

## 第3部

### 特集3 DIX-IE3:次世代の光ネットワーク相互接続を見据えたIX

西野 大、加藤 良輔、豊田 安信、遠峰 隆史、関谷 勇司

---



---

#### はじめに:要約 この報告書の内容

---



---

DIX-IEとは、Distributed Internet eXchange In Edoの略であり、WIDEプロジェクトのサブプロジェクトであるNetwork Service Provider Internet eXchange Project Working Group (NSPIX WG) が運用するInternet eXchange (IX)である。

この報告書では、NSPIX WGが運用する現状のDIX-IE3までの変遷を理解するために、DIX-IEのこの5年間の経緯について述べる。

序章では、NSPIXプロジェクトの1994年から30年の歴史を振り返り、この歴史的な文脈の中で述べることにする。

NSPIX WGでは、他にも大阪地域のIXであるNSPIX3を、東京と大阪の両地域のIXを接続するIXであるNXPIX23も運用を行っている。なお、本稿では、DIX-IEの変遷を中心に述べることで、これらについては多くは触れないこととする。

---



---

#### 第1章 DIX-IEの歴史～ NSPIXからDIX-IE3までの30年

---



---

##### 1.1 NSPIXからDIX-IEまで

本原稿執筆時点から遡ること約30年前の1994年にNSPIX WGはNSPIXの運用を開始した。1994年は日本の商用インターネットサービスプロバイダ (ISP) が多数誕生した時期であり、当初のNSPIXは、商用ISPが日本で展開する際のトラフィックの課題を解決するために構築された。NSPIXは、日本のインターネットトラフィック

を捌くことに成功したが、想定を超えるISPのトラフィックが集中したため、その成功の負の側面として、トラフィックが輻輳するまでの状況になってしまった。そして、この課題を解決し、より安定なインフラを構築するために、NSPIX2が構築された。NSPIX2はKDDI大手町ビルのコモンキャリアのファシリティ (コロケーションサービス) を利用することにより、1996年から運用を開始した。

だが、NSPIX2も次の課題に直面することになる。1990年代のIXにおける大きな課題は、LAN技術によって接続できる範囲の単一サイト内 (データセンター内) でしかサービスを展開できないことであった。そのため、当時の日本のインターネットインフラストラクチャは、「KDDI大手町『一局』集中」と呼ばれるような弊害に直面した。これは、建物で発生する障害に弱い構成になってしまったこともさることながら、ファシリティ事業者の競争事業者に対する中立性が課題になっていた。この課題を解決するために誕生したのがDIX-IEの頭文字であるDistributed (分散) であり、複数のデータセンターにおける分散IXサービスを2003年から運用を開始した。

DIX-IEのスイッチシステムは、当初のスイッチシステムからスイッチのアップデートを行いながら、最終的にDIX-IE3と呼ばれる新しいシステムが稼働する2023年11月まで運用された。現在は新しいシステムに完全に切り替わっている。

##### 1.2 SDNの勃興とPIX-IE

2010年の前後に、OpenFlowを中心としたSoftware Defined Networking (SDN) の技術が目されるようになった。NSPIX WGとしても、Hop-by-HopのSDN技術であるOpenFlowの利用を検討し、プログラムしたフロー

以外のものは通さないコントローラブルなIXサービスをコンセプトにProgramable Internet eXchange In Edo (PIX-IE)のプロトタイプを構築した。

2014年に完成したPIX-IEのプロトタイプは、イベントネットワークであるINTEROP 2014 ShowNetに対して試験的にサービス提供した。その後、NSPIX WGでは、より安定的なOpenFlowを利用したスイッチシステムの開発を進め、SDNコントローラとしてFAUCETを利用したUmbrella方式で、DIX-IEのコアスイッチのシステムから移行を行うことを目指した。

2021年には、PIX-IEのスイッチを当時の拠点であるKDDI、NTT-COM、BBTowerの各拠点サイトに展開し、旧DIX-IEのスイッチシステムからの移行に挑戦した。しかしながら、2021年4月と8月に実施した移行作業においては、その都度課題が生じ、PIX-IEのスイッチシステムへの移行は断念されることになった。このため、PIX-IEに完全に切り替える予定のDIX-IEのスイッチシステムは、結果として2023年11月まで継続して利用されることとなった。

### 1.3 クラウドアーキテクチャとPIX-IE++

現状のDIX-IE3と呼ばれるIXのスイッチシステムは、2019年8月にNSPIX WGの拠点運用者の一つである株式会社ブロードバンドタワー (BBTower)の加藤と西野からアイデアが提案された。これは、PIX-IEの後継という意味からPIX-IE++の名称で、NSPIX WGにおいても、PIX-IEと同時並行的に研究開発が行われた。

PIX-IE++のアイデアは、クラウド・サービスにおけるネットワークのスケラビリティを確保するために使われている技術である「Clos networkアーキテクチャ」と、「IPファブリック」、「EVPN/VXLANプロトコルを利用したL2トンネル」をIXのスイッチシステムで利用する、というものである。これらの要素は、世界最大級のIXの一つであるLINXにおいて実践された。大型のシャシー型のスイッチの更改に伴い、ピザボックス(1U)スイッチによるIPファブリックへの切り替えを目的に2018年から実践と運用されており、DIX-IEが注目した当時としては大きな新規性があるものではなかった。なお、当時は日本の

商用IXでは採用はされる予定はなかったように思われるが、2025年の現在では国内商用IXの一社であるJPIXが全面的に採用している。

PIX-IE++は、2021年のPIX-IEのDIX-IEへの導入の断念と、その後の研究・開発の難航を背景に、2022年中頃からサービス開発を加速させる必要が生じていた。

### 1.4 DIX-IE3の誕生

2022年から2023年までは、NSPIX WGにとって多難の年であった。NTT-COM大手町拠点本館の閉鎖に伴い、2022年12月末までにNTT-COM大手町拠点の移設に対応する必要があった。しかも、それまでNTT-COM大手町拠点で利用してきたスイッチを継続利用することが出来ず、暫定的な移設対応を行うことになった。また、2023年の6月に判明したKDDI大手町拠点のユーザ接続用スイッチMLX-3の保守上の課題を回避するために、急遽、MLX-3の代替スイッチの導入をすすめる必要に迫られた。このため、PIX-IE++として開発されてきたスイッチシステムの一部を切り出して、一次的なスイッチ群として構成することになった。

さらに、DIX-IE接続ユーザの一部からは、接続ポートの100Gbpsイーサネット対応のリクエストへ対応にも迫られていた。NSPIX WGでは、高度な機能を実現できるPIX-IEへの移行が最優先としてきた。しかしながら、PIX-IEとして開発してきたシステムでは、スイッチが100Gbpsイーサネットに対応することが出来ず、これに対して、PIX-IE++の接続スイッチとして100Gbpsポートが提供できる目処が立ったこともあり、PIX-IE++として開発されてきたシステムをDIX-IEの後継として導入する決断に踏み切ることになった。2023年9月のWIDE研究会でPIX-IE++のスイッチシステムへの移行についてコンセンサスが成立した。

2023年11月からは、PIX-IE++から切り出したネットワークの一部がDIX-IEのコアスイッチとして利用されるようになり、2024年6月には、完全にPIX-IE++として開発されてきたネットワークに移行することになり、100Gbpsのサービスを提供を開始した。

DIX-IE2は欠番である。「次世代のものは、WEB3になど  
 になって、バージョン3とする業界風習があるので、DIX-  
 IE3と呼ぼう」（関谷）という発想から、PIX-IE++はDIX-  
 IE3と呼ばれることになった。

## 第2章 DIX-IE3のコンセプトと設計

### 2.1 DIX-IE3の概要

DIX-IE3は、EVPN/VXLANによるスイッチ間接続をもつ  
 Clos networkトポロジのIPファブリックによって、オー  
 バーレイ型のレイヤ2のIXサービスを提供している。

2019年における提案時において、イメージ・構想はある  
 程度は存在していたが、現在の設計と全く同一ではない。  
 本来であれば、スイッチなどの全てのシステムを新規に  
 並行して構築し試験を行い、移行を実施していれば当初  
 の設計の通りに構築出来たのであろう。だが、実際にはそ  
 うはいかず、順次に移行を行う必要があったからである。

何よりも、新しく導入する技術(プロトコル)・装置(ス  
 イッチ、伝送装置)などが、想定通りの機能や性能で動作  
 するのかの確証はがく、順次、簡単な単体試験から実証  
 的に導入するしかない、という現実があった。また、さま

ざまりソース、たとえばラック、スイッチ、サーバ、階間  
 光ファイバ、サイト間光ファイバなどが限りある環境で  
 利用されており、運用環境と同時に構築する余裕がない、  
 という物理的制約があった。

### 2.2 DIX-IE3のゴールとアップグレードへの課題

DIX-IE3のゴール

- ✓ 既存の拠点間L2ファブリック構成の10Gbpsスイッチ  
 の撤廃
- ✓ 事業者へ提供するサービスの高速化(100Gbps化)

DIX-IE3のゴールを実現するためには、まずは100Gbps以  
 上に対応したスイッチへと刷新する必要がある。エンド  
 ユーザへのサービスの高速化に対応して、複数のスウィ  
 ッチの間の接続も同等かそれ以上に高速化する必要があ  
 る。そして、スイッチ間を高速化するには、サイト間の  
 回線が最もボトルネックとなる。

ボトルネックとなる主な要因はサイトとサイトを接続す  
 るダークファイバ等の回線である。これは、回線のコス  
 トが月次で発生するため導入のハードルが高い、という  
 側面もあるが、IX事業者が利用するデータセンター毎に  
 キャリアの制約がある場合がある。

## PIX-IE++ (DIX-IE3) コンセプト

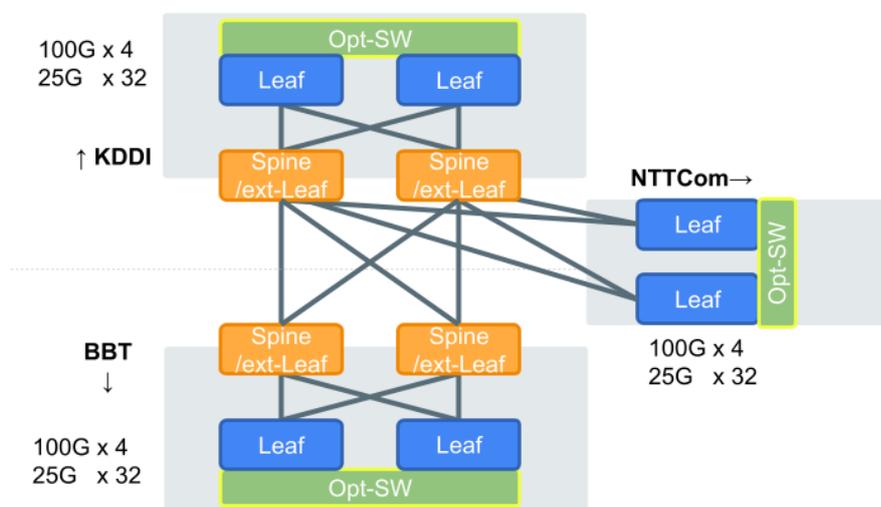


図 DIX-IE3 (PIX-IE++) コンセプト

サイト間の回線、特に、ダークファイバを潤沢に入手することは、現実には困難であるので、回線を多重化して利用する必要があった。さらに、100Gbpsの接続をを多数用いる設計の中、100Gbpsのイーサネットを多重する方法が多くはないことが技術的な課題であった。

配線システムの課題もあった。「IXを構成するリソース、重要なリソースは、パケット交換用のスイッチではなく、配線システムである」と一部のIX事業者では言われていた。WIDEプロジェクトでは、配線システムを利用するカルチャがなかったために、接続するISP事業者の配線について、まとまった記録がなかったことは課題であった。

ラックのリソースも課題であった。DIX-IEがコアサイトで利用しているラック数は合計で6ラックであったが、これはIX事業者がコアで利用するには足りない数字である。IX事業者のリモートサイトでは十分であるが、コアサイトで配線が集中することや、スイッチの世代交代と考えると厳しい数字である。何よりも、現用で利用しているラックの場合での移行はかなり厳しいものであった。

このように主に物理的な制約をどのように取り払うのが課題となった。

### 2.3 100Gbpsインターフェイス規格の選択肢

本節では、利用可能な100Gbpsの規格を概観し、DIX-IEのスイッチ間接続やISPとの接続などで選択した接続規格とその理由を述べる。

100G-CWDM4の規格は、IEEEではなくメーカー間の取り決めで作られたMSA (Multi Source Agreement)規格であり、製品が安価となることを狙って規格が制定されているため、大量導入で大幅なコストダウンが望める。単一のサイト内(単一のデータセンター内の500mを想定)では利用可能であるように規格が制定されており、光減衰のバジェットが十分であるなら選択したいところである。KDDIサイトにおいては、構内の異なるフロアに設置したIXのコアスイッチ間を接続するために利用することにした。一方、ISPなどのIX以外との接続インターフェイスとしての100G-CWDM4の利用は、IEEEではないことから、対向装置の対応も万全ではなかったことも懸念

され、使われることは少なく、IEEE規格の100GBASE-LR4が利用されることが多かった。しかしながら、近年では十分に選択肢であると判断し、DIX-IE3においては、ISPとの接続インターフェイスのラインアップに100G-CWDM4も加えている。

100GBASE-LR4の規格は、LANで利用される100Gbpsイーサネット規格であり、標準的な減衰率のSMFで10km届くように規格が制定されている。商用IX事業者においては、IXスイッチとISP事業者との標準的な接続インターフェイスとして定義されており、DIX-IE3においても、ISP事業者との標準的な接続インターフェイスの1つとして規定している。また、100GBASE-LR4は、良好な品質のダークファイバが潤沢に入手できるようなキャンパス間接続においては利用可能で選択肢である。DIX-IEのコアとなるKDDIとNTT-COM拠点の間接続では、バジェットに大きな問題がなかったため、DIX-IE3では一部の拠点間接続でも利用している。

100GBASE-ER4の規格は、接続距離が長く減衰が大きいサイト間接続であれば、選択しうる接続規格の一つである。しかしながら、100GBASE-ER4は、Oバンド(1260nm～1360nm)の光を利用しているため、通常のSMFではCバンド(1530nm～1565nm)と比較して光ファイバ中での減衰率が大きく、増幅が必要な局面では高価なOバンド用の半導体アンプ(Ramanアンプ)などが必要となる。また、100BASE-ER4のデバイスは一般的に消費電力が大きい製品が多く、利用するネットワーク装置によってはポートの電力供給不足で利用できないケースもあるので注意が必要である。そもそも、距離が長いダークファイバに対して、減衰が大きいOバンドであること、既に100Gの実現のために4多重化されて根本的に多重化できない設計の100GBASE-ER4/LR4がIXのサイト間接続が求める長距離かつ多重接続の選択肢となることは希である。

なお、100GBASE-SR4が、上記のIXの構内接続、対外接続の選択肢に入っていないのは、配線に多芯のマルチモードを要求しているからである。100GBASE-SR4はインターフェイスのコストは非常に安価であるという利点はあるものの、世代が変わり続けるマルチモードファ

イバであり、さらにMPO終端が必須のコネクタ形状であるため、データセンター内の構内配線としては採用されにくく、100GBASE-SR4に必要なファイバを用意するのはハードルが高い。100GBASE-SR4の利用は単一ラック内での利用に留めることが現実的である。DIX-IEのIXのスイッチ間接続でも複数のラックに渡る場合が多く、100BBASE-SR4を選択することは少ない。また、DIX-IEのIXのスイッチ間接続のうち単一のラック内で収まる短距離の100Gbps接続では100BASE-SR4を採用せず、AOCを採用している。これはコネクタとファイバが一体型のデバイスであり、MPOコネクタ接続のメンテナンス用品など(クリーナー、端面検査器)のロジスティックスを考慮してのことである。

LANで利用されるSMFの100Gbpsイーサネットの規格でよく使われる100GBASE-LR4、CWDM4は25Gbpsの4レーンにて構成されている。波長分散を押さえるために短距離向けのOバンドの波長のレーザーを利用し、かつ、LAN WDMで1レーン25Gbpsを1波長として、4つの波長を多重して100Gbpsを実現している。このためCバンドの波長帯による長距離かつ多重伝送を行う事は出来ない。サイト間の接続は、100Gbpsは、多重化と光の増幅(減衰の補償)が要件となってくるので、Cバンドのレーザーを利用することになる。次節ではこれについて述べる。

## 2.4 サイト間100Gbpsインターフェイス

Clos networkのトポロジを実現するためには複数の100Gbps回線が必要となる。サイト間を100Gbpsで接続しようとした場合に、長距離接続、回線の多重化を要件を満たす選択肢は少ない。

DIX-IE3 (PIX-IE++) 提案時の2019年時点において、100Gbpsイーサネットを伝送する光伝送のデバイスとしては、200Gbpsコヒーレントデバイスによるトランスポンダ型の伝送装置が利用されていた。200Gbpsのコヒーレントデバイスはイーサネットスイッチ向けに標準化が進まなかったという背景があり、また、光トランシーバのフォームファクタもCFP2であり、コヒーレントデバイス全体のコストの低廉化がすすんでいなかった。当時、100Gbps以上の長距離多重伝送を検討すると、基本的には専用の光伝送装置を購入する必要があり、システム全

体として高価になり、複数の100Gbps回線を必要とするDIX-IE3としては選択が出来なかった。

結論として、DIX-IE3のサイト間回線では、100G COLOR-Zと呼ばれるinphi社(M&Aにより、2025年現在、Marvell社)が開発・販売しているトランシーバ型の伝送装置を利用している。これは、PIX-IE++の提案時の2019年時点で利用可能であった100Gbpsを多重化するためのデバイスで安価なものが、他に選択肢が存在しなかったためである。COLOR-Zという選択肢が存在したことによってはじめて、PIX-IE++の提案に繋がった。COLOR-Zは、Cバンドのレーザーを利用し、モジュールのフォームファクターがQSFP28であるので、スイッチなどの装置が対応していれば、伝送装置のトランスポンダが不要で多重伝送が可能となる。残念ながら本プロジェクトで導入したMellanox社(現、NVIDIA社)製のスイッチではトランシーバに必要な電力を供給対応できなかったため、メディアコンバータを利用する必要があった。COLOR-Z対応メディアコンバータとして、SmartOptics社の製品を利用している。

COLOR-Zは安価であるという半面、技術的努力が必要な側面が多い。COLOR-Zでは、Cバンドの一つの100GHzグリッドの波長幅に、グリッド中心波長から約0.2nmずらした2波長を100GHzグリッド内に収め、それぞれの2波長の光を2値のNRZ (Non Return Zero)ではなく、4値を持つPAM4 (Pulse Amplitude Modulation 4)の信号を扱うことで実現している。このため、ファイバ長が長くなると、距離に比例して二系統のPAM4信号は、波長分散による到着時間の差が出るために、距離に応じた波長分散を補正する必要があり、COLOR-Zの場合、5～10kmを目安に分散補償モジュール(DCM)をDWDM回線に導入する必要がある。分散補償デバイスは、一般的には分散値が固定的で変更出来ないデバイスが多く、動的に変更する分散補償ユニットは高価である。また、分散値を測定することは非常に困難であるので、多くの場合にはファイバ長を測定することにより分散値を推定して導入する必要がある。つまり、ダークファイバの回線長をあらかじめOTDRで測定して距離を決定してからモジュールを購入する必要がある。システム導入後の微妙な調整は困難という弱点がある。

また、COLOR-Zのデバイスは、Cバンドで利用する波長が変更できない固定波長であり、運用上DWDMのシステム全体の設計の自由度が損なわれることも弱点である。COLOR-Zのモジュールは、DSPを内蔵しているために、モジュール全体の消費電力が非常に大きいことも課題であり、この結果として、対応することが出来ないスイッチが多くなってしまっている。本プロジェクトにてメディアコンバータを使用しているのもこれが原因である。また、消費電力を抑えるためなどから、光アンプでの光の出力レベルが-8dBm程度と低く、かつ、要求される光の入力レベルが0から+2dBmのため、DWDMで他の(10Gbps DWDMなどの)メディアや光アンプ挿入時・送出時に均一になるようレベル調整が必要になる。また、光アンプの利用を前提としているために、同一デバイスを直結することによる簡易試験が実施出来ないことも不便な点である。

本報告書を執筆している2025年現在、長距離伝送の選択肢は広がっている。特にデジタルコヒーレントのQSFP56-DD/OSFPトランシーバ型デバイスである400G ZR、および、400G ZR+ (以下、400G ZR/ZR+)は、2019年にプロダクトとして注目され始めて、商業的に大きく成功し、2025年の現在では安価に入手可能となりつつある。。400G ZR/ZR+は、デジタルコヒーレントによる伝送可能なDPSを搭載し、COLOR-Zで困難とされる、波長分散の補償はDSPの処理にて不要になることは大きい。また、利用しているレーザーは可変波長であり、DWDMシステムの運用上の自由度が広がることは大きい。近年ではトランシーバ内部にEDFAを内蔵して、±0dBm近辺の出力が可能なBright、High powerといった高出力の意味合いを冠した400 ZR/ZR+デバイスが存在するので、レベル調整という側面でも容易になることが期待される。

ただし、400Gbpsに対応したスイッチをDIX-IE3の構築時点である2022年では安価に入手することが出来なかったために、現在のDIX-IE3では利用していないが、将来のDIX-IE延伸の為に長距離多重伝送の際に最有力な選択肢である。

DWDMシステムの回線の導入も課題である。前述したように、DIX-IE3で利用しているCOLOR-Zは、細かい光レベル

調整や分散補正が要求される。具体的にはトラフィックテストでBERT試験をしながら、エラーが出ないように分散補償デバイスの導入や、光増幅レベルの調整する必要がある。しかし、運用中のダークファイバの回線では、このような調整はトラフィックに影響が出る事もあり非常に難しい。影響がないように構築するには、新規のダークファイバ回線が必要になる。DIX-IE3では、本格的に導入が決まる前の時点である、2021年末と2022年末に、新規のダークファイバ回線を利用して、WDM回線の導入と調整を行うことが出来た。このことは幸いであった。

## 2.5 スwitchの選択

DIX-IE3でのアーキテクチャの最大課題は、サイト間の回線の調達であるが、もちろん、導入するSwitchの選択も課題である。

DIX-IE3のClos network用Switchとしては、次の2種類のSwitchが要求された。

- Spine Switch
  - 100Gbps Ethernet(QSFP)を8ポート以上接続可能
- Leaf Switch(小規模拠点)：
  - 1/10Gbps Ethernet (SFP/SFP+) を10ポート以上 (ユーザ向け)
  - 100Gbps Ethernet(QSFP)を3ポート以上 (Spine向けアップリンク2ポート、ユーザ向けを1ポートに設定)
- Leaf Switch(大規模拠点=KDDI-COM大手町拠点)：
  - 1/10Gbps Ethernet (SFP/SFP+) を20ポート以上 (ユーザ向け)
  - 100Gbps Ethernet(QSFP)を6ポート以上 (Spine向けアップリンク2ポート、ユーザ向けに4ポートを想定)

ポート数の設計には、IXの将来にわたるキャパシティプランニングが不可欠であるが、各拠点では大きくユーザが増えず、10～20%の増加という想定を行った。また、Clos networkの特徴としてスケールアウトが可能であるため、Leaf Switchの増設を行って対応すれば無問題であろう、と想定した。

Clos networkの設計ではオーバーサブスクリプトの比率の設計が重要である。全てのユーザのポートの合計とアップリンクのポートの合計の比率が、50%を切ることもやむを得ないと考えた。まずは、100Gbpsポートのサービスの提供を開始することが重要という判断である。

スイッチのハードウェアとして、IXのトラフィック交換用のスイッチとしては、バッファが深いスイッチが望ましいとされてきた。この点については、Clos networkのアーキテクチャ導入に伴う最も大きな課題であった。しかしながら、バッファの深いスイッチは高価であり、調達に課題があった。このため、バッファが少ないスイッチを試験的な導入を行い、実証的に問題がないことを確認しつつ、進めることとした。

スイッチのソフトウェアNOSとしては、EVPN/VXLANをハードウェアに対応することを要求してはいたが、優先度としては大きくはなかった。

DIX-IE3のスイッチとして、Mellanox社製のスイッチを導入した。Mellanox社製のスイッチを選択したのは、多分にロジスティックス的な要素が大きかった。Mellanox社製のスイッチは、世界的にもクラウドサービスにおいてClos Networkとして利用されている実績があり、BBTower社のクラウドサービスで大量に利用されていた。このため、BBTower社から潤沢なスイッチの供給を受けることが出来た、というのが理由である。

供給されたスイッチのうちSpine用はMellanox SN2100であり、100Gbpsが16ポート収容可能であった。Leaf用のスイッチは、当初入手できた機材はSN2010であり、100Gbpsが4ポート(1/10/25Gbpsが18ポート)のものであった。最終的には、SN2410を入手することが出来たので、これを大規模拠点用のLeafスイッチとすることになった。Mellanox SN2410は、1/10/25Gbpsが48ポートに加えて、100Gbpsが8ポートのスイッチである。上流ポートとして、100Gbpsを2ポート利用した場合には、100Gbpsのユーザを6組織まで接続が可能である。オーバーサブスクリプトの比率を1対1とした場合にも、4組織までの接続が可能になる。スケーラビリティ的な課題は存在したとしても、なにより100Gbpsのユーザの接続

を開始することが重要という判断である。

## 2.6 光スイッチの導入

DIX-IE3では、レイヤ1の光スイッチを全IXユーザ収容向けに新規で導入することを目指した。光スイッチとは、信号を正常系のスイッチとバックアップ系のスイッチとを切り替えるための光学的な機構である。光スイッチと言う名称であるが、IP的なレイヤの動作とはまったく関係ないデバイスである。

日本のIX事業者においては、2003年にマルチフィード社の運営するJPNAPサービスで、NTT-AT社製の光スイッチが最初に導入された。現在では、同様な光スイッチがJPIX・BBIXなどでも採用され、ほぼ全ての事業者で導入されている。国外の一部のIXにおいてコア・スイッチの切替に利用している例はあるが、光スイッチが導入されている例は少ないと言われている。

商用IXにおける光スイッチの導入では、日本独自の信頼性・耐障害性の向上というクオリティ上のユーザメリットが強調される。DIX-IE3でユーザ接続に光スイッチの導入した意味は、クオリティの向上というユーザメリットよりも、運用上のメリットを想定して導入をしている。これまでのDIX-IEでは、サイト間接続の一部で回線断が発生した時の回線保護装置として光スイッチを導入していた。

DIX-IE3においては、光デバイスの故障時の自動切り替え、(駆けつけの対応が出来ない場合などの)緊急時の遠隔での切替など、24×7の体制を補完する目的が第一である。何よりも、メンテナンス時の停止時間を減少させることを目的としている。DIX-IE3では、クラウド・サービスで利用されているスイッチを導入することになった。クラウドサービスを前提とするNOSでは、アップデートを頻繁に行うことが前提となっている。このため、ユーザを接続しているスイッチの再起動を頻繁に行う想定をしなければならず、光スイッチの導入が不可欠であると考えていた。

レイヤ1の光スイッチは、2023年10月の時点で、NTT-COM大手町拠点とBBTower大手町拠点に導入されてい

た。KDDI大手町拠点で、BBTowerで導入して稼働実績があったNTT-AT社製の機材をBBTower大手町拠点から移設し導入した。これは、この後に移行が想定されていた100Gbps対応のスイッチの導入を安全に行うためには必須の要素である、との認識であった。

課題としては、光スイッチがSMF対応のものしか確保できなかったことがある。全ての接続ユーザの構内回線がSMF接続ではなく、一部のユーザのMMF接続であることが課題であった。このうち、9階へMMF配線で接続していたユーザはメディアコンバータを介してSMFへ変換して接続することで、光スイッチを導入した。ただし、5階からMMFで接続しているユーザが5組織存在したが、これらのユーザについては当メンテナンスでの光スイッチの導入は諦め、次のメンテナンスに持ち越すことにした。

長期的にはユーザを1000BASE-LX 10GBASE-LRなどのSMFでの接続へと誘導することとして、SMFに配線への移行と、光スイッチの導入へめざしている。

#### 注記:

光スイッチを導入する場合には、コアのパケットを交換するスイッチも二系統導入する必要がある。コストとしては光スイッチの機構分だけ増加することになる。それならばいっそのこと、二系統の独立したポートを加入者に提供すればスイッチを提供した方が良いのではないか、という指摘もある。これはIXのビジネスモデルによるところが大きい、と筆者は考える。多くのIXでは、IXのポートごとに課金を行っている。IXのビジネスモデルとしては、トラフィックという単位で課金を行ってはならず、ポート数による課金を行っている。このことはユーザにメリットをもたらしている。だが、トラフィックと言うリソースへの課金は、ポート数からフィードバックされている。過去、ルータが非常に高価であり、加入者もルータのインターフェイスを複数できなかった時代には、このような構成が合理的であったかもしれない。

## 2.7 配線システムの整備

IXのサービスを構成する要素において、一般には、パケッ

トを交換するスイッチなどに脚光が当たるが、最も重要なものは配線システムである。ネットワーク装置のトポロジーなどの構成を柔軟な構成変更に対応できるようにするために、配線システムは重要なインフラの役割を果たす。スイッチは長くても5～6年で更新されてしまうが、光ファイバの耐用年数は10年以上である。配線システムはより長い期間利用されるので、慎重に設計する必要がある。

DIX-IE3では構造化配線の導入を行い、配線をSMFのLCコネクタに統一し、配線をパッチパネルで集中管理することを目指した。配線についての情報は、TWO WGが利用しているnetboxを共用させていただくことにした。これは、KDDI大手町の階間ケーブルやKDDI～NTT-COMなどの回線など、NSPIX WGと両方で共用するリソースが多いことによる。さらに、TWO WGの方式に準拠して、パッチパネルでは配線の意味がわかるラグを貼付て、管理を行うことにした。いわば、これらの配線管理に関する「みえる化」である。

DIX-IEにおいては、最近の10年間では配線システムは整備されていなかった。DIX-IEの構築がはじまる前の2020年ごろ、KDDI大手町での配線システムは、SCコネクタである前提であった。もちろん、DIX-IEもSCコネクタであった。階間ケーブル、異サイトから入線ファイバ、実験協力組織との接続もSCコネクタで入線されていた。もちろん、SMFに限らずMMFもSCコネクタであった。しかしながら、装置の接続も、DIX-IEのコアスイッチは、すでに大半がLCコネクタになっていたが、LC-SCのパッチコードを利用し、SCのパッチパネルを設置し、パッチコードとしては、両端がSCコネクタであるSC-SCパッチコードを利用する、というポリシーが採用されていた。これは、2000年前後の状況としては当然の設計である。だが、さすがに2020年台においては、機材の実装密度の関係からLCコネクタによる接続が主流になり、装置においてもSCコネクタの機材をみつけるのが難しい状況である。

このために、全体をLCコネクタを利用するポリシーへ変更することにし、2023年3月のWIDE研究会のNSPIX WGのBoFにおいて、コンセンサスとなり、実行に移されることになった。

DIX-IE3では、ユーザの入線管理を確実に行うことをめざした。ユーザの入線システムを整理、つまり、できる限り決まったラックの決まったパッチパネルを経由させ、構造化廃線によりスイッチへ接続とすることにした。なお、このユーザの入線を集めるラックのことをTWO WGでは「ランディングラック」と呼んでおり、DIX-IEでもこれにならうことにする。

これまでのDIX-IEのユーザの回線の多くは、都度都度の配線が行われており、配線もラベリングが行われていないものも多く、さらには、J-Jなどでジョイントされているケースなどもあり、構築担当者以外が配線を理解することは困難な状況であった。また、構造化廃線もなく、配線に関するデータベースも不在だったために、その配線システムの調査が都度必要になっていた。

IXでの理想的な配線システム状況としては、データセンターと協力連携することで、事前に配線をバルクで構成することが望ましい。ユーザの接続が決まった時点でパッチパネルの番号を指定するだけで、プロビジョニングを実施することができる。

## 2.8 マネージメント・システム

IXにおいては、コアのスイッチなどのアーキテクチャに目が行くが、それを管理するためのマネージメント・システムの存在も重要である。IXサービスを正常に運用し続け、その状況をモニタリングするためには、次のような機能が必要である。

- IXのトラフィック測定・可視化
- IXの正常性監視・障害通知
- マネージメント・システム自身の正常性監視・障害通知
- マネージメント・ネットワーク自身の正常性監視・障害通知

具体的には、cactiのようなSNMPによるトラフィック測定、zabbix / xymon / muninなどのシステム・モニタリングや監視や通知を構築している。障害通知には、いまのところslackを利用している。日本のIX事業者においては、運用上のトータルのトラフィックを公開する慣例がある。現在、DIX-IEではトータルトラフィックの公開行っ

ていないが、これはポリシーではなく手順の問題だけであるので、今後は公開できるように整備を進めていきたい。

DIX-IE3では、扱っている機材が非常に多くなるため、機材の管理用のIDやパスワードの認証・認可の仕組みも必要になる。メンテナンスするメンバの加入と離脱を想定すれば、そのたびに共通パスワードの変更を行うのは、トラブルを惹起することになり得る。共有パスワードは、セキュリティの観点においても問題であり、管理者ごとに個別のIDとパスワードを利用するような管理を行うべきであり、DIX-IE3ではできる限り共有パスワードは使用しない管理方法を実現している。

DIX-IE3においては、オープンソースでのID管理サービスの389dsを利用し、LDAPで集中管理を実施している。ネットワーク関係の機材では、LDAPの情報をベースにradiusプロトコルによるUnix系のOSでは、各マシンがsssd経由でLDAPの情報を参照している。また、sshの公開鍵の管理もLDAPで集中管理しており、LDAPのポータルであるLDAP managerから登録することで、全ての管理リソースにアクセスすることが可能になっている。

また、前記のようにシステムが多数になるため、多くのサーバ(インスタンス)が必要になるため、Proxmox VEによる仮想化システムを導入してシステムを仮想化している。

これまでの反省から、1台のサーバの上に複数の機能を実装するようなモノリシックな構成とするのではなく、サーバの機能ごとに分離した仮想インスタンスを構築することにした。

## 2.9 マネージメント・ネットワーク

マネージメント・サービスのシステムが存在するのであれば、当然ながら、マネージメント・システムのネットワークが必要である。IXのスイッチシステムは複数の拠点間を結ぶ必要がある。

一般に、管理ネットワークとしては、大きく分類すると、実際のトラフィックが流れている「インバウンド・ネットワーク」で行う手法と、そうでない専用のネットワーク

で行う「アウトバンド・ネットワーク」で行う手法がある。DIX-IEでは、従来からアウトバンド・ネットワークで実施する手法が採られてきた。DIX-IE3では、DIX-IEを踏襲しアウトバンド・ネットワークでの管理を行っている。

また、管理ネットワークとしては、拠点ごとにサブネットを分割することも選択肢ではあるが、従来からのDIX-IEの管理ネットワークの構成がフラットなレイヤ2のネットワークであったので、これも現状を踏襲した。

無論、いずれもクリーンスレートから構築していれば異なるデザインもあり得るが、これらも既存のDIX-IEの管理ネットワークも、現在のDIX-IEネットワークも合理的ではあり、移行時のトラブルを最低限しシームレスな移行を実現するために、管理ネットワークのアーキテクチャはそのまま採用することにし、現状を踏襲した。

DIX-IE3の管理ネットワークが、DIX-IEの管理ネットワークと異なるのは、拠点間の回線に対する考え方と、冗長性に対する考え方である。

#### ● 拠点間の回線

- DIX-IE: コモンキャリアの専用回線による
- DIX-IE3: DF上の自前のDWDMシステムにて構築

#### ● 冗長化に対する考え方

- DIX-IE: DIX-IE/PIX-IE側では、あえて冗長化は行わず、事業者の信頼性に頼る
- DIX-IE3: RSTPにより、自前での

DIX-IEのマネージメントネットワークの拠点間の回線としては、商用イーサネット回線を利用していた。DIX-IEの運用・監視の目的のためには拠点間を接続するネットワークが必要であり、かつ、DIX-IEから移行予定のPIX-IEにおいてはトラフィックを流すデータプレーンとは別にコントロールプレーンが必要なアーキテクチャであったために、信頼性が高いネットワークが必要であったことが、商用イーサネット回線を利用していた主な理由である。これは、DIX-IEからPIX-IEへの移行を見据えたものでもあった。

DIX-IE3でも、DIX-IEと同様に運用・監視の目的のために、マネージメントネットワークが必要である。ただし、DIX-IE3においては、データ交換のために構築したDWDM網の有効性・実効性を確認するためにも、DWDM網から管理用に回線を払い出し、WIDEプロジェクトでネットワークを構築することを想定していた。しかしながら、ここで構築したDWDM網は、シンプルな構成を旨としたDWDM網であるので、商用ネットワークに導入されているような高度なプロテクションなどの機能はもっていない。このため、拠点間を異方路かつ複数回線を利用し、冗長性を確保するプロトコルとを組み合わせることにより、全体で信頼性を確保する設計である。

DIX-IE3の管理ネットワークが、DIX-IEと大きくことなるのは、RSTPを利用していることであろう。これは、DIX-IEがコアのネットワークでRSTPを利用していたので、管理ネットワークでRSTPを利用することができなかった、という点が多い。DIX-IE3のコアネットワークが冗長化の手法としてEVPN/VXLANを導入しRSTPフリーとなったことで、管理ネットワークにおいてRSTPを利用することが出来るようになった。

DIX-IE3の管理ネットワークは、複数のVLANを重畳して利用している。VLANごとのSTPを動作させるような、MSTPやPVSTPを利用する、という選択肢もあり得た。しかしながら、DIX-IE3のマネージメントネットワークにおいては、これらのプロトコルは利用していない。愚直に全VLANを通し、スイッチ間のリンクは、RSTPによりポートごとの冗長性を担保するポリシーである。これは、VLANの数が多くはない状況であるので、スイッチの相互接続性を重視し、多くの安価なレイヤ2スイッチを選択できるようにするためである。

---

## 第3章 DIX-IE3への移行

---

### 3.1 DIX-IEからDIX-IE3まで

前章では、DIX-IE3の設計と実装に至る背景を記述した。本章では、DIX-IEからDIX-IE3へいたるまでの時間軸での変遷について報告する。

最終的なDIX-IE3のアーキテクチャと構成の変遷だけを見ても、その理由を理解できないことも多いと思われる。本報告では、「なぜ、このような順番で、このような構成変更を行ったのか」という「理由」を明確に記録しておくことが目的の一つだからである。また、これまでのDIX-IEプロジェクトと比較すると、かなり速い時間的なスピードで構成が変更されており、年次単位の報告だけでは不適當である、との判断である。

### 3.2 提案からPoCの構築へ(2019年8月～2022年3月)

前述のように、PIX-IE++の基本的なアイデアは2019年8月に提案された。2021年1月に開催されたJANOG 47において、ダークファイバの活用という発表の中で、PIX-IE++のコンセプトについて加藤が紹介を行った。この発表が、PIX-IE++についてのWIDE外部での初出となる。

提案の翌年2020年からPIX-IE++のPoCとして、EVPX/VXLANのプロトコルの動作検証を開始した。当時は、KDDI拠点でのみPoCを行っていた。

#### ● KDDI拠点

- Leaf : FS.Com N8500-48B6C (NOS:Cumulus Linux)
- Spine: DELL S5212F-ON(NOS: Cumulus Linux)

他の拠点でのPoCは、2023年3月にBBTower拠点での開始を待つ必要があった。BBTower拠点開始時の構成は次のとおりである。

#### ● BBTower拠点

- Leaf :Mellanox SN2010 (NOS: Cumulus Linux)
- Spine:Mellanox SN2100 (NOS: Cumulus Linux)

CLOS Networkのデザイン上の課題は、サイト間で複数の100Gbpsの回線が必要なことであり、BBTower拠点の開設もこの課題が解決されることを待つ必要があった。また、予算の都合上、回線の調達には年単位の時間を要することになる。

### 3.3 DWDM回線の準備(2021年5月～2022年1月)

CLOS Networkの導入には複数かつ大量の回線が必要のため、複数の100Gbps回線のリンクを構築するためには、ダークファイバを入手し、DWDMシステムを設置する必要がある。しかしながら、サイト間のダークファイバは高価であり、簡単に入手できない。

2021年5月に、BBTower社のDIX-IE大手町拠点をNTTデータ大手町(以下では、「旧センター」と呼ぶ)から新設された新大手町センター(以下では、「新センター」と呼ぶ)へ移設する計画が持ち上がったことが、この課題のブレークのきっかけとなった。旧センターから新センターへ移行する際に、並行して回線を敷設するための予算的な手配が可能になったのである。

回線の切替までに2～3ヶ月の同時並行できる期間も確保が可能になったため、新規のダークファイバを利用してDWDMのシステムを構築することが可能になった。2021年11月から新規の回線が利用開始、2022年1月末に旧来の回線の利用停止、というスケジュールで作業を実施した。

新規での回線調達であったので、これまでとは異なり、異経路で異キャリアでの冗長回線を調達することをねらった。のBBTower大手町拠点からNTT-COM大手町拠点への2回線(DF4芯)を調達していたものを切り替えることになった。

この時のDWDM回線への要求仕様は次のとおりである。

#### ● DWDM回線への要求

- IXサイト間リンク(旧来用) : 10Gbps × 2回線
- IXサイト間リンク(新規用) : 100Gbps × 2回線
- その他の利用:可能であれば、1Gbps、10Gbps、100Gbpsを増設可能。
- 機材:加えて、エイリアン的な装置構成であること。

2021年11月29日、BBTower旧拠点から新拠点への移設を完了した。この次期は、新型コロナウイルスの蔓延期であり、外出が制限され、スタッフの稼働時間を十分にとれない中での作業で困難を伴うものであった。

### 3.4 NTT-COM拠点の移設(2021年12月～2023年1月)

前節で述べた回線移転計画開始と同時期である2021年12月、当時のNTT-COM大手町拠点(以下では「本館」と呼ぶ)の閉鎖の報が伝えられた。2023年3月末に本館のコロケーションサービスが廃止され、その数年後に取り壊しが行われる、ということであった。WIDEプロジェクトとしても利用していたNTT-COM大手町拠点は、その拠点

の移転先として、隣接するビルであるNTT-COM大手町拠点別館(以下では、「別館」と呼ぶ)と決まった。この拠点移転の対応のために多くの人的なリソースが費やされることになったため、DIX-IE/PIX-IE/PIX-IE++のプロジェクトとしても、2022年は大きな進捗のない1年となった。この拠点移転のプロジェクトは、WIDEバックボーンを

## ファイバ接続: ~2021年10月

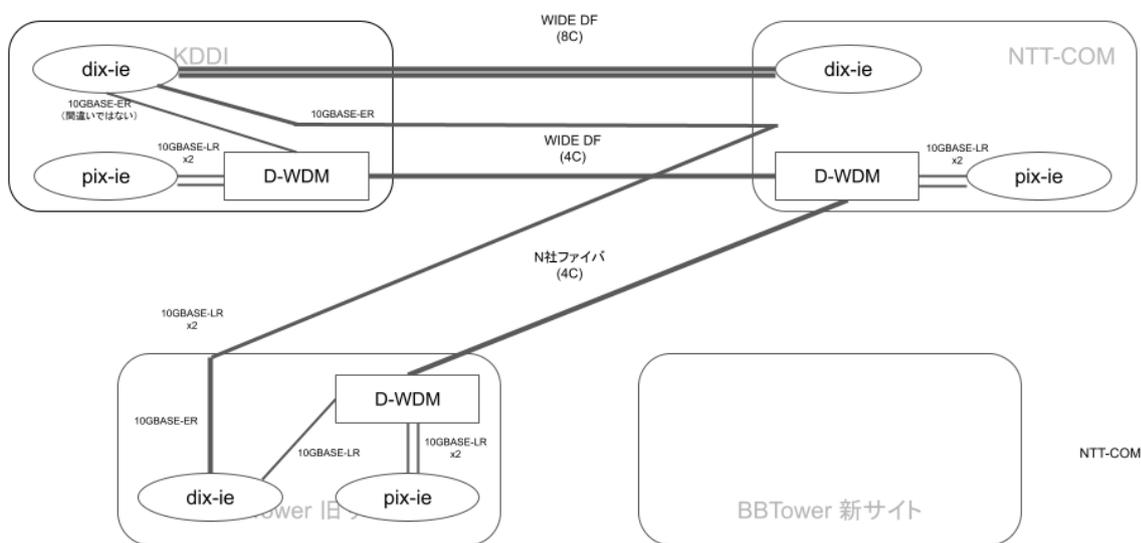


図 2021年10月

## ファイバ接続: 2022年01月～

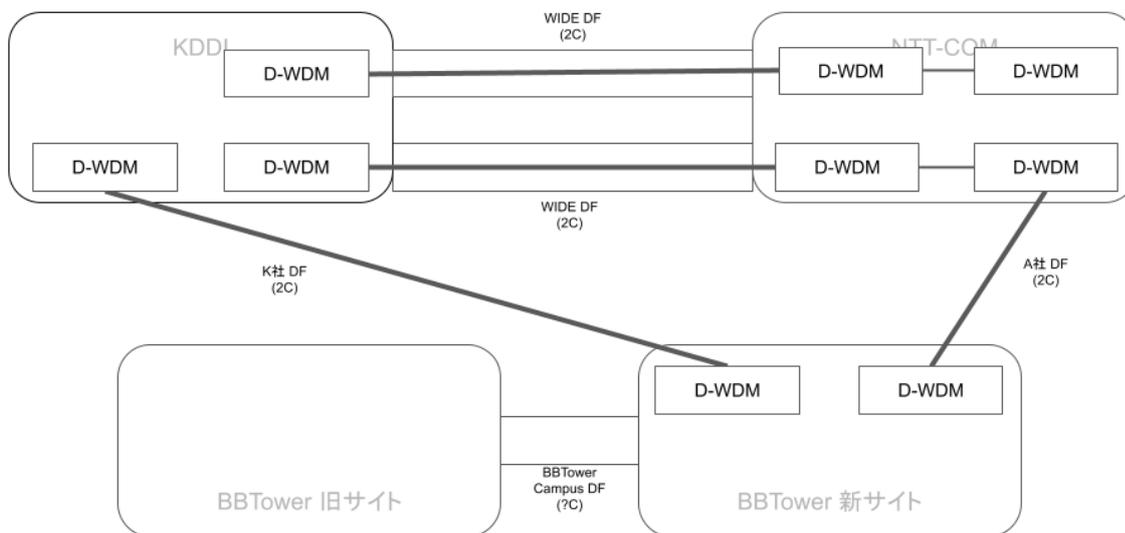


図 2022年01月～

運用するTWO WGにおいても非常に困難なプロジェクトであったが、NSPIXPプロジェクト(DIX-IE)としても、困難なプロジェクトであった。

ネットワークやシステムなどの拠点移設の一般的な方法としては、停止時間を最小限にする場合、新拠点にネットワークを構築し、新旧の両拠点を並行稼働させつつ、個別のシステムや接続先の回線などを移転を順次実施していくものが多いと思われる。しかしながら、並行稼働するだけの新規の機材を調達できないため、正副系の両系のネットワークのうち、片系のネットワークを先に移転して実施することになる。この同時稼働リソースに制限が大きかったであったこともさることながら、ネットワークの機材や回線などのリソースがどれだけあるのか？という全貌が不明であり、長年に亘るWIDEバックボーン、NSPIXPプロジェクト全体のリバースエンジニアリングを必要としたことが、困難を究める要因となった。

2021年10月から2022年8月にかけて、TWO WGでは拠点移設計画が作成され、移転の準備が進められていたが、NSPIXPプロジェクトでは拠点移設の計画立案は遅れていた。2022年7月7日に、DIX-IEの接続ユーザに対して拠点移転に関するアナウンスが実施された。しかしながら、この時点ではDIX-IEの拠点移転の方法やスケジュールなどの詳細は計画されていなかった。NSPIXPプロジェクトでは計画立案が遅れてしまったことはプロジェクトとしての大きな反省点である。TWO WGでは、2021年から1年にわたって移転方法について議論を積み重ねてきており、2022年9月からは実際に移転作業を開始していたことと比較するとなおさらである。

2022年の10月にDIX-IEのNTT-COM大手町拠点の移転計画が策定された。旧来の拠点スイッチと光スイッチを流用し、片系ずつ移行することを想定した計画であった。しかしながら、2022年の11月にNTT-COM大手町の旧拠点ユーザ接続用スイッチと光スイッチを移転に伴い継続利用できないことが判明し、計画を全面的に見直す必要が発生した。2022年12月6日に、NSPIXP WG関係者の緊急会議が開催され、次の方針が決定した。

#### ● DIX-IEの移転方法の概要

- 旧来の拠点スイッチを利用することが前提だったため、スイッチの購入・調達が間に合わない。このためコンティジェンシープランを発動する。
- DWDM回線によって多重化した接続を使い、NTT-COM大手町拠点の接続ユーザを他のDIX-IE拠点で収容する、というコンティジェンシープランを採用。
- 接続先の拠点として、DIX-IEのコアスイッチを設置しているKDDI大手町を想定。

しかしながら、KDDI大手町拠点においてスイッチのポート数の不足や、コアスイッチの構成変更が間に合わないことや不明な点が多いことなどの問題が浮上した。2022年12月23日の関係者の打合せの結果、DWDM回線での接続先はBBTower大手町拠点に変更された。BBTower大手町拠点は、NTT-COM大手町拠点と同程度のユーザを収容が可能なように、そもそもリソースが整備されていたことが功を奏したことになる。

2022年12月の年末から2023年1月にかけて、BBTower大手町拠点からNTT-COM別館へのDWDM回線を再整備し、ユーザ接続の準備を行った。2023年1月18日から20日にかけて、接続ユーザの変更作業が行われ、DWDM回線によりBBTowerに接続が実施された。当初予定よりも遅れること約1ヶ月であったが、2023年1月20日に、DIX-IEのNTT-COM大手町から移設が完了した。

このコンティジェンシープランの採用により、BBTower拠点の小型スイッチであるS4048-ONとDWDM回線について、DIX-IEの総トラフィックの半分が流れることになったが、サービスに影響なく正常に動作した。期せずして、1Uの小型スイッチにIXコアスイッチ相当として運用する実績を積むことが出来たことは、後のPIX-IE++にとって幸いであった。

#### 3.5 KDDI拠点のクリーンアップ(2023年2月～5月)

NTT-COM拠点の移設が完了した後は、DIX-IEにおいてクリーンアップ作業が必要であった。NTT-COM拠点の移設が非常にタイトなスケジュールでの計画の策定と実行であったために、いくつもの課題が残されていた。多

くがNTT-COM拠点ではなく、KDDI拠点におけるクリーンアップ作業であり、2023年2月から5月にかけては、この課題に取り組むことになった。

問題点の一つとしては、BBTower大手町拠点のスイッチがSPoFになっていた課題の解消がある。

それまで、DIX-IEの各拠点のスイッチはKDDI大手町拠

点に設置された正副二系統のコアスイッチMLX-1とMLX-2に接続していた。KDDI大手町拠点も例外ではなく、KDDI大手町拠点でユーザが接続するスイッチであるMLX-3は、MLX-1とMLX-2に接続して、RSTPにより冗長化を図っていた。

NTT-COM拠点の移設前、BBTower大手町拠点にある正系と副系の両者のスイッチは、KDDI大手町拠点のコア

### 2023年1月時点

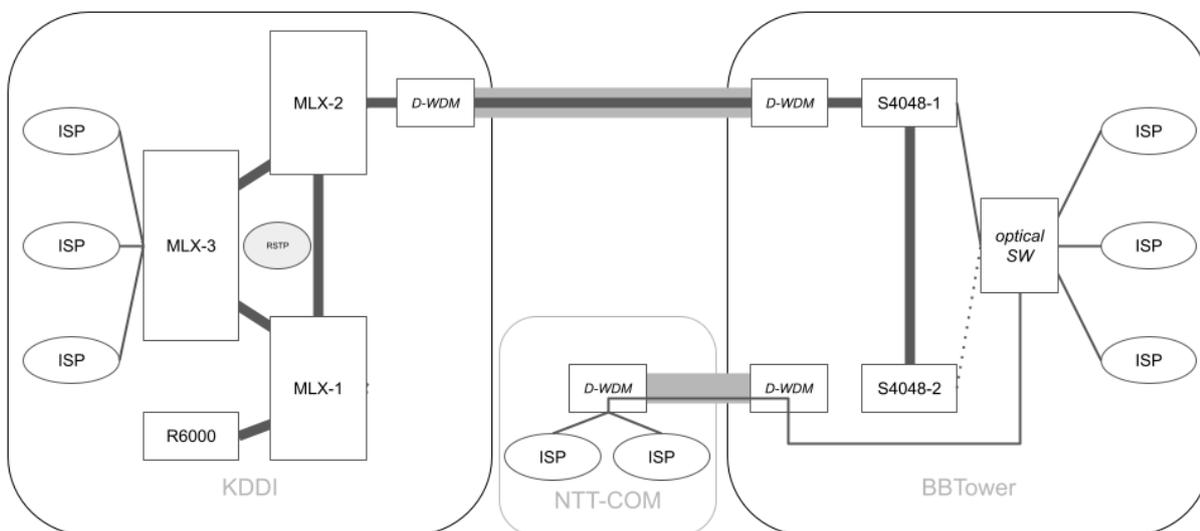


図 冗長化前

### 2023年5月12日時点

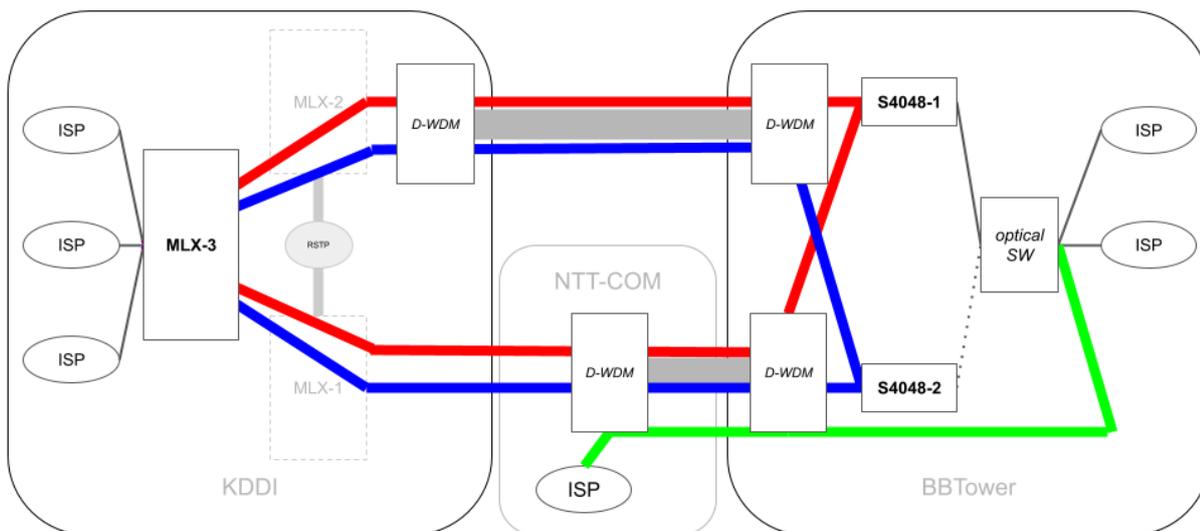


図 冗長化後

スイッチであるMLX-1とMLX-2にそれぞれ接続されていた。このうち一方は、NTT-COM大手町本館を経由することで異経路冗長の構成をとっていた。本館の移転により、シングル構成になっていた点の解消である。

また、NTT-COM拠点の移転の結果、拠点スイッチはBBTower拠点のスイッチだけになった。このため、BBTowerとKDDI拠点のスイッチを直結の方がシンプルであり、DIX-IEのコアスイッチ・MLX-1とMLX-2が事実上は不要になっていた。コアスイッチのMLX-1とMLX-2 (MLXe-8)を退役させることで、ネットワークの見通しをシンプルにして、今後の作業に備えるという意味のほかには、今後のPIX-IE++のデプロイにおいて必要と被われるラックスペースを捻出する意図もあった。

さらに、KDDI大手町拠点の顧客接続スイッチであるMLX-3には、すでに使われていないユーザの配線が多数残置されており、これが見通しを悪くしていたので、撤去を実施した。

KDDI大手町拠点では、9階にコアスイッチ、5階のリーフスイッチが稼働しているため、配線上の課題も存在した。過去、KDDI大手町拠点では旧来のNSPIXPのスイッチは5階のコロケーションに設置されていたため、多くのユーザが5階のスイッチまで構内配線を調達し接続していた。DIX-IEのコアスイッチとユーザ接続用のスイッチを9階のコロケーションに移設し、1000BASE-SX・1000BASE-LX・10GBASE-LRで接続していたユーザは、WIDEプロジェクトが提供する構内配線を利用して、9階のスイッチへ接続した。しかしながら、大手町5階の同一フロア内で100BASE-TXや1000BASE-Tなどのメディアで接続している一部のユーザのために、スイッチ(CES-1)を設置して対応していた。

過去の経緯から、このスイッチへの配線がJ-Jを利用して延長されており、一部のユーザではリンクが不安定であるなどの課題もあった。また、9階のMLX-3と同様に利用していない配線が多数残置されていたため、見通しも悪くトラブルシュートの妨げにもなっていた。2023年3月14日のメンテナンスにおいて、これらの整理を実施し、スイッチやメディアコンバータを一つのラックに集

約し、同時にパッチパネルを導入して配線管理を行えるような仕組みを導入した。

KDDI大手町拠点とBBTower大手町拠点を結ぶDWDMシステムの強化も行った。2023年4月17日から5月8日までという3週間におよぶ長期間のメンテナンスであったが、10Gbpsと100Gbpsの収容数を増強するための処置であった。当初、利用していたDWDMの光アンプでは固定出力が小さかったため、COLOR-Zのポート数を増やした際に必要な出力を得ることができなかったのである。両拠点を接続するDWDMのシステムを一次的に停止し、光アンプの交換と調整を実施した。この間、シングル構成となり、NTT-COM大手町拠点とBBTower大手町拠点の回線のみはSPoFとなってしまったが、やむを得ない判断であった。

### 3.6 DIX-IEコアスイッチの障害と回避策の決定(2023年5月～9月)

2023年の5月時点においては、PIX-IE++は、SDNのIXシステムであるPIX-IEの後継であり、DIX-IEの後継はPIX-IEである、という位置づけであった。もちろん、100Gbpsのポートサービスを実現するために、PIX-IEの後継として意欲的に実装をすすめていた。だが、あくまでも後継であり、2024年1月での試験サービスの導入を想定していた。

ところが、2023年6月に状況が変わった。KDDI大手町拠点のユーザ接続用のスイッチMLX-3 (MLX16)で障害が発生し、特定のスロットであるが、10～20%程度のパケットの破棄が発生した。スロット単体の故障かスイッチファブリックの故障か特定出来なかったが、すでに保守契約がなく、予備機や検証機もない状態であり、対応が困難であった。回避策として特定のスロットを使用不可にする以外の対応はできなかった。また、コアスイッチとして利用していたMLX-1とMLX-2の同型のコントロールボードでも故障が多発していた(4枚のコントローラ中で2枚で、起動用のフラッシュメモリを認識できない故障が発生していた)。このため、MLX-3を再起動することできない、という問題も内在していたため、速やかな対応が必要であった。

100Gbps Ethernetのポートサービスの提供について、ユーザから望む強いリクエストがあった。PIX-IEとして準備していたスイッチであるアライドテレシス社製x950では、100Gbpsのポートは2ポートしかなく、100Gbps Ethernetのポートサービスの対応が困難であった。PIX-IE++として用意していたLeafスイッチのSN2010では、100Gbpsポートは2ポートであり、各拠点での収容数は2社が限度であった。だが、PIXIE++では、新たにLeafスイッチとして、Mellanox社製SN2410の確保が可能になったために状況は好転した。SN2410は、アップリンクの100Gbpsが8ポートある10G/25Gbpsのスイッチであった。Spine-Leaf間でのオーバーサブスクリプションを許容すれば、100Gbpsのユーザを6社までの収容が可能になった。IXのシステムとしては心許ないが、長期的なアップデートを想定すれば、PIX-IE++でも100Gbpsポートを提供可能である、という判断に至った。

2023年の9月のWIDE研究会において、DIX-IEのコアスイッチのPIX-IE++への移行がコンセンサスとなった。しかしながら、PIX-IE++へ全面的な移行を行うには時間を要することが想定されていたために、MLX-3の障害に対応するための暫定的で迅速な対応を短期的には必要であった。このため、KDDI大手町拠点のユーザ接続用スイッチを暫定スイッチへと交換を主目的とするメンテナンスを実行することになった。

DIX-IEのトラフィックを運ぶKDDI大手町拠点のユーザ接続用の暫定スイッチをどの機種にするかは悩ましい問題であった。昨年のNTT-COM拠点の移設後からのBBTower拠点でDIX-IEのトラフィックの半分を担っていた実績から、DELL社製S4048-ONを利用することになった。しかしながら、BBTower拠点で動作実績がある機材の捻出は難しく、別のS4048-ONの機体を利用することになった。ただし、この機体ではライセンスの関係で、BBTower拠点で実績が出来たDELL社製のNOSが搭載された機体ではなく、Cumulus Linuxを搭載した機体であった。ただ、スイッチエンジンとしてのASICは同様であることなどや、BBTower社内別プロジェクトの検証環境でのCumulus Linuxの実績を鑑みて利用することにした。

また、NTT-COM大手町拠点にユーザ接続用のスイッチ

を導入することにしたNTT-COM大手町の移設時から1年近くDWDM回線にユーザ接続を依存する状態が続いており、この問題も同時に解消し正常化したい、という動機であった。

形としては、PIX-IE++としてPoCを実施してきたシステムから、KDDI大手町拠点とNTT-COM拠点を切り出し、運用系を構成することで準備することにした。

万が一のスイッチの障害を想定すると、メンテナンスは一刻も早く移行しなければならなかったが、準備期間と告知期間としては1ヶ月をみる必要があること、ユーザ影響を想定すると年末年始は大規模メンテナンスを避けるべきであろうこと、などから最終的に2023年11月に大規模メンテナンスを実施することを決定した。ただし、段階を追って実施する必要があることから、メンテナンス作業を3度にわけ、11月に2週間ごとに3回実施した。

### 3.7 大規模メンテナンスの準備とEVPN/VXLANの導入 (2023年9月～10月)

2023年9月のWIDE研究会でのコンセンサスを得た計画を実行に移すためにメンテナンスの準備を開始すると、想定外の課題や変更点が判明した。

最大の誤算は、NTT-COM拠点用に準備していたスイッチの動作に問題が生じたことである。当初想定では、NTT-COM拠点用にはDELL社製S5212F-ONの導入を予定していた。S5212F-ONは、1Uに2台を並べて設置することができる非常にコンパクトなスイッチであり、NTT-COM大手町拠点のようなスペースが狭いところでは(当DIX-IEプロジェクトには1ラック分しかスペースの割当がなく、伝送装置などでラックの半分以上が専有されていたので)有用であると思われていた。このS5212F-ONでは、NOSとしてCumulus Linuxを利用する予定であったが、1GbpsのイーサネットのSFPを認識することができない不具合があることが試験期間中に判明した。それまでのPoC試験では、10Gbps接続以上でのみ試験を行って居たために気がつかなかったと思われる。この不具合の回避するために労力を割く考えもあったが、時間的に浪費する可能性、および、最終的に不具合を回避できない可能性を鑑み、早期に導入を諦めた。対応策として、

同様に(1Uで2台のスイッチを搭載可能な)コンパクトなスイッチであるMellanox社製SN2010を導入することにした。SN2010は、前記のS5212F-ONと同様なCumulus Linuxを採用しており、将来のBBTower拠点のLeafスイッチ用にと、試験を行ってきた実績があり、1Gbpsイーサネットのバグはないことが判っていたので、役割を交換

することとなった。

拠点間の接続には、2回線の10GbpsをIEEE802.3adでLAG化した20Gbpsのリンクで利用した。DIX-IEの総トラフィックから十分であると判断した。また、当初は、拠点間の接続はsimpleなレイヤ2での接続を行うことに

## 2023年10月

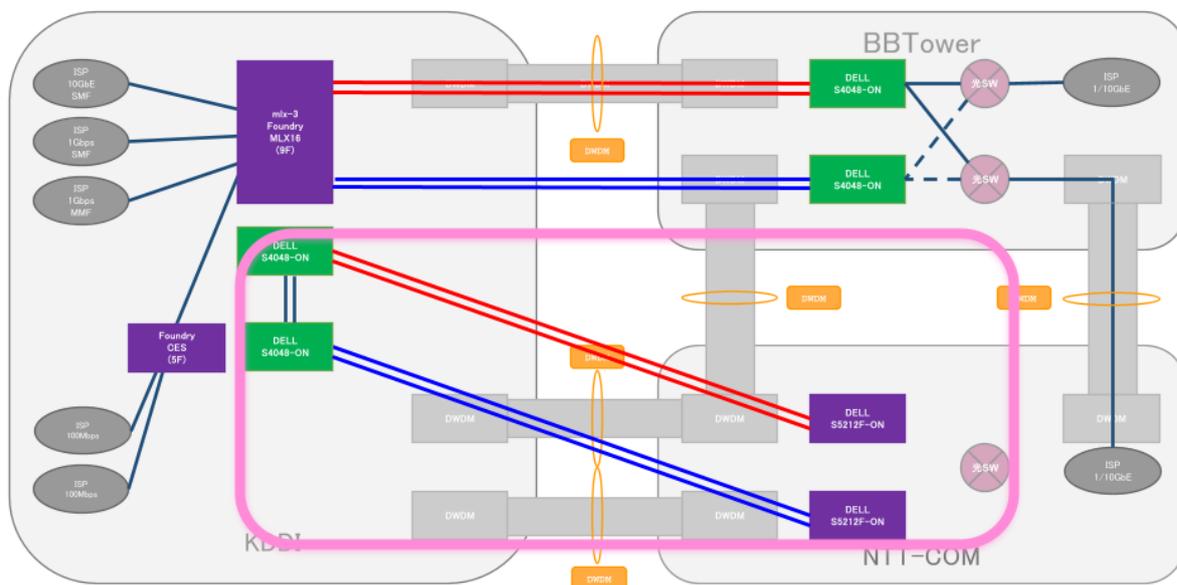


図 2023年10月

## 2023年11月1日

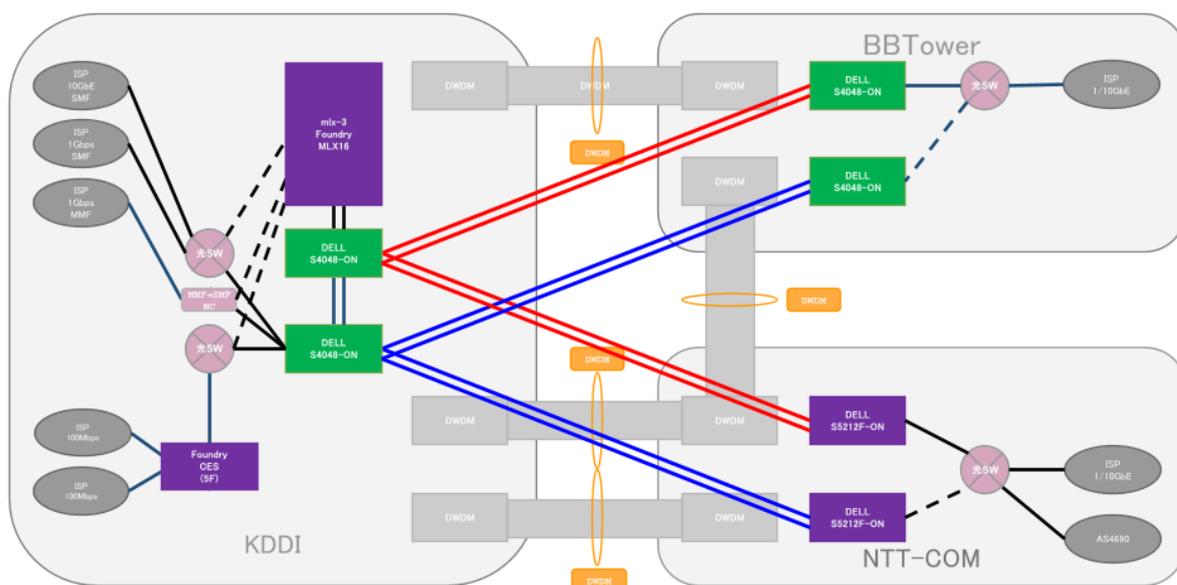


図 2023年11月

していたが、KDDI大手町拠点とNTT-COM大手町拠点とCumulus Linuxを導入したスイッチ群になったことから、EVPN/VXLANによるL2トンネルによる接続を行うことにした。sshログイン可能なプローブと呼ぶ試験端末を各拠点ごとに設置し、十分な試験時間をとることが出来たので、この点では問題がないという判断であった。

また同時に、本メンテナンスにおいては、NTT-COM大手町拠点とBBTower大手町拠点のユーザを移行する際には、レイヤ1の光スイッチを導入することにした。レイヤ1の光スイッチとしては、FS.com社のFMTシリーズのシャーシ対応の光回線保護用のスイッチ(OLP)を利用し、1Uに4ユーザの収容が可能であった。これまで利用していたNTT-AT社製の光スイッチは、1Uに16ユーザを収容することが出来たため、ラックの1Uあたりの収容効率を比較すると劣ってはいるが、ユーザが多くないNTT-COM大手町拠点とBBTower大手町拠点では、コストパフォーマンスの観点から選択した。

2023年11月1日に実施した第1回のメンテナンスでは、NTT-COM大手町拠点に接続するユーザを新規スイッチに移行することに主眼をおいて実施した。

想定外のトラブルとしては、1Gbps Ethernet(1000BASE-LX)接続のユーザにおいて、新規のスイッチであるSN2010に接続先を切り替えるとリンクが上がらない、という問題が発生した。本来ならば顧客と連絡をとって障害切り分けを行う必要があったが、オートネゴシエーションに関わる問題と想定は出来ていた。障害対応の時間を天秤にかけ、DIX-IE側のスイッチとユーザ側のルータの間に1000BASE-LXと1000BASE-LXを変換するメディア・コンバータを設置して暫定的に対応を行った。

NTT-COM大手町拠点は、10Gbps接続が1ユーザ、1Gbps接続が3ユーザであったため、総作業量としてはそれほど大きくなかったため、メンテナンス・ウィンドウの時間内に作業を終えることができた。

### 3.8 テンポラリスイッチの導入とMLX-3の退役(2023年11月)

2023年11月1日のメンテナンスは無事完了した。この11月1日のメンテナンスの終了後から、NTT大手町拠点のトラフィックは、KDDI大手町拠点で導入予定のテンポラリスイッチS4048-ONを経由して、従来のKDDI大手町のユーザ接続用スイッチであるMLX-3に到達することになった。これは、S4048-ONにトラフィックをかける

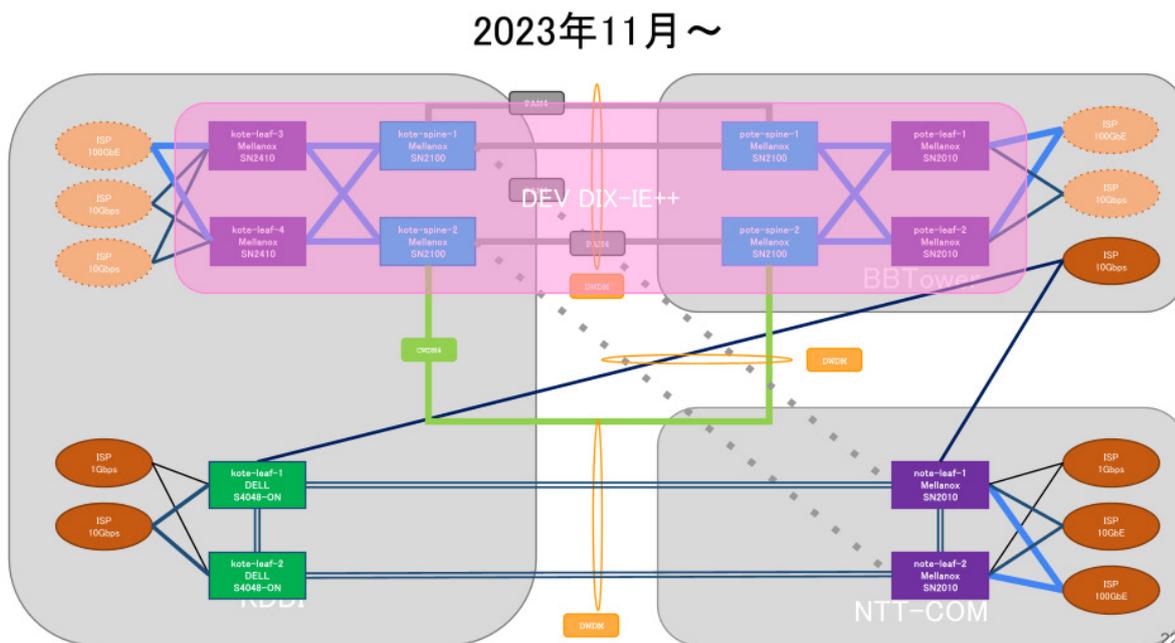


図 リング構成

シェークダウン試験を実施していることと同義であり、ここでS4048-ONのトラフィック交換上の問題が無いことを確認出来たことは、次のステップに進む上で重要な点であった。

また、移行の結果、BBTower拠点で利用していたNTT-AT社の光スイッチが捻出することが出来たため、この後のKDDI大手町拠点で利用することが可能になった。

次のメンテナンスは、2023年11月15日にKDDI大手町拠点で実施した。KDDI大手町拠点のユーザ接続用スイッチであるMLX-3に接続しているユーザをテンポラリスイッチのS4048-ONへと接続変更を行うことであった。本メンテナンスではスイッチが異なることによるトラブルがないことを確認するために、光ファイバの接続の問題を極力回避することを選択した。

このメンテナンスに引き続き、テンポラリスイッチでのトラフィック交換に問題がないことを確認した2日後の2025年11月17日に、MLX-3の物理的な撤去を実施した。じつは、ここが本プロジェクトのゴールでもあり帰還不能点であった。新規のテンポラリスイッチでのトラフィック交換に問題が生じた最悪の場合では、MLX-3への切戻しも想定すべきであった。本来であれば、一ヶ月程度の時間をあけた後に、MLX-3の撤去を行うべきであったことが反省点である。

### 3.9 光L1スイッチと構造化配線の導入(2023年11月)

2023年11月の第3回の大規模メンテナンスは、2023年11月30日に実施した。このメンテナンスでは物理層での整備に主眼を置き、次の事項を実施した。

- 配線システムの整理
- 構造化配線の利用開始
- レイヤ1の光スイッチの導入

配線システムの整理とは、ユーザの入線システムを整理し、ランディングラックを経由した接続とすることである。NTT-COM大手町拠点では、2022年末の本館から別館への移転の際にランディングラックの概念が導入され、すでに配線システムは整理されていた。KDDI大手町拠点にランディ

ングラックを導入するのが、このメンテナンスの主眼であった。

前述した配線システムの整理のためにも構造化配線の導入が必要であり、当該メンテナンスの直前までに構築を完了する必要があった。ところが、導入予定の構造化配線において、パッチを集中させるラック位置には、旧ユーザ収容スイッチのMLX-3のマウントされていた。MLX-3の撤去を急いだ背景には、この構造化配線の構築を行う必要があったことがある。MLX-3の撤去から一週間で、構築するには工夫が必要であった。MPOからLCへのMPOカセットとMPOケーブルで構造化配線を利用することで、構築に要する時間を3日程度と最小限に抑えることに成功した。

なお、当該メンテナンスでは、事前に新規の配線システムを並行して敷設し、当日は切替のみを行うことにより、トラブルを最小限にするように努めた。

レイヤ1の光スイッチは、2023年10月の時点で、NTT-COM大手町拠点とBBTower大手町拠点に導入されていた。KDDI大手町拠点で、BBTowerで導入して稼働実績があったNTT-AT社製の機材をBBTower大手町拠点から移設し導入した。これは、この後に移行が想定されていた100Gbps対応のスイッチの導入を安全に行うためには必須の要素である、との認識であった。

課題としては、光スイッチがSMF対応のものしか確保できなかったことがある。全ての接続ユーザの構内回線がSMF接続ではなく、一部のユーザのMMF接続であることが課題であった。このうち、9階へMMF配線で接続していたユーザはメディアコンバータを介してSMFへ変換して接続することで、光スイッチを導入を可能にした。ただし、5階からMMFで接続しているユーザが5組織存在したが、これらのユーザについては当メンテナンスでの光スイッチの導入は諦め、次回のメンテナンスに持ち越すことにした。

長期的にはユーザを1000BASE-LX 10GBASE-LRなどのSMFでの接続へと誘導することとして、SMFに配線への移行と、光スイッチの導入へめざしている。

### 3.10 マネージメントシステムの再構築(2023年12月～2024年3月)

2023年11月の大規模メンテナンスの後の次に実施したのは、マネージメントシステム(サーバ群)の再構築であった。暫定的なスイッチの導入の後には、本格的なClos networkへの移行と100Gbpsのサービスインなどが想定されていた。その過程で問題なく移行するためには、さまざまな監視を行う必要があり、そのためにはそれなりの支援システムが不可欠である。最低でも次のような機能が必要としていた。

- ユーザのポートへの到達性監視
- IXのトラフィックを交換するシステムの正常性監視
- IXのトラフィックの可視化
- IXの機材などのログ収集

また、認証統合のマネージメント・システムの構築も急務であった。この背景としては、DIX-IE3では管理すべき機材が非常に増えることが想定され、これまでの共有パスワードによる運用を継続するには無理が生じてしまったことによる。それまでのDIX-IEではトラフィック交換用のスイッチは全拠点で数台のオーダーであったが、DIX-IE3のアーキテクチャではPoCの段階でも16台のスイッ

チが稼働予定であり、かつ、マネージメント関連のインスタンスも含めると、相当数の機材を稼働させる必要があった。本格的なDIX-IE3の稼働の前に、radiusプロトコルなどを利用した認証統合の導入が不可避であった。

本来であれば、マネージメントシステムの導入が先にあり、その後にメンテナンスを実施すべきであったが、残念ながら、2023年11月に実施したメンテナンスは意味合いとしては緊急非難的な要素が強かったために、暫定構成からマネージメントシステムで運用し、時間差で作り込む形となったのはやむを得ないところであった。

2023年11月の大規模メンテナンスの終了後の2023年12月から構築を開始し、2024年2月には、最低限に必要な機能の実装を完了した。

### 3.11 トポロジの変更: KDDI大手町5階拠点の構築(2024年5月)

2024年5月に実施したメンテナンスでは、KDDI大手町拠点の5階にユーザの収容拠点を新たに構築した。

2003年にDIX-IEのユーザ収容拠点を9階に構築した際には、KDDI大手町9Fにユーザが移行することを想定して

2024年05月22日

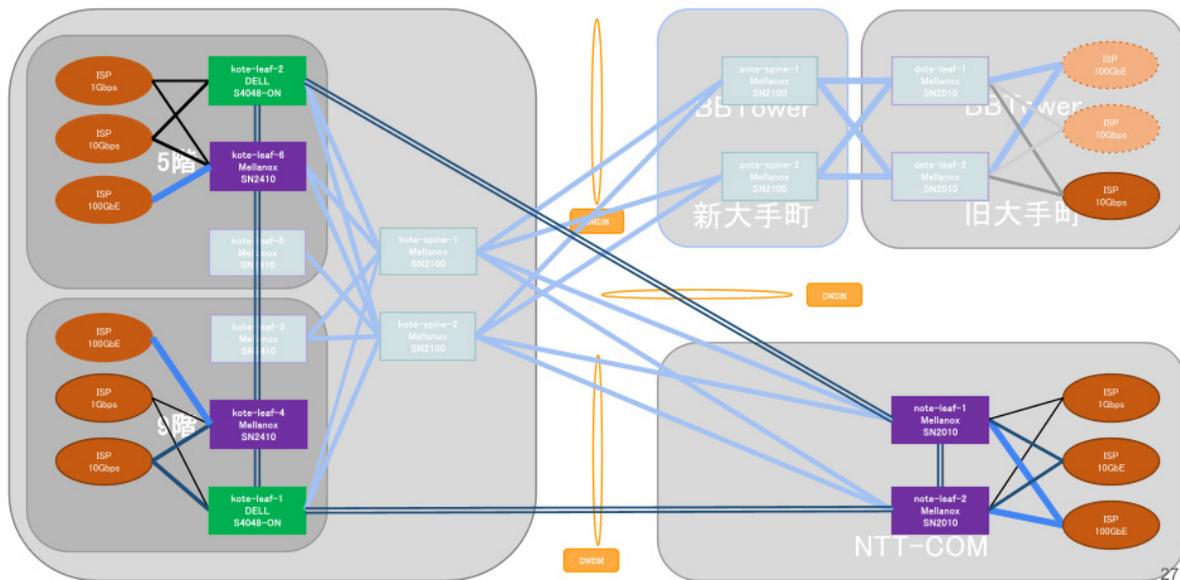


図 KDDI大手町5階拠点の構築

いたと思われる。しかしながら、当初の想定以上にユーザが5Fでの接続のままに残り続けており、KDDI大手町拠点におけるユーザの半数は、5階での接続であった。DIX-IE3の当初の想定では、9階で全てのユーザを接続する予定であったが、5階から9階までの階間ファイバは、WIDEのバックボーンなどでも使用しており、階間のファイバが不足することが想定された。階間のファイバは、サイト間ファイバと同様に高価であるので、予算を確保出来ていない短期間での増設は難しい。

このため、CLOS Network構成とEVPN/VXLANならではの利点を活かして、KDDI大手町拠点の5階にユーザ接続用のLeafスイッチを置くことで、上記問題を解決した。接続用のスイッチとしては、KDDI大手町拠点の9階でバックアップとして利用していたスイッチを移動することで対応を行った。

それまでは、NTT-COM大手町9階拠点とKDDI大手町拠点にて、ユーザ接続用のスイッチがそれぞれ2台ずつあるのリング構成であったが、この2024年5月のメンテナンスにおいて、NTT-COM大手町拠点とKDDI大手町9階拠点と5階にスイッチがあるリング構成に変更した。KDDI拠点内の構内でのスイッチ間は40Gbps Ethernetで接続し、NTT-COM大手町拠点のスイッチが関係するところでは10Gbps×2回線のアグリゲーション(20Gbps)で接続し、総トラフィックが20Gbpsまで無問題なように帯域を確保した。

### 3.12 トポロジの変更: 100Gbpsサービスの開始(2024年5月～7月)

DIX-IE3への移行におけるゴールの一つは、100Gbpsサービスの提供であった。このために必要となることは、100Gbpsに対応するスイッチの導入、および、100Gbpsに耐えるトポロジーへの移行である。

一点目のスイッチのデプロイについては、SN2410のDIX-IEでの稼働実績がなかったため、2024年5月に実施した大規模メンテナンスにおいてデプロイを実施し、1Gbpsと10Gbpsの動作についての実績をみることにした。試験的なデプロイであるので、光スイッチを利用することにより、旧来のDELL S4048-ONへいつでも切戻し

ができる状況でデプロイするようにした。

100Gbps接続ポートのサービスのデプロイについても、段階を踏むことにし、提供先を次のように行った。

#### 1. WIDEプロジェクト(AS2500)

- 問題が生じた際に即座に切り戻せるASへの提供
- WIDEプロジェクトの自AS2500のボーダールータ juniper1.otemachi.wide.ad.jpの接続

#### 2. INTEROP Tokyo(AS290)

- イベントトラフィックでの試験的提供
- INTEROP Tokyo 2024 (2024年6月12日～2024年6月14日:幕張メッセ)への試験的な提供を実施

これら試験提供を経てスイッチの問題ないことを確認し、本格的なサービス提供を開始することにした。

二点目のトポロジーについては、2024年5月のメンテナンスにおいて、40Gbpsおよび10Gbps×2回線のアグリゲーション(20Gbps)のループ構成となっていた。しかしながら、DIX-IEの総トラフィックが20Gbpsに抑制されてしまうため、これを改善する必要があった。

移行については、次のようなプロセスを踏んだ。KDDI大手町5階と9階、および、NTT-COM大手町拠点におけるLeafスイッチから、KDDI大手町9階に設置した二台のSpineスイッチ(Mellanox SN2100)まで、あらかじめ100Gbps回線を接続。これらの回線の物理的な接続(レイヤ1的なリンク)に問題がないことは事前に確認した。さらに、EVPN/VXLANのSpine-Leaf間のBGP設定はあらかじめ設定を投入し、BGPセッションの確立などを確認した上で、大規模メンテナンスまではBGPをshutdownするようにした。

2024年6月26日に大規模メンテナンスを実施した。このメンテナンスにおいては、Leafスイッチについて、一台ずつSpine側とのBGPセッションをno shutdownし、さらに、隣接するLeafスイッチとのBGPセッションをshutdownすることにより、トポロジーの移行を実施した。移行の際には、cactiでモニタしているトラフィック

がSpine側に移行し、各ポートにErrorsやDiscardsが出ていないことを確認した。

この移行により、DIX-IEはLeaf-SpineのClos networkのトポロジに移行し、トラフィック的には最大200Gbpsが可

能となった。これにより、正式にKDDI大手町拠点とNTT-COM大手町拠点での100Gbpsサービスを開始することになった。

2024年6月26日のメンテナンスの時点では、BBTower拠

## 2024年06月26日

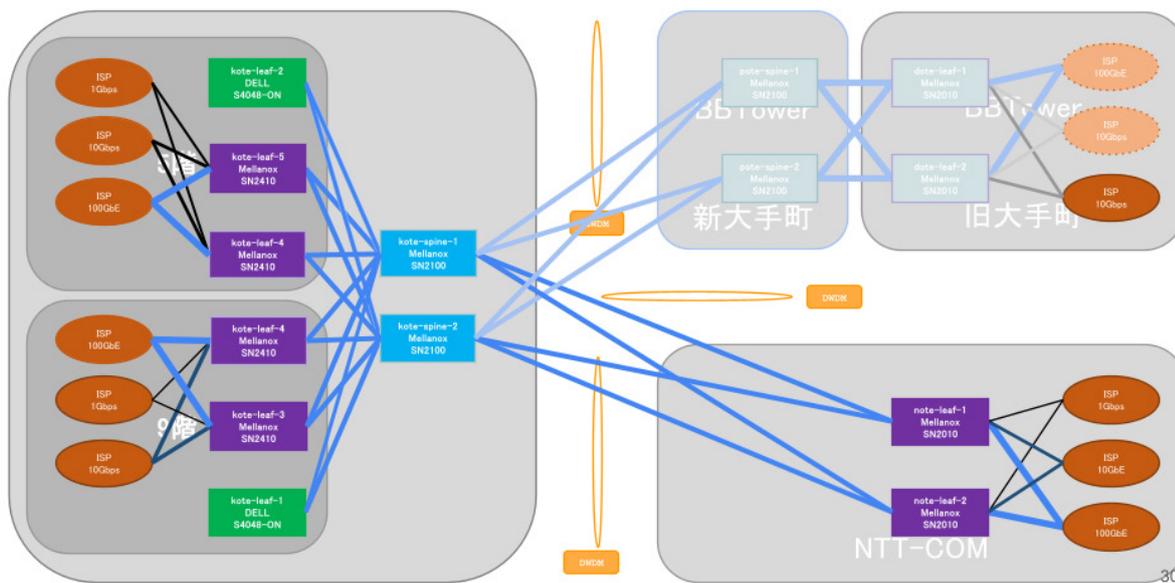


図 2024年06月26日

## 2024年07月

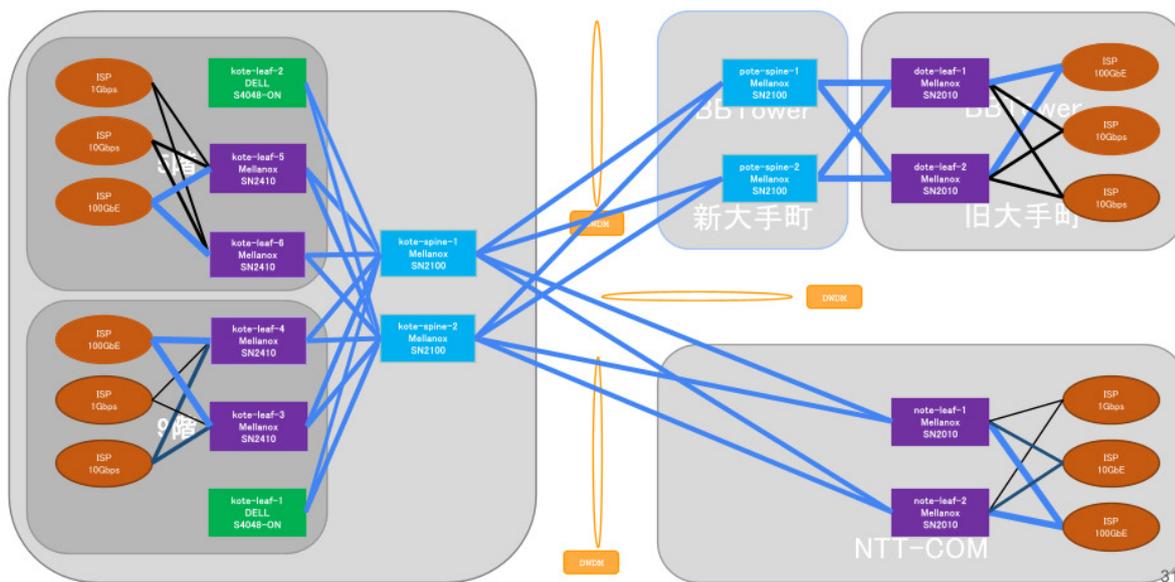


図 2024年07月

点についてのLeaf-Spineへの移行は完了していなかった。BBTower大手町拠点のユーザは、DWDM回線でKDDI大手町I拠点とNTT-COM大手町拠点へ接続する構成は残存していた。2024年7月に、BBTower拠点にNTT-COM大手町拠点と同じMellanox SN2410を設置し、同様のLeaf-Spineのスイッチ構成を構築した。ここでのユーザの移行は、関連するASが少数のため、スケジュールを調整の上で実施できるものであった。

以上のメンテナンスにより、KDDI大手町(5階・9階)、NTT-COM大手町、BBTower拠点において、Leaf-SpineのClos networkの構成に移行し、100Gbpsサービスを提供可能となった。これにより、DIX-IE3への移行が完了することになった。

### 3.13 マネージメント・ネットワークの刷新(2024年1月～11月)

DIX-IE3のIXセグメントの構築と同時に、マネージメント・ネットワークの刷新も順次行った。大きい事象としては、拠点間の商用イーサネット回線の廃止、マネージメント・サーバ・システムの構築である。

DIX-IE3のマネージメント・ネットワークでは、スイッチ

間の冗長を制御するプロトコルとしてはRSTPを採用する設計であった。これは、安価なマネジメント用のスイッチでも実装がされており、異ベンダでの相互接続性が十分に担保されていることから、マネジメントスイッチの選択肢が多様になることが理由であった。ところが、旧DIX-IEでは、トラフィック交換を行うIXセグメントでRSTPを冗長構成のプロトコルとして利用してきた。このために、もしもIXセグメントとマネージメント・ネットワークとを接続してしまった場合に、想定外のスイッチのポートがブロックされてしまう事故が発生する可能性があった。このため、MLXの退役を行った2023年の11月までは、マネージメントネットワークにおいてRSTPを利用することができなかった。また、PIX-IEでは、信頼性の高い商用イーサネット回線によるコントロールプレーンを前提としていたために、PIX-IEへの移行が想定されている状況では、商用イーサネット回線をなかなか廃止することはできなかった。このため、DIX-IE3の稼働が確実にされた2024年の6月に、ようやく、商用イーサネット回線の廃止を行う事ができた。

2024年6月に商用イーサネット回線を廃止した段階では、マネージメントネットワークには各拠点間を結ぶトライアングル(ループ)のトポロジーであり、RSTPを利用

## DIX-IE3マネージメントネットワーク(2023年6月)

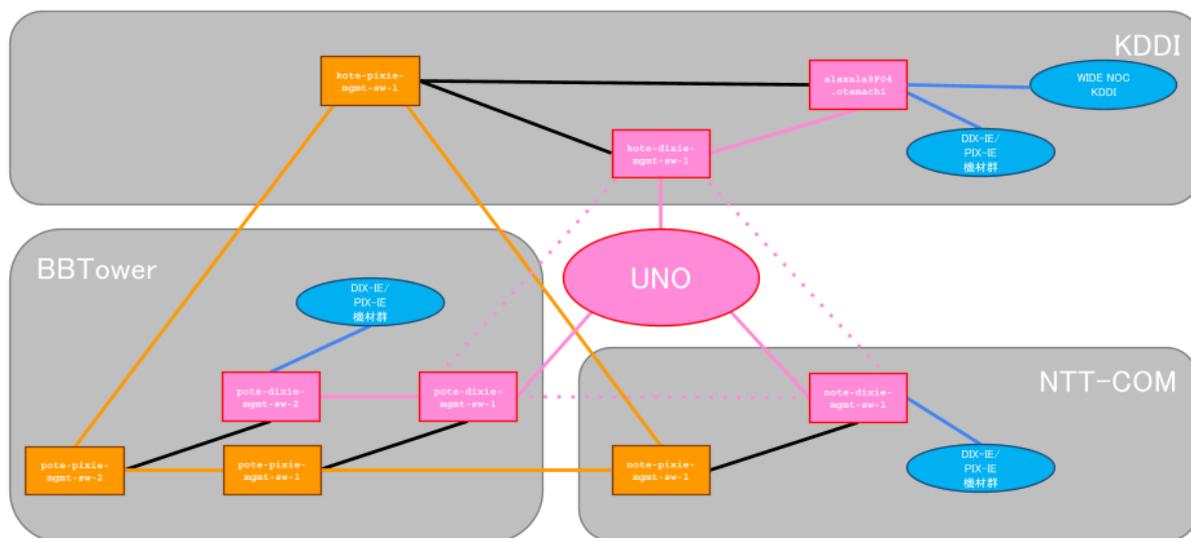


図 マネージメントネットワーク

して冗長化を図っていた。ただし、KDDI大手町拠点のマネージメントスイッチが、SPoFとなってしまうため、この時点から冗長性を確保するように変更を加えていった。

マネージメント・ネットワークの中心となるスイッチには、2台のDELL S4048-ONを利用している。これは、DIX-IEのBBTower拠点で2023年末までトラフィック交換を行っていたスイッチである。これらのスイッチをマネージメントのサーバシステムのToRスイッチとして利用し、かつ、ネットワークのコアスイッチとなるようにRSTPのルートブリッジとなるようにRSTPのpriorityを設定している。

また、マネージメント・ネットワークでは境界防御のために、ファイアウォールを導入している。これは、OSのアップデート、各種の監視、および、トラフィック情報の公開などのために、インターネットとの接点を持つ必要性からである。このため、2023年1月からマネージメントのサーバシステムを構築する段階において必要となったので、SOHOルータであるNEC IX2215の2台を暫定的に導入した。その後、DMZなど複数のマネージメントセグメントなどの高度なトラフィック制御が必要となったために、2024年11月にFortigate 200EのHA構成へと更新を行った。

---

## 第4章 DIX-IEの成功とPIX-IEの失敗

---

### 4.1 PIX-IEへの移行の失敗分析

2021年4月と8月、DIX-IEのスイッチシステムからPIX-IEのスイッチシステムへの移行を実施したが、最終的に移行を実施することは出来なかった。PIX-IEの移行の断念は、NSPIX WGとしては痛い失敗であることは確かであるが、冷静に「失敗の分析」を行い、反省(ふり返り)を行うことは重要なことであろう。

以下の内容は、あくまでも、筆者が考える「失敗の分析」であることを注記しておく。本来ならば、NSPIX WG全体で反省を行う必要があるが、そもそも反省を行う事が出来ていないことが、PIX-IEプロジェクトの課題である

かもしれない。

### 4.2 「マイグレーション」の概念の不在

PIX-IEの移行における最大の反省は、移行プロセスを軽視したことである、と筆者は考える。たしかに、PIX-IEは、その特質上、スイッチシステムの移行をホットカット的に行う必要があった。つまり、一度のメンテナンスにおいて、移行を実施し、全て成功しきる必要があった。これが、集中型のコントローラでトラフィックのコントロールを行う、という特質上、当時としては避けられない課題でもあった。

一般に、ホットカットでの移行を成功裏に実施するためには、いくつかの条件があると思われる。それは、切戻しのシナリオなどのコンティジェンシプランを含めた移行シナリオの作成と綿密な検討、詳細なシステムのヘルチェック、そして、また、ホットカットで課題が発生した場合に(本番システムに対応した)試験システムの存在である。しかしながら、これらはPIX-IEの移行では存在しなかった。

さらに、失敗の背景には、「必ず成功する」との前提による思考が多いように思われる。ホットカットのシナリオを選択する心理的な背景には、この発想を行っている結果なのではないだろうか？

失敗を前提に移行のシナリオを作成した場合には、マイグレーションというキーワードが重要になる。必ず失敗がある前提に立ってシナリオを作成する場合に重要なことは、最低での戦術的な成功を定義し、最悪での撤退プランを作成することである。システムの移行では、本当に最悪な状況は、全面的な切戻しを実施することになる。だが、このような切戻しは、しばしば成立しない。一般的には、とくにデータベースの変更を伴う移行では、完全な切戻しが困難な「帰還不能点」が存在する。幸いにも、ネットワークの場合は、切戻しが容易であるため、このような意識が低いものと思われる。

### 4.3 「移行プラン」のドキュメンテーションの不在

システムやネットワークの移行を行う場合には、移行作業に関する手順書などのドキュメント作成が必要であ

り、また、検討会の実施などによる複数人による確認作業が重要であると筆者は考える。ただし、これらの作業の多く、とくに、検討会の実施は形式ともなりがちなために、確かに空虚なものに感じることもあるであろう。WIDEプロジェクトの研究開発IXであるDIX-IEでは優秀なネットワークエンジニアが揃っており、これらの実施は不要であると考えていたのではないだろうか？

筆者は、この点について疑問を投げかけたい。確かに、そう言った側面はあることは確かにあるかもしれない。商用サービスでは、WIDEプロジェクトと比較すると、優秀ではないエンジニア達でサービスを実施することが必要であるために、より多くの時間を消費する必要があり、優秀なエンジニアからすると無駄な時間にも感じられるだろう。では、優秀なエンジニアが、同等なプロセスを実施したら、どうなのだろうか？よりエクセレントな移行を実施できるのではないだろうか？「優秀な人間の時間」という、タイムパフォーマンス、イマドキの言葉で言う「タイパ」を重視し過ぎだったのではないだろうか？過信はなかったのか？

現状のWIDEプロジェクトは、自身よりも運用が劣っていると考えている商用サービスから学ぶことは、非常に多いであろう。

#### 4.4 PIX-IEへの移行の失敗とDIX-IEの移行での対策

PIX-IEの移行の失敗をまとめると、次の3点である。

- マイグレーションの概念の不在
- 移行プランのドキュメンテーションの不在

ただ、これらの課題は、2020年ごろWIDEバックボーンでも同様であったので、PIX-IEチームだけの問題と捉えるべきではないだろう。

DIX-IEからDIX-IE3への移行においては、これらの反省を踏まえて、反省に基づいたアクションを実施した。

マイグレーションについては、DIX-IE3への移行を一気に実施しなかったことには留意してもらいたい。大規模メンテナンスを複数回に分けて実施しており、その都度

には、大きなゴールを設定し、その到達を確認しつつ進めた。本報告書でも、時間軸の変遷を詳細に記述し、その際の移行の「ゴール設定」を強調しているのは、この考え方に基づいている。

移行プランのドキュメント化に努めた。移行のためのタスクが多い大規模メンテナンスにおいては、詳細のアクションアイテムをリスト化も行った。ドキュメント化を行うことで、事前の準備などの抜けや漏れを最小限にすることが出来たと考えている。

---

## 第5章 DIX-IEとDIX-IE3の課題と今後

---

### 5.1 今後に向けて

まず、DIX-IEの今後を検討する上で、商用IX事業者を含めた日本のインターネットの現状を俯瞰して見ることは重要である。さらに、現状でNSPIX WGが持つ課題を正面から見据えることも重要であろう。

### 5.2 日本のIXの歴史的な俯瞰

DIX-IEの今後を検討する上で、まずは、商用IXを含めた日本のIXの状況について、30年におけるDIX-IEと商用IXの位置づけを振り返る必要がある。

NSPIXとDIX-IEにとって、時代は3区分されると筆者は考える。

1. NSPIX中心の時代
2. 商用IXの勃興期：DIX-IEの衰退のはじまり
3. 商用IX中心の時代

1990年代から2004年ごろまでの約10年間は、NSPIX中心の時代であった。そもそも、NSPIXは1994年からIXサービスを開始しており、1996年にはNSPIX2において100Mbpsのポートサービスを商用IXに先んじて開始していた。

1997年に、日本初の商用IXとして日本インターネットエクスチェンジ(JPIX)が創業したが、商用IXといってもNSPIXとのサービスの違いは打ち出すことは難しく、

既存のユーザ数がNSPIX2に劣る状況での営業活動であったと伝えられている。2001年にはマルチフィードがJPNAPサービスを、2003年にはソフトバンクグループのBBIXがIXサービスを起ち上げた。

分散IXについても、2001年にDIX-IEの構想が発表し研究開発が進められ、2003年に運用を開始した。分散IXの展開においてもNSPIXがリードしていた。10Gbpsのポートサービスの開始も、DIX-IEとJPIXは同時期の2003年であった。

これまでは、NSPIX中心の時代であったと言える。

2003年にBBIXのサービスが開始され、日本のキャリアグループごとのIXが出そろい、その後、商用IXの3社は営業活動を競いあってユーザを増やしていった。WIDEプロジェクトの2005年の報告書によると、2003年から2004年にかけて総トラフィックの伸びの変化が観測されるようになった。DIX-IEのユーザ数も、2005年の総ユーザ数75ネットワークをピークに減少に転ずる。DIX-IEの総トラフィックも、2008年に30Gbpsを越えるが、これをピークにトラフィックは減少傾向に転じた。この間、ユーザ数は大きくは変わらない中で、トラフィックの減少があることから、日本のISPがDIX-IEをトラフィック交換の場として活用しなくなる傾向が生じたとみるべきであろう。商用IXのJPIXでは、この当時100Gbps以上のトラフィックがあり、この時点で日本のインターネットエクスチェンジの主役の交代が起こったと解釈できる。

2010年代に入ると、この差は著しくなった。2010年の総トラフィックは約10Gbpsであり、その後の2024年現在のトラフィックが半分以下の3～5Gbpsである。2011年のWIDE報告書からは、ユーザ数と総トラフィックの記述さえ掲載されなくなったので、この間の変化を追うことは出来ないが、長期間で減少していると想像できる。この間も、商用インターネット3社は成長を続けて、2024年現在では各社のユーザ数は300社前後であり、その総トラフィックも1Tbpsを越えている。これに対して、減少を続けているDIX-IEでは、ユーザ数で30組織と商用IX各社の10分の1であり、大きな差が生まれてしまった。2010年台以降は、ユーザとトラフィックという意味から

は、商用IXの中心の時代と言っても良いだろう。

### 5.3 DIX-IE3の課題

現状のDIX-IE3をIXサービスとしてみると課題は、あまりにも多い。

そもそも、IXサービスとしてみた場合、商用IX事業者が展開するIXサービスと比較して加入ユーザが少なく、ピアリング先が少ない、という問題は大きい。この結果、トラフィックも少なく、ユーザがIXとして新規にDIX-IEを選択するトラフィックや経済的なインセンティブは、まったくないと言って良いだろう。

また、もしもトラフィックを流そうとユーザが考えたとしても、DIX-IEが高帯域のポートへの対応が遅いことも問題である。2012年には、JPNAPが100Gbpsのポートサービスを開始したが、DIX-IEが100Gbpsのポートサービスを開始したのは、その12年後の2024年である。このことは、DIX-IEの立ち後れに関する典型的な事例の一つである。

過去、DIX-IEの建て直し策として、「100Gbpsポートを安価に提供する」というアイデアもあった。しかしながら、日本のIXサービス市場は、すでに過当競争下にある。日本での商用IXサービスは、コモンキャリアの資本をバックグラウンドに持つ商用のIX事業者が3社と、外資系のIX事業者が提供を行っている。1国の中で4社の商用IXサービスが展開されている状況は、日本独自の競争環境や歴史的な事情に由来する特殊な状況であり、他の国では類をみない。「100Gbpsを安価に提供する」というアイデアはマーケティング用語で言うと、「フォロワーによるディスカウント戦略」であり、成立する可能性はある。だが、同様な戦略は、過去にBBIXが選択したものである。BBIXはサービスを開始した時点で、他の商用IXに遅れること3年から7年の時間差があった。IXサービスの優位性は、ユーザ数という規模のメリットが効いてくるため、先行者と圧倒的な差が開いてしまうものである。フォロワーとしてのBBIXとしては、ディスカウント戦略と多大な営業リソースにより、この差を逆転することに成功した。だが、それには20年近くの年月を要したのである。

DIX-IEのサービス面においては、現在のIXに必要となる機能などが欠けており、課題は大きい。

- Route Serverによる経路交換
- RTBHによるDDoS対策
- Flowによるトラフィックモニタリング
- 運用Portal
- サービスのWebサイト

DIX-IE3の運用体制としての課題も多い。次の点には改善が必要である。

- 24×365運用のための自動化
- 新規のユーザのプロビジョニング
- ユーザのサポートの体制
- サポートのための情報共有システム

マネージメントシステムなども、最小限の機能しか実装出来ておらず、まだまだ未完成であり、解決すべき課題は多い。

#### 5.4 将来イメージの共有

運用において最も重要なことは、運用者どうしにおける「将来イメージの共有(コンセンサス)」である、と筆者は考える。

たとえば、ラックの利用方法、将来の移行するシステムのイメージ、システムの移行の計画(タイミング)などである。これらが存在しなければ、システムはちぐはぐなものになってしまうことは容易に想像できるだろう。一般的な通信事業者では、将来計画として立案したイメージをチーム内で共有していることが多いであろうが、実験プロジェクトであるDIX-IEには、このコンセンサスが存在しなかった。

シンプルに運用者が少なかったためにコンセンサスを形成する習慣がなかったのかもしれない。この点は、現在のDIX-IEでも課題であると思われる。現在のDIX-IEに必要なのは、将来イメージの共有であろう。

#### 5.5 研究開発の場としてのDIX-IE

商用IXの各社が、手堅い技術的な運用を行い、トラフィック量は営業力によるユーザ数の勝負となった2010年代。運用の組織力と営業力を持ち合わせていないNSPIX WGは、その得意領域とする研究開発に活路を見いだしたのは自然なことである。

2014年には、OpenFlowを利用したSDN IXであるPIX-IEの研究開発を開始し、早いスピードでPoCを構築し公開した。問題はその後の開発が停滞したことである。2017年3月に新規のメンバが加わり、新しいアーキテクチャとしてFaucet/Umbrellaに開発の中心を移行することになった。ただし、さらに問題は、その後のデプロイに時間を要したことである。4年後の2021年には、DIX-IEをリプレイスする目的で導入をトライしたが、最終的には断念することになった。

2019年にプロジェクトがスタートしたPIX-IE++も、その後、デプロイまでには多くの時間を要した。PIX-IE++は、2024年にDIX-IE3として100Gbpsポートのサービスをローンチすることが出来たが、振り返ると、そのスピードには問題があった。とくに、機材や回線の獲得に大きな課題があった。100Gbpsの回線を十分に調達できたのは2022年1月であり、デプロイのためのPoC以上のための機材を調達できたのは2023年1月であった。また、PIX-IE++の導入の最終的な意志決定は2023年9月であり、その時点から本格的なデプロイが開始された。構想発表から構築開始まで4年を要したのは、主に機材と回線の予算獲得と意志決定に時間を要したことである。

対する商用のIXの一社のJPIXでは、2018年のLINXでの成功をみて、水面下で研究開発を進めていた。JPIXでは、DIX-IEよりも早く、2023年から同様のアーキテクチャに移行を開始し、2024年12月の段階で東京地区は、すでに移行を完了している。

以上のような状況を鑑みると、研究開発のスピードでも、現状のDIX-IEは商用IXの後塵を排していると考えべきである。これは非常に重い問題である。とくに、機材や回線などのリソースの調達に要する研究開発の場合には、DIX-IEプロジェクトでは長い時間を要してしまってい

る。今後のNSPIX WGの研究開発のテーマは、スピードを優先した内容を選定すべきであろう。

## 5.6 DIX-IE3の選ぶべきテーマ

---

今後、NSPIX/DIX-IE3が進むべき研究開発の方向はどちらであろうか。前節までにみたように、商用IXの各社が圧倒的な予算とマンパワーで研究開発を行っている中で、他の商用IXが手を出すことができない領域が、NSPIX/DIX-IE3が目指すところとなるであろう。

たとえば、IETFでの標準化が関係する方面などや、IXでデプロイするにはリスクがある技術の研究開発になるだろう。たとえば、以下のようなテーマがあり得る。

- APNなどの光ネットワークと連携したIX
- RPKIとの連携したIX
- RFC8950 IX(IPv6 only IXの実験)
- RFC8327 IX(BGP session cullingの導入)

2025年現在、All Photonics Networkというキーワードで光ネットワークが注目を集めている。IOWN構想などをNTTが推し進めているなど、光伝送のネットワークでの飛躍が期待されている。

また、安定的なIX運用としては手を出しにくい、リスクがあるようなサービスも検討すべきであろう。

- IX～IX間の相互接続技術の開発
- 100Gbpsポートでのスピードテスト
- 教育目的のIX(アマチュアASホルダーの相互接続)

100Gbpsスピードテストでは、iNoniusプロジェクトとの連携などの模索を開始している。

## 5.7 NXPISP/DIX-IE3の今後

---

DIX-IEは、2024年6月にDIX-IE3として100Gbpsサービスを開始し、一応の区切りを迎えた。今後のNSPIX/DIX-IE3においては、研究開発のIXとしての役割を担っていくことを目指したい。