
■ Interdomain Routing Security Workshop 11
■ 日時 : 2006/12/11 14:30-16:45
■ 会場 : Cisco Systems 赤坂オフィス

- Agenda
  1.BGP Loop よもやま話@さくら 大久保さん
  2.20万経路に挑戦@MEX 高田さん
  3.CEF 苦労話@上川さん

1.BGP loop よもやま話
  発表者 : 大久保修一@さくらインターネット

  - ネットワークの不思議な現象
    経路広報をストップしても経路が残る?
    eBGP ピアを落としても経路が残る?

    -> 実際に発生しました

  - 顧客で bgp peer down が発生した時に、経路が残る現象が発生
    -> debug ip bgp で調べた
       ループ経路受信時/送信時で実装が違った。各2通り。

  - AS_PATH に経路送信先の AS が既に含まれている経路情報が BEST に
    なった場合に、

    1-1) withdrawn を送信するケース
    1-2) 自ASを付加し、AS_PATHにピア先ASが含まれた経路情報を送信するケース

    -> 1-1) の場合には特に問題ないが、1-2) のケースで 次の問題が発生。

    - AS_PATH に 自AS が含まれた経路情報の受信時に、
      2-1) 過去受信した同一経路の prefix 情報を RIB-in から削除する
      2-2) 削除しない

      -> 2-1) の場合は問題ないが、2-2) の場合には消えるべき経路情報が
         残留してしまう。
         RIB-in から削除しないのは bug ですよね?

  - 対応済みルータ
    - F10 Eシリーズ
      -> bug だと認めて修正した
    - Foundry NetIronシリーズ
      -> 最近修正された
         Foundry は RFC で明記されていない事象は、bug と認めない!?
     - GRシリーズ
      問題なし
    - zebra(quagga)
      問題なし

  - ワークアラウンド
    route-map で out 方向で 対向AS を落としてあげると withdrawn されるので問題な
い

  - Originator_ID ループによる経路残留について

    - Originator_ID ループのワークアラウンド
      - route-map で originator_ID で match して deny できれば。。
      - wellknown problem ?
      - バグですよね?
      - 使用だと言い張るメーカの説得方法

- 議論
  - Q: AS_PATH に 自AS が含まれた経路情報の受信時に、RIB-in から削除しないのは b
ug か?
    A: bug 。先にループチェックをして、NG なら RIB-in に格納しない実装であるべき
。
       ただ、RFC にはそこまでの明確な記述がない為、ベンダにより実装が違うかもし
れない。

  - Q: Foundry で BGP 使ってる人はどれくらい?
    A: 使っている人は 1,2 人。使い込んでいないのかもしれないが、
       特に問題は出ていない。

  - ワークアラウンドの route-map を書く方法は疲れる
    -> 書くなら bug を潰してもらう

- YAMAHA ルータは BGP 使えるがループ検知をしていない?
  -> IP-VPN などへ接続するためだけに、BGP が動くようになって
     おり、ループが発生するような可能性がないところで使われる
     想定の作りになっている!?
     また、意図的にループ検知をしていないという話も聞いたことがある。

- Q: 通常の update における withdrawn でも問題は出るのか?
  A: 発生するのは 自AS を含むループ経路受信時のみ

- Originator_ID の不具合は治った?
  - F10 は治っている。Foundry はまだ。

- 運用者からすると、気づきにくい bug というのが嫌なところ。
  -> 今回の事象も顧客からの申告によって気づいた。

- Q: 不具合時にどうやってベンダに伝えたら効果的?
  A: - janog で話すと言う
     - 自分の所だけでなく 2,3 企業からベンダを攻める
     - ベンダに公開質問票を投げる

- 一番にいいのは RFC を直しもらう。メーカによって解釈が違うのはよくない。
  -> RFC4271 で BGP 関係は一旦ブラッシュアップされたので、
     今から盛り込み直すのは難しいかも。。

- IRS で今回のような情報を共有できるようにしていきましょう。

2.20万経路に挑戦
  発表者 : MEX 高田さん

- 経路受信の状況
  - verio 201421
  - kddi 193809(国際のみ)

- ルータの Forwarding Table
  CEF 問題 23万9千経路まで

- Hitachi
  - GR2000-BH
    約25万経路
    -> 入らなくなると、入らない。何も言わない。
  - GR4000
    39万経路(v4のみの場合)

- Juniper
  - 60万くらいのる? メモリーがあるだけ!?

- Foundry
  256000 まで。メモリーがあっても制限がかかっている。

- 鬼のように/24
  /24 大杉
  /19 なのに /24 たくさん流すのやめて
  /8 なのに、/16 とか流すのもやめて
  パンチングホールも多すぎ
  /32 とか /24 より長いやつはさすがに聞かない
  -> 問合せがあったが、到達性は確保している。

- このままのいきおいで増えつづけると
  -> v4/v6 デュアルスタックなんて無理。

- で、どうすんのさ
  - bgp を使っているけど、実は multihome じゃない所
    - 適当に aggregate
    - 4/8 とか

  - multihome だと
    - longest match しちゃうから、削ると負ける
    - 自分で複数の上流を持っている場合、両方やらないと偏る
    - 国際売ってくれる人は合わせてやってくれないと

- 議論
  - 7206VXR が死んだ

- メモリが 512 MB しかない GR2000-BH も死ぬ。メモリが高い。
  Q: 普通の DIMM じゃだめか? 見た目は普通の DIMM っぽいが。。
  A: 高さが低いので普通のだとダメなはず。低いのを探すか、高いのを
     買うしかない。

- Q: 7206VXR が死んで NPE-G1 に対策した人はどのくらい?
  A: 2,3人が挙手。

- Q: 経路を削るとトランジット提供経路数も減るのか?
  A: そこを減らしてしまうと他の ISP に longest match で負ける。

- Q: どういう基準で aggregate するのか?
  A: 例えば verio / kddi が upstream だとして、verio 方向に
     bestpath が一箇所向いている連続した CIDR は aggregate
     -> ただし、状況は変わるかもしれないしのでメンテは大変。

- 頑張って第1オクテットの若番(4/8とか) の方は aggregate できる経路ををを調べてみた。
  -> 200番台の aggregate を誰か調べて教えてください。

- aggregate して良い経路を fullroute から抽出するスクリプト誰か作って。
  -> aggregate 箱なんていうのがあったら売れる?

- ルータの経路上限が近いので ospf 経路は流さないでね。
  -> 昔 uunet であった。たしか約5万経路くらい。

- Q: そもそもどの程度で経路数が増加している?
  A: 1年間で 2万5千経路 程度増えている
     残りが少なくなると、どんどん細かくなる。
     過去からの計算で 1.15 倍/年で推移している。2013年で約 60 万経路。

- Q: v6は 100万経路入るくらいで設計されている?
  A: 25万程度で作っているはず。

- community を使って aggregate して良い経路がわかるようにしてほしい。

- 困ったというときにルータは買えない。
  そういう場合に コマンド手法 などの指針があればいい。
  -> 割当 block で aggregate する?
  -> どのコマンドを入力して・・、どれを aggregate するかが分かる手順。

- Q: ルータの経路数エントリ上限が来たらどうする?
  A: 買うという人が多数。経路の aggregate を考える人は少数。。

3.CEF 苦労話
  発表者 : 上川さん

- 事象紹介
  - 特定のサイトに接続できない
  - そこに ping を打つとロスる
  - ロスは BGPルータで発生していた

- 原因
  - CEF でキャッシュあふれしたルートが CPU エスカレ処理できないバグにヒットしていた。
    -> CIDR が大きいのがあふれると塊でトラフィックが落ちる

  - 現象発生時は show mls cef summary で 197Kルート
    -> sh mls cef maximum-routes では 192K 。5K くらいあふれていた!?

- 対応
  - CEF テーブルサイズを大きくして対応
    (192K->220K)
  - バグ対応ファームアップ

- mls cef maximum-routes
  - XL-mode とnon-XL mode がある

- cef 関連 URL
  # 資料を参照

- 議論
  - Q: 機種は?
    A: Cat6500 で sup720-xx

  - 対策してもどこかが漏らして 239K を超えると NG 。

  - Q: mls cef maximum-routes で変更した後は reboot が必要?
    A: 必要。再起動して変更が有効になる。
       他メーカのルータもそれはほぼ同じ。

  - Q: mls cef maximum-routes の変更での値はどう決める?
    A: ベンダに確認し、推奨値で設定した。

  - Catalyst6500 や Cisco7600 は IPv6 や Multicast を切り捨てて
    IPv4 用のエントリを増やす。
     -> show mls cef maximum-routes で default のエントリ上限を確認するべし。

    Q: multicast が 0 にしたら ospf が使えないんじゃ?
    A: multicast のエントリは 1k くらいは残る
       0 になったら ospf 使えない。そうならないようになっている。

- 次回
  - Agenda
    募集中!

  - 日程
    2007/3/9(Fri)