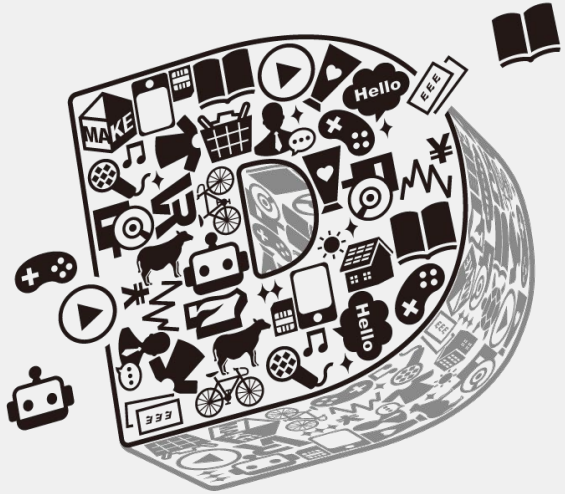


倉庫で構築するNWで苦労したこと

合同会社DMM.com インフラ部ネットワークグループ
佐々木 航平

Agenda

- 自己紹介
- DMM.comについて
- DMMオンクレについて
- DMMオンクレNWについて
  - 初期構築
    - 構成概要
      - 物理
      - 論理
    - DHCP
    - 構築について
    - 苦労ポイント
  - FW導入
    - 構成概要
    - 苦労ポイント
- 討論

自己紹介

名前(年齢): 佐々木 航平(28)
趣味:

- アメ車
- アガベ
- コーヒー
- 筋トレ
- 競馬
- プロ野球観戦

経歴:

- 2016/5 ~ 2017/6 ITベンチャー(SES事業)
- 2017/7 ~ 2020/3 DC事業者(SES事業)
- 2020/4 ~ 合同会社DMM.com(コンテンツ)

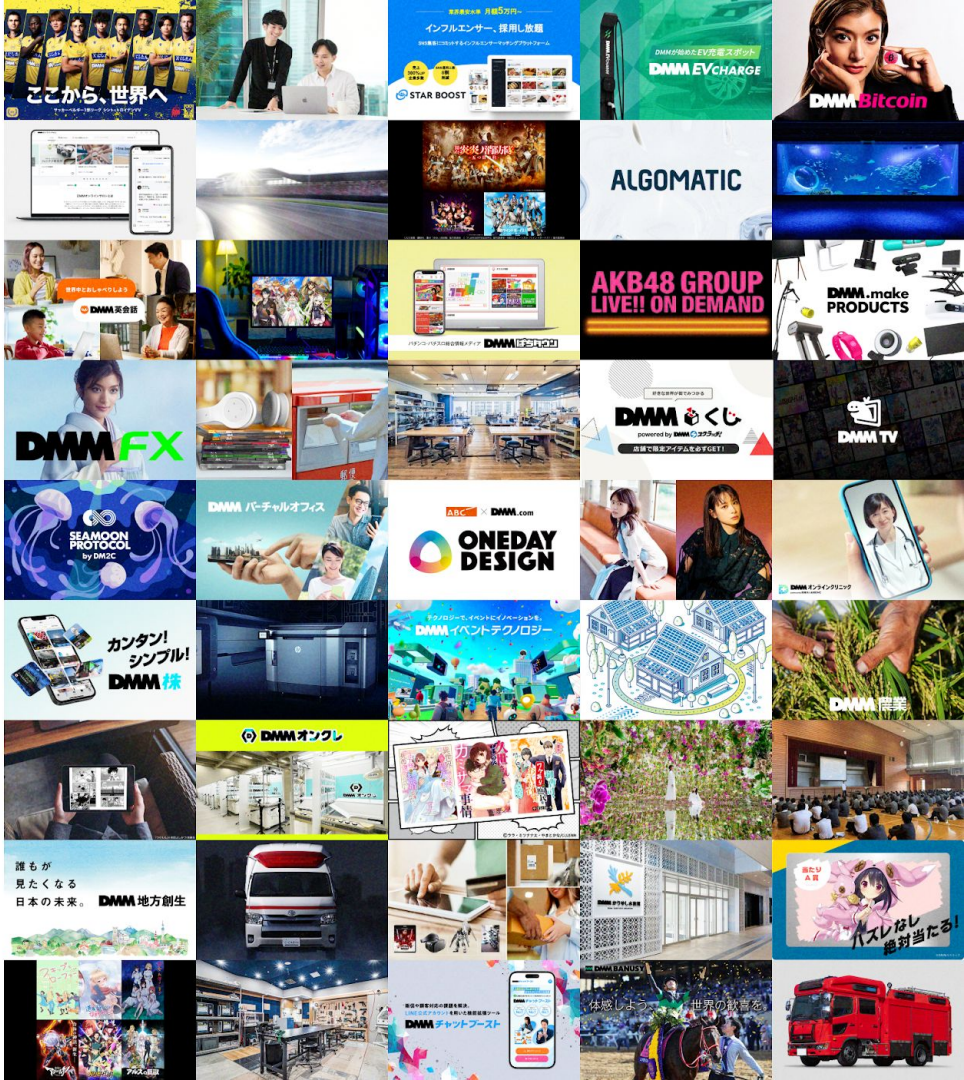
業務範囲:

- NW構築/運用(バックボーン~拠点)
- ツール?開発(Ansible/Python)

DMM.comについて

事業
BUSINESS

60以上の事業を
20以上のグループ会社で運営



主な事業

DMM TV

月額550円のDMMプレミアム会員に登録することでアニメ約5,400作品を中心に17万本以上※1の映像作品や、漫画、2.5次元、声優にフォーカスしたオリジナル番組など多彩なエンタメコンテンツを楽しめるサブスクリプション動画配信サービス。

DMMブックス

話題のコミック、雑誌、小説、写真集等の電子書籍など98万冊以上を、スマホやパソコンで読めるプラットフォーム。

DMM pictures

日本が世界に誇るコンテンツ「アニメ」の企画開発、ライセンスビジネスや製作委員会への参画。

DMM STAGE

2.5次元作品を中心とした舞台を制作。DMM picturesやDMM GAMESなど自社エンターテインメント領域のコンテンツを、グループシナジーを用いて舞台化。舞台以外にもLIVE、映像作品などを展開。

DMM GAMES

ユーザーに快適にゲームを楽しんでもらえるプラットフォームづくりや、ユニークなゲームの開発・パブリッシングを行っています。

DMMオンクレ

スマートフォンやパソコンを使って実物のクレーンゲーム機を遠隔操作し、24時間どこからでもクレーンゲームを楽しめるサービスです。

DMMスクラッチ

限定エンタメグッズやお得な雑貨・家電が当たるハズレなしのオンラインくじ。

DMMオンラインサロン

ビジネスから趣味まで1500以上の多彩なサロンがある日本最大級の「学べる・楽しめる」会員制コミュニティサービス。

DMM通販

DVD・Blu-ray、CD、本・コミック、ホビー、玩具、家電、日用品など、豊富な商品を展開。DMMによる仕入れ販売に加えて、法人や個人が出品できるサービスも提供し、コレクター商品など、希少性の高い商品も。

DMM宅配レンタル

DVD・Blu-ray／CD／コミックの宅配レンタルを送料無料で展開。

DMMイベントテクノロジー

Webブラウザ上で誰もが手軽に参加できる、テクノロジーを活用した次世代のイベントの提供。

DMMオンラインクリニック

「DMM.com」と医療社団法人DMCが提携したオンライン診療事業です。

※1 ： 2023年7月時点

主な事業

デジタルコミック事業

DMMグループが保有するコミックのコンテンツIPを創作している事業です。DMMブックスをメインプラットフォームに、国内や海外に向けて良質なデジタルコミックを発信しています。

DMMチャットブースト

LINE公式アカウントの機能を拡張させた誰でも・簡単に顧客対応やマーケティングを自動化できる「LINE公式アカウント機能拡張ツール」です。

STAR BOOST

InstagramのPR投稿が依頼し放題のインフルエンサーマーケティング。

DMMバーチャルオフィス

“スマホに Office”をコンセプトに、テックカンパニーが提供する利便性の高いツールと都内一等地のこだわりの築浅物件のオフィス住所を割安な価格でご提供。

DMM英会話

24時間365日、世界120ヶ国・1万名以上の講師とマンツーマンレッスンができるオンライン英会話サービス。

DMM FX

操作性抜群の取引ツールが好評のFX取引（店頭外国為替証拠金取引）サービス。

DMM 株

国内株式に加えて米国株式も取り扱う、お得な手数料が魅力の株式取引。

DMM Bitcoin

豊富な暗号資産種類・DMMグループで培われた高い技術力で、お客様が安心して暗号資産（仮想通貨）の取引ができる環境を提供。

DMMかりゆし水族館

最新の映像表現と空間演出を駆使した新しいカタチのエンターテインメント水族館。

チームラボプラネッツ TOKYO DMM

「4つの巨大な作品空間と、2つの庭園からなる「水に入るミュージアムと、花と一体化する庭園」。はだしとなって、超巨大な作品に、他者と共に、身体ごと、圧倒的に没入していく。

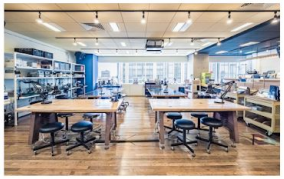
ベルリング

軽量化技術を活用した新しい概念の消防・救急車両開発。

ハッシャダイ ソーシャル

全国の高校などを中心に、350校以上でキャリア教育プログラムを提供。

主な事業



DMM.make AKIBA

つくり手が描く未来を共に実現するハードウェア開発環境と、技術やビジネス面でサポートするスタッフを備えたコワーキングスペース。

DMM.make 3Dプリント

国内最大規模の法人・個人向けオンライン3Dプリント出力代行サービスを中心とした3Dプリントトータルプラットフォームサービス。

DMM.make PRODUCTS

世界中の家電・日用品の中から、暮らしを心地よくする厳選アイテムをお届け。

Algomatic

大規模言語モデル等生成AI技術を活用したサービスを開発・提供し、時代を変える事業を生み出します。

DM2C Studio

web3事業の企画開発・運営。

DMMバヌーシー

競走用馬の成長を見守り、喜びを分かち合える感動共有型のファンドサービス。

DMMエナジー

太陽光発電を基盤に自然エネルギー事業を多角的に展開。

DMM EV CHARGE

全国の商業施設や宿泊施設、公共施設、マンション等へ、EV充電サービス（設置・運営）を提供しています。

DMM music

DMM.comとA-Sketchによる音楽レーベル。

DMMぱちタウン

全国47都道府県の情報を掲載。パチンコ・パチスロを楽しむための総合エンターテインメントアプリ。

DMM競輪

競輪投票を楽しむアミューズメントインフラ。

ABC × DMM.com
ONEDAY DESIGN

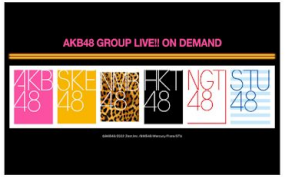
ONE DAY DESIGN

朝日放送グループHDの持つテレビの発信力と、DMMの持つデジタルの展開力や60以上の事業シナジーを組み合わせ、新しい価値提供を行うジョイントベンチャー。

主な事業

DMM WEBCAMP

キャリア開発型プログラミングスクール。2014年よりサービスをスタートし、8,000名以上のIT人材を輩出。



DMM AKB48 グループ

AKB48グループの劇場公演がみられるLIVE&オンデマンド配信サービス。

シント=トロイデンVV

世界最高峰と言われる欧州サッカー若手選手の育成リーグとしても定評のあるベルギー「ジュピラー・プロ・リーグ」1部所属クラブの経営権を取得。

DMM 地方創生

DMMのリソースを組み合わせ、ワンストップで事業を展開するプロフェッショナル集団。誰もが見たくなる日本の未来を目指して。

DMM 農業

「新しい価値を作り、農業の未来に貢献する」をミッションに、ジビエ流通・農業業界のDX推進を展開。

DMMオンクレについて

DMMオンクレとは？

DMMオンクレ(以後オンクレ)はPCやスマートフォンを使って実物のクレーンゲーム機を遠隔操作し、ユーザーが24時間どこからでもクレーンゲームを楽しめるサービスです。

埼玉県加須市の倉庫でクレーンがみなさまの Playを待機中です！
その数なんと約500台！
ぜひお試し下さい！

オンクレのNWについて

NW構築の全体像

話自体は新サービスが立ち上がるとは聞いていたがそのサービスで使う NWを
インフラ部で構築するとは事前に聞いていなかった
その中で実際に拠点構築タスクとして動き出してからのスケジュールは以下のイメージ

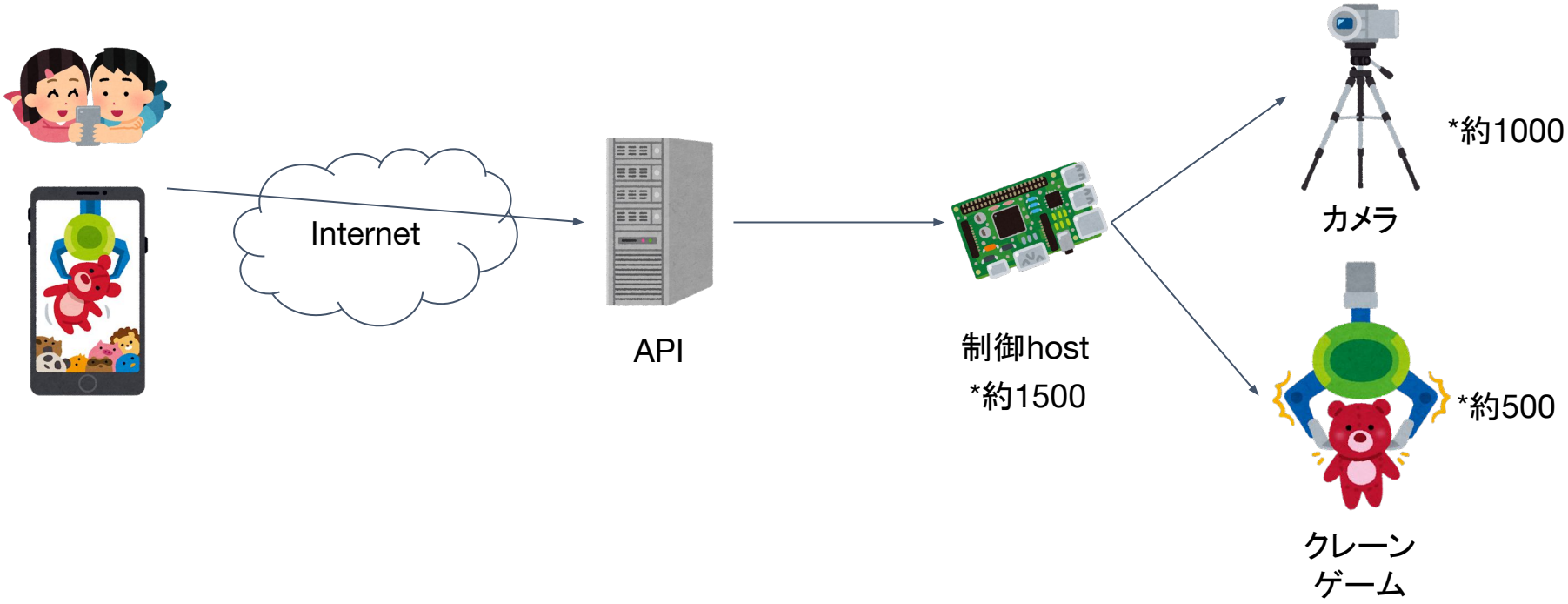
スケジュール

2021/9月 NW構築Kick off & 構成設計(2ベンダー4構成)
2021/9~11月 機器/構成検証
2021/12~2022/1月 DHCP/ZTP検証
2022/2月 現地構築
2022/3月 引き渡し

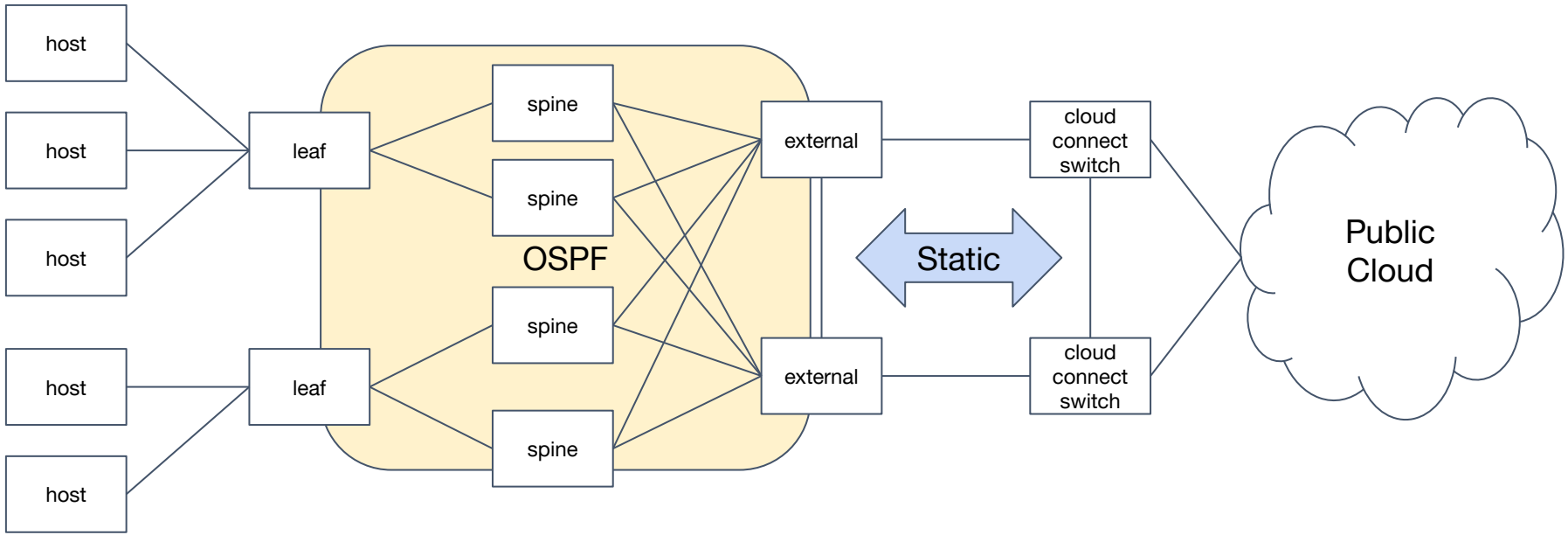
メンバー構成

PM / PL(私)※ / PMO / member *2
※Playing Leader(?)として設計/検証/構築まで主担当として実施

サービス全体像



初期構成概略図


host
host
host
host
host
leaf
leaf
spine
spine
OSPF
spine
spine
external
external
Static
cloud
connect
switch
cloud
connect
switch
Public
Cloud

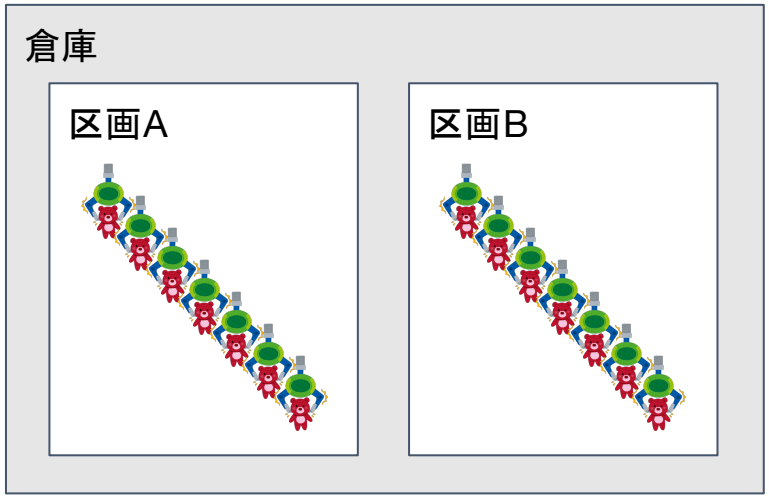
初期構成(物理)①

- 収容ノード数は1500程度、すべてStandalone
- 倉庫を2区画レンタルしてサービス環境として利用、倉庫間は高所配線が必要
- クレーンゲーム筐体が置かれるので UTPケーブルをあまり長く伸ばせない
  - ラックから伸ばそうにも1区画あたり750本のUTPケーブルがラックに
- すべてNorth-SouthでPublicCloud経由での通信

など諸々の諸条件があったため、spine-leaf構成
それぞれ、

- external(バックボーン接続)
- spine(倉庫集約)
- leaf(host収容)

加須からはバックボーンに専用線を引き、
PublicCloudまでは既存のCloudConnect回線を使用



初期構成(物理)②

external

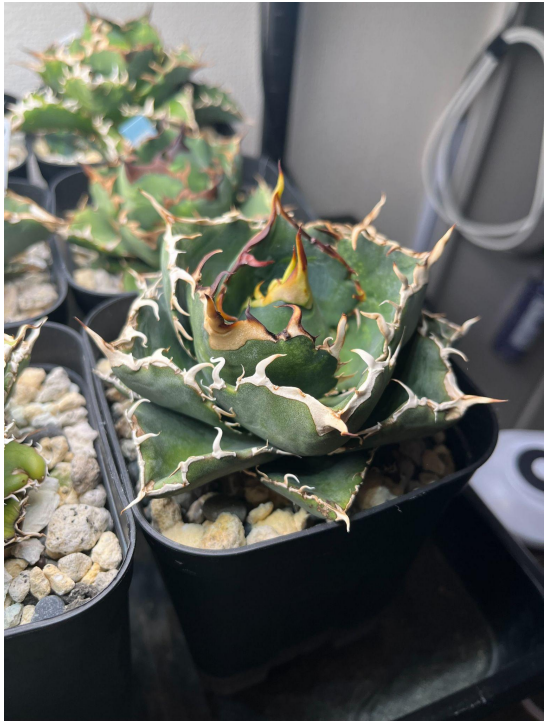
- 弊社バックボーンに接続する回線と spineを収容
- spineとの接続で倉庫間を接続
- 拠点間がフルメッシュではないのでここだけ渡りあり
- これは46Uラックを立てて収容

spine

- 倉庫単位でleafを集約
- こちらもラックに収容

leaf

- クレーン筐体の合間に簡易ラックを立てて設置
- UTPを集約、Uplinkは光に
- 収容ノードがstandaloneなので単体扱い



こちらは我が家の spine-leaf(?)

初期構成(論理)①

物理構成でも記載の通りspine-leaf構成
ただし、拠点間部分のみ回線費用の都合上 spine-leafではなく四角形での接続
また、拠点内のRoutingはOSPF、拠点間はStaticRouteでの接続

extermal

- 拠点間の接続と倉庫の Trafficの集約を担当
- 拠点間はバックボーン全体への経路広報をしたくなかったので Staticで接続
  - DC以外の拠点をOSPFで繋げない慣例だったのも理由
- Cloud(専用線接続)向けのStaticを再配送してleafに経路を持たす

初期構成(論理)②

spine

- 各倉庫ごとのleafの集約をしてNorthSouthのPacket転送を担当
- ただOSPFでRoutingしてるだけ

leaf

- 収容ホストが冗長化を考慮されていないため leafもstandalone
  - そのためFHRPはなし
  - 同じ設定のクレーンが大量にあることでサービスとしての冗長性を担保
- leaf単位で単一NWを保持
- DHCPリレーを行う
  - PortBasedDHCP(Interface単位でIPを付与する)を実現するためにDHCPサーバにoptionをつけて転送
- OSPFでhost収容segmentを広報

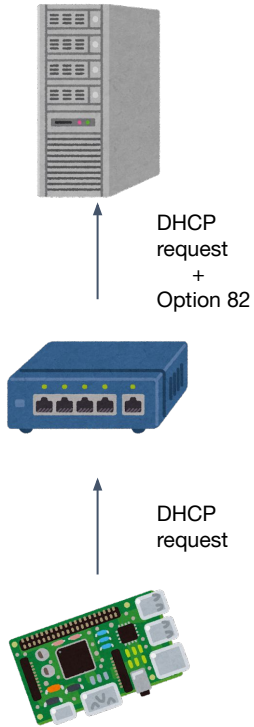
DHCPの設計

- 現地のクレーンは約500台
- クレーン1台あたりのhost台数は3
  - 現地で約1500host設置
- 固定IPなどを手作業で運用していくのはなかなか大変

などの意見があったのでDHCPの検討も合わせて実施
購入を検討しているSW自体でアドレスの払い出しはできないため、

- DHCP option82でrelay時に情報を追加してDHCPサーバに転送
- DHCPサーバ側でoptionの内容+segment情報でアドレスを固定払い出し

となるように全体を設計し、設定 /動作検証や構築
これにより、クレーンで使用している host故障時も入れ替えるだけで同じ IPを利用でき、
現地のオペレーションの簡略化を達成



DHCPの問題点

前スライドの通り動作自体は問題ないが、host故障時にアドレスのleases情報は残る為、
新規で払い出せないという事象に遭遇
都度DHCPサーバ側でleases情報の手動変更が必要
→MAC address baseでのアドレス払い出しと手間が変わらなくなる

その中で、そもそもどのhostに払い出したかはSWのInterfaceに紐づく→管理不要では？
と発想の転換で、leases情報を破棄する運用に変更

これにより、故障交換時も問題なく同じアドレスが利用可能に

初期構築/ZTPについて

この拠点は加須市内の駅から離れた場所にあり、
構築のために通うと移動時間だけで往復 4時間以上かかる
→ZTPで構築期間を短くして現地の回数を減らすことを検討

当初はDHCPでKitting用IPとssh用の設定のみ配布後、Ansibleで設定投入を想定
しかし、購入した機器の上位モデルは Ansibleのmoduleが存在するが、
購入モデルでは利用できないと判明

そのため、以下の方式で実装

- AnsibleのTemplates moduleでconfigを一括で作成し、ZTP用中間fileをあわせて更新
- 作成したconfigをZTPで配布する

→1台あたりの構築時間を 10分程度に短縮、同時に複数台構築で 3時間程度で機器設定が完了

初期構築時の苦労ポイント①

- スケジュール感
  - スタートアップ企業のスケジュールだった分、機器調達など考えるとなかなかタイト
- 半導体不足ともろに重なったために SWの納期が軒並み長い
  - サービスインは決まっていたので SWの選定が納期ベースに
- 遠方過ぎて移動だけで4~5時間取られる
  - 現地に行くと他のタスクは触れられない状態に
- 段々要件が増える
  - 全体的に手探りで始まった分仕方ない
- 回線の選択肢が少ない
  - 倉庫団地ということで10Gbps通せる回線がなかなか見つからない

初期構築時の苦労ポイント②

- ZTP増設hostはhost_varsとplaybook実行で設定が済むが、
  既存hostへの変更がAnsibleではできない
  - 主にleaf追加時のspineが該当
    物理的に拡張の余地がほぼなく増設もあまり見込めないので手作業で実施

- ZTPの挙動の誤認
  - 中間ファイルでOSを指定するのでconfig内のOSの記述を省略したところ、
    一番古いOSでの記述と筐体が理解したうえで自動で現 OSの形へと再解釈
    結果SSHがデフォルトで無効になりRemote接続ができない事になった
    - 検証期間にこの辺までは確認できず …

- アプリ開発チーム側との意思疎通
  - “ポート”などの言葉1つでもSWのNICやTCP Portなどで認識齟齬が
  - 極力同音異義語にならないようにすり合わせ

初期構築完了後

予定通り2022年3月に環境引き渡しをしてクレーンとクラウドの接続なども確認
アプリの試験も進み、2022年4月に一度サービス開始したが
その数日後、サーバ負荷が問題となり一旦サービスクローズして
アプリ開発チームにて遅延の原因になりそうな箇所を再開発

2022年6月にサーバ調整した状態で再度リリース
サービス提供は可能だが、操作遅延が最優先課題として残る形に

再リリース後にNWというよりインフラで問題がちらほら …

サービス再開後の苦労ポイント

- クレーン筐体とブレーカーが共通なことで、クレーン筐体の漏電に巻き込まれて電源断
  - SWは電プチOKなので無事だが、サーバが …
  - 構築時にUPSの相談はしたが様子見 (とりあえずなし)だった

- 北関東特有(?)の雷で電源断
  - 倉庫の電源系統は1系統のため、雷でその系統が影響を受けると全断
  - 漏電+雷のWパンチを受けるのでUPSが必要と判断して導入
    - ただし、ラック内の機器のみで leafやクレーン、hostは…

- 空調のない環境でのSW動作の不安
  - 倉庫でエアコンがないので機器の負荷状況によっては動作温度を超えないかと
    - サーバの温度は38℃程度の日も…

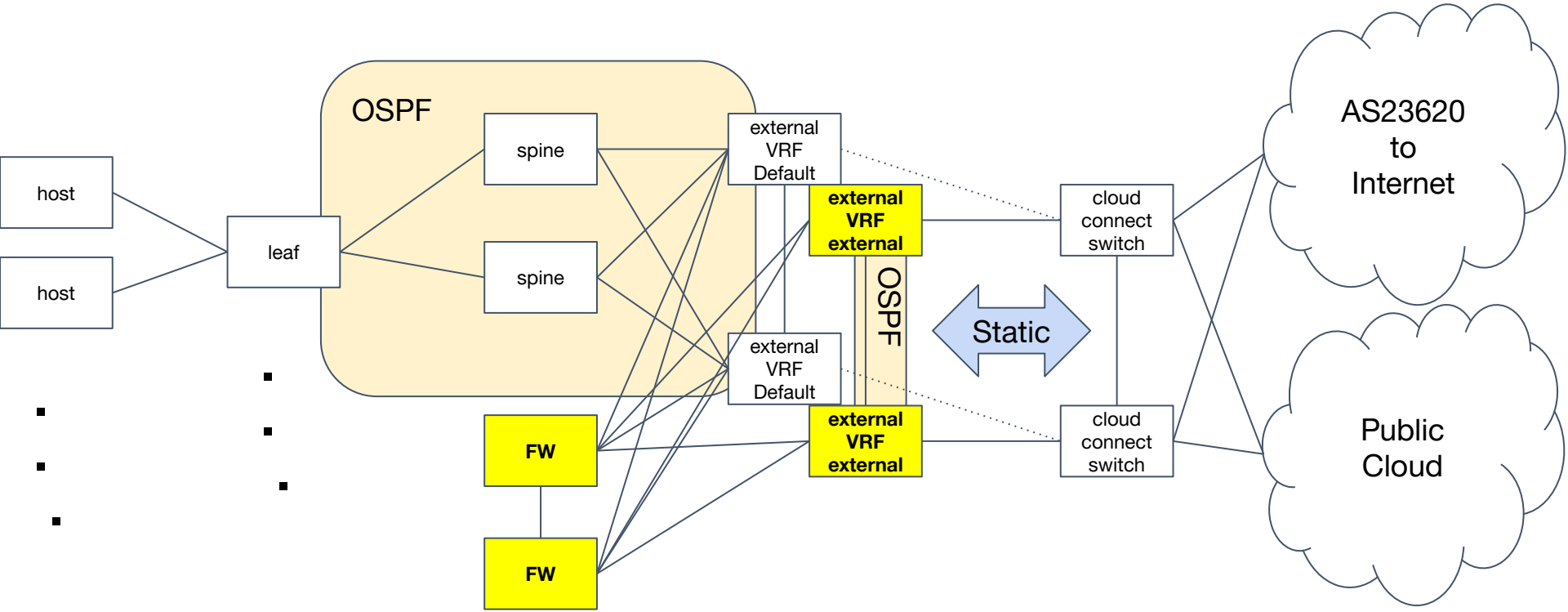
FW導入へ

サービスとして稼働がする中で、さらなる遅延改善に向けて映像配信方式の変更の話が浮上
CloudのProxyから配信していた構成からWebRTCのP2Pで直接配信する方式へ
これに伴いInternet向け通信を想定していない環境から
Internetへの通信を行えるようにFWを導入する流れに

- 一部Cloud経由の通信が残り、これ以上の遅延の悪化は避けたい
- 当初Internet経由の通信が必要になることはないと聞いていた為、
  構成上FWを入れる箇所を検討していない
- スケジュール感がクラウド、設計や検証を含めて 2ヶ月後にFWを導入

のような形でFW導入プロジェクトが始動
今回もPL(?)として設計~導入まで担当

FW導入構成概略図



FW導入構成について

FWを導入するにあたり、物理面については構成上 externalに収容で決定
ただし、HAが起きると500msec程度ダウンタイムが出るだろうということで接続だけ特殊に

- FW-external間はたすき掛けで10Gbpsを2本ずつ接続
  - LAGで束ねるがmin-links1で1本断を許容
- externalをVRFで分割
- FW-external間はexternal2台、それぞれのVRF2つの計4segmentで接続
- FWでECMPを有効化してexternal向けのtrafficを分散
- Internet向けはVRFでインラインに見える構成
- Cloud向けはFWを経由しないようにRoutingを制御
  - RouteLeakのイメージ
- spine以下は変更なし

FW導入時の苦労①

- 2ヶ月しかない
  - 納期は改善されていたが、相変わらず設計と検証の時間がほぼ取れない
  - 3月末スタート→4月半ばには購入→5月半ばに導入のスケジュール感
    - 結果的に在庫から転用したので購入自体は後ろ倒し

- FWのHAすら減らす構成という正直やったことない構成での検証
  - 結果的にはpingでの計測で200msec断に収まる

- 4月に検証したので新卒がたまたま社内ツアーで遊びに来て
  無垢な顔で「クラウド中心の時代にオンプレもつ意味ってなんですか？」と聞かれる
  - クラウドも中にはこういう機器があるし、オンクレみたいにそもそも
    クラウドだけじゃ用意できないサービスもあるんだよ、と優しく (?)伝える

- 現地が遠いなか導入メンテが AM7時開始
  - ホテルが倉庫付近になく、2駅先のホテルから早朝出発することに …

FW導入時の苦労②

- 初期構築でも記載の通り、Internet通信をそもそも考慮していない環境からのアクセス
  - externalの先にFWを単純に置くとCloud通信のhopが増えて遅延が増えるジレンマ
    - VRFって技術があって本当に良かった
- FWのZTPは検証含めて時間がないので現地に行く必要が出た
  - 在庫転用のお陰で大手町で構築してから配送したので現地オペレーション自体はほぼなし
- 初期構築でも記載の通り、既存 hostへの設定変更が手作業の必要
  - 更にZTPのJinja2をFW用の設定にも対応できるように書き換えが発生
    - まあ、当分使わないものですけど …(機器故障時くらい)
- WebRTCの知識の無さ
  - 通信要件ではCloudのWebRTCサービスへの通信許可のみと聞いていたが、
    うまく動作せずトラシューに時間が取られた

NWの設定時にどこまで気にするか

WebRTCの通信、という考え方ではなくL4の通信許可でただポート単位の通信許可を実施したところ
P2P接続ができない問題に遭遇

- WebRTCの動作をアプリ開発チームと調査したところ ICE Candidateに従ってTURN経由に
  - WebRTC(ICE Candidate)の流れをお勉強
  - VoIPと同じだと気がつく
    - STUN経由の通信を許可する設定を追加して P2Pで通信可能に

L3/4あたりまで考えて設定していたため WebRTCの動作を理解していなかったが
構築したNWを流れていく通信についての知識がないと設定自体ができないことを実感

IP:Port番号のACLでもL7の知識が必要な時代に…？

まとめ

オンクレを通しての苦労と伝えたいこと

- クラウドの速度感でインフラ 1から準備するのは大変
  - VPCと同じスピードでNWは作れない
- 通信要件は構築後に増える
  - 倉庫に限らずあるある
- UPSはほぼ必須
  - 電プチOKな機器しかないなら要らないかも …
- 空調は偉大
- ZTPは構築台数が多いほど恩恵を受けられる
- WebRTCはVoIP
  - 挙動が近いものだと雑に解釈すると飲み込みやすい

討論したいこと

- NWを普段作らないような箇所での構築時の検討事項やハマったエピソードなどあれば
  - 今後のためにもぜひ
- 遠隔地のZTPやリモートハンズなどについて
  - DCであればサービスとしてありますが、それ以外の場所の場合はどうするか
- AnsibleでOSモジュールがないときはどうしますか？
  - コマンド流すだけならPythonであればparamikoで書けばいいって話ではある
  - 他になにか方法などあれば
- 作ったNWを流れるProtocolの挙動って気にしますか？
  - L4で頭が止まってハマった経験などあれば
- NW構築/運用の自動化、どのように進めていますか？
  - 若干登壇内容とずれますが