレスを用いたIPv6トランスポートによるサービスを開始した。この際、13組織によって運用されているRoot DNSサーバのうち、4組織がIPv6トランスポートによる運用を開始しており、M-Root DNSはその中の1組織として、いち早くIPv6トランスポートによるサービスを開始した。ただし、Seoul拠点を除く他の拠点での開始となった。その後2013年にM-Root DNS全拠点においてIPv6トランスポートによるサービスを開始した。

また、当初はWIDE Project単独によって管理運用されていたが、2005年12月から日本レジストリサービス（JPRS）との共同運用となった。

各拠点の概要を次に述べる。

東京拠点

東京拠点では、2004年から、Anycastを用いてDIX-IE、JPIX、JPNAPのそれぞれに別々なクラスタを割り当てて運用しており、他のサービスとはネットワークやサーバを共有せず、独立した設備にてサービスを提供している。

大阪拠点

2002年から運用を始めた大阪拠点は、当初は東京のバックアップとして、東京拠点がダウンした場合のバックアップとしての運用であった。これは、BGPでAS-PATHを複数個prependすることにより実現していた。しかし、これだけではトラフィックは皆無で、BGP peeringの保守や障害検出にも問題があるため、同じルータを利用して別なサーバを設置し、RFC1918 (いわゆるプライベートアドレス) 等の逆引きに対してMXDOMAINエラーを返すAS112サービスの運用も行っている。なお、2013年からはAS-Path prependを廃し、大阪拠点におおいても定常的なサービスを提供している。

Seoul拠点

Seoul拠点は、2004年7月に韓国の商用IXであるKINXに接続する形態で運用を開始した。韓国国内のISP事情により、規模の大きなISPとはIXにおいてpeeringすることが難しいため、トラフィック流量は多拠点に比べると少ない。しかし、M-Root DNS海外拠点における最初のAnycastの運用拠点として、2004年3月にサーバ基盤を設置し運用経験を積んできた。Seoul拠点では国際トランジットを提供してもらっていないため、主に韓国国内に対するサービスを提供している。しかし、KINXに接続する東南アジアのISPも存在するため、必ずしも韓国国内限定というわけではない。また当初はIPv4のみの運用で

あったが、2013年にはIPv6のサービスを提供している。

Paris拠点

Paris拠点は、2004年9月に運用を開始した。当初はデータセンター内部にスイッチを持っていた、フランスの研究開発ネットワークRenaterが運用するSFINXと、France Telecomが運用するPariXの２つのIXに接続していた。その後、接続するIXも徐々に増え、Equinix Parisを始め、NL-IXとFrance-IXが加わり、現在では５つに増加している。サービス範囲は主にヨーロッパ諸国が中心となっているが、アフリカの都市からのアクセスも少なくない。各Root DNSサーバへのRTTの調査では、最もRTTが短いのがM-Root (Paris) という都市もあることがしばしば報告されている。

San Francisco拠点

San Francisco拠点は、従来から存在したWIDE San Francisco NOC（現在は廃止）に併設する形態で2004年10月から運用を開始している。この拠点では、IPv6に特化したISPからIPv6のtransitの提供を受けているため、全体の平均としては6%前後であるIPv6トラフィックの比率が、本拠点では25%前後と大きくなっている。また、2017年12月にSan Francisco拠点は従来のデータセンターから新たなデータセンターに移設された。詳細に関しては後述する。

Orly拠点

Orly拠点は、Paris拠点に続くフランス国内2番目の拠点として2016年10月に機材が設置され、2016年11月からサービスが開始された。Paris拠点とともに、M-Root DNSを構成するコア拠点となっている。

San Jose拠点

San Jose拠点は、2017年12月に機材が設置され、運用が開始された。詳細に関しては後述する。

第２章　2017年のトラフィック傾向

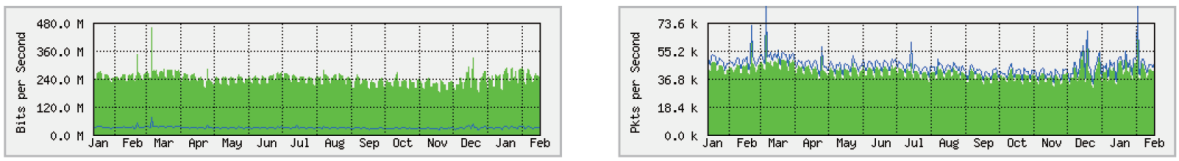
図1にM-Root全体に対するトラフィックの2017年における推移を示す。2017年も前年に比べて大きな変動も無く、トラフィック量ならびにクエリ量とも安定した推移を示している。パケット数のグラフにおいていくつか突発的な増加が観測されているが、短時間における散発的な攻撃が観測されたものと思われる。

第３章　2017年の変更点

本節では、2017年における主な変更点や発生したイベントに関してまとめる。

3.1　KSK Roll Over

Root ZoneをDNSSECで署名して保護する体制は、2009年後半に整備が進められ、2010年前半にDURZ - Deliberately Unvalidateble Root Zone - として、その妥当性がフィールドでチェックされながら進められた。これは、適当なタイミングで、A ～ Mの13あるRootサーバに対して、最初は署名なしのZoneであり、２～３週の間隔を空けて、署名付きのZoneに基づいて動作するletterを徐々に増やしていき、そのタイミングでpacketの収集解析によって不具合が発生していないことを確認する、というものである。署名付きzoneは、公開しない - 従って検証はできない - 鍵によって生成されたもので、DNSSEC情報によるDNSサーバの挙動あるいはパケットが大きく



(a) トラフィックの推移　　(b) パケット数の推移

図1　2017年におけるM-Root DNS全体の問合わせ数の推移

なることに起因する問題をチェックする、ということで、その結果、2010年7月15日に、別のKSK (Key Signing Key) で署名したzoneに切り替え、そのKSKの公開鍵を公表することができた。これによって、KSKの公開鍵を設定すれば、DNSSECによるDNS情報の検証が可能になった[1]。

それ以来、Root ZoneのZSKは年に4回、アメリカ合衆国の東海岸と西海岸で交互に、証人の目の前で生成されるKey Signing Ceremonyで生成されており、3ヵ月に一回交換されてきた。しかし、KSKは2010年に生成されたものが引き続き使われてきた。これは、KSKはZoneの鍵情報を保持するDNSKEY RRset (同一の名前で同じClass、Typeのレコードの集合)の署名のみに使われたこと、署名の頻度も一日に2回程度と少ないことによる。しかし、同一の鍵を永続的に使えるわけではなく、当初は7年程度で更新する、ということになっていた。

2017年7月に新しいKSKが公開され、2017年末には古い当初の鍵は退役するはずであったが、2017年9月時点で約4%のfull service resolverは古い鍵しか保持していないことが観測されており[2]、KSK Rolloverの作業は一旦中断された。

RFC5011に基づいた実装であれば、新しいKSKを、古いKSKに基づいて自動的に導入することができるため、full service resolverの運用者のマニュアル操作は不要である。しかし、RFC5011に対応していないソフトウェアを使用していたり、意図的かどうかは分からないが、RFC5011の機能を停止していたりした場合、新しい鍵は手作業でのインストールが必要になる。

古い鍵のみを保持している場合、その鍵が署名に使われなくなった時点で、DNSSECの検証が不能になり、DNSの名前解決ができなくなる。もっとも、DNSSECの検証をしない設定にしている場合にはこの限りではないが。どの鍵が保持されているかどうかは従来のDNSSECプロトコルでは報告することはできなかったが、2017年4月に出版されたRFC8145 - Signaling Trust Anchor Knowledge in DNS Security Extensions (DNSSEC) - では、保持している鍵を報告する機能が定義され、それに基づいた実装も行われてきた。これに対応して、Root Server側でもこの報告情報を収集し、それをVerisignに送る仕組みが実装されている。

現在のところ、ICANNでは2018年10月にKSK Rolloverを計画しており、パブリックコメントを含めた調整作業が行われている。

なお、KSK更新作業そのものに関して、Rootサーバのオペレータは Zoneの中身は関与しないので、特段の作業は発生しない。ただし、上記の計測作業や、期間中のパケット長の増加によるTCPによる問い合わせの増加などの影響は考えられるし、更新作業終了後に古いKSKしか保持していないfull service resolverからの平常時以上のトラフィックの発生も懸念されるところである。

3.2 San Francisco拠点の移設

San Francisco拠点は、WIDE ProjectのSan Francisco拠点があった200 Paulデータセンタをベースに2004年に設置されたものである。当初の1/2ラックではM-Rootの運用に必要な機器を収容することはできなかったが、幸いその下段が空いていたため、大きな問題なく導入することができ、翌2005年からM-Rootの3番めの海外拠点としての運用を開始した。M-Rootは当時Switch and Dataが運用していたPalo Alto Internet Exchange (PAIX) のスイッチがあったため、3Fから5FにUTPで接続し、サービスしていた。後にスペースの関係で、同データセンターの、現在はUnitedLayerが運営しているスペースに移設している。San Francisco拠点は、WIDE Project以外にF-Rootも運用しているInternet Systems Consortium (ISC) 経由でもアクセスすることができた。

この間に、当時のデータセンターの運営はDigital Realtyに移管され、またPAIXもEquinixに買収されるなどの変遷があった。また、データセンターのラックの契約には、機

＊1 http://root-dnssec.org/
＊2 https://www.icann.org/news/blog/update-on-the-root-ksk-rollover-project

器および作業中のオペレータに対する保険への加入が必須になってきた。しかしながら、U.S.に法人格がない団体が保険契約をすることは必ずしも容易ではない。そのため、F-Rootも運用しているISCが提供していたHosted@ISCプログラムにより、保険契約のみならず、ラックの費用やPAIXのポート費などを含めたone stop shoppingによってサービスを継続することができた。

Hosted@ISCは小規模なサービスを簡便に開始するのに適していたため、いくつかのnon-profitな団体が活用していたが、その数も減ってきたため、2017年に入って、ISCは同プログラムを終了することを決めた。そのため、現在利用しているサービスを提供する会社に個別で契約を引き継ぐか、別なデータセンターに引っ越すかの選択を迫られることになった。200 Paulは、San Francisco International Airportから車で15分程と近いため、立地は悪くないが、近年は館内のファイバの維持費が高額であること、保険の問題を解決できる見込みがなかったことから、移転することにした。

アメリカ合衆国、特にカリフォルニア州ではデータセンターの需要は旺盛で、ラックの価格も急騰していた。さらに、空きラックがない、あるいはスペースはあっても電源容量の余裕がないなどの問題で、候補探しは簡単ではなかった。

2017年10月にカリフォルニア州内のデータセンターを２箇所下見し、Hurricane Electricの運用するFremont2データセンターが選択された。Fremont2では、過去に発電機が故障したため停電時に一部に給電が止まるなどの事故もあったが、A+B構成で同時に障害が発生したことはないこと、また、館内の相互接続が容易で、館内にSFMIXおよびAMSIX-Bayという二つのIXが存在すること、ラック費用やおよび保険が必須ではない点が決定に役立った。

当初はSan Francisco拠点は重要であることから、もう一クラスタを追加することで準備が進んでいたが、納期の関係から一度に作業することは断念し、12月18日に、200 Paulの機材を全てshutdownし、Fremont2に移設した。IPアドレスの変更や、それに伴うACLの変更なども必要であったが、翌12月19日の夕刻（現地時間）にはサービスを復活することができた。なお、M-Root San Francisco拠点に付随して運用されていたE.DNS.JPのSan Francisco拠点も同時に移設されている。

その後、もう一拠点分の機材も概ね配送されたため、2018年1月22日の週に作業を行った。1月22日の東京圏での降雪による飛行機の遅延や、納入された機器の障害による交換保守作業、作業員の体調不良などの問題はあったが、現地時間の1月25日の夕刻、２クラスタ体制での運用を開始することができた。新たなクラスタはSan Jose拠点と命名された。

最初のクラスタはHurricane Electricにサービスのtransitの提供を受け、またFremont2にスイッチがあるSFMIXに接続予定で、現段階では書類作業中である。一方、新しいクラスタはLevel3からtransitの提供を受け、またAMSIX-Bayにも接続されており、従来に比べて安定したサービスの提供が可能になった。

第４章　今後の展開

2017年は目立ったDDoS攻撃は観測されていないが、インターネットの接続基盤が強化されている現在、その懸念まで払拭されたわけではない。そのため、一層のサービス容量の強化およびAnycast拠点の増強が必要である。

アジア太平洋地域のRegional Internet RegistryであるAPNICでは、Rootサーバの地域内への展開をサポートする活動も行っている。基本的には現地のデータセンターやISPとの橋渡しであるが、予算によってはAPNICの機器で運用することも可能とのこと（この場合でもAPNICは運用には携わらない）であり、このような制度も活用しつつ、footprintの拡大を検討していきたい。

第５章　謝辞

機材更新に関して、機材の提供を頂いたM-Rootの共同運用をしている日本レジストリサービスおよび同社のの

関係技術者諸氏に感謝します。M-Rootの運用は、東京拠点ではDIX-IE、JPIX、JPNAP、大阪拠点ではNSPIXP3、JPNAP-Osaka、JPIX-Osakaの、Seoul拠点ではKINXおよびKREONET2、Paris拠点ではTelehouse Paris、Renater、SFINX、PariX、FranceTelecom、Equinix Paris、NL-IX/OpenTransit、France-IX、TINET、Nerim、およびAFNIC、San Francisco拠点ではHurricane Electric、SFMIX、Level 3、AMS-IX BA等この他数多くのISPや諸氏のご協力のもと運用されており、これらの団体および関係者に感謝します。