JPNAP Service

高速光切替素子を用いたネットワークの信頼性向上

2006/1/20 JANOG17

インターネットマルチフィード㈱
菊池　之裕
yuki-k@mfeed.ad.jp

JPNAP Service

目次

- 背景
- 今回の内容
- 信頼性のあるネットワーク
- 高速切替素子の特徴
- 開発で考慮した点
- 応用例
  - IXにおけるUNIの冗長化
  - VRRPを利用した冗長
- まとめ
- 今後の課題

JPNAP Service

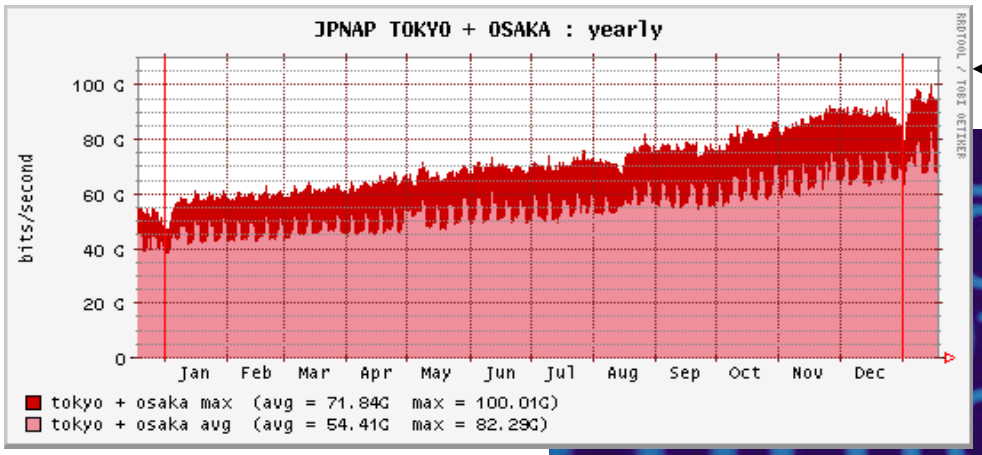
背景

- ネットワークの高速、広帯域化

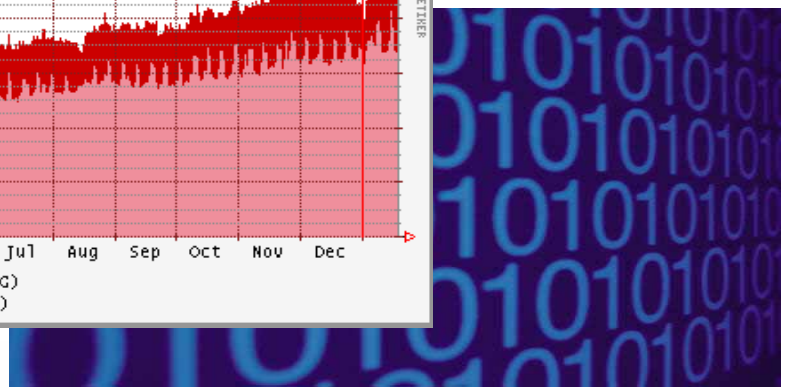
  1G -> 10G …

  – 単位時間当たりの流量が増える

    → 障害時間当たりの失われるデータが増える。



←なんと約100Gbits/sec



All communication flows through here.

全てのコミュニケーションはここを通る

JPNAP Service

今回の内容

お題「高速光切替素子を用いた
ネットワークの信頼性向上」

- 高速光切替素子を使ったLayer1による切替
- 高速光切替素子の遠隔操作による障害時間短縮

信頼性向上手法の1つとして、実例を交えて、議論をしていきたい。

信頼性のあるネットワーク

- ＭＴＴＲ
  - 平均回復時間に焦点を置いてみる

Layer4～ (ロードバランサ)	○
Layer3(OSPF/BGP)	○
Layer 2 (STP/RSTP)	○
Layer1 (物理的な二重化)	数十分～数時間 ×

うれしくない (泣

All communication flows through here.

全てのコミュニケーションはここを通る

JPNAP Service

物理的二重化の必要性

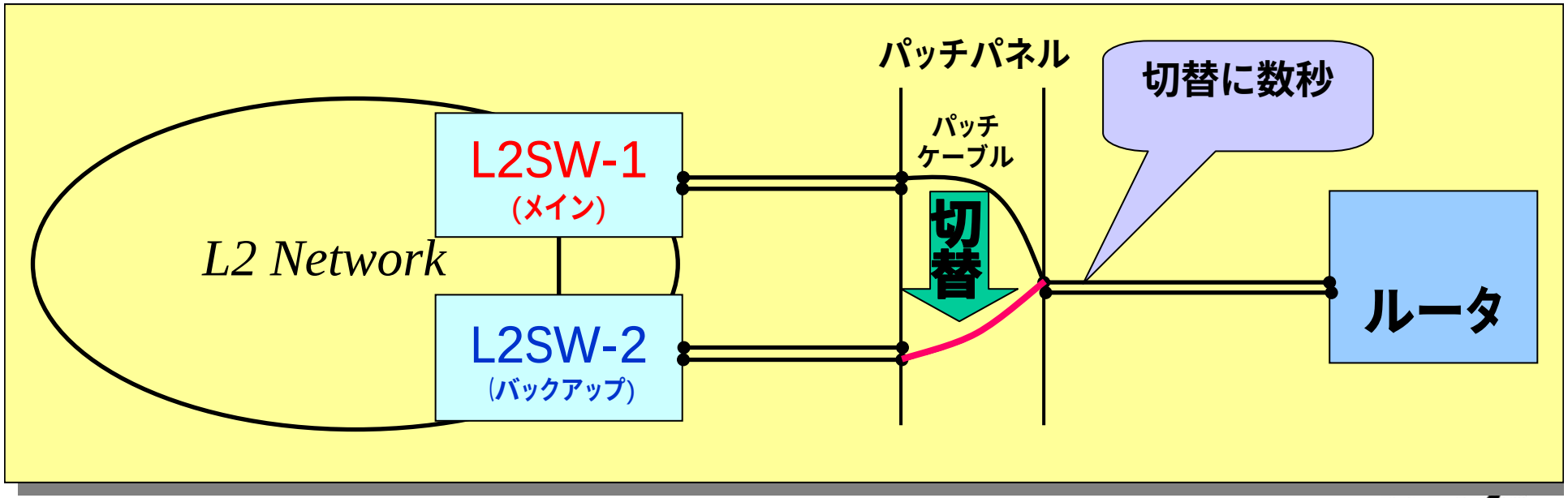
- スイッチ/ルータや回線を
  完全二重化するのが理想
- どうしても二重化できない制約が存在
  - 専用線顧客への接続点
  - IXの接続点
  - データセンタのお客様接続点
    - コスト的な制約、、、

→ 接続提供者側で二重化
パッチパネルの利用

JPNAP Service

物理的な切替には、、

- ✓ メンテナンス時、パッチ切替が必要
- ✓ 数秒かかる → 高速化したい。



JPNAP Service

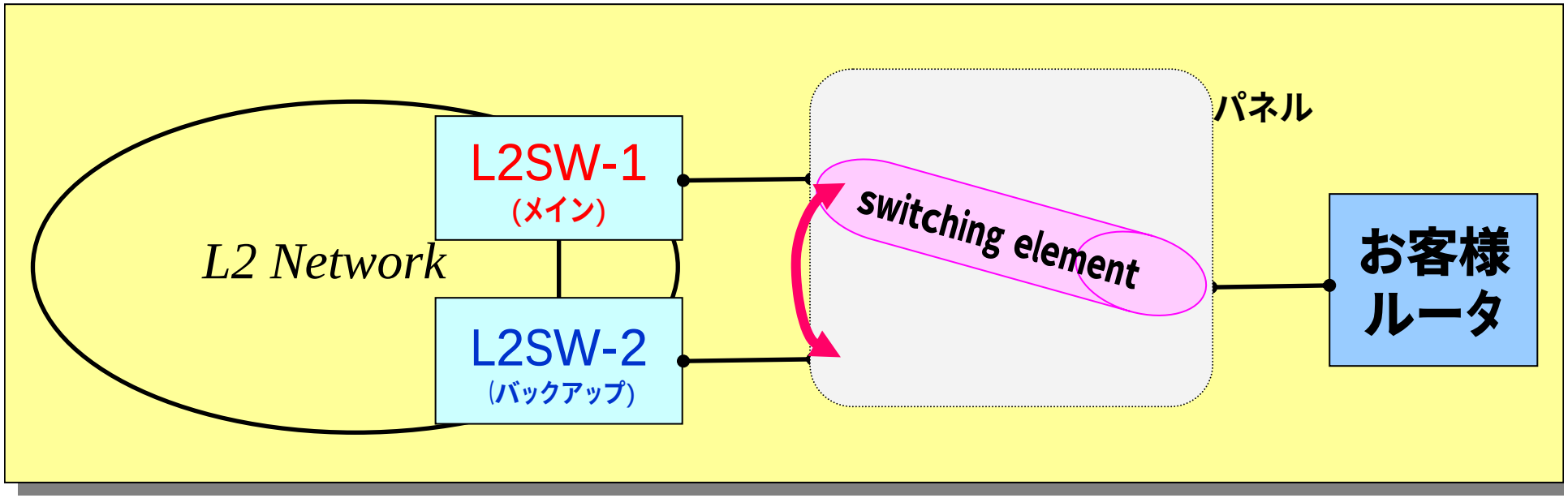
切替高速化のために

パッチパネルの代替として、光切替素子を導入

- ✓ シンプル＆パッシブ → 故障しにくい
  - ✓ 自分で判断条件をもって勝手に切り替えてくれるもの。
  - ✓ 規模が大きくないこと、100ポート収容できるとかだと、、
    - ✓ 壊れると100ポート巻き込んで壊れると痛い。

3DMEMSはこれがあるので、、、没

- ✓ 高速切替可能 → 「切れない」ネットワークの実現

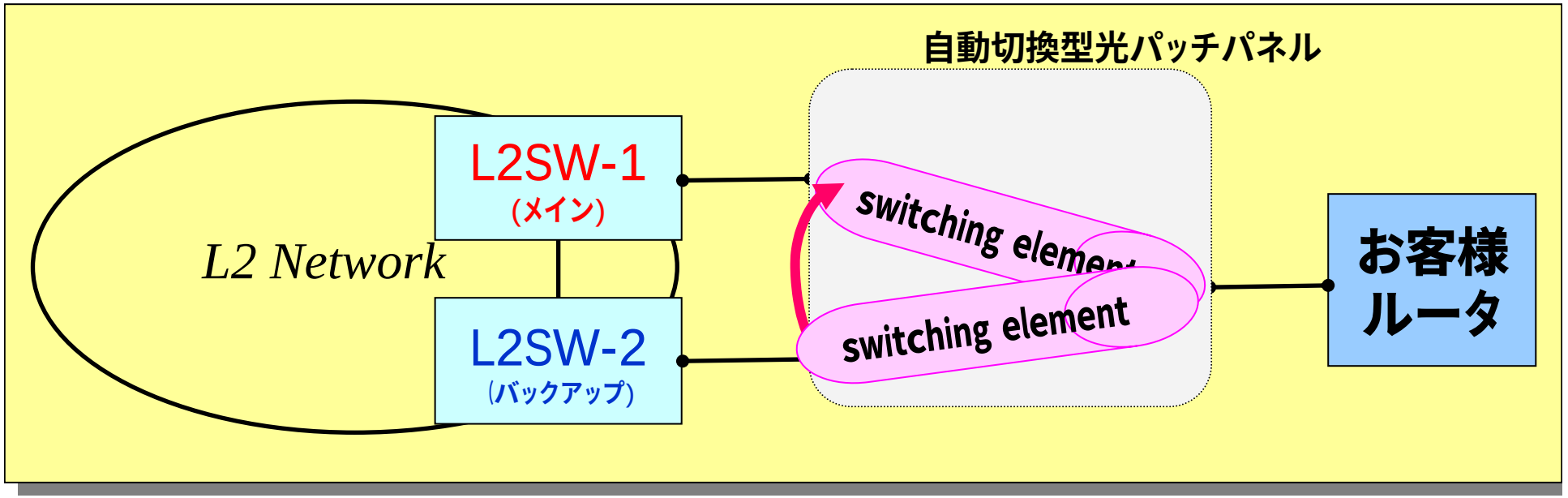


JPNAP Service

高速光切替素子の特徴

- ✓ 方式としては、ファイバ駆動型機械式光スイッチ
  電磁石により高精度で機械的に素子内部のファイバの接続先を変更
  → 手動による切替と機構的には同一
- ✓ 一回路一素子、壊れるときも一回路だけ、とってもシンプル

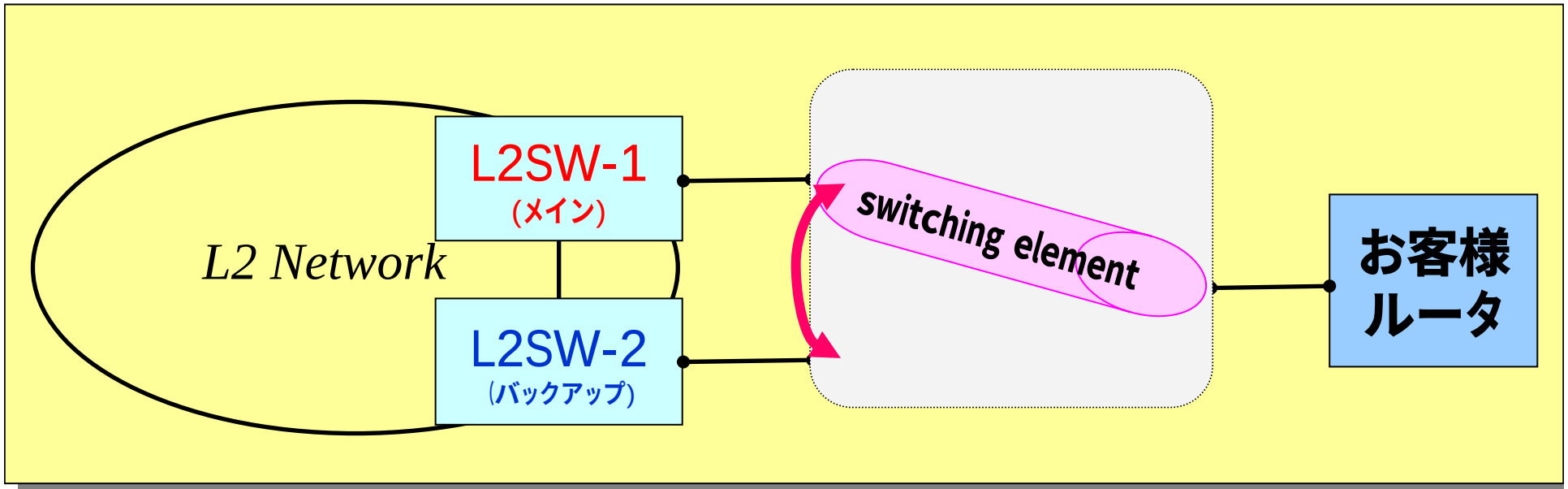
- ✓ 数十ミリ秒程度の高速切替
  → ルータ側でリンクダウンに気づかない。
  → BGP,OSPF等のプロトコルが切断されない。



JPNAP Service

高速光切替素子の特徴

- ✓ スイッチの電源が断となっても自己保持
  - ✓ → 電源が断となっても通信に影響がない

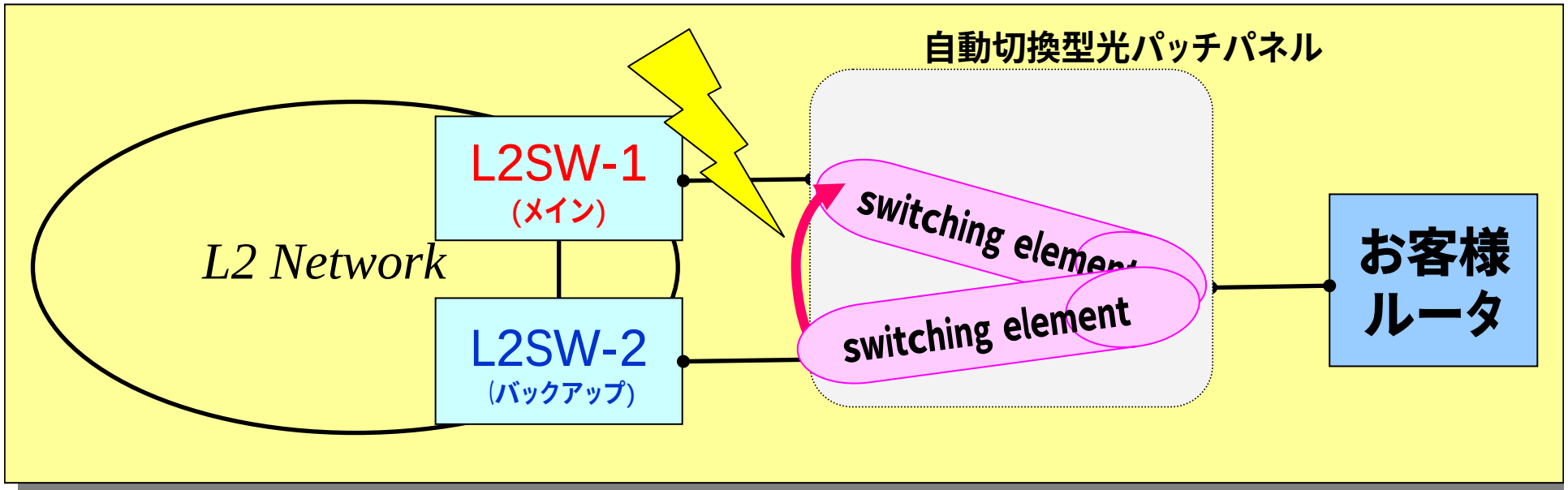


JPNAP Service

障害検出時の動作

① メイン系の光が断となると、バックアップ系に高速に切替る
→ 対向ルータでLink downを検出しない場合が多い

② 仮にメイン系の断の後、メイン系の光が回復しても、バックアップ系状態を保持



JPNAP Service

そんなわけで

NTTーAT社と共同開発

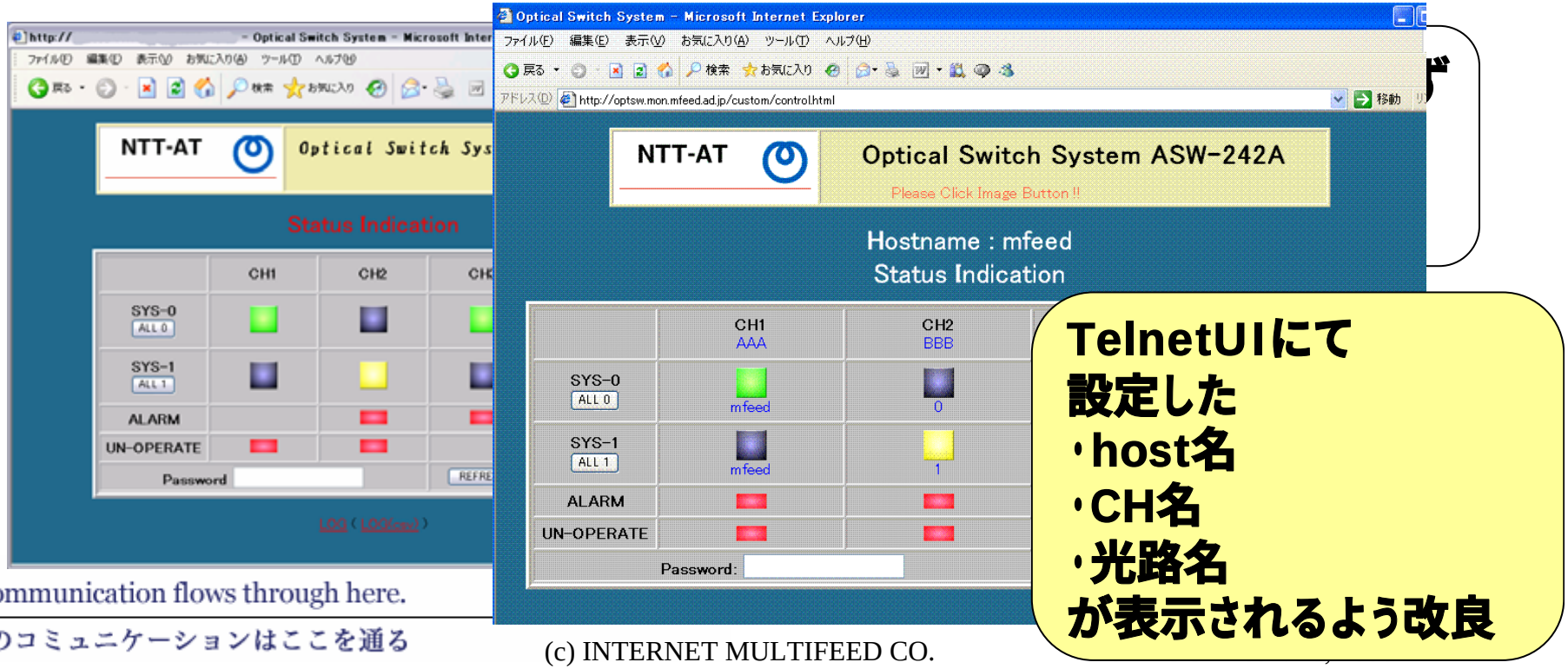
- 既存の1x2のスイッチをベースに2x4スイッチ版を製作
- 2本のファイバを同時駆動することによりEthernetに対応
- 1000Base-SX対応のためMMF版も開発
- Management LANを持ち、WebUI,Telnet、Serial から遠隔操作可能
  - 光断とならない障害への対応
  - メンテナンス時の切替に対応
- SNMP対応、切り替わり時、光断時にはTRAP発出

http://keytech.ntt-at.co.jp/optic1/prd_1021.html から引用

JPNAP Service

開発で考慮した点

- WebUIの改良
  - 大量導入による　機器操作誤りミス防止



JPNAP Service

目次

- 背景
- 今回の内容
- 信頼性のあるネットワーク
- 高速切替素子の特徴
- 開発で考慮した点
- 応用例
  - IXにおけるUNIの冗長化
  - VRRPを利用した冗長
- まとめ
- 今後の課題

JPNAP Service

IXにおけるUNIの冗長化

- IXの特徴
  - ISP間の相互接続点
  - Layer2 IXにおいては、一台 1BGP セッション
    - 冗長化には複数台のルータが必要
    - → ユーザに対して高いコストのハードル
  - 接続点がシングルポイントとなる。
    - → 接続断になったときにＩＳＰ内の大きく経路を揺らす可能性
      - →東京経由のものが大阪まわりになる
      - →Peer復旧時にＢＧＰ経路の収束に時間がかかる

JPNAP Service

IXにおけるUNIの冗長化

- IXの特徴

  従来はパッチパネルを使用

  →高速化手法として、光切替装置を導入

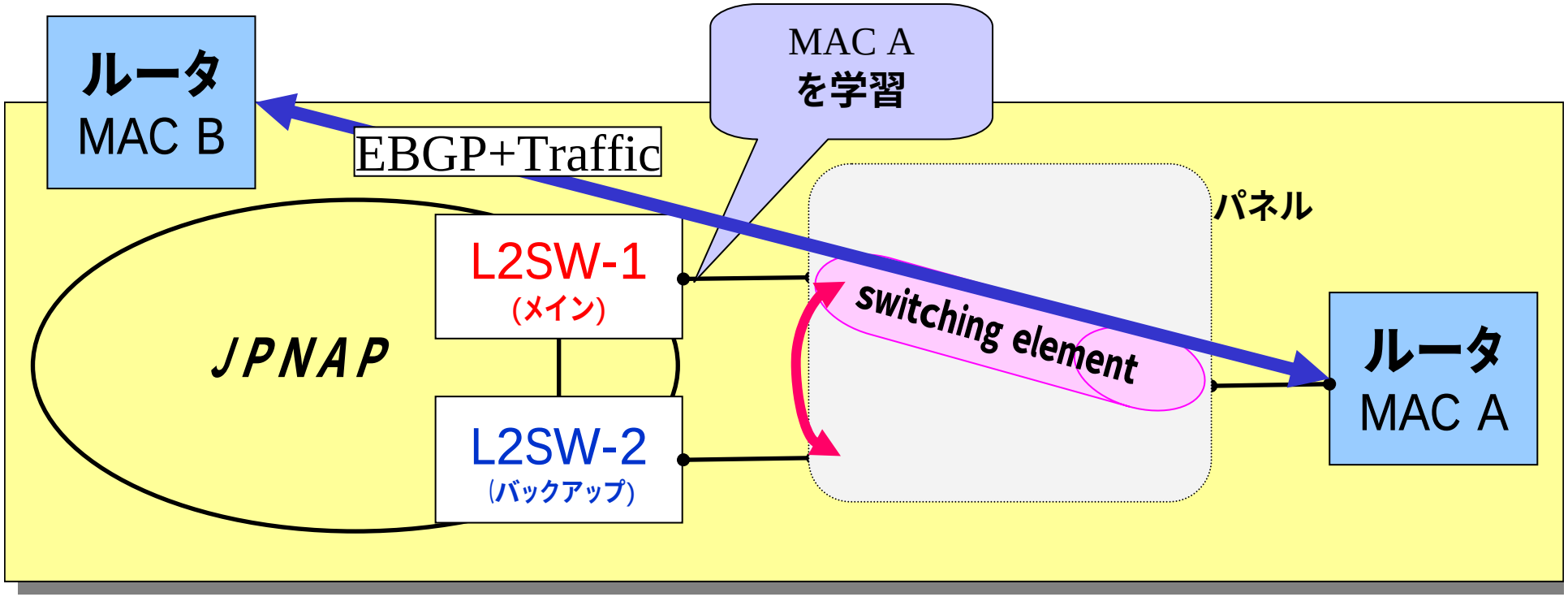
  →障害、メンテナンス時にリンクダウンしずらいネットワークを実現

  →Link FlapによるBGP収束を無くすことに成功

JPNAP Service

IXにおけるUNIの冗長化

パッチパネルの代替として、光切替素子を導入



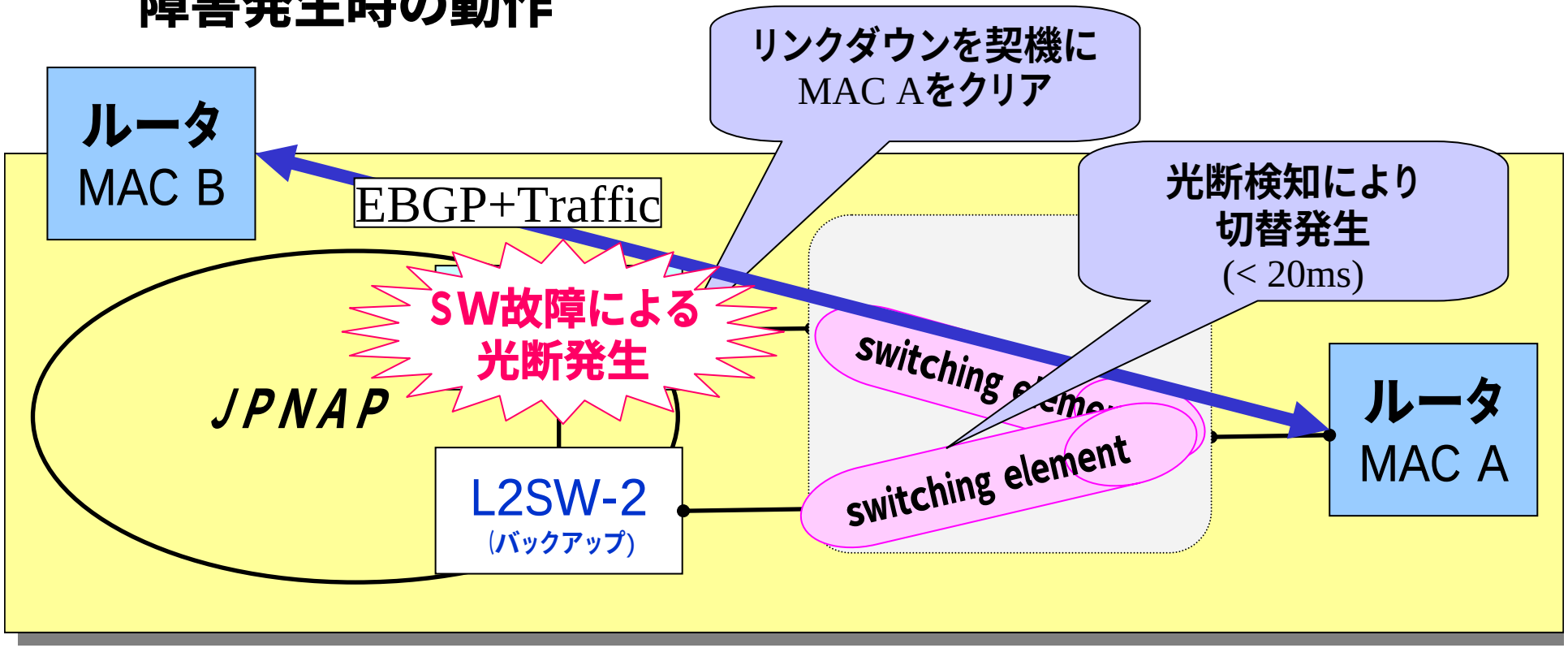
All communication flows through here.

全てのコミュニケーションはここを通る

JPNAP Service

IXにおけるUNIの冗長化

障害発生時の動作



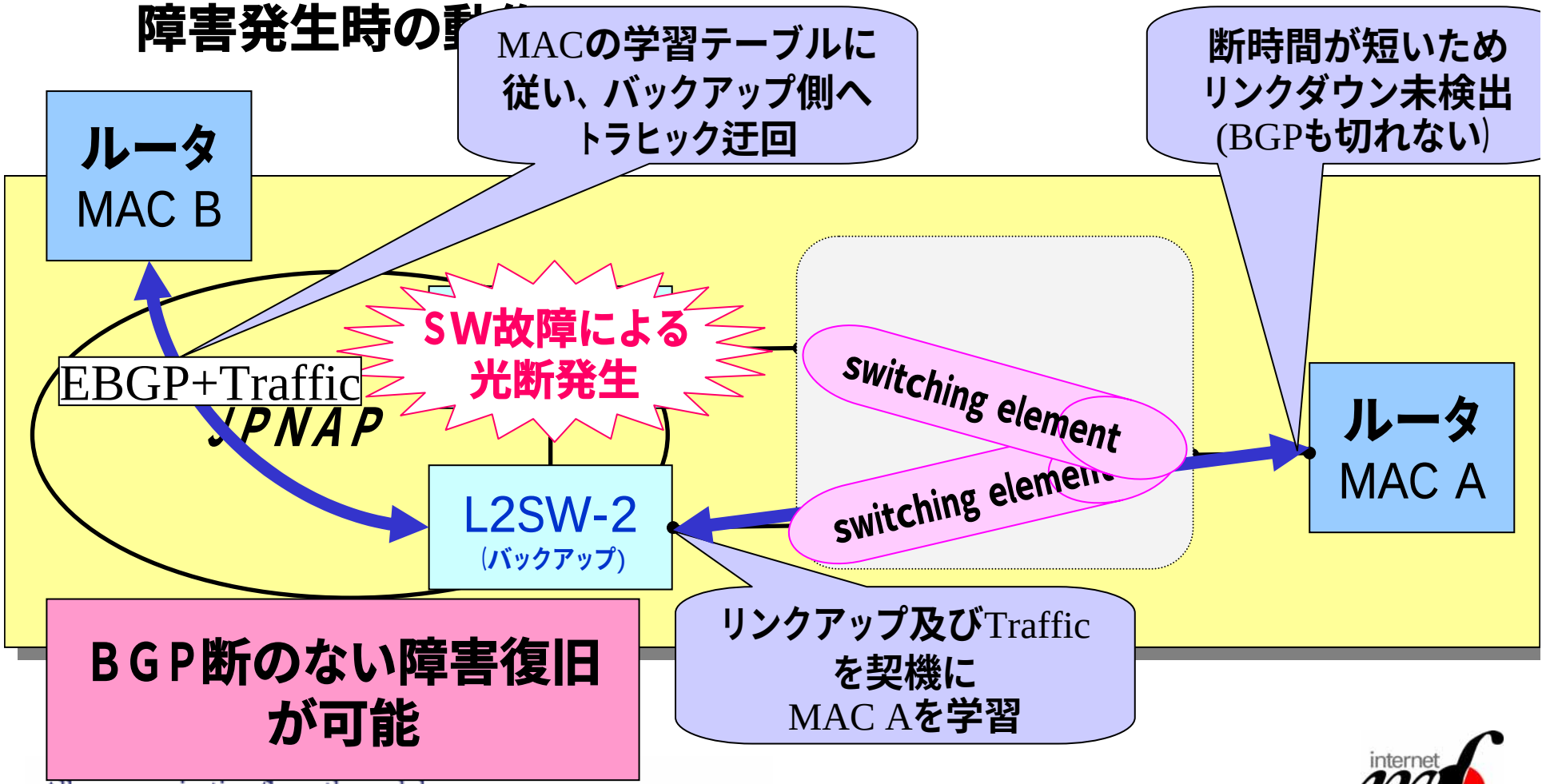
All communication flows through here.

全てのコミュニケーションはここを通る

JPNAP Service

IXにおけるUNIの冗長化

障害発生時の動



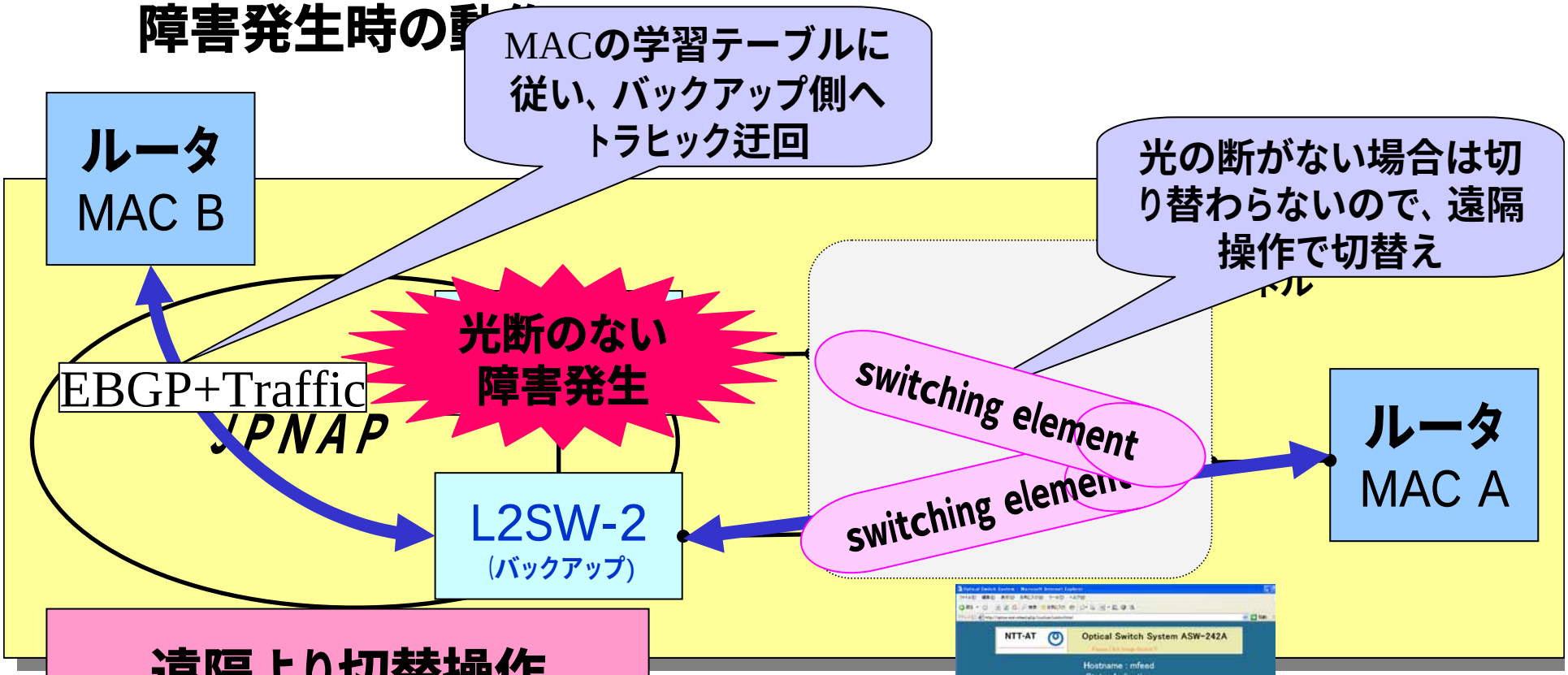
All communication flows through here.

全てのコミュニケーションはここを通る

JPNAP Service

IXにおけるUNIの冗長化

障害発生時の動



All communication flows through here.

全てのコミュニケーションはここを通る

JPNAP Service

実際の運用

・ こんな感じで動いてます

All communication flows through here.

全てのコミュニケーションはここを通る

JPNAP Service

VRRPを利用した冗長

- 環境
  - お客様拠点の冗長化を想定
  - 制約条件として、お客様からの線は1本だけ
- 目的
  - 拠点からスイッチへの経路を複数にする

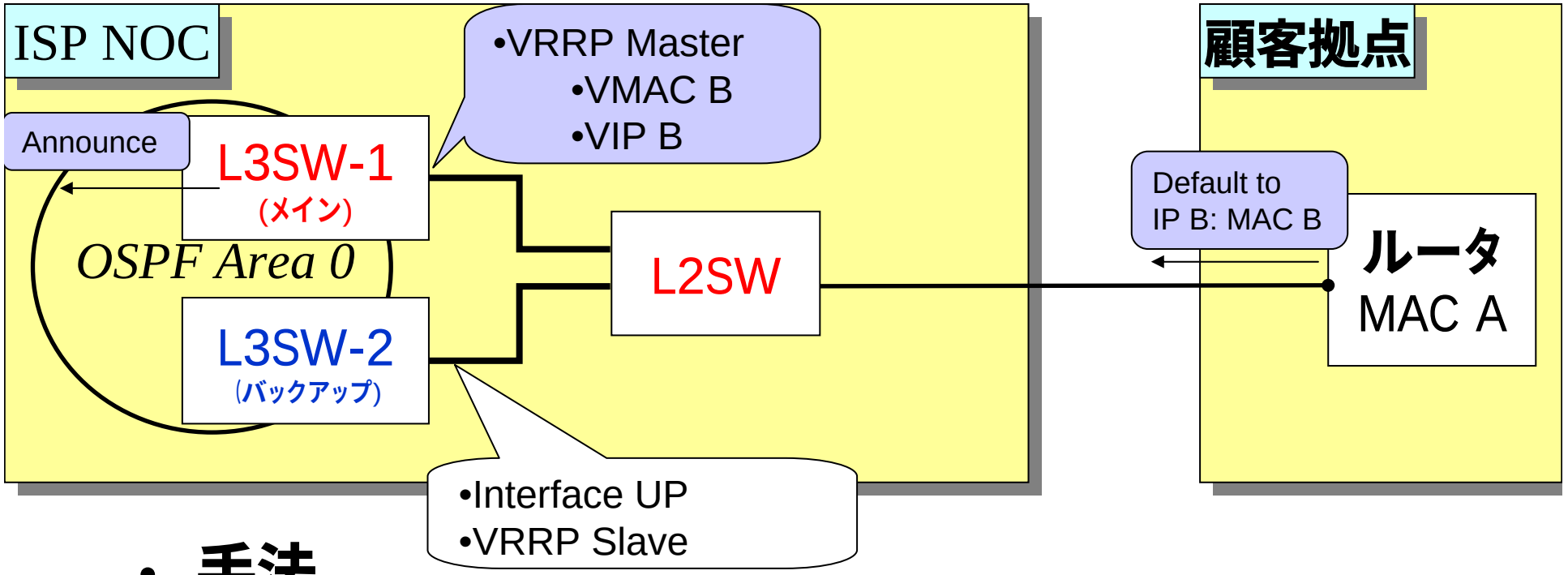
例)

- L3 +L2 switch
- L3 switch+光スイッチ

を比較しながら紹介したい

JPNAP Service

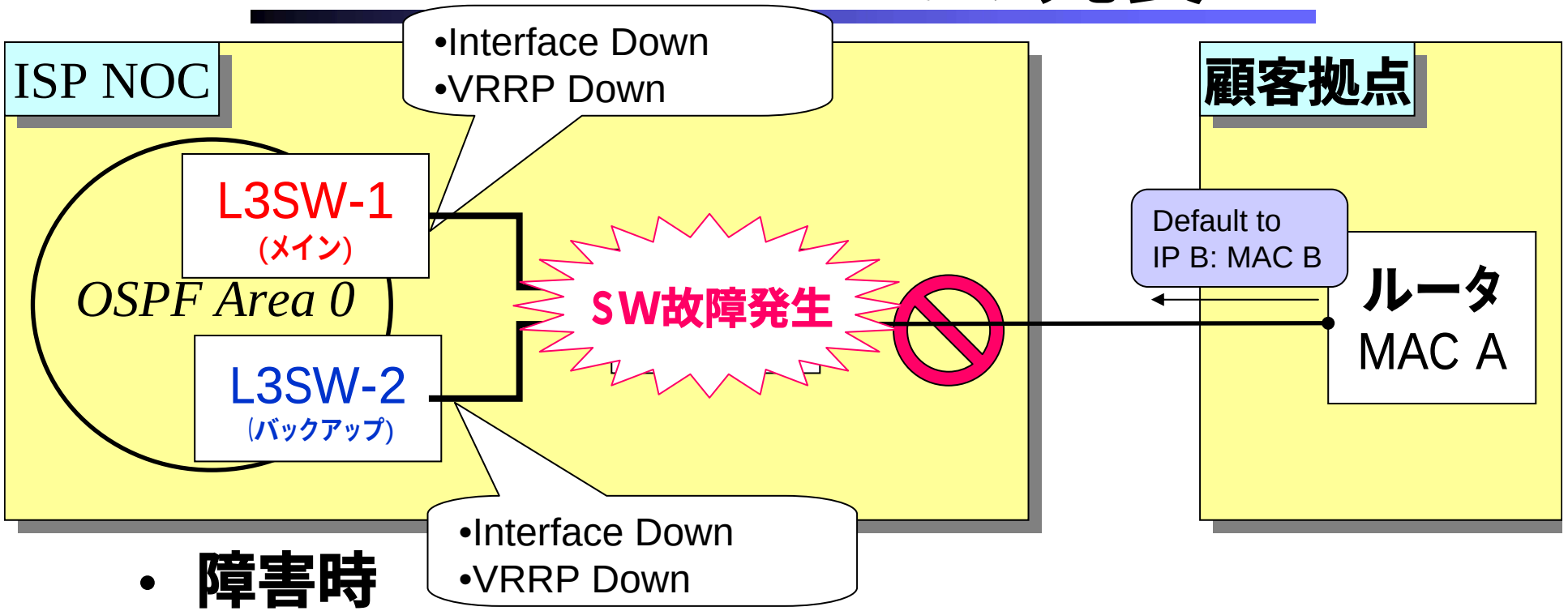
L3 +L2 switchでの冗長



- **手法**
  - **拠点側**: static route**のみ**
  - ISP**側**:static route**設定し、**Static to OSPF**でアクティブな経路のみアナウンス**
  - L3+L2 switch VRRP**で二重化**

JPNAP Service

L3 +L2 switchでの冗長

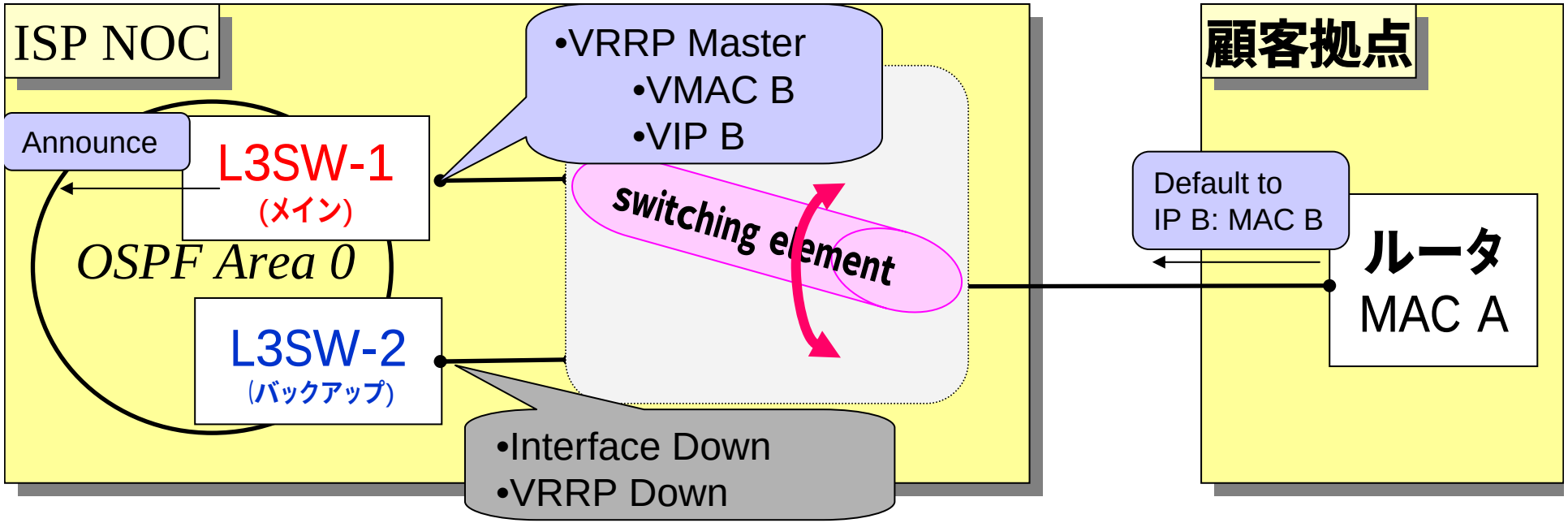


- **障害時**
  - L2スイッチが故障
  - 顧客への経路が無くなり、接続断
  - 駆けつけまたは常駐者による保守による切替

→ 障害時間の長期化

JPNAP Service

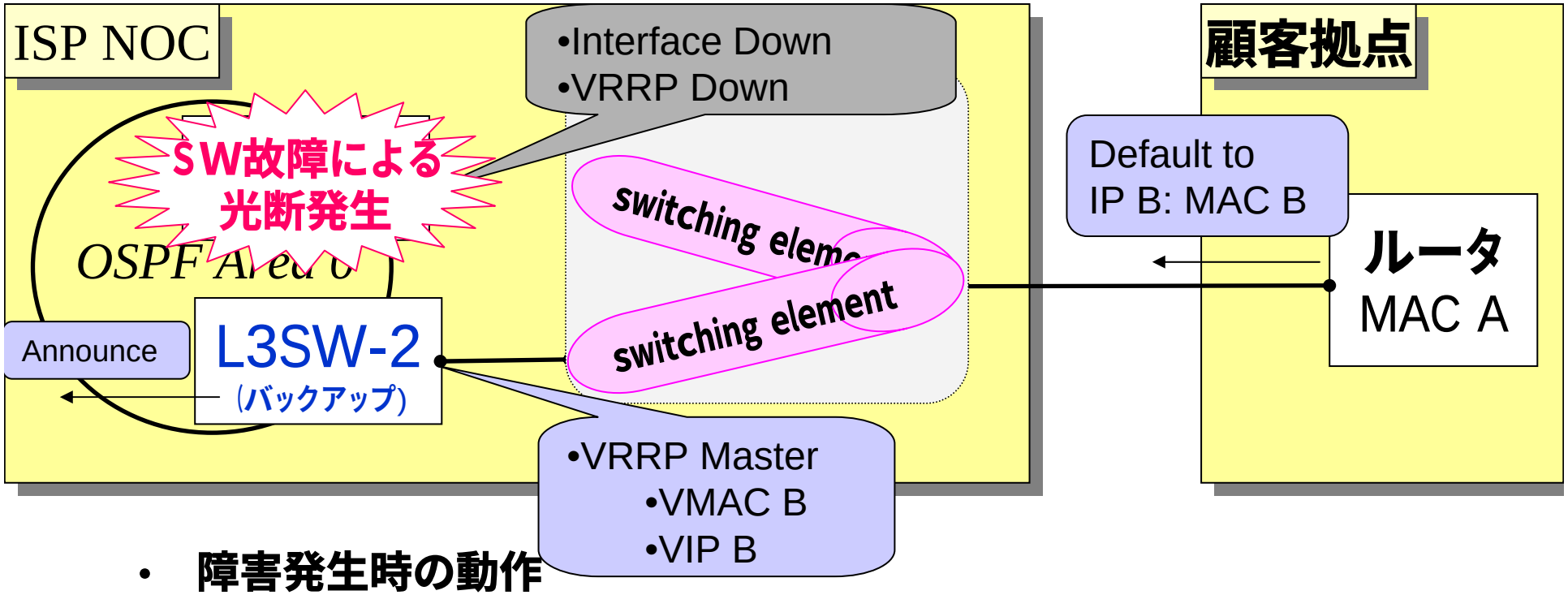
L3スイッチ +光スイッチでの冗長



- L2SWの代わりに光スイッチを使用
  - 構成はほぼ同じ
  - スタンバイ側は光が来ていないのでDown

JPNAP Service

L3スイッチ +光スイッチでの冗長

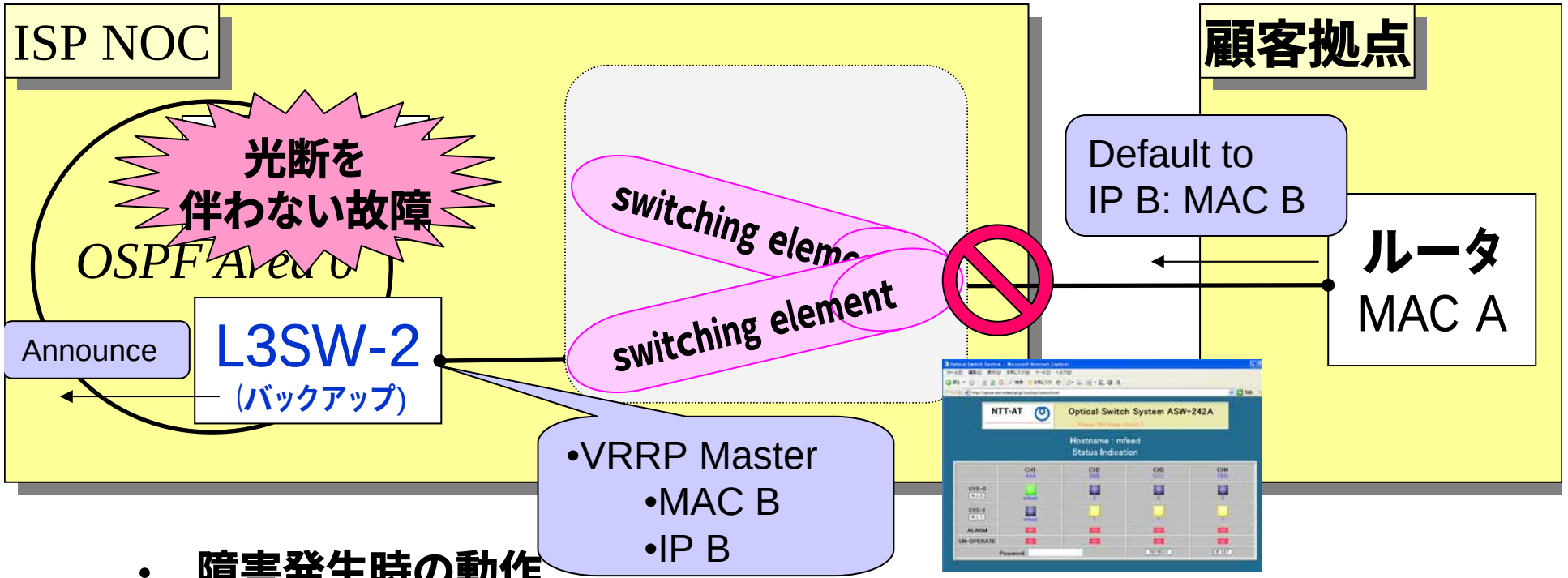


- **障害発生時の動作**
  1. **光切替装置が光断を検出し切替**
  2. L3SW-1**の**Interface/VRRP**が落ちる**
  3. L3SW-2**の**Interface**が**UP/**ネゴシエーションの後**VRRP**が**UP
     → OSPF**によるアナウンス元の変化**

  拠点側ルータはMAC**が変わらないので、リンクダウンに気づかずにトラヒックが移る**

JPNAP Service

L3スイッチ +光スイッチでの冗長

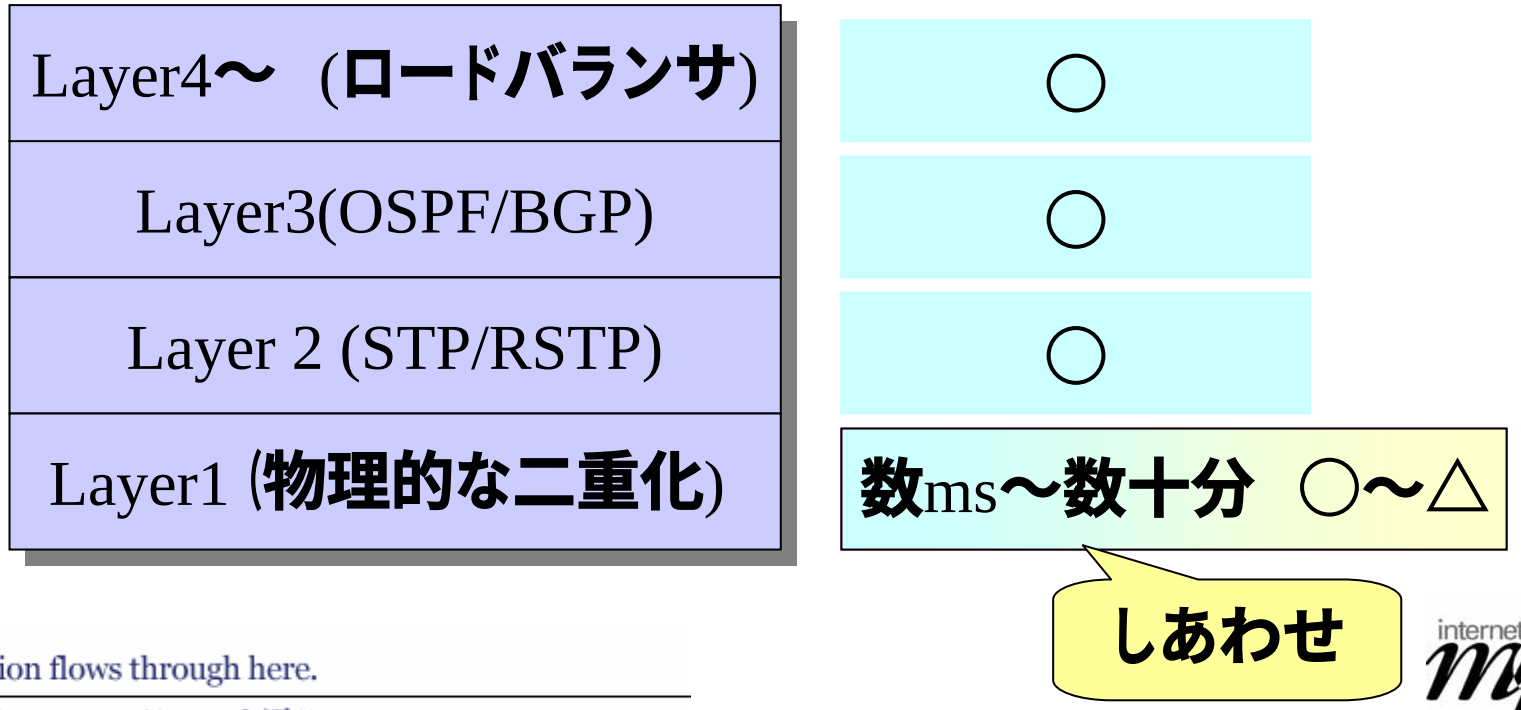


- **障害発生時の動作**
  1. **光スイッチが断検知ができないので　到達性はなくなる**
  2. **光スイッチを遠隔操作することにより、切替**
     がUP/ネゴシエーションの後VRRPがUP
     ウンス元の変化

遠隔より切替操作
駆けつけ対処より高速

信頼性のあるネットワーク (もう一度)

- MTTR
  - 平均回復時間



JPNAP Service

目次

- 背景
- 今回の内容
- 信頼性のあるネットワーク
- 高速切替素子の特徴
- 開発で考慮した点
- 応用例
  - IXにおけるUNIの冗長化
  - VRRPを利用した冗長
- まとめ
- 今後の課題

JPNAP Service

まとめ

- ネットワーク信頼性向上のため、Layer1による高速光切替装置を導入する事例を紹介しました。
  - ベストパターンでリンクダウン無しの10ms以下での切替
  - ワーストパターンでも遠隔操作により駆けつけよりも早く対処できます。
- プロトコルだけに頼るだけでなく、物理的な切替装置を導入すると、全体の障害時間を短くすることができます。
- 機械は想定通りに壊れない、、、

JPNAP Service

今後の課題

- 802.1adで複数本トランクを組んで帯域増強
  → トランクを組んでいる同士で多数決論理が必要

- トランク版も開発したが、、あまり4ch,8chとか束ねると密度的にきびしい、、
- 3DMEMSのほうが相性がいいかも、、、
  - うーんでもラッチ付きは無いのですよね、、
  - ベンダさんの参入に期待、、、

- **最後に**
  - ほかにこんな手法があるかな、とかありましたら会場からもぜひお願いします。

JPNAP Service

ご静聴ありがとうございました。

Question & Comments?

JPNAP Service

おまけ

- **各ルータの**IF-down**確認結果**

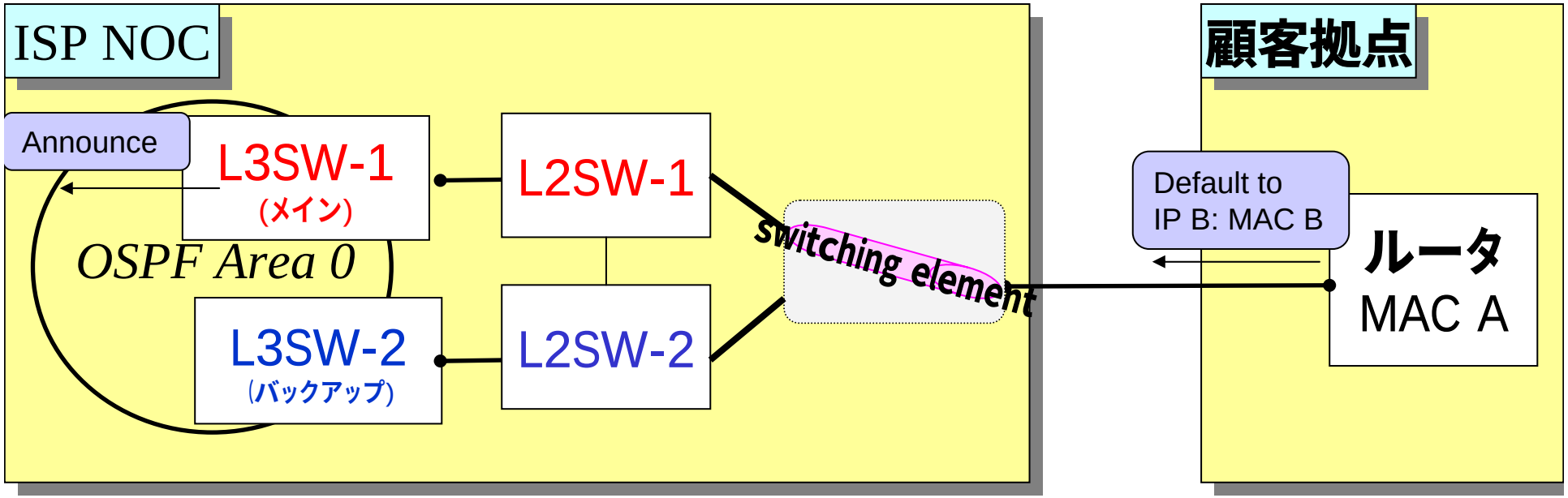
1. **切り替え時間**
   - ＜50ms**であれば**link-down**を検知しない。**

GS4000-10GE ※1	<100ms
BD-10GE	
Summit-GE	
Juniper-10GE ※2	<200ms
GSR-10GE(default)	<100ms
GSR-10GE(delay=0s)	<50ms

※1,2 **机上値よりも長い、これはデバイス側の仕様とのこと**

JPNAP Service

贅沢にやると



- L3SW+L2SW**二重化+光スイッチ**
  - L2sw、**光スイッチ両方の利点を取れる。**