JANOG44.5 LT

続！Goodbye IPv4 On L3 ToR

XFLAG
開発本部 インフラ室 上竹 嘉史

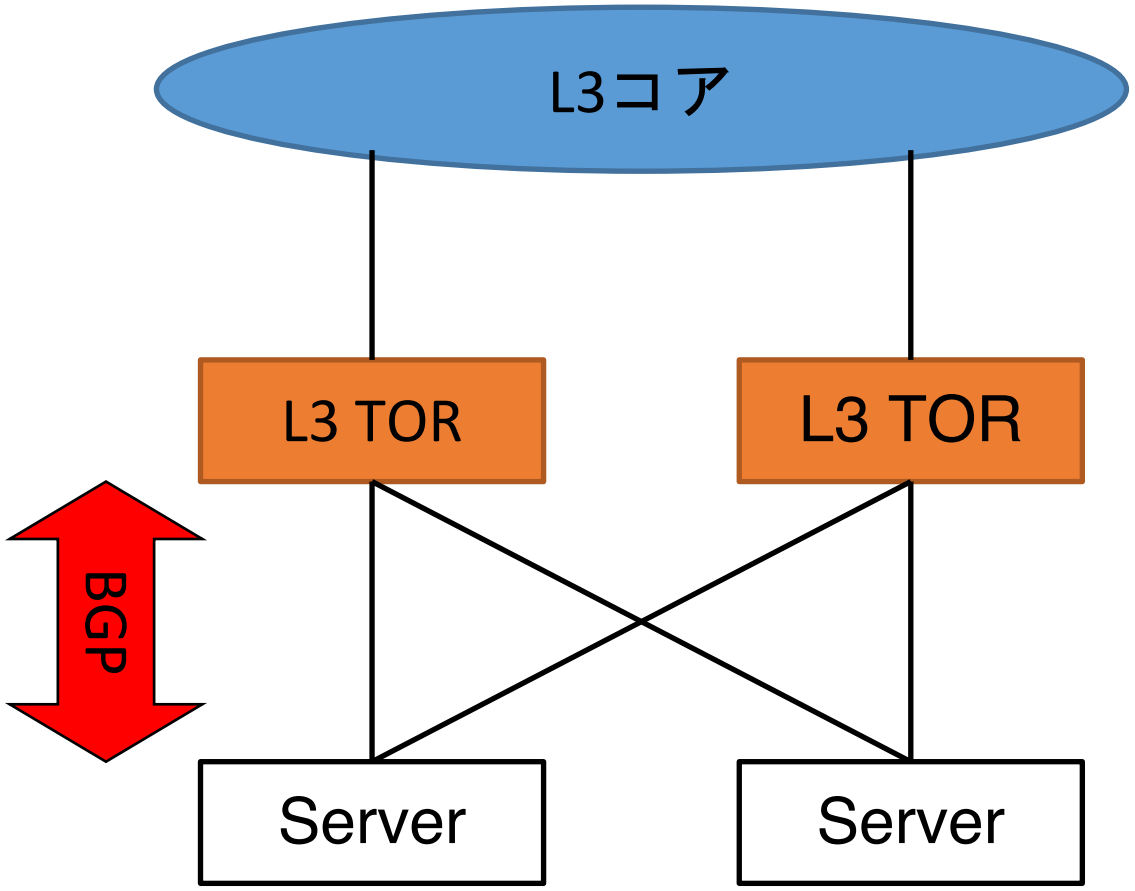
背景

- JANOG44 のぺちゃくちゃ枠にて発表
  - 「目指せ！Goodbye IPv4 on L3 ToR」
- 管理対象最小でサーバープロビジョニングをどう実現するかに焦点
- 今回続編ということで運用についてフォーカス
  - 将来の展望を含め

- まずは軽くおさらいから

おさらい

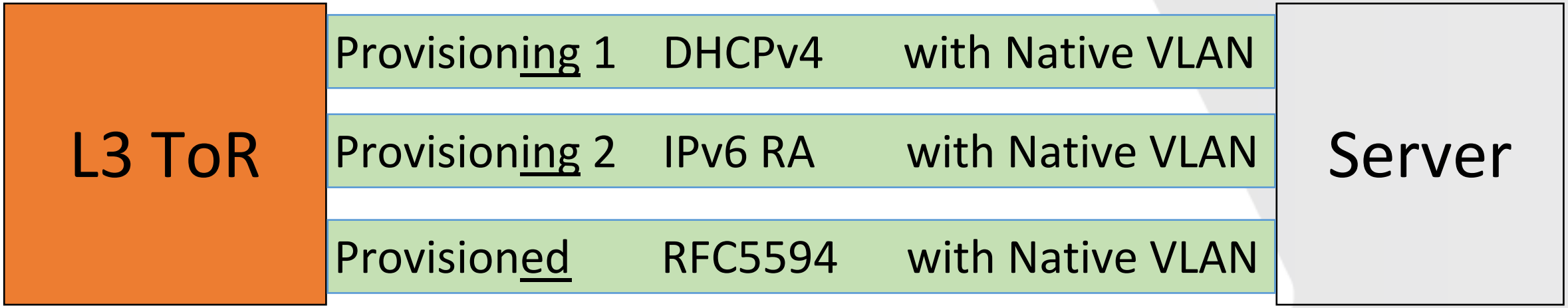
構成



- IP CLOS Network
  - OSPFでToRまでL3にしてみた(J40)
- ToRとサーバの接続 L2 => L3
  - フルL3化
- ルーティングプロトコルにBGPを選択
- ToR のメンテナンス性の大幅な向上

- ただし設定に関して楽をしたい
  - 管理が必要なパラメーターを最小限に！

プロビジョニングの流れ



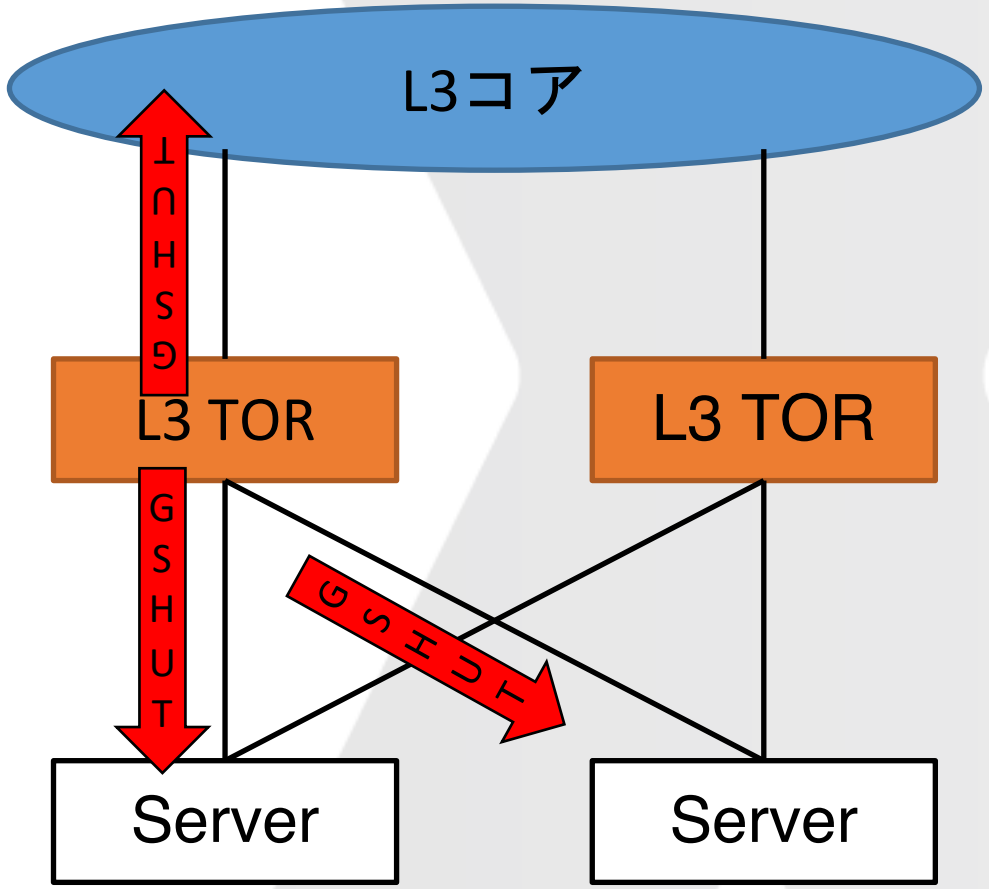
- インストーラ取得まではDHCPv4
- インストール中はIPv6 RA + NAT64
- 起動後はRFC5594
  - Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop

本題

迂回について

Community による迂回

- GSHUT Community(65535:0) を利用
- 迂回時に上流コアとホストに対して、コミュニティ付与してUPDATE
- コミュニティ付きの経路のLocal Preference を0 に



メンテナンスモードの利用

- ToR にArista社製品を採用
- EOS にはメンテンナンスモードという機能がある
  - 設定要素で構成されるユニットを定義
  - ユニット単位でメンテンナンスon/off 可能
- 弊社で定義していてよく使うユニット
  - eBGP の広報経路にGSHUT communtyを付与するユニット
  - 接続リンクをshutdownするユニット
- on-boot を使って起動時の動作を制御することも可能
  - 上流とのBGP が安定してから、ホストとの通信可能に など

設定例

```
group bgp eBGP
   neighbor HOST
   neighbor CORE
   exit
!

group interface all
   interface Et1-54
   exit
!

route-map gshut permit 10
   set community GSHUT
additive
!
```

```
maintenance
   profile bgp gshut
      initiator route-map gshut out
   profile bgp gshut default
   profile interface shutdown default
   !
   profile interface shutdown
      shutdown max-delay 0
   !
   unit bgp
      group bgp eBGP
      profile unit gshut
   !
   unit link
      group interface all
      profile unit shutdown
!
```

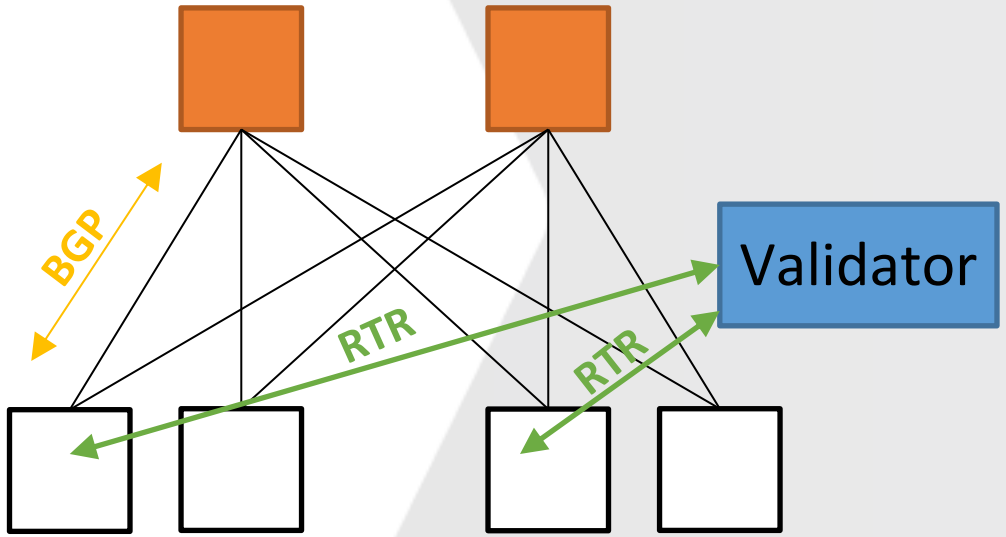
経路ハイジャックについて

つらみ

- 経路ポリシーをホスト向け全neighbor で書くのは辛い
  - neighbor group は共通にしたい
  - /32 exact で書かないと本来守りたいことは実現できない
  - 我々のモチベーションは、管理対象を最小化すること
    - 自動化より管理対象削減
- 何もいじらなくてもどこにつないでもプロビジョニングできて疎通できるのがメリット

RPKIでOrigin Validation

- データセンター内に独自のRPKI作成
- Validator とRTR を使って経路検証
- Invalid/Not Found な経路を検知
- AS番号とPrefix の組み合わせは一意なので実現可能
  - AS番号は資産管理上のホスト名から自動採番
    - a06248 => 4200006248
  - Prefix はホストに対して最初に与えたら永続的に固定
    - 物理的な成約が一切ないから



ホスト側BGPデーモンのメンテについて

BGP LLGRへの期待

- Long-Lived Graceful Restart(LLGR)
  - 古い経路を低優先度で保持できる拡張機能
  - LLGR_STALE(65535:6) well-knownコミュニティを付与
- 影響を与えずホスト側のBGPデーモンを再起動できる
  - メンテンナンス
  - Verup など

- 最終的にはGSHUT/RPKI/LLGR を組み合わせて
  - LP 強弱を 「GSHUT < RPKI < LLGR < 通常」のようにLP 設定しようと考えている

聞いたみたいこと
(時間が余ったら)

会場へ

- 迂回こうやってるよ
- 経路ハイジャックに対してはこう対策しているよ
- ホスト側ルーティングデーモンをメンテナンスする際はこうやってるよ

- ぜひ実際例やご意見をお聞かせ下さい

Thank You!!!