

2 IOWN Global Forum の PoC 活動を通じた実証

郊外型データセンタを活用する AI 分析の遅延と消費電力を IOWN 技術により大幅に削減

IOWN (Innovative Optical and Wireless Network) 構想の実現を目指しアジア・米州・欧州を含む 138 の組織・団体が参画 (2024 年 1 月時点) する IOWN Global Forum (以下、IOWN GF) では、メンバー企業が協力し実証実験 (PoC) を推進している。本稿ではその枠組みに沿い NTT を含む 4 社が共同で実施した PoC の概要とその成果について紹介する。

大規模データセンタ (以下、DC) における AI 分析のリアルタイム性能や消費電力が課題に

スマートシティで監視カメラ映像などのセンサデータを AI 分析するようなユースケースにおいて、センサ設置拠点で AI 分析を行う方式を採用すると維持管理コストが高く、AI モデルやハードウェアの進化に追従しづらい。AI 分析を大都市圏のエッジ DC で行う方式を採用しても、DC の用地や電力不足が課題となり AI 処理を収容可能な DC を見つけることが難しい。一方、そうした問題の少ない郊外型 DC で AI 分析を行う場合には、ネットワーク (以下 NW) 遅延や大量データ収集のオーバヘッド増加といった理由によりリアルタイム性能が問題となる (図 1)。

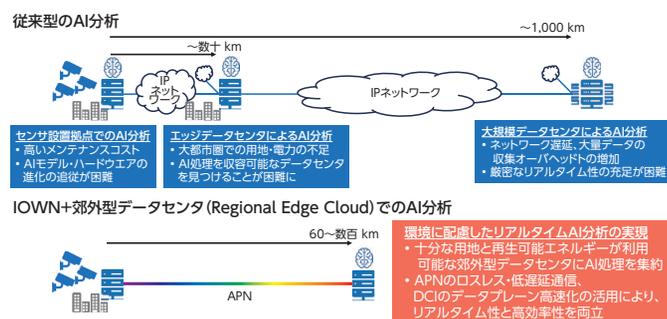


図 1 郊外型データセンタによる AI 分析の必要性

課題解決につながる PoC が IOWN GF から認定

NTT はこの問題の解決に向け富士通、NVIDIA、Red Hat と共同で PoC を行った。その概要は 100km 以上の遠隔拠点で AI 分析を行う際の遅延と消費電力を IOWN 技術により削減するというものであった。この PoC により、従来技術と比較しセンサ設置拠点にてデータ取得後 DC において AI 分析を完了するまでの時間を最大 60%、また AI 分析に要する消費電力を最大 40% 削減できることを確認した。

この PoC は IOWN GF が発行している PoC リファレンス (計画書) のうち、スマートシティを題材とした大規模映像推論とそのデータ共有における IOWN 技術適用を目的とする

“CPS Area Management Security” に基づき実施された。その結果は IOWN GF の基準を満たすことが認定され、IOWN GF 公式



NTT ソフトウェアイノベーションセンター
AI 基盤プロジェクト
研究員 原田 臨太郎氏

Web サイトにおいて詳細なレポート^{*1}も公表されている。

IOWN GF が PoC を推進する目的は大きく 2 つある。1 つめは IOWN GF の各文書で示した技術の実施可能性・有効性を実証しアピールすることで、ビジネス機会の創出やデファクト化の足がかりにすること。2 つめは更なる技術課題の発見と議論に役立てることだ。認定される PoC は製品化につながる「Milestone」、製品化の前段階である「SSF (Significant Step Forward)」の 2 つに分類されており、今回の成果は SSF に認定されている。

IOWN 技術によるリアルタイム性と高効率性の両立

本 PoC で用いた AI 分析基盤には、

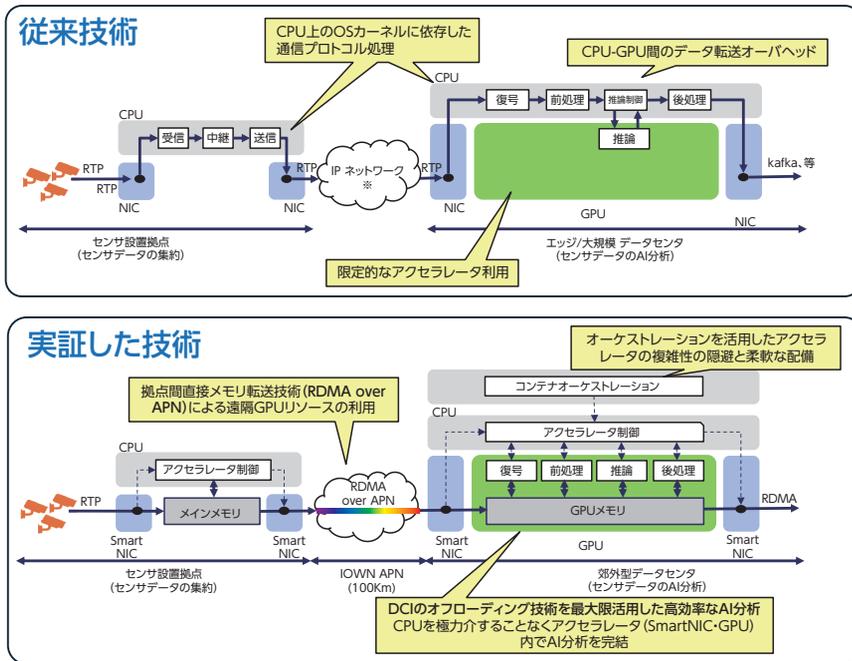


図2 従来技術と実証した技術

大きく2つのIOWN技術が活用されている。1つめは光波長の占有により拠点間の通信における大容量化、低遅延化、さらに将来的には低消費電力化を実現するAPN (All Photonics Network)だ。センサ設置拠点とDCをAPNで接続することにより、低遅延・ロスレス通信を実現する。

2つめはさまざまな性質をもつデータを効率よく転送、処理、蓄積するためのデータ中心 (data-centric) の考え方に基づくコンピューティング基盤DCI (Data Centric Infrastructure)におけるデータ処理高速化手法だ。より具体的には「アクセラレータを用いたオフローディング技術」を活用する。AI分析を高速に行うためGPUのようなアクセラレータを使用する際、従来技術ではCPU上のOSカーネルにおける通信プロトコル処理やCPU/アクセラレータ間のデータ転送におけるオーバーヘッドにより、

大規模データを収集・分析する際に遅延増加やボトルネックの要因となっていた。アクセラレータを用いたオフローディング技術によりCPUを極力介さずデータを転送しアクセラレータ内でAI分析を完結することで、大幅な効率化、低消費電力化を実現する。

多くの製品を組み合わせ、AI分析基盤を実現し、解析時間と消費電力を大幅に削減できることを確認

横須賀のセンサ設置拠点からファイバ長で100kmの距離がある武蔵野のDCをAPNで接続し、AI推論にはNVIDIAのGPUとSmartNICを搭載した富士通のサーバを利用した。アクセラレータへのオフローディング技術として、NVIDIAのソフトウェアやオープンソースソフトウェア (OSS) を活用した。またRed Hatが提供するコンテナオーケストレーションソフトウェア

OpenShiftを活用するなど、多くのハードウェアとソフトウェアを組み合わせ、NTTソフトウェアイノベーションセンタ (以下、SIC) がシステムをインテグレートした。

「多くの製品・OSSを組み合わせ、AI分析基盤を実現しました。互換性等の問題により設計が難しい部分もありましたが、最終的にはAPNとDCIを活用したシステムを実現し、性能改善を達成できました」(原田氏)。

本PoCでは前述のように解析に要する時間を最大60%、1カメラあたりのAI分析に要する消費電力を最大40%削減できることを確認した。この結果を受けた見積もりでは、1,000台のカメラを収容した場合の消費電力を最大60%削減可能であるという。

EXPO 2025 大阪・関西万博でよりリッチなAI解析システムを紹介予定

本PoCの成果に対する反響は大きく、各種メディアで紹介された。今後の取り組みについて、原田氏は次のように述べている。

「IOWN GFが公表するレポート^{※1}にも記載されていますが、PoCを通じていくつかの課題も抽出されました。今後SICではシステムのスケラビリティ向上などに取り組んでいく予定です。また、2025年に開催される大阪・関西万博にはNTTも出展します。本PoCを通して得られた知見をもとに、よりリッチなAI解析システムを紹介できるよう、NTT内で研究開発が進められています」。

※1 : <https://iowngf.org/wp-content/uploads/2024/02/PoC-Report-Sensor-Data-Aggregation-and-Ingession-Phase-1.pdf>