JANOG26　IXPセッション
～頑張ってトラフィック支えてます～

IXの信頼性編

任田　大介　touda [at] mfeed.ad.jp
Internet Multifeed / JPNAP

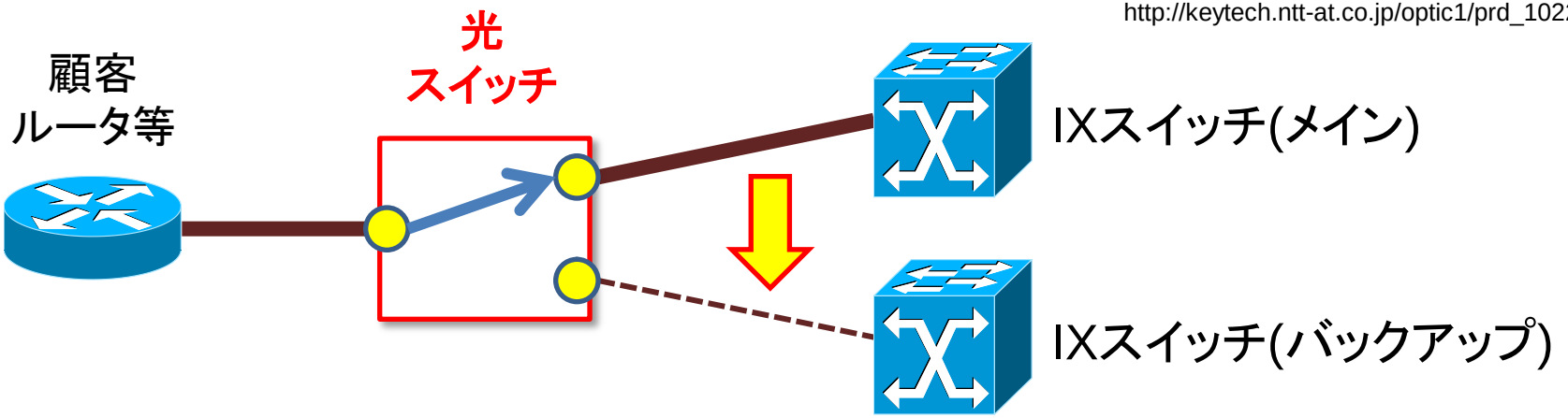
IXにおける品質向上のとりくみ

- ネットワーク構成
  - バックボーン機器冗長
    - レイヤ2プロトコルにより数秒で障害回避
  - 拠点間ファイバの異キャリア・異ルート冗長
    - JPNAP東京I　4拠点間接続
    - バックボーン2面化、異キャリア・異経路接続
  - LAG(リンクアグリゲーション)活用
    - 耐障害性
    - 10GbE ×最大8本のLAGを複数セット使用
    - 但し、設計・オペレーション上の課題は色々あり
  - 大容量L2SW
    - 10GbE × 128ポートなどの高密度収容L2スイッチ
    - ノード集約、シンプルネットワーク化

IXにおける品質向上のとりくみ

- 光スイッチによる可用性向上
  - 光レベルで接続を冗長
  - 切替時間：数10msec内

NTT-AT社HPより
http://keytech.ntt-at.co.jp/optic1/prd_1022.html



- 顧客ルータ ～ IXスイッチ(2台)間に設置
- IXスイッチ故障で自動切替

IXにおける品質向上のとりくみ

- **得られる効果**
  - **格段に早い切り分け・復旧**
    - 手動ケーブル差替え不要
    - 現地かけつけ不要　(リモート操作可)
  - **シングルポイントの故障率低減**
    - L2スイッチに比べて低い故障率
  - **顧客間トラヒックへの影響の極小化**
    - 『顧客ルータがLink断検知しない』⇒ **BGP peer 断しない**
      - 顧客ルータでLink断検知のconfig調整が必要な場合あり

Cisco	carrier-delay msec <miliseconds>
	link debounce time <miliseconds>
Juniper	hold-time up <milliseconds> down <milliseconds>

IXにおける品質向上のとりくみ

- **レイヤ2ループ対策**
  - ブロードキャストストームによるNW不安定
  - 不正なMAC学習による、通信影響

【いくつかの対策】 ※IXによって方式は異なります

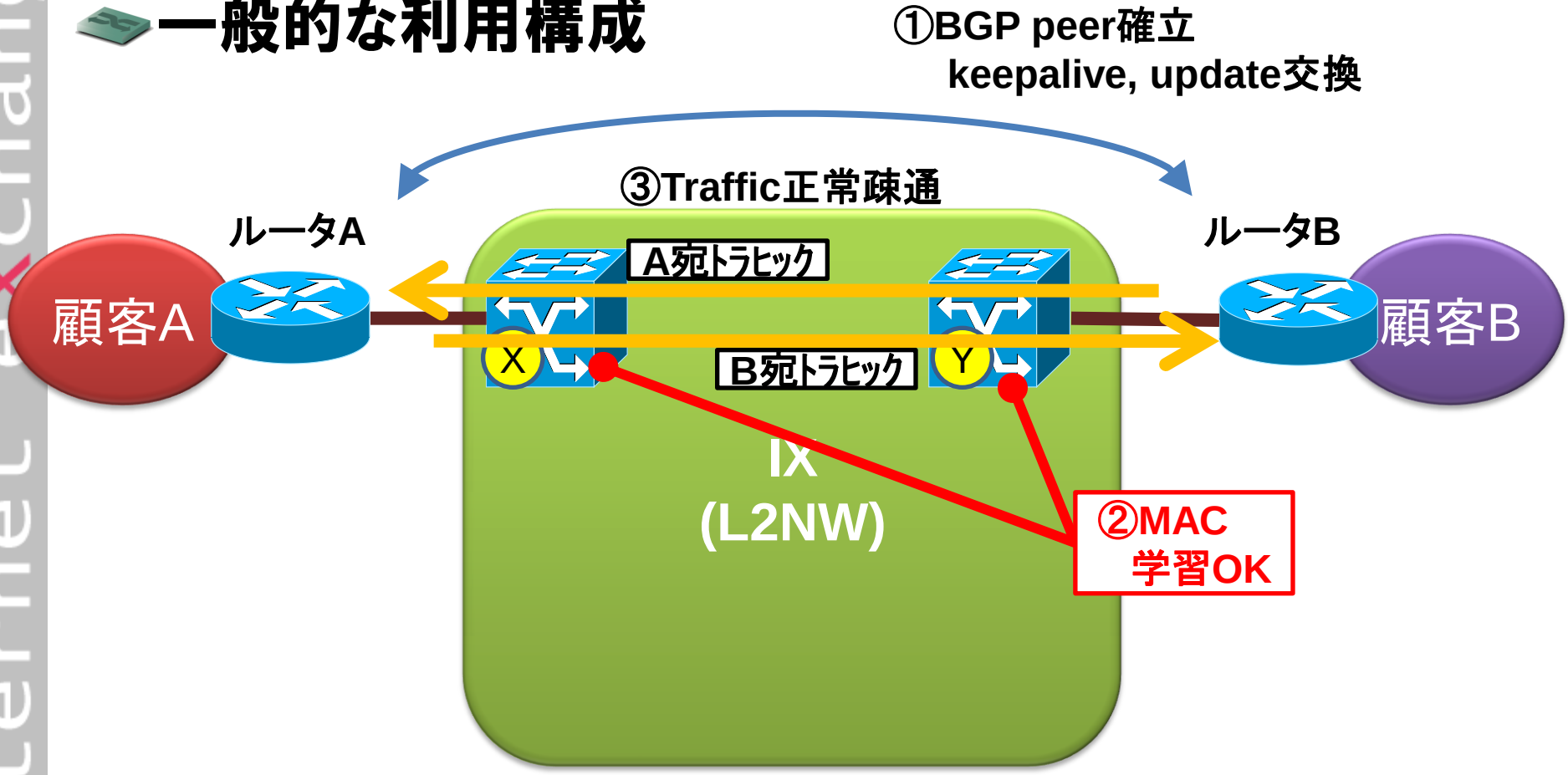
- **学習MAC制限**
  - 顧客収容ポートで学習できるMAC数に上限設定
  - 超過時には、自動でInterface Shutdown
- **MACアクセスリスト**
  - 顧客ルータの正しいMAC アクセスリスト許可
  - レイヤ2ループ等で生じるフレームは、不正Source-MACフレームとして廃棄

IXにおける品質向上のとりくみ

- JANOG Comment 1005
  - **IX接続時に考慮すべき事項をとりまとめ**
    - http://www.janog.gr.jp/doc/janog-comment/jc1005.txt
  - **インタフェース設定**
    - 流してはいけないフレーム　CDP,IGP,STP,Multicast, etc...
    - proxy-arp 、icmp redirectなどの無効化
  - **BGP peer設定関連**
    - 不要なpeer設定の削除
  - **ルーティング関連**
    - IXセグメントアドレスの取り扱い(他ASへの広告不可)
    - BGP next-hop-selfの推奨

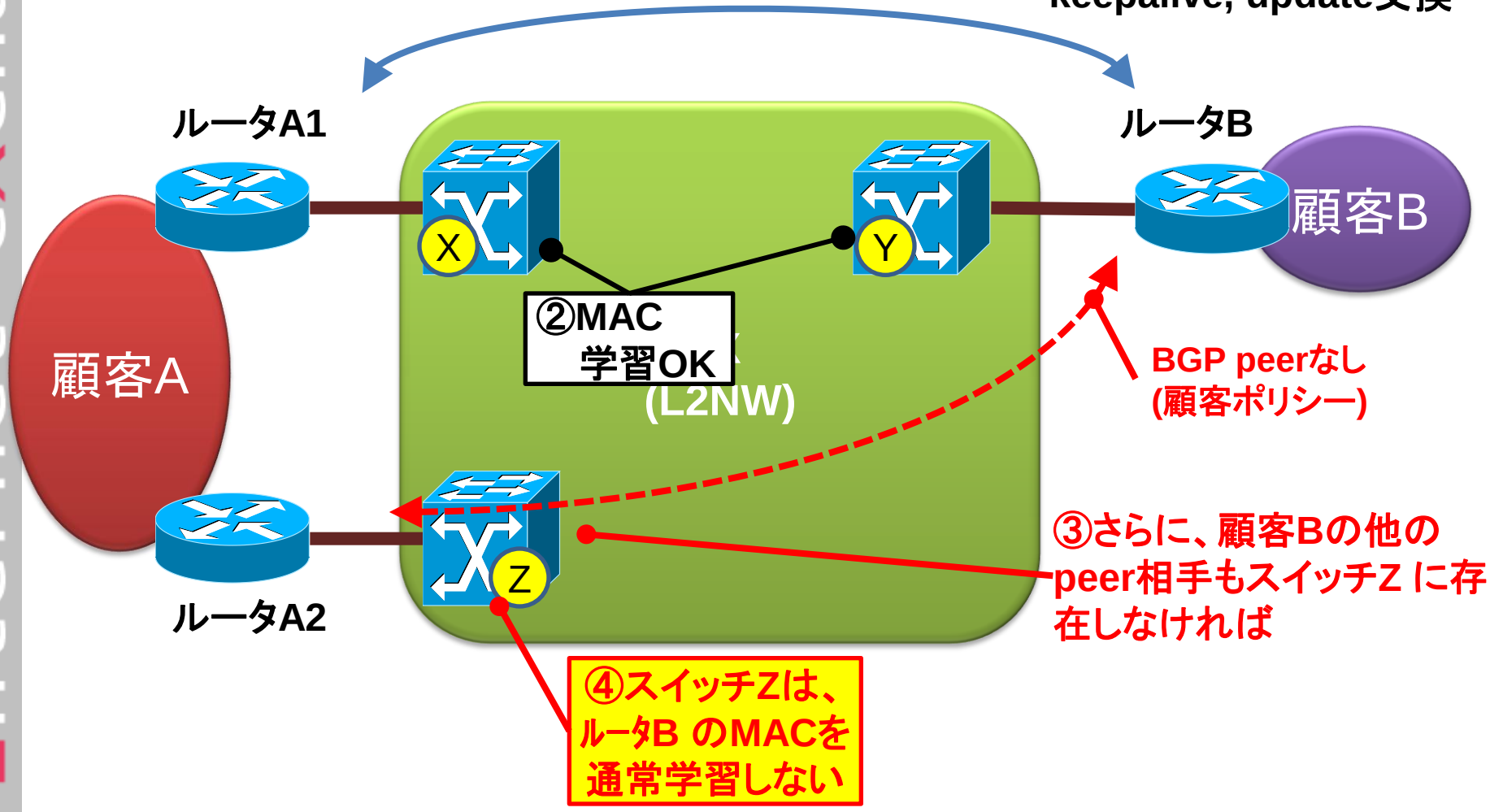
IXにおけるトラブル事例 ～MAC Address Learning～

一般的な利用構成

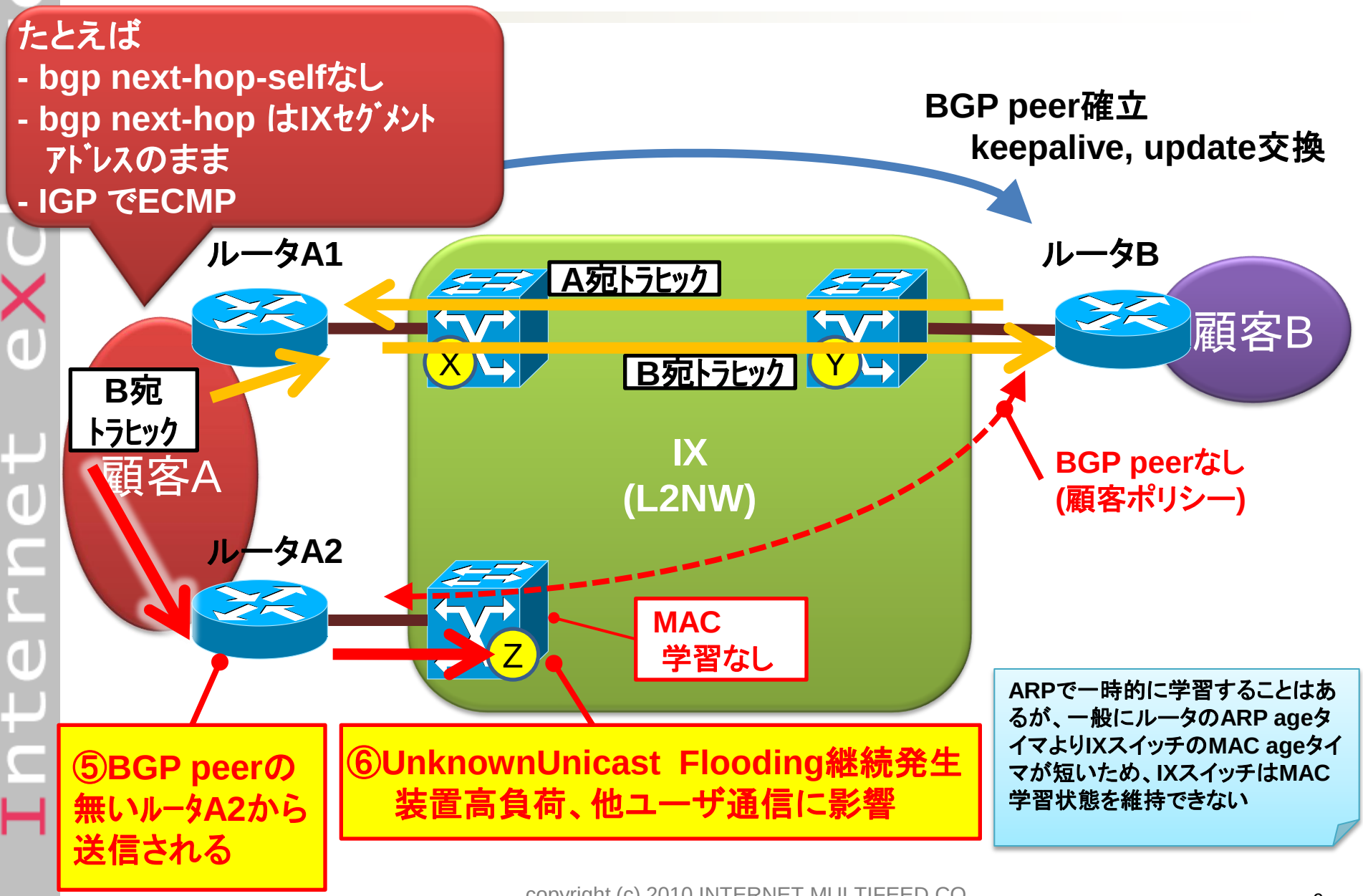


IXにおけるトラブル事例 ～MAC Address Learning～

顧客ルーティング起因で影響を受ける例

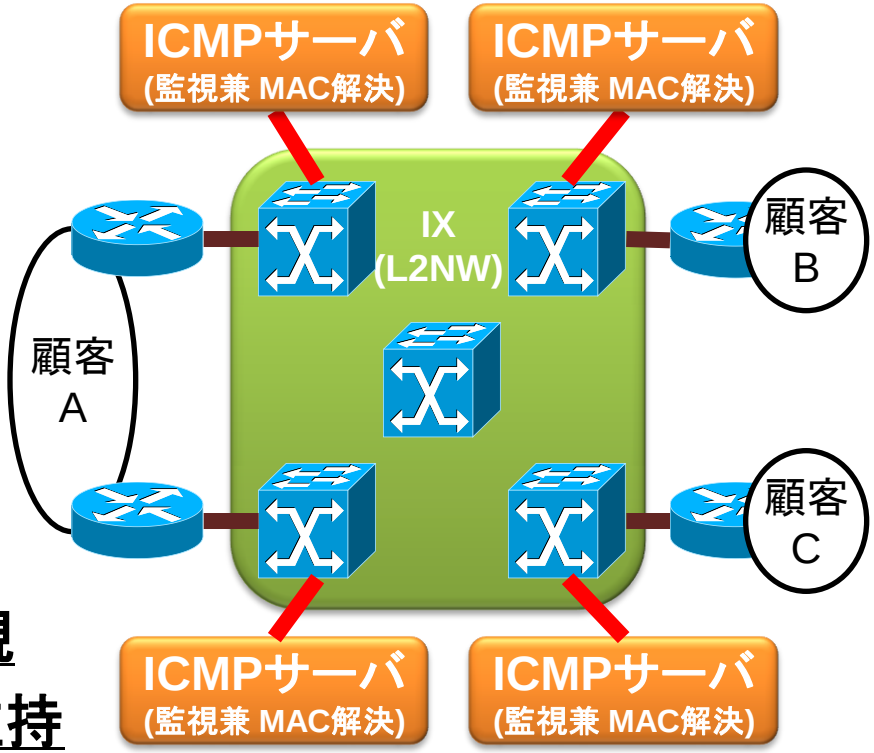


IXにおけるトラブル事例 ～MAC Address Learning～



IXにおけるトラブル事例 ～MAC Address Learning～

- 顧客からみれば
  - IXは、『巨大な1つのL2スイッチ』
  - Trafficはどこから入れても運んで欲しい
- 安定通信のための施策
  - 監視用pingサーバをNWのエッジ部の全てに設置
  - 全顧客をフルメッシュでping監視し、継続的なMACラーニングを維持



- 運用の立場からは
  - BGP peer のあるルータから送信して頂けるのがBestです
  - JC1005にある、iBGP next-hop-self をご利用頂くと、実は解決します！

今後の取り組み

- **IXご利用上の情報提供　(引き続き)**
  - ルーティングや、接続インタフェースの設定・考慮事項
  - configサンプルなどの情報提供
    - 光スイッチ切替時に、断検知しないconfig、etc

- **100GbE-IFへの取り組み**
  - シンプルな網構成
  - 増加するTrafficへの対応
  - LAG運用からの解放

次はBBIX 越智さんから、地域・分散IXのお話です

- **Ether-OAM活用への取り組み**
  - Ethernetレベルでの疎通チェック、品質確認