SINET5 へのネットワーク移行

岩本 光夫[1)], 宇野 篤也[1)],

1) 理化学研究所 計算科学研究機構

iwamoto@riken.jp

概要：計算科学研究機構では「京」と「HPCI 共用ストレージ」を SINET 経由でユーザに提供している。昨年度まで利用していた SINET4 では 40G でデータセンターと接続していたが、本年度より運用を開始した SINET5 では 100G で接続を行っている。本稿では SINET5 へのネットワーク移行作業と、移行後に発生した 100G の LAN 機器と伝送路装置間のリンクダウンについて報告する。

1 はじめに

理化学研究所 計算科学研究機構（以下、AICS）では「京*1」と「HPCI 共用ストレージ*2」をユーザに提供しており、大量のデータ転送が行われている。インターネットへは SINET 経由で接続しており、昨年度まで利用していた SINET4 では 40G(10GBASE-LR×4 本)で大阪のデータセンターと接続していたが、本年度から運用が開始された SINET5 では 100G(100GBASE-LR)で兵庫のデータセンターへ接続を行っている。本稿では SINET4 接続時に発生していたリンクアグリゲーションの問題点と、SINET5 移行作業について解説し、移行後に発生した 100G の LAN 機器と伝送路装置間のリンクダウンについて述べる。

2 SIENT4 接続時の構成と問題点

SINET4 と AICS 間のネットワーク構成を図 1 に示す。SINET4 では 40G(10GBASE-LR×4)と 10G で接続を行っていた。40G はデータ転送用として、10G はデータ転送の待機系と生活用として利用していた。40G は 40G の帯域を最大限活用するため最寄りのデータセンターではなくコアノードである大阪のデータセンターへ直接接続していた。SINET4 では 10G までの物理インターフェイスのサポートしかなかったため、40G は 10G×4 本のリンクアグリゲーションで構成していた。しかし、1 本の物理帯域以上のトラフィックが特定の回線に集中しパケットが廃棄されるという問題がしばしば発生した。これを解決するために、SINET5 では新規にサポートされる 100G の物理インターフェイスを採用することとした。

3. SINET5 への移行作業

SINET5 と AICS 間のネットワーク構成を図 2 に示す。SINET5 ではデータセンター間が 100G に増強されたため、100G(100GBASE-LR)と 10G で最寄りのデータセンターに接続を行っている。100G はデータ転送で、10G はデータ転送の待機系と生活用である。SINET5 への移行作業は 2 日間で実施した。1 日目は 10G アクセス回線を SINET4 から SINET5 へ切り替える作業で、2 日目は 40G アクセス回線を SINET4 から外し、新た

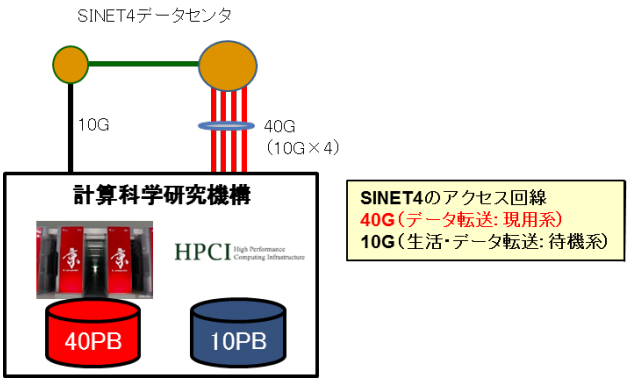

図 1 SINET4 とのネットワーク構成

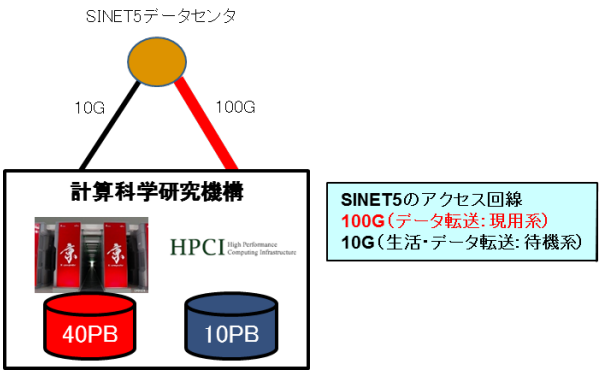

図 2 SINET5 とのネットワーク構成

に調達した 100G LAN 機器と 100G アクセス回線を SINET5 へ接続する作業であった。以下に SINET5 で導入した 100G 回線の効果と移行作業で発生した接続問題について紹介する。

3.1 100G アクセス回線の効果

前述のとおり、SINET4 時に採用していた 40G 回線（10G×4 本のリンクアグリゲーション）では、図 3 に示すように物理帯域以上のトラフィックが特定の 1 本に集中して廃棄されることがしばしば発生していた。このトラフィックの偏りは、複数の物理ポートを 1 本の論値ポートとして束ねる装置がどの物理 I/F を利用するかを送信時に自動決定することが原因で発生していた。これを回避するために、SINET5 への接続では従来の 10G の 10 倍の転送性能を持つ 100G を採用することにした。これにより、入力トラフィックが 100G を超えない限り、フレーム破棄という問題は発生しない（図 4）。

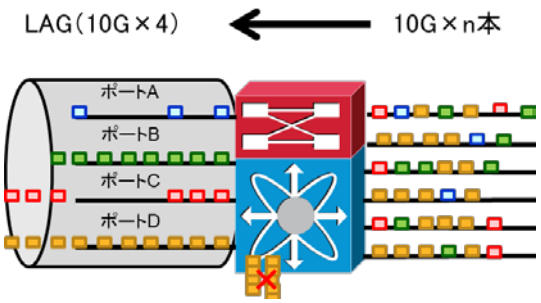

図 3　従来の 40G 回線のフレーム廃棄

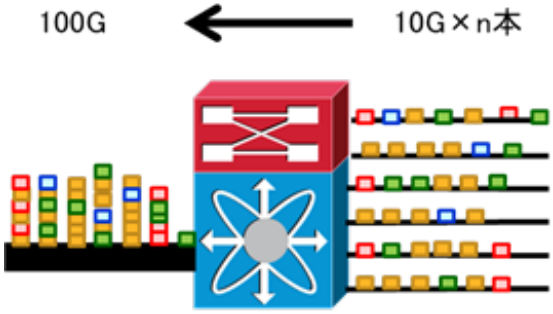

図 4　100G 回線の効果（イメージ）

3.2 100G アクセス回線接続時の問題

SINET5 への移行から数日が経過したあたりから、100G 回線で 3 秒程度のリンクダウンが発生するようになった（図 5）。調査の結果、AICS の 100G LAN 機器の出力で何らかのエラーが発生し、AICS の伝送路装置がエラーを検知してリンクダウンを起こしていることが判明した。予防保全のため、伝送路装置の光ケーブルの清掃、交換、アッテネータ（光減衰器）の挿入、QSFP 交換、ラインカード交換と 100G LAN機器のQSFP交換、本体交換を順次実施したが（図 6）、このリンクダウンは解消することなく発生し続けた。そこで図 7 に示すようにSINET5 データセンター側へAICS の 100G LAN機器を直接持込み、SINET5 の 100G LAN機器と伝送路装置なしで接続して原因の調査を行った。3 日間の調査期間中、双方の 100G LAN機器でリンクダウンは発生しなかった。また、電子計測器で光出力を計測したが、規格の範囲内であった。以上のことからAICSの 100G LAN機器と伝送路装置間になんらかの問題があると推測された。AICSの 100G LAN機器について調査した結果、海外で同様の事例が 1 件発生していたことが判った。その事例を参考に、図 8 に示す物理層の電気/光変換の送信パラメタを調整したところ、リンクダウン事象を解決することができた。このことから、受信側のジッタ[i]に問題があったと推測される。余談ではあるが，今回の事象を反映したファームウェアが後日リリースされている。

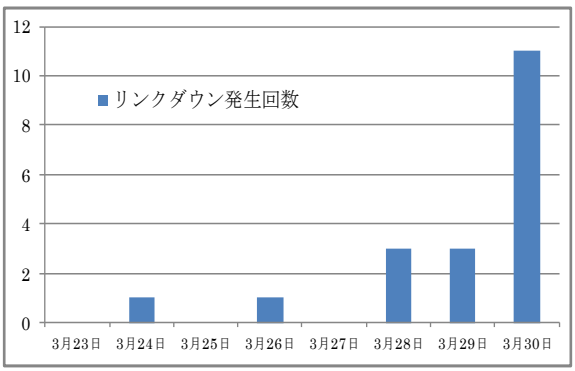

図 5　リンクダウンの発生回数

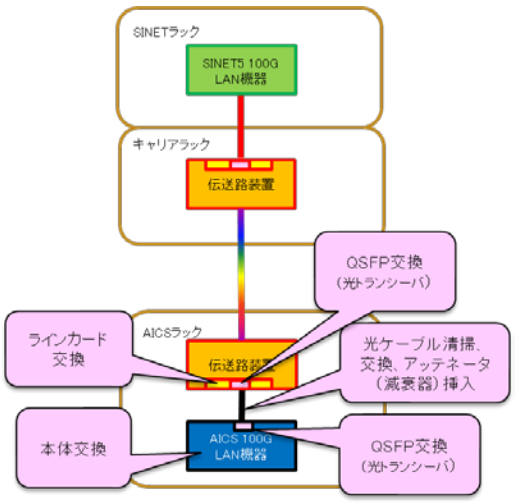

図 6　交換した部位

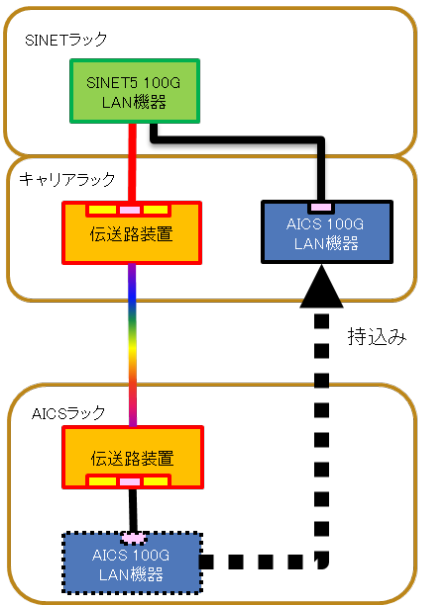

図 7　伝送路装置なしで接続

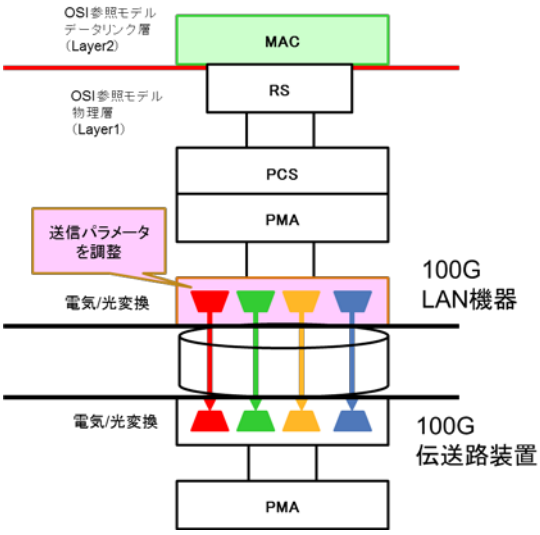

図 8　100GBASE-LR の物理層（イメージ）

4. おわりに

SINET5 への移行は、全般としては大きなトラブルはなく順調に完了した。SINET5 では県間がすべて 100Gのバックボーンで構成されており、伝送遅延が大幅に縮小されたことから、AICSのネットワーク環境はSINET4 時と比べて非常に快適になったと言える。また、SINET5 の移行後に遭遇したリンクダウン事象は、伝送路装置と 100G LAN機器の受信側のジッタ耐性のちょっとした違いにより発生したものであった。100GBASE-LRは従来の強度変調[ii]とは違い、ディジタルコヒーレント（QPSK）*3と呼ばれる通信技術が採用されており、非常にセンシティブ*4であるため、100G機器同士の相互接続時には、長期安定試験やネットワークテスター等を用いて、ワイヤーレートの帯域試験を行うといった基本的な作業が重要と改めて教えられた経験であった。

参考文献

[1][online]http://www.aics.riken.jp/jp/k/system.html（参照 2016-10-19）

[2] 原田ら，「HPCI 共用ストレージの構築と運用」,AXIES（2013）

[3] 鈴木ら,「総合報告　光通信ネットワークの大容量化に向けたディジタルコヒーレント信号処理技術の研究開発」,電子情報通信学会誌 vol.95, No.12（2012）

[4] 大内宗徳（2011）「IX と 100GBit Ethernet ー運用を想定した場合に気になる点についてー」,[online]http://www.janog.gr.jp/meeting/janog28/doc/janog28-100g-IIJ-after.pdf（参照 2016-9-26）

用語集

[i] ジッタとは、信号波形の時間的な揺らぎであり、アナログ信号では厳密に一定周期で繰り返されるべき波形が部分的に早くなったり遅くなったりすることや、受信側で再生した場合にそれによって引き起こされる品質低下の要素を指す。デジタル信号では、基準クロックや信号データの波形の位相の揺らぎによって起き、最悪の場合には受信側でのデータエラーなどの原因となる。デジタル信号のジッタは、ランダムジッタとデターミニスティックジッタに分類できる。ランダムジッタ(Random jitter、RJ) は正規分布に従う時間軸方向での信号波形の揺らぎであり、データ信号やクロック信号の波形にランダムな時間的揺らぎが含まれることで生じる。高速化したデジタル信号の伝送では、信号波の立ち上がりと立ち下がりの傾きの変化も位相揺らぎの要素となるために、電源電圧やグランド電圧の乱れもランダムジッタの原因となる。デターミニスティックジッタ

（Deterministic jitter、DJ、ディタミニスティクジッタ、確定的ジッタ、決定論的ジッタ、限定ジッタ、Bounded Jitter）はデータやクロックに依存して受信信号の波形タイミングが変化するジッタであり、同一のデータ／クロックでは常の一定のジッタが生じる性質のものである。シンボル間干渉（Inter-symbol interference, ISI）とも関連する。RJとDJを合わせてトータルジッタと呼ばれる。ジッタは高速デジタル伝送におけるシグナルインテグリティ（Signal integrity, SI）に関わる要素である。

[ii] 強度変調とは、デジタル符号の”1″と”0″を光の”ON”と”OFF”に変換して情報を伝達する方式。