

短期間で開催される大会/イベントにおける バックボーン構築について

ICTトラブルシューティングコンテスト運営委員
in JANOG42



アジェンダ

1. 自己紹介
2. ICTトラブルシューティングコンテストとは
3. 各回の会場について
4. チームビルディング
5. バックボーン設計
 - ネットワークトポロジの紹介
 - Wi-Fiの苦悩
6. ICTSC 2018



議論したいこと

- Wi-Fiについて
 - 準備期間中のテストどうしてますか？大人数を再現できない！
 - 2.4GHzとうまく付き合えていますか？
- 短期間プロジェクトのチームビルディングって……
- ICTSC 独自技術 NAVT (詳細後述) はイケてますか
- 新しい技術導入のノウハウ
 - 安定を取るか技術を取るか
- イケてる技術ないですか？
 - 例) データセンタアーキテクチャ(RFC5549など)

自己紹介



中西 建登

- 電気通信大学 大学院 2年 情報学専攻
- JANOG 初参加
 - 若者支援プログラムありがとうございます！
- 第4回からほぼ全部を担当
- 大学ではブロックチェーン/スマートコントラクトの研究



中川 稜

- 情報科学専門学校 情報セキュリティ学科3年
- JANOGには若者支援プログラムで39,40,41(インフルで欠席)
- ICTSCには前回大会から運営
- 企業・自治体・学校のネットワーク構築をする零細企業でバイトしている






黒岩 稜

- 東京大学大学院総合文化研究科広域科学専攻修士1年
- JANOG 初参加
 - 若者支援プログラム本当にありがとうございます！
- 第7-9回でクラウド基盤を担当
- 大学院では人工知能を研究





川原 大輝



-  HEARTBEATS 技術開発室 バイト
- 電気通信大学 総合情報学科 4年
- JANOG 初参加

- 第6～9回で運営として参加
- 自宅ラック勢 (右図)
 - たのしい





川原 大輝

-  HEARTBEATS 技術開発室 バイト
- 電気通信大学 総合情報学科 4年
- JANOG 初参加
 -  ありがとうございます
- 第6 ~ 9回で運営として参加
- 自宅ラック勢 (右図)
 - たのしい



ICTトラブルシューティングコンテストとは



ICTトラブルシューティングコンテストとは

- 2014年から実施しているインフラ技術を競う学生向けコンテスト
 - 2018年3月に9回目(ICTSC9)を実施しました
- 目的: 人材の発掘、育成やエンジニアコミュニティの形成を推進すること
- 学生が運営委員、社会人が実行委員 (補助)

ICTトラブルシューティングコンテストとは: <http://icttoracon.net/>トラブルシューティングコンテストとは
前回行われたコンテストのレポート: <http://icttoracon.net/archives/5831>



数字で見る ICTSC

参加者数 : 約 100 人

VM数 : 約 600 台

物理機材 : 約 150 台

消費電力 : 20250 W



開催までの流れ @ ICTSC9

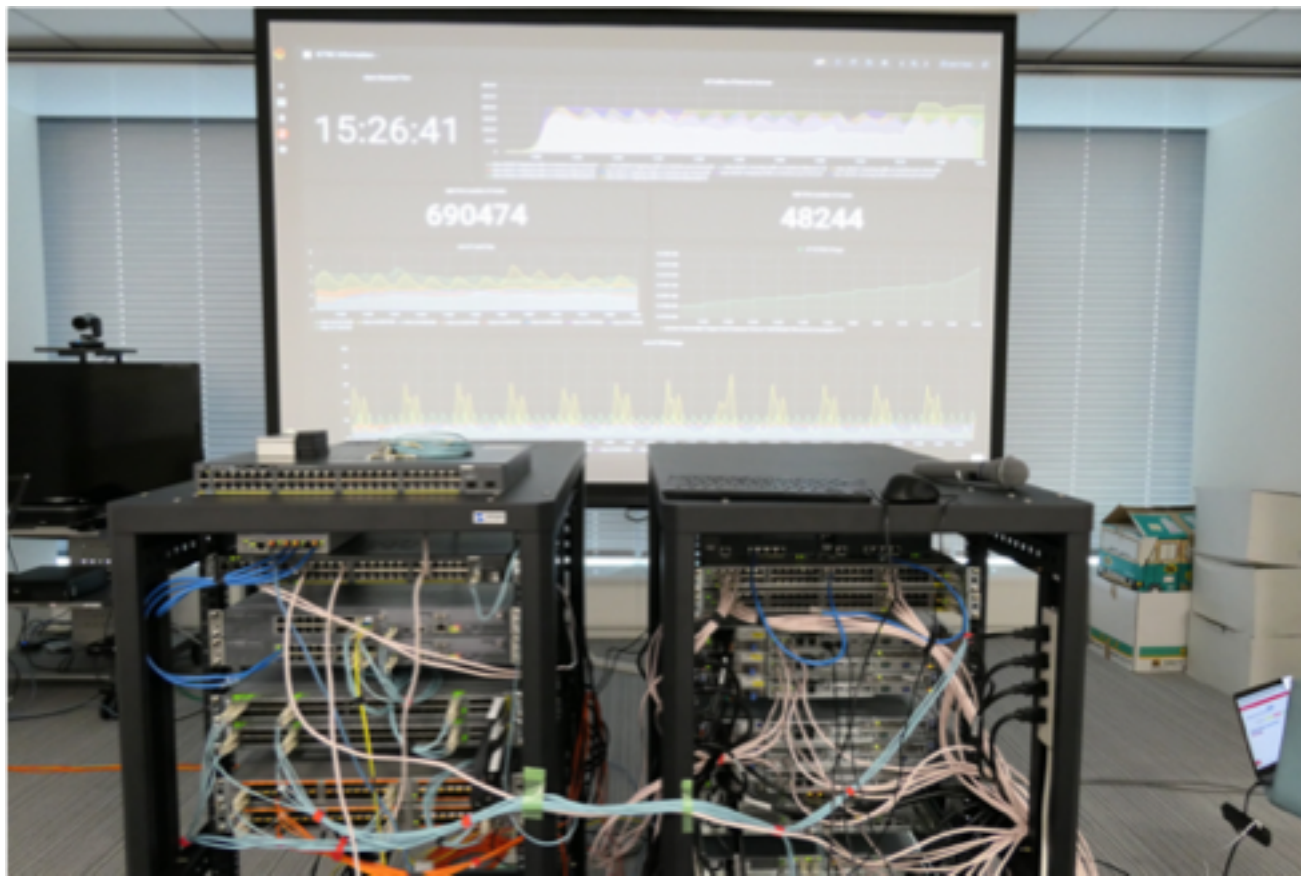
2017/10	2017/11	2017/12	2018/01	2018/02	2018/03
10/28 キックオフ					
	11/10 参加チーム募集開始				
		12/16 オンライン予選			
				02/16 ~ HotStage	
				03/03~04 コンテスト本戦	
					03/05 撤収日



ICTSC Hotstageの写真



ICTSC Hotstageの写真 (ラック)





ICTSC 本戦の写真 (参加者)





ICTSC 本戦の写真 (参加者)





ICTSC 本戦について

- 1チーム最大5人
- 本戦は全国から集まった15チームで競う
 - 最近ではオンライン予選を実施
- 出題範囲
 - ネットワーク問題
 - 物理的な機材や、VyOS、VXLANなど
 - サーバ問題
 - Webアプリケーション、OpenStack、Postfixなど
 - その他インフラ技術に関する問題



ICTSCのルール

- トラブルが起きた環境を全チームに提供
- 参加者は原因の特定と解決、ドキュメント形式で報告
- 採点結果の合計が高いチームが優勝
- 2日間午前午後で解ける問題を分割 @ ICTSC6
- 問題のジャンルでグルーピングし、自分のチームがとけた問題に応じて順番に開放 @ ICTSC7~8
- 問題のジャンルでグルーピングし、どこかのチームがとけた問題に応じて順番に開放 @ ICTSC9



問題の紹介：サーバ問題（ICTSC6より）

あなたはログイン不要で自由にファイルをアップロード/ダウンロードできるサービスを運用しています。

このサービスは今まで順調に稼働していたが、ある日を境に突然アップロードができなくなってしまった。原因究明と対処を行ってほしい。



問題の紹介：サーバ問題

- 情報：
 - SSH 先アドレス: 192.168.x.x
 - ユーザ名: admin
 - パスワード: hogehoge
- 解説：
 - `df -i`コマンドで inode 不足に気づく
 - ディスクを新規に作成して問題を解決し、アップロードを可能にする



問題の紹介：サーバ問題

XX学校の〇〇です。

サービスが動いてるサーバを調べたところ、アップロードができなくなった原因はinode不足によるものと思われます。

`df -i`でinodeの使用率を見たところ保存先のディスク /dev/sda1 のIfreeが0となっておりました。

ですので空いていたディスク /dev/sda2があったのでそちらに保存先を変更する応急的な処置をしました。

ご確認ください。



問題の紹介: ネットワーク問題 (ICTSC7より)

C社は攻撃を受けやすいサービスを運用しているため、攻撃をしてきたIPをすべてブラックリスト方式でカットする運用を行っている。

ある日、あるエラーが表示されたのち期待通りのパフォーマンスでL2SWが動作しなくなってしまった。原因を突き止めてほしい。

```
access-list 101 permit tcp 192.168.0.1 0.0.0.0 10.1.1.1 0.0.0.0 eq 1101
access-list 101 permit tcp 192.168.0.2 0.0.0.0 10.1.1.1 0.0.0.0 eq 1101
access-list 101 permit tcp 192.168.0.3 0.0.0.0 10.1.1.1 0.0.0.0 eq 1101
access-list 101 permit tcp 192.168.0.4 0.0.0.0 10.1.1.1 0.0.0.0 eq 1101
access-list 101 permit tcp 192.168.0.5 0.0.0.0 10.1.1.1 0.0.0.0 eq 1101
.
.
.
access-list 124 permit tcp 192.168.0.250 0.0.0.0 10.1.1.1 0.0.0.0 eq 1124
```



問題の紹介: ネットワーク問題

- 解説
 - エントリが多すぎてTCAM（スイッチのメモリ）で処理できずソフトウェアフォワーディングになるため通信が遅くなる
 - TCAM使用率を下げる
 - 無駄なACLエントリーを集約する
 - 無駄なACLのインターフェース割当を無効化する



ICTSC 最近の状況

- Wi-Fi、DNSなどのライフサーバの提供
- VM を使う問題
 - チーム数 * 問題 VM = $15 * 38 =$ 約 600 VMの展開 (7回)
 - 基本的に SSH で接続する
 - 物理機材や参加者のPCから L2 で接続させたい問題もある
- 物理機材を使う問題
 - チーム数 * 問題機材 = $15 * 9 =$ 約 150台の物理機材 (9回)
 - 基本的にコンソールで接続する



ICTSC 最近の状況

- 機材が増えているため電源設計がシビア
 - チーム数 * (問題機材 + PC) = 15 * 7.5 A = 112.5 A
 - バックボーン: (サーバ + ネットワーク) = 60 A
 - 協賛: 15 A
 - 運営: 15 A
 - 合計: 202.5 A
- バックボーンで障害が起きた際に一部のチームだけ解けるような構成にしてはいけない
 - サーバやネットワークの公平性の確保



ICTSC 最近の状況

- コンテストサイトに問題内容と各チームの得点を掲載する
 - 回答の送信と採点もコンテストサイト上で行う
- **参加者**は問題へ回答したい
 - コンテストサイトと問題環境とインターネットへ接続
- **運営**は出題・採点・監視したい
 - すべてのネットワークへ接続
- **見学者**は観戦したい
 - コンテストサイトとインターネットへ接続



機材について

- Cisco Networking Academy から派生した大会
 - 現在も問題に使う機材を多数お借りしている
- ~~要件に合わせて機材をお借りしたり~~
機材をお借りしてから要件を組み立てたり
 - いつもお力をお借りしています
 - 学生が交渉して機材をお借りすることも
- 具体的な機材については後ほど

学生 「こんな機材使ってみたい！」

大人 「任せとけ！」



コンテストサイトの開発開始 @ ICTSC6

- 過去大会では Redmine を用いて問題を提供
- より使いやすいインターフェース / システムを求め自作開始
- github.com/ictsc/ictsc-score-server にて OSS として公開
 - Ruby / Sinatra / Vue.js
 - 現在 (2018/07)もメンテナンスを行っている
- リアルタイムなスコア公開 / 問題のグルーピング等
- ICTSC6は負荷に耐えられずその場でチューニング
 - 何とか耐えきった！
- 毎回ルールに合わせて更新・修正



コンテストサイトの現在

ICTSC

ガイド メンバー チーム

トップ 問題 質問 解答 グラフ ログアウト

問題一覧

+ 新規グループ

+ 新規問題

各部署に複数の問題があります。

それぞれの問題に設定された基準を満たすと、次の問題が解放されます。

各部署から問題を選んで解答

解放されている問題を選択し、解答を行ってください。

運営が採点 (20分間)

運営が採点中はその問題を解けません。

採点結果を確認

満点でない場合は、高得点を目指して追加の解答も可能です。

問題解答の注意点

- 採点結果は、採点依頼を送信してから20分後に返ってきます。
- 採点中はその問題へ新たに解答することはできません。

電子通信網局 経路安定課

<p>プロバイダ業務に自信あります! (admin)</p> <p>内訳 満点 100 基準点 80</p> <p>▼0チーム正解</p>	<p>新規プロジェクト 1 (admin)</p> <p>内訳 満点 150 基準点 100</p> <p>▼0チーム正解</p>	<p>明日から有給取るからよろしく!! (admin)</p> <p>内訳 満点 250 基準点 200</p> <p>▼0チーム正解</p>
---	---	---

電子通信網局 仮想機密伝送路課

<p>ドキッ!新人だらけのトラブル開発部署(admin)</p> <p>内訳 満点 100 基準点 80</p> <p>▼0チーム正解</p>	<p>ダイエット失敗(admin)</p> <p>内訳 満点 200 基準点 160</p> <p>▼0チーム正解</p>	<p>Pingは返ってくるのにwebが見れない!(admin)</p> <p>内訳 満点 300 基準点 240</p> <p>▼0チーム正解</p>
---	---	---

07:37:09



各回の会場について



会場 @ ICTSC6

NTT西日本大阪研修センター本館(PRISM)

- 電源: 詳細は禁則事項ですが、いっぱい
- スペース: 潤沢 (とっても広い)





会場 @ ICTSC7

NTT中央研修センタ

- 電源: 詳細はないですが、いっぱい
- スペース: 少し狭い
 - (何気ない一言がNTT東の方を傷つけてしまったらしく反省)





会場 @ ICTSC8

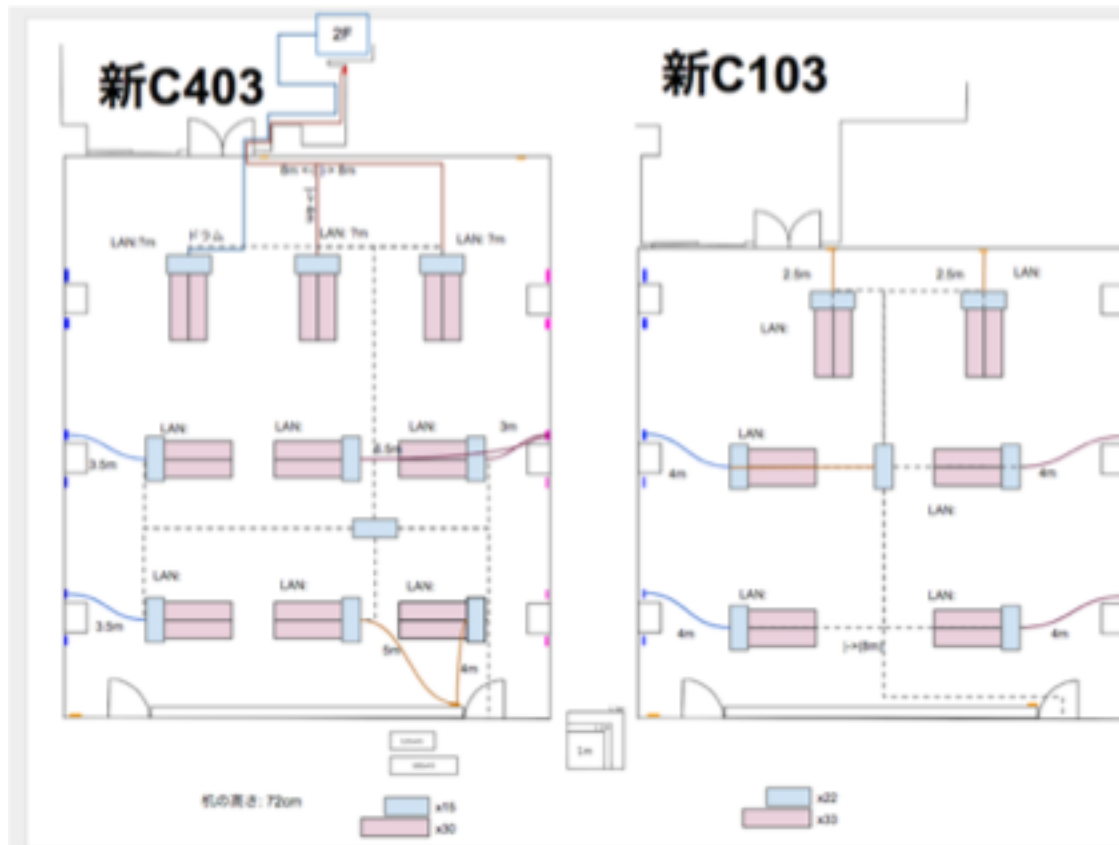
電気通信大学新C棟

- 電源: $15A * 3系統 * 4階層 = 180A$
- スペース: 4階層に別れていて、作業上少し不自由
 - ファイバーを階段沿いに敷いたり
 - 窓からUTPケーブル出して階移動したり





会場 @ ICTSC8





会場 @ ICTSC8





会場 @ ICTSC8





会場 @ ICTSC8





※ 電源容量が足りず本番でダウン @ ICTSC8

- 2日目の開始1時間後、突然サーバの入ったラックが異音と共に静かになる
- サーバと反比例してざわつく学生
- 様々な要因を考慮して、電圧を計測した結果がこちら





※ 電源容量が足りず本番でダウン @ ICTSC8





※ 電源容量が足りず本番でダウン @ ICTSC8





※ 電源容量が足りず本番でダウン @ ICTSC8

学生「うーん、まだ足りない…」



※ 電源容量が足りず本番でダウン @ ICTSC8

@ibucho神の一声



「女子トイレ入れるやつ居る？」



※ 電源容量が足りず本番でダウン @ ICTSC8





※ 電源容量が足りず本番でダウン @ ICTSC8

解決 😊



会場 @ ICTSC9

- さくらインターネット本社 グランフロント大阪タワー
 - 電源：禁則事項です
 - スペース：まあまあ広い





※ まるで成長していない... @ ICTSC9



チームビルディング



チームビルディング

- 「本業(=学業) ではない」 「フルリモートな」 「20名」の学生
- もちろんリーダーもマネジメント経験は薄い
- スキルセット、現場(≒修羅場?)経験もバラバラ

その中で「**大会開催**」という**結果**を出す必要

→なにより、楽しくやりたい！
良い思い出にしたい！



※ Hotstage 中に設計の大幅変更 @ ICTSC6

- 若気の至り
 - 原因はコアネットワークとクラウド基盤のそれぞれの設計者がコミュニケーション不足
 - IPv4 アサインが全て手戻り
 1. キレたコアネットワーク担当が (なぜか) IPv6 RA を流す
 2. (なぜか) CloudStack 崩壊 🎉
 3. クラウド基盤担当ブチギレ 🎉🎉
- 運営内の不和につながる
- ※ 睡眠不足で行動に整合性も取れていない



チームビルディング

- 今も正解があるのか模索中
 - 学生なので非常に流動的
 - ノウハウを引き継ぐ、温度感も引き継いでいく
- ロール毎にリーダーを設けてみたり
- HotStageから、ふせんを使ってカンバンを設置してみたり
- 「人間DNS」としてとにかく情報のポイントを渡してみたり
- マネジメント論の本を真面目に数冊読みあさったり
 - 「あるあるあるある～～～」となったリーダーが……



チームビルディング





チームビルディング

- 今も正解があるのか模索中
 - 学生なので非常に流動的
 - ノウハウを引き継ぐ、温度感も引き継いでいく
- ロール毎にリーダーを設けてみたり
- HotStageから、ふせんを使ってカンバンを設置してみたり
- 「人間DNS」としてとにかく情報のポイントを渡してみたり
- マネジメント論の本を真面目に数冊読みあさったり
 - 「あるあるあるある〜〜〜」となったリーダーが……

バックボーン設計



免責事項

- 断言的な表現をすることがありますが、
「当時の学生が」「当時の精神状態で」
「当時の技術力で」辿り着いた結果によるものです
 - 現在の事実とは異なる事があります





クラウド基盤

- CloudStack v4.8 @ ICTSC6
 - 必ず CloudStack のソフトウェアルータを作る必要がある
 - オーバヘッドが発生
 - 参加者に L2 で接続させたいので無理やりブリッジした
 - 直接 DB を編集する必要があった
 - VM 数のクォータを増やす UI/API が見つからない
 - 消した VM の IP アドレス割り当てが消えない
- OpenStack Newton - Pike @ ICTSC7 – ICTSC9
 - Packstackで毎回構築を行っている
- Kubernetes @ ICTSC9

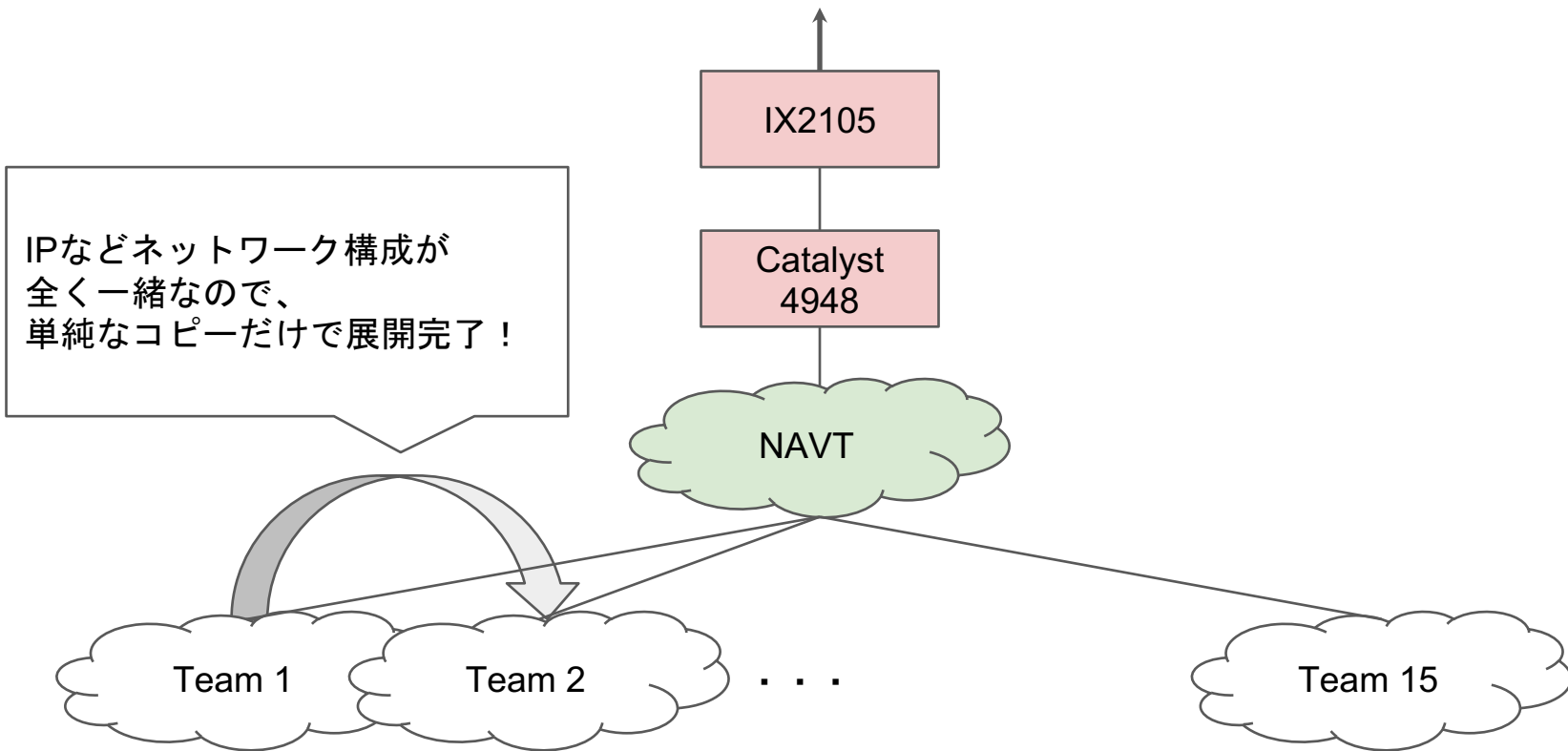


NAVT 誕生 @ ICTSC6

- VM を15チーム分プロビジョニングすることに課題があった
 - 全ての問題の構成ファイルを完成させる余裕がなかった
 - CloudStack が構成管理ツールに対応していなかった
 - IP をハードコードしたいなどの問題 VM 特有の特殊な要件
- NAVT という技術が構成管理ツールの使用を避けるために誕生
 - テンプレートイメージからの複製による VM の展開が可能になった

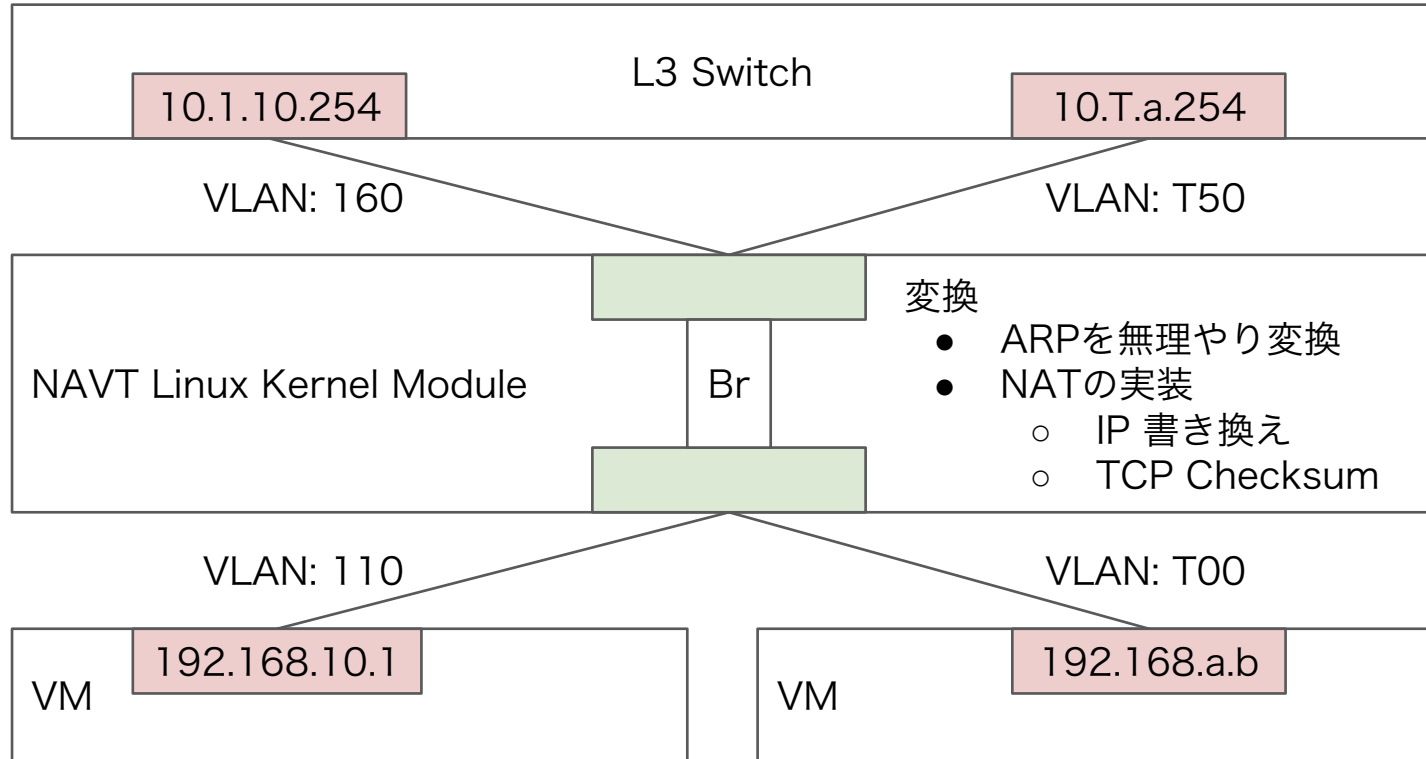


トポロジ図 (論理) @ ICTSC6





NAVT v6 (Network Address Vlan Translation)



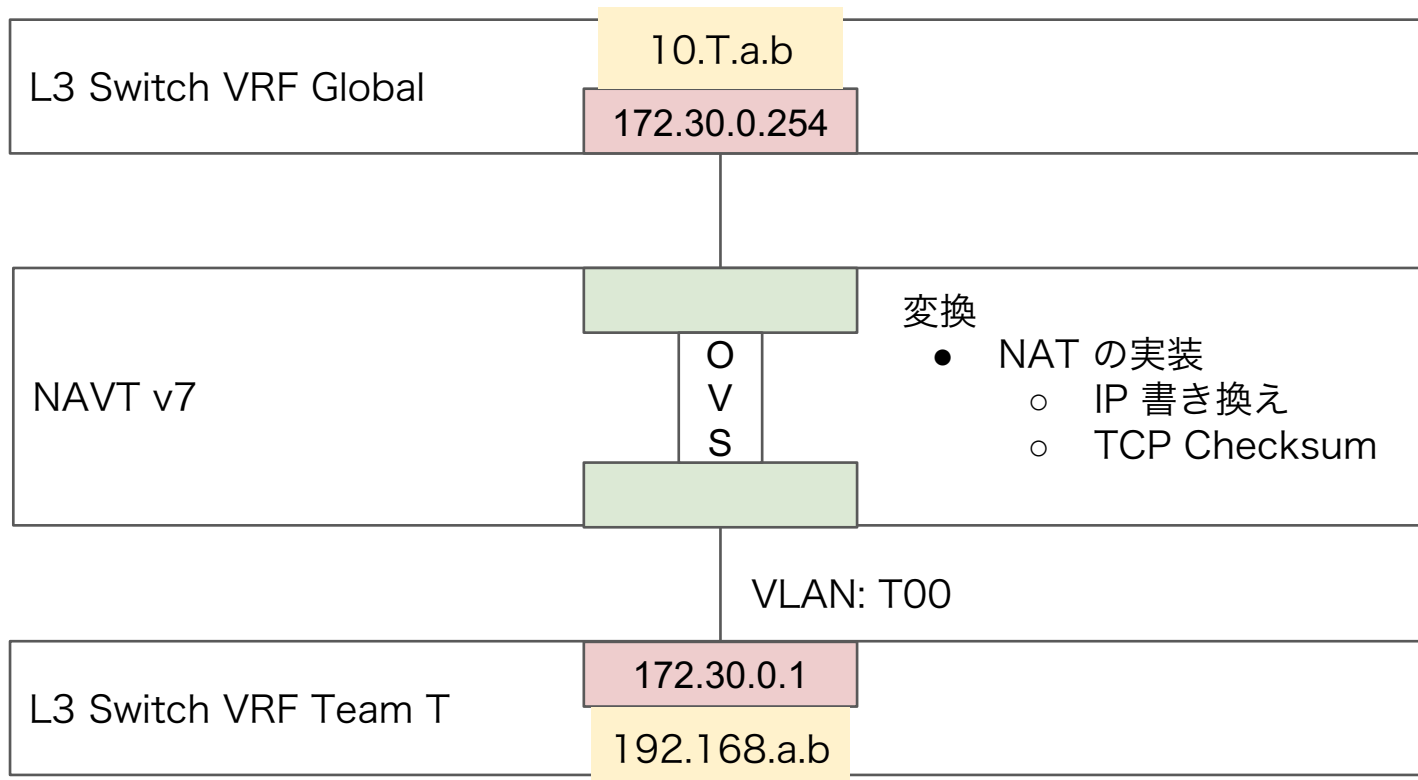


NAVT v7 (Network Address Vlan Translation)

- 第7回は OpenFlow 1.3 (Ryu SDN Controller)
+ Open Vswitch で実装
- ARP は静的解決をさせることで簡略化
- VRF も併用
 - 各チームへの通信を Null route を書くだけでよく
ACL が必要なくなる
 - 10.0.0.0/12, 192.168.0.0/24 -> Null



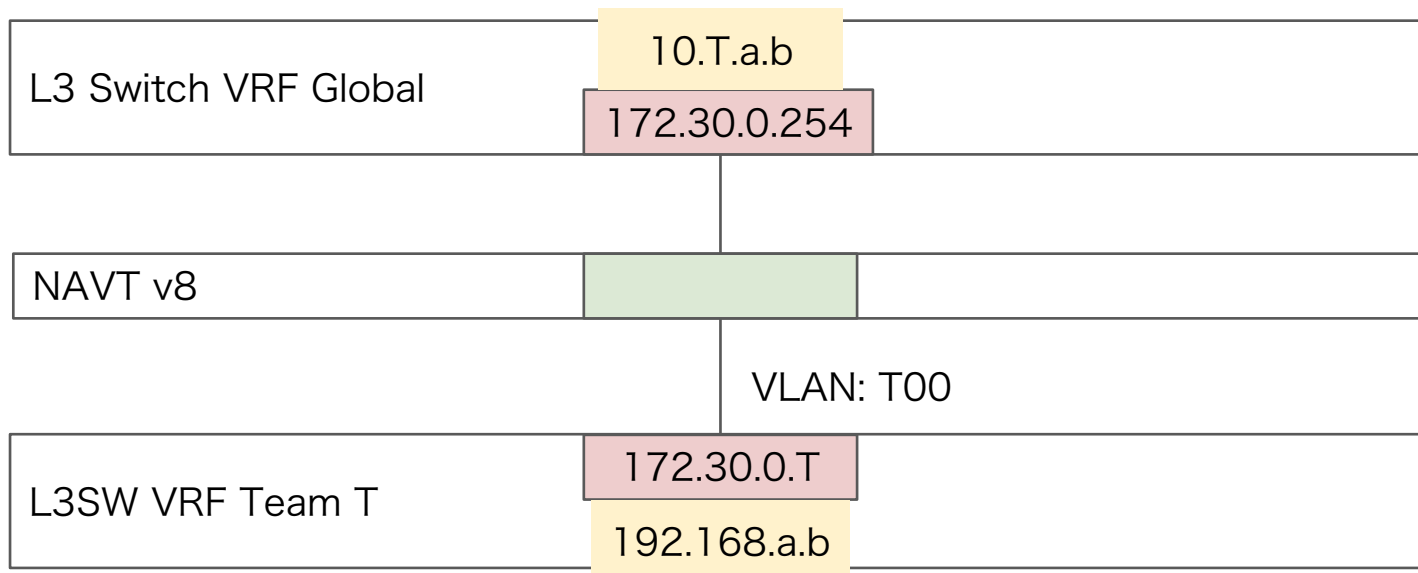
NAVT v7 (Network Address Vlan Translation)





NAVT v8 (Network Address Vlan Translation)

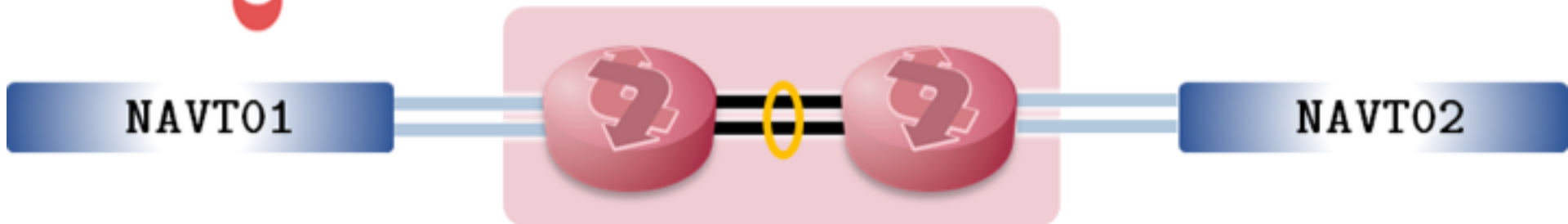
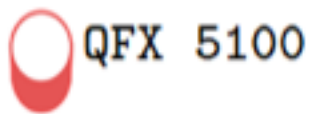
- 第8回は第7回の拡張
- ARP の動的解決に対応





NAVT v9 (Network Address Vlan Translation)

- 冗長化に挑戦
- コードは何も変更せず
- ステートレスなので QFX5100 にモジュールとして組込



ネットワークトポロジの紹介



とくになし



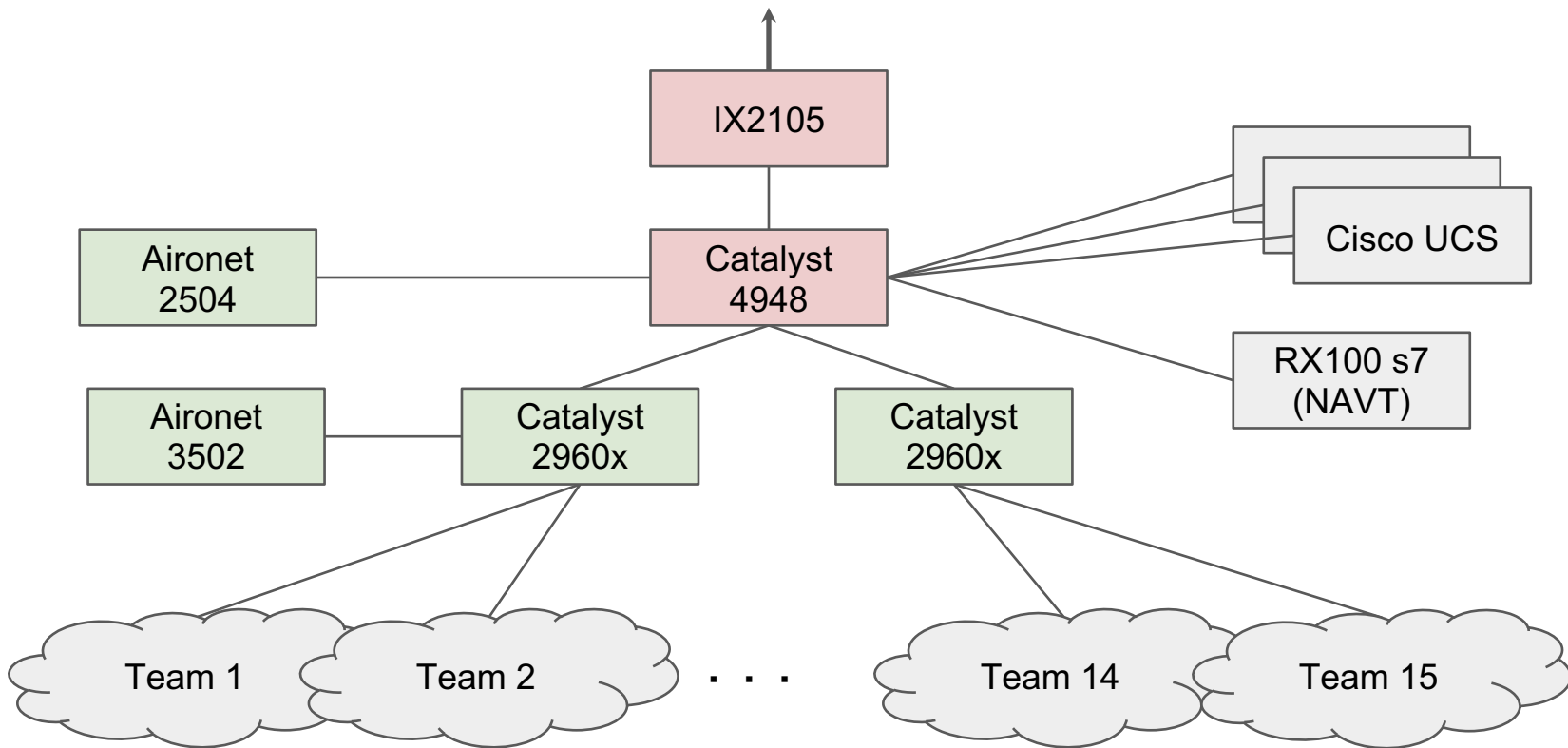
機材 @ ICTSC6 (敬称略)

- Cisco Aironet 2504 (Cisco)
- Cisco Aironet 3502 * 4 (Cisco)
- Cisco Catalyst 2960x * 4 (Cisco)
- Cisco Catalyst 3750x (Cisco)
- Cisco Catalyst 4948 (CyberAgent)
- Cisco UCS C220 M3 * 3 (CyberAgent)
- Fujitsu RX100 S7 * 4 (NHNテコラス)
- NEC IX2105 (私物)





シンプルなスター型トポロジ @ ICTSC6





安定感！！



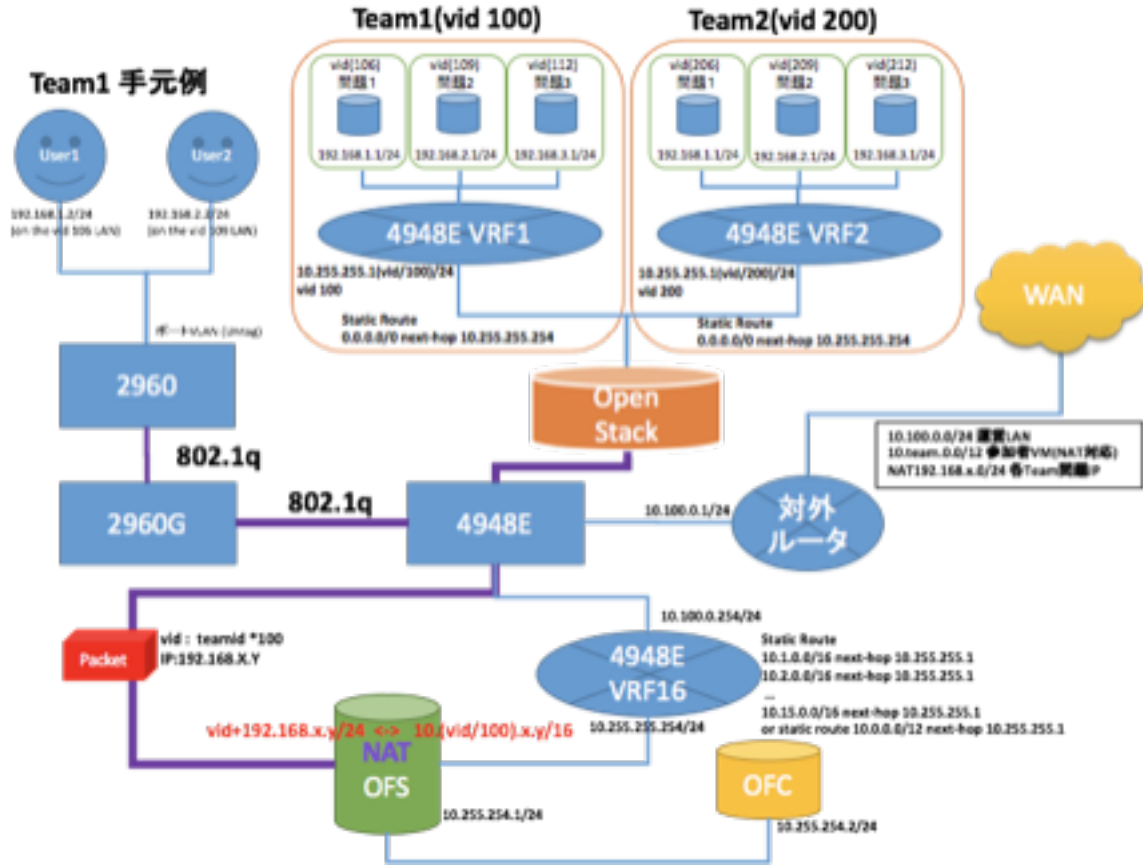
機材@ICTSC7 (敬称略)

- Centec V330 * 2 (Sakura Internet)
- Cisco Aironet 2504 (Cisco)
- Cisco Aironet 3502 * 4 (Cisco)
- Cisco Catalyst 2960x * 4 (Cisco)
- Cisco Catalyst 4948 (CyberAgent)
- Cisco UCS C220 M3 * 3
(CyberAgent)
- IBM x3530 m4 * 3 (CyberAgent)
- Fujitsu RX100 S7 * 4 (NHNテコラス)
- NEC IX2105 (私物)





VRF + NAVT の導入 @ ICTSC7





仮想化！！



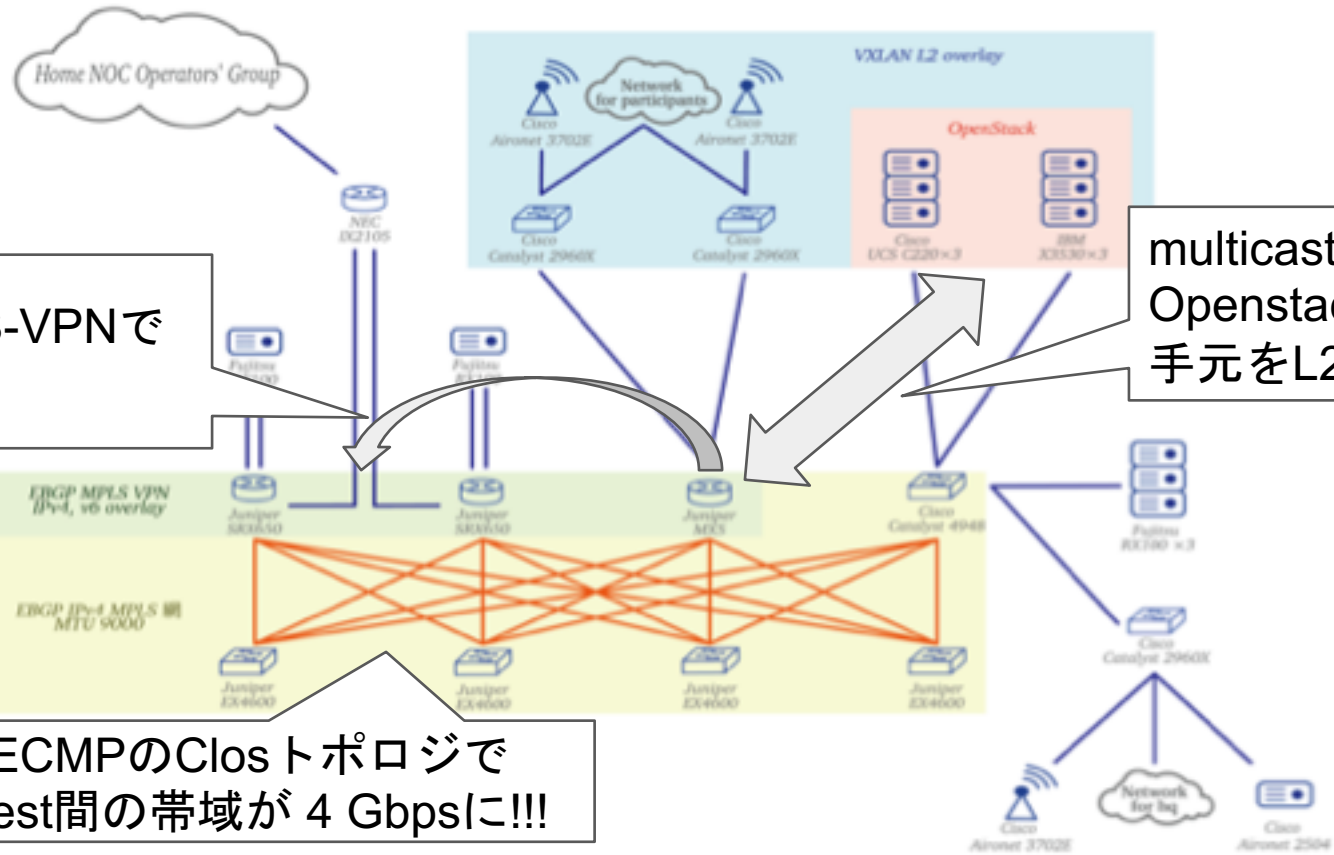
機材@ICTSC8 (敬称略)

- Centec V330 * 2 (Sakura Internet)
- Cisco Aironet 2504 (Cisco)
- Cisco Aironet 3702 * 4 (Cisco)
- Cisco Catalyst 2960x * 4 (Cisco)
- Cisco Catalyst 4948 (CyberAgent)
- Cisco UCS C220 M3 * 3 (CyberAgent)
- IBM x3530 m4 * 3 (CyberAgent)
- Fujitsu RX100 S7 * 4 (NHNテコラス)
- Juniper EX4600 * 4 (Juniper)
- Juniper MX5 (Juniper)
- Juniper SRX650 * 2 (Juniper)
- NEC IX2105 (私物)





Clos トポロジ + VXLAN + MPLS @ ICTSC8



BGP MPLS-VPNで
L3延伸

multicast VXLANで
OpenstackのVMと
手元をL2接続

eBGP+ECMPのClos トポロジで
East-West間の帯域が 4 Gbpsに!!!



冗長化！！



※ L3SWが突然の再起動 @ ICTSC8

- 睡眠環境を良くするため、家から作業できるように設定して帰宅
- 深夜作業中、突然サーバへの接続が断
- 最も家が近いメンバーが深夜2時にタクする事態に🙄
 - 1680円は闇夜に消えました
- 理由としてはToRのL3スイッチが突然再起動し、bootmanで止まっていた…
 - “boot [Enter]”の5キーで終了
- 勿論電車もないのでそのまま木のベンチで睡眠 🙄
- 掃除のおばちゃんに死んだか確認される事態に



機材 @ ICTSC9 (敬称略)

- ネットワーク機器
 - Juniper SRX 1500 * 2 (Juniper)
 - Juniper QFX 5100 * 2 (Juniper)
 - Cisco Nexus 5548 * 2 (CyberAgent)
 - Cisco Nexus 2248p * 2 (CyberAgent)
 - Cisco AIR-CT2054 * 2 (Cisco)
 - Cisco AIR-CAP3702E * 3 (Cisco)
 - NEC IX2105 * 1 (私物)
- サーバ
 - IBM x3540 m4 * 6 (CyberAgent)
 - Fujitsu PRIMERGY RX100 S7 * 4
- ドライブ
 - Fusion-io * 6 (CyberAgent)

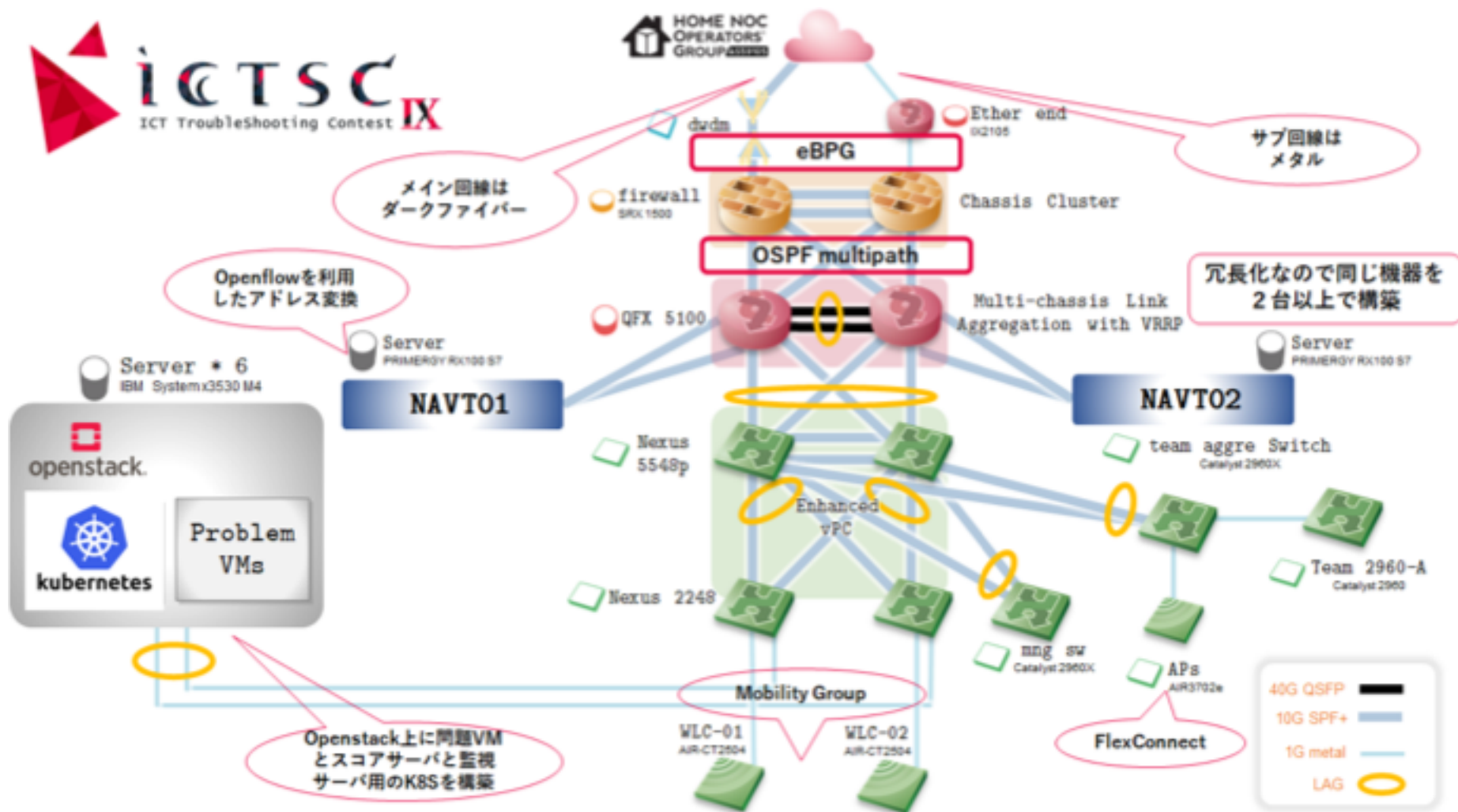




機材 @ ICTSC9 (敬称略)

- モジュール
 - SFP+モジュール 52個
 - マルチファイバー 26本
 - シングルファイバー 3本
 - QSFP+ 2本
 - FEXモジュール 8個

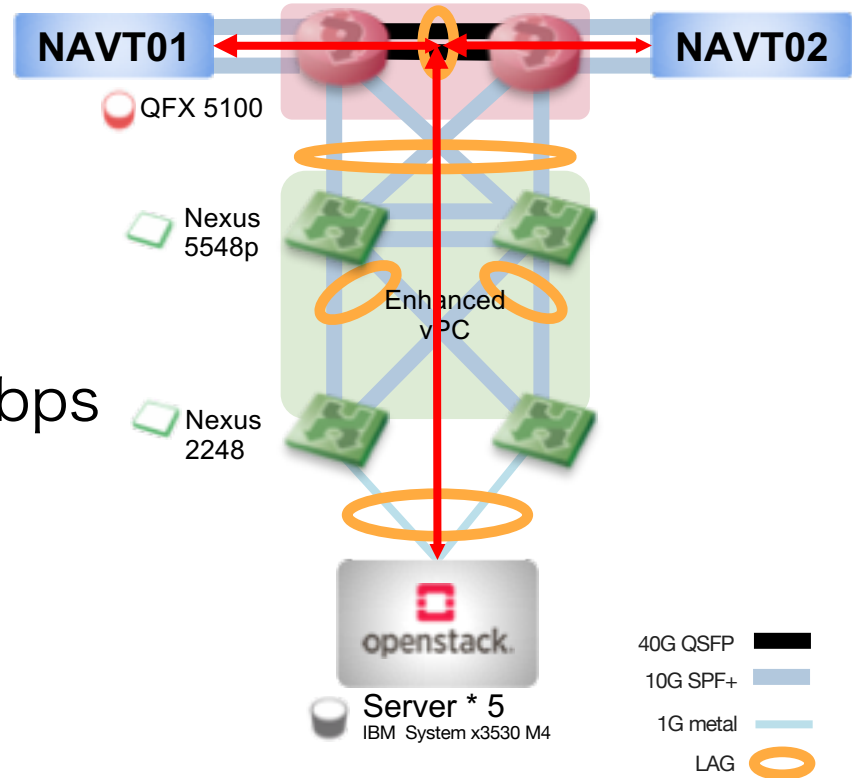






ベンチマーク @ ICTSC9

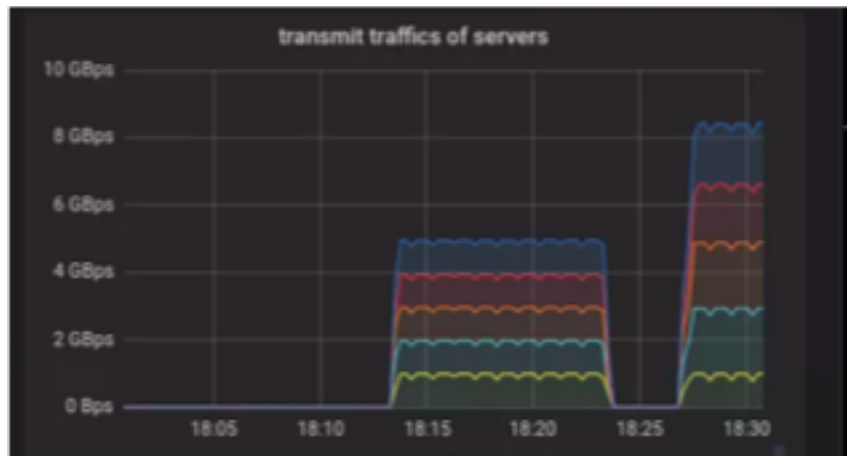
- iperf3を用いて赤矢印の経路上の帯域ベンチマーク
 - 10対のVM
 - 15対のVM
- 理論値の計算
 - $2 \text{ Gbps} * 5 \text{ VMホスト} = 10 \text{ Gbps}$





ベンチマーク結果

- 実測8Gbps出た
- 理論値が出なかったのはサーバとスイッチ間のLACPの偏り
- その他のネットワーク機器でボトルネックは発生しなかった
VMホスト



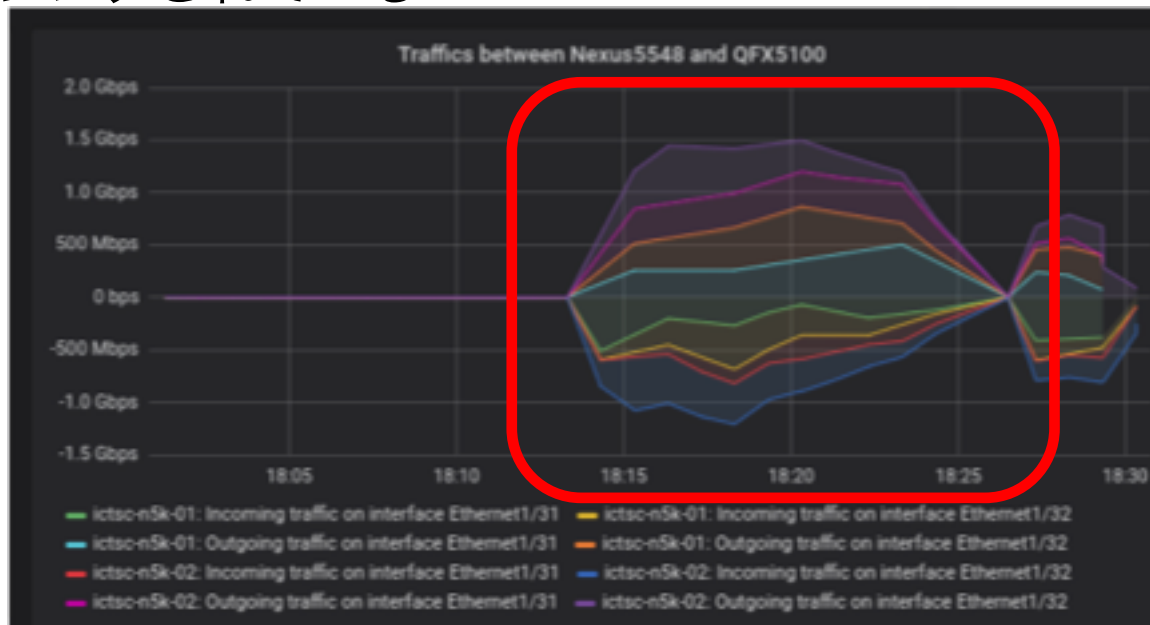
2台のNAVTサーバのトラフィック(上下2台)





Nexus 5548 - QFX 5100 のトラフィック

- 4インターフェイスの上りと下りトラフィックがそれぞれ最高1.5Gbytes/sec = 12Gbps 流れていることがわかる
- Juniper MC-LAG(上)とCisco 拡張vPC(下)を利用して等コストロードバランシングされている





Wi-Fiの苦悩

- Radius + ダイナミックVLAN → 1 個のSSID @ ICTSC7
 - 不安定だった
 - 他のチームの packets が見えていた
- 15個(チーム数分)のSSIDを作った @ICTSC8
 - 安定した
- またRadius使った@ ICTSC9
 - 2.4GHzに引っ張られる人多かった
 - 2日目は2.4GHzを弱めて対応



ICTSC 2018



ICTSC 2018 について

ICTSC 10 ではなく ICTSC 2018 として新たにスタートし、
1年を通して大会を行います！

参加者には1年にわたってチームで技術を磨いてもらいます

2018/08: 第一次予選 @ オンライン

2018/12: 第二次予選 @ オンライン

2019/03: 本戦 @ 東京



協賛もお待ちしております

- ICTSC では様々な形での協賛を募集しています！
 - スポンサー / 機材提供 / IP 提供…？ / ASN 提供…？
- 具体的には…
 - コンソールサーバ(15台ぐらい)
 - ケーブルマネージャー / UPS / PDU / ラックファシティ系
 - 睡眠の取りやすい宿 / 美味しいごはん / 潤沢な電源



まとめ

今後とも頑張ります

色々ご意見・ご協力お願いします！！

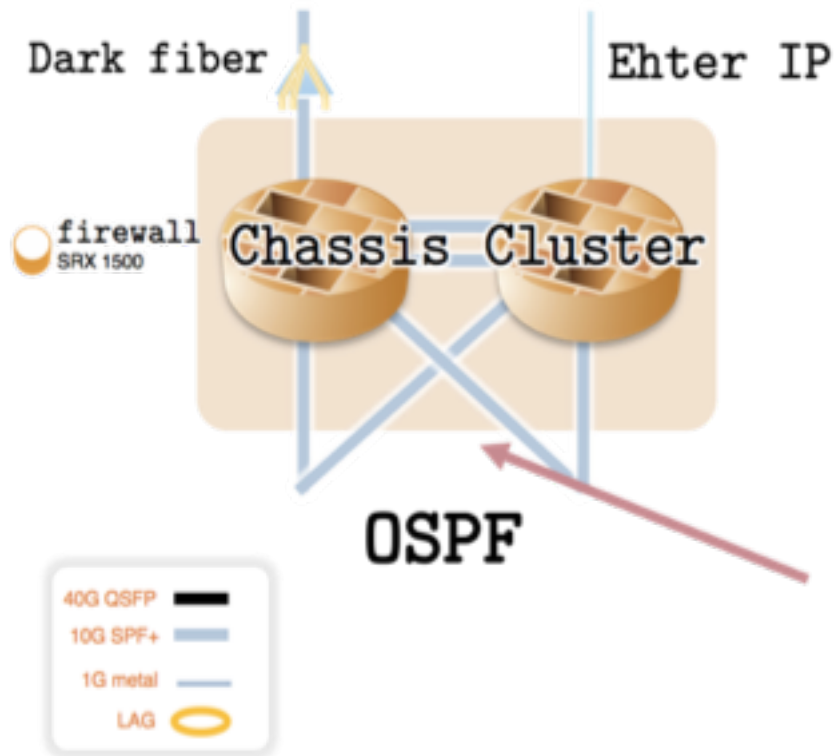


議論したいこと（再掲）

- Wi-Fiについて
 - 準備期間中のテストどうしてますか？大人数を再現できない！
 - 2.4GHzとうまく付き合えていますか？
- 短期間プロジェクトのチームビルディングって……
- ICTSC 独自技術 NAVT はイケてますか
- 新しい技術導入のノウハウ
 - 安定を取るか技術を取るか
- イケてる技術ないですか？
 - 例) データセンタアーキテクチャ(RFC5549など)



Juniper SRX1500



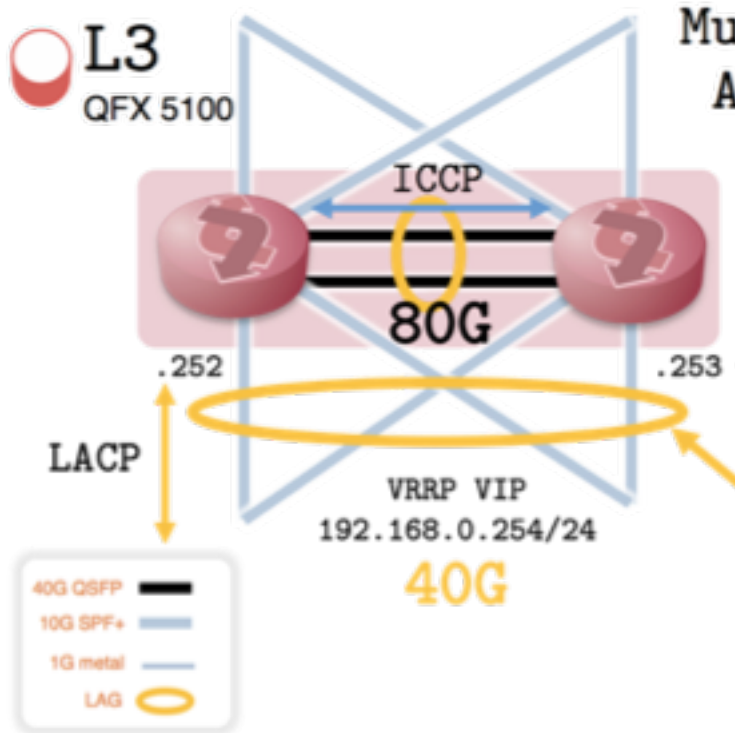
```
Cluster ID: 1
Node   Priority Status      Preempt Mar
-----
Redundancy group: 0 , Failover count: 1
node0  1         primary    no       no
node1  1         secondary  no       no
```

- 仮想化機能を活用した一つの論理デバイス
- Active/Active 動作
- 各デバイスで同じconfigを持つ

OSPFのマルチパスで負荷分散



Juniper QFX5100



Multi-chassis Link
Aggregation with
VRRP

- ・ 対向からみたL2では1台に見える
- ・ 対向からみたL3では2台に見える
- ・ Active/Active 動作
- ・ 別々のconfigになる

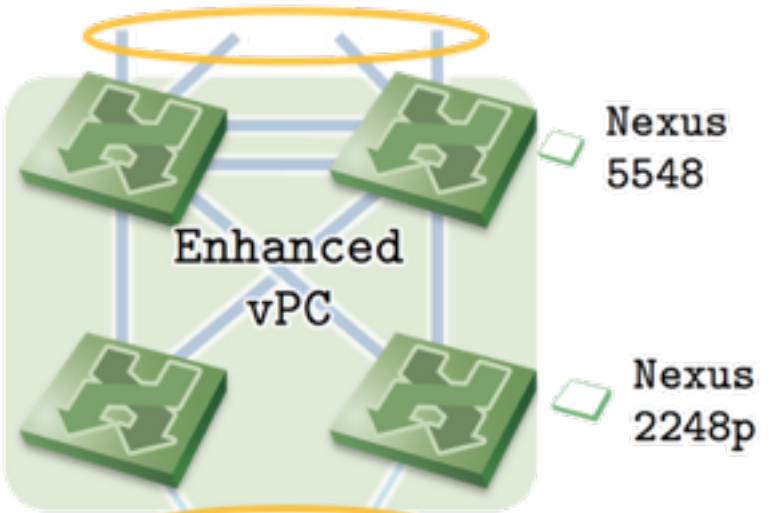
L3はVRRPで冗長

L2はMC-LAGで冗長



Cisco Nexus 5548 & 2248p

Enhanced vPC (virtualPortChannel)



40G QSFP	—
10G SFP+	—
1G metal	—
LAG	○

- Nexus 2248一つのラインカードとして使用しています
- Nexus 5548のvPCを使用してデュアルホーム接続
- またNexus 2248(FEX)がNexus 5548にデュアルホーム接続
- スイッチをまたいだport-channelを実装
- 2つの独立ノード