IRS26
longer経路を一番安全にwithdrawする方法選手権

2017.03.24

NTTコミュニケーションズ株式会社 西塚
アクセリア株式会社 柴崎
株式会社FORNEXT 篠宮

背景: DDoS対策はより上流へ

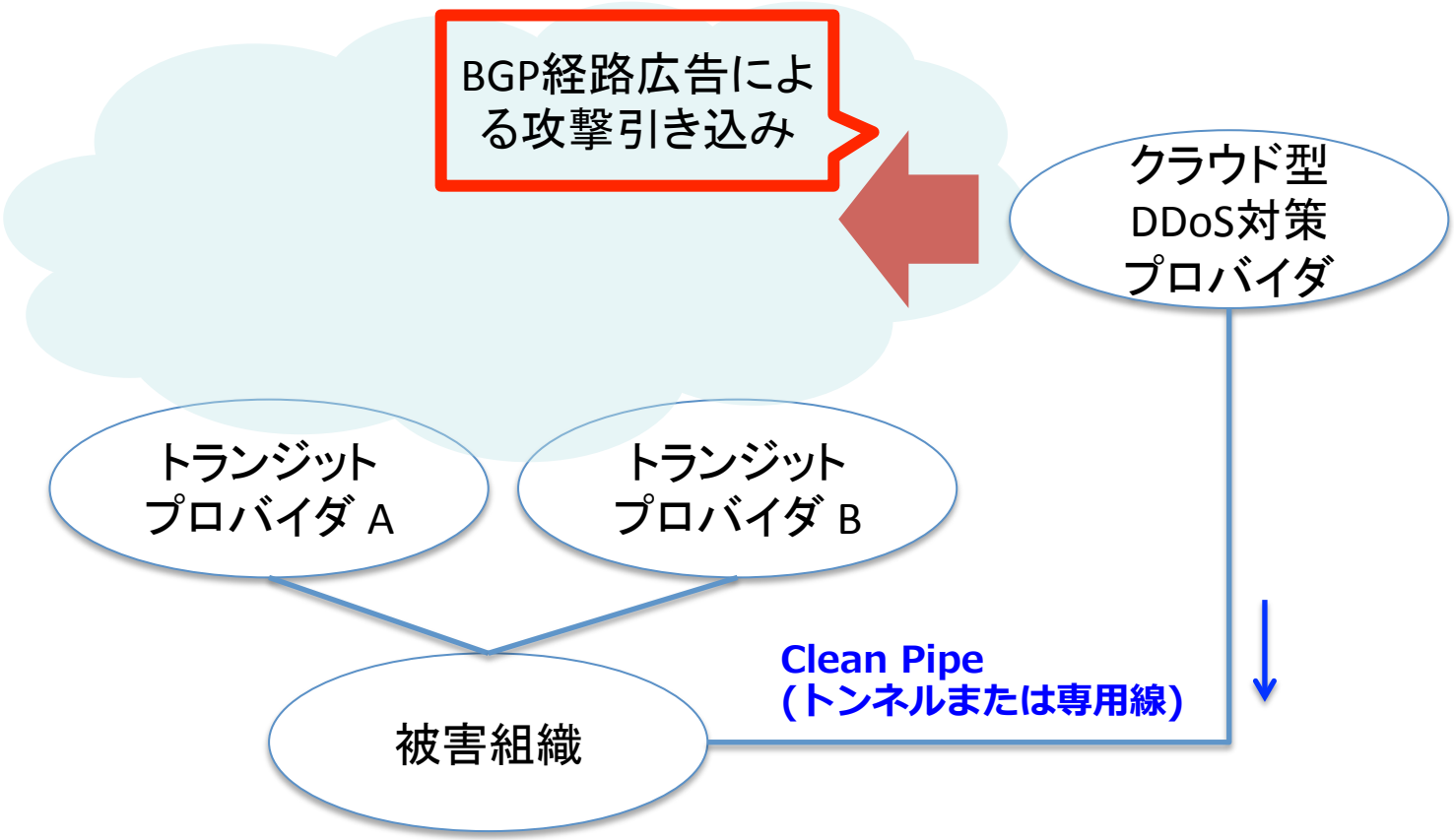
Amp攻撃やIoT BotによるDDoS攻撃の大規模化
(100Gbpsを超える攻撃)

上流回線の輻輳による被害を避けるには

1. クラウド型DDoS対策プロバイダ
2. トランジットプロバイダが提供するDDoS対策サービス

クラウド型DDoS対策プロバイダ

攻撃引き込みにはDNSやBGPが使われるが、ここでは、BGPによる攻撃引き込みに注目する。



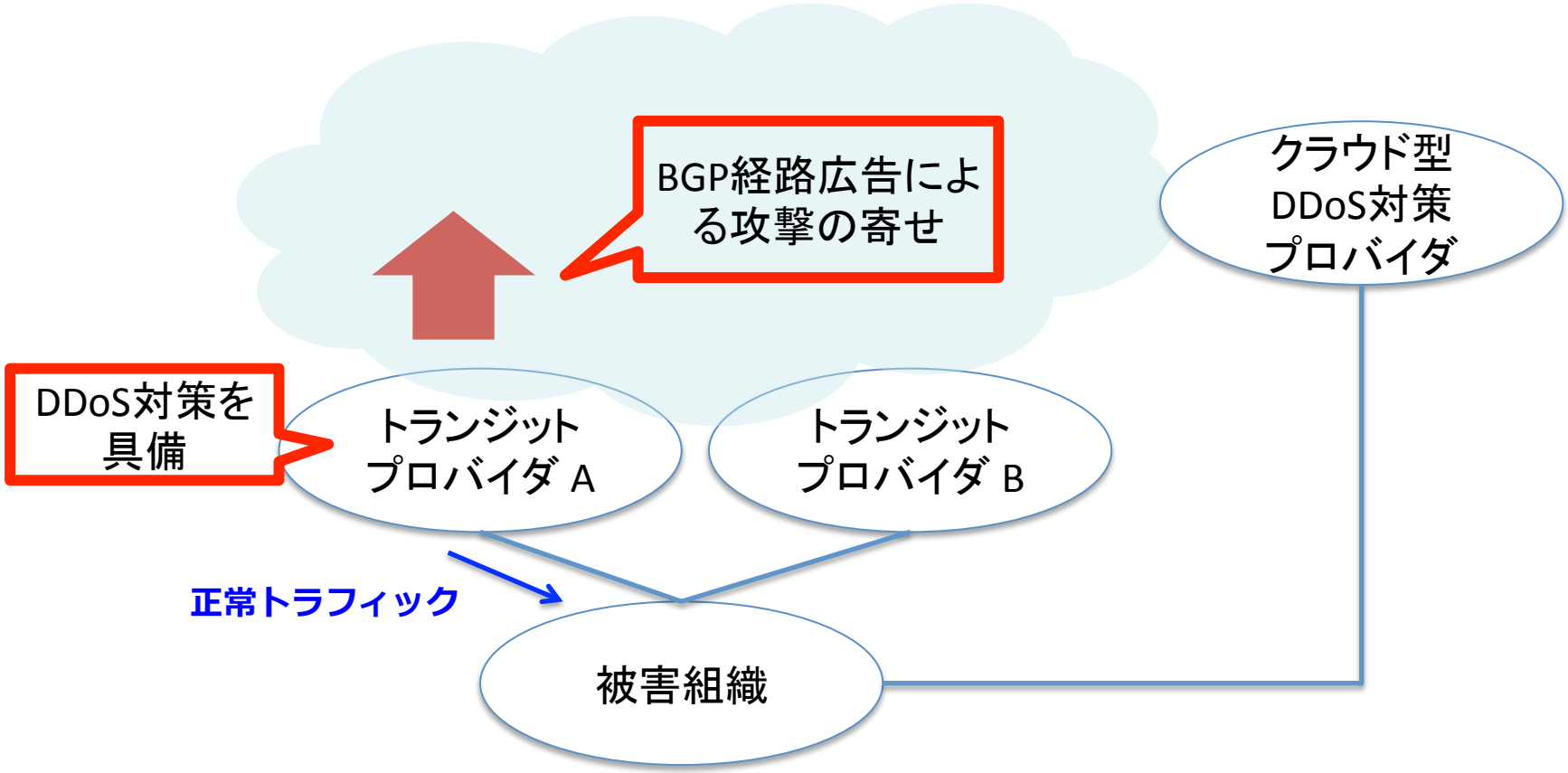
クラウド型DDoS対策プロバイダ

- 引き込み手法
  - 顧客のCIDRブロックよりもlongerなBGP経路を広報することによって、longest-match規則によって顧客のトラフィックの引き込みを行う
  - DDoS攻撃が終了した時には、longer経路を消去し、通常のルーティングに戻す

NOTE: トラフィックの常時引き込みのオプションがある対策プロバイダもあるが、クラウド上にスクラビングセンタがあることから引き込み時には遅延が印可されるため、顧客は常時引き込みを嫌う傾向がある

トランジットプロバイダが提供する DDoS対策サービス

トランジットプロバイダ自身がDDoS対策を具備しており、不正パケットを網内でドロップする



トランジットプロバイダが提供するDDoS対策サービス

- トラフィックの片寄せ
  - マルチホームしている場合、DDoS対策サービスが具備されているトランジットプロバイダに攻撃トラフィックを一時的に寄せて、集中的に対策ができるとよい
  - 攻撃トラフィックだけを片方のトランジットに確実に寄せるオペレーションとしては、先ほどと同様に、longer経路を広報する手法が一般的

NOTE: DDoS対策が必要な顧客は、冗長化でサービスを守るためにマルチホームをしているケースの方が多いと考えられる

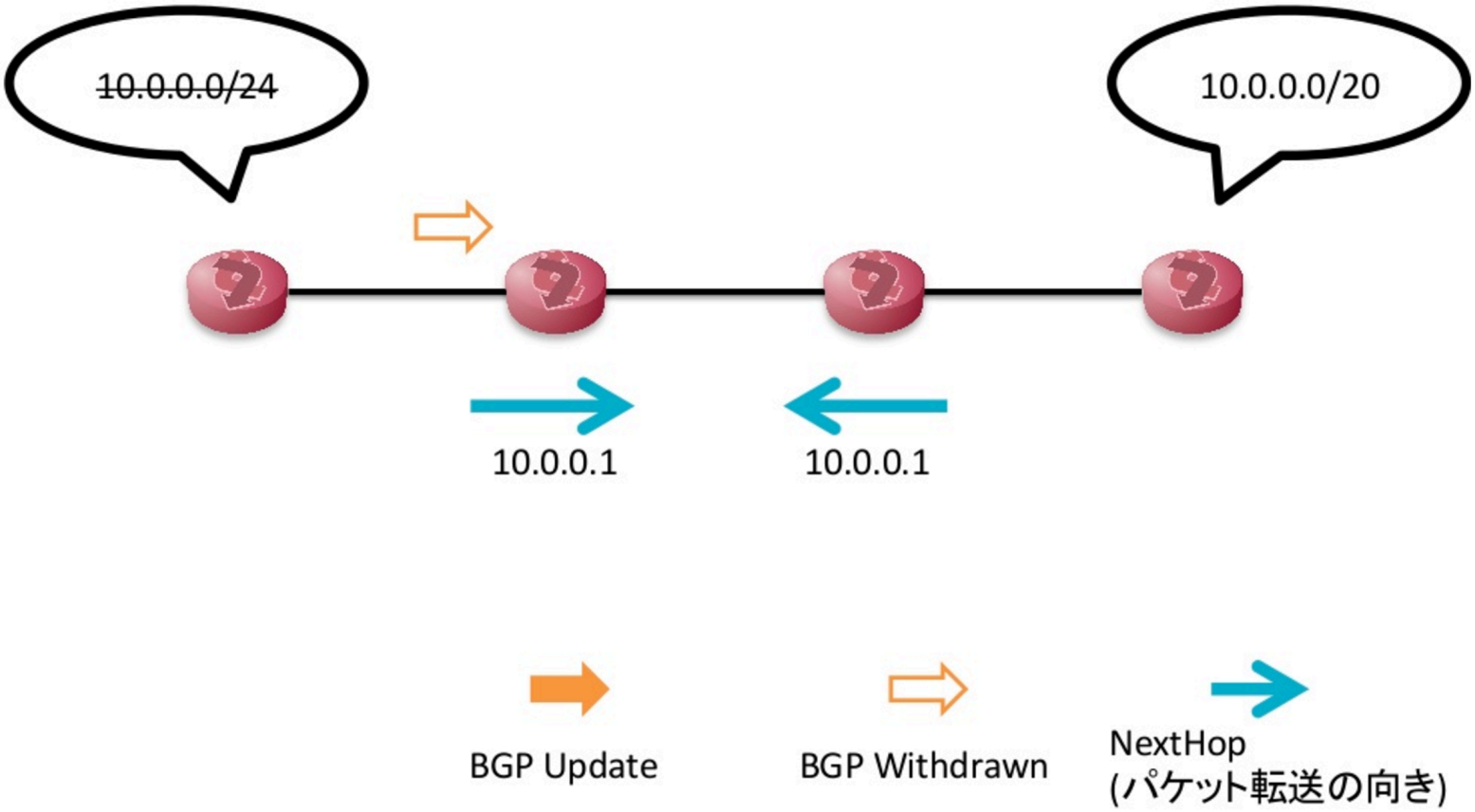
longer経路の憂鬱

- longer経路の広報(Update)および消去(Withdrawn)
  - インターネット上にBGP経路情報の伝搬が発生し、(どこかで)一時的なループを起こす可能性がある
    - パケットロスが発生する
  - ループの継続時間は、一般的に、

    BGP Withdrawn >> BGP Update

経路の消去(withdraw)は危険工程と心得よ！

BGP withdrawによるループ



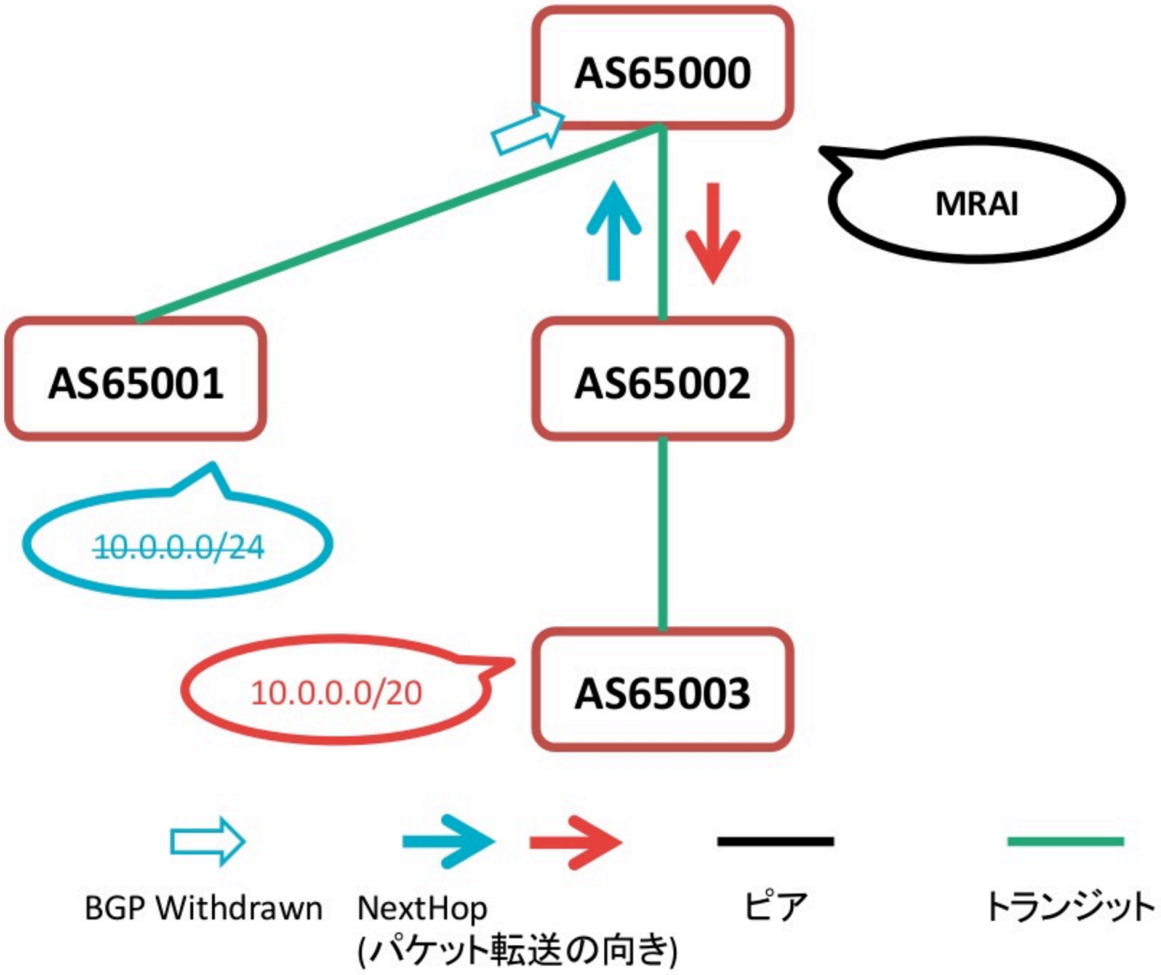
MRAI とBGP Withdrawn とパケットロス
http://codeout.hatenablog.com/entry/2016/08/22/215757
スペシャルサンクス: @codeout

MRAI

- MRAI: Minimum Route Advertisement Interval
  - 経路変化を一定時間バッファリングしてからネイバー(iBGP/eBGP)に送るしくみ
  - ベンダによってデフォルト値やふるまいが異なる
- メリット
  - 無駄なフラッピング経路広報の緩和
- デメリット
  - 経路伝搬時間の長期化

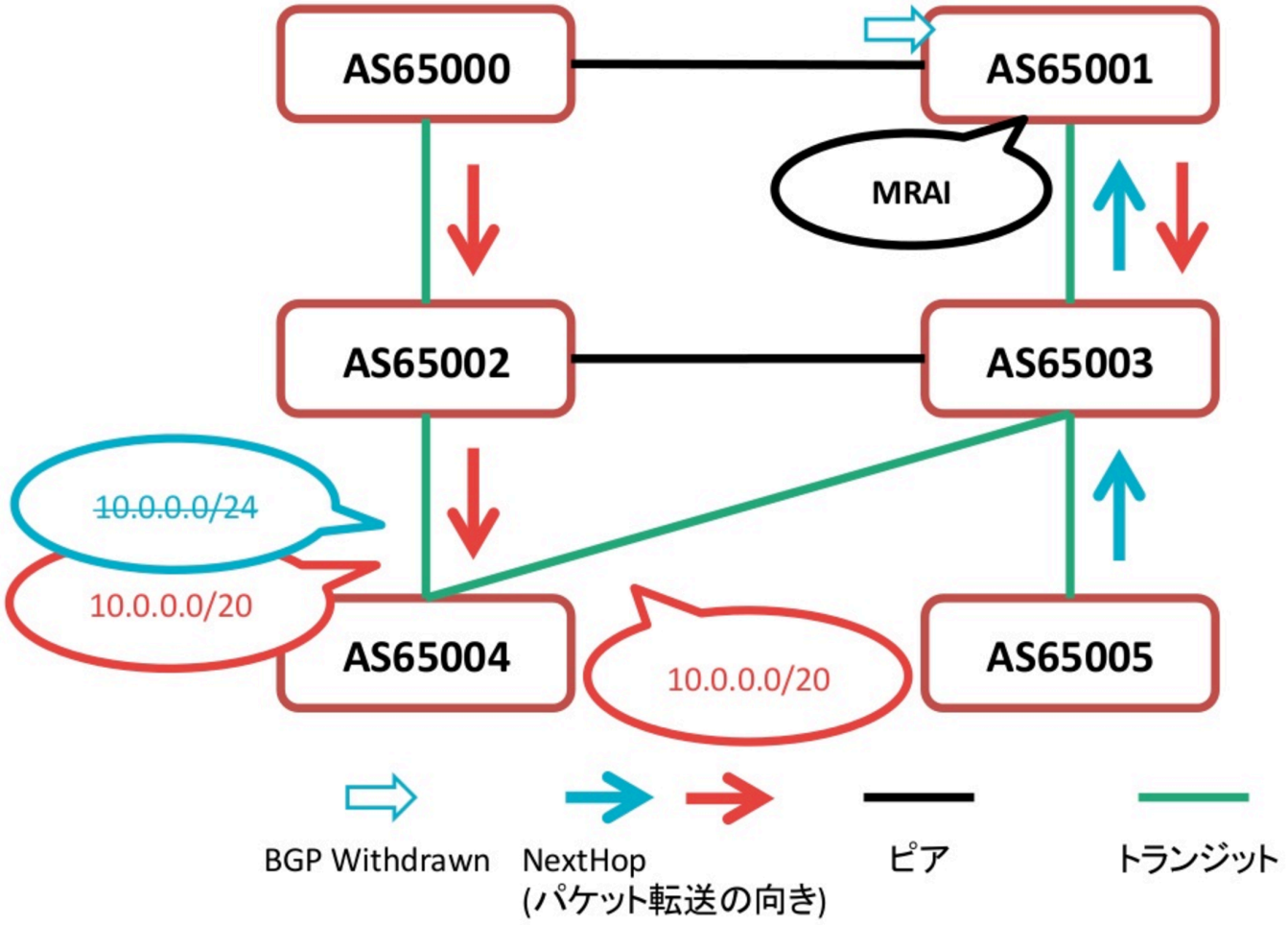
MRAI の代表的なふるまい
http://codeout.hatenablog.com/entry/2016/07/22/220601

MRAI とループ



MRAI とBGP Withdrawn とパケットロス
http://codeout.hatenablog.com/entry/2016/08/22/215757

遠くのASからの影響が残る



MRAI とBGP Withdrawn とパケットロス
http://codeout.hatenablog.com/entry/2016/08/22/215757

安全なwithdrawのために

- longer経路操作(update/withdraw)によるパケットロスの影響を計測した。
- 外部クラウドからのトレースにより原因を分析して、対策案の効果を評価した

1. トランジットプロバイダ型: 西塚
   - トランジットからCIDR経路に重ねて/24を広報
2. クラウド型:柴崎、篠宮
   - クラウド型のDDoS対策プロバイダから/24を広報

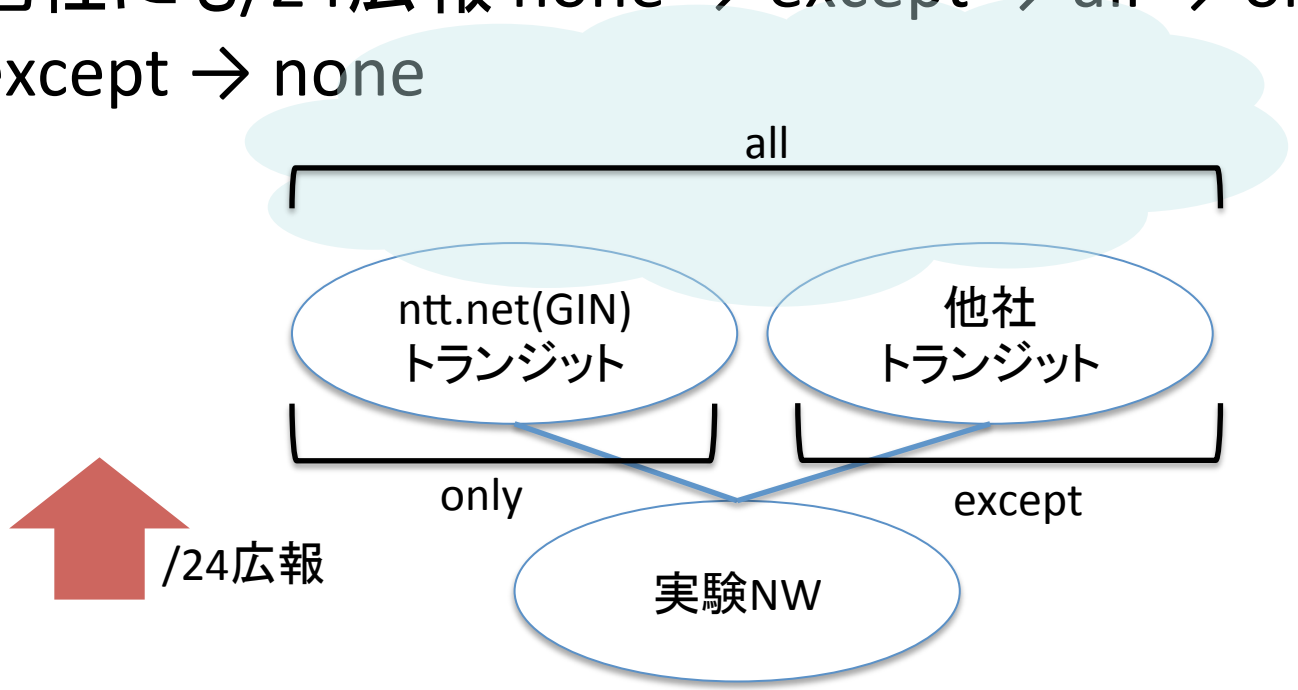
経路操作シナリオ

/24の広報方法

A: 無策 none → only → none

A’: 緩やかな撤退 none → only → asia-only → none

B: 他社にも/24広報 none → except → all → only → all → except → none



測定方法1: ping

実験NWから、ALEXA top 50 サイトにpingを打ち、ロスをカウント(ping応答しなかった24サイトは除いた)

www.2ch.net	AS13335	CLOUDFLARENET - CloudFlare_ Inc._ US
www.asahi.com	AS20940	AKAMAI-ASN1 Akamai International B.V._ US
www.nhk.or.jp	AS16625	AKAMAI-AS - Akamai Technologies_ Inc._ US
www.apple.com	AS16625	AKAMAI-AS - Akamai Technologies_ Inc._ US
www.dmm.co.jp	AS9607	BBTOWER BroadBand Tower_ Inc._ JP
www.livedoor.biz	AS17707	DATAHOTEL-JP AS for DATAHOTEL_ which is one of iDC in Japan_ JP
www.livedoor.com	AS17707	DATAHOTEL-JP AS for DATAHOTEL_ which is one of iDC in Japan_ JP
www.blogimg.jp	AS17707	DATAHOTEL-JP AS for DATAHOTEL_ which is one of iDC in Japan_ JP
www.biglobe.ne.jp	AS2518	BIGLOBE BIGLOBE Inc._ JP
www.goo.ne.jp	AS2914	NTT-COMMUNICATIONS-2914 - NTT America_ Inc._ US
www.live.com	AS8075	MICROSOFT-CORP-MSN-AS-BLOCK - Microsoft Corporation_ US
www.yahoo.co.jp	AS23816	YAHOO Yahoo Japan Corporation_ JP
www.wikipedia.org	AS14907	WIKIMEDIA - Wikimedia Foundation Inc._ US
www.twitter.com	AS13414	TWITTER - Twitter Inc._ US
www.impress.co.jp	AS4694	IDC Yahoo Japan Corporation_ JP
www.msn.com	AS8068	MICROSOFT-CORP-MSN-AS-BLOCK - Microsoft Corporation_ US
www.so-net.ne.jp	AS2527	SO-NET So-net Entertainment Corporation_ JP
www.sakura.ne.jp	AS9371	SAKURA-C SAKURA Internet Inc._ JP
www.google.co.jp	AS15169	GOOGLE - Google Inc._ US
www.google.com	AS15169	GOOGLE - Google Inc._ US
www.blogspot.jp	AS15169	GOOGLE - Google Inc._ US
www.nifty.com	AS2510	INFOWEB FUJITSU LIMITED_ JP
www.rakuten.co.jp	AS16625	AKAMAI-AS - Akamai Technologies_ Inc._ US
www.hatena.ne.jp	AS9370	SAKURA-B SAKURA Internet Inc._ JP
www.facebook.com	AS32934	FACEBOOK - Facebook_ Inc._ US
www.japanpost.jp	AS22822	LLNW - Limelight Networks_ Inc._ US

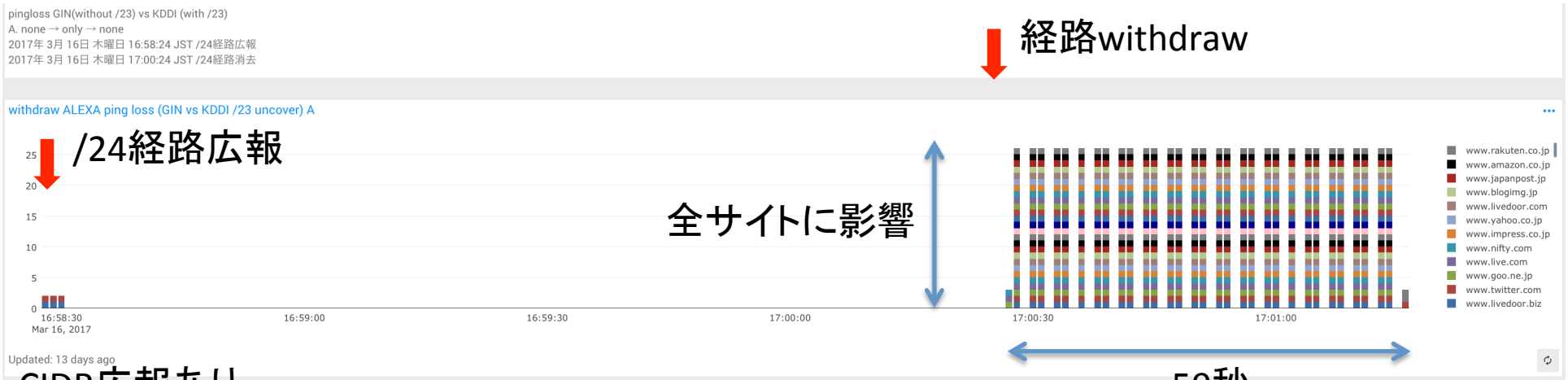
測定方法2: traceroute

各クラウドから実験NWに向けてtracerouteを打って、
hop, AS-PATH, latency を調べた

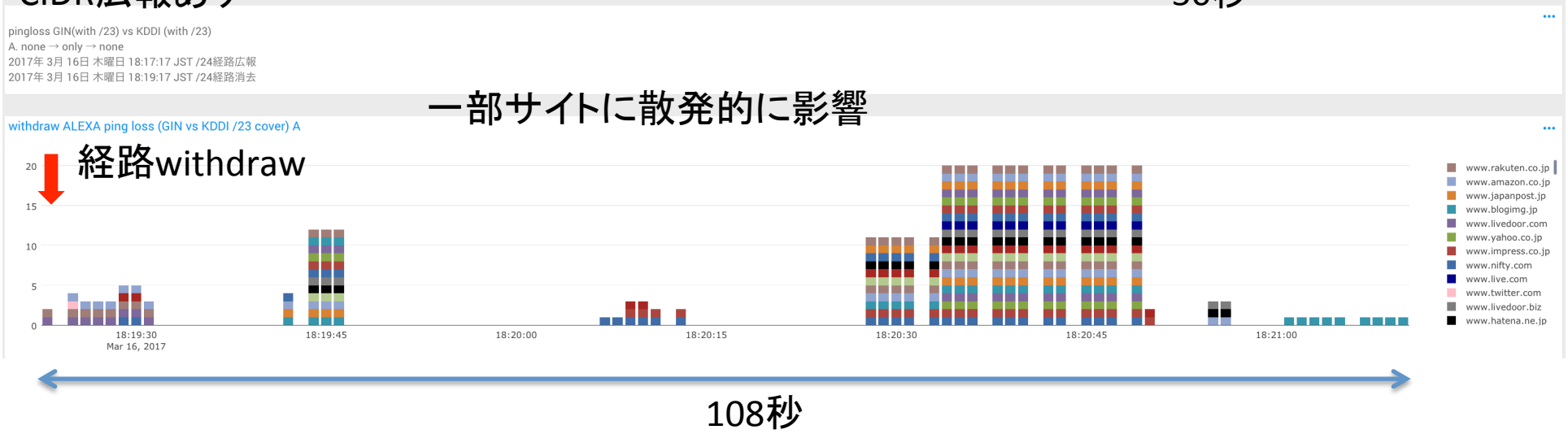
webarena	JP	AS2514
biglobe	JP	AS2518
cloudn	JP	AS4713
nifty	JP	AS2510
kddi	JP	AS9597
iij	JP	AS2497
sakura_cloud	JP	AS7684
sakura_vps	JP	AS9371
aws_verginia	US	AS16509
aws_california	US	AS16509
aws_ireland	IE	AS16509
aws_singapore	SG	AS16509
aws_japan	JP	AS16509

- GINからのCIDR広報の有無の影響
  - 縦軸: packet lossのあったサイト数

CIDR広報なし



CIDR広報あり

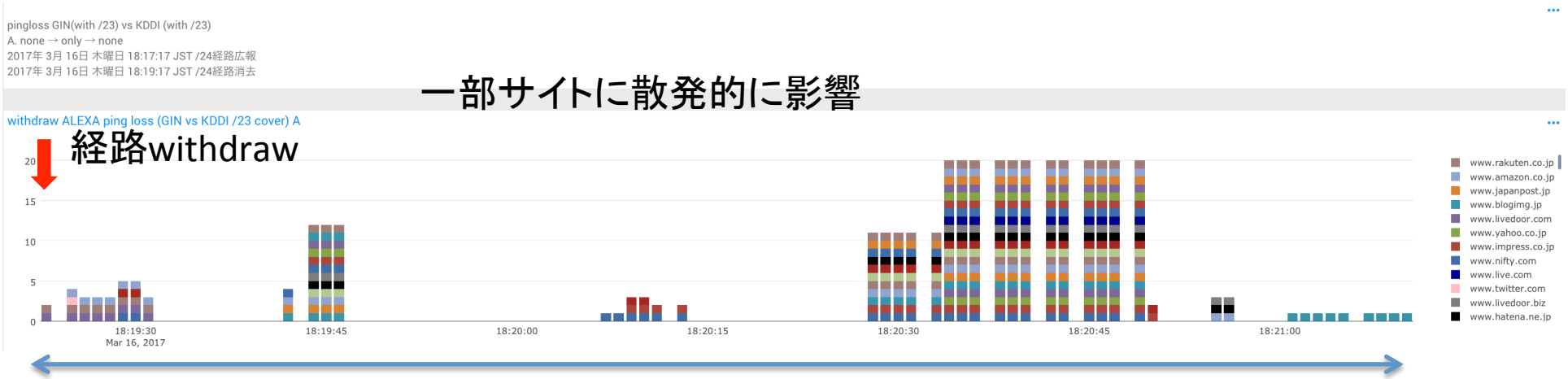


一部サイトに散発的に影響

108秒

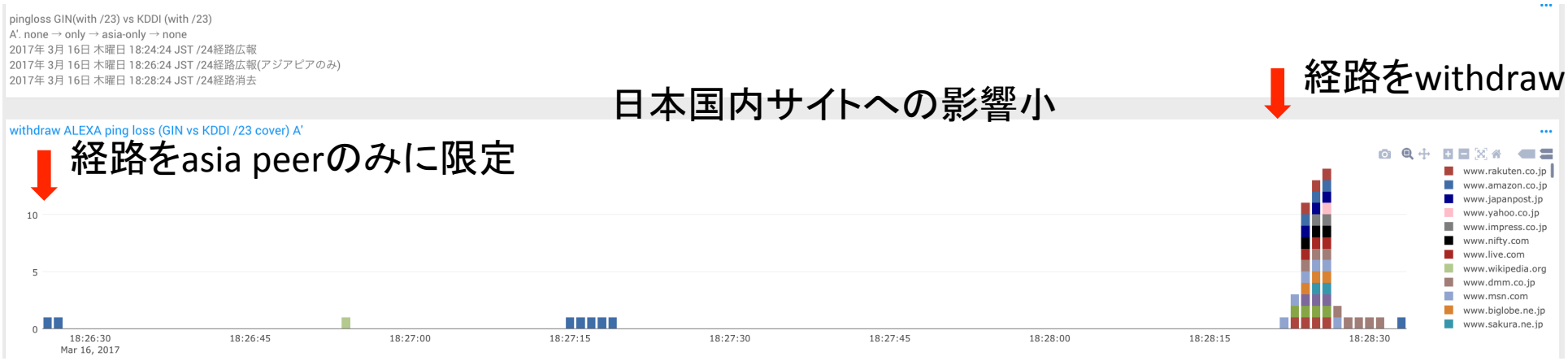
- /24広報範囲の段階的な撤退
  - 縦軸: packet lossのあったサイト数

一気にwithdraw(前ページ下段再掲)



一部サイトに散発的に影響

108秒



日本国内サイトへの影響小

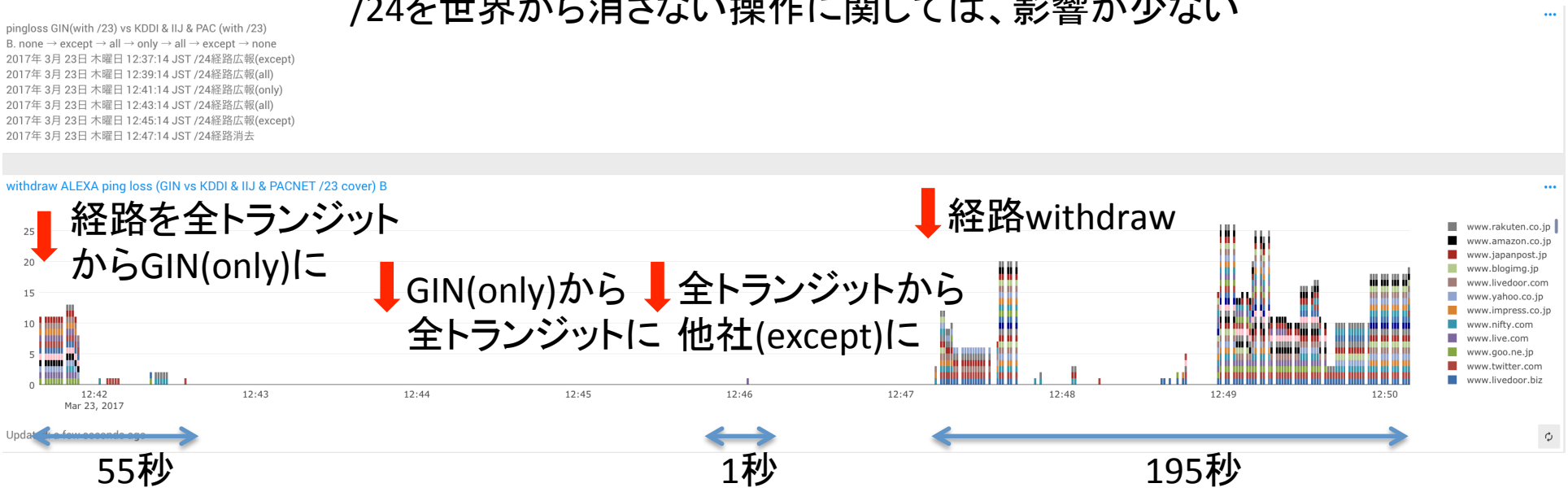
経路をwithdraw

54秒

12秒

- 他トランジットにも/24広報
  - 縦軸: packet lossのあったサイト数

/24を世界から消さない操作に関しては、影響が少ない

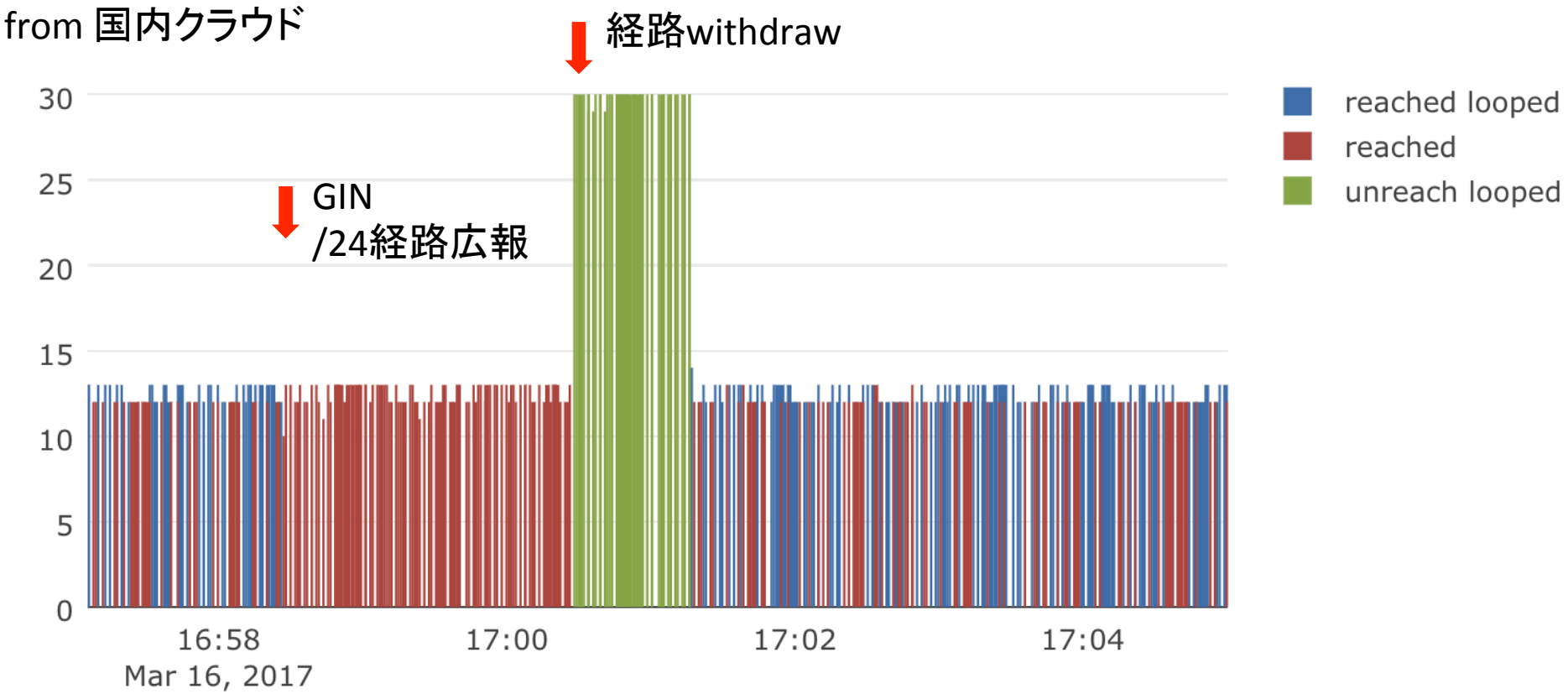


traceroute: hop数(国内から)

reached: 宛先に到達

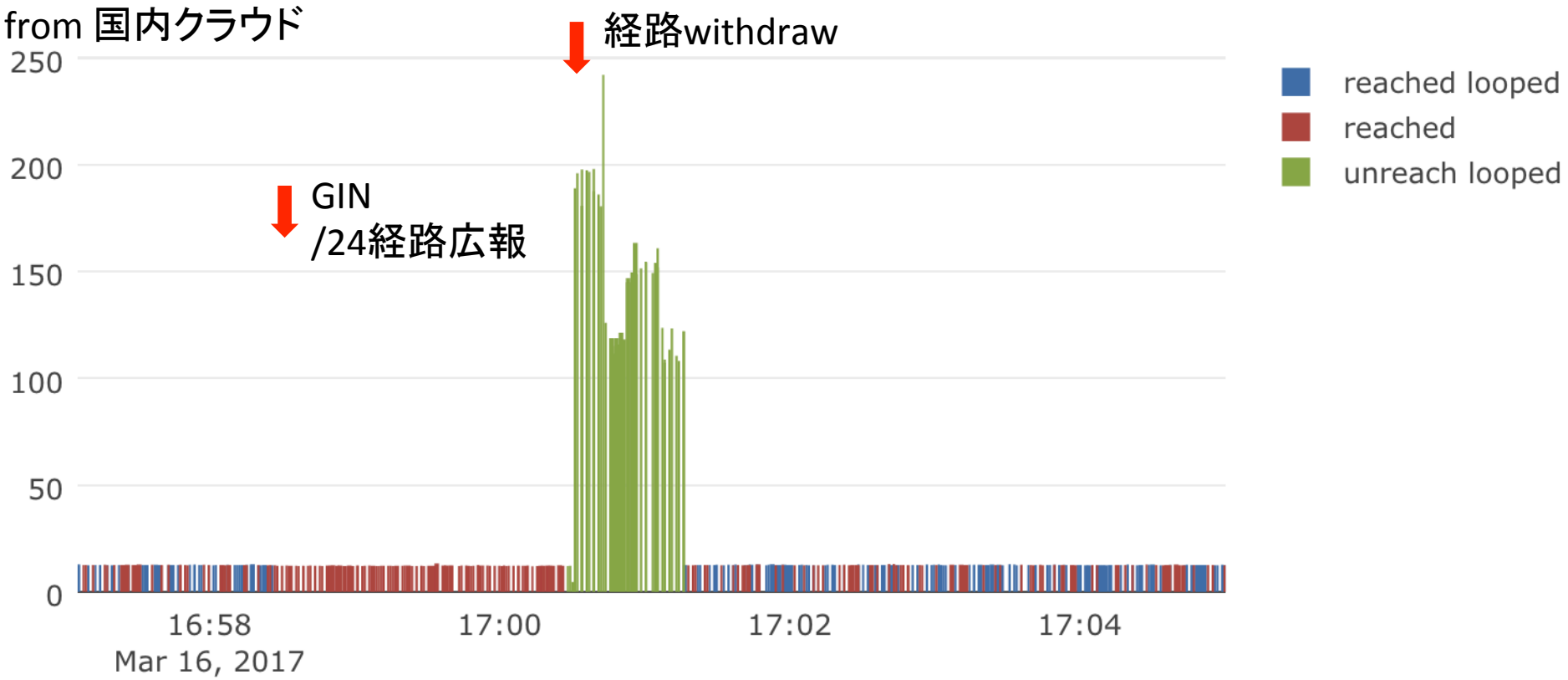
unreach looped: 宛先に到達せず、途中に重複IPを観測

reached looped: 宛先に到達しているが、途中に重複IPを観測(ECMPの影響)



traceroute: latency(国内から)

海外ASに引っ張られ、海を渡った先でループしている
(非到達の場合の latency は 30hop目(最大hop数)の値)

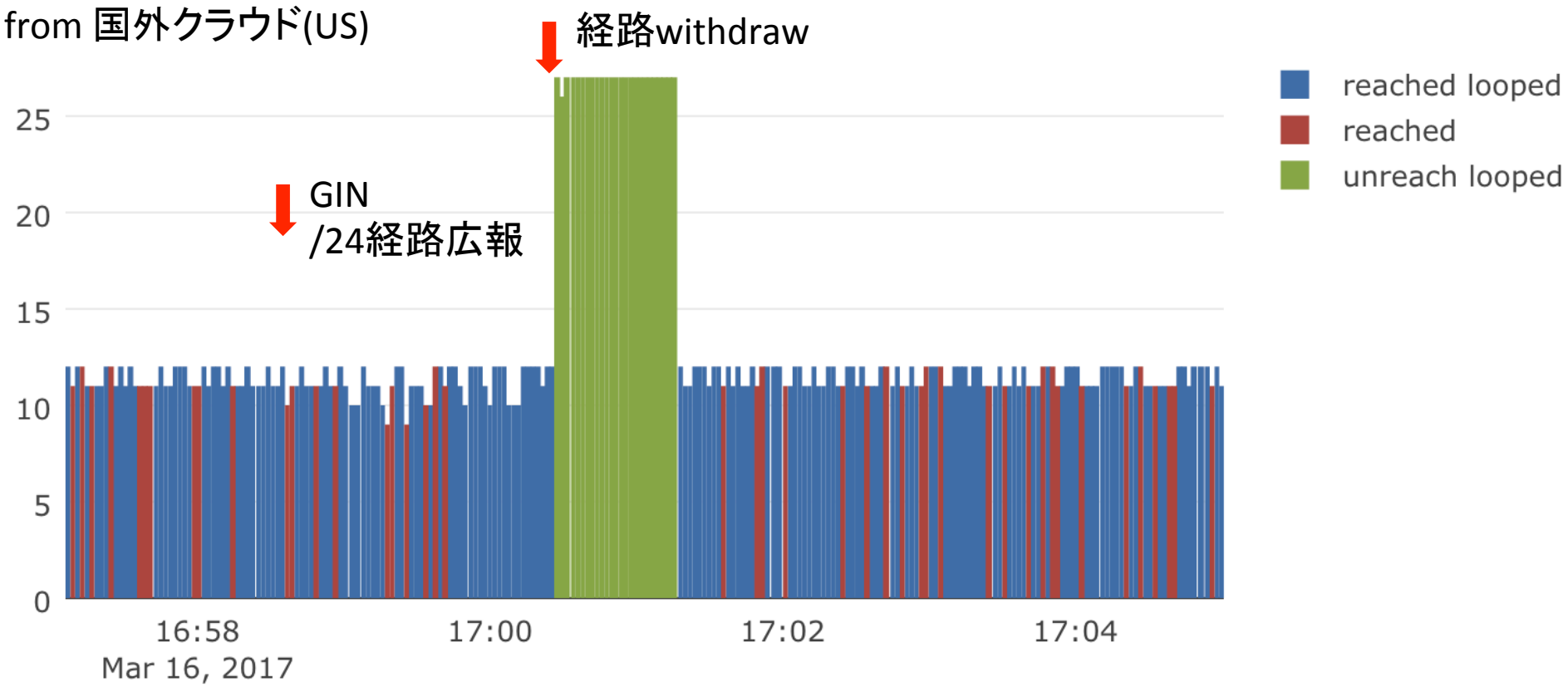


traceroute: hop数(国外から)

reached: 宛先に到達

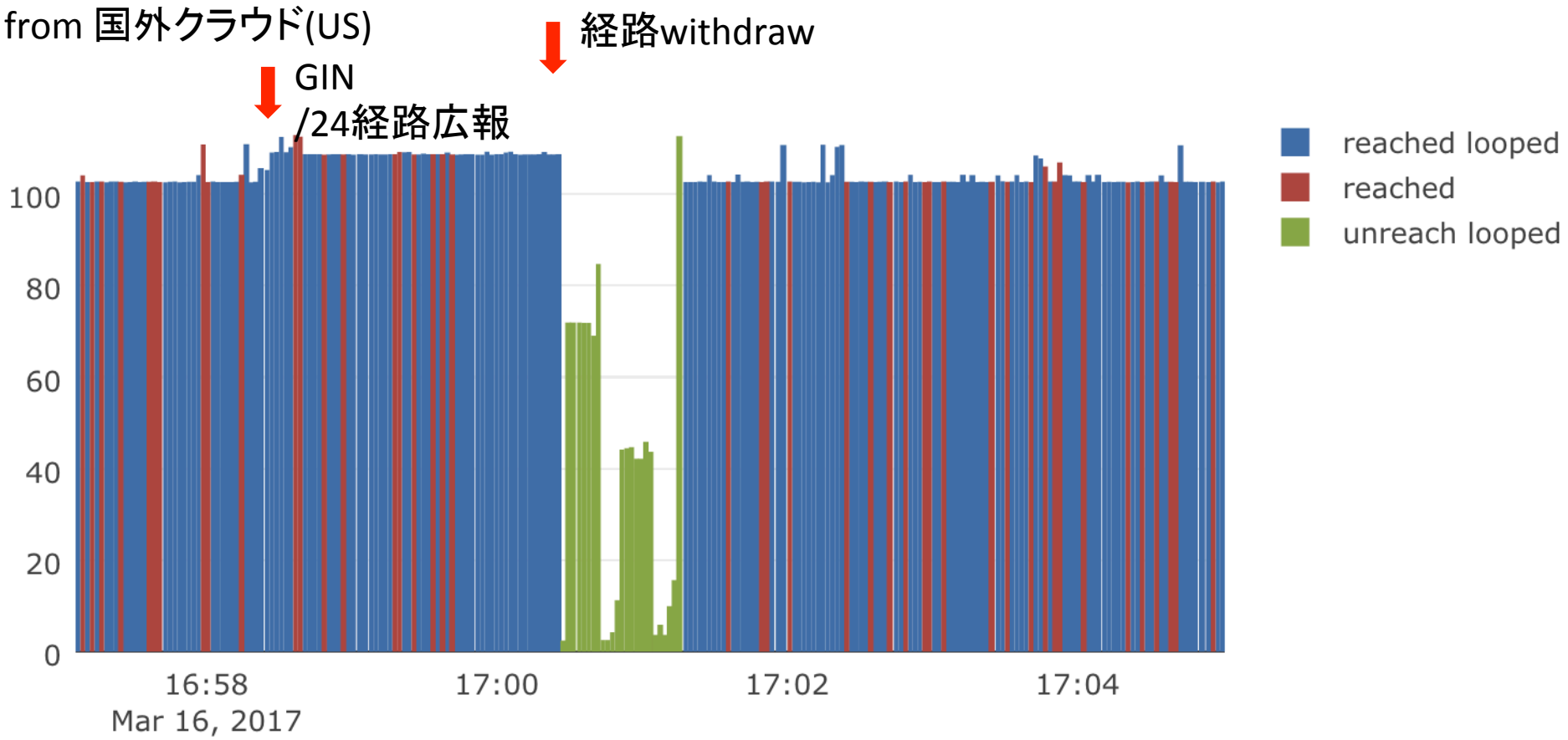
unreach looped: 宛先に到達せず、途中に重複IPを観測

reached looped: 宛先に到達しているが、途中に重複IPを観測(ECMPの影響)



traceroute: latency(国外から)

- 国外でループしているため、30hop目のlatencyの値は、逆に小さくなる
- 山が分かれているのは、ループ箇所の移り変わりによる(MRAIによる30秒間隔の可視化)



まとめ

- CIDR広報の効果
  - CIDR同士の比較で引きが強ければ影響が緩和される
  - 影響範囲は少なくなるが、全体の影響時間が短くなる効果は見られなかった
- 段階的に広報範囲を狭めることの効果
  - 影響範囲を狭めるとともに、段階ごとの影響時間が短くなった
- 他トランジットへのlonger経路の広報の効果
  - longer経路がある状態を通常運用状態と見なせば、ループの発生を一番抑えることができた(ただし、ゼロにはできない)
  - AS prependによって、さらに影響を少なくできる(篠宮さんの資料参照)