仮想化基盤収容NW
設備の4年がかりの大規模更改

2024/7/4

NTTドコモ

■ 自己紹介

1. ドコモのNW仮想化基盤の概要
2. ドコモの仮想化基盤収容NWの更改における前提条件
3. 更改手順
4. 更改の苦労話
5. 総括/議論ポイント

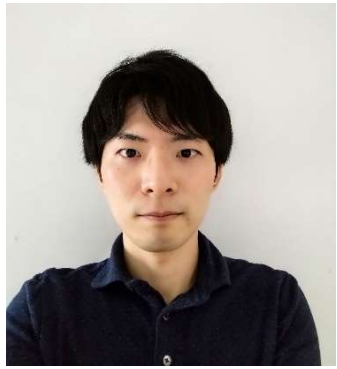
■ 自己紹介

1. ドコモのNW仮想化基盤の概要
2. ドコモの仮想化基盤収容NWの更改における前提条件
3. 更改手順
4. 更改の苦労話
5. 総括/議論ポイント

自己紹介

小野 健一

西川 哲平



渋谷彰寿

■ 自己紹介

1. ドコモのNW仮想化基盤の概要
2. ドコモの仮想化基盤収容NWの更改における前提条件
3. 更改手順
4. 更改の苦労話
5. 総括/議論ポイント

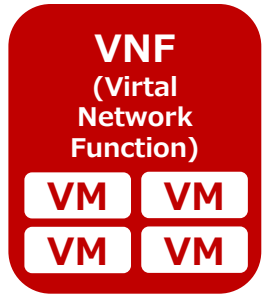
1. ドコモのネットワーク仮想化基盤の概要 1/6

- ドコモでは2016年以来、コアNWシステムを収容するETSI ISG NFVに準拠したNW仮想化基盤を構築・運用しています。

Data Centers **24**

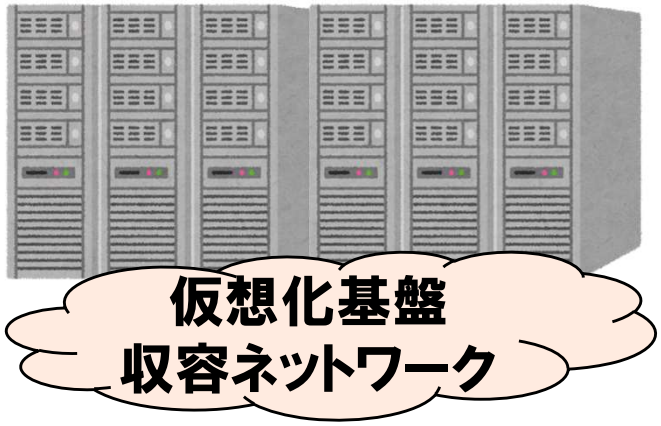
Virtual Network Function **40＋**

Virtual Machines **250,000＋**



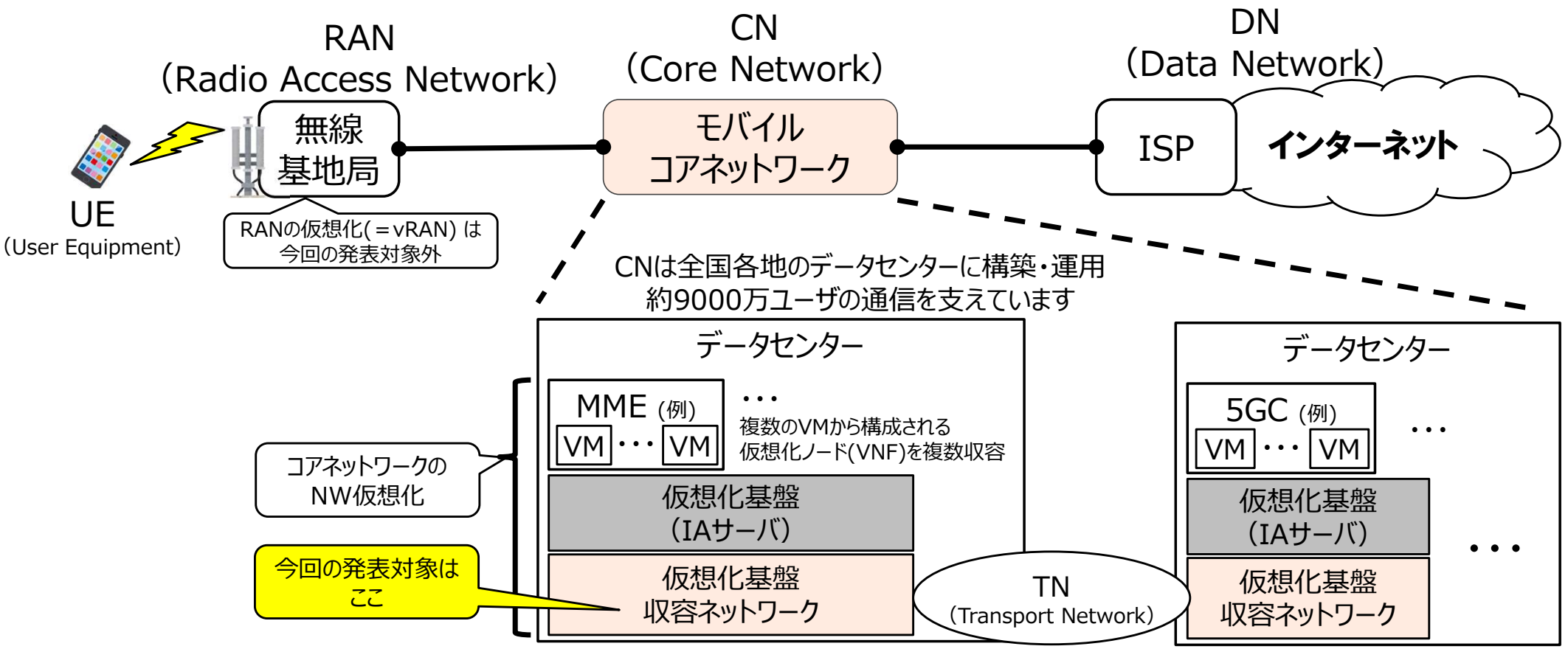
Servers **9,000＋**

NW機器 **3,500＋**



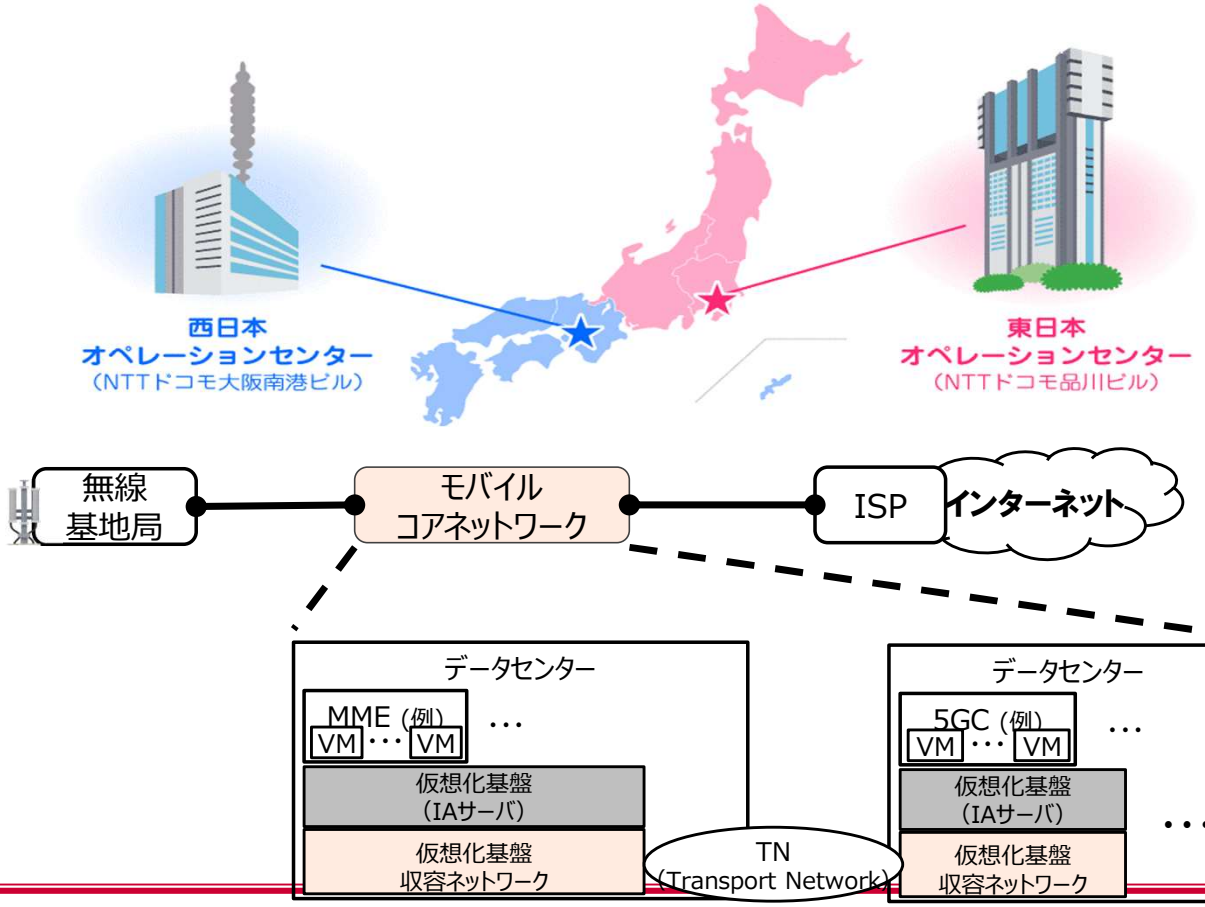
1. ドコモのネットワーク仮想化基盤の概要 2/6

- 今回の発表の対象はNW仮想化基盤において物理的なサーバを収容するNW機器が対象です。

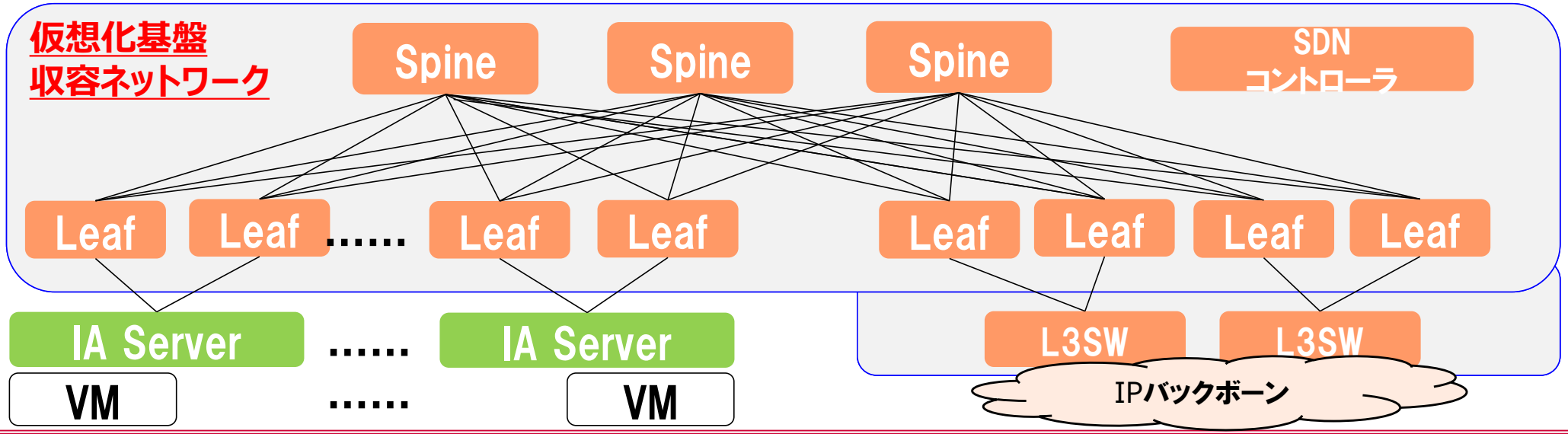


1. ドコモのネットワーク仮想化基盤の概要 3/6

■ 全国システム横断的な監視・運用部門、各システム毎の保全担当や工事担当、拠点ごとの現地部門など、多くの関係部門が連携して運用しています。



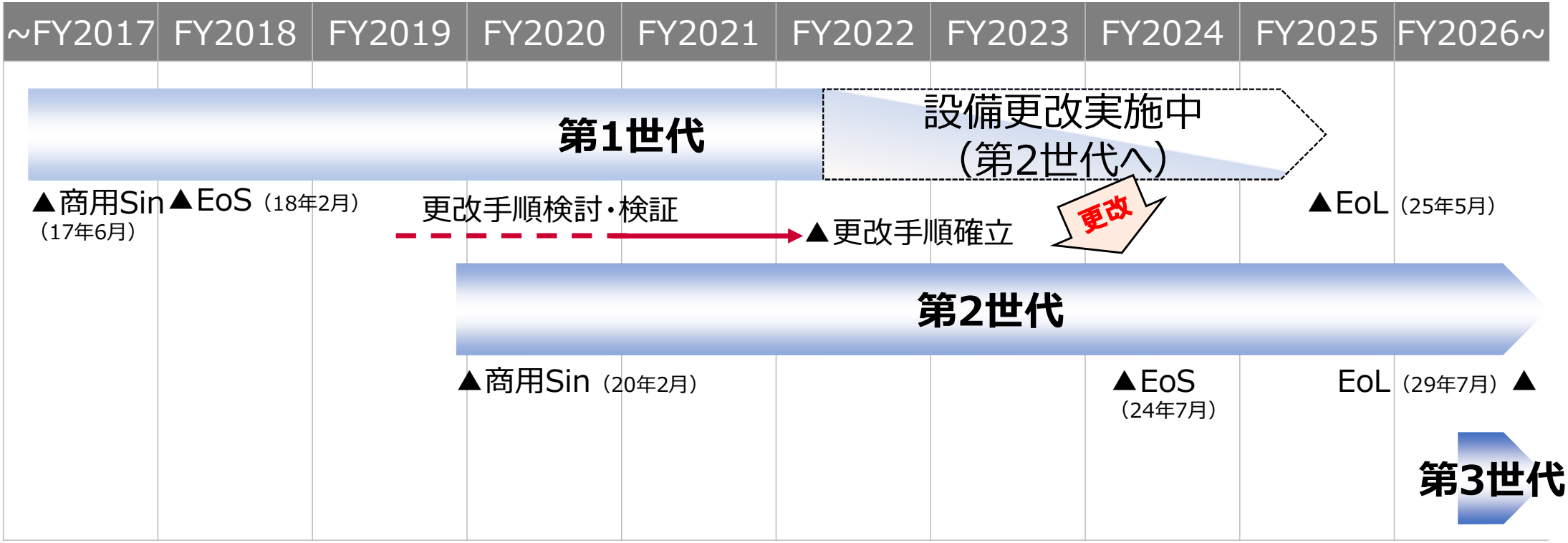
- 仮想化基盤収容NWの物理的な構成は、データセンターネットワークとしては一般的なファブリックNWです。
  - 1基盤当り数百万加入者収容であり、高い信頼性が求められる
    - 故障発生の都度、駆け付け即時対応せずに済むよう余裕を持たせた冗長構成
    - 機器台数が多いため、SDN（Software Defined Network）コントローラで一元管理
  - 監視部門や保全部門、現地作業者など、多くの関係者が運用、工事に関与
  - 収容アプリケーション(VNF)は主管部門が異なるため、多くの部門間での連携必須



1. ドコモのネットワーク仮想化基盤の概要 5/6

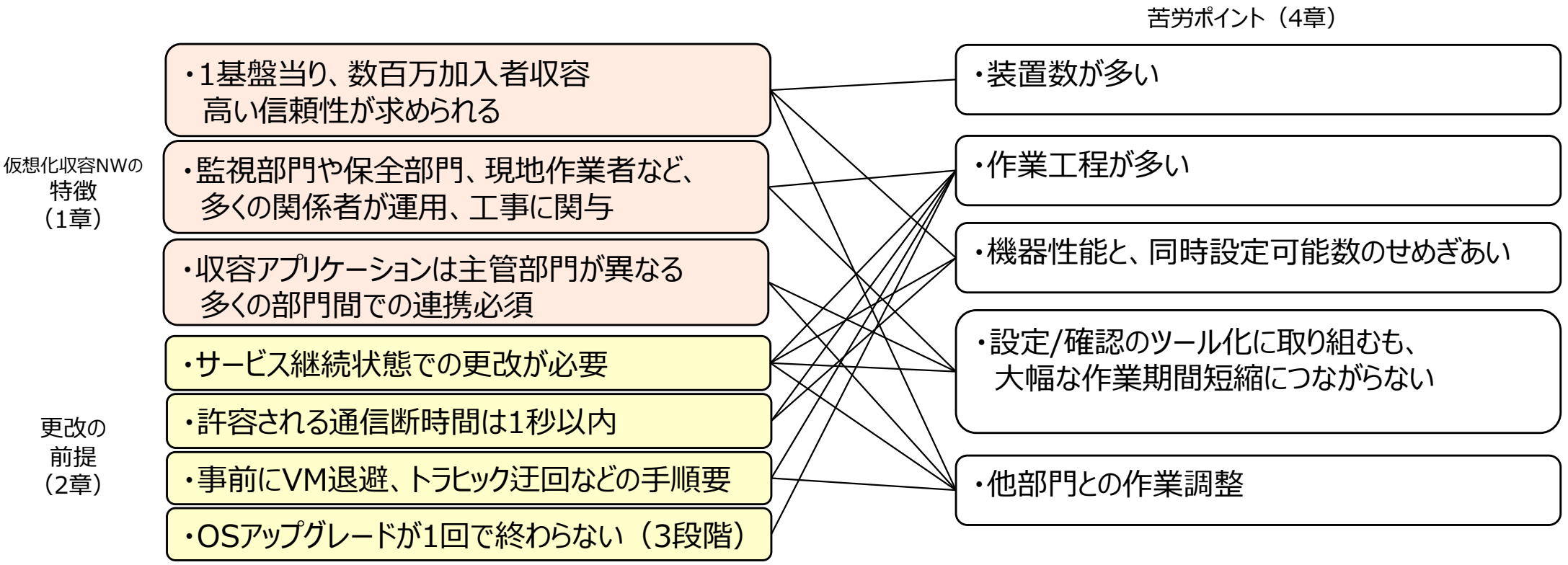
- 2016年度から商用構築していた第一世代HWは2025年度にEoL予定
- 2022年度から第一世代HW⇒第二世代HWへの更改を実施中です。

EoS：End of Sales＝販売終了
EoL：End of Life=サポート終了



なぜ単なるハードウェア更改に4年もかかるのか？

- 更改の前提・制約、策定した更改手順、苦労ポイントをご紹介します



■ 自己紹介

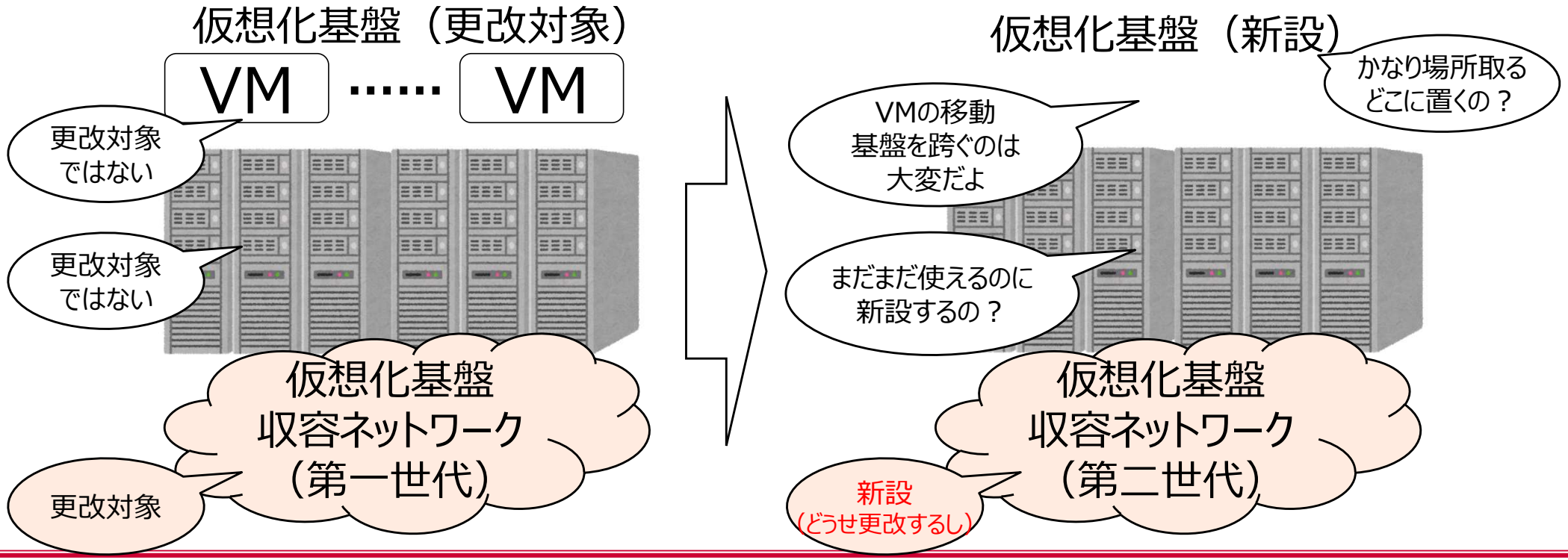
1. ドコモのNW仮想化基盤の概要
2. ドコモの仮想化基盤収容NWの更改における前提条件
3. 更改手順
4. 更改の苦労話
5. 総括/議論ポイント

2.ドコモのネットワーク仮想化基盤　更改の背景

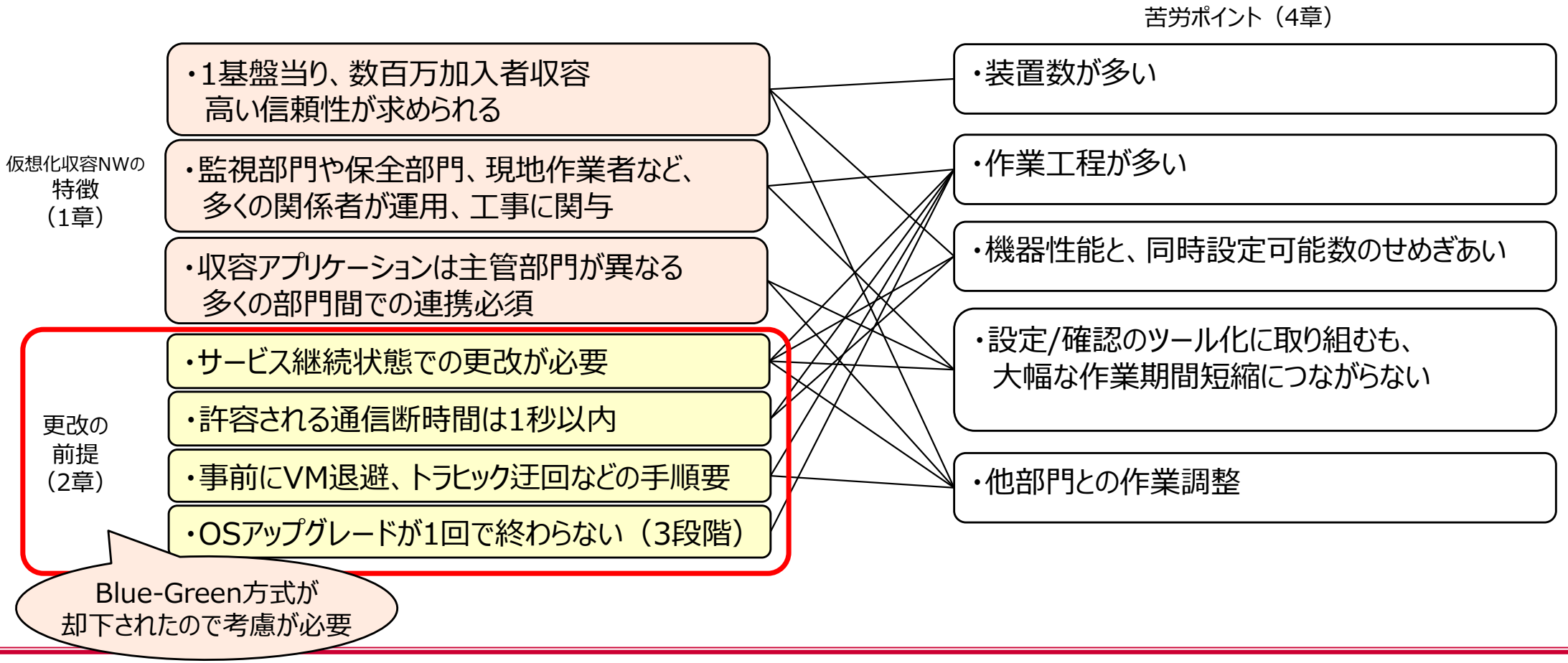
- なぜ更改が必要なのか？
  - EoLとなる設備の更改
  - 安定したサービス提供継続のため

Blue-Green方式（却下された案）

- サーバも含めて新規に仮想化基盤を構築して、その後にVMを移動する。
- 手順はとてもシンプルだが、フロアスペース確保等の課題も多い。検討の結果、却下された。
- このため、基盤としてサービス継続しながら更改する手順の検討が必要となった。（次スライド以降）

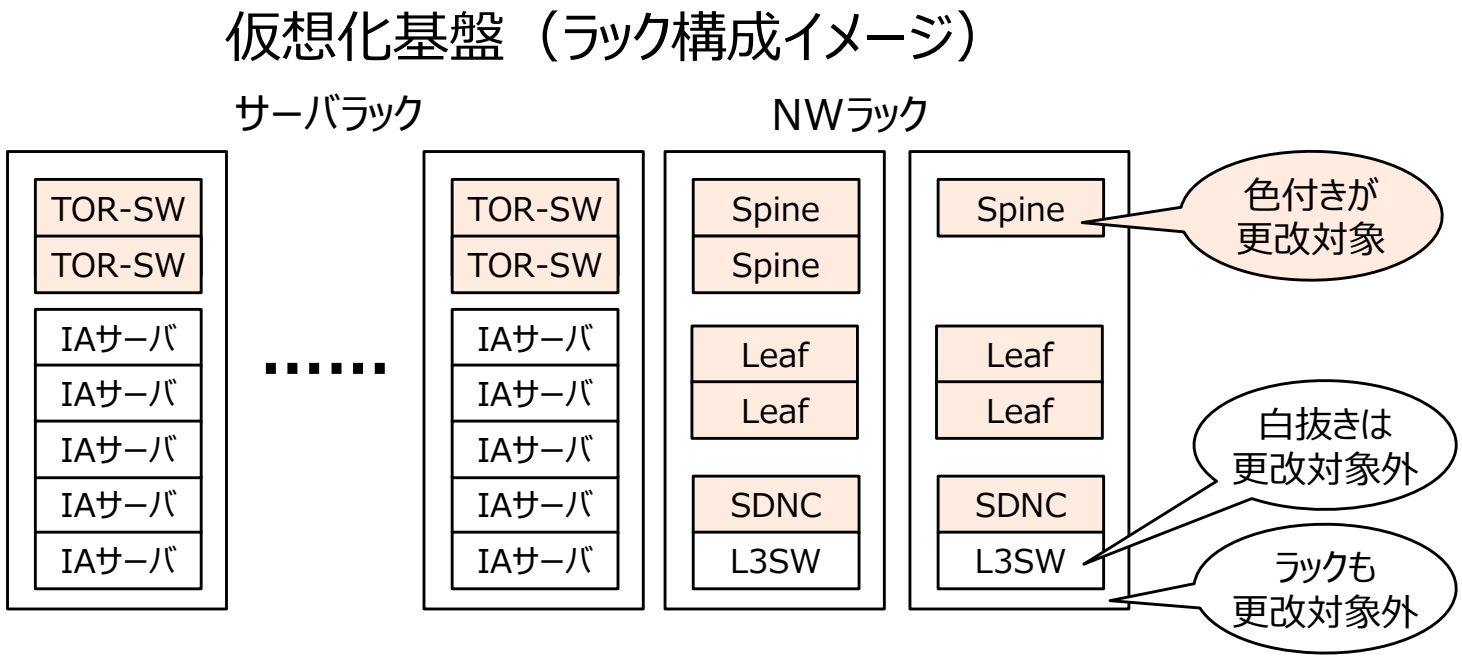


Blue-Green方式が却下されたことで、考慮すべき前提・制約が多数発生



2.更改手順検討の前提条件 1/5

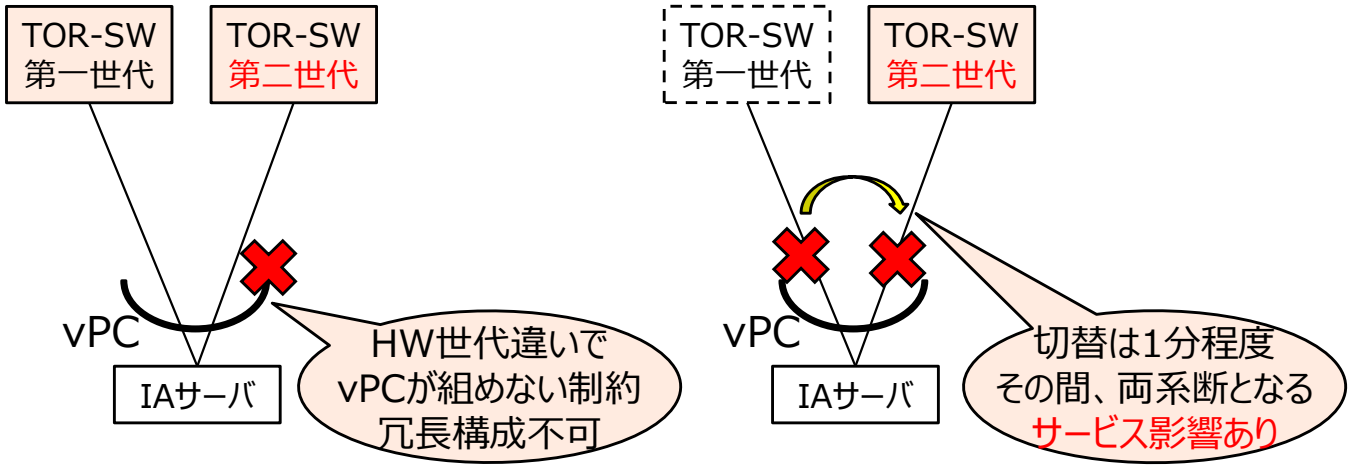
- サービス継続状態でNW機器だけ更改するという要求条件
- マシン室フロアの余裕がない局舎が多いので、ラック新設を必要としない手順とする必要がある



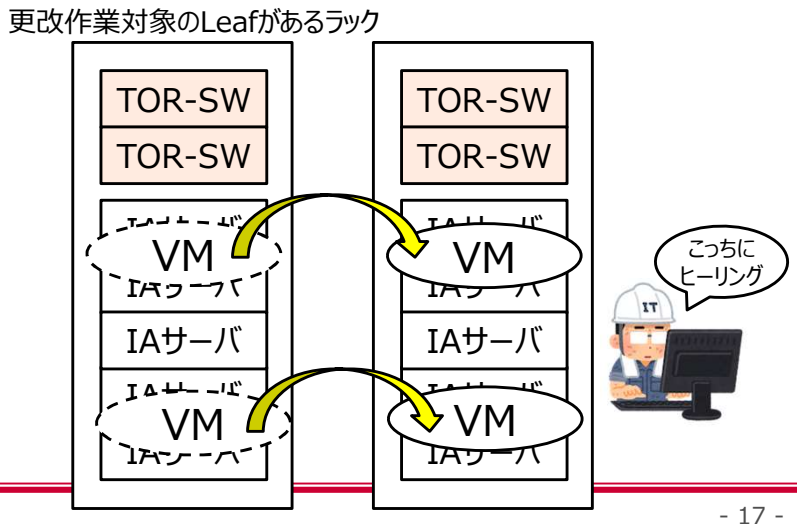
2.更改手順検討の前提条件 2/5

- ToR-SW（Leaf）とIAサーバ間の接続：vPCによる冗長確保する設計としている区間あり
- しかし、ハードウェア世代が異なるLeaf間ではvPCが組めない制約あり
  - 更改中の冗長構成が取れない
  - 第一世代から第二世代LeafへのvPC切替に1分程度のサービス断が生じる
  - 作業対象ラックのVMを事前に別ラックへヒーリングによる退避を行う手順とすることで対処する必要あり

ToR-SWの更改過渡期で生じる
第一世代･第二世代Leaf間で冗長確保できない問題



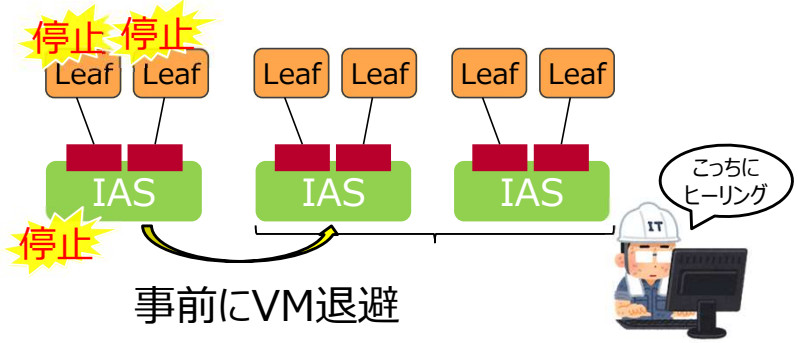
対策
VMを事前に別ラックに退避



2.更改手順検討の前提条件 3/5

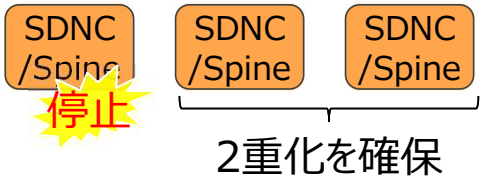
- サービス継続状態での更改のため、冗長構成を確保したまま作業実施する必要がある
  - 更改作業の事前/事後作業として、切替・切離し・トラヒック迂回等の手順を入れる必要あり

【2重化装置】
・IAサーバ収容Leaf



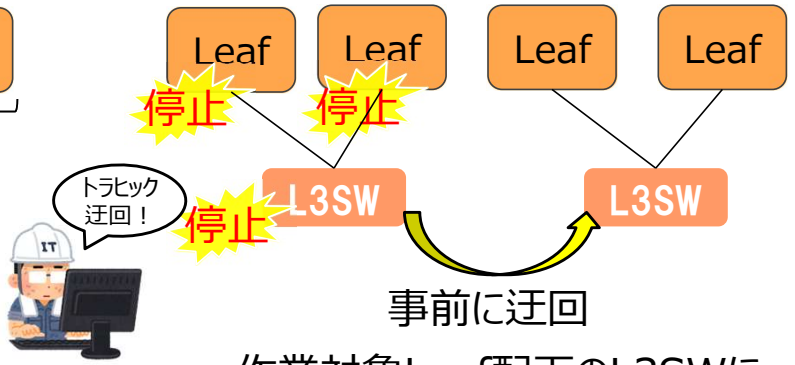
作業対象Leaf配下のIAサーバに
収容されているVMを事前に退避し、
サービス影響ない状態で作業する。

【3重化装置】
・SDNC、Spineなど



作業対象装置以外で冗長構成
が確保できていることを確認し、
サービス影響ない状態で作業する。

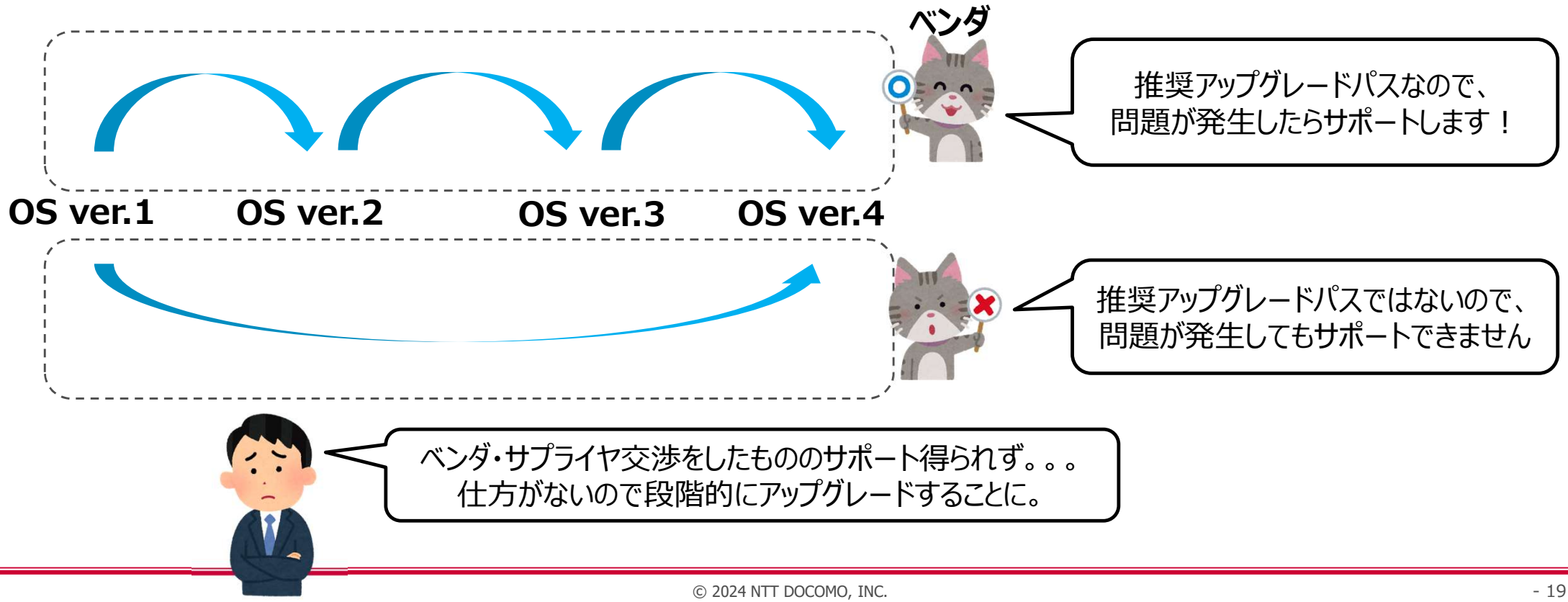
【4重化装置】
・外部接続用Leaf



作業対象Leaf配下のL3SWに
流れているトラヒックを事前に迂回し、
サービス影響ない状態で作業する。

2.更改手順検討の前提条件 4/5

- 第一世代のOSバージョンから、第二世代で利用の最新OSバージョンへのアップグレードパスなし
- 運用中OSバージョンは出来るだけ揃えておきたいので、3段階のアップグレード作業が必要



2.更改手順検討の前提条件 5/5

- OSアップグレードの順序性
  - SDNコントローラ(SDNC) ⇒ Leaf/Spine の順に実施する必要がある。
  - LeafとSpineには順序性なし
- SDNコントローラ更改における考慮事項
  - 第一世代SDNCは、OS ver.1~4全てで動作が可能
  - 第二世代SDNCは、OS ver.1での動作不可
  - ⇒ 第一世代SDNCでOS ver.4までのアップグレードを実施後に、第二世代へのハードウェア更改をする必要がある
- Leaf更改における考慮事項
  - ハードウェア交換と同時にOS ver.1⇒2へのアップグレードを行う必要がある
- Spine更改における考慮事項
  - ラインカード/制御部/電源モジュールのハードウェア更改を行う
  - ハードウェア交換とOSアップグレードの順序性なし
- その他Fabric全体での考慮事項
  - Fabric全体への通信影響が生じる設定変更が必要な場合、許容される通信断時間は1秒以内

■ 自己紹介

1. ドコモのNW仮想化基盤の概要
2. ドコモの仮想化基盤収容NWの更改における前提条件
3. 更改手順
4. 更改の苦労話
5. 総括/議論ポイント

3.更改手順

■ 更改手順検討の前提条件を踏まえて下記の手順とした。

- 各Phase、装置数分ひたすら繰り返し。（Phase②は17セット、Phase③は2セット、など作業が多い）

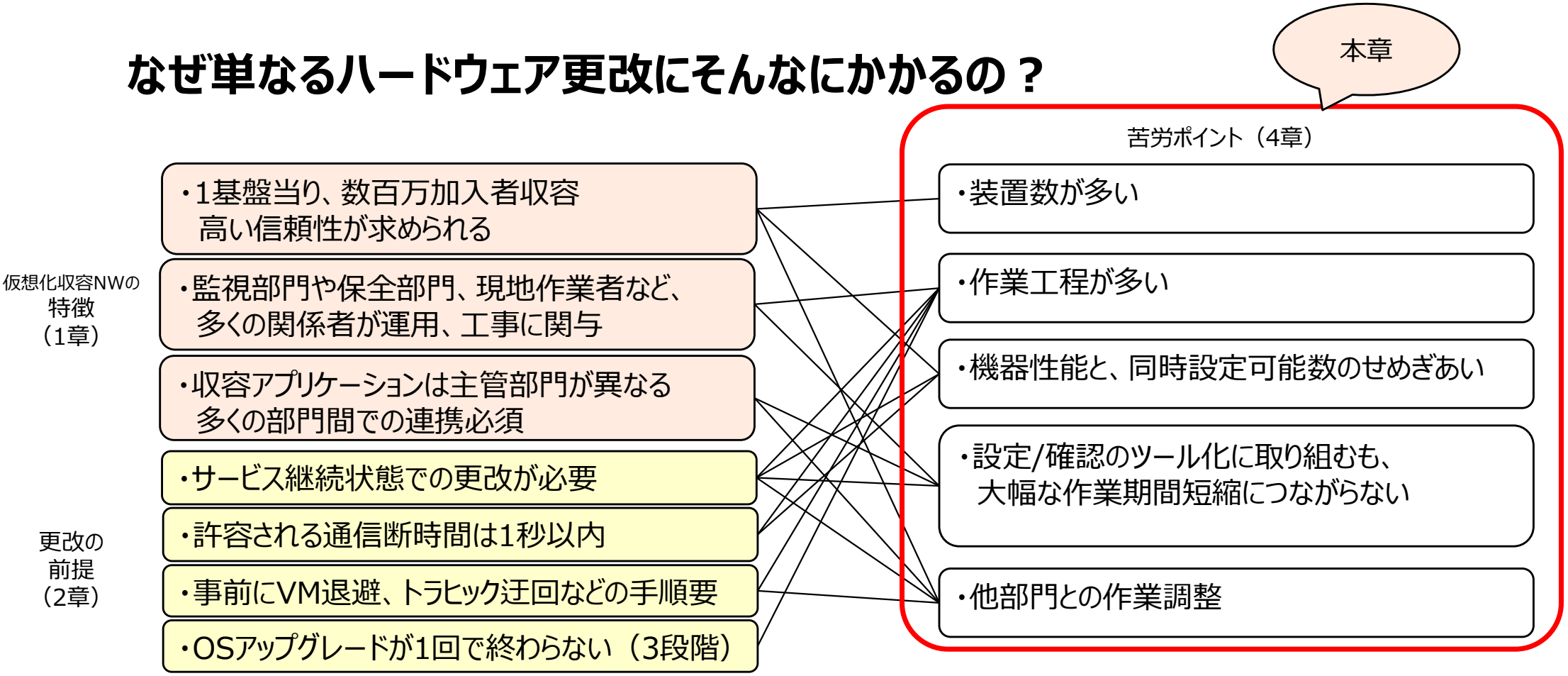
Phase	作業概要	SDNC		IAS収容Leaf		外部接続Leaf		Spine		FabricIF Speed	設定			WAツール	
		HW	SW	HW	SW	HW	SW	HW	SW		Flood化	DOM	監視登録	ツール１	ツール２
⓪	初期状態	Gen.1	OS ver.1	Gen.1	OS ver.1	Gen.1	OS ver.1	Gen.1	OS ver.1	40G	未	有効	—	作動	—
①	SDNC SW更改	〃	OS ver.2	〃	〃	〃	〃	〃	〃	〃	〃	無効	—	〃	〃
繰り返し ②-1	VM 作業対象ラックから退避	〃	〃	〃	〃	〃	〃	〃	〃	〃	〃	〃	—	〃	作動
繰り返し ②-2	Leaf HW更改[装置交換]	〃	〃	Gen.2	OS ver.2	〃	〃	〃	〃	〃	〃	〃	該当Leaf 再登録	〃	〃
繰り返し ③-1	外部接続機器 トラヒック迂回	〃	〃	〃	〃	〃	〃	〃	〃	〃	〃	〃	—	停止	〃
繰り返し ③-2	Leaf HW更改[装置交換]	〃	〃	〃	〃	Gen.2	OS ver.2	〃	〃	〃	〃	〃	該当Leaf 再登録	〃	〃
④	Spine SW更改	〃	〃	〃	〃	〃	〃	〃	OS ver.2	〃	〃	〃	—	〃	〃
⑤	全体 SW更改	〃	OS ver.3	〃	OS ver.3	〃	OS ver.3	〃	OS ver.3	〃	〃	〃	—	〃	〃
⑥	全体 SW更改	〃	OS ver.4	〃	OS ver.4	〃	OS ver.4	〃	OS ver.4	〃	〃	〃	—	〃	〃
⑦	Flood化	〃	〃	〃	〃	〃	〃	〃	〃	〃	済	有効	—	〃	〃
⑧	Spine HW更改[ﾓｼﾞｭｰﾙ交換]	〃	〃	〃	〃	〃	〃	Gen.2	〃	100G	〃	〃	fabricリンク 再登録	〃	〃
⑨	SDNC HW更改[装置交換]	Gen.2	〃	〃	〃	〃	〃	〃	〃	〃	〃	〃	—	〃	〃
事後	設定管理システム データ移行（OS ver.1⇒4）	Gen.2	OS ver.4	Gen.2	OS ver.4	Gen.2	OS ver.4	Gen.2	OS ver.4	100G	済	有効	—	停止	作動

■ 自己紹介

1. ドコモのNW仮想化基盤の概要
2. ドコモの仮想化基盤収容NWの更改における前提条件
3. 更改手順
4. 更改の苦労話
5. 総括/議論ポイント

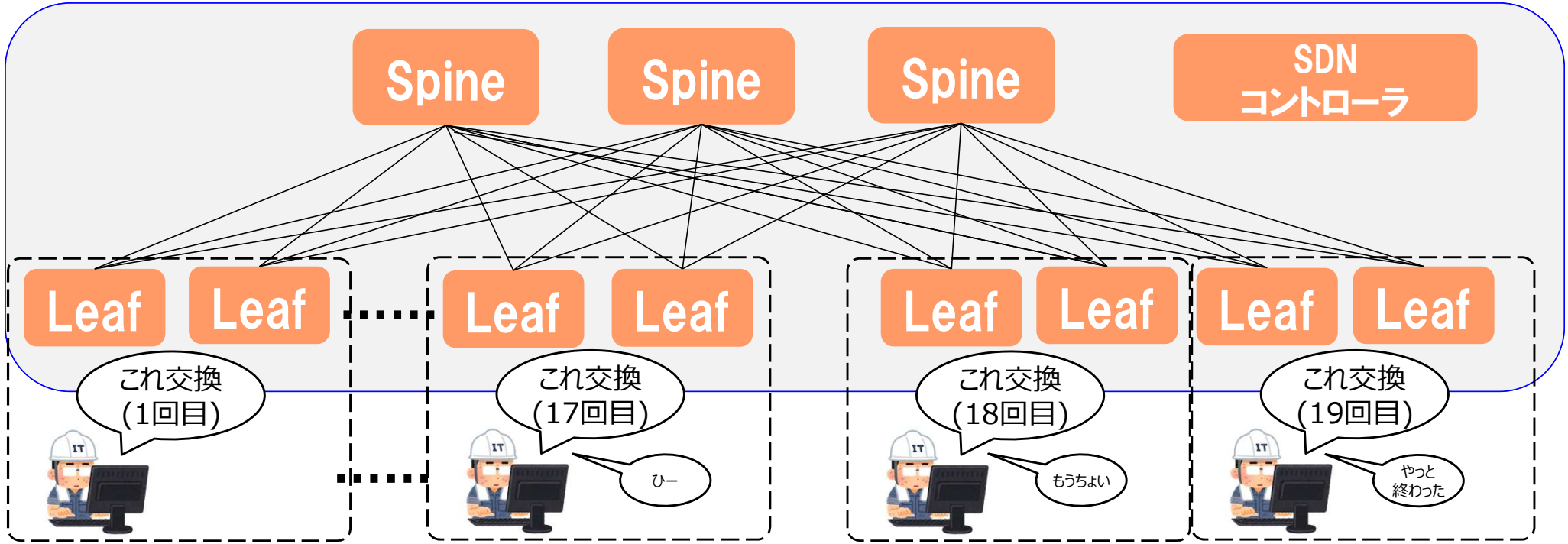
■ 2022年から2025年までの4年がかりで更改に取り組んでいます。

なぜ単なるハードウェア更改にそんなにかかるの？



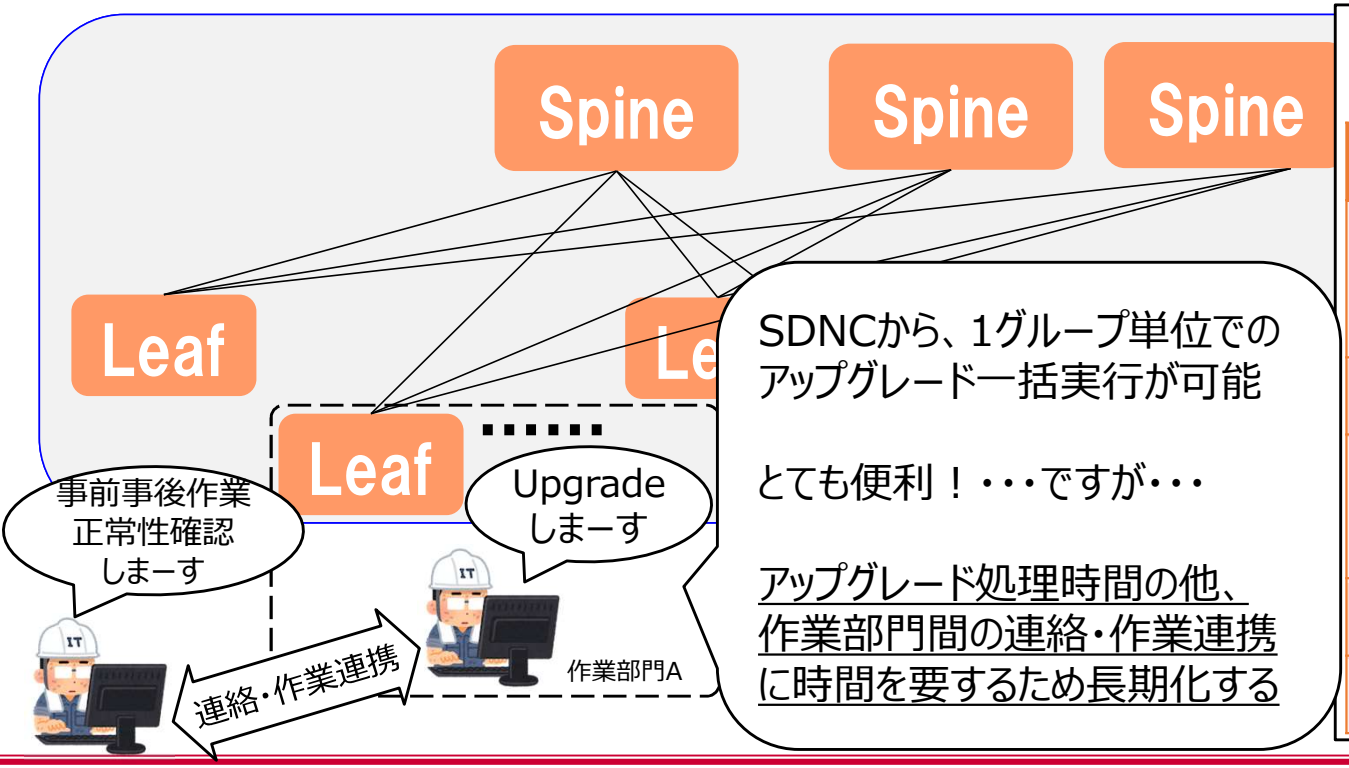
4.更改の苦労話 1/5　Leaf HW更改の作業工程が多い（Phase②、③）

- 安全・確実な作業実施のため、１つのPhase内でも細かく作業を分けて実施
- 1作業工程に1営業日を割り当て
  - ⇒ LeafのHW更改であるPhase②、③だけで19営業日かかる



4.更改の苦労話 2/5　OS更改の作業工程が多い（Phase⑤、⑥）

- 安全・確実な作業実施のため、同時にアップグレードする機器を17のグループに分けて実施
- 1営業日に1～3グループを割り当て、1回のアップグレード完了までに5~6営業日かかる
- このアップグレードを2回繰り返し。 10～12営業日を要する



SDNCから、1グループ単位での
アップグレード一括実行が可能

とても便利！・・・ですが・・・

<u>アップグレード処理時間の他、
作業部門間の連絡・作業連携
に時間を要するため長期化する</u>

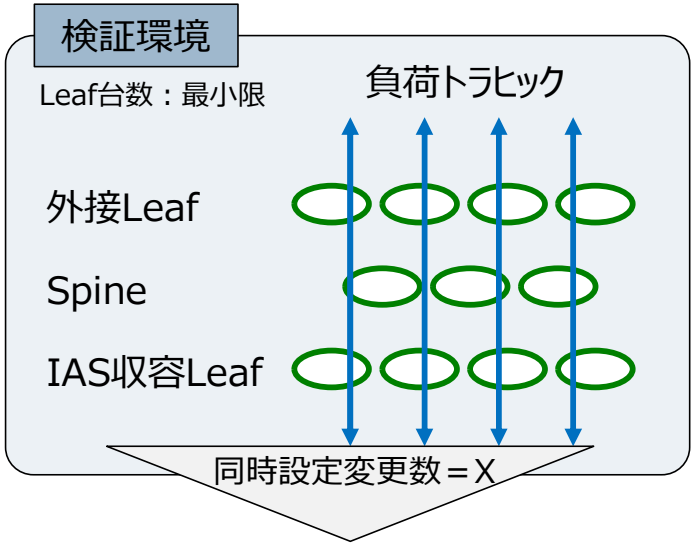
収容している機器の特性を踏まえて、
冗長性を失わないようにグループ分け

対象	グループ	
IAサーバ収容Leaf（1系）	5グループ	1営業日
		1営業日
外部接続Leaf（2,3系）	2グループ	
IAサーバ収容Leaf（0系）	5グループ	
外部接続Leaf（0,1系）	2グループ	
Spine	3グループ	1営業日

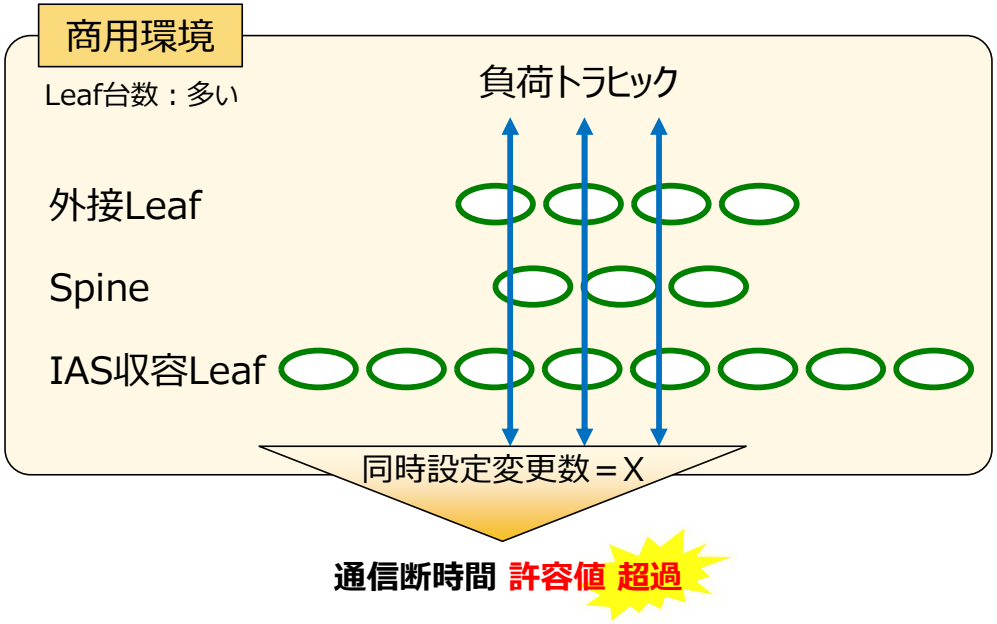
4.更改の苦労話 3/5　機器の性能との戦い１（Phase⑦）

- 前提
  - ARP解決されていないL2 Unknown Unicastパケットの扱いを第一世代から第二世代で変更した。
  - この設定変更は、仮想NW単位の設定である。（ARPをfabric全体にFloodする＝Flood化）
  - 同時に設定変更する仮想NW数が増えると、通信断時間が増加することが分かっていた。

- 許容通信断時間以内となる同時設定変更数を確認していた。・・・が、商用では許容値超過!!



通信断時間 許容値 以内
(0.2～0.9秒/中央値0.6秒)

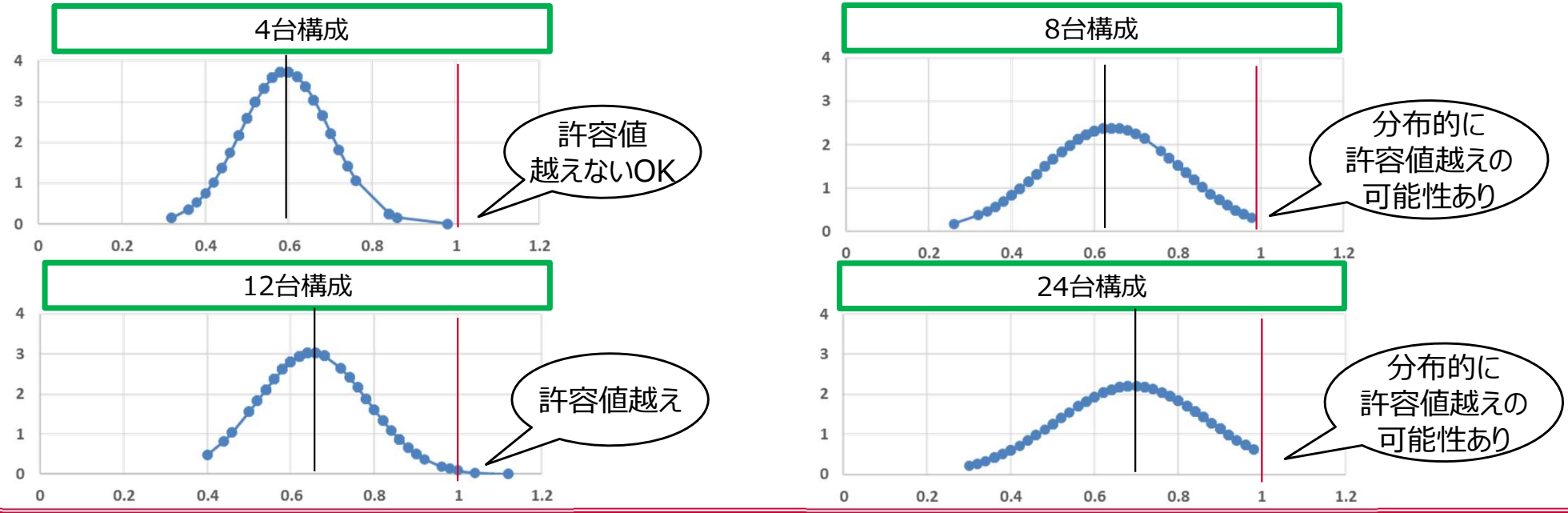


通信断時間 許容値 超過

4.更改の苦労話 3/5　機器の性能との戦い２（Phase⑦）

- 許容通信断時間を超過するのはLeaf台数の違いによるものと仮説を立て、追加検証を実施
  - Leaf台数増で、通信断時間の平均、標準偏差ともに増加し、許容値を超える確率が増加することを確認
  - 許容値以内に収まる同時設定変更数を改めて見極め、作業手順にフィードバック
    - 問題は解消したが、作業量は増えてしまった。

IAサーバ収容Leaf数の変化と通信断時間（同時設定変更数＝X）



4.更改の苦労話 4/5　作業効率化の取り組み 1

- 更改支援ツールによる効率化
  - 装置交換前後のshowコマンド結果取得および正常性判定、事後の差分チェックをツール化
  - 保全部門によるトラヒック確認など手作業・目視確認・結果報告など**ツール化できない作業は別途必要**
  - 作業実施部門の作業負担軽減・人為ミス防止には寄与。 作業日数削減へのインパクトは限定的

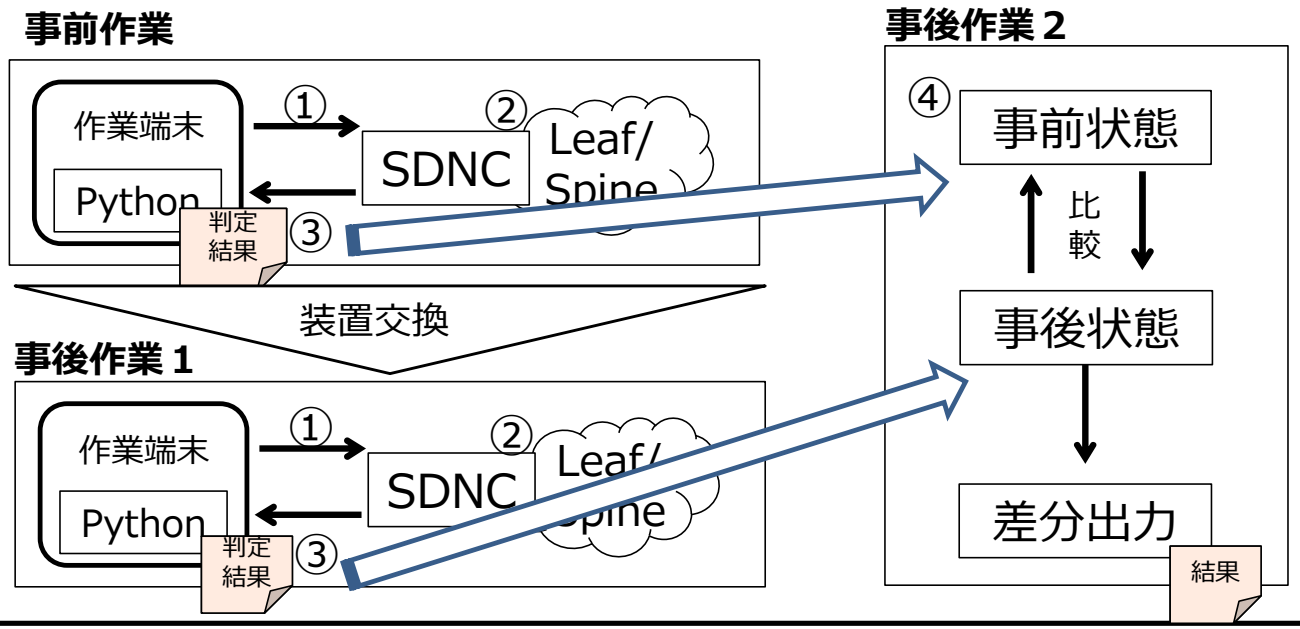
更改支援ツールのイメージ

①作業端末からSDNコントローラへ接続

②更改対象機器に対してshowコマンド実行

③showコマンド結果から正常性を判定

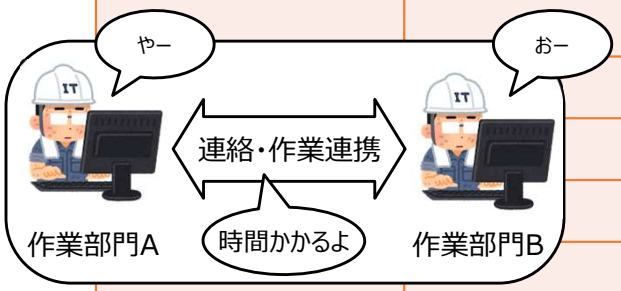
④事前事後のログを比較し差分を確認



4.更改の苦労話 4/5　作業効率化の取り組み２

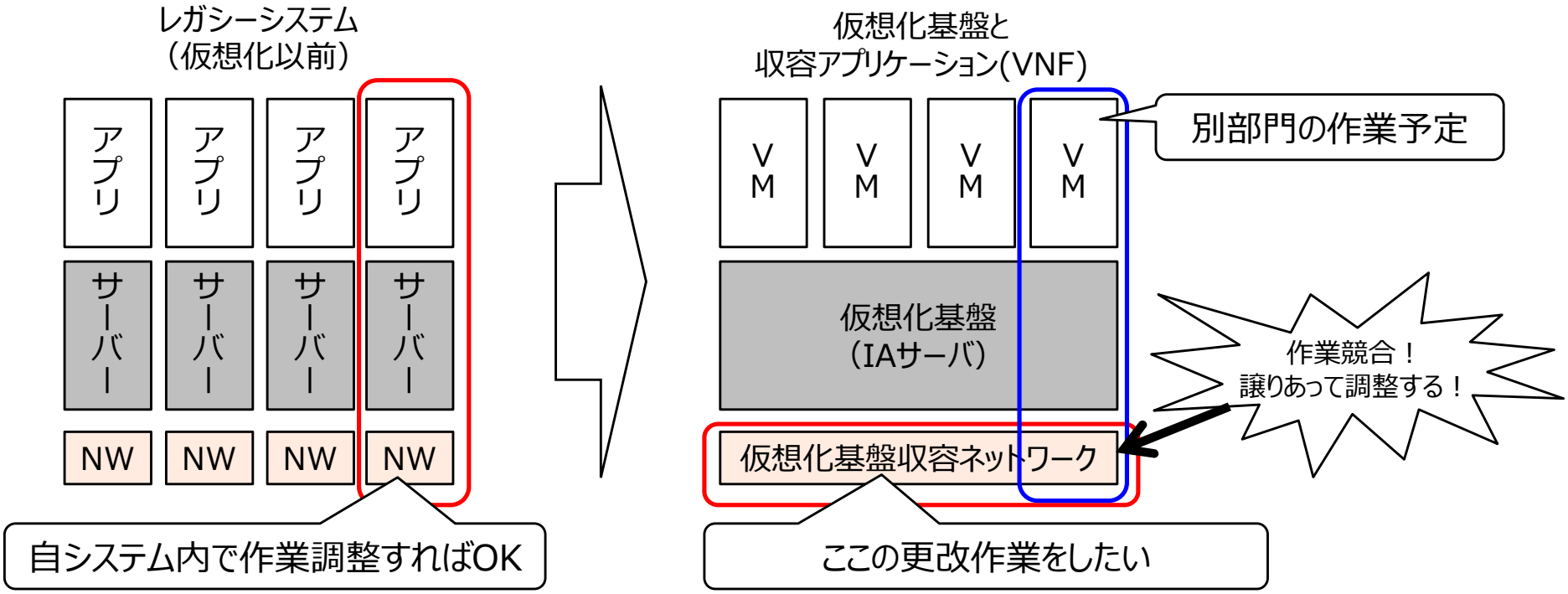
- サービス継続状態での作業のため、１つの作業Phaseに複数部門が関与する必要がある
- 部門を跨る連絡・作業連携は、簡易ツールで自動化できず、作業時間短縮のボトルネック

作業Phase	作業	作業実施部門
IAS収容Leaf HW更改 [装置交換]	作業対象ラックからのVM退避	収容VM担当
	正常性確認（事前）	保全部門、現地作業部門
	作業開始連絡	保全部門⇒監視部門
	IF閉塞	保全部門
	ケーブル抜去、電源断	現地作業部門
	装置交換	現地作業部門
	電源投入、立ち上げ、ケーブル挿入、正常性確認	現地作業部門
	交換後装置の組み込み	保全部門
	IF閉塞解除	保全部門
	正常性確認（事後）	保全部門
	作業完了連絡	保全部門⇒監視部門
	退避していたVMの戻し作業など	収容VM担当



4.更改の苦労話 5/5　他部門との作業競合１

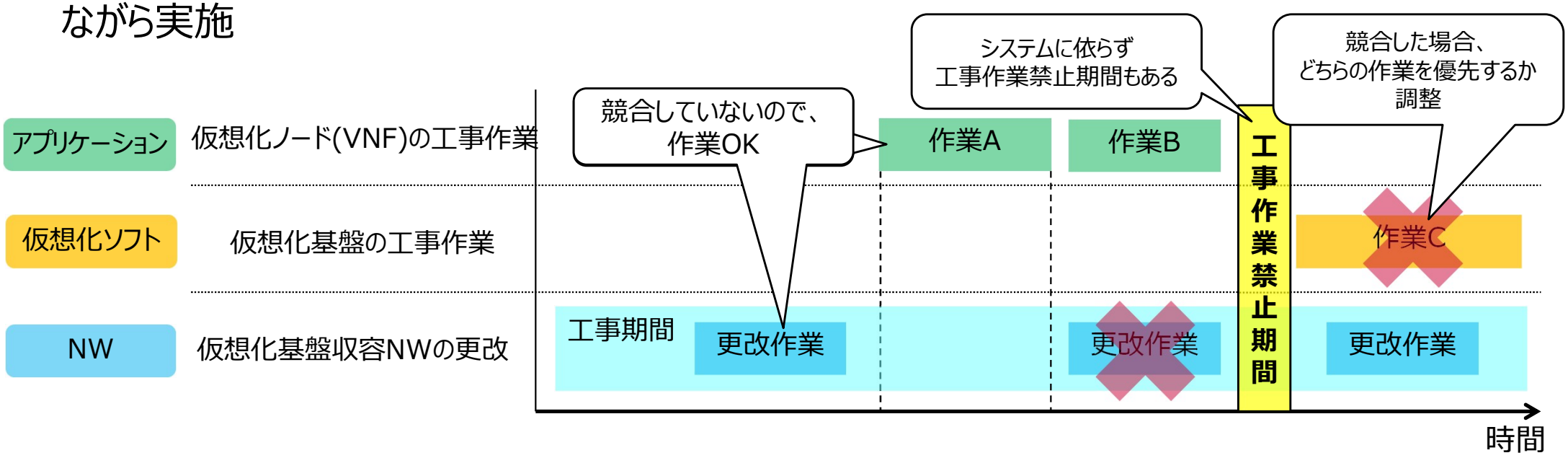
- 収容システムとの作業競合を考慮したスケジューリングが必要
- 仮想化基盤収容NWの更改作業は、実作業日数で40営業日程度が必要であり、収容システムとの作業競合が頻発する



4.更改の苦労話 5/5　他部門との作業競合２

- 他部門との作業競合を考慮した作業計画を立てる必要あり
- 選挙や大規模イベントなどの作業禁止期間も存在する
- 作業実施部門の稼働的な制約も考慮が必要。

- 実作業日数 40営業日/基盤　だが 更改期間として半年程度を見込み、スケジュール調整しながら実施



4.更改の苦労話 5/5　他部門との作業競合３

- 少しでも調整がしやすくなるように、SDN更改と同時実施可能な作業を整理
- 作業競合調整に活用中

SDN更改作業 ＼ 仮想化ノード(VNF)作業		VNF作業1	VNF作業2	VNF作業3	VNF作業4	VNF作業5	VNF作業6	基盤作業
①	SDNC SW更改	×	○	○	×	×	○	×
②-1	VM 作業対象ラックから退避	×	×	×	○	×	○	×
②-2	Leaf HW更改[装置交換]	×	×	×	×	×	○	×
③-1	外部接続機器 トラヒック迂回	×	×	×	×	×	○	×
③-2	Leaf HW更改[装置交換]	×	×	×	×	×	○	×
④	Spine SW更改	×	×	×	×	×	○	×
⑤	全体 SW更改	×	×	×	×	×	○	×
⑥	全体 SW更改	×	×	×	×	×	○	×
⑦	Flood化	×	×	×	×	×	○	×
⑧	Spine HW更改[ﾓｼﾞｭｰﾙ交換]	×	×	×	×	×	○	×
⑨	SDNC HW更改[装置交換]	×	○	○	×	×	○	×
事後	設定管理システム データ移行（OS ver.1⇒4）	×	○	○	○	○	○	○

■ 自己紹介

1. ドコモのNW仮想化基盤の概要
2. ドコモの仮想化基盤収容NWの更改における前提条件
3. 更改手順
4. 更改の苦労話
5. 総括/議論ポイント

- データセンター等における大規模なNW機器の設備更改を行う場合、どのくらい時間・期間をかけて実行しているでしょうか。

- 今回は、更改対象のNW機器が収容しているVMを事前に退避させたり、トラヒックを迂回させるなど作業工程が多く必要となりました。

- 各社みなさまの設備更改に対する計画の立て方、何を優先してどこを割り切るのかといったアプローチの仕方などの考え方を議論できればと考えています。

Special Thanks to いらすとや （いらすと16点利用）