「どんなテストしてる?」BoF

# ルータをいじめる話

さくらインターネット研究所　大久保 修一
ohkubo@sakura.ad.jp

# ルータと私

SAKURA Internet

- 2003年～2009年
  - AS9370, AS9371 Backbone担当
  - ▽検証＋導入してきたもの
  - ルータ、スイッチ、回線、ルートリフレクタ
  - DNSサーバ(ANS)、ロードバランサ
- 2009年～2011年
  - さくらインターネット研究所
  - IPv4/v6共存技術(トランスレータ,6RD,他)
  - 仮想スイッチ/アプライアンス
- 2011年～
  - 「さくらのクラウド」インフラ開発を担当
  - L2/L3、InfiniBand、iSCSIストレージ、VPN機器

# サービス用機器導入～提供の流れ

SAKURA Internet

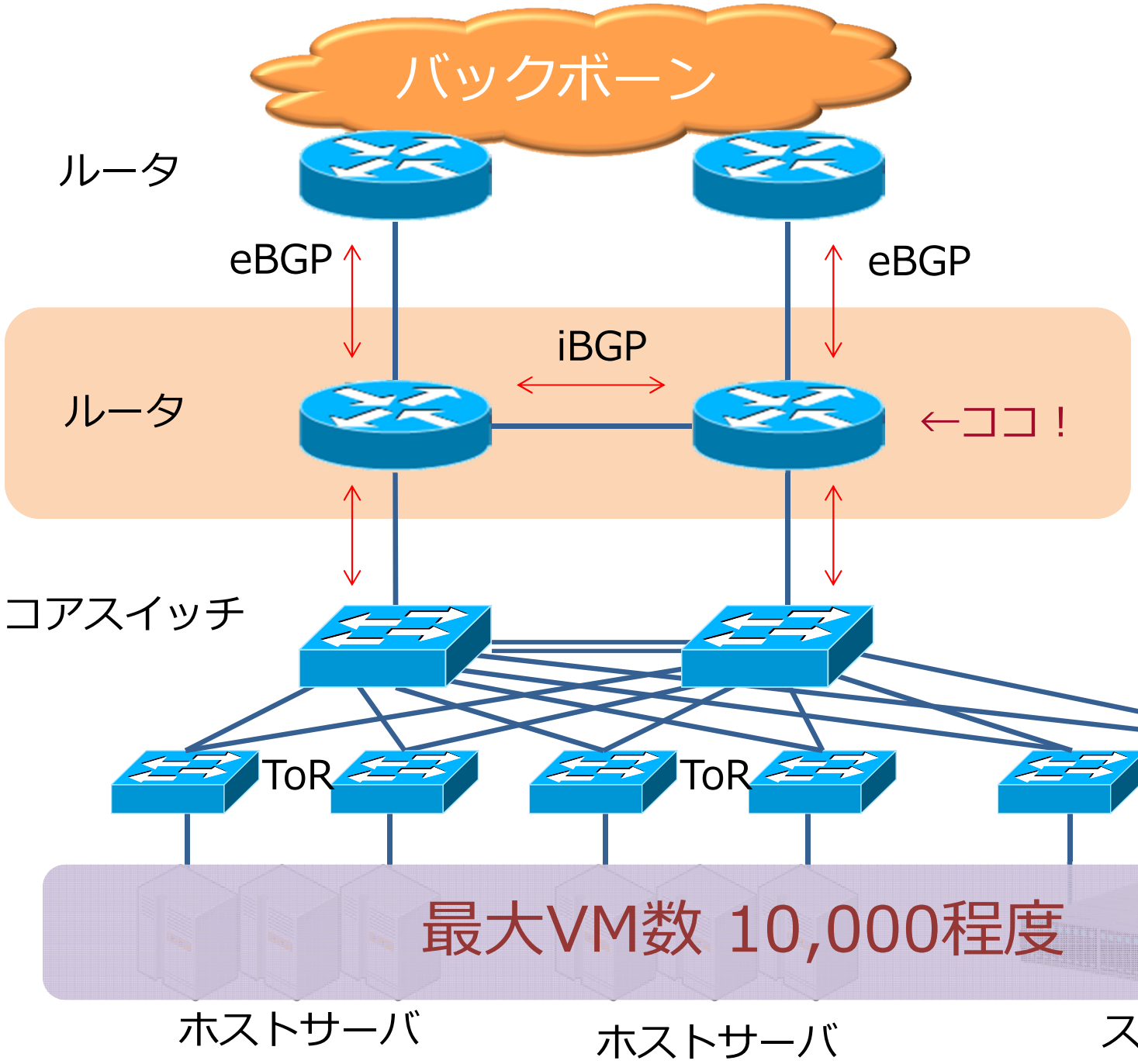
1. サービス仕様策定、システム全体設計
2. 各機器の想定機能要件、想定収容要件洗い出し
3. RFP
   - 要件を満たしそうで価格感がマッチする機種をいくつか選定
   - メーカさん、ベンダさんに相談したり
4. 機器検証
   - 検証項目洗い出し、検証構成検討
   - 検証機材をお借りしたり、ベンダさんのラボに訪問したり
5. 導入判定、購入
   - バグ修正、追加インプリの条件をお願いすることも
6. 構築、サービスイン前試験

場合によっては
設計を見直し

# ケーススタディ
# ～クラウド(IaaS)収容用途のルータ～

# 某クラウドサービスのネットワーク構成

SAKURA Internet




上流とはプライベートASを用いたBGP接続

VRRPを用いた冗長化

3,500VLANくらい

# 要件、検証項目のリストアップ

SAKURA Internet

機能要件

性能、スケーラビリティ要件

運用要件

# クラウド収容ルータの例：機能/収容要件

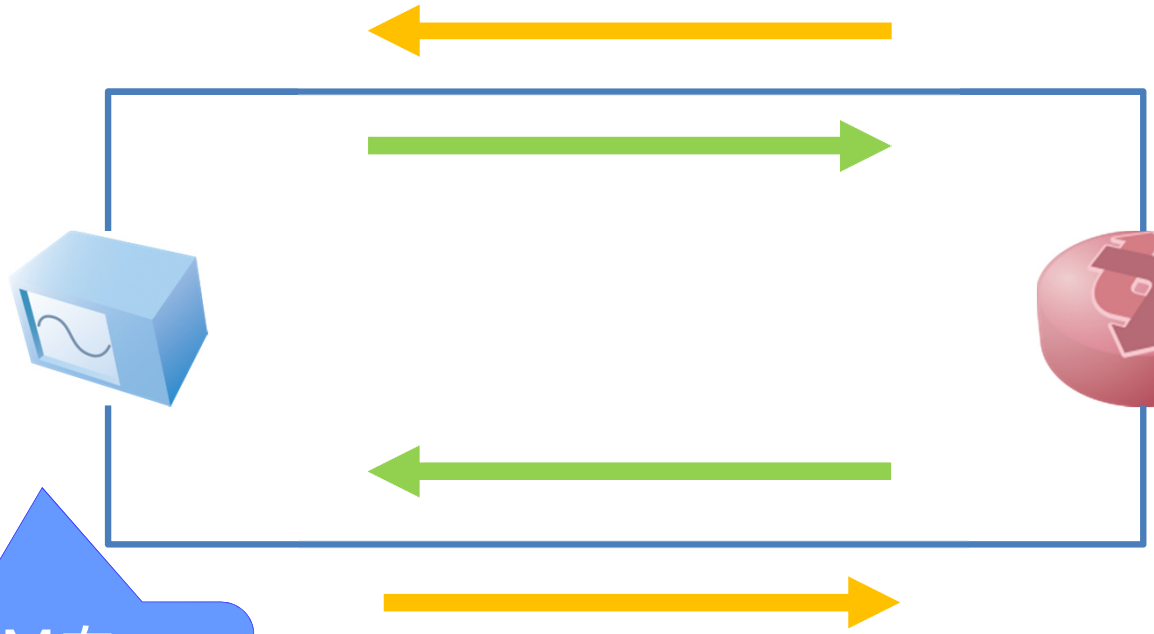
SAKURA Internet

- 機能要件
  - ルーティングプロトコル: OSPF, BGP IPv4/IPv6デュアル
  - 冗長化プロトコル: VRRP (もしくは同等のもの)
  - その他: IPv6 RA, VLAN単位での帯域制限, ACL
- 性能、スケーラビリティ要件
  - スループット: 10Gbps以上、そこそこMpps
  - SVI: 2,000以上
  - VRRP数: 2,000以上
  - ARP/NDPエントリ数: 12,800以上
- 運用要件
  - 管理、監視しやすいか？消費電力・サイズ・ファシリティ面
  - 大量のConfigをいれてもサクサク動くか？

SAKURA Internet

## ルータ単体試験

2,000ユーザ分の
設定を投入



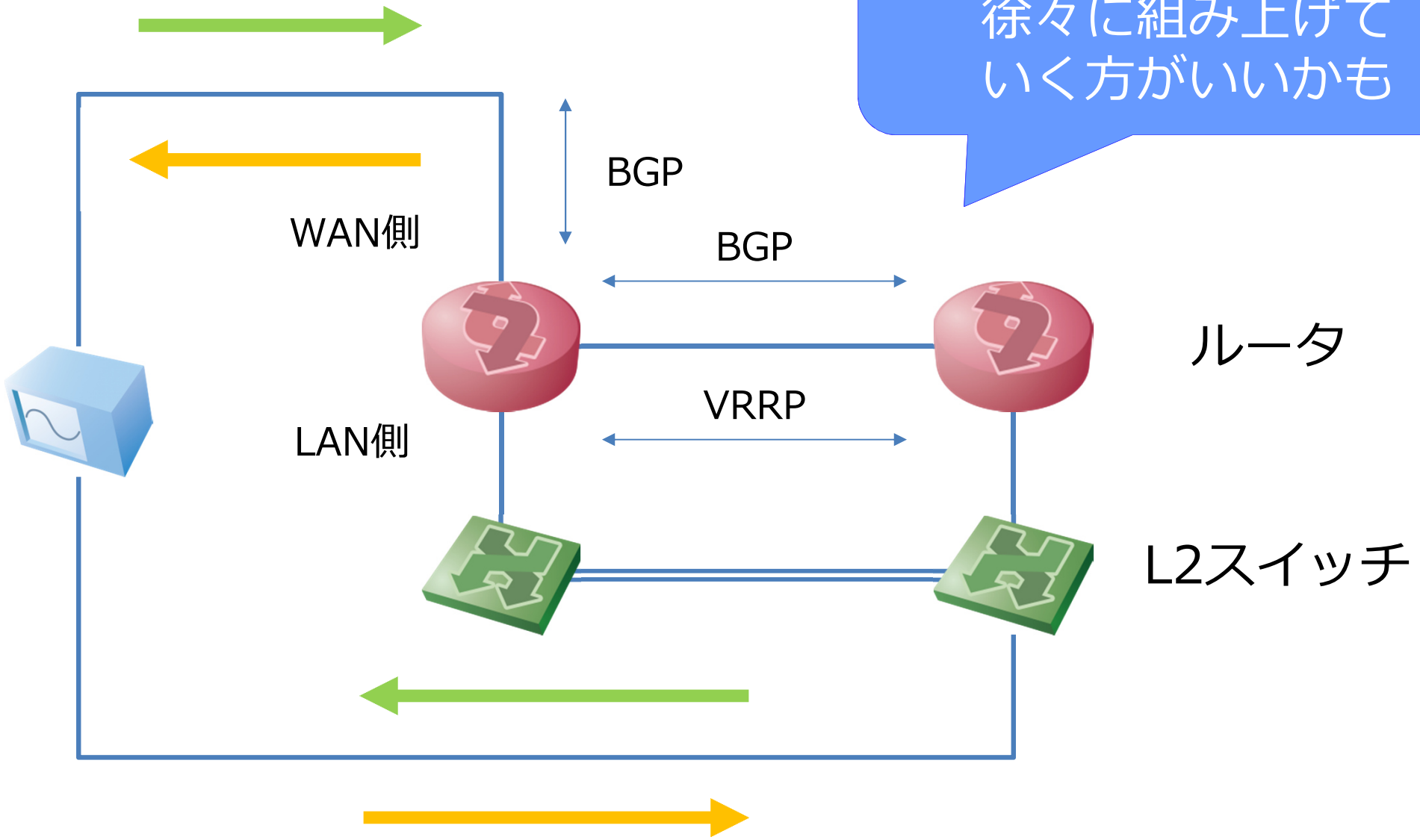

12,800VMを
エミュレーション

- ユーザ→インターネット向け
- インターネット→ユーザ向け

※ 上りと下りでは別世界

## 結合試験

単体でテストしておき、
徐々に組み上げて
いく方がいいかも

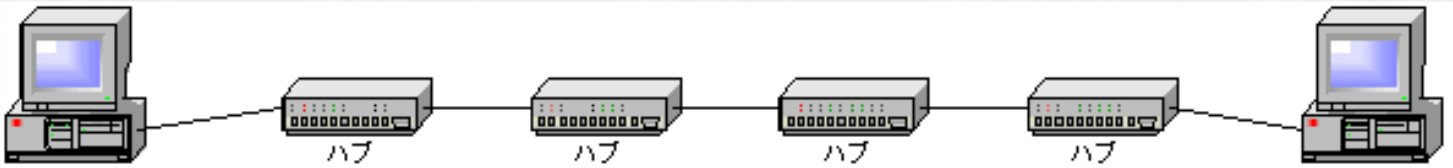


# ショートパケットワイヤレートの不条理

SAKURA Internet

- 64バイトショートパケットワイヤレートの性能にこだわる無意味さ(10GbEだと14.8Mppsとか)
- そもそもなんで64バイトなんだっけ？

ご承知のとおり、**イーサネットの媒体アクセス制御方式は CSMA/CDです**。そのCSMA/CD で考えなければいけない重要なことは、**「最悪 の条件の下ですべての端末が衝突を検出」**できなければいけません。最悪な条件とは、一番はなれた2台の端末間での通信のときです。10BASE-Tイーサネットでは、4台のハブを経由する場合でしかも各 ケーブルが100mのときが最悪の条件です。




http://www.n-study.com/network/minframe.htm より

- たまたま100mのコリジョン検出のために最小フレーム長が64バイトに決まっていた。
- 200m(128B)だったら7.4Mppsの性能で済んでたかも

# ルータの弱いところを突く！

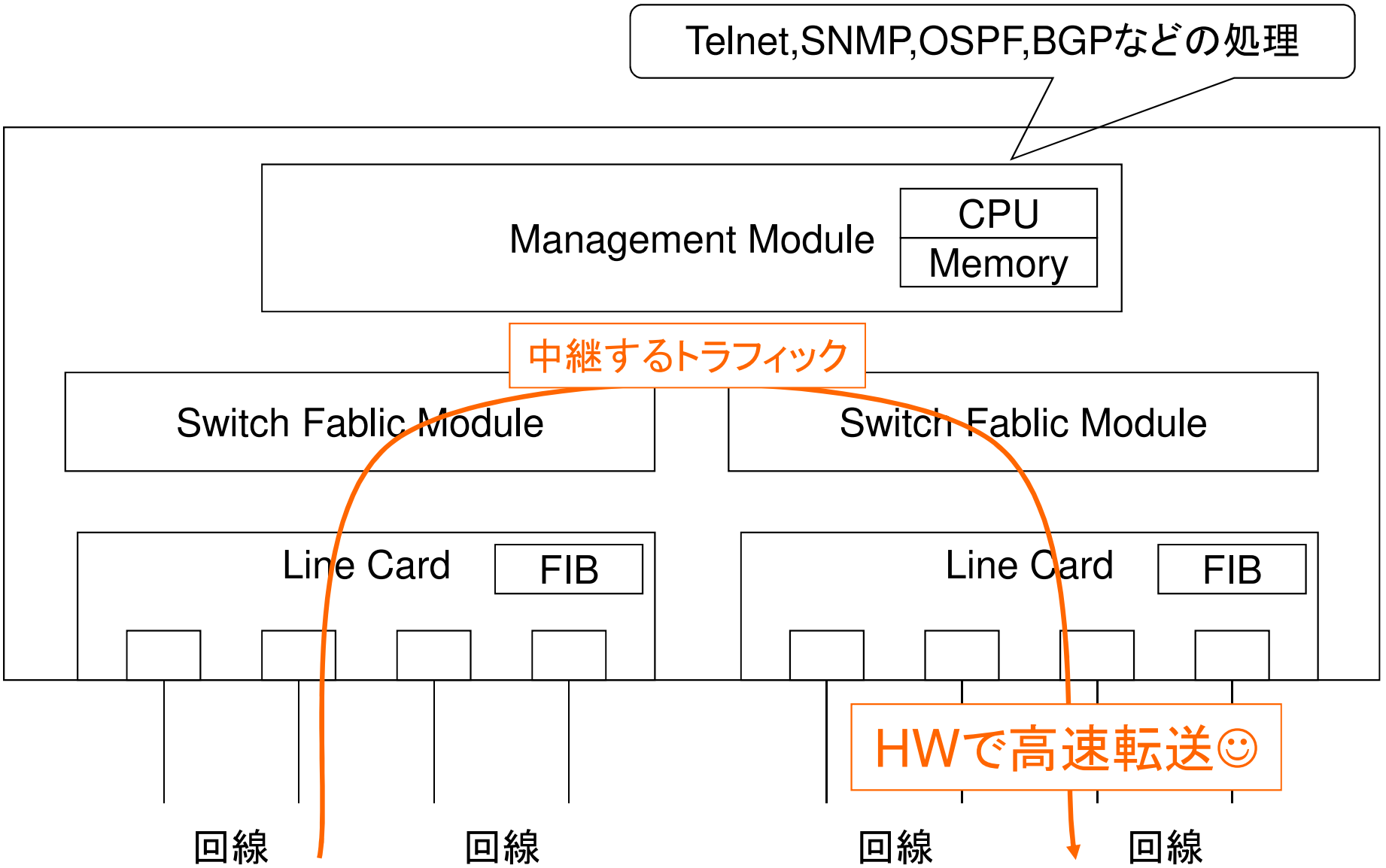
SAKURA Internet

- ルータ宛てのDoSアタック
  - UDPフラッド、ICMP echo request
  - SYNアタック(22番/23番/179番ポート)
- CPUエスカレーションされるパケット
  - TTL=1のパケット(TTL=0でICMP Time Exceeded生成)
  - ARP/NDP解決不能な宛先のパケット
  - トラフィックを流している状態でclear ip arp/clear ipv6 neighborsする
- 上記を10Gワイヤレートでぶつける

ルータの作り込まれ具合、
叩かれ度合があからさまに

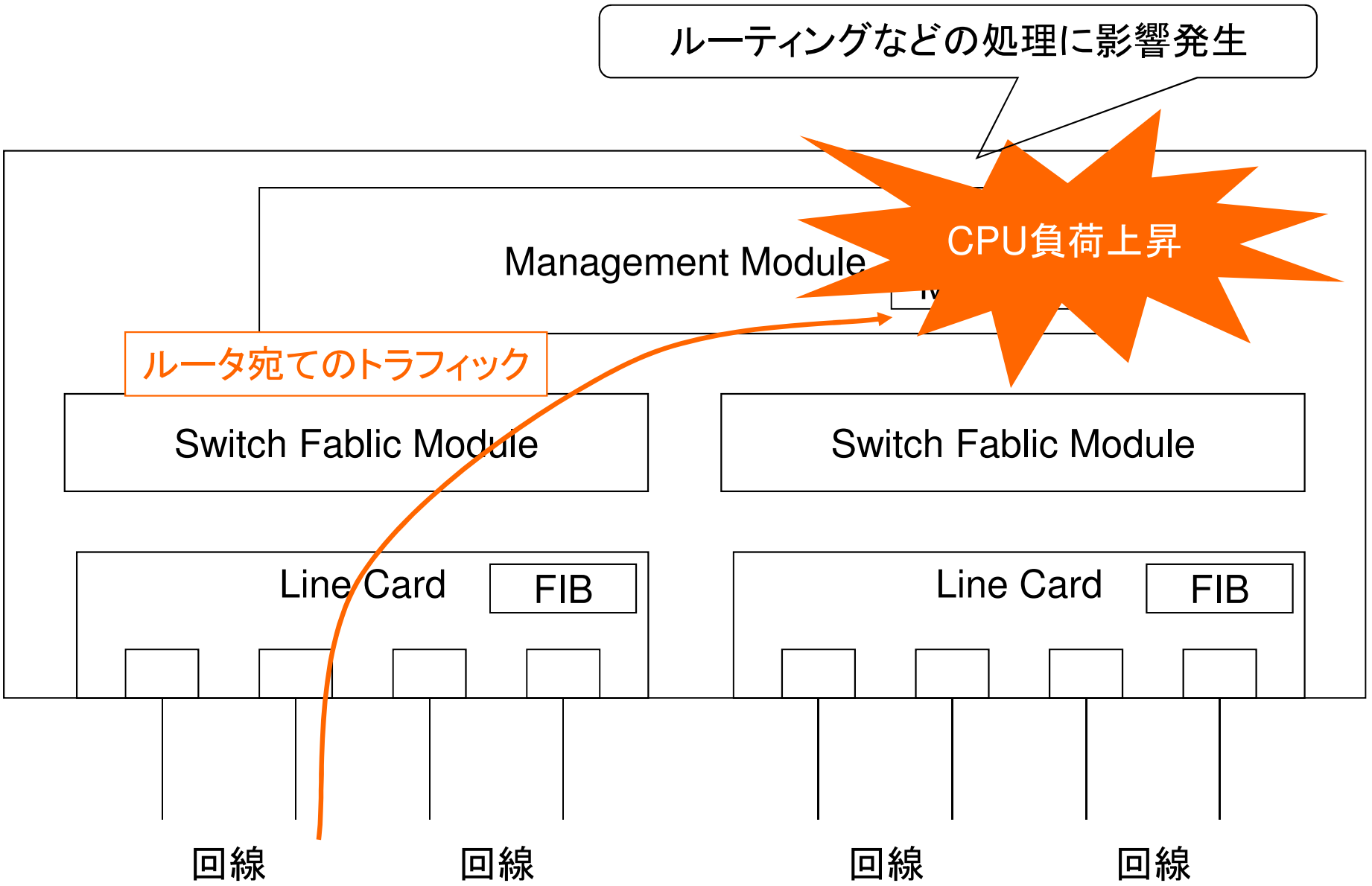
# ルータの弱いところを突く！

SAKURA Internet




http://irs.ietf.to/past/docs_20090521/ より

# ルータの弱いところを突く！




http://irs.ietf.to/past/docs_20090521/ より

# 発生した不安定事象の例

SAKURA Internet

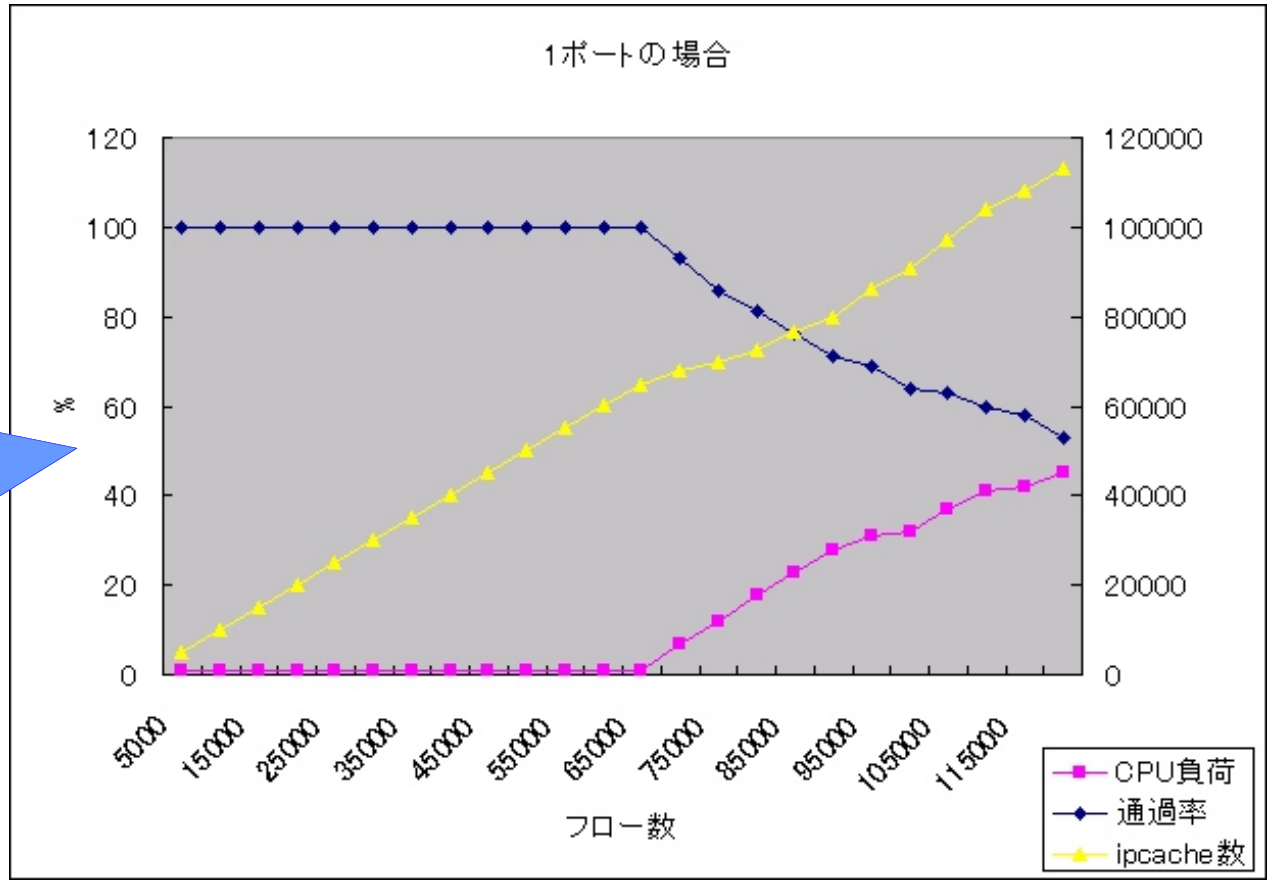
- VRRP Act/Stbフラップ
  - Standby側がVRRP Hello受信時に取りこぼし
- CLI操作が不能/重くなる
  - 23番/179番ポート宛てのDoSアタック
  - ARP/NDPエントリを大量に持った状態
- BGPピア断/OSPF Neighborダウン
  - clear ipv6 neighborsを実行
- LACPダウン

# ルータの弱いところを突く！

SAKURA Internet

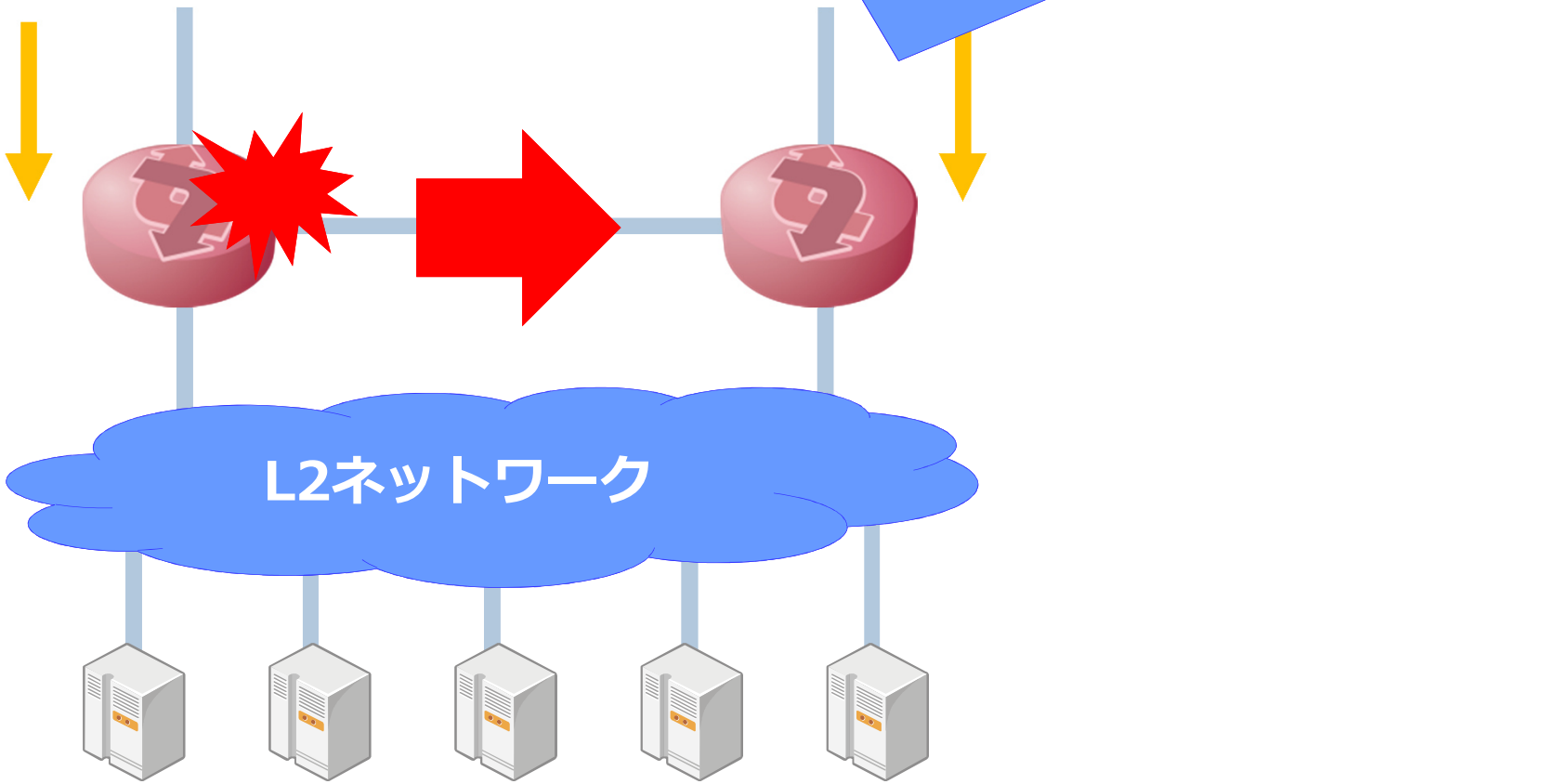
- ランダムな宛先IPアドレスのパケット
  - フローベースのルータだとパフォーマンス低下が発生
  - 最近のルータだとほぼ大丈夫なはず。

昔、散々痛い目にあったので、今でもやってます。

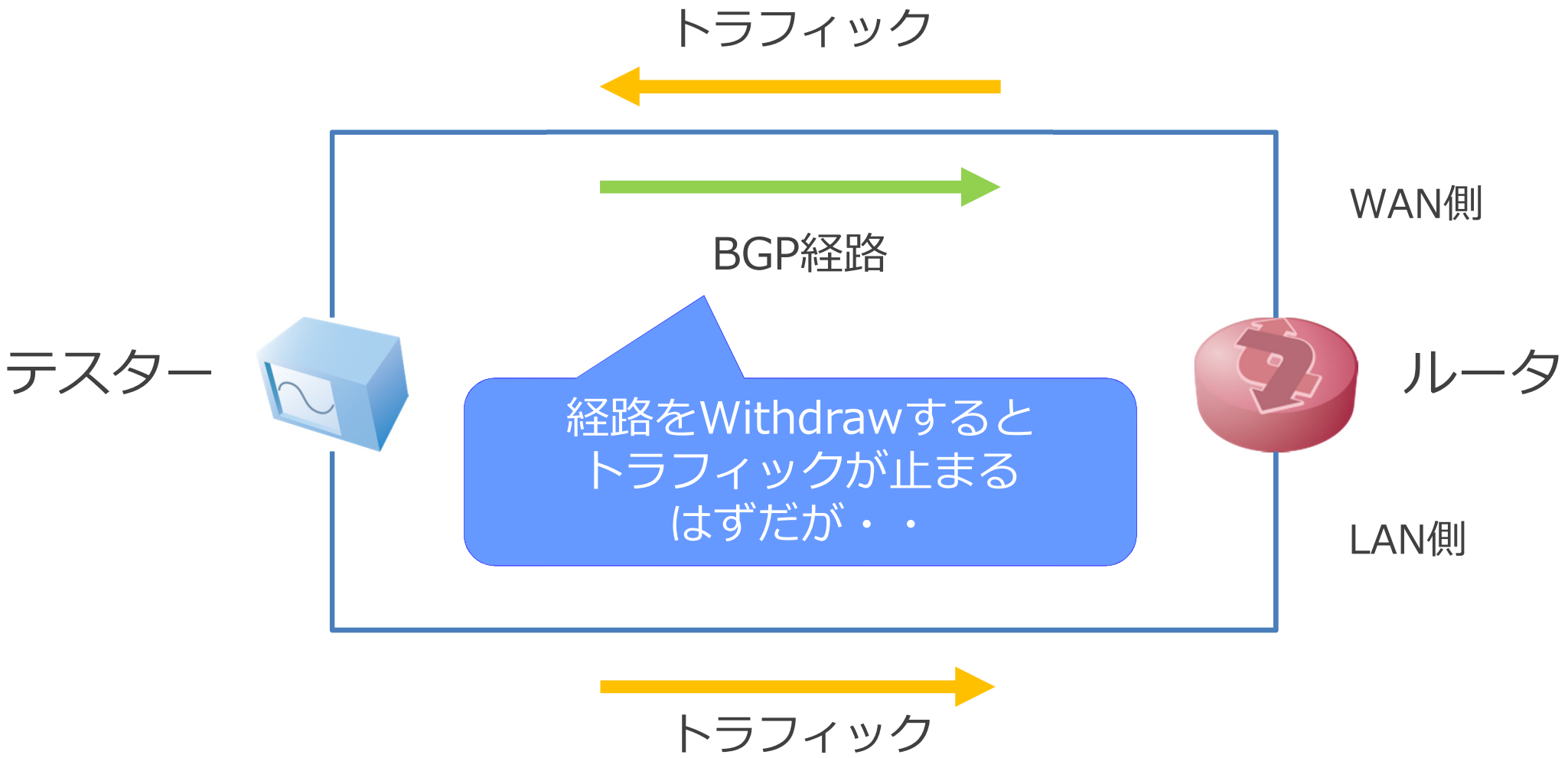



パフォーマンス低下が発生する例

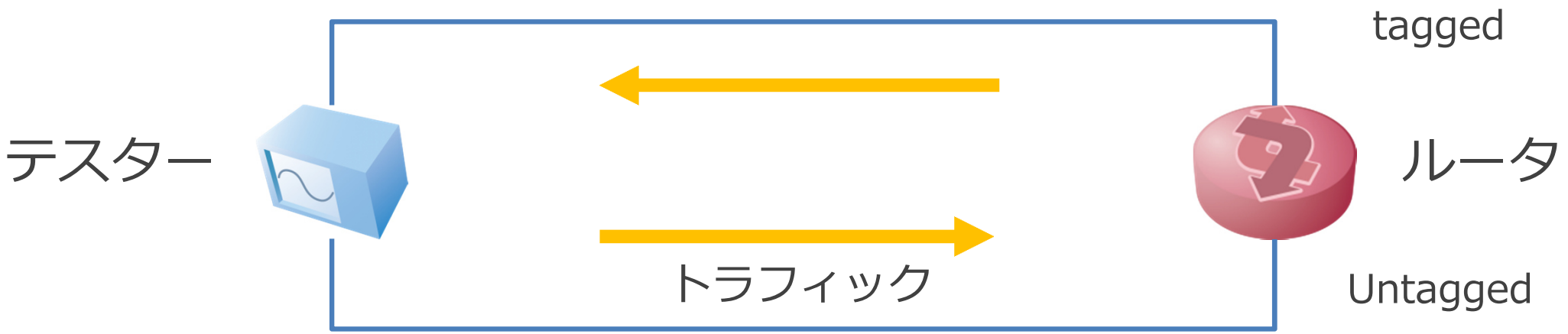
- 大量のConfigを入れたり抜いたり
  - クラウドだとお客様操作で自動でConfigされる
  - OSがクラッシュしないか確認
- ARP/NDPの学習スピード
  - 切り替え時間に影響



- RIB(Routing Table)とFIBの不一致がおきないか？
  - 多数経路の更新時にForwardingに不整合が発生することも



- パケットバッファサイズの測定




Longパケットで99%、100%負荷時のレイテンシの差を求める
例: 99%:2.77us　100%:849.99us　差:847.22us

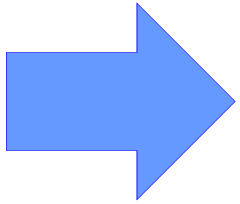
10Gbpsで847.22us分のパケットをバッファリング可能

10,000,000,000/8*0.00084722=1,059,025Bytes

# ずっと悩んでること

SAKURA Internet

- 試験を綺麗に実施/試験結果を綺麗にまとめることの不毛さ
- いくら試験項目をクリアしてもサービスイン後、何が起きるかわからない
- インターネットからは、予想もできない変な経路やパケットがばんばん飛んでくる
- 疲れ果てて適当に打ったコマンドが、バグ発見のきっかけになったり
- ○×表は飾りでしかない
- それでも導入の可否を決めなければならない



現場の担当者が「いける」と思えたかどうか

# 検証作業の心得

SAKURA Internet

- とりあえず溢れさせてみる
  - カタログスペックは信じない(言わずもがな)
  - 実運用で発生する多くの問題は、負荷や収容数に起因。
  - 収容上限を超えるConfigを突っ込んだり、ありえない負荷をかけてみたりする。
- 機器をダウンさせるあらゆる方法を考え、やってみる
- 不具合が見つからないのは検証が足りてない証拠
  - Configに魂を込め、精根尽き果てるまで戦う
- メーカ/ベンダさんのレスポンスも要確認
  - 実運用に入って困ったときに助けてもらえそうか？
  - 何がおきてもなんとかなりそう、という感触を得られるか？
- 完全に要求仕様を満たす装置は無い
- 導入の最終決断は自身の覚悟をもって

# ご清聴ありがとうございました。