



# SONiC ZTPでデータセンターネットワークを作った話



アーキテクト  
畔柳 駿一

KDDI xG Networks



HaaS Management NW PM

浅野 慶太  
KDDI



NWエンジニア  
伊藤 暎二

KDDI xG Networks



NWエンジニア  
遠藤 雅也

KDDI xG Networks



NWエンジニア  
片岡 祐介

KDDI xG Networks

KDDIグループの社内副業制度を用いて本案件に参加（兼務出向）

Private HaaSとHaaS Management NWの紹介

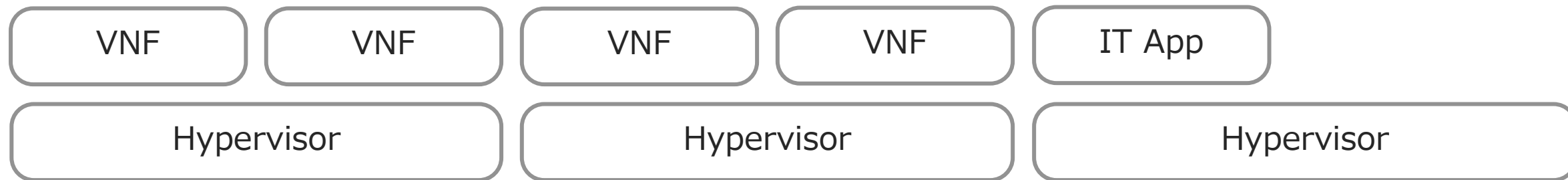
SONiCを用いたZTPの紹介

HaaS Management NWの開発/検証の進め方

検証で見つけたSONiCの課題の紹介（主にIPv6周り）

まとめ・皆さんと議論したい事

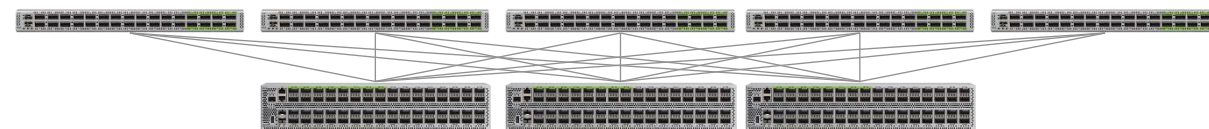
## ■ 仮想化基盤(Tenant)に対し、全国でPrivate HaaSを構築、提供



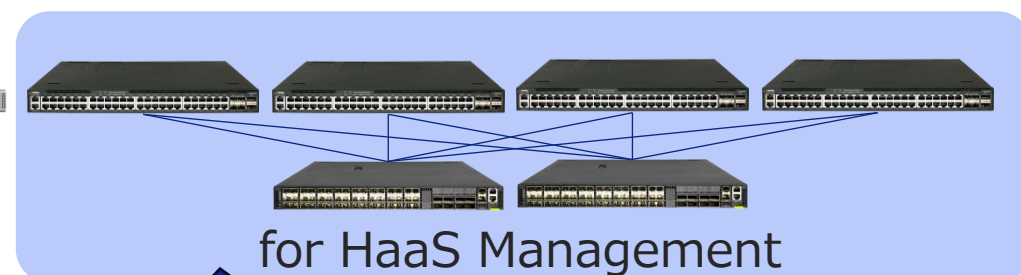
6300+ Servers



Data Center Network



for Tenant Traffic



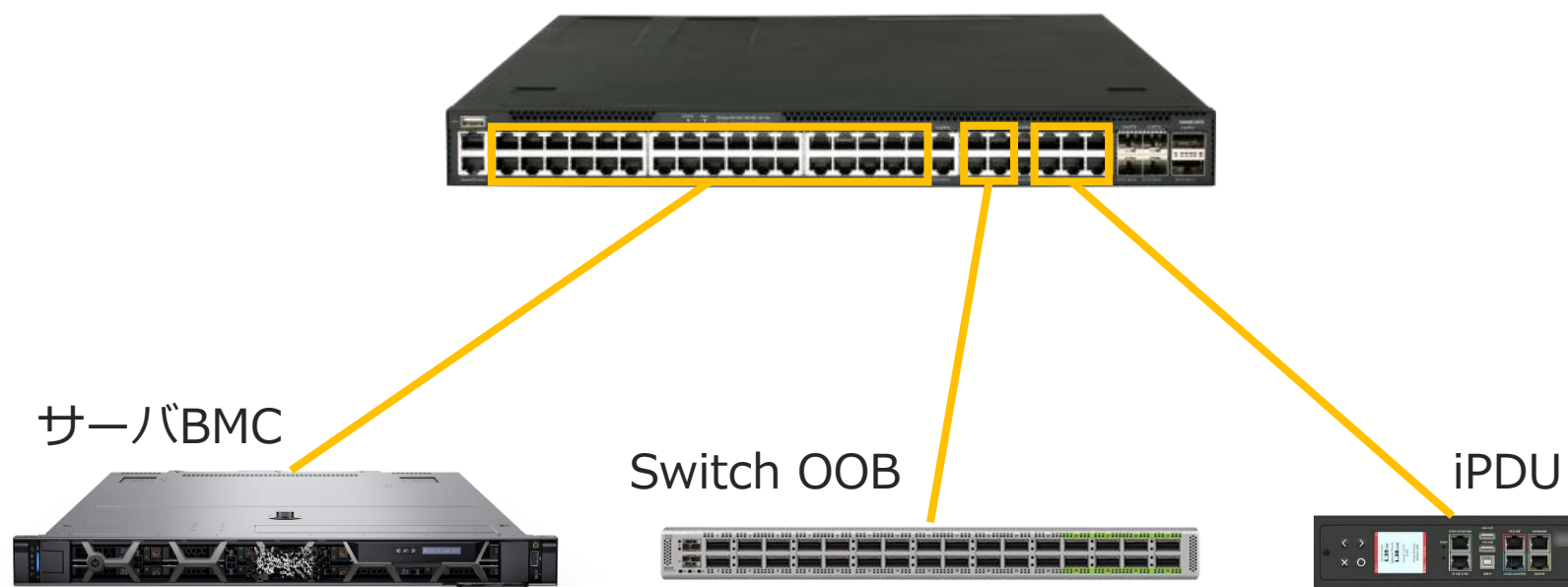
for HaaS Management

Private HaaS

このNWが本日のテーマ

## ■ HaaSのHWの構築、プロビ、保守運用を行うためのNW

- サーバBMC、ストレージやSwitchのOOB、Intelligent PDUの管理ポート、HaaSの管理用コンポーネントなどを接続し、構築や保守運用に利用
- 各ラックに1台設置。サーバ等の機器数量や搭載位置は全ラック共通となっており、Switchの設定パターンは数個。構築後はほとんど設定変更がないことが特徴

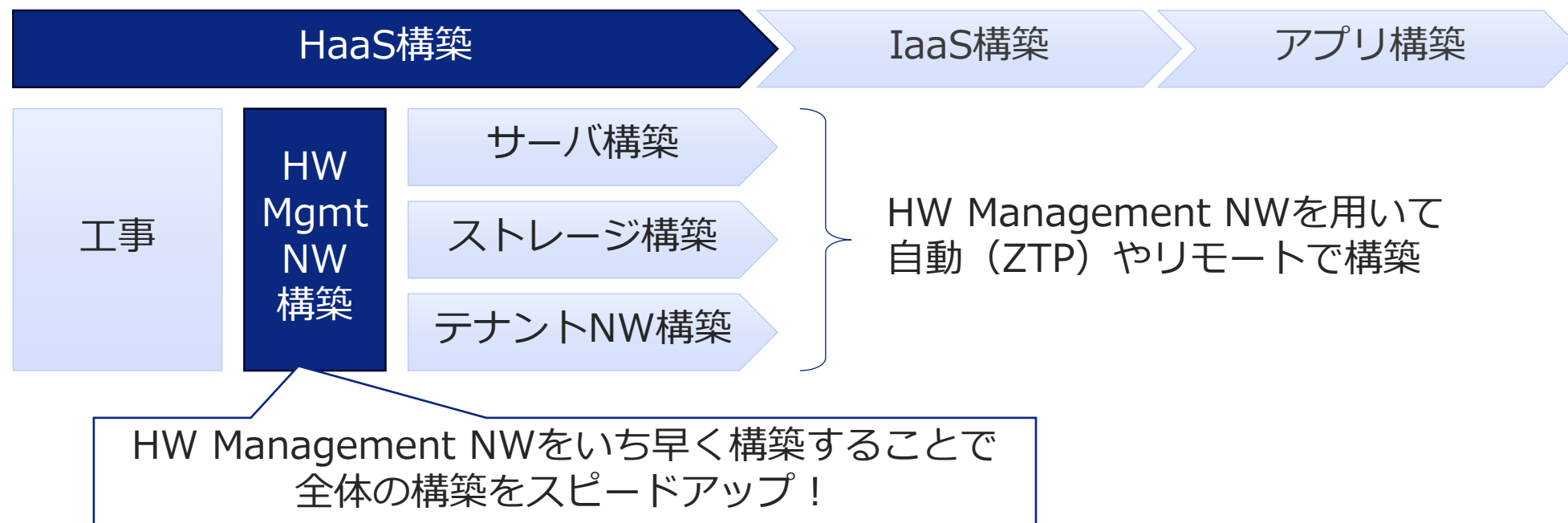


## ■実現したかった3つのこと

- ① Private HaaSを早く利用者に提供したい
- ② 保守作業の品質を維持しつつ効率を高めたい
- ③ NWエンジニアとしてスキルアップできるNWにしたい

## ① Private HaaSを早く利用者に提供したい

- HaaS構築の起点となるNWであり、いち早く構築することが重要



## ②保守作業の品質を維持しつつ効率を高めたい

- HaaSのSwitchは現在1600+、毎年+300程度。定常的に交換作業が発生
- 作業品質を維持しつつ、全国の保守担当がより簡単に作業を行える仕組みが必要に





## ③NWエンジニアとしてスキルアップできるNWにしたい

- せっかく新しいNWを作るので、みんなでスキルアップできる経験にしたい
- いままでやったことないコンセプトのNWを作りたい

みんなで勉強し、スキルアップできるNW



## ■ なぜWBS + SONiC？

### ✓ 構築自動化の仕組みがGood！

- ONIEとSONiCのZero Touch Provisioning (ZTP) 機能の活用により、構築のスピードアップと、保守での作業簡易化が可能に

### ✓ 新しい技術へのチャレンジ！

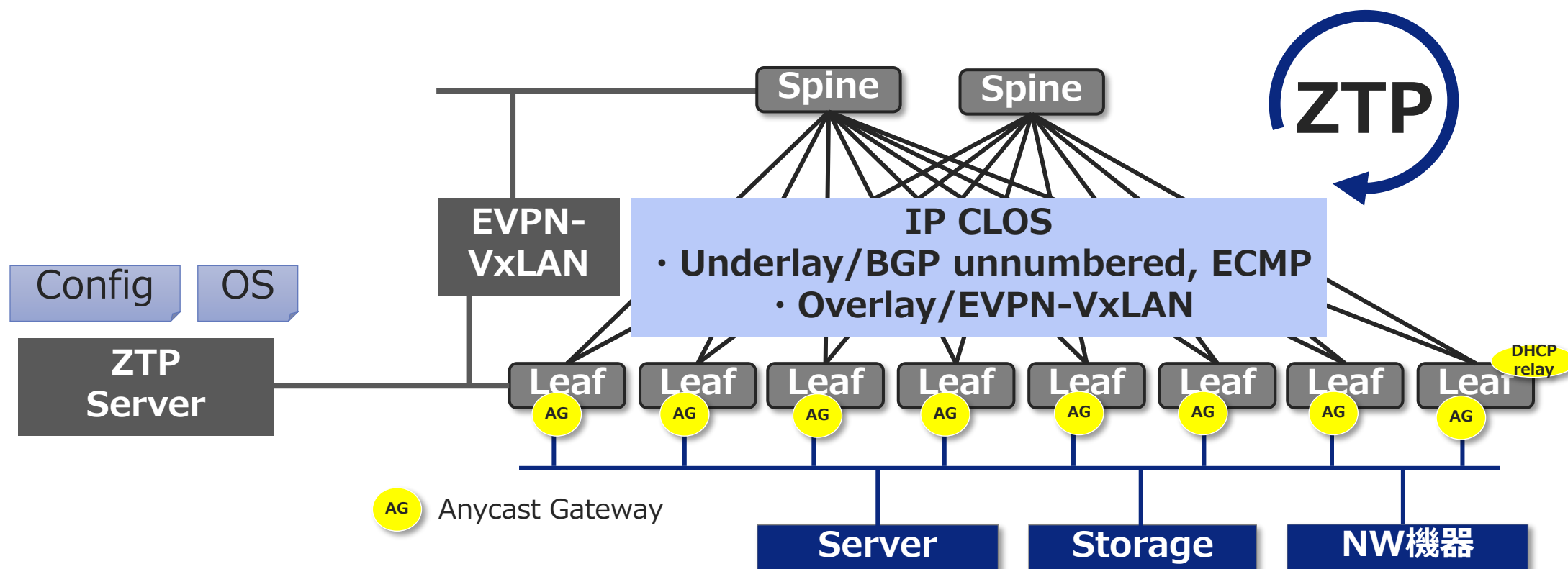
- HyperscalerやOTT各社で導入が進むSONiCの商用導入を通して、今後のNWエンジニアに求められるスキルやナレッジを習得したい

### ✓ 納期がGood！

- EdgecoreのWBS納期が希望のスケジュールにミート
- SONiCを活用していくことで、将来的にWBSのマルチベンダー化による納期問題の解決の足掛かりに

## ■ HaaS Management NWのZTP

- 機器電源ONにより、OSインストール、Underlay/ Overlay設定、構築後の疎通試験とログの保存が人の手を介さず、自動で実施される





## ■ 起動からの流れ

- SONiC起動まではONIE ZTPによるインストール

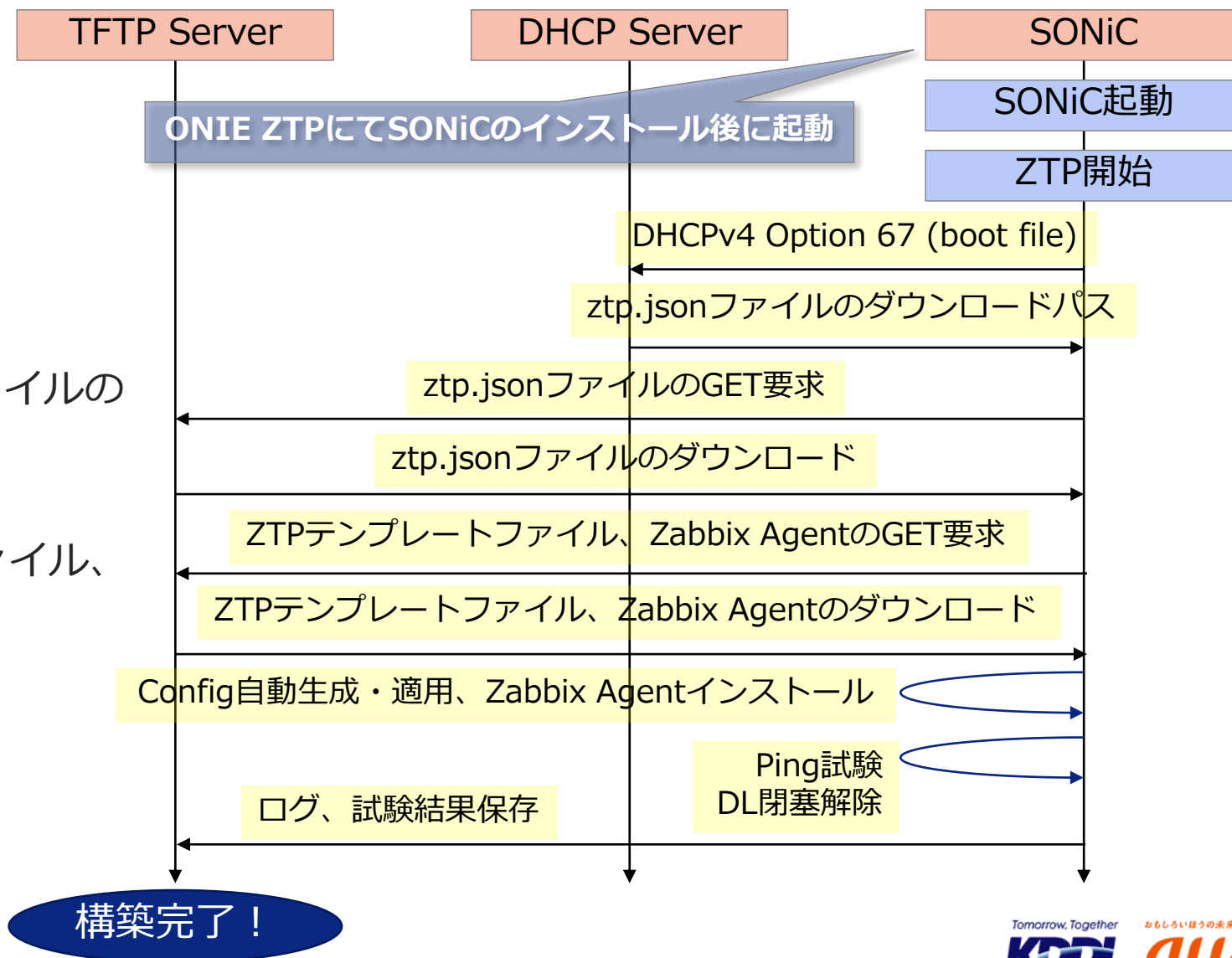
- SONiC起動後は右図のようなシーケンスにより構築

- ztp.json

- ZTPのシーケンス及びスクリプトファイルのパスを記載

- ZTPテンプレートファイル

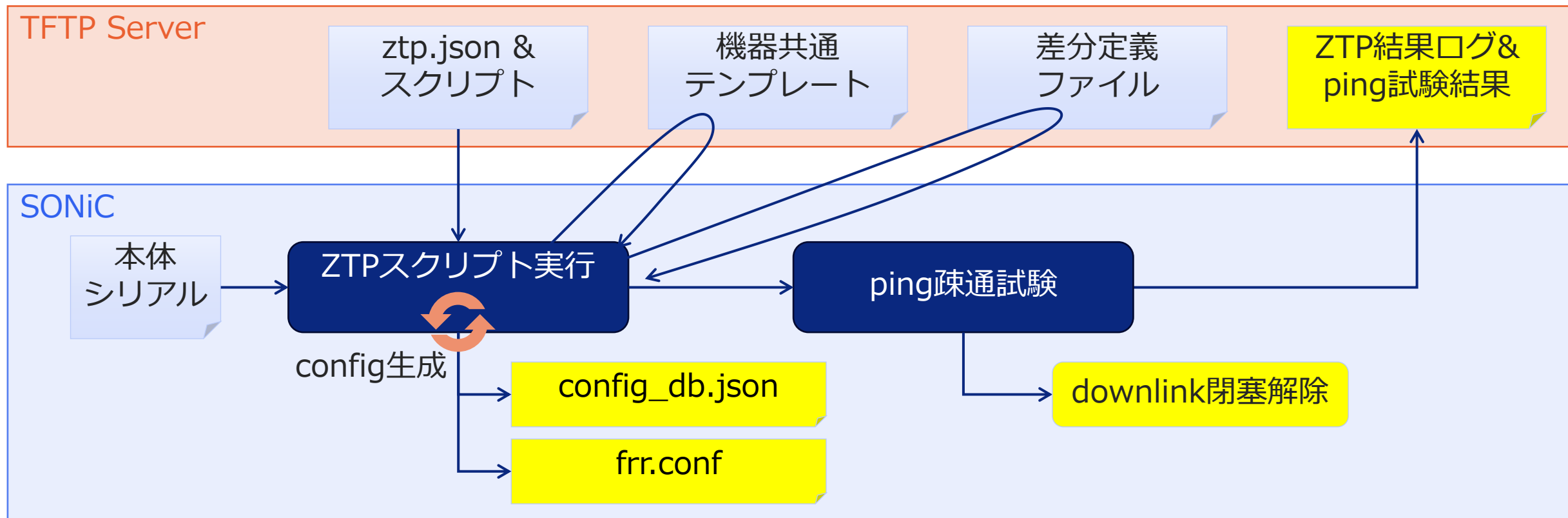
- 機器共通テンプレートや差分定義ファイル、変換スクリプトなど



# 機器個別のconfig設定から疎通試験まで

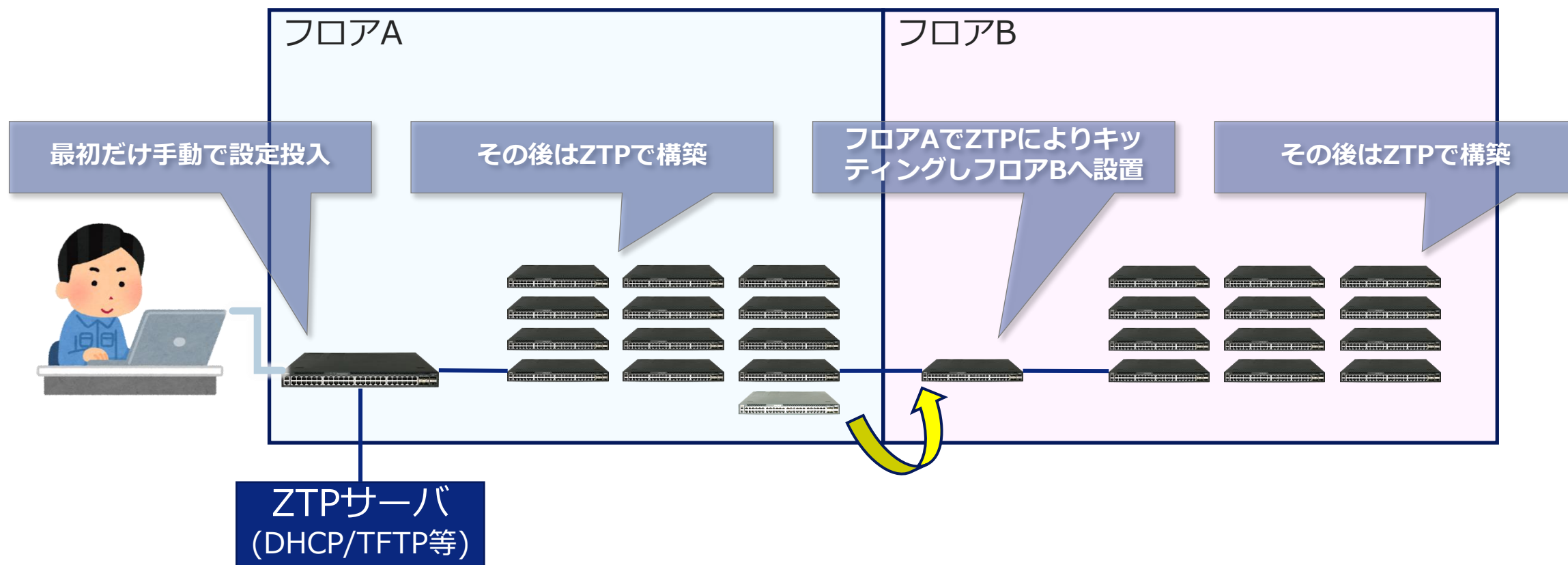
## ■ 機器固有の設定はztp.jsonの実行で生成

- 機器個別の情報をサーバー上の“差分定義ファイル（CSV）”に集約
- SONiCがZTPスクリプトの中で自身のシリアルをもとに検索しConfigを生成
- ZTP結果ログと試験結果をサーバにアップロード



# ZTPによるフロア単位での構築の流れ

- 基本は結線と電源投入で構築は完了する
  - 最初の1台だけは手動で設定が必要
  - 別フロア展開する場合はZTPで先にキットティング



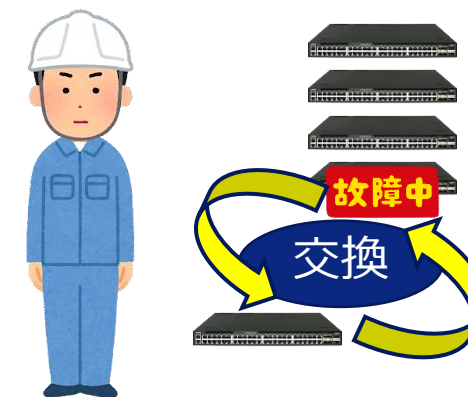
## ■ 交換対応がほぼ現地の交換作業だけに

- 設定に関わる部分は差分定義CSVファイルのS/Nを交換前⇒交換後に修正するだけ
- その後は電源ONでZTPにより故障前と同一の設定で起動する

S/N	hostname	router-id	ASN	...
AAAAAA	Spine-1	1.1.1.1	4200000000	...
AAAAAB	Spine-2	1.1.1.2	4200000001	...
...	...	...	...	...
hogehoge ⇒fugafuga	Spine-X	x.x.x.x	4200000XXX	...
...	...	...	...	...



① 差分定義ファイルのS/Nを  
交換品のもの書き換える

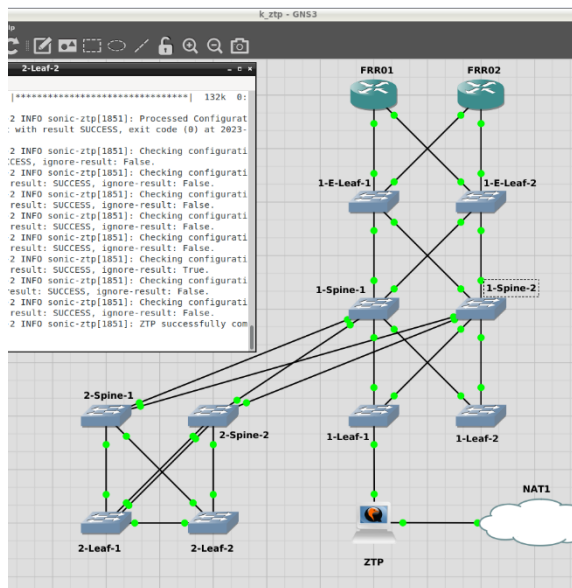


② 良品と交換し結線&電源ON



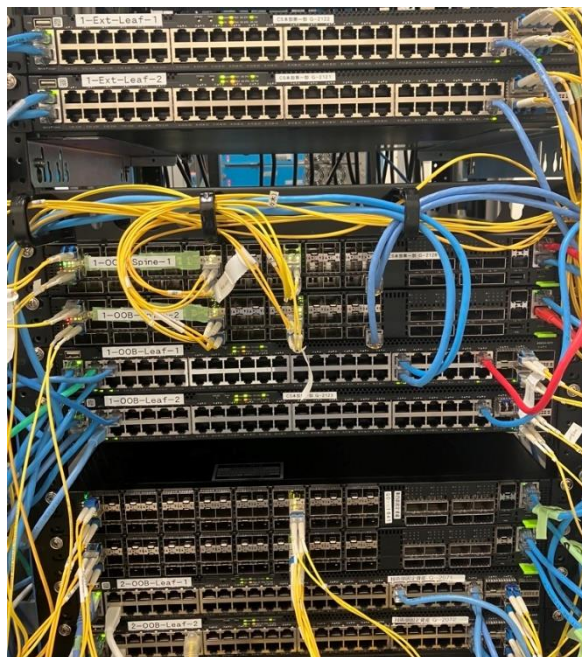
## ■ 導入に向け仮想環境及び用途毎に複数の検証環境を構築し並行して検証を実施

- 検証→課題検出→分析→パッチ→検証を繰り返し

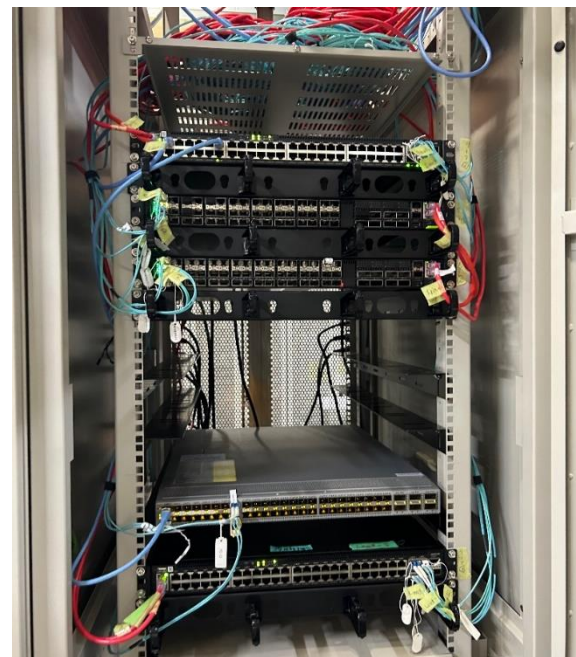


### 仮想検証環境

機能の動作や性能の検証の他、商用疑似・テナント相互接続検証環境で検出した課題の再現試験やOSパッチリリース後のリグレッション検証など



### 機能・性能検証環境



### 商用疑似検証環境

商用環境を疑似した検証や他のNW機器との相互接続検証、Upgrade試験など



### Private HaaS検証環境

サーバなど他HaaS機器を含めたE2E検証や運用保守の検証、検証環境の運営等

※一部、ベンダ/SIer拠点に構築した検証環境を含みます。

## ■ 各社で検証を進める中でいくつかの課題を検出し対処中（一部抜粋）

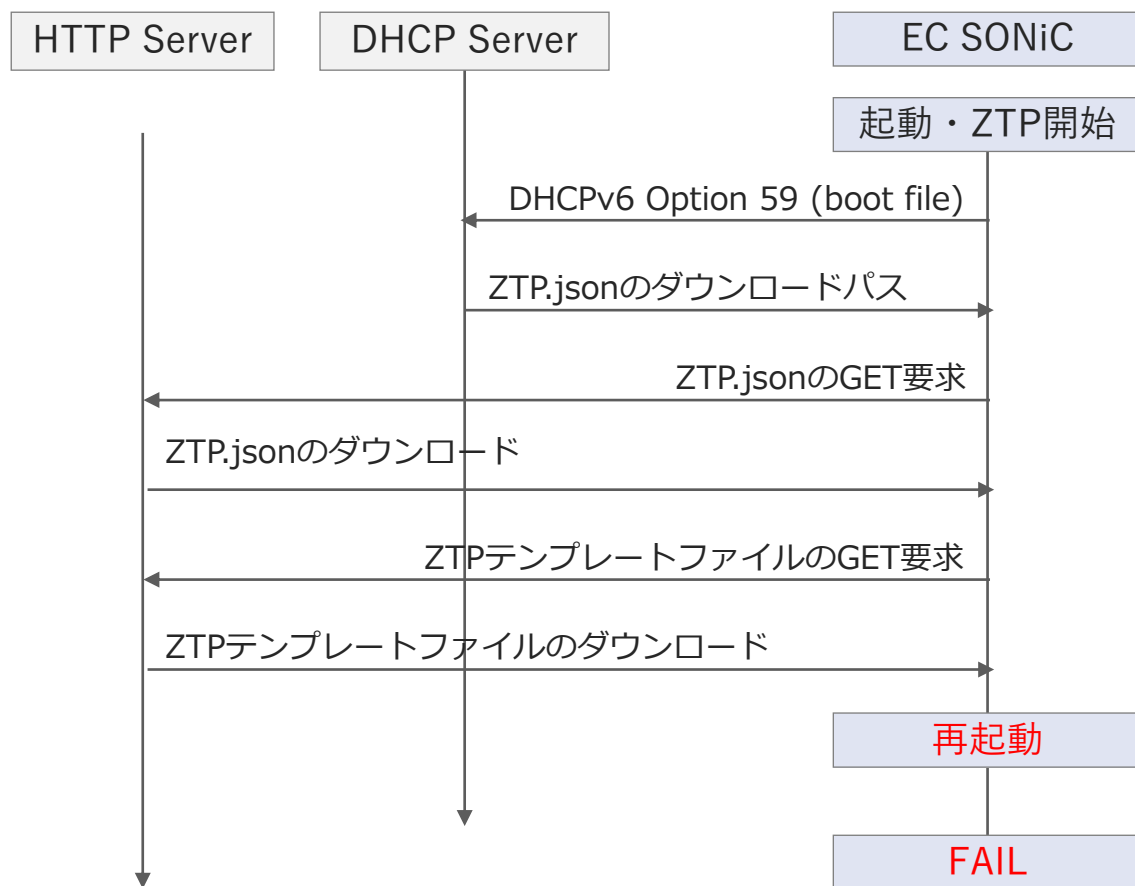
通番	内容	ステータス
不具合①	MACアドレス変更スクリプトをダウンロードした後、再起動時にDHCPv6リクエストが送出されず、ZTPプロセスがFAIL	修正済
不具合②	ポートのリンクアップ時、Anycast GatewayのIPv6 neighbor情報がアップデートされず、Neighbor状態がFAILEDとなる	修正済
不具合③	DHCPv6 Relayを設定したAnycast Gatewayに接続されるDHCP ClientでIPv6アドレス割り当てに失敗	修正済
不具合④	Leafで動作するdhcp6relayサービス/コンテナが再起動を繰り返す	修正中

上記①～④の不具合は、機能・性能検証環境（実機検証）にて検出

実機にパケットを流し、HWの中継処理を確認する過程で検出

## 不具合事象

MACアドレス変更スクリプトをダウンロードした後、再起動時にDHCPv6リクエストが送出されず、ZTPプロセスがFAIL



## 原因

eth0 (管理ポート) がupした後に、DHCPv6 clientが起動していない

## 改修方法

管理ポートがリンクアップしてから、DHCPv6 clientが起動するように処理の順番を制御

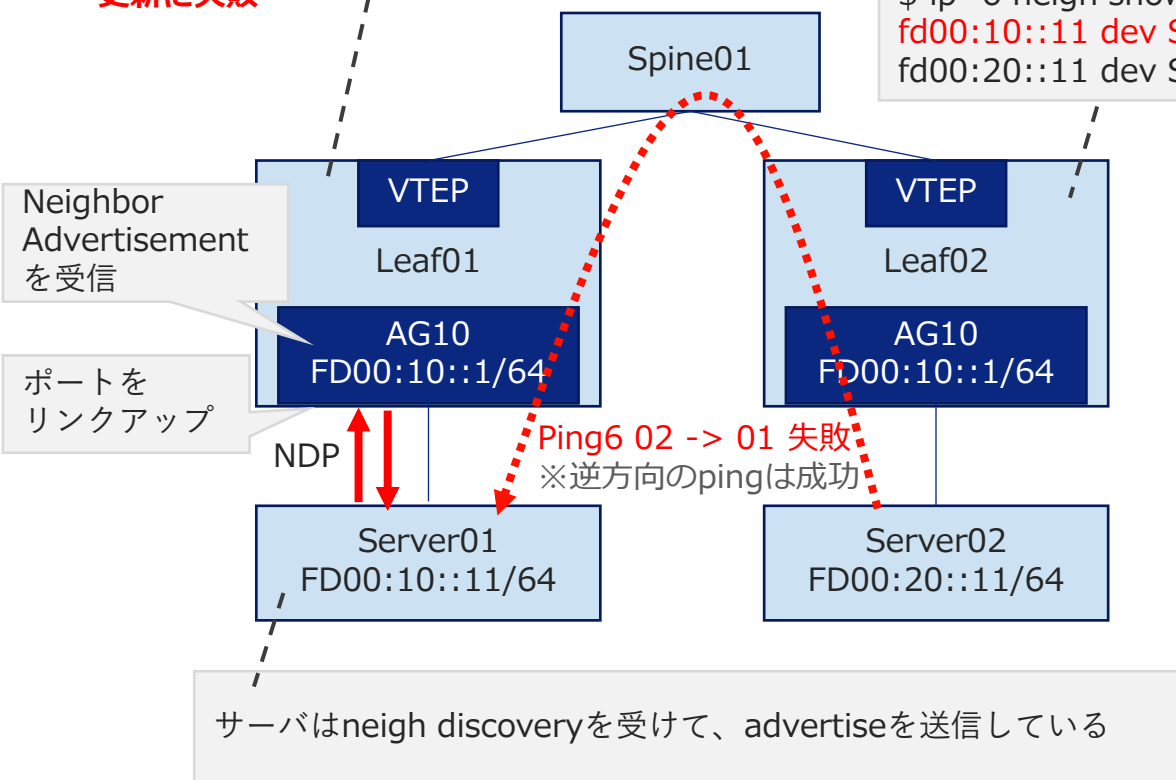
## 不具合事象

ポートのリンクアップ時、Anycast GatewayのIPv6 neighbor情報がアップデートされず、Neighbor状態がFAILEDとなる

```
$ ip -6 neigh show vrf Vrf1 | grep fd00
fd00:20::11 dev Vlan20 lladdr 00:0a:f7:c8:5a:a0 extern_learn NOARP proto zebra
fd00:10::11 dev SAG10 FAILED
```

更新に失敗

```
$ ip -6 neigh show vrf Vrf1 | grep fd00
fd00:10::11 dev SAG10 INCOMPLETE
fd00:20::11 dev SAG20 lladdr 00:0a:f7:c8:5a:a0 REACHABLE
```



## 原因

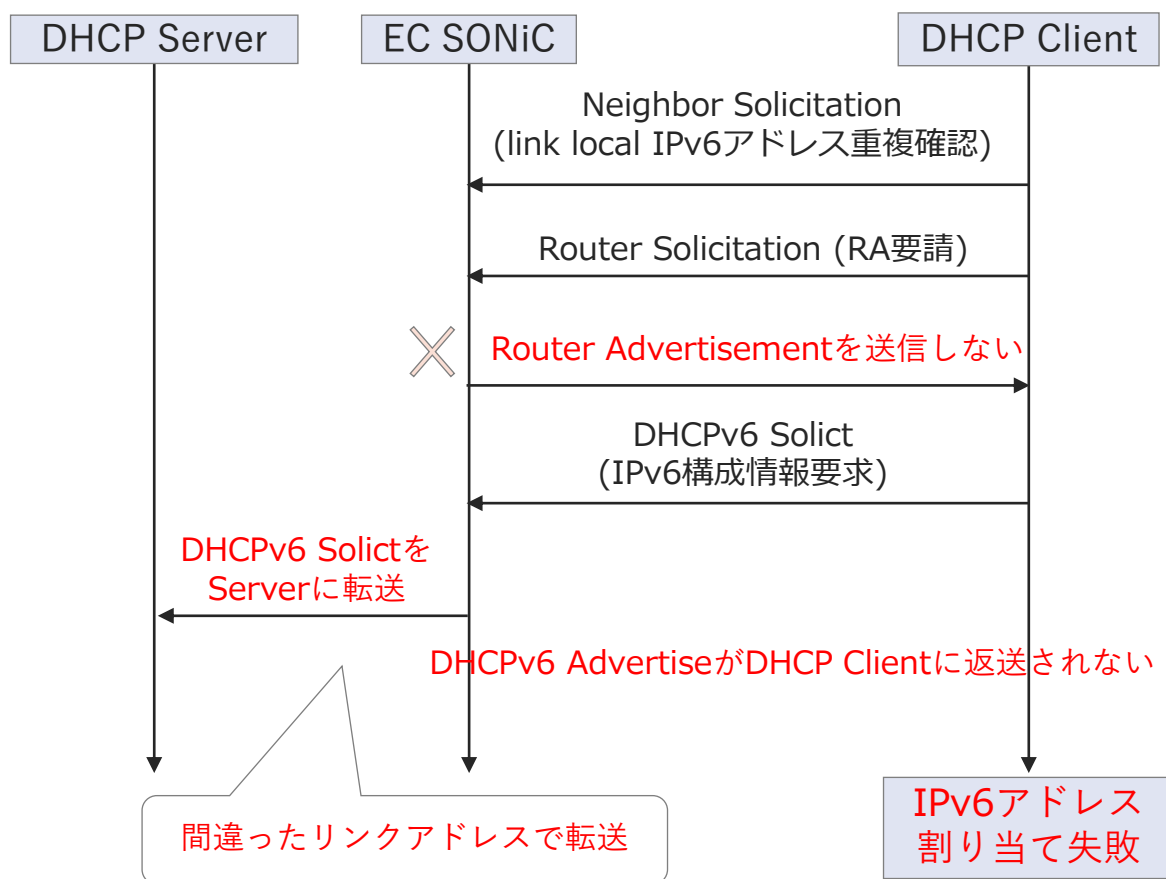
ポートのリンクアップのイベントを上位アプリに通知する手段がない (SONiCコミュニティの実装)  
そのため、neighbor情報のアップデートがされない

## 改修方法

ポートのリンクアップ時に、Anycast Gatewayにて neighbor情報を更新できる状態へ遷移する仕組みを追加

## 不具合事象

DHCPv6 Relayを設定したAnycast Gatewayに接続されるDHCP ClientでIPv6アドレス割り当てに失敗



### 原因

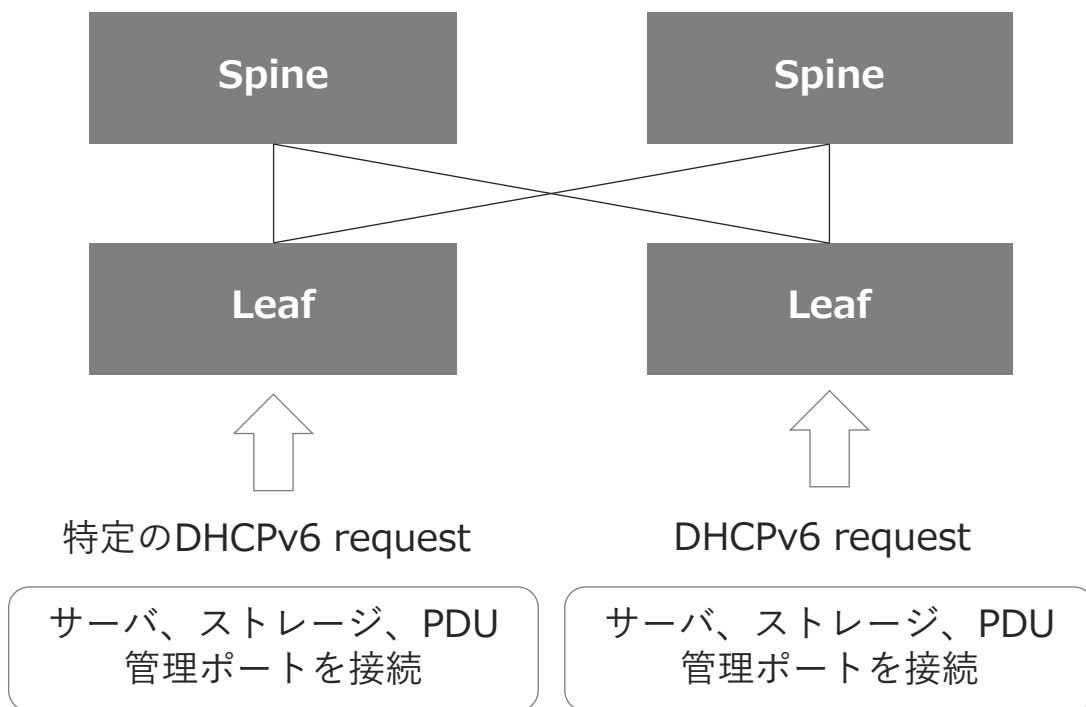
DHCP ClientからRSを受け取ってもRAを送信しない  
DHCPv6 Solicitを転送する際、間違ったリンクアドレスを送信元にして転送

### 改修方法

Anycast GatewayにてDHCPv6 Relay動作を修正

## 不具合事象

Leafで動作するdhcp6relayサービス/コンテナが再起動を繰り返す



[dhcp6relay] Segfault can be seen when updates are made to DHCP\_RELAY table #12496

Closed vivekrmv opened this issue on Oct 26, 2022 · 0 comments · Fixed by sonic-net/sonic-dhcp-relay#25

<https://github.com/sonic-net/sonic-buildimage/issues/12496>

## 原因

DHCPv6 information-requestを受信した際に、dhcp6relayコンテナが処理できずクラッシュする

## 改修方法

改修版パッチOSは**7/10リリース予定**  
(Communityでは202205で改修済)

## ■ HaaS Management NWで実現できたこと

- 構築や保守交換は電源投入後、約10分で完了
- これまでエンジニアがやっていた内容をZTPに取り入れ、保守交換フローが簡易に
- NWエンジニアが自ら考え、工夫し、実装するように

## ■ まだまだ課題も、、、

- 構築失敗は能動的に検知する必要あり（ログが上がらない、SSH出来ない等）
- IPv6やL3VPN+Anycast Gateway関連でバグにあたるが多かった印象。使い方に応じて検証し修正していく必要あり。





- ZTPに取り組んでいる？やってる場合はどこまでZTPで出来てる？  
あえてZTPをやってないなどありますか？  
また、ZTPを進めすぎて困ったことなどありますか？
- 完全なZTP実現のためにこういった機能が必要だと考えだと思いませんか？
- HW/NWインフラ構築のスピードアップの工夫はどういうことをやっている？
- IPv6化を進めるモチベーションはありますか？  
進めている場合はどこまで進めてますか？

「つなぐチカラ」を進化させ、  
誰もが思いを実現できる社会をつくる。

# KDDI VISION 2030

