IIJ Internet Initiative Japan

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)

株式会社インターネットイニシアティブ
ネットワーク本部アプリケーションサービス部DNS技術課
其田　学

Ongoing Innovation

自己紹介

其田　学

所属

- 株式会社インターネットイニシアティブ
- ネットワーク本部アプリケーションサービス部**DNS技術課**
- **2014年1月入社 今月で１０年目突入しました**

仕事内容

- お客様向けのDNSサービス・設備の設計、構築、運用
  - 権威DNSサービス、IIJ回線向けのフルリゾルバ、フルリゾルバサービス
- **外部での情報発信活動**

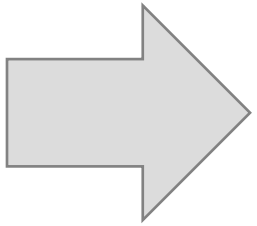
IIJの最近のDNSの変化

権威DNSサービスを一新しました

2019年に新サービスのIIJ DNSプラットフォームサービスを開始
2023年に旧サービスのマイグレーションも完了しました

DNSアウトソースサービス
権威DNSを２系統提供する
マネージドDNS

DNSセカンダリサービス
顧客のプライマリDNSの
セカンダリDNSを提供



IIJ DNSプラットフォーム サービス(略称DPF)

プライマリ/セカンダリ両対応
DNSSEC標準対応
ゾーン転送対応
セキュリティ重視の設計

複雑構成の解消 – 親子同居の解消

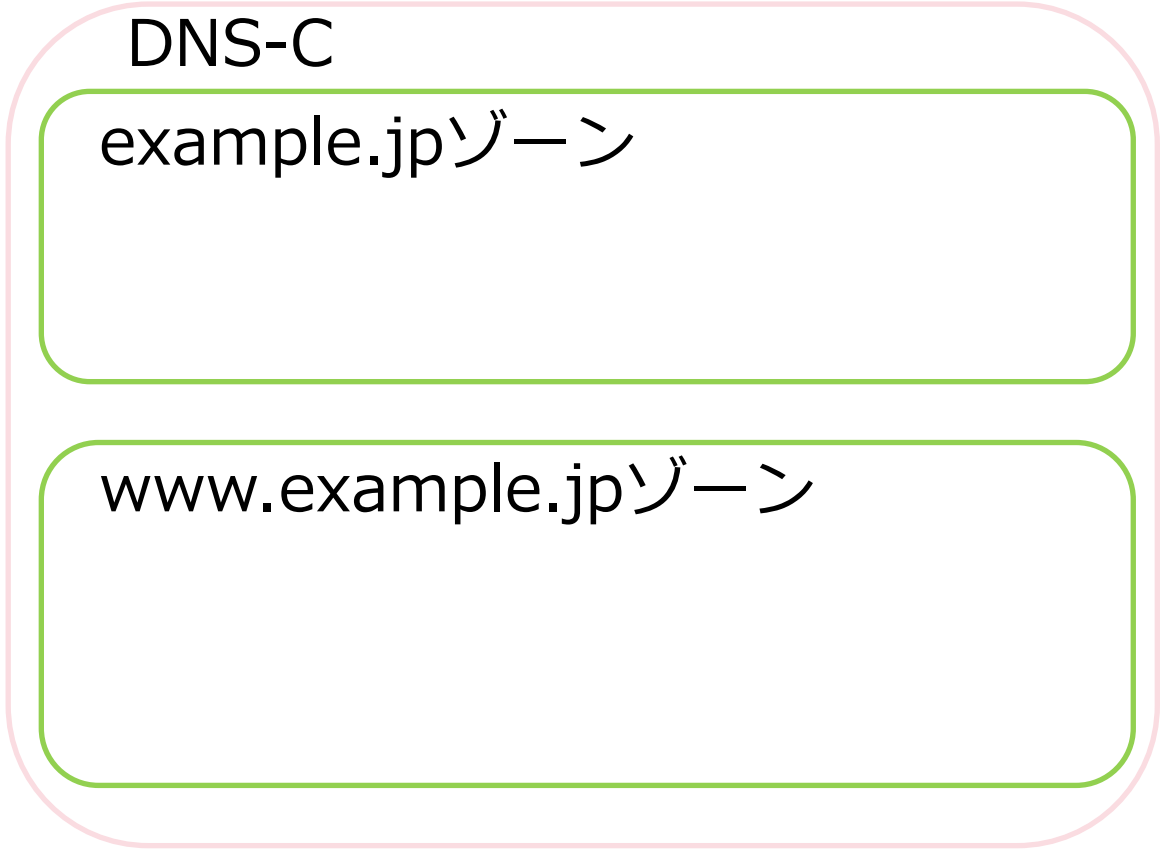
複雑構成の解消

- 親子同居とは
- 親子同居による問題
- 親子同居解消によるメリットと新たな課題

複雑構成の解消ー親子同居の挙動（1/3）

親子同居とは

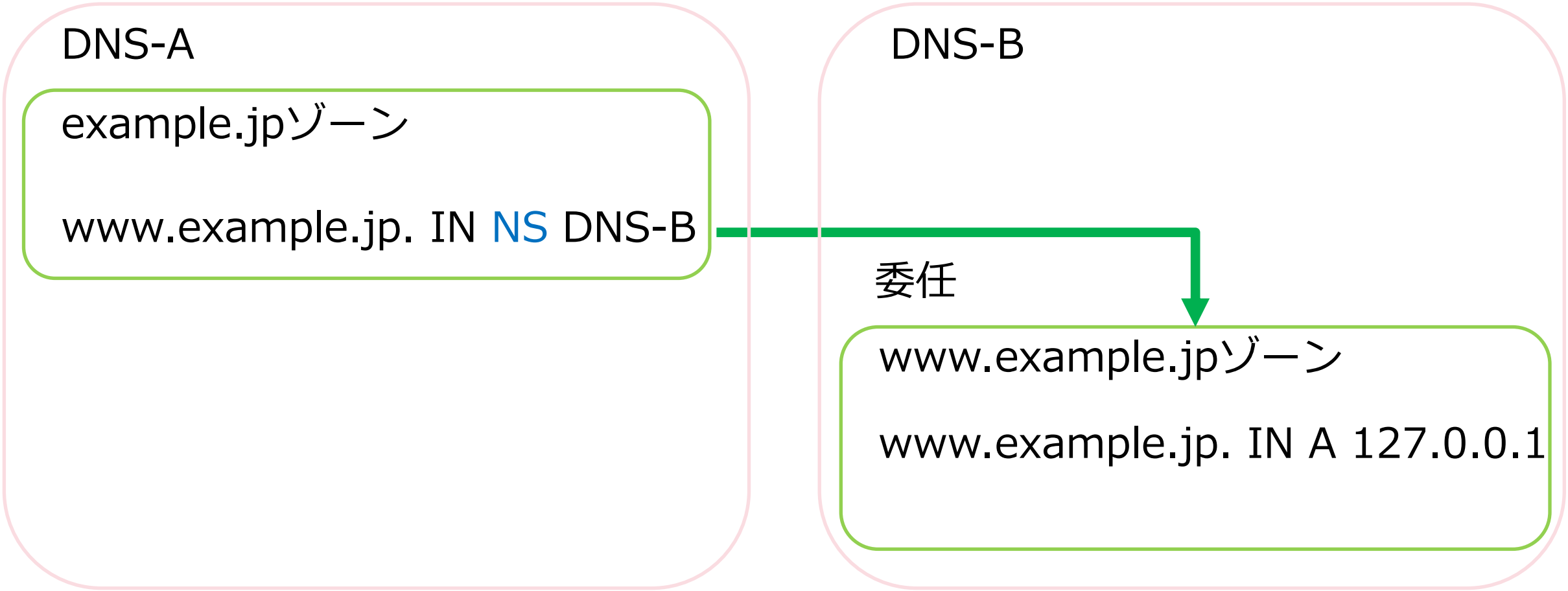
同じ権威DNSサーバに親ゾーンと子ゾーンが同居している状態



複雑構成の解消ー親子同居の挙動（2/3）

通常のDNSプロトコルが想定している委任

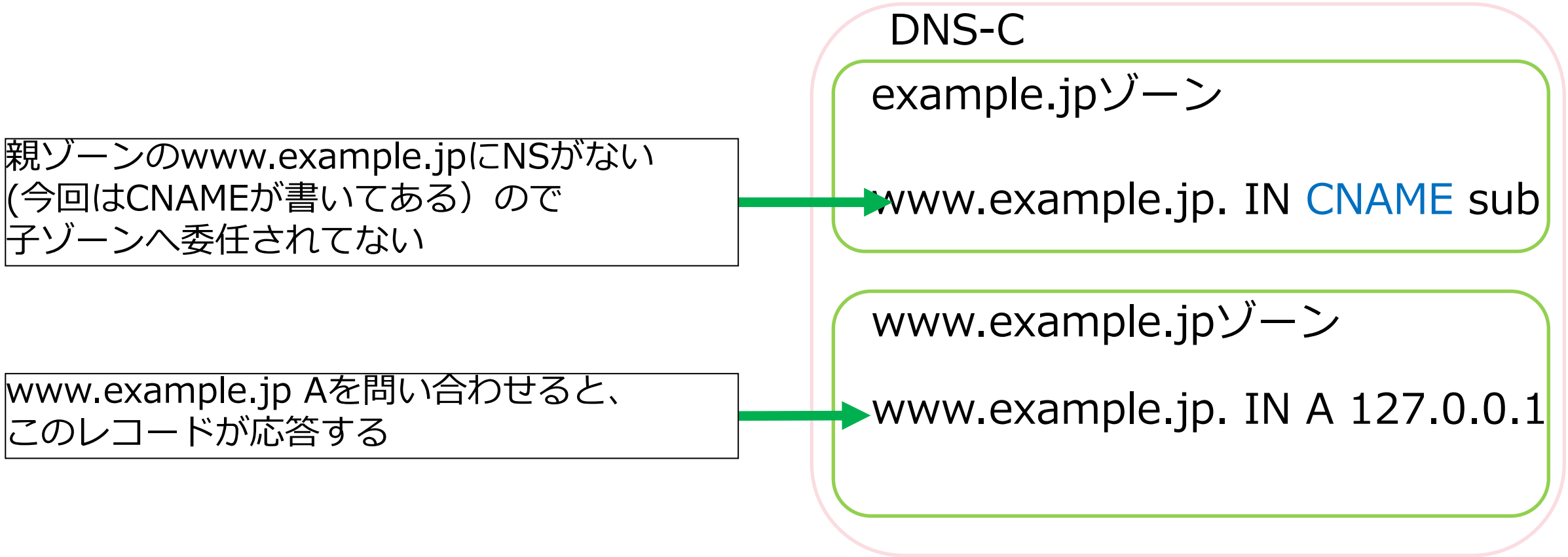
親ゾーンのGlue NSで権威DNSサーバを指定



複雑構成の解消ー親子同居の挙動（2/2）

親子同居による意図しない委任

親子同居していると、委任なし(Glue NS無し)でも
最長一致し、子ゾーンが応答してしまう



複雑構成の解消ー親子同居とその問題（1/2）

親子同居による意図しない委任の問題点

1. DNSSECが有効だと署名検証エラーになる

- 親ゾーンから見ると、Glue NSが無い
  - 子ゾーンは委任されてない=子ゾーンは存在しないのが正しい
- しかし、存在しない応答が返ってくるので、検証エラーになる

2. Glue NSがない時の挙動が実装に依存する

- プロトコルレベルでは、同居が考えられていない（と思われる）
- そのため、ゾーンカット(委任名)の挙動が実装によって異なる
- 実装によっては想定外動作をすることがある

Glue NSが無い親子同居はDNS的に壊れている

複雑構成の解消ー親子同居とその問題（2/2）

課題: 親子同居に起因する問題対応するための機能の複雑化

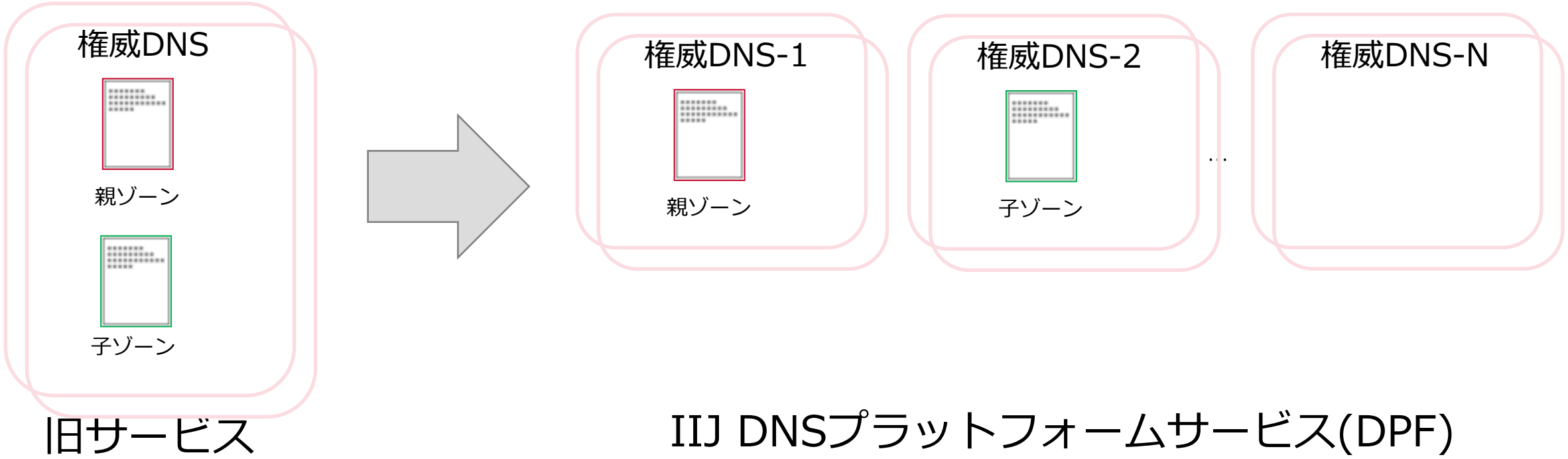
Glue NSの問題以外にも、問題があり、それをシステム側で対応

- 新規ゾーン追加時のGlue NSの追加
- サブドメイン名ハイジャック対策
- 空ゾーン公開対応
- 子ゾーン削除時に親ゾーンにレコードをコピーする機能
- etc…

システムが複雑化し、メンテナンス性や、拡張性が問題に

複雑構成の解消ー親子同居解消について

対応: 新サービスでは、権威DNSのセットを複数用意し
親子同居しないようにゾーンを配置



旧サービス

IIJ DNSプラットフォームサービス(DPF)

複雑構成の解消ー親子同居解消について

親子同居解消により得られたもメリット

- システムの設計がシンプルになった
- DNSとして想定外な挙動をすることが少なくなった
  - 顧客が操作しない限りは委任されない安心感

親子同居解消により発生した新たな課題

- 権威DNSサーバのインスタンス数が10倍以上増加
- 構築・運用面の負荷が激増
- コンテナ化を検討し、最終的にkubernetesを採用
  - インスタンス数による運用負荷はほぼない
  - しかし、ライフサイクルが早いk8sの運用が辛い状況
    - (k8sあるある)

攻撃への対応

攻撃への対応

- IIJのDDoS対策について
- 大規模化するDDoS
- 新サービスでの対策

攻撃への対応ーIIJのDDoS対策について（1/2）

権威DNS設備は攻撃を受ける可能性の高い設備です
ここ数年は特に国内でもさまざまな攻撃がありました

IIJも2012年に権威DNS設備にDDoS攻撃を受け、他のIIJサービスを巻き込んで障害が発生しました

2012年の障害を教訓として、今日までDNS設備の攻撃対策を進めています

攻撃への対応ーIIJのDDoS対策について（2/2）

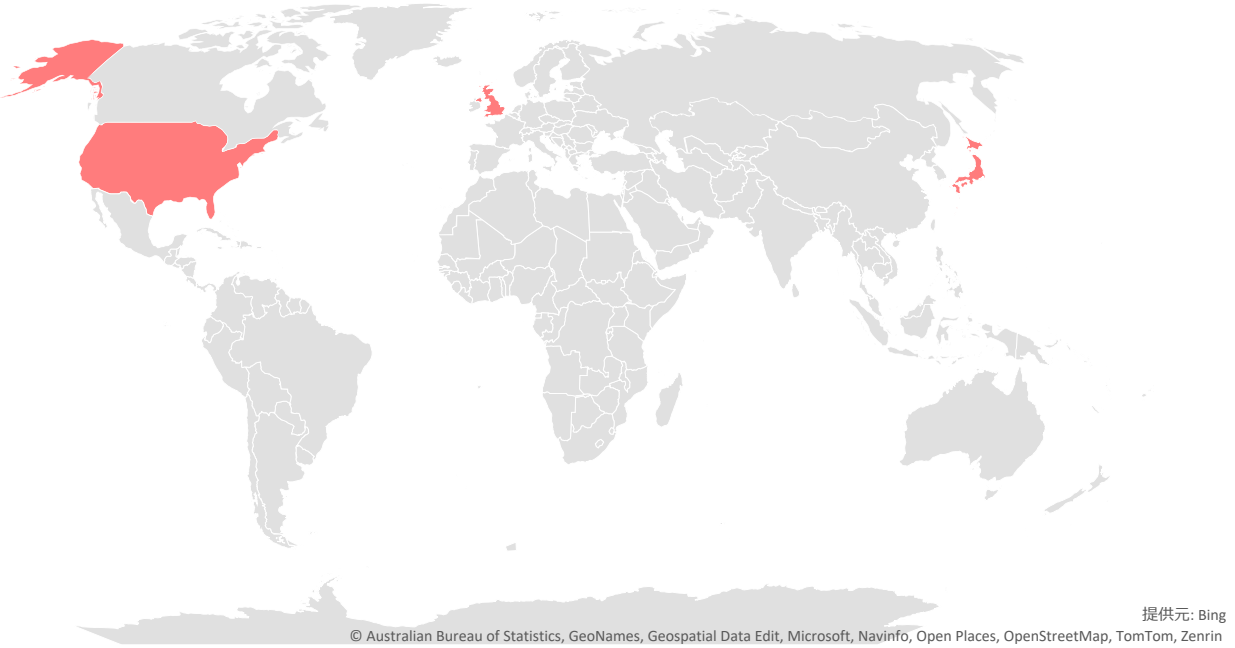
2012年の障害を受けて行った対応

IP Anycast
- 日米英３カ国 計4箇所

他サービスとの分離
- 巻き添え防止のためNWを分離

キャパシティの増強
- 攻撃されても、ある程度は
  打ち返せるだけのHWリソース

DDoS緩和装置の導入



詳しくは「IIJとDNSの30年」という動画・ブログで

攻撃への対応ー大規模化するDDoS

課題: TB級のDDoSに対して、重要なゾーンをどう守るか

2016年にDyn(現Oracle)がDDoSを受けて大規模障害が発生
著名なサイトが軒並み影響を受けました

国内の重要なゾーンを多く預かるIIJとして、Dynのような大規模DNS事業者でも耐えられないような、TB級のDDoSから顧客ゾーンをどう守るかが課題になりました

**従来のDNSサービス設備の強化という方針から
パラダイムシフトが必要**

もちろん新サービスでもサービス設備強化はやってます

攻撃への対応ー新サービスでの対策（1/2）

対応: マルチプロバイダ構成が取れるサービスにする

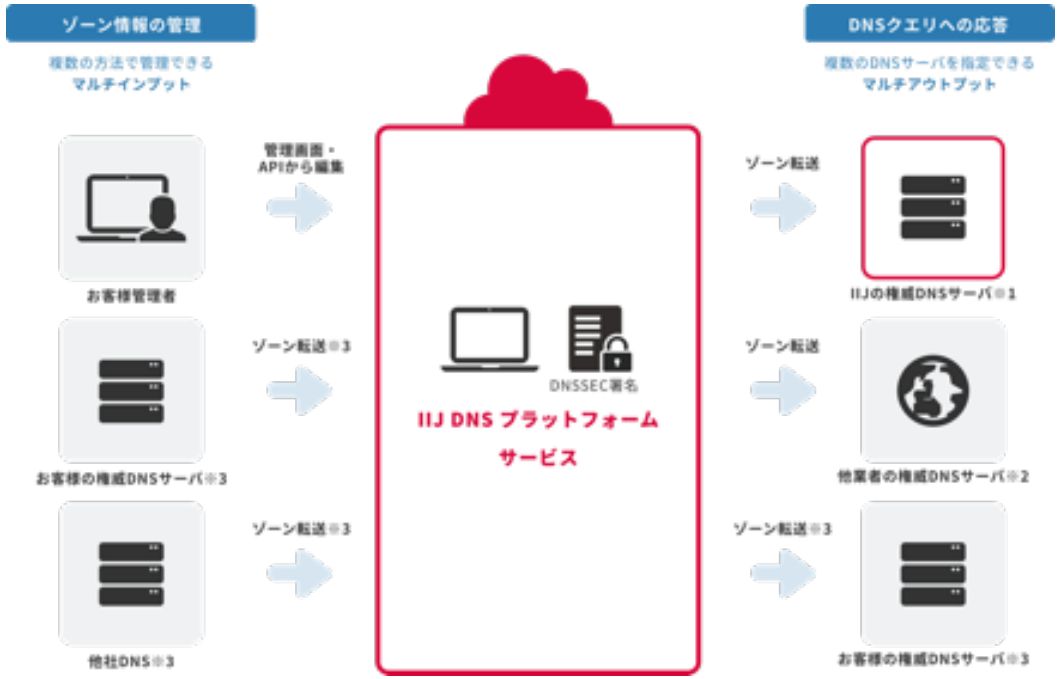
Dyn障害時に複数のDNSプロバイダを利用している場合、被害が少なかったことが改めてわかりました
これを受け、新サービスではマルチプロバイダ向けの機能を追加しました

1. ゾーンの受送信ができる機能

- 従来はセカンダリDNSのみ
- プライマリDNSにも対応

2. 他社DNSとセットのサービス

- IIJが他社プロバイダーを選定
- IIJで一括して運用
- DNS-LB的も対応



攻撃への対応ー新サービスでの対策（2/2）

マルチプロバイダ構成により得られたもの

- DDoSに強いゾーンをユーザが作成可能に
- 他社DNSサービスを統合したサービスにより、ユーザはDNSサービスの知識がなくても、簡単にDDoS耐性を上げることが可能になった

マルチプロバイダ構成を採用したことによる新しい課題

- 他社DNSサービスの選定
  - 実は親子同居していないサービスがあまりなく、簡単に連携できないパターンが多い

今・これからの課題

今・これからの課題

技術の複雑化

DNS設備を構築、運用するための技術スタックの変化

- Linuxの基礎的な知識
- DNS
  - DNSSEC
- ネットワーク
  - TCP/IP
  - BGP
  - BFD
- コンテナ
  - k8s
  - CNI
- CI
  - Github action
  - Gitlab CI
- IaC
  - Ansible
  - Terraform

赤字がここ５年で新しく加わったもの

今・これからの課題

人材不足

DNSに興味を持ってくれる人が少ない

- DNS以前にインフラに興味を持ってくれるひとが少ない印象

DNSの知識は後からでもどうにかなる
技術スタックに刺さる人募集中

Appendix

Glue NSが無い委任問題

- Glue NSが無い状態だと問題が多すぎたので、新規ゾーン作成時にGlue NSを追加する仕組みを作った
- しかし、この処理が複雑でやめたかった
  - 既存が親で、新規が子であれば、親にGlue NSを追加するだけ
  - 既存が親と孫で、新規が子の場合は。。
  - DNSなので、TTLも考慮しなければならない
  - 明らかに無理なケースは契約自体と止めるパターンもあった
    - これも分岐になるので複雑性がましていく

空ゾーン公開

- 子ゾーンを契約し、親ゾーン側に書かれたレコードを転記しないまま、空のゾーンが公開される事故
- 旧サービスでは空ゾーンの場合はゾーン公開しない制御を導入した

サブドメイン名ハイジャック対応

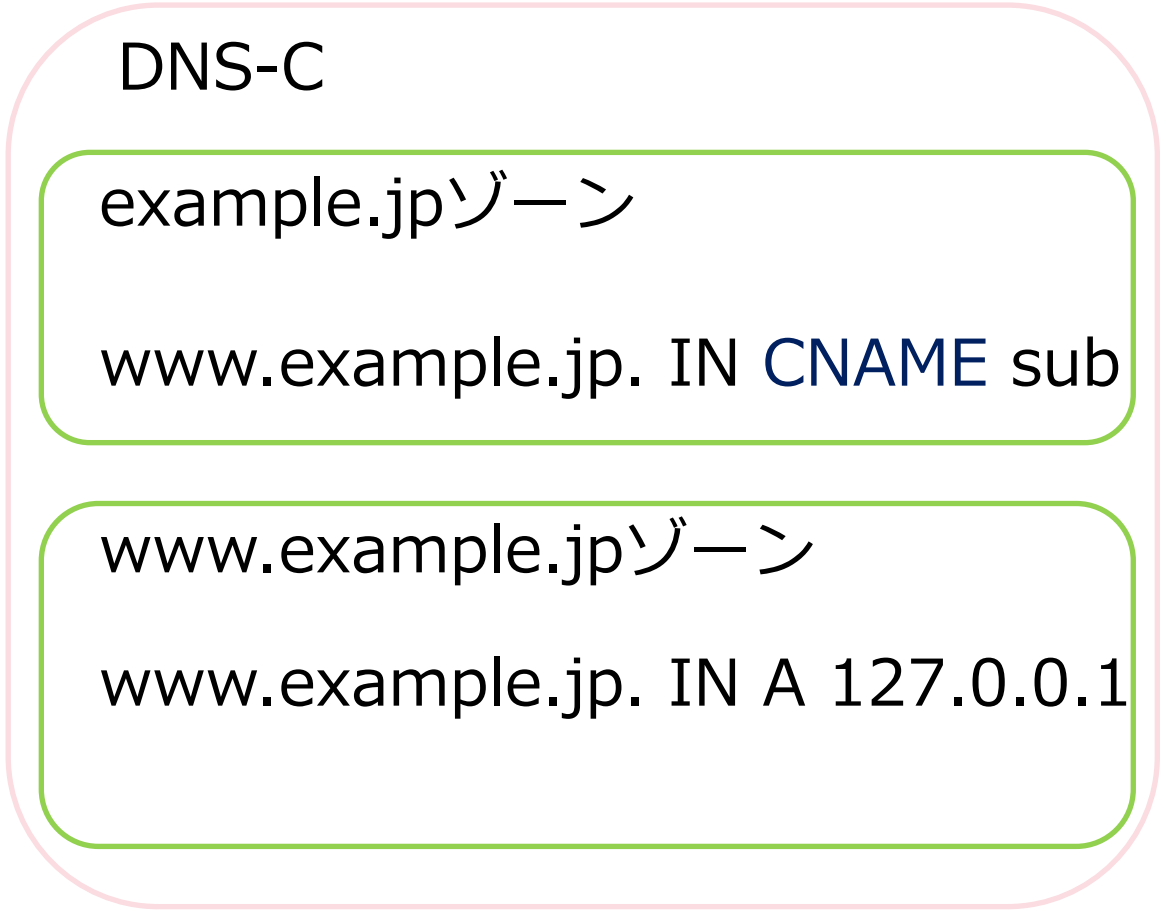
- マネージドサービスのような不特定多数の人がゾーンを追加できる環境かつ、親子同居する構成の場合に、子ゾーンを第三者が追加すると、簡単にレコードを乗っ取れてしまう問題
- **サービス事業者としては、追加を許可するか判断が必要**
- 機械的に同一契約者のみ登録可能にするのが楽だが、IIJでは不可能
  - IIJの場合、親子の契約者が別のパターン(例えば親会社・子会社・SIer等）が普通に存在する
- 運用的にも、確認後に契約を許可が必要など、とにかく大変だった

実装依存の例

DNSでは、DS RRTypeという、親ゾーン側に権威がある
レコードタイプが存在する。

右の例で
　QTYPE=DS,QNAME=www.example.jp
を引くと、親ゾーン側のCNAMEを返す実装が存在する。

フルリゾルバに
　Type=A Name=www.example.jp.
のキャッシュがない場合
DSのレスポンスでCNAMEを受け取り、キャッシュしてしまうと、Qname=www.example.jpの応答にCNAMEを返し続ける。



他サービスとの分離

DNS設備は、相互接続点に近い拠点でサービスしている。
打たれるなら相互接続点に近いほうが内部への影響が少ない

IP Anycast

複数の拠点で、同じIPアドレスをBGPで広報して、近い拠点に吸い込ませる。
レイテンシーの改善と、攻撃影響の局所化がねらえる。
BYOIPして、バックボーンがないところにも足出したいなとか妄想中

DDoS緩和装置導入

正常なクエリを誤作動で落とすこともあるため、すべてのネームサーバには導入していない。誤作動してもAnycast系でリトライしてくれることを期待している。
他社DNSサービスのセキュリティ機能も同じ理由で強めの設定を入れてる。

IIJ Internet Initiative Japan

Lead Initiative

日本のインターネットは1992年、IIJとともにはじまりました。以来、IIJグループはネットワーク社会の基盤をつくり、技術力でその発展を支えてきました。インターネットの未来を想い、新たなイノベーションに挑戦し続けていく。それは、つねに先駆者としてインターネットの可能性を切り拓いてきたIIJの、これからも変わることのない姿勢です。IIJの真ん中のIはイニシアティブ――IIJはいつもはじまりであり、未来です。

Ongoing Innovation