変化するDNS運用と
これからの課題について
(DNS設計/運用者の目線から)

Kento KAWAKAMi

LINEヤフー

発表者

川上 けんと

LINEヤフー株式会社

其田 学

株式会社インターネットイニシアティブ

小坂 良太

NTTコミュニケーションズ株式会社

太田 健也

LINEヤフー株式会社

このセッションについて

- (建前)
  - DNSは多くの環境で長期間運用されているシステムであり
    重要な基盤として動作している
  - また、現在のインターネットでは、IPアドレスを知る前に欲しい情報を
    保管する分散KVSとして重要な役割を持つようになっている
  - DNSに求められる役割がさらに大きくなっていくタイミングで
    現在のDNSに関する運用など最新の知見をまとめて、
    DNSのこれからについて議論を行いたい
- (本音)
  - JANOG53回目だし、
    DNSでお祭りっぽい事がしたい

Agenda

01	最近の各社のDNSの変化について
02	これまであった課題への対応
03	これからの課題

Agenda

01 最近の各社のDNSの変化について

02 これまであった課題への対応

03 これからの課題

About me

- 川上けんと
- LINEヤフー株式会社
- 職歴
  - 2019年~：LINE株式会社(新卒入社)
  - 2023年~：LINEヤフー株式会社
- お仕事
  - DNSの設計、運用、開発
    - 最近はDNSaaSの開発をしている時間が多め
  - LB Serviceの開発、運用
- 趣味
  - 神頼み

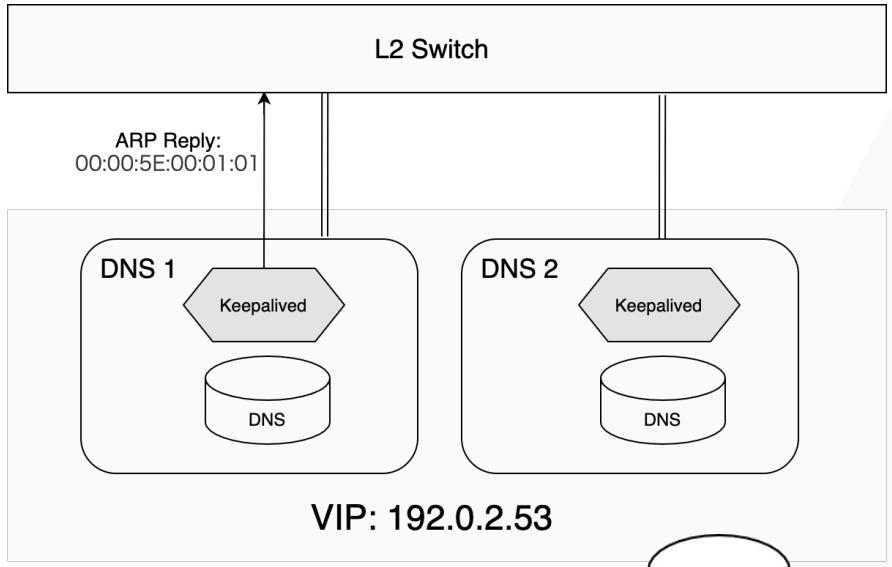
以前のDNS構成

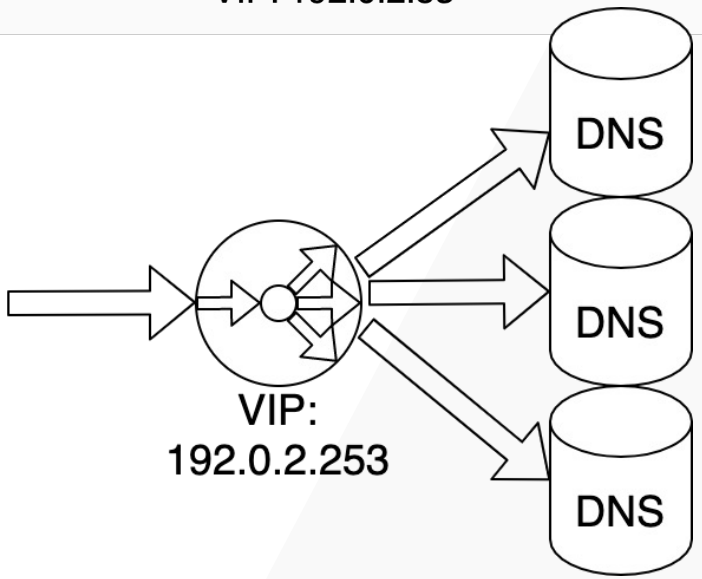
DNS Auth(VRRP)

- Act-Stb構成
  - サーバのスケールを行えない
- L2構成が必須
- Act-Stbの切り替えパケットロスが発生する可能性がある

DNS Cache(HWLB)

- サーバ単体でDNSのメンテナンスが難しい場合がある
- VIPのライフサイクルがHWLBに依存する
  - EoL
  - EoS



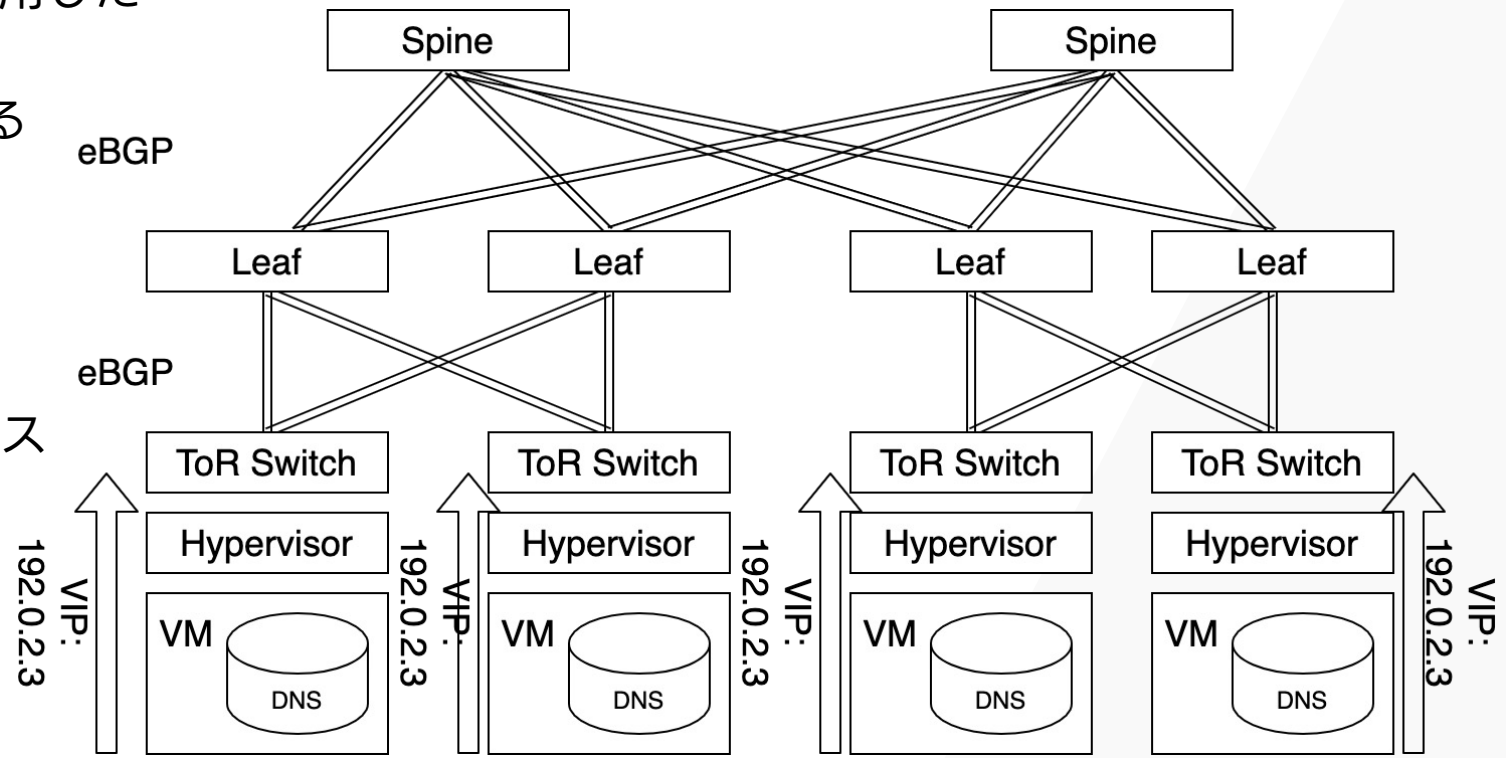


ANCYSAT+VMを用いたDNSの構築

- DNS CacheとAuthで同じ構成
- DNSサーバをVMを用いてBGPを利用した IP ANYCASTで構築している
- VMから /32でDNSのVIPを広報する

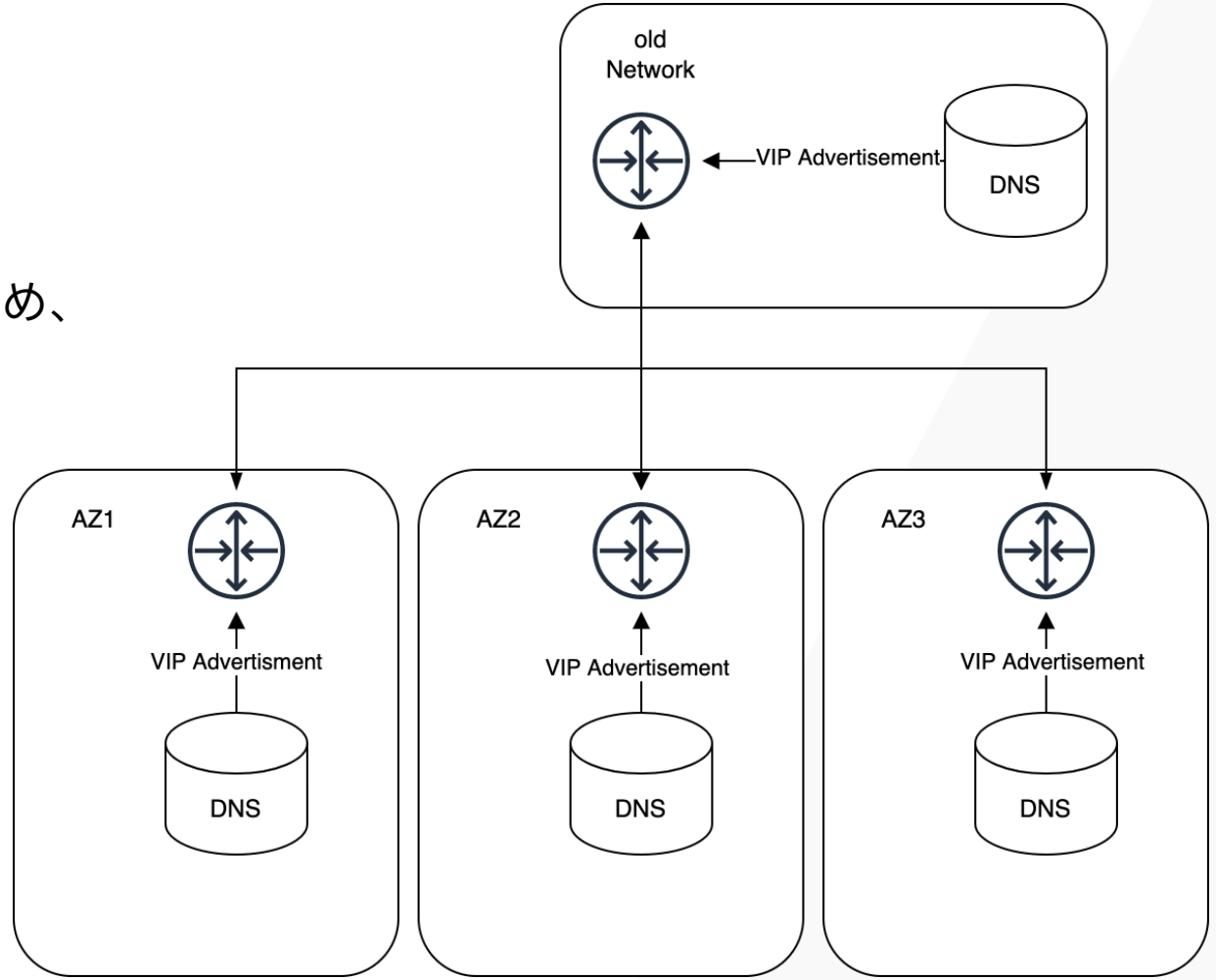
解決した事

- ほぼ無制限にスケール
- DNSオペレータのみでのメンテナンス
- FullのL3構成で構築



Multi-AZs対応

- ANYCSATを用いたDNS構成もちいて
  DNSサービスのMulti-AZs対応を行った。
- BGPにより経路を広報することで、
  VIPの持ち運びが楽に行える。
- DCのNWがBGPでAZ間も含めて制御されるため、
  BGPを用いて経路を制御する事で
  詳細なトラヒックコントロールが可能



Agenda

01 最近の各社のDNSの変化について

02 これまであった課題への対応

03 これからの課題

これまであった課題への対応

01	HWLBについて
02	DNSサーバの管理コストについて
03	IPアドレスが変えられない問題について
04	複雑な構成について
05	攻撃について

これまであった課題への対応

01 HWLBについて

02 DNSサーバの管理コストについて

03 IPアドレスが変えられない問題について

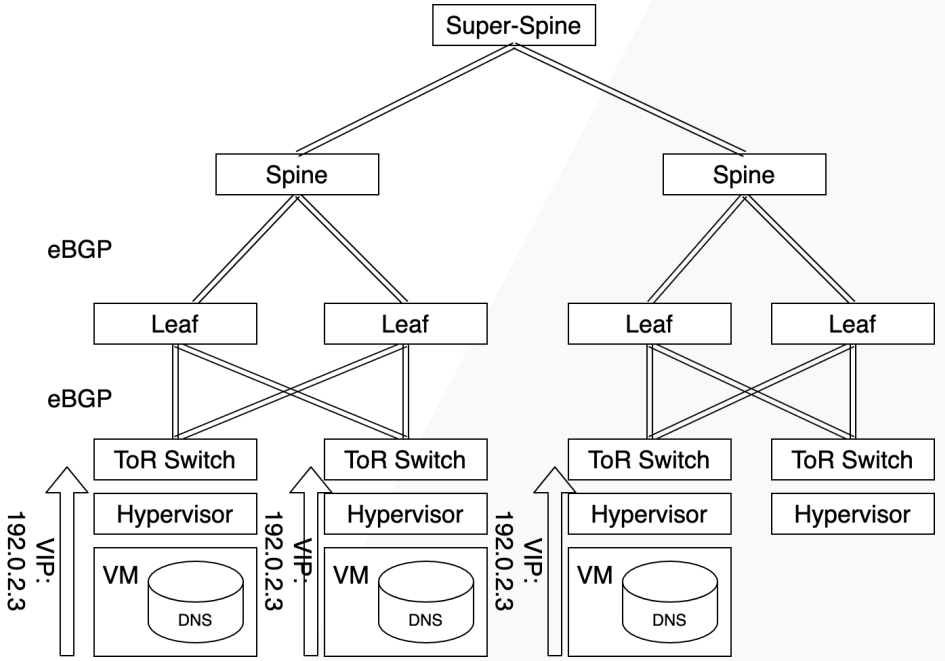
04 複雑な構成について

05 攻撃について

ANYCASTでDNSを構築してHWLBなしの構成に

- HWLBの課題
  - DNSの増設やメンテナンスの度に
    HWLBを管理しているチームとOperationが必要になる
  - HWLBのEoLによりVIPの変更などを行う必要がある
  - HWLBに関連したOutageを経験
- ANYCSATでDNSを構築してHWLBなしの構成にした事によって生まれた課題
  - トラヒックのバランシング
    - BGPで経路を吸い込んでいるため、
      Clos NW内での位置やDCの位置によってトラヒックのBalanceが崩れる
    - NW構成をある程度理解してスケールを行う必要がある

ANYCSATを用いて構築



HWLBについて

LINEヤフー株式会社(ex-LINE株式会社)
- BGP ANYCASTを用いてHWLBのない構成を構築
- BGPのANCYST化した事により、Balancingなどの別の問題も発生
- HWLBよりはDNSオペレータとして運用しやすい構成

株式会社インターネットイニシアティブ
- もともとHWLBは使っておらず、lvs+keepalived構成
- VRRP（を含めL2冗長化）って実際の障害時にうまく切り替わらない
- DPFはBGP+BFDで構成して、GWがECMPする構成

NTTコミュニケーションズ株式会社
- HWLB増設の際のサーバ調達や複数LBのトラヒックの平準化が課題
- 更にIPoE普及に伴い増設だけでは対応できず高性能LBの導入が不可欠
- これらHWLBに起因する諸々の課題を解消するためLBなし構成を設計し導入

LINEヤフー株式会社(ex-ヤフー株式会社)
- 過去の実績から使用中
- ヘルスチェックによるサーバの切り離しに利点
- コンテンツプロバイダとして他の用途でのHWLB 利用実績が豊富

これまであった課題への対応

01 HWLBについて

02 DNSサーバの管理コストについて

03 IPアドレスが変えられない問題について

04 複雑な構成について

05 攻撃について

DNSサーバの管理コストについて

- DNSサーバを物理サーバで構築する事による課題
  - サーバの調達に時間がかかる
  - メンテナンス対応を自分で行う必要がある
  - Multi-AZs対応の環境でPMで構築すると過剰なresourceを保持する事になる
    - 各AZに冗長構成でPMを配置する事になる
    - HWLBを利用した構成でないので、サーバの増設も全AZで同時に検討する必要がある

VMを用いてDNSサーバを構築

- VMを用いてDNSサーバを構築した事による課題
  - VMになる事によって性能の低下
    - 多くの余剰resourceを持つよりは安い
  - 管理台数の増加
    - 管理を基本的に自動化しているので台数の増加に対する課題はある程度解決

DNSサーバの管理コストについて

LINEヤフー株式会社(ex-LINE株式会社)
- PMからVMに変更
- VMにした事で1 node単位での性能はさがるがキャパシティの効率化などを行いやすくなった
- DNSのオペレーションに関しては全てAnsibleで実行

株式会社インターネットイニシアティブ
- 台数増えるより種類（設定・実装）が増えるのが辛い
- 設定の共通化やメトリクスの共通化を進めているところ

NTTコミュニケーションズ株式会社
- LBなし構成1setに収容させたことで分割損が少なくサーバ台数を大きく削減
- サーバはLB と比べ安価なため将来を見越して多めに増設可能
  - EoS時に前倒し調達を実施

LINEヤフー株式会社(ex-ヤフー株式会社)
- DNSの性能向上や、Query log収集方法の改善により台数の削減
- 構成管理ツールによりconfigの自動配布

これまであった課題への対応

01 HWLBについて

02 DNSサーバの管理コストについて

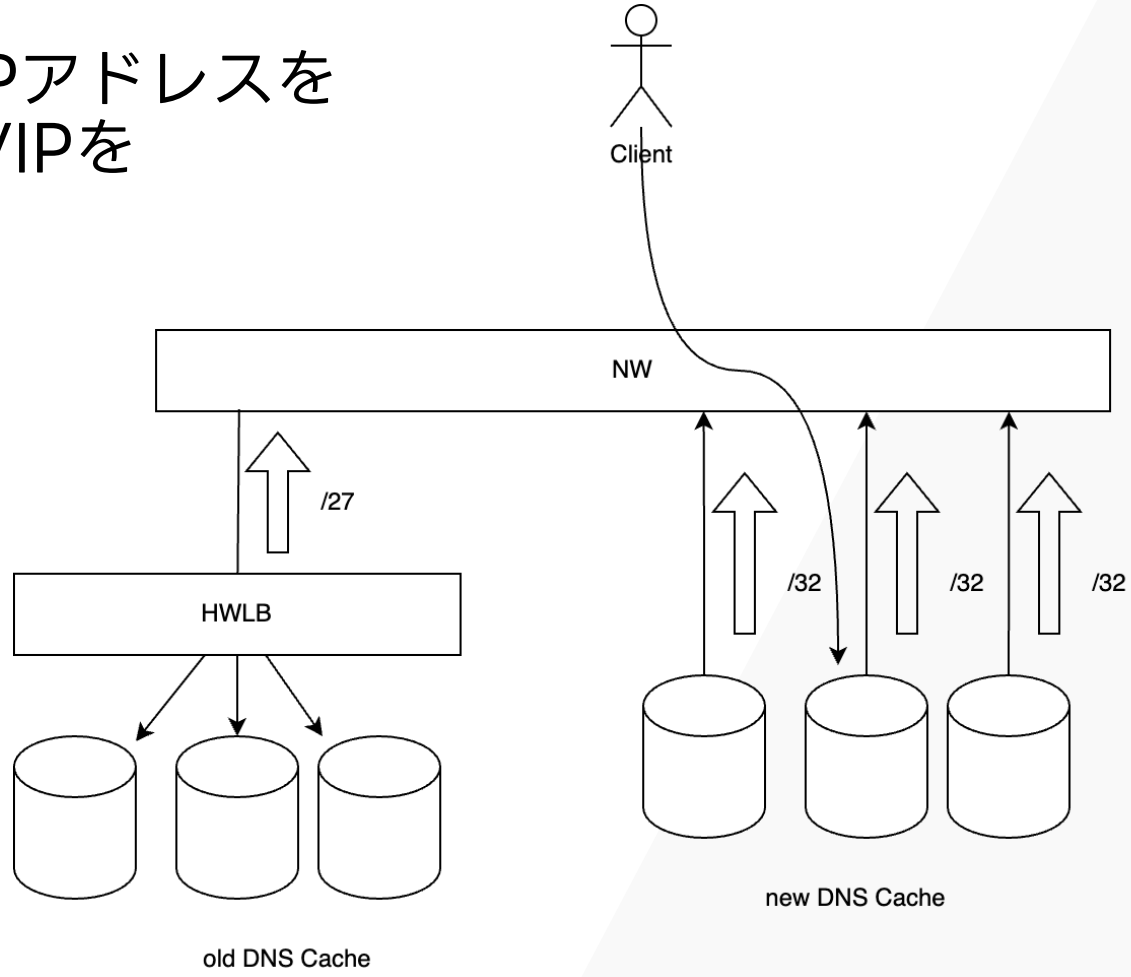
03 IPアドレスが変えられない問題について

04 複雑な構成について

05 攻撃について

IPアドレスが変えられない問題について

- IDC内の多くの場所でDNS CacheのIPアドレスを直接利用しているためDNS CacheのVIPを簡単に捨てる事は難しい
  - VMのresolv.conf
  - Dockerfile
  - CI/CD
  - NW機器
- BGPでVIPを/32で直接広報する事で古いVIPをHWLBの寿命とは関係なく持ち運ぶ
  - 古いVIPを保持する事になるがDNSの構成としては新しい環境に随時持ち運んで行く事が出来る



IPアドレスが変えられない問題について

LINEヤフー株式会社(ex-LINE株式会社)
- BGP ANYCASTでDNSを構築した事によりIPアドレスのポータビリティが向上
- 運用中のオペレーションコストはそれほど上がらない構成
- DNS CacheのIPアドレスを変えることは難しい

株式会社インターネットイニシアティブ
- 顧客に通知しているアドレスはBGPで1アドレス単位で広報してます
- フルリゾルバのIP、セカンダリのNotify先のIP、ゾーン転送を許可するネットワーク等

NTTコミュニケーションズ株式会社
- これまでは複数のLBに収容するためDNSアドレスも複数用意
- LBなし構成化により細かく分ける必要性がなくなったため今後は集約するポリシー
- 既存のIPアドレス削除はスグには実施できないがADSLサービス終了等と共に削減予定

LINEヤフー株式会社(ex-ヤフー株式会社)
- 内製ツールによる resolv.conf の自動更新
- 更新できない環境は担当者へ変更依頼
- IP Anycast の導入も検討中

これまであった課題への対応

01 HWLBについて

02 DNSサーバの管理コストについて

03 IPアドレスが変えられない問題について

04 複雑な構成について

05 攻撃について

複雑な構成について

LINEヤフー株式会社(ex-LINE株式会社)
- 社内でのみ利用されるDNSaaSなので、Zoneの構成などを自分達で選択出来るためそれほど複雑にならないようにコンサルティング

株式会社インターネットイニシアティブ
- 親子同居ダメ絶対！

NTTコミュニケーションズ株式会社
- 2020年6月まで権威DNSとキャッシュDNSが一部相乗り(同一IPで運用)
- 権威とキャッシュでトラヒックパターンが違うためrate-limitが困難
- お客様1万人以上を移行しフルサービスリゾルバ機能を停止したことで相乗り構成を解消

LINEヤフー株式会社(ex-ヤフー株式会社)
- Zone の管理もしているため、委譲などが適切になるようコントロール

これまであった課題への対応

01 HWLBについて

02 DNSサーバの管理コストについて

03 IPアドレスが変えられない問題について

04 複雑な構成について

05 攻撃について

攻撃について

LINEヤフー株式会社(ex-LINE株式会社)
- 処理性能の高いDNS実装を利用する
- 過去の攻撃に遭わせてDNSをスケールさせる

株式会社インターネットイニシアティブ
- 数万規模のものは、ほぼ毎週きている
- DPF運用に関わるゾーン自体もDDoSを考慮して設計されている

NTTコミュニケーションズ株式会社
- 2021年以前と比べ桁が1-2つ違うDDoS攻撃が複数回発生(攻撃パターンもそれぞれ異なる)
- DDoS対策を行っており影響なし
- UDP53トラヒックをRate-limitするとDoS(サービス拒否)が成立することがあるため今後も対策を検討

LINEヤフー株式会社(ex-ヤフー株式会社)
- DDoS 検知の仕組みと、キャパシティで受け切るアーキテクチャの併わせ技を検討中

Agenda

01 最近の各社のDNSの変化について

02 これまであった課題への対応

03 これからの課題

これからもDNS Cacheの負荷が増える事について(1/2)

最近増えたDNSの運用に関わるRFC

- RFC9460: Service Binding and Parameter Specification via the DNS (SVCB and HTTPS Resource Records)
- RFC9210: DNS Transport over TCP - Operational Requirements

Draft

- TLS Encrypted Client Hello
  - https://datatracker.ietf.org/doc/draft-ietf-tls-esni/
  - https://datatracker.ietf.org/doc/draft-ietf-tls-svcb-ech/
  - ESNIのためにDNSの名前解決の時点で鍵が欲しい
- TLS Key Share Prediction
  - https://datatracker.ietf.org/doc/draft-davidben-tls-key-share-prediction/
  - Downgrade攻撃などを回避するために、DNSでKeyをShareしたい

これからもDNS Cacheの負荷が増える事について(2/2)

- 基本的にインターネットにおいてIPアドレスと同時に欲しい情報はDNSが返答するという方向性
- セキュリティ周りではDoH + DNSSECが前提として話が進んでいる
  - 例: ESNIではDoHで検閲側がDNSのQueryが見れない+ DNS Cache側のデータはDNSSECにより改ざん防止されているという前提
- HTTPSレコードに乗るデータもどんどん増えていく
  - もし全てのWebサイトがHTTPSレコードを付けてKeyのshareなどをした場合にDNS Cacheが保存するCacheも同様に膨らんでいく

Discussion Points

- DNS人材について
  - DNS人材はいっぱいる説といない説がある
  - 一緒にDNSをやる人が居て欲しい
  - DNSに入って貰うにはどのくらいのDNSちからがあれば良い？
- DNSのConfigを事業者間で共有出来る仕組みってないですかね？
  - Bindのversion up時などでは有事によるポイントが共有されている
  - 設定入れて良さそう？
    - min-ttl入れると負荷が下がるけど入れて良いの？
    - QNAME Minimization
  - フルリゾルバにどういう設定をいれてる？
    - DNSSEC, QNAME Minimization, Root Server Local to Reoslver, minimal-response, etc….
- DNSの負荷について
  - DNSに求められる機能がどんどん増えていく
  - ブラウザの挙動によっても負荷が急激に変化する
    - Prefetch
    - ANY
    - HTTPS
    - TCP failback
  - ブラウザレベルで大規模にデプロイされるとISP側の負担が急激に増える事について
- DoHって
  - ISP側でやるべき？
  - MITMや盗聴とかには強くなりそう
  - Free Wi-FIなどでユーザが使う意味などが大きいが、ISP網を信頼出来るかどうかだしなぁ
- DNSプロセスの動かし形について
  - DNS Cacheはベアメタル？VM?コンテナ?
  - DNS Cacheは CPU/Mem/Network/Diskをフルに使う。(Tuningで性能が2~3割変わる)
  - VMだとしても1ベアメタルに1VMみたいな形になるのでは？
    - CPUはそれでいいけど、VNICが遅いよね
  - コンテナ使ってる?
- その他、登壇者や会場に聞いてみたり喋りたかったりする事何でも

触らぬDNSに祟りなし、寝たDNSを起こすな(1)

→ DNSに求められる事が増え変わっていくしかない

→ 眠らないように触り続ける

(1):レジストラとDNSプロバイダの世界 （後日公開Ver） @garnet_yn P.11
http://ck.rozen.jp/docs/dnsonsen/20230909-dnsonsen8-registrar-and-dnsprovider.pdf

LINEヤフー