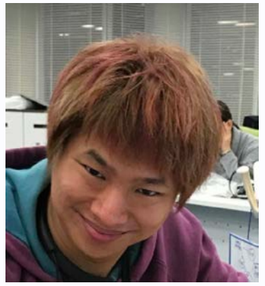
Yahoo! JAPAN Site Operations

ヤフー IP CLOS ネットワーク その後

ヤフー株式会社 テクノロジーグループ
システム統括本部 サイトオペレーション本部
深澤 開

自己紹介



- 名前
  - 深澤 開 (ふかざわ かい)
- 経歴
  - 2013 4月~ ヤフー新卒入社後、全社Hadoopの設計・運用
  - 2014 10月~ データセンターネットワーク
  - 2018 7月~ アメリカ赴任(予定)
- 業務
  - データセンター内ネットワークの設計・運用
- 趣味
  - Splatoon2
    - S+底辺
    - メイン武器：プライムシューター
  - 髪を染めること
    - 赤 → 赤&オレンジ&ピンク → 緑 → シルバー (1.5ヶ月 2万円)

Yahoo! JAPAN Site Operations

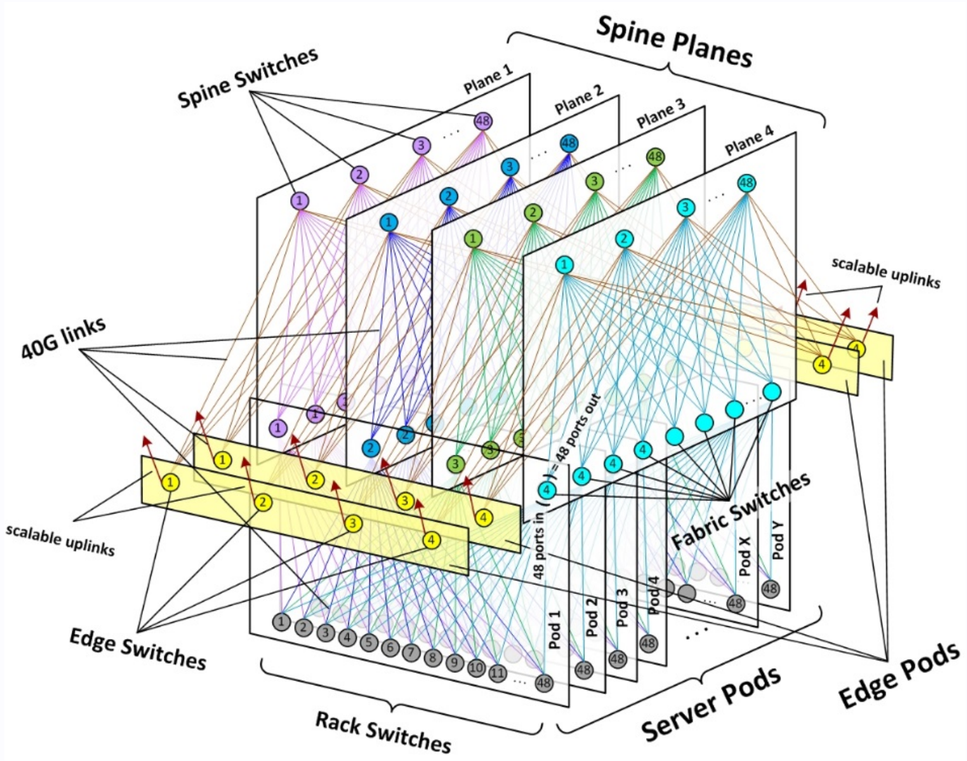
アジェンダ

- JANOG38 振り返り
- アメリカ拠点のその後
- IP CLOS ネットワーク 国内導入事例
- IP CLOS ネットワーク 全面展開への課題
- Network Lab
- 今後

JANOG38 振り返り

JANOG38 振り返り

- IP CLOS ネットワークとは
  - Google, Facebook, Amazon, Yahoo…
  - Hyperscale が採用しているDCネットワーク構成

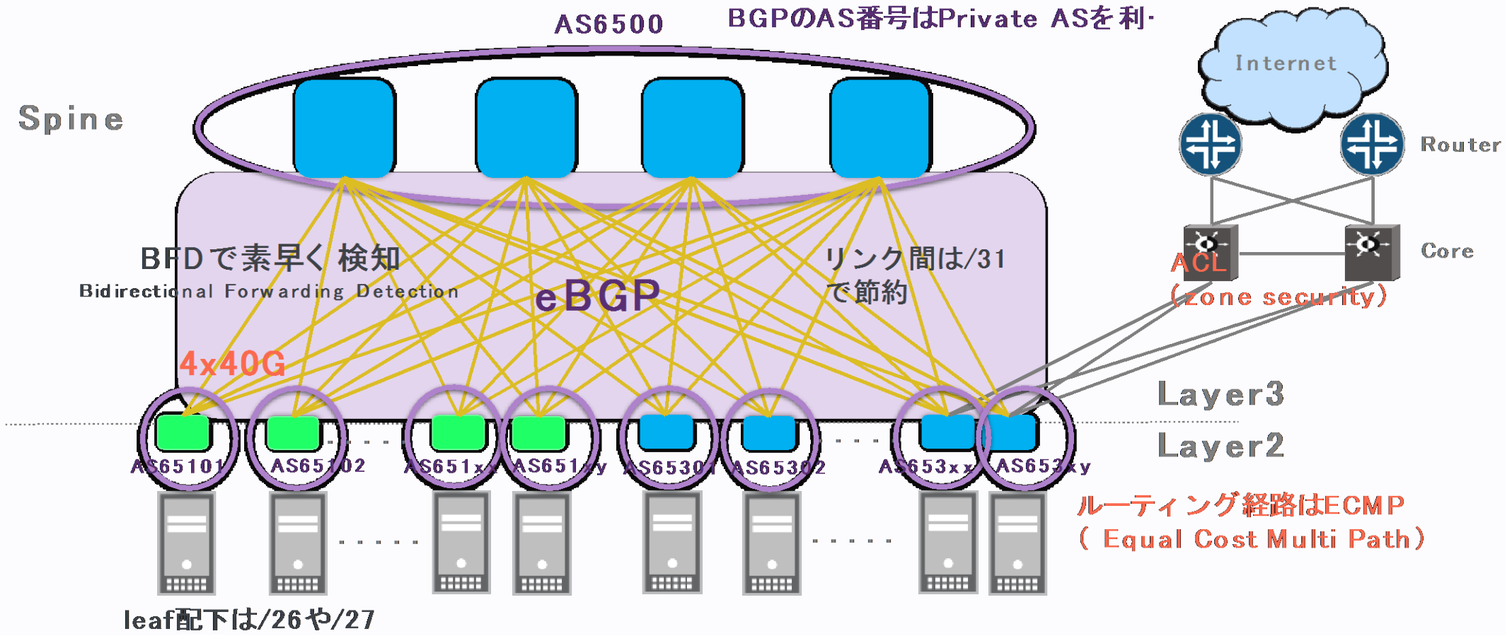


"Introducing data center fabric, the next-generation Facebook data center network". Facebook Code. https://code.facebook.com/posts/360346274145943/introducing-data-center-fabric-the-next-generation-facebook-data-center-network/. (10/06/2016).

SiteOPS
Yahoo! JAPAN Site Operations

JANOG38 振り返り

- 概要 / 構成
  - US DC
  - Spine:某A社シャーシ / Leaf:某A社とWhite Box半々
  - Spine - Leaf間は BGP
  - Leaf の Uplink は 40Gx4 = 160G



Yahoo! JAPAN Site Operations

JANOG38 振り返り

- Hadoop テスト

5TB Terasort

40TB Distcp

SiteOPS
Yahoo! JAPAN Site Operations

JANOG38 振り返り

- 構築方法
  - Zero Touch Provisioning (ZTP) による構築
  - 社内管理ツールと連携し、個別の設定を作成する必要がある
  - 作成した設定をDHCPで取得し設定反映
  - Spine や Mlag 構成は手動での設定

JANOG38 振り返り

- これからの課題と展望
  - ACL問題
    - 社内間の通信はセグメントごとにSVIでACL管理
    - コアスイッチで膨大なACL設定が必要
    - Spine-LeafのLeaf側へ設定をもっていくか、あるいはホスト単位か
  - 今後の展望
    - Hadoopネットワークのみではなく、その他のProductionへ展開
    - SpineやLeafのアップリンクが落ちても深夜対応しない構成へ！

JANOG38 振り返り

- これからの課題と展望
  - ACL問題
    - 社内間の通信はセグメントごとにSVIでACL管理

詳細は「 JANOG38 ヤフーのIP CLOSネットワーク 」を参照
(https://www.janog.gr.jp/meeting/janog38/download_file/clos.pdf)

  - 今後の展望
    - Hadoopネットワークのみではなく、その他のProductionへ展開
    - SpineやLeafのアップリンクが落ちても深夜対応しない構成へ！

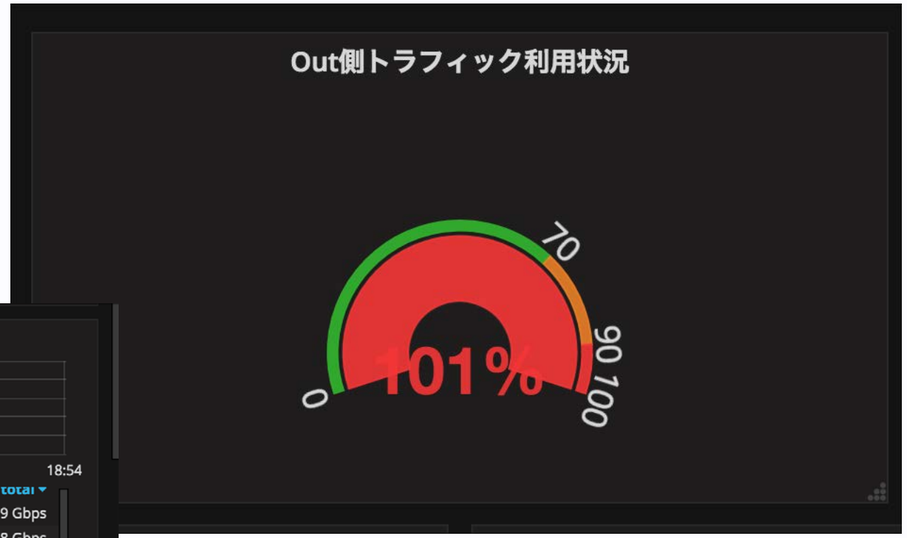
アメリカ拠点のその後

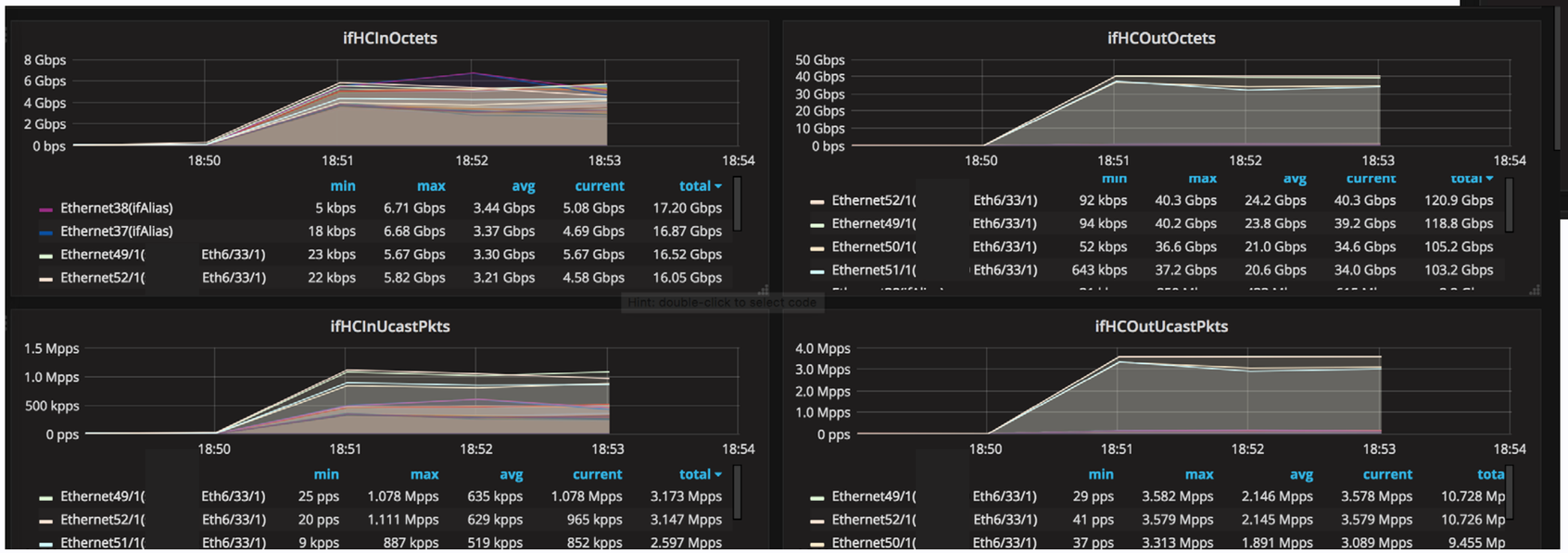
アメリカ拠点のその後

- 40ラック増強
  - 2017年10月にHadoopのコンピュートノードを1400台 (1ラックあたり約40台)
  - サーバはOCP
  - プレスリリース
    - http://www.ctc-g.co.jp/news/press/20171013a.html

アメリカ拠点のその後

- 増強したラックでのトラフィックテスト
  - コンピュートノードのみのため、データのローカリティが低い構成
  - 構築時の検証と同じく、
    Uplinkの帯域をFullで出せるのを確認





SiteOPS
Yahoo! JAPAN Site Operations

アメリカ拠点のその後

- 遭遇したトラブル
  - ラック納品されたMgmtスイッチのコンソールケーブルのMicro USBアダプタが一緒に納品されず。。。。。。
  - MgmtスイッチもZTPで設定予定だったため、コンソールの配線がなくても問題ない想定だったが想定通りにはいかず。
  - しかし、何もトラブルシューティングができないので、納品一日目はただ納品を見つめるだけで過ごすことに。

来ないなぁ。

SiteOPS
Yahoo! JAPAN Site Operations

アメリカ拠点のその後

- これからの課題と展望
  - ACL問題
    - 社内間の通信はセグメントごとにSVIでACL管理
    - コアスイッチで膨大なACL設定が必要
    - Spine-LeafのLeaf側へ設定をもっていくか、あるいはホスト単位か
  - 今後の展望
    - Hadoopネットワークのみではなく、その他のProductionへ展開
    - SpineやLeafのアップリンクが落ちても深夜対応しない構成へ！

アメリカ拠点のその後

- これからの課題と展望
  - ACL問題

**冗長性が通常の構成よりも高いため、
実際の運用でも1本程度Uplinkが落ちても
翌営業日対応で問題なし**

- Hadoopネットワークのみではなく、その他のProductionへ展開
- SpineやLeafのアップリンクが落ちても深夜対応しない構成へ！

IP CLOS ネットワーク
国内導入事例

IP CLOS ネットワーク 国内導入事例

P 18

- 国内拠点事例1
  - BCP Hadoop用ネットワーク(新規構築)
- 国内拠点事例2
  - サービス専用Hadoop用ネットワーク(リプレイス)

IP CLOS ネットワーク 国内導入事例

P 19

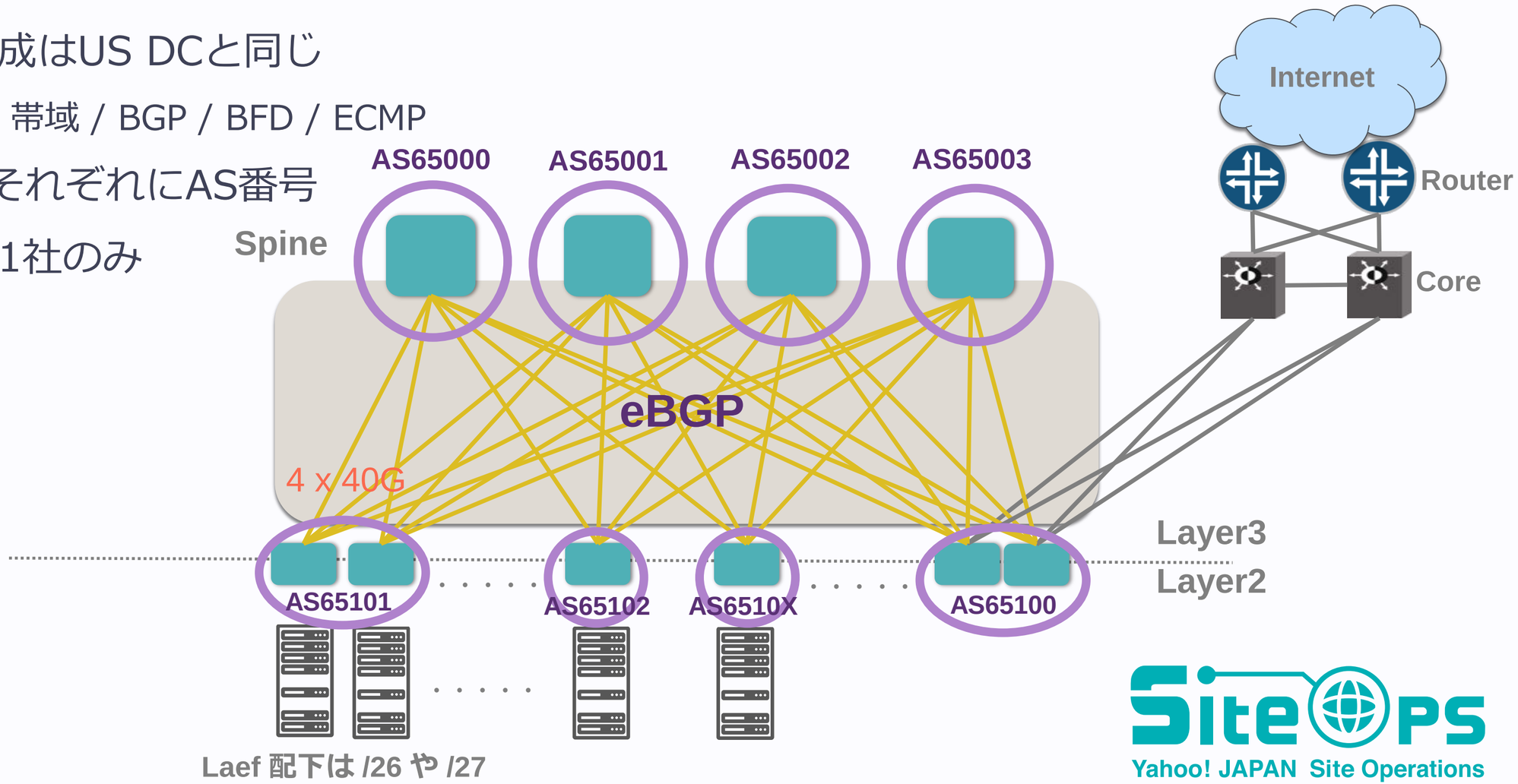
- 国内拠点事例1
  - BCP Hadoop用ネットワーク(新規構築)
- 国内拠点事例2
  - サービス専用Hadoop用ネットワーク(リプレイス)

IP CLOS ネットワーク 国内導入事例

- 国内拠点事例1(要件)
  - BCP Hadoop用のネットワーク
  - レスポンスが求められるため国内に必要
  - 利用が一部のサービスに限られる
    - 大規模なものは必要としない
  - サービス展開によっては拡張する可能性があるため、
    拡張可能な構成が望ましい

IP CLOS ネットワーク 国内導入事例

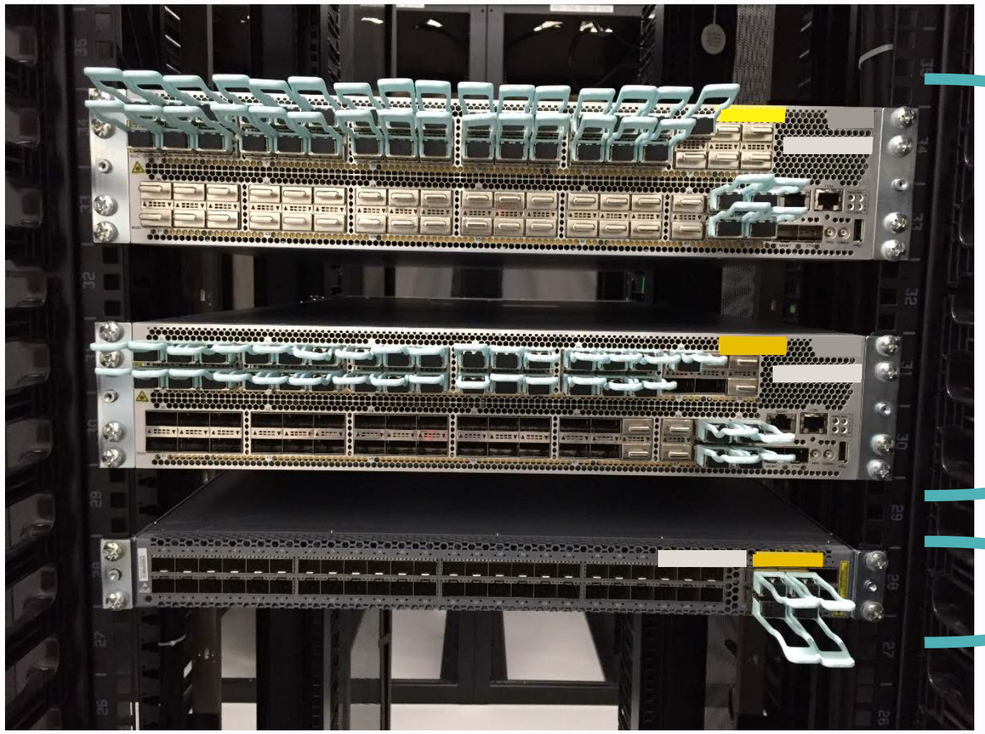
- 国内拠点事例1(構成)
  - 基本的な構成はUS DCと同じ
    - 2層構造 / 帯域 / BGP / BFD / ECMP
  - 各SpineにそれぞれにAS番号
  - 採用機種は1社のみ



Yahoo! JAPAN Site Operations

IP CLOS ネットワーク 国内導入事例

- 国内拠点事例1(構成)



配線前

配線後

Yahoo! JAPAN Site Operations

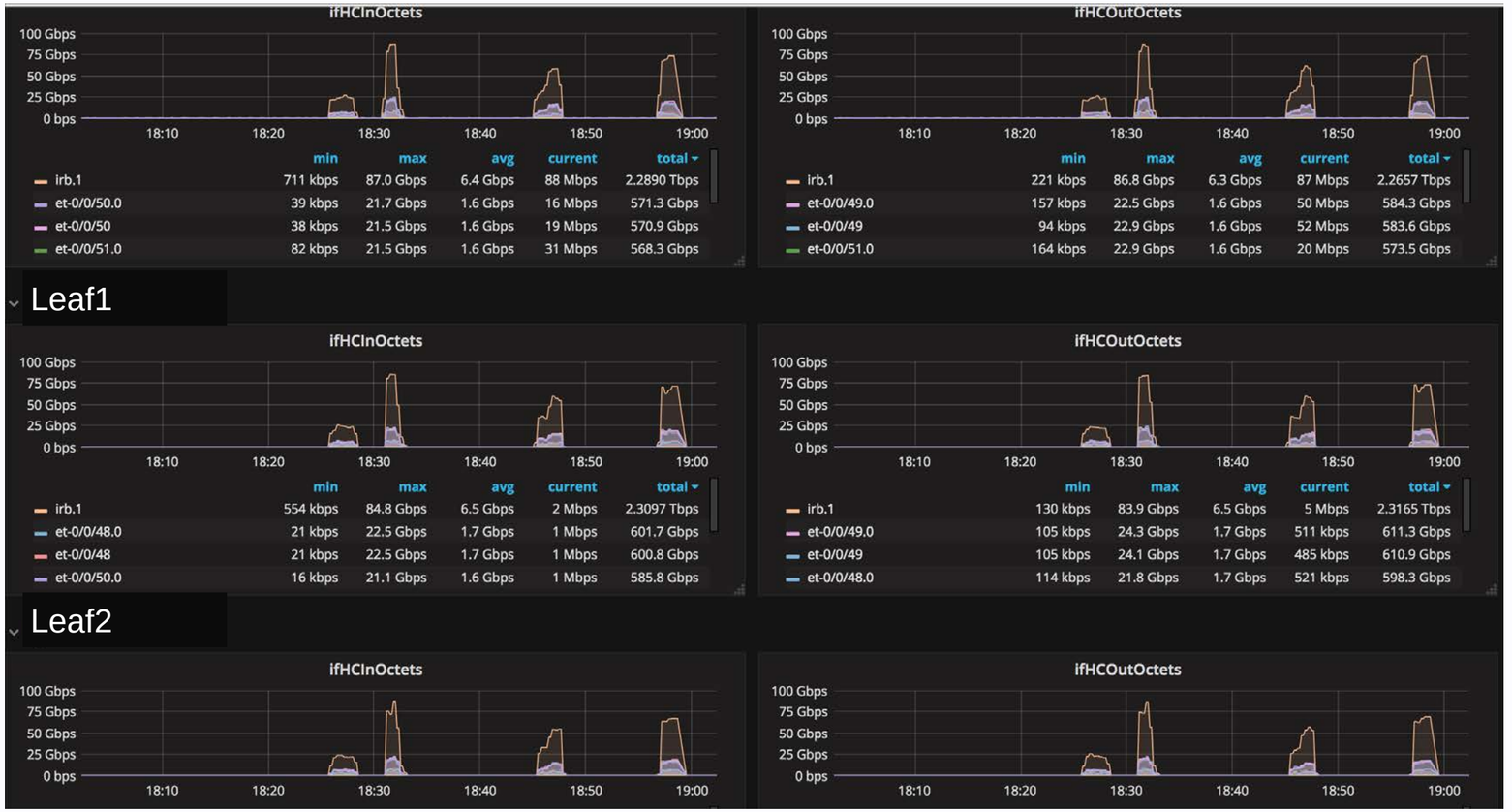
IP CLOS ネットワーク 国内導入事例

P 23

- 国内拠点事例1(構築)
  - 機器を採用した企業のOSSツールを用いて構築
  - US DC とは違いコンフィグを台数分作成せず、
    プールしたIP・AS番号や基本の設定を定義すれば設定を作成してくれる
  - Spineは手動で設定が**不要**になった
  - ツールで対応しきれていない部分は改修

IP CLOS ネットワーク 国内導入事例

- 国内拠点事例1(トラフィックテスト)
  - 初採用の機種だったが、問題なくトラフィックが出せることを確認



IP CLOS ネットワーク 国内導入事例

P 25

- 国内拠点事例1
  - BCP Hadoop用ネットワーク(新規構築)
- 国内拠点事例2
  - サービス専用Hadoop用ネットワーク(リプレイス)

IP CLOS ネットワーク 国内導入事例

P 26

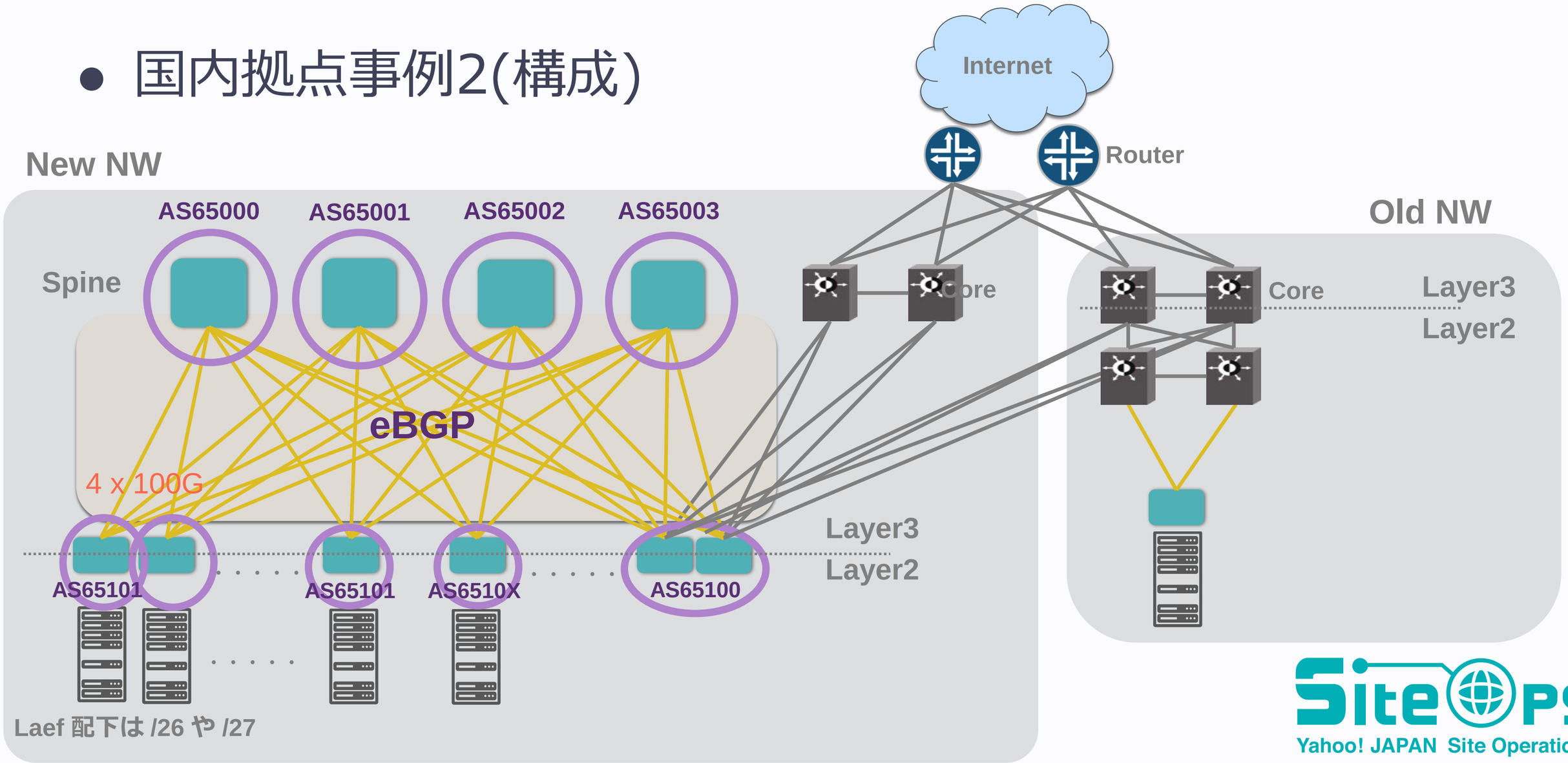
- 国内拠点事例2(要件)
  - 特定サービス向け Hadoop用ネットワーク
  - すでに動いているネットワーク (L2 Fabric)を
    IP CLOS ネットワークの構成にしたい
  - 現在のラック数で200ラック
    - 拡張の可能性はあり

IP CLOS ネットワーク 国内導入事例

- 国内拠点事例2(構成)
  - 具体的な構成はいままでと基本一緒
  - Uplink が **4 x 100G = 400G**
    - Uplink は 6 x 100G まで増速可能
  - 採用機種は1社のみ

IP CLOS ネットワーク 国内導入事例

- 国内拠点事例2(構成)



Yahoo! JAPAN Site Operations

IP CLOS ネットワーク 国内導入事例

- 国内拠点事例2(構成)
  - 今後増強予定
  - サーバも順次移行

Yahoo! JAPAN Site Operations

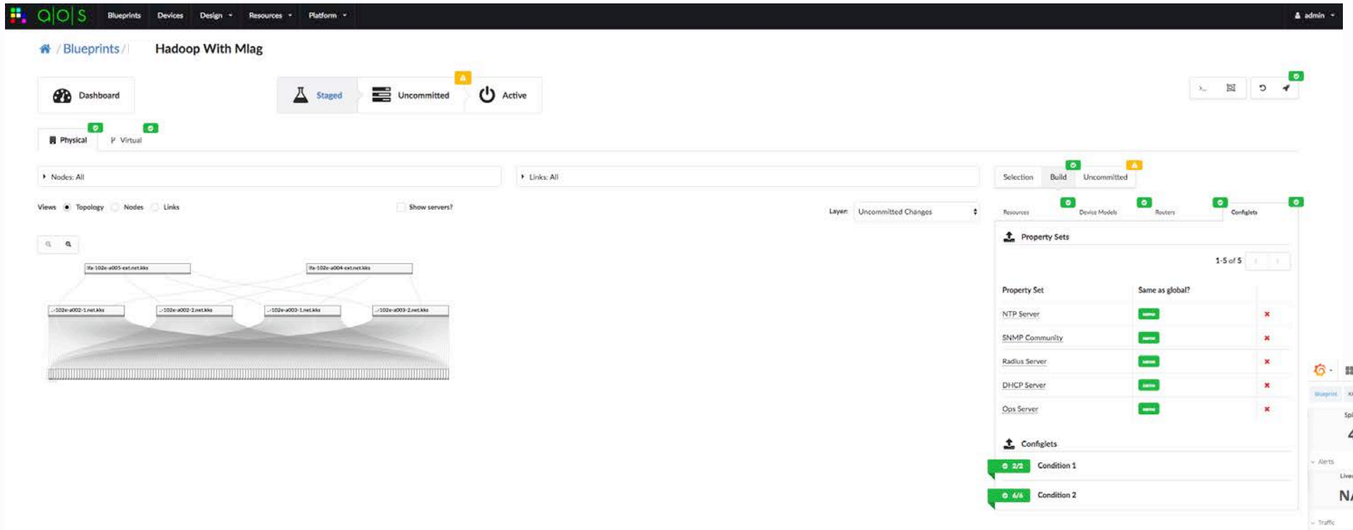
IP CLOS ネットワーク 国内導入事例

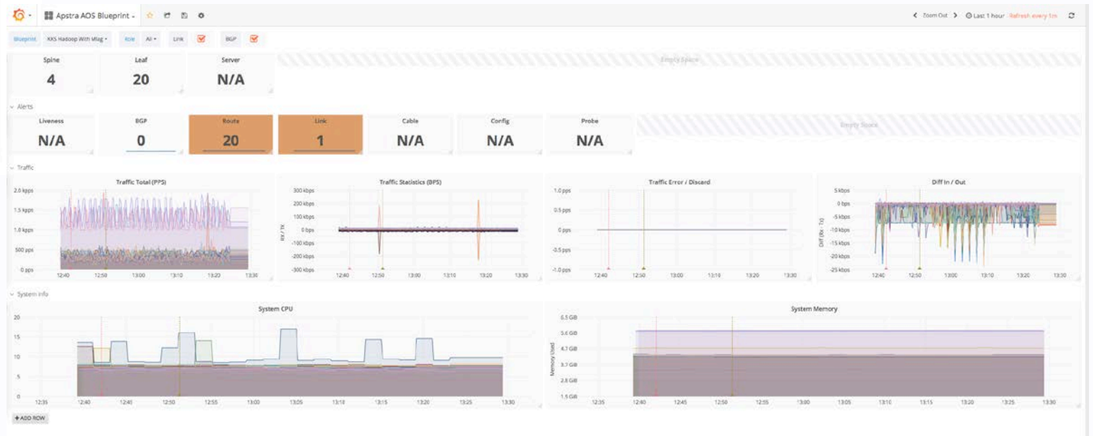
P 30

- 国内拠点事例2(構築)
  - 商用ツールを用いて構築
  - プールしたIP・AS番号や基本の設定を定義すれば設定を作成し、設定投入まで行ってくれる
  - SpineだけでなくMlagの手動での設定も**不要**になった
  - マルチベンダー対応

IP CLOS ネットワーク 国内導入事例

- 国内拠点事例2(構築)





IP CLOS ネットワーク
全面展開への課題

IP CLOS ネットワーク 全面展開への課題

- これからの課題と展望
  - ACL問題
    - 社内間の通信はセグメントごとにSVIでACL管理
    - コアスイッチで膨大なACL設定が必要
    - Spine-LeafのLeaf側へ設定をもっていくか、あるいはホスト単位か
  - 今後の展望
    - Hadoopネットワークのみではなく、その他のProductionへ展開
    - SpineやLeafのアップリンクが落ちても深夜対応しない構成へ！

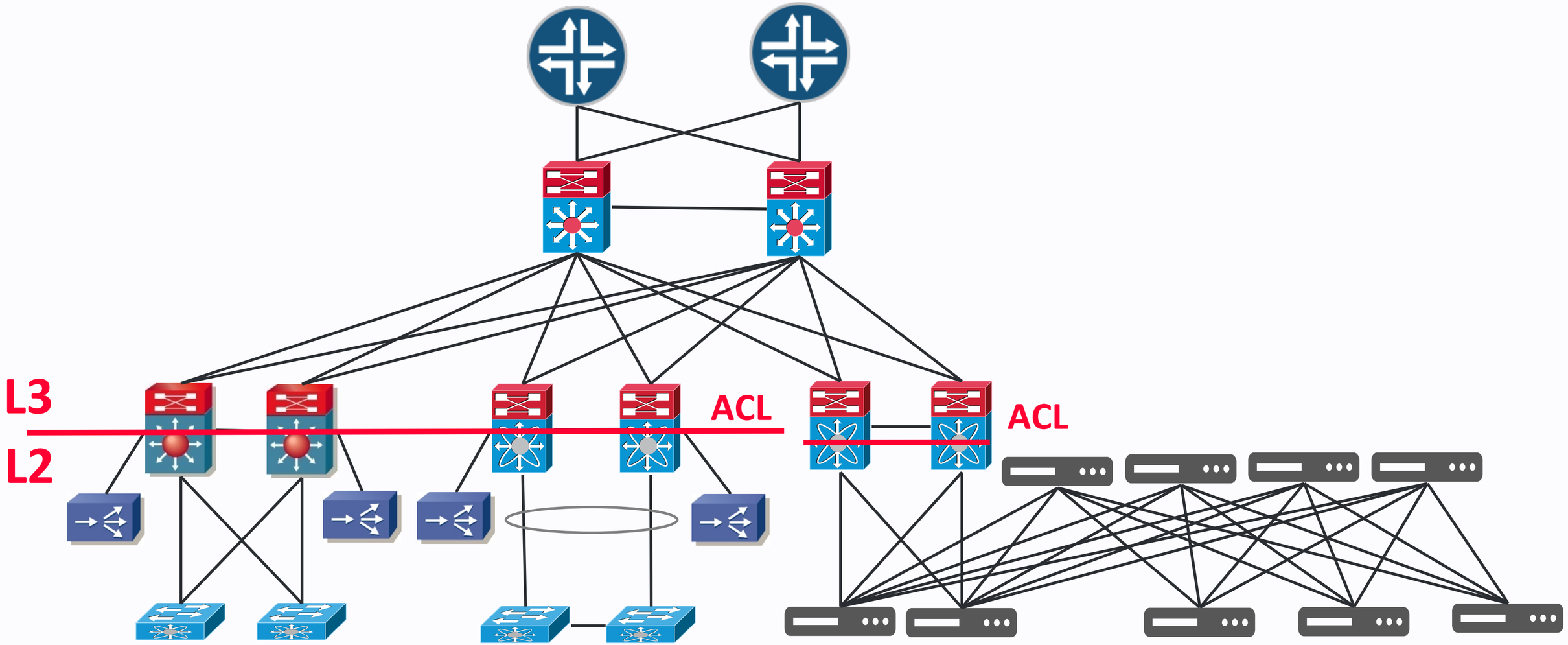
IP CLOS ネットワーク 全面展開への課題 P 34

- これからの課題と展望
  - ACL問題
    - 社内間の通信はセグメントごとにSVIでACL管理
    - コアスイッチで膨大なACL設定が必要
    - Spine-LeafのLeaf側へ設定をもっていくか、あるいはホスト単位か
  - 今後の展望
    - Hadoopネットワークのみではなく、その他のProductionへ展開
    - SpineやLeafのアップリンクが落ちても深夜対応しない構成へ！

Yahoo! JAPAN Site Operations

IP CLOS ネットワーク 全面展開への課題

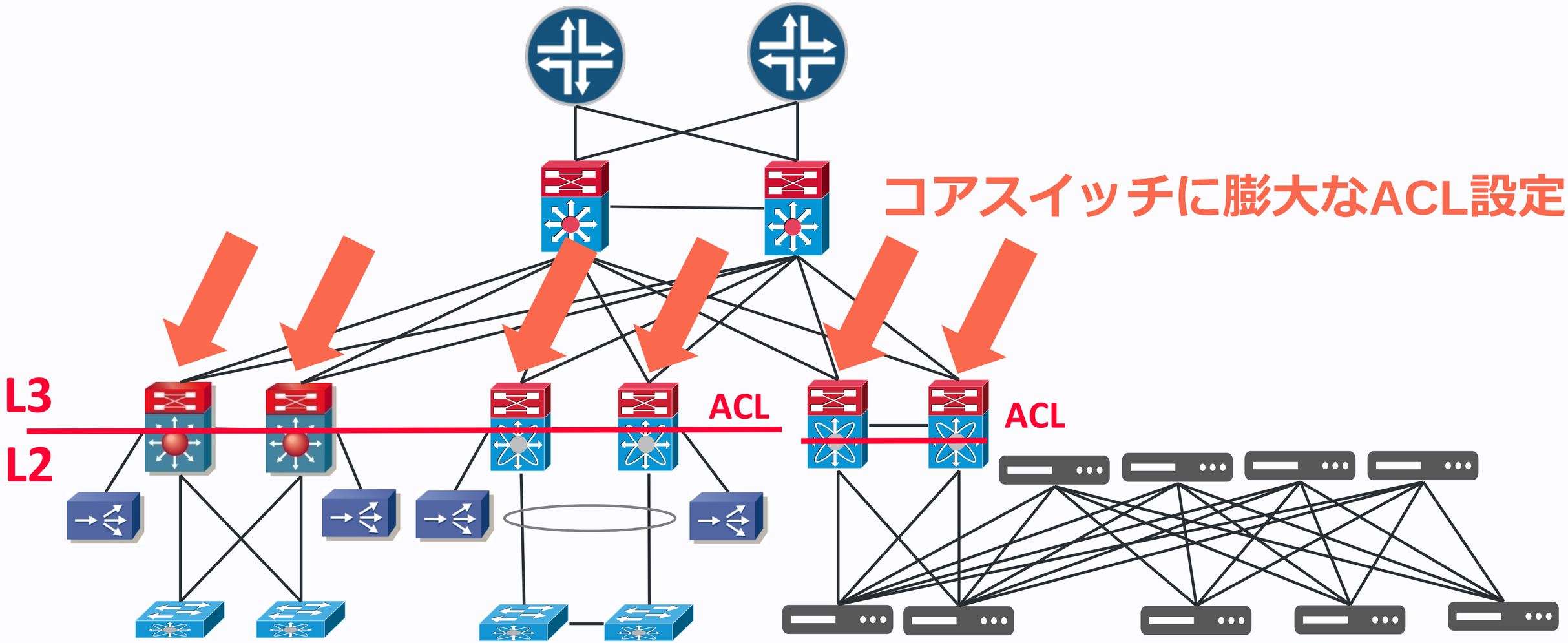
P 35



Yahoo! JAPAN Site Operations

IP CLOS ネットワーク 全面展開への課題

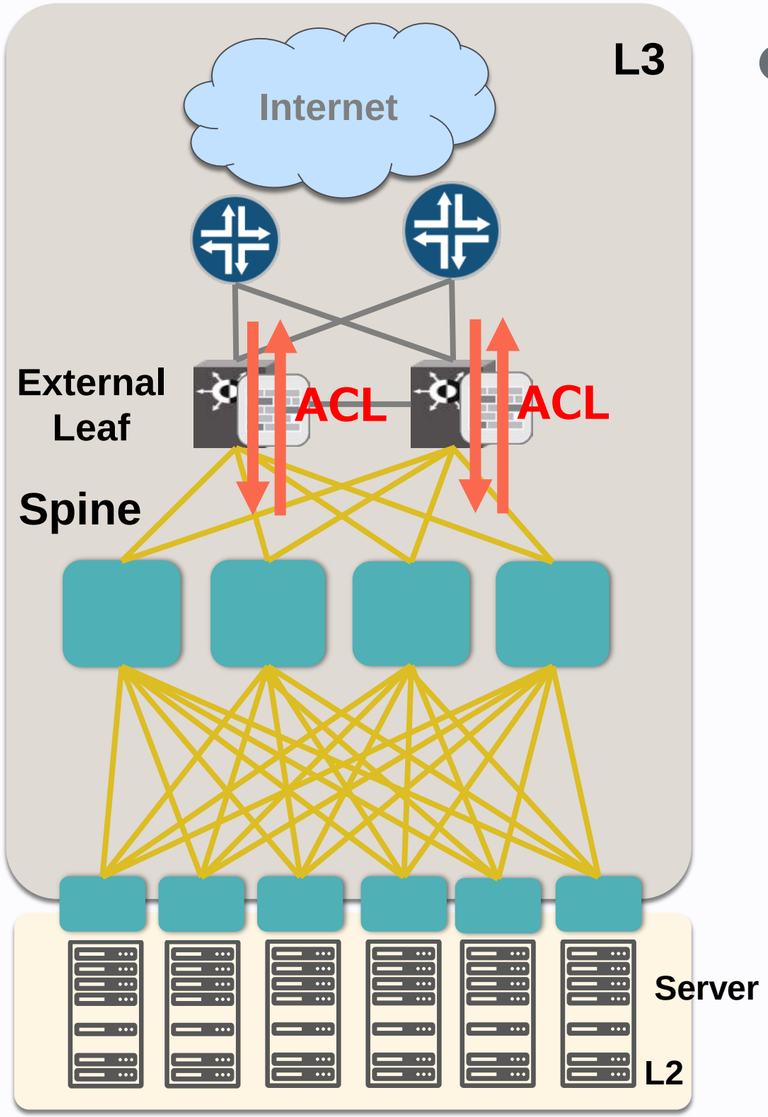
P 36



Yahoo! JAPAN Site Operations

IP CLOS ネットワーク 全面展開への課題

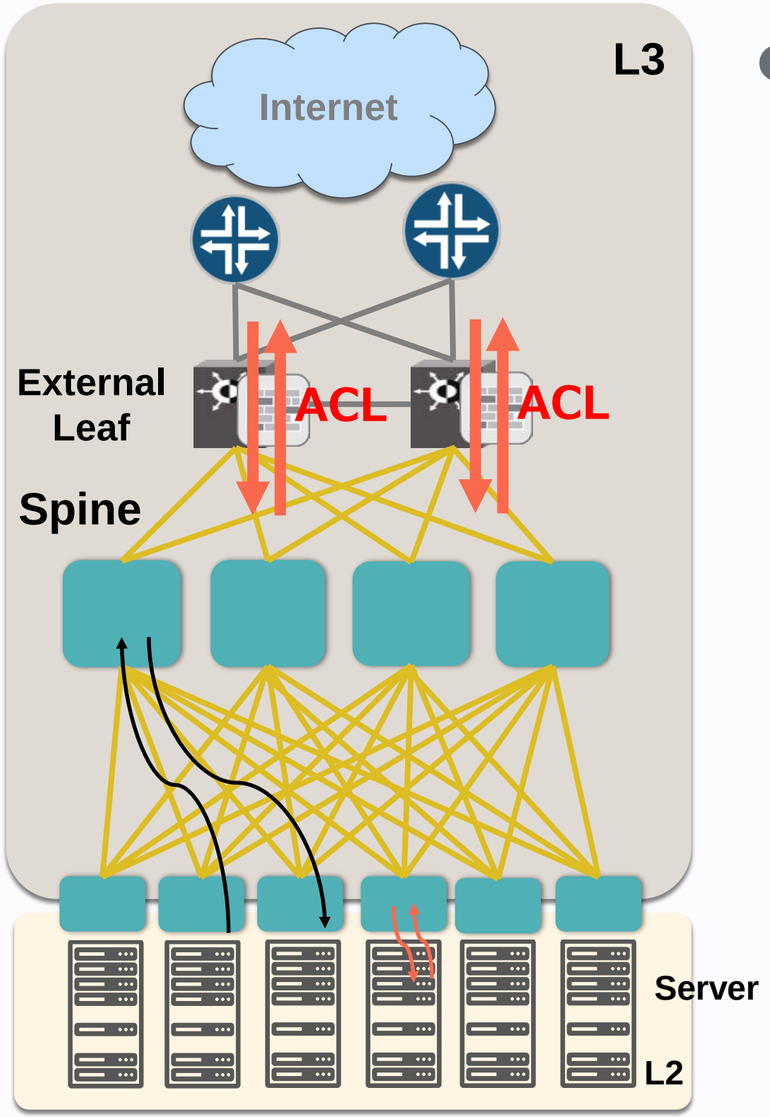
P 37



- 現状
  - 単一サービス用のIP CLOS ネットワーク
    - 基本的にHadoopで利用
  - ACL設定箇所
    - External leaf = 外部との境界

IP CLOS ネットワーク 全面展開への課題

P 38



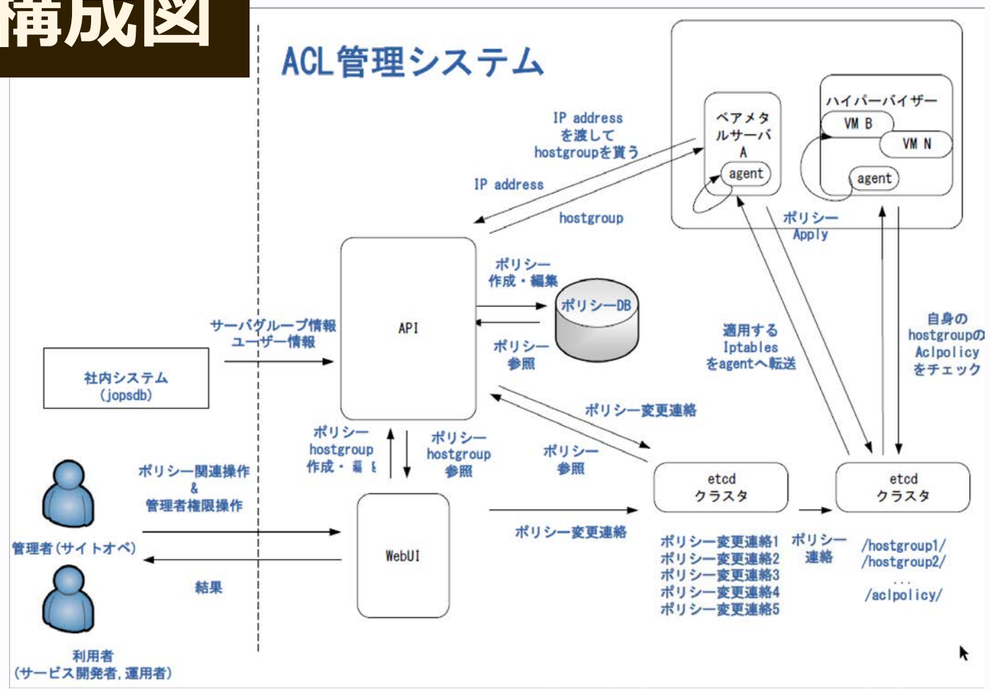
- 現状
  - ACLがIP CLOSと外部との境界の場合
    - 構成内のサーバ同士の通信制御ができない
  - External leafでのACLは必要最低限で数多くは設定できない
    - Spine, leaf等でたくさんACLを入れるには高性能スイッチが必要で高コスト

IP CLOS ネットワーク 全面展開への課題

P 39

- この問題を解決するためにヤフーでは
  サーバ・仮想マシンにACLをかけるためのシステムを開発中
- **コアやIP CLOSネットワークでのACL問題が解決**

システム構成図



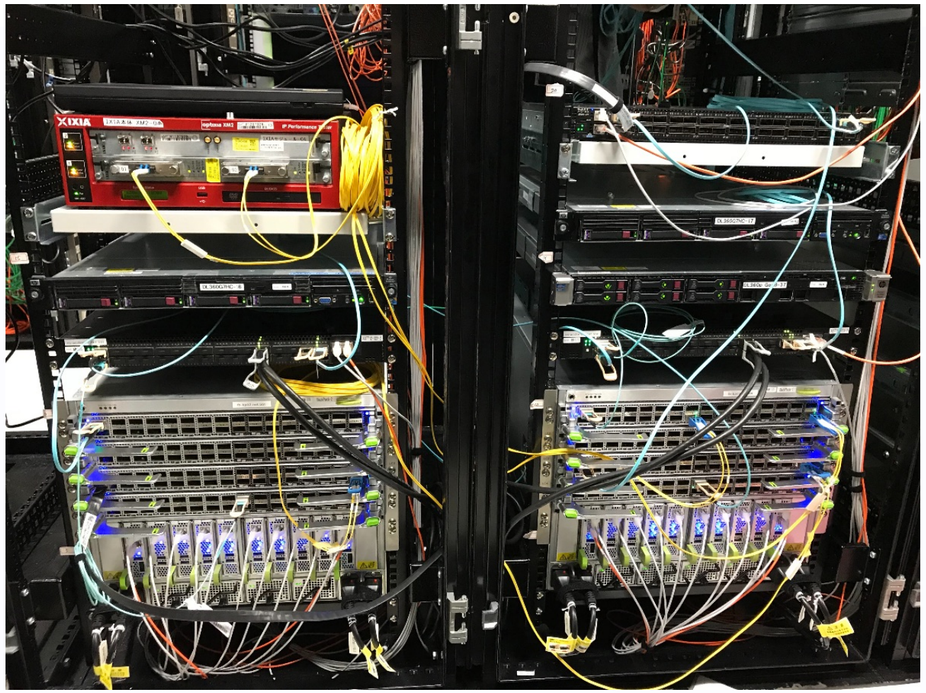
Network Lab

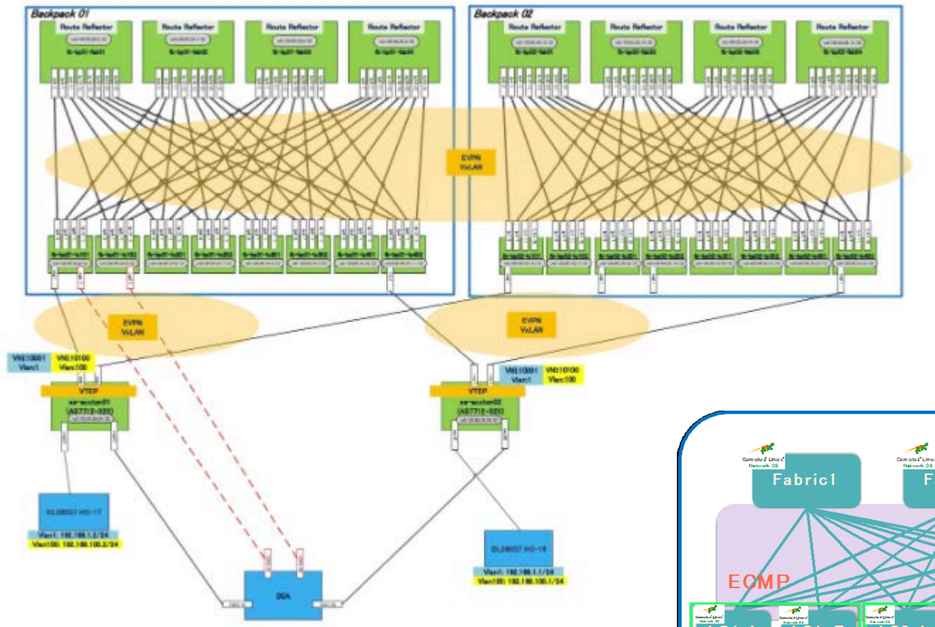
Network Lab

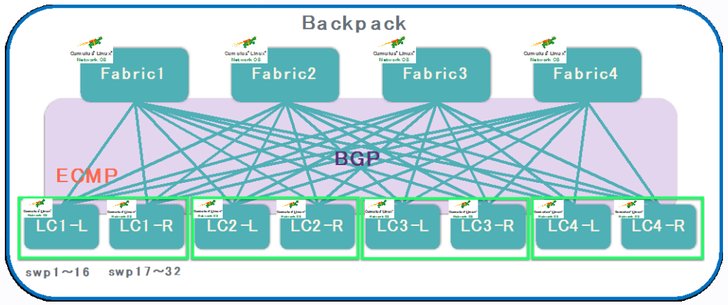
- BackPack

Yahoo! JAPANにOCPネットワークスイッチ「Backpack」を提供

Facebookで稼働する最先端機器を導入し、ITインフラの拡張性を高める





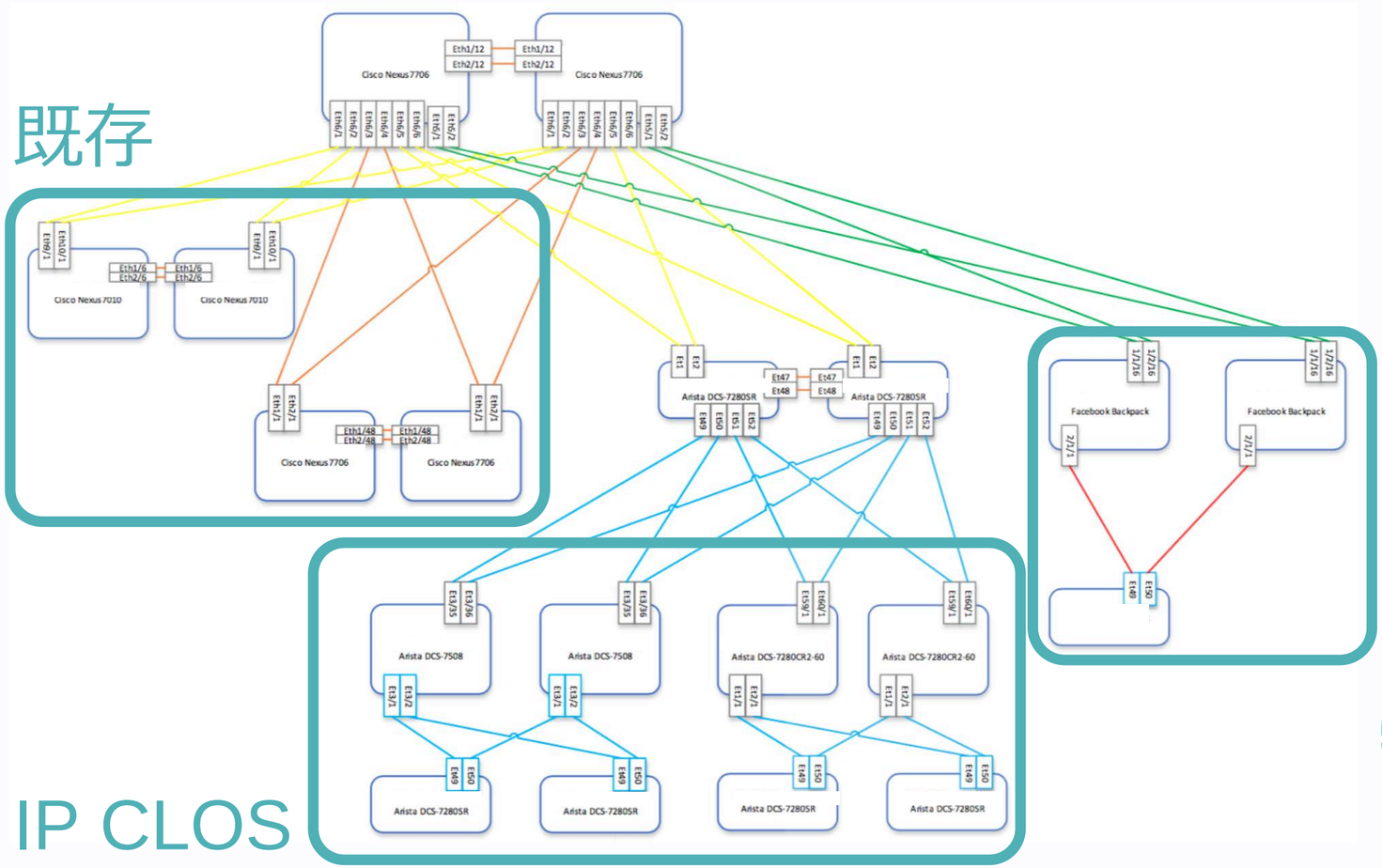


- http://www.ctc-g.co.jp/news/press/20170616a.html
- https://code.facebook.com/posts/864213503715814/introducing-backpack-our-second-generation-modular-open-switch/
- https://techblog.yahoo.co.jp/advent-calendar-2017/datacenternetwork_backpack/

SiteOPS
Yahoo! JAPAN Site Operations

Network Lab

- 新Network Lab環境
  - 既存・IP CLOSネットワーク環境をLabに再現



まとめ

- アメリカ拠点その後
  - 深夜対応の軽減 / Leaf増強 / トラフィックテスト
- 国内導入事例
  - Hadoop 領域で国内でも順調に導入
- 全面展開への課題
  - ホスト単位のACLシステムを開発中
- Network Lab
  - 既存/IP CLOS/BackPack を含めた構成で今後も様々な技術を検証予定

今後

今後

- IP CLOS ネットワークが

  有効なプロダクトには積極的に導入予定

  - Hadoop Eco System
  - Storage

Thank you for your kind attention.

Yahoo! JAPAN Site Operations