ccTLD 及び ccSLD 権威サーバの運用に関する報告

大谷亘 alt@sfc.wide.ad.jp
ルーク，コリー koluke@sfc.wide.ad.jp
豊田安信 yas-nyan@sfc.wide.ad.jp
近藤賢郎 latte@itc.keio.ac.jp
島慶一 keiichi@iijlab.net
石原知洋 sho@c.u-tokyo.ac.jp

Dec. 2021

1 概要

WIDE Project では 2021 年度より以下の DNS ゾーンの権威サーバを運用している．

- lb.
- com.lb.
- edu.lb.
- gov.lb.
- net.lb.
- org.lb.

本稿では，これらの DNS 権威サーバの運用にかかる経緯，設計，運用について記述する．

2 背景

各ゾーンはレバノン共和国に割り当てられた ccTLD とそのサブドメイン (ccSLD) のゾーンであり，ドメインレジストリは LBDR によって運用されている [1]．権威サーバ群はベイルート・アメリカン大学をはじめとする学術組織及びコミュニティによって運用されており，WIDE Project では 2021 年 8 月より権威サーバとして ns-jp.lbdr.org.lb. の運用を開始した．

ccTLD 権威サーバを割り当て国以外で運用することにより，地政学的観点から可用性の高い ccTLD 運用を実現した．また，アジア太平洋地域のフルサービスリゾルバ・エンドユーザに対し，より低遅延な名前解決サービスを提供している．

設計から運用にあたり，TWO ワーキンググループ内で m-root の運用者や 20 代の若手研究者を含むサブグループを組織した．これまでの経験や root DNS ゾーンの運用知見に基づく効率的な運用を行うだけでなく，情報交換や本運用における実務経験を通して高度な DNS オペレーションを行うことのできる若手人材の育成にも貢献している．

3 設計

運用開始にあたり，必要なシステム設計を行った．

3.1 サーバ

サーバ設計においては，可用性を重視し，一つのサービスアドレスに対して複数台のサーバを用意することとした．また，物理障害に対する耐障害性を確保するため，複数台の物理サーバ上に複数台の仮想マシン (ノード) を構築することとした．

3.2 DNS 委任関係

DNS 的な委任関係の設計においては，後述のネットワーク設計を考慮し，Figure 1 に示すように ns-jp.lbdr.org.lb. という単一のホスト名に対して委任することとした．このホスト名には単一の A/AAAA レコードを紐付けた．これにより，運用事故発生時に lame delegation が発生する蓋然性を低減した．

3.3 ネットワーク

ネットワーク設計においては，冗長性のあるサーバ設計を前提とし，可用性を確保する設計を行った．具体的には，NS レコードに紐付く A/AAAA レコードとして公開されるサービスアドレス 1 つを複数のノードに対して割り当て，AS 内の経路広告による動的経路制御を行うこととした．

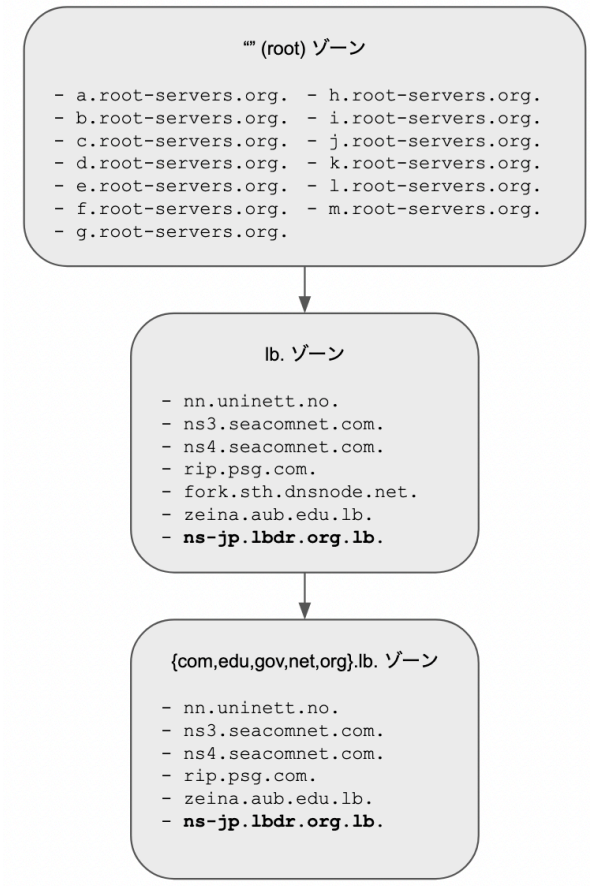

Fig. 1: 委任関係の構図

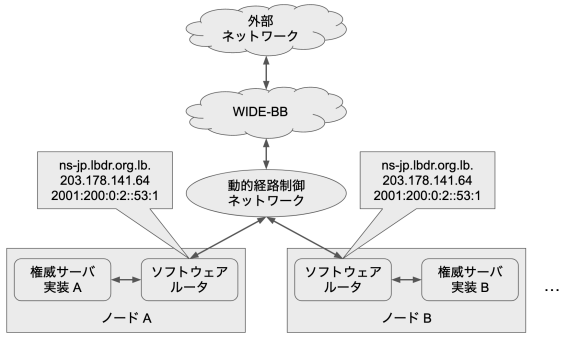

Fig. 2: L3 ロードバランサレスアーキテクチャ

また，各ノード内で権威サーバ実装とソフトウェアルータが稼働し，ノード自らが経路広告を行うことにより，Figure 2に示す L7 及び L3 の双方の観点で単一障害点を排除した (cf. [2])．これにより，状況によって柔軟にトラフィックを分配したり，特定ノードへのトラフィックを止めることで容易にメンテナンスしたりと，運用コストの低減を見込むことができる．

3.4 権威サーバ実装

権威サーバ実装はメンテナンス性や開発・運用コストを考慮し，オープンソースな既存の実装を用いることとした．また，実装ダイバーシティを確保するため，ノードにより異なる複数の実装を採用することとした．実装の選定は，大規模な権威サーバを運用する事業者での採用事例やそれぞれのコミュニティが公開しているパフォーマンス測定結果を基に実施し，比較的信頼でき ccTLD 及び ccSLD の運用に耐えうるパフォーマンスを持つものを採用した．

3.5 サービス安定性

ccTLD 及び ccSLD のゾーンホストという条件につき，セキュリティ的に安定したサービス稼働が求められる．そのため，基本的にノードへの手動での設定変更等は行わない方針とし，人為的ミスによる運用事故を減らすことを目的に運用自動化ツールである Ansible [3] を導入することとした．これにより，平時のオペレーションで必要な設定変更や，障害児やメンテナンス時のトラフィック制御に必要な設定変更を安全に行うことを見込む．

また，物理セキュリティについては，各物理マシンの設置場所を物理アクセス制限システムの整った国内データセンタから選定した．また，具体的な設置場所・データセンタについては非公開とした．

`ns-jp.lbdr.org.lb.` はプライマリ権威サーバからゾーン転送を受けてゾーンホストを行うセカンダリ権威サーバとして機能する．そのため，ゾーン転送においては TSIG [4] による認証を用いることとし，サービスアドレスとは異なる IP アドレスでの転送に限定し，転送元サーバ運用者にその IP アドレスのみへの転送に限定するように依頼した．

4 運用

前項までの設計を元にサーバ・ネットワークのデプロイをすすめ，2021 年 8 月 7 日より `lb.` ゾーンを除く ccSLD のゾーンについて，同月 10 日より `lb.` ゾーンについて，親ゾーンからの委任及び権威応答を開始した．

4.1 トラフィックの分析

ネットワーク上のトラフィック監視機構で収集したデータをもとに，2021 年 12 月 6 日から 12 日までの

1 週間における 5 tuples (送信元 IP アドレス，送信元ポート番号，送信先 IP アドレス，送信先ポート番号，プロトコル) に基づくトラフィックフローの分析を行った．なお，一部の統計値を付録に示す．

4.1.1 概況

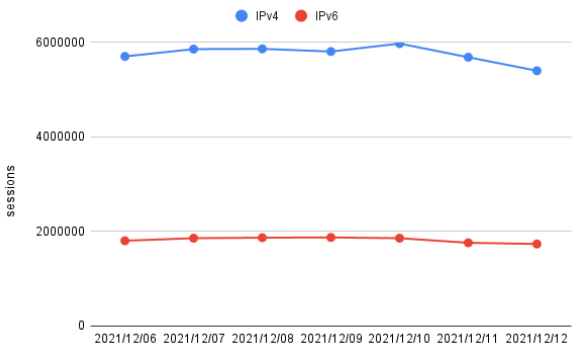

Fig. 3: クエリ数の変化

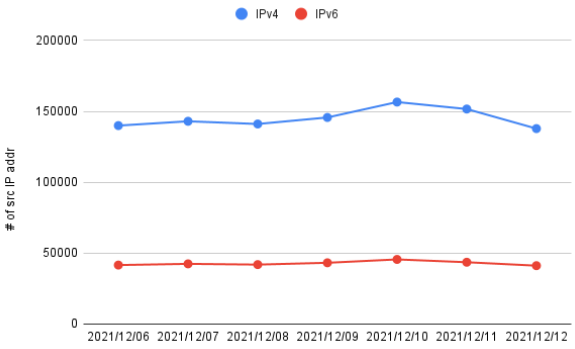

Fig. 4: クエリ元 IP アドレス数の変化

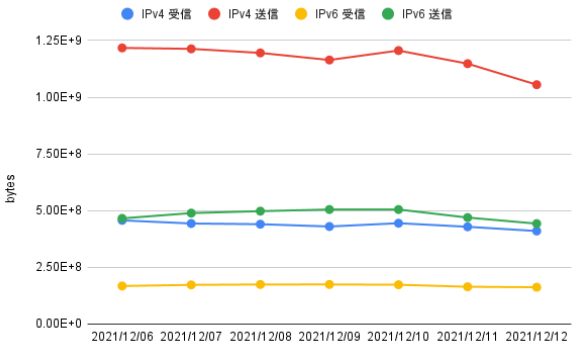

Fig. 5: 送受信バイト数の変化

Figure 3，Figure 4，Figure 5はそれぞれ権威サーバのサービスアドレスに対する DNS クエリと応答について，セッション数，クエリ元 IP アドレス数，送受信バイト数を日ごとに集計したものである．いずれも曜日による大きな変化はない．クエリ数 (Figure 3) をみると，1 日あたりの平均で IPv4 では約 576 万クエリ，IPv6 では約 182 万クエリ，あわせて約 758 万クエリを受けている．クエリ元 IP アドレス数 (Figure 4) をみると，1 日あたりの平均で IPv4 では約 14.5 万ホストから，IPv6 では約 4.28 万ホストから，あわせて約 18.9 万ホストからクエリを受けている．また，同様の仮定において，IPv4 でクエリしているホストは IPv6 でクエリしているホストよりも約 3.39 倍多く，IPv6 の普及が十分進んでいないことがわかる．送受信バイト数 (Figure 5) をみると，IPv4 と IPv6 をあわせて 1 日あたりの平均で約 578MB を受信，約 1.54GB を送信している．トラフィックの増幅については次章で述べる．

4.1.2 トラフィックの増幅率

Figure 5からもわかるとおり，DNS 権威サーバでは往々にして受信バイト数よりも送信バイト数の方が大きい．ある受信クエリと送信応答のセットがあったときに，受信バイト数に対する送信バイト数の割合を増幅率と定義し，増幅率がそれぞれ 0–5 倍，5–10 倍，10–15 倍，15–20 倍，20 倍以上のケースに分類して分析を行った．

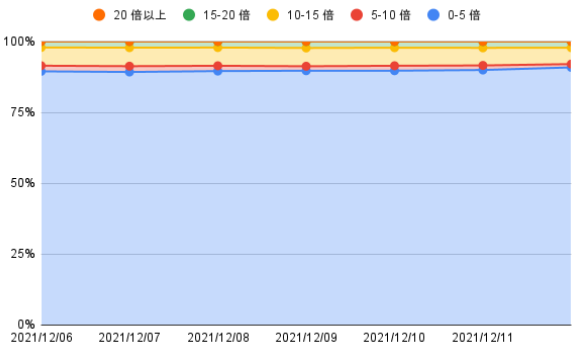

Fig. 6: 各日のクエリ数に対する増幅率ごとのクエリ数の割合 (IPv4)

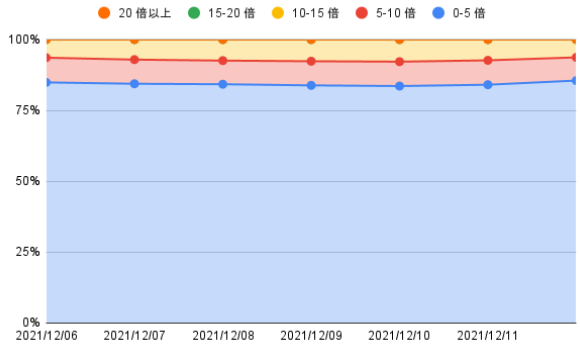

Fig. 7: 各日のクエリ数に対する増幅率ごとのクエリ数の割合 (IPv6)

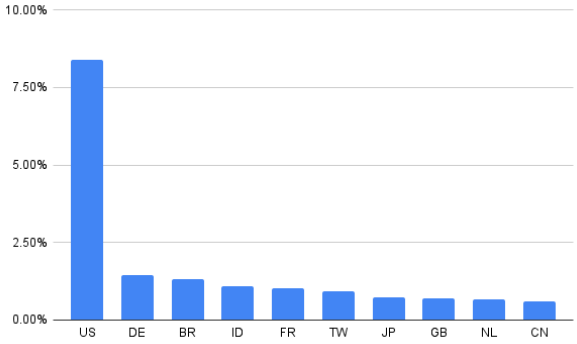

Fig. 9: 国・地域別送信元 IP アドレス数の割合 (IPv6) (上位 10 カ国)

Figure 6，Figure 7にそれぞれ IPv4 と IPv6 における，各日の合計クエリ数に対する増幅率ごとの応答数の割合を示す．どちらもクエリの大半が 0–5 倍の増幅率に収まっており，20 倍を超すクエリはほとんどないことがわかる．増幅率が 0–5 倍であるクエリ数の割合は平均して IPv4 で約 91.9%，IPv6 で約 88.1% となっており，IPv6 のほうがやや増幅率の高いクエリが多い．一方，増幅率が 15 倍以上であるクエリ数の割合を平均すると IPv4 では約 2.24%，IPv6 では約 0.00% となっており，異常に増幅率の高いクエリは IPv4 にやや多く見られる．

4.1.3 国・地域別トラフィック量

GeoIP2 [5] を使い，送信元 IP アドレスを国・地域ごとに分類した．

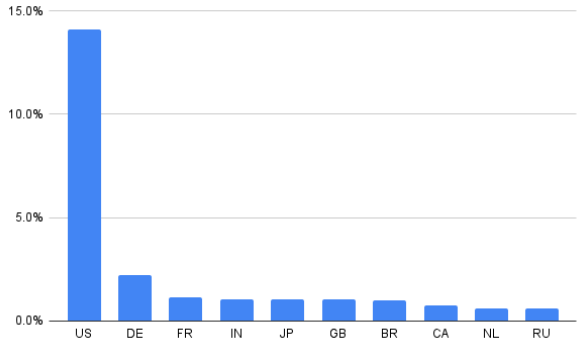

Fig. 8: 国・地域別送信元 IP アドレス数の割合 (IPv4) (上位 10 カ国)

Figure 8, Figure 9にそれぞれ IPv4 と IPv6 の国・地域別送信元 IP アドレス数の割合を示す．IPv4・IPv6 共にアメリカからのクエリが多いことがわかる．アジア太平洋地域では，インド，日本，ロシア，台湾，中国からのクエリが多いことがわかる．

4.2 監視

迅速な障害検知のため，監視システムによる継続的な監視を行っている．また，障害発生時における障害点・障害原因の迅速な把握のため，AS 内と AS 外に複数の監視点を設置している．監視点の設置場所により監視項目を分担しており，AS 外では外形監視としてサービスの死活監視や応答時間の監視などを行っており，AS 内ではゾーン転送異常検知のための SOA レコード SERIAL 値の監視やノード自体の OS レベルでの死活監視などを行っている．

4.3 障害対応

本稿執筆時点において特筆すべき障害は発生していないが，前述の監視体制による障害検知や後述する組織外からの情報提供に基づき，常に障害に対応できる体制を敷いている．また，WIDE-BB [6] 上の障害に関しては TWO ワーキンググループ内で連携して対応を行っており，重大な障害発生時には他の権威サーバ運用者を介して IANA への報告など必要な措置を行う．

4.4 セキュリティ

セキュリティ運用に関しては，WIDE Project の組織 CSIRT である WIDE Incident Response Team (WIRT) [7] と連携し，ネットワーク上に設置したト

ラフィック監視システムによる異常な通信の監視を行っている．

4.5 組織外との連携

複数組織での権威サーバ群運用にあたり，他の権威サーバ運用者とメーリングリストやオンラインミーティングを介した情報交換を行っている．

5 まとめと展望

本稿では，WIDE Project で運用を開始した ccTLD 及び ccSLD 権威サーバについて 2021 年度に行った活動を報告した．WIDE Project での権威サーバの運用は，地政学的及びグローバルな名前空間である DNS の可用性・多様性の観点から価値ある活動であると考えられる．運用開始前のフェーズにおいては，可用性や多様性を重視したサーバ・ネットワーク・セキュリティ設計を行った．運用開始後のフェーズにおいては，WIDE Project 内外の組織と連携して持続的な運用を行っている．今後は現在の運用を維持するとともに，運用しているインフラストラクチャを利用した研究活動も行いたい．

付録

トラフィック分析における統計値

Table 1: IPv4 における統計値

	合計 (/週)	平均 (/日)	標準偏差
セッション数	40,304,007	5,757,715.3	171,717.11
送信元 IP アドレス数	1,015,807	145,115.3	6,258.238
受信バイト数	3,053,498,733	436,214,104.7	13,867,233.56
送信バイト数	8,197,774,011	1,171,110,573	53,060,719.1

Table 2: IPv6 における統計値

	合計 (/週)	平均 (/日)	標準偏差
セッション数	12,764,376	1,823,482.3	52,078.282
送信元 IP アドレス数	299,459	42,779.9	1,387.76
受信バイト数	1,188,687,601	169,812,514.4	4,683,587.632
送信バイト数	3,373,971,024	481,995,860.6	21,724,408.57

References

[1] “LBDR LLC.” [Online]. Available: https://www.lbdr.org.lb/

[2] M. Prince, “Load Balancing without Load Balancers,” Mar. 2013. [Online]. Available: https://blog.cloudflare.com/cloudflares-architecture-eliminating-single-p/

[3] Ansible, Red Hat, “Ansible is Simple IT Automation.” [Online]. Available: https://www.ansible.com

[4] F. Dupont, S. Morris, P. A. Vixie, D. E. Eastlake 3rd, . Guðmundsson, and B. Wellington, “Secret Key Transaction Authentication for DNS (TSIG),” Internet Engineering Task Force, Request for Comments RFC 8945, Nov. 2020, num Pages: 22. [Online]. Available: https://datatracker.ietf.org/doc/rfc8945

[5] “GeoIP2 Databases | MaxMind.” [Online]. Available: https://www.maxmind.com/en/geoip2-databases

[6] “WIDE Backbone.” [Online]. Available: http://two.wide.ad.jp/

[7] WIDE Incident Response Team, “WIDE Incident Response Team (WIRT) - wirt.wide.ad.jp.” [Online]. Available: https://wirt.wide.ad.jp/