GMO INTERNET

DNS運用、その時どうする？

JANOG33 Meeting, 2014/1/24

GMOインターネット株式会社
永井祐弥

GMO INTERNET

自己紹介

名前

永井　祐弥（ながい　ゆうや）

所属

GMOインターネット株式会社

システム本部 サービス開発部

担当

2012年からGMOで、お名前.comのDNS関連サービスなど、DNS関連の開発、運用を担当

経歴

2003年（11年前）にdjbdnsと出会う。それ以来、前前職、前職でもDNS関連の開発、運用を経験。

GMOインターネットって？

こんなサービスを提供しています

- ドメイン名事業「お名前.com」
- サーバ事業「アプリクラウド」「ConoHa」など

公式キャラクターの美雲あんず（左）と、美雲このは（右）

GMOインターネットって？（続き）

DNSサービス「レンタルDNS」

- 登録ドメイン名数は70万件程
- 登録レコード数は600万件程
- 1日の総クエリ数は15億クエリ程
  - 1秒間あたりの平均クエリ数は18000/qps
  - DoS/DDoSなどの異常トラフィックを除く
- システム構成
  - BGP Anycast
  - マルチロケーション
  - 権威DNSサーバだけで24台
  - DNS実装ダイバーシティ（BIND 9 + NSD 3）

突然ですが
タイトル変更の
お知らせです

GMO INTERNET

~~DNS運用、その時どうする？~~

権威DNSサーバの
デュアルスタック化による
BIND 9のキャッシュDNSサーバに
発生する問題について

JANOG33 Meeting, 2014/1/24

GMOインターネット株式会社

永井祐弥

変更の内容

- 当初予定していた発表内容を急遽変更して、「権威DNSサーバのデュアルスタック化によるBIND 9のキャッシュDNSサーバに発生する問題」について報告します
- BIND 9のキャッシュDNSサーバは、ドメイン名の権威DNSサーバが全てデュアルスタック化している時に、SERVFAILエラーが発生して名前解決に失敗する問題を抱えていることが判明しました

変更の背景

- 当初はDNSサービスの運用をテーマに、弊社のIPv6対応事例についてトラブル事例を交えてお話する予定でした
- しかし、発表資料を作成する中で詳細な検証と原因究明を行った結果、問題の根が深いということが判明しました
- プログラム委員の方々とも相談した結果、今回の問題の原因と対策についてきちんと伝え、かつ資料として残すため、発表内容を変更させていただくこととなりました

そもそものきっかけ

- VPSサービス「ConoHa」のIPv6対応に合わせて、DNSサービス「レンタルDNS」のIPv6対応を実施
- 権威DNSサーバをデュアルスタック化してリリースしたところトラブルが発生
- 調査の結果、BIND 9のキャッシュDNSサーバの実装が原因と判明

どういう問題か？

1. あるドメイン名の権威DNSサーバが全てデュアルスタック化されている時に
2. その権威DNSサーバのAレコードのTTLが短いと
3. 到達性のないIPv6アドレス（リンクローカルアドレス等）を持っているBIND 9のキャッシュDNSサーバで
4. そのドメイン名の名前解決を行うとSERVFAILエラーが発生し、名前解決がエラーになる可能性が高くなる

既知の問題ではありませんでした

- JP DNSサーバの構成について - 2008年10月版 -
  - http://jprs.jp/tech/jp-dns-info/2008-10-06-jp-dns-servers.html
- 権威DNSサーバのデュアルスタック化による問題とその報告
  - http://dnsops.jp/bof/20081125/2008-11-25-dnsops.jp-BoF-dual-stack-01.pdf
- 上記はそれまでのBIND 9のキャッシュDNSサーバにおいて、名前解決に時間が掛かる可能性があるという話でした
  - 名前解決そのものはエラーにはならない

この問題の発生経緯

- 2008年11月頃に、BIND 9のキャッシュDNSサーバの応答が遅延する問題について修正される
  - 2468. [bug] Resolver could try unreachable servers multiple times. [RT #18739]
  - 到達不能な権威DNSサーバが存在することによる応答遅延のための修正
- この修正の副作用が原因と考えられる
  - 上記の修正が原因であることを特定出来たのは資料作成時にバージョン毎の違いを検証していて偶然気がつく

対象のBIND 9のバージョン

- 2008年11月以降にリリースされたBIND 9が対象
  - 9.3.6～（2008/11/20）
  - 9.4.3～（2008/11/20）
  - 9.5.1～（2008/12/24）
  - 9.6～（2008/12/24）
- それより前のバージョンでは、少なくともSERVFAILエラーは発生しない
- BIND 9以外（Unbound、dnscache、PowerDNS）では発生しないことを確認済み

発生条件(再掲)

1. あるドメイン名の権威DNSサーバが全てデュアルスタック化されている時に
2. その権威DNSサーバのAレコードのTTLが短いと
3. 到達性のないIPv6アドレス(リンクローカルアドレス等)を持っているBIND 9のキャッシュDNSサーバで
4. そのドメイン名の名前解決を行うとSERVFAILエラーが発生し、名前解決がエラーになる可能性が高くなる

つまりこういう権威DNSサーバの設定

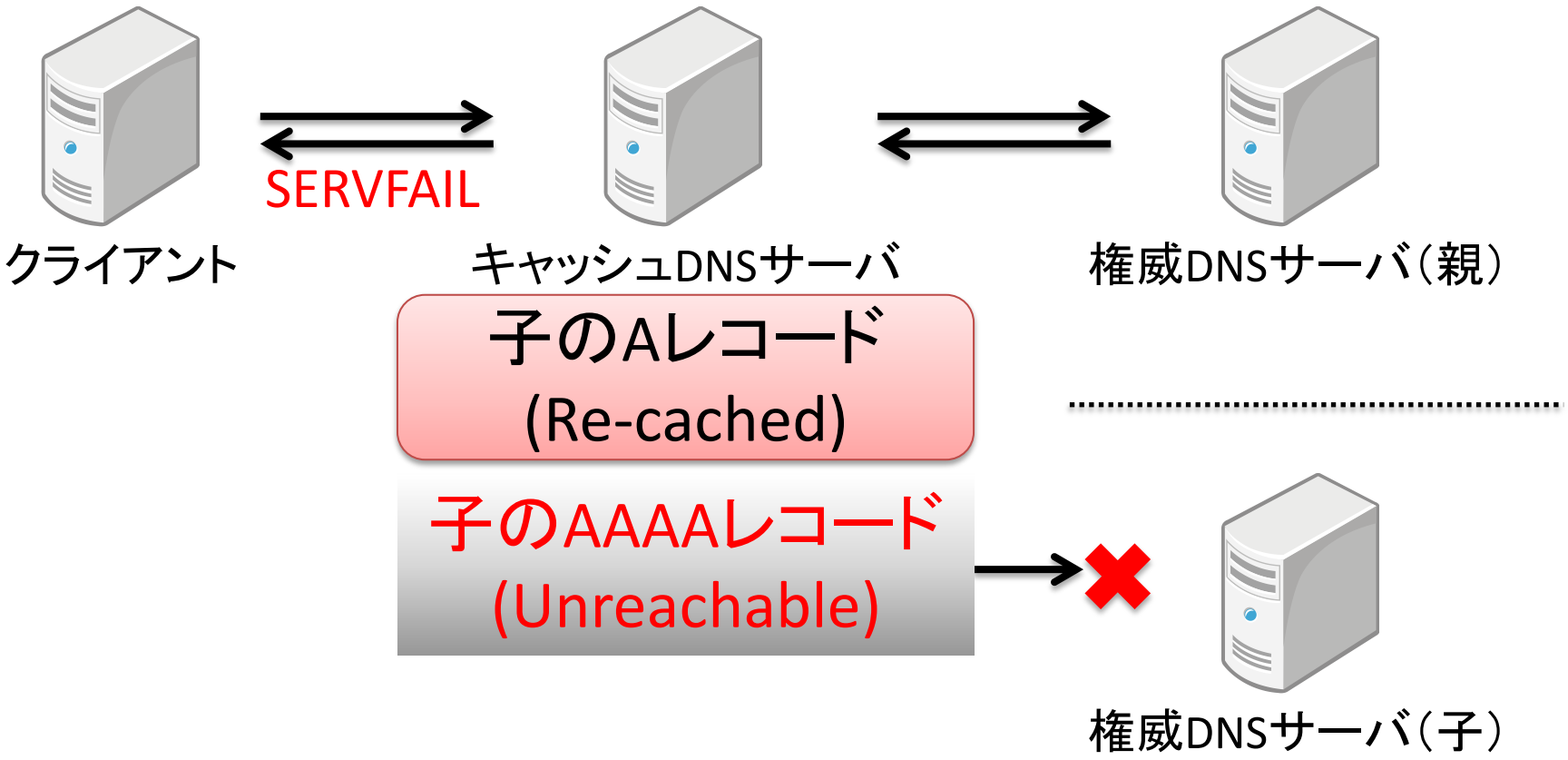
所有者名	TTL	RRタイプ	リソースデータ
example.jp.	3600	NS	ns1.example.jp.
example.jp.	3600	NS	ns2.example.jp.
ns1.example.jp.	300	A	192.0.2.1
ns2.example.jp.	300	A	198.51.100.2
ns1.example.jp.	300	AAAA	2001:db8:1000::1
ns2.example.jp.	300	AAAA	2001:db8:2000::2

※上記は説明のためのサンプルであり、
実際の設定とは異なります

何故問題が発生するのか？

- キャッシュDNSサーバ上での動作
  1. 権威DNSサーバのAレコードが、AAAAレコードよりも先にキャッシュから消滅する
  2. AAAAレコードがキャッシュに残る
  3. 権威DNSサーバにIPv6で接続しようとするが、到達性が無いので接続に失敗する
  4. Aレコードを再検索する
  5. IPv4にフォールバックせずSERVFAILエラーを返す

SERVFAIL発生時のイメージ図



何故Aレコードが消えるのか？

- キャッシュDNSサーバで権威DNSサーバのAレコードが先に消滅する条件
  - 親から得たグルーレコード（権威のない回答）のキャッシュが、子から送られてきた権威ある回答により上書きされることで、A/AAAAレコードのTTLに差異が発生する
  - 子のAレコードのTTLを親のAレコードよりも短い値に設定していると、特に差異が発生しやすい

顕在化しそうな問題ではあるが…

- この問題は、あるドメイン名の権威DNSサーバが全てデュアルスタック化されていて、かつ、AレコードのTTLが短い場合に発生する可能性が高くなる
  - 権威DNSサーバのA/AAAAレコードは、大体86400秒（1日）など長めの値を設定することが多い
  - しかし、権威DNSサーバの切り替えの際などにTTLを短くすることは実際にありうる
  - 加えて、全ての権威DNSサーバをデュアルスタック化しているケースは少ないため、顕在化していないのではないか？

セキュリティリスクについて

- 仮に、権威DNSサーバのAレコードのTTLが長くても、A/AAAAレコード間でキャッシュTTLを意図的にずらすことはDNSの仕様上可能
  - グルーレコードがキャッシュされた状態で、権威DNSサーバのAレコードを問い合わせればよい
  - DNSキャッシュポイズニング（毒入れ）とは違い、何度も問い合わせる必要がない

セキュリティリスクについて（続き）

- つまり、攻撃者がキャッシュを意図的にコントロールすることでサービス妨害出来る可能性がある
- 利用者が多いキャッシュDNSサーバは攻撃者の対象になりやすいため、セキュリティリスクが上がる可能性がある
  - 大手ISPのキャッシュDNSサーバ
  - オープンリゾルバ

問題の再確認

1. 応答遅延のバグを修正した結果
2. 本来はIPv4にフォールバックされる動作が
3. 途中で諦める奇怪な動作を起こした

どう考えてもバグとしか思えない

- 最近のBIND 9にありがちなクラッシュが起こらないとはいえ、名前が解決できなくなることは致命的
- 今後IPv6の普及が進むにつれて、問題が発生する可能性は上がると考えられる

ISCへバグレポート中

- 2013年4月26日に初投稿
- ISC-Bugs #33327
- 何通かやり取り後、放置されてます…

一時的な対応策

- いわゆる「運用でカバー」というもの
- 権威DNSサーバ、キャッシュDNSサーバのどちらも設定の変更で一時的に問題を回避できる
- 問題が修正されるまではこれを実施しておけばとりあえず大丈夫

権威DNSサーバの一時対応

- 同じリソースレコードセット（RRset）はRFC 2181の仕様により、TTLに差異が発生しない
- NSレコードにシングルスタックの権威DNSサーバが存在すると、その権威DNSサーバに名前解決要求を送るようになる
- つまり、権威DNSサーバの一部にシングルスタックのものを混在させる（残す）ことで、今回の問題を回避できる

つまりこういう権威DNSサーバの設定

所有者名	RRタイプ	リソースデータ
example.jp.	NS	ns1.example.jp.
example.jp.	NS	ns2.example.jp.
example.jp.	NS	ns3.example.jp.
ns1.example.jp.	A	192.0.2.1
ns2.example.jp.	A	198.51.100.2
ns3.example.jp.	A	203.0.113.3
ns1.example.jp.	AAAA	2001:db8:1000::1
ns2.example.jp.	AAAA	2001:db8:2000::2

※上記は説明のための（以下略）

キャッシュDNSサーバの一時対応

- キャッシュにAAAAレコードだけ残っても、IPv4に正しくフォールバックするようにしたい
- 次の環境では問題が発生しない
  - 正しくデュアルスタック化されたキャッシュDNSサーバ
  - OS上でIPv6を無効化してるキャッシュDNSサーバ
- namedにもIPv6を無効にするオプションがある
  - 「-4」オプション
  - IPv6での反復検索を止める

一時的な対応策まとめ

- 権威DNSサーバ
  - Aレコードだけの権威DNSサーバを混在させる
- キャッシュDNSサーバ
  - キャッシュDNSサーバーを正しくデュアルスタック化する
  - IPv6が不要なBIND 9環境ではIPv6機能を無効にする（named -4）

ご清聴ありがとうございました