

広帯域データ伝送システム ULTRA の研究開発

自然科学研究機構国立天文台 大江 将史 <masafumi.oe@nao.ac.jp>

1. はじめに

国立天文台(以下、本台)での種々プロジェクトにおいて、10 ギガビット毎秒を超える広帯域データ伝送が必要となっている。これに対して、10 ギガビット毎秒を超える高帯域ネットワーク基盤を安価に構築することが可能となったことや、汎用 PC(Intel Architecture: IA)サーバの処理能力が飛躍的に向上するなど、データ伝送処理を安価に実現できる環境が整うようになった。

そこで、本台では、低コストかつ広帯域データ伝送を実現する ULTRA の研究開発に取り組んでいる。ULTRA は、汎用ネットワーク機器と汎用 PC サーバを基盤とした超高速 IP ルータやストレージキャッシュシステムであり、2013 年には第 3 世代「連雀」を開発し、ピーク時に 200 ギガ級の処理能力を発揮する。

本発表では、ULTRA の紹介に加え、科学的成果を生み出すために必要なキャンパス基盤ネットワークについても述べる。

2. 背景

本台の各天体観測システムや計算機システムなどは、その最適な拠点に整備され、本台の拠点間を結ぶ WAN を通じて相互接続している。本台の主要拠点とその WAN 構成は、図 2 の通りである。また、必要とする帯域や、データ伝送が必要なるタイミング、リアルタイムの必要性などの要求要件は、システムごと異なる(表 1)。故に、各システムがデータ伝送を無秩序に行った場合、WAN の輻輳などにより、全体の利用効率が著しく低下する点が課題となっている。

3. ULTRA の開発

この課題に対して、ULTRA では、伝送データを超高速ストレージによりキャッシュし、各システム(アプリケーション)の稼働スケジュールや必要帯域、(プロパティ)、WAN のトラフィックウェザーに基づいた WAN 伝送スケ

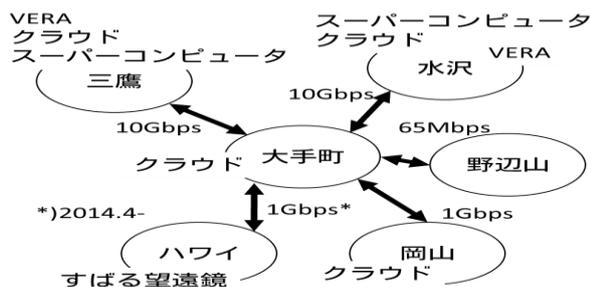


図 1 WAN 構成図

表 1 システムごとに異なる要件例

システム名	データ帯域	要求要件
すばる HSC(Hyper Supreme Cam)	一晩で 250GB~1 夜 (ハワイ→三鷹)	観測計画に基づく伝送、観測後に伝送
VERA プロジェクト VLBI 観測	2Gbps~1 局 (各局→三鷹)	リアルタイム性必要、観測計画に基づく伝送、局数に応じた帯域が必要
スーパーコンピュータ	500M~2Gbps (水沢→三鷹)	定常伝送、計算バッチ終了(最大 8 時間周期)のタイミングで伝送

ジュールに従って、TCP/UDP を組み合わせせた高効率 WAN 伝送により解決をはかっている。

この実装のため、ULTRA では、アプリケーション要求要件を満たす要素技術の検証結果から、汎用 PC サーバと高性能なネットワークカード(NIC)を組み合わせる実装を行っている。この理由は、検証を行った 2011 年当時、ギガビット級レベルでは FPGA による実装が最も高性能であったが、10Gbps 以上の帯域では、クラウドの伸長から汎用 PC と NIC の高性能化と低価格化の速度が、FPGA のそれを超えていくことが予想されたためである。

現在、ULTRA では、超高速 SSD ストレージと 100Gbps IP ルータの 2 つのトラックに分けて、開発を行っており、前者は野川(2012 年)、後者は、大沢(2012 年)、連雀/連雀+(2013 年)を発表し、2014 年現在、その両者の統合を進めている。

野川は、Intel Nehalem マイクロアーキテクチャ上に、RAID0 接続された合計 16 台の SATA 汎用 SSD を組み合わせ、32Gbps の連続書き込みに耐えうるストレージ性能を有している。

一方、IP ルータのである連雀+は、Spirent 社の計測器により、120Gbps 以上の IP フォワーディング性能と 10 μ sec 以下の低遅延伝送の両立を実現している。このシステム構成は、Intel 社 Sandy Bridge-E マイクロアーキテクチャ上に、PCI-e 3.0 対応の 40GbE ネットワークカードを搭載しており、OS は、CentOS6.3 をベースとしたものとなっている。

4. 国立天文台基盤ネットワーク

本台では、国立天文台基盤ネットワークシステムにおいて、各所のサーバールーム内にて、10/40 ギガビットイーサネットをアクセスインターフェースとして提供している。この目的は、L3/L2 スイッチング機器向けの汎用 LSI の登場により、安価に高速ネットワークの構築ができるようになった点と、先に示した ULTRA の開発事例のようにその帯域を生かせるシステムを安価に構築できるようになった点にある。

また、クラウド技術の成熟により、各システムは、これまでのシステムに紐づく計算機やストレージといった物理的な要素の組み合わせで構築されるのではなく、要素をリソースとして、たとえば、仮想サーバ、ストレージサーバ、セキュリティプラットフォーム、LAN、WAN などを集約化し、各システムの目的などに応じて、リソースから仮想化技術により、組み合わせで構築されるようになった。

この場合、リソースの結合には、広帯域・低遅延なネットワークが必要不可欠であり、このようなトレンドにも対応できるよう本台の基盤ネットワークは設計されている。

5. まとめ

汎用 PC サーバや NIC の高性能化は、そのデータ処理能力を飛躍的に向上させ、だれもが、100Gbps 以上のデータ処理が可能になった。また、その成長を支えるネットワークも汎用化により高性能低価格化が進んでいる。これらの利活用は、ノウハウを要するが、そのコスト対効果は非常に高く応用性も高いことから、組織内の生産性を高めることができる。

本台の事例からも、現在において、組織内のリソースを縦横無尽に結ぶための広帯域情報ネットワークの導入と活用は、生産性を高めるうえで、重要であることがわかる。