データセンターと光技術

Data center and optical network technologies

大下 裕一　　村田 正幸
Yuichi Ohsita　　Masayuki Murata

大阪大学 大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University

1　はじめに

データセンター内では、著しく大きなデータの処理を、多数のサーバ間の連携により行っている。そのため、データセンター内では、広帯域・低遅延でサーバ間を接続することが求められる。さらに、データセンターが大規模化するにつれ、データセンターの消費電力は大きくなっている。そのため、データセンターの低消費電力化も大きな課題であり、サーバ・空調だけでなく、ネットワークについても低消費電力化が求められている。

その一方、データセンター内のネットワークは単一の機関により管理、構築されるものであり、接続が必要なサーバも建屋内に配置されているため、任意のネットワーク構造を適用でき、また、任意のネットワーク技術を用いて構成が可能であるため、広域ネットワークと比べ、新たな技術の導入が容易である。

そのため、光ネットワーク技術を適用した、新しいデータセンターネットワーク構成に関する研究が進められている。

2　データセンターのための光ネットワーク技術の研究開発

以下に、データセンター向けに研究開発が進められている代表的な光ネットワーク技術を挙げる。

2.1　光パススイッチを用いたネットワークの研究開発

光パススイッチは、事前に行った設定に従い、入力ポートと出力ポートを接続することができるスイッチである。マイクロミラーを用いた MEMS 型のスイッチは商用化され、広く利用されている。データセンターネットワークにおいても、光パススイッチを用いたネットワーク構成に関する検討が進められている。光パススイッチにおける入力ポートと出力ポートの対応付けの変更には時間がかかるため、パケットといった小さな単位でのスイッチングはできない。そのため、データセンターネットワークにおいて、光パススイッチを用いる場合には、電気パケットスイッチとの併用が考えられている。

その代表的な例は、図 1 に示される Helios である [1]。Helios では、電気パケットスイッチで構築した従来型のデータセンターネットワークに、光パススイッチを追加する。そして、各サーバラックから光パススイッチへの接続を追加する。本構成では、多量のトラヒックが流れるサーバラック間を光パススイッチの設定により直結する。これにより、多量の通信が必要なサーバラック間には広帯域の通信路を確保することができ、それ以外の通信は、従来型の電気スイッチを用いて通信することにより、パケット多重が可能となる。

本ネットワーク構成では、光技術は、従来型の電気パケットスイッチで構成されたネットワークを補完する形で用いられており、多量の通信が必要なサーバラック間のトラヒックをバイパスさせるといった限定的な役割のみを担っている。そのため、各サーバラックが通信する相手が頻繁に変わるような状況では、光パススイッチを用いてバイパスできるトラヒックが少なく、電気パケットスイッチの負荷を十分に削減できない。

2.2　大規模光パケットスイッチの研究開発

多数のサーバ間を接続できる大規模な光パケットスイッチの開発が進められている [2, 3]。これらの研究では、ポート数の大きな光パケットスイッチを構築し、それらをサーバ、あるいは、サーバラックと接続することが提案されている。各サーバ、あるいは、サーバラックから到着する電気パケットを光パケットに変換し、大規模光パケットスイッチを経由して宛先ポートまで転送し、宛先ポートで電気パケットに再度変換することにより、サーバラック間の通信を中継する。

これらの光パケットスイッチでは、内部にバッファは持たない。そのため、光パケットスイッチの内部でパケットの衝突が発生しないように、各ポートからのパケット送出のタイミングは、集中的に管理される。つまり、各サーバからのパケットは、一度光パケットスイッチ外でバッファされ、光パケットスイッチのコントローラに従って光パケットスイッチへ送出される。この構成では、スイッチの規模が大規模になり、また、流入するパケットの到着レートが大きくなると、スケジューラが扱う必要のあるパケット数が膨大となるという問題がある。

2.3　光電子融合型パケットルータ

光パケットスイッチと電気バッファを組み合わせた光電子融合型パケットルータについても研究開発が進められている [4]。

2.3.1　光電子融合型パケットルータ

図 2 に光電子融合型パケットルータを示す。光電子融合型パケットルータは、光ポートと電気ポートの二種類のポートを持ち、光ポートは、他の光電子融合型パケットルータとの接続に用いられ、電気ポートは各サーバラック内に設置された電気スイッチとの接続に用いられる。

各サーバラックから電気ポートを介して光電子融合型パケットルータに流入したパケットは、共有電気バッファに蓄えられたのち、光パケットに変換された上で送出される。また、光電子融合型パケットルータに直接接続しているサーバラック宛のパケットは、光パケットから電気パケットに変換した上で、電気バッファに蓄えられたのち、宛先サーバラックに送出される。

他の光電子融合型パケットルータから到着した他の光電子融合型パケットルータ宛の光パケットは、パケット

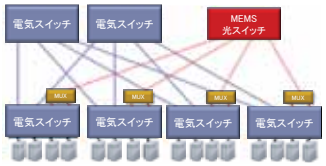

図 1　Helios

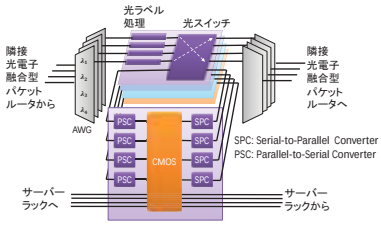

図 2　光電子融合型パケットルータ

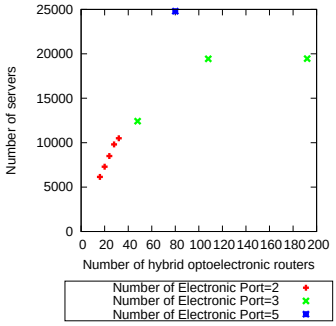

図 3　光電子融合型パケットルータ数と収容可能なサーバ数の関係

内のラベルに合わせて転送先ポートが決定され、転送先ポートが空いていれば、電気に変換されることなく、宛先光ポートを介して転送される。転送先のポートが空いていない場合は、光/電気変換を行った上で共有バッファに一旦保存したのち、再び電気/光変換を行った上で転送を試みる。

光電子融合型パケットルータでは、パケットの衝突が発生した場合であっても、電気バッファに一旦保存したのち、再度転送を試みることが可能であるため、パケットの衝突を避けるような集中制御は不要であり、大規模なネットワークへの拡張が容易である。

2.3.2　光電子融合型パケットルータを用いたデータセンターネットワーク

我々は、光電子融合型パケットルータを用いて、サーバラック間を接続するためのネットワーク構成の検討を行った [5]。本ネットワーク構成では、光電子融合型パケットルータを用いてデータセンター内のコアネットワークを構成し、各サーバラックから複数の光電子融合型パケットルータに接続した。各サーバラックからの光電子融合型パケットルータへの接続の際には、同一の光電子融合型パケットルータに接続したサーバラックは、その光電子融合型パケットルータ以外では、同一の光電子融合型パケットルータと接続することはないようにした。これにより、各サーバラックから一つの光電子融合型パケットルータのみを経由して接続可能なサーバラックの数を最大化することができる。そして、光電子融合型パケットルータ間の接続を決める際には、Valiant Load Balancing (VLB) を用いて負荷分散を行った際に、最も多くのトラヒックを帯域不足を起こさずに収容可能な接続構成を求めた。

上述のネットワーク構成方法を用い、光電子融合型パケットルータ数と、十分な帯域を確保して接続可能なサーバ数の関係を調べた。本調査では、VLB により負荷分散をされているという条件下において、各サーバ当たり 1Gbps のトラヒックを生成した。そして、100Gbps の帯域を持つ光電子融合型パケットルータ間のリンクにおいて、輻輳を発生させないという制約を満たす範囲内で、各光電子融合型パケットルータに接続するサーバ数を最大化することにより、各ネットワーク構成において、収容可能なサーバ数を求めた。

図 3 に、光電子融合型パケットルータの光ポート数が 12 の場合の結果を示す。図では、サーバラックあたりの電気ポートの数 2、3、5 の場合の結果を示す。図より各サーバラックからの配線数が 2 や 3 の場合は、20,000 台以上のサーバを接続可能な構成を構築することができないことが分かる。それに対して、各サーバラックあたり 5 台の光電子融合型パケットルータと接続することにより、80 台の光電子融合型パケットルータを用い、25,000 台のサーバを接続することが可能となる。つまり、大規模なデータセンタネットワークを構成するためには、光電子融合型パケットルータ間の接続構成のみではなく、サーバラックから光電子融合型パケットルータへの接続も合わせて設計することが必要となる。

3　むすび

光技術を用いたデータセンターネットワーク構成技術に関しては、様々な検討が進められている。今後、より大規模、高速、低消費電力なデータセンターを構築するためには、ポート数の大きな光スイッチの構成といった光ネットワーク技術の進展とともに、それらの新しい機器を導入する際には、サーバやサーバラックといった電気パケットを送出する機器と光ネットワーク機器の接続方法も合わせたデータセンターネットワーク全体の構成についても検討が必要である。

参考文献

[1] N. Farrington, et al., “Helios: a hybrid electrical/optical switch architecture for modular data centers,” in *Proceedings of ACM SIGCOMM*, pp. 339–350, Oct. 2010.

[2] H. J. Chao and K. Xi, “Bufferless optical clos switches for data centers,” in *Proceedings of OFC*, Mar. 2011.

[3] C.-T. Lea, “A scalable AWGR-based optical switch,” *Journal of lightwave technology*, vol. 33, pp. 4612–4621, Nov. 2015.

[4] S. A. Ibrahim, et al., “100-Gb/s optical packet switching technologies for data center networks,” in *proceedings of Photonics in Switching*, July 2014.

[5] Y. Ohsita, et al., “Data center network structure using hybrid optoelectronic routers,” in *Proceedings of Cloud Computing*, Mar. 2016.