

光電子融合型パケットルーターを用いた データセンターネットワーク向け低消費電力化制御の一検討

西島 孝通[†] 小泉 佑揮[†] 大下 裕一[†] 村田 正幸[†]

[†] 大阪大学 大学院情報科学研究科

〒 565-0871 大阪府吹田市山田丘 1-5

E-mail: †{t-nisijm,ykoizumi,y-ohsita,murata}@ist.osaka-u.ac.jp

あらまし データセンターネットワークへの要求として、高通信性能に加えて省電力化が求められている。低消費電力で高通信性能なネットワークを実現するため、我々は、光通信技術と電気技術を融合した光電子融合型パケットルーターを開発した。光電子融合型パケットルーターでは、ルーターの電源のオン・オフのみではなく、バッファについても不要な箇所の電源を落とすことによる低消費電力化が可能である。しかし、バッファ数を削減するとトラヒックを集約することができず必要なルーター数が増加してしまうなど、必ずしもバッファの電源を落とすことが消費電力削減に繋るとは限らない。そこで本稿では、光電子融合型パケットルーターを用いたネットワークにおいて、遅延の性能要件を満たした上で、電源を投入する必要があるルーターおよびバッファの総消費電力を削減するようにトラヒックの経路を選択する低消費電力化制御を提案する。簡単な評価の結果、グリッドトポロジのネットワークにおいて、伝搬遅延最小化を目的とした経路選択法と比較して、本手法が遅延の性能要件を満たした上で消費電力を最大 34 % 削減可能なことを示した。

キーワード 低消費電力化制御、データセンターネットワーク、光電子融合型パケットルーター、経路選択

A Study on Energy Saving for a Data Center Network with Opt-Electronic Packet Routers

Takamichi NISHIJIMA[†], Yuki KOIZUMI[†], Yuichi OHSITA[†], and Masayuki MURATA[†]

[†] Graduate School of Information Science and Technology, Osaka University

1-5 Yamadaoka, Suita, Osaka, 565-0871 Japan

E-mail: †{t-nisijm,ykoizumi,y-ohsita,murata}@ist.osaka-u.ac.jp

Abstract In data center networks, not only high communication performance but energy saving is important. For high communication performance and energy saving, we have developed an opt-electronic packet router. In a network with the opt-electronic packet routers, it is possible to reduce energy consumption by turning off not only the routers themselves but also their buffers. However, there is a trade-off between reducing the number of necessary routers and reducing the number of necessary buffers. In this paper, on a data center network with opt-electronic packet routers, we propose a route selection to reduce total power consumption on necessary routers/buffers and to achieve required performance. Through a simple evaluation, we show that our method reduces at most 34 % energy consumption in a grid topology network by comparing to a route selection for minimizing propagation delay.

Key words energy saving, data center network, opt-electronic packet router, route selection

1 はじめに

近年、データセンターネットワークへの要求として、高通信性能に加えて省電力化が求められている [1]。データセンター内では、複数のサーバが連携して一つの大きなデータの処理を

行っており、サーバ間を結ぶデータセンターネットワークは、データセンターの処理性能に大きな影響を与える。そのため、広帯域・低遅延でサーバ間を接続するネットワークが必要となる。その一方、ネットワーク機器の消費電力がデータセンターの消費電力全体に占める割合が大きくなっており、データセン

ターの消費電力を削減するためにネットワーク低消費電力化も必須である。

低消費電力・低遅延・広帯域なデータセンターネットワークを構築する一つの手法として、光通信技術を有効活用することが考えられる。光通信技術を用いることにより、消費電力の小さい光デバイスにおいて光信号を電気変換することなく直接スイッチング可能になり、低消費電力・低遅延な通信が可能になる。また、光波長多重通信により電気と比べて広帯域を実現できる。光通信技術を用いてデータセンターネットワークを構築する方法として、光パススイッチを用いる手法 [2] やバッファレスな光パケットスイッチを用いる手法 [3] が検討されている。しかし、パス設定に時間がかかるためトラヒックパタンの変化が激しい環境への適用が困難である問題や、パケットの衝突を回避するための制御が必要でありネットワークの大規模化が困難であるという問題がある。一方、光信号のままバッファリング可能な光パケットスイッチには大容量の光バッファが必要であり、ネットワーク機器の実現が難しいという問題がある。

これらの問題を解決するため、我々は、光通信技術と電気技術を融合した光電子融合型パケットルータを開発している [4]。光電子融合型パケットルータは、他のルータと接続に用いる光ポートと、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る電気ポートや電気/光変換器、光/電気変換器、電気バッファを持つ。光電子融合型パケットルータでは、光パケットを中継する場合、パケットの衝突が発生しない限り、変換器やバッファを用いずに消費電力の小さい光デバイスのみを用いてスイッチング処理を行なえる。一方、パケットの衝突が発生した場合でも、パケットをバッファに一旦保存したのち、再度転送を試みる事が可能であり、パケットの衝突を回避するための制御は不要である。そのため、光電子融合型パケットルータを用いることにより、パケットの衝突を回避するための制御を行なうことなく、光通信技術の利点である低消費電力・低遅延を実現したデータセンターネットワークの構築が可能である。

これまで、ネットワークの低消費電力化制御として、発生したトラヒックの経路を選択する際、必要な機器のみ電源を投入し、それ以外の機器の電源を落とすことによる低消費電力化がいくつか提案されている [5-8]。文献 [5] では、IP ルータで構成されたコアネットワークにおいて、トラヒックエンジニアリングにより、必要な IP ルータのポート数を削減し、不要な IP ルータの電源を落とすことで、低消費電力化を実現している。文献 [6-8] では、IP over WDM において、最適化問題を解くことにより、必要な IP ルータや光/電気変換器等の数を削減し、低消費電力化を実現している。

光電子融合型パケットルータでは、ルータの電源のオン・オフのみではなく、バッファについても不要な箇所の電源を落とすことによる低消費電力化が可能である。光電子融合型パケットルータでは、光パケットの衝突が発生しないルータの電気バッファを落とすことが可能であり、バッファの電源を落とすとしても光パケットの中継は可能である。そのため、適切にトラヒックの経路を選択することにより、トラヒック集約による必要なルータ数の削減に加えて、パケットの衝突回避による

必要なバッファ数の削減による消費電力の削減も可能である。しかし、単純にバッファ数を削減するとトラヒックを集約することができず必要なルータ数が増加してしまうなど、必ずしもバッファの電源を落とすことが消費電力削減に繋るとは限らない。バッファの電源を投入すべきか、それとも落とすべきかは、バッファの電源を投入したときの消費電力の増加量や発生したトラヒックパタン、ネットワークに要求される性能により決まる。ルータ数の削減とバッファ数の削減を両立させるためには、必要最低限のバッファのみを用いてトラヒックを集約し、ルータ数を削減する必要がある。すなわち、多くのトラヒックで利用可能なバッファのみ電源を投入することが望ましい。

そこで本稿では、光電子融合型パケットルータを用いたネットワークにおいて、遅延の性能要件を満たした上で消費電力を削減するため、電源を投入する必要があるルータおよびバッファの総消費電力を抑えるようにトラヒック経路を選択する低消費電力化制御を提案する。本手法では、トラヒックを収容するための経路を決める際に、遅延の性能要件を満たした経路候補から、新たに電源の投入の必要があるルータおよびバッファの総消費電力が最小の経路を選択する。また、トラヒックを収容するためにルータまたはバッファの電源投入が必要な場合に、他のトラヒックにも使われる可能性が高いルータおよびバッファのみ電源を投入する。これにより、性能要件を満たした上で、消費電力を抑えたトラヒックの収容が可能となる。さらに、本手法による消費電力削減の効果を明らかにする第一歩として、グリッドトポロジのネットワークにおける簡単な評価を行なう。その結果、遅延最小化を目的としたトラヒック経路選択法と比較して、本手法が遅延の性能要件を満たした上で消費電力を削減可能なことを示す。

本稿の構成は以下の通りである。2章で本稿で対象とする光電子融合型パケットルータおよびそれを用いたデータセンターネットワークを紹介する。3章において、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における低消費電力化制御を提案し、4章で提案手法の有効性を明らかにする。最後に、5章にて本稿をまとめる。

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

2.1 光電子融合型パケットルータ

図1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

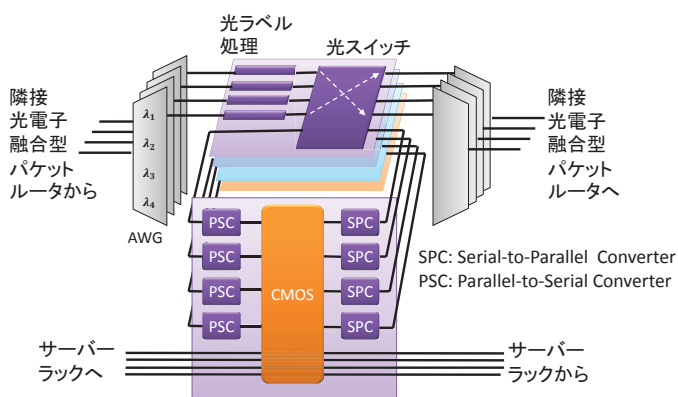


図1 光電子融合型パケットルータ

融合型パケットルータ宛の光パケットは、パケット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。

この光電子融合型パケットルータは以下の利点を持つ。(1) パケットの衝突が発生しない場合は、光/電気変換が不要で、光パケットをそのまま中継が可能であり、低消費電力・低遅延・広帯域の通信が可能である、(2) 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、再度転送を試みる事が可能であるため、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可能である。

2.2 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の構成

光電子融合型パケットルータ間は、低遅延で広帯域の通信が可能である。そのため、光電子融合型パケットルータは、データセンター内のネットワークのコアに配置し、各光電子融合型パケットルータが多数のサーバーラックからの通信を束ねて転送するネットワーク構造が適切だと考えられる。その一方、光電子融合型パケットルータが故障したとしても、サーバーラック間の接続性を確保することができるネットワーク構造が必要とされる。そのため、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の構造としては、図2に示すように、光電子融合型パケットルータを複数台相互接続してデータセンター内コアネットワークを構築し、各サーバーラックからは複数台の光電子融合型パケットルータに接続するという構成が適切だと考えられる。

3 光電子融合型パケットルータネットワーク上の低消費電力化制御

データセンター内の通信需要は時々刻々変化し、発生した通信を收容するのに必要な機器のみ電源を投入し、それ以外の機器の電源を落とすことにより、低消費電力化を行うことができる。光電子融合型パケットルータでは、バッファの電源を落と

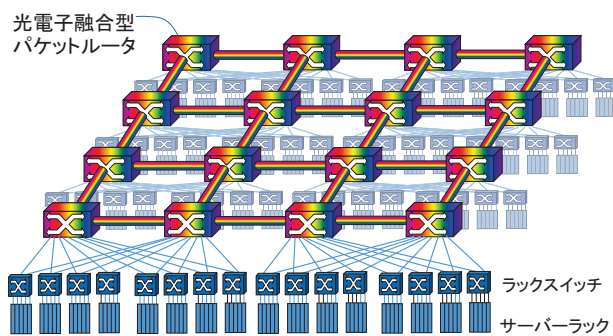


図2 光電子融合型パケットルータを用いたデータセンターネットワーク

したとしても光パケットの中継は可能であり、ルータの電源のオン・オフのみではなく、バッファについても不要な箇所の電源を落とすことによる低消費電力化が可能である。

以下の2つの条件の両方を満たす光電子融合型パケットルータは、バッファの電源を落とすことが可能である。(1) サーバラックからの流入トラヒックやサーバラックへの流出トラヒックが存在しないこと。光電子融合型パケットルータでは、サーバラックからの流入トラヒックやサーバラックへの流出トラヒックは、一度バッファに蓄えられたのちに送出される。そのため、サーバラックからの流入トラヒックやサーバラックへの流出トラヒックが存在する光電子融合型パケットルータはバッファの電源を落とすことができない。(2) 異なる入力ポートから流入したトラヒックは必ず異なる宛先ポートから出力されること。異なる入力ポートから流入したトラヒックが同一の宛先光ポートを利用する場合は、パケットの衝突が発生する可能性がある。この場合、パケットロスを防ぐためのバッファが必要となる。

バッファの電源を落としたとき、バッファの電源を投入したとき、それぞれの利点欠点は以下の通りである。

- バッファの電源オフ

バッファの電源を落としたことによりルータ単位での消費電力を抑えることができる。また、キューイング遅延が存在しないため低遅延でエンド間を接続できる。一方、エンド間で1波長を占有するなど、資源の利用効率が低く、多数のルータが必要となる。その結果、大規模なネットワーク構成になり、伝搬遅延の増加に繋がる。
- バッファの電源オン

リンクを複数のエンド間で共有できるため資源の利用効率が高く、必要なルータ数を削減でき、消費電力を抑えることができる。また、小規模なネットワーク構成になり、小さい伝搬遅延でエンド間を接続できる。一方、バッファの電源を投入したことによりルータ単位の消費電力は増加し、キューイング遅延の増加も予想される。

このように、必ずしもバッファの電源を落とすことが消費電力削減に繋るとは限らない。バッファの電源を投入すべきか、それとも落とすべきかは、バッファの電源を投入したときの消費

電力の増加量や発生したトラヒックパターン、ネットワークに要求される性能により決まる。

ルータ数の削減とバッファ数の削減を両立させるためには、必要最低限のバッファのみを用いてトラヒックを集約し、ルータ数を削減する必要がある。すなわち、多くのトラヒックで利用可能なバッファのみ電源を投入することが望ましい。

3.1 低消費電力なトラヒック経路選択手法

ネットワークの低消費電力化と低遅延化という相反する目的を両立させるため、以下の2つの方針によりトラヒックの経路を選択する。(1) 遅延を制約条件以下に抑えるため、リンク負荷が低くキューイング遅延が小さい経路およびホップ数が短く伝搬遅延が小さい経路のみを候補とする。(2) 消費電力を抑えるため、経路の候補から追加で電源を投入する必要があるルータおよびバッファの総消費電力が最小の経路を選択する。多くの場合、その経路は他のトラヒックが収容されている経路を再利用した経路となる。

本手法では、短時間でトラヒックの経路を決めるため、最適化問題を用いず、物理ネットワーク上のホップ数が短いトラヒックから順に収容先の経路を確定する。その際に、既にトラヒックが収容されている経路を部分的に再利用することにより、既に電源が投入されているルータおよびバッファを効率的に利用し、新たに電源を投入するルータおよびバッファの数を抑える。また、新たなルータおよびバッファの電源投入が必要な際には、その後に収容先が決定されるトラヒックにも使われる可能性の高いルータおよびバッファのみ電源を投入する。

このルータおよびバッファの利用される可能性を調べる指標として、以下のようにノード n の再利用容易性を定義する。

$$\sum_{s,d} \frac{N_{s,n,d} \lambda_{s,d}}{N_{s,d}} \quad (1)$$

ここで、 $\lambda_{s,d}$ は $s-d$ 間のトラヒック量、 $N_{s,d}$ は $s-d$ 間の最短ホップ経路数、 $N_{s,n,d}$ は $s-d$ 間の最短ホップ経路のうち、 n を経由するもの数である。

各送信元宛先間のトラヒックの経路決定の手順は、以下のよう決定される。

(1) 性能要件を満たす送信元・宛先間の経路の候補 P を取得する

(2) 経路の候補 P のうち、新たに電源の投入の必要がある機器・バッファの総消費電力が最小の候補を選択する

(3) 新たに電源の投入の必要がある機器・バッファの総消費電力が最小の候補が複数ある場合、再利用容易性を計算し、再利用容易性が最も大きい経路を選択する。

4 性能評価

4.1 評価環境

本章では、特にネットワークの消費電力が問題となるような、大規模データセンターを想定した環境における本手法の有効性を明らかにする。

図2に示したネットワーク構造において、トラヒックの収容を試み、その収容に必要な電力を評価した。各パラメータは、大

規模データセンターを想定して設定した。具体的には、10,000個のサーバ、1個のToRスイッチに25個のサーバを接続するとして400個のToRスイッチ、1個のルータに8個のToRスイッチを集約し1個のToRスイッチを2個のルータに接続するとして100台のルータが存在するとする。ルータは 10×10 のグリッド上に配置した。各リンクの帯域は10 Gbit/sとし、波長数は8とした。また、一般に電気技術の消費電力が光電子融合型パケットルータの消費電力の大半を占めることから、消費電力のモデルとして、各光電子融合型パケットルータのバッファの電源を落とした場合の消費電力を60 W、バッファの電源を投入した場合の消費電力を180 Wとした。

本評価では、ToRスイッチに接続された複数サーバからのトラヒックを集約し、ToRペア間にトラヒックが発生しているものとして扱う。各サーバは、送信元サーバが接続されたToRスイッチ(送信元ToRスイッチ)と送信先サーバが接続されたToRスイッチ(送信先ToRスイッチ)を介して、他のサーバと通信をしている。評価を簡単にするため、送信元・送信先ToRスイッチが同じトラヒックは、送信元・送信先のサーバを区別せず、一つのトラヒックとして扱う。

トラヒックパターンとして、ランダムに選ばれた一部のToRペア間でパレート分布に従う量のトラヒックが発生した状況で評価を行なった。一般に、ネットワークを流れるトラヒック量を削減するため、通信頻度の高いサーバは同一サーバラックに格納されることが多い。そのため、トラヒックを送信・受信するToRペア数は全体の30%とした。発生したトラヒック量は、多数の小さいトラヒックと少数の大きいトラヒックが存在すると仮定し、パレート分布に従い総トラヒック量が240 Gbit/sになるように正規化した値を用いた。

ネットワークの性能要件として、伝搬遅延を想定し、物理ホップ数が最も遠いToRペア間の最小ホップ数の1.5倍のホップ数以内で全トラヒックを収容することとした。 10×10 のグリッド上のネットワークでは、最も遠いToRペア間の最小ホップ数が18のため、本評価では1.5倍の27ホップとした。

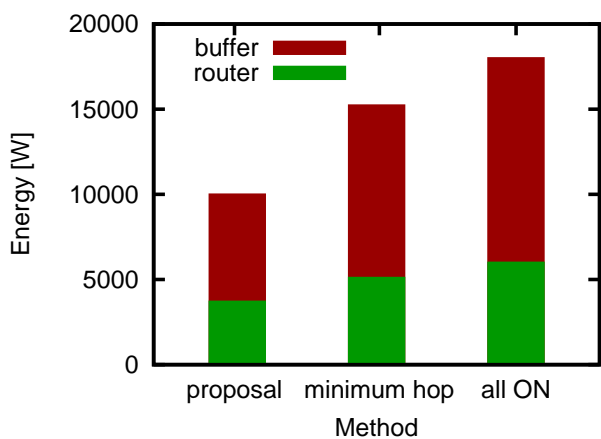
比較対象として、(1)全機器の電源を投入した場合、(2)各サーバラック間の経路として最短ホップ経路を用いて遅延を最小化し、不要なルータ・バッファの電源のみオフとした場合を用いた。

以降、断りが無い限り、上記の設定においてトラヒックパターンをランダムに100通り生成し、提案手法および比較手法を用いたときの消費電力の平均を求めた。

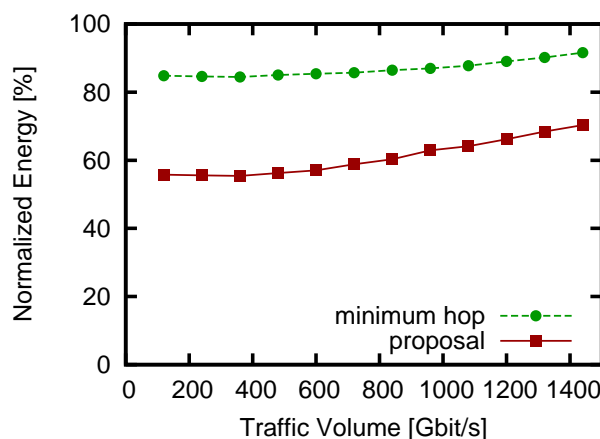
4.2 消費電力の削減効果

まず、本手法がデータセンターネットワークにおいて有効に動作するかを調査する。

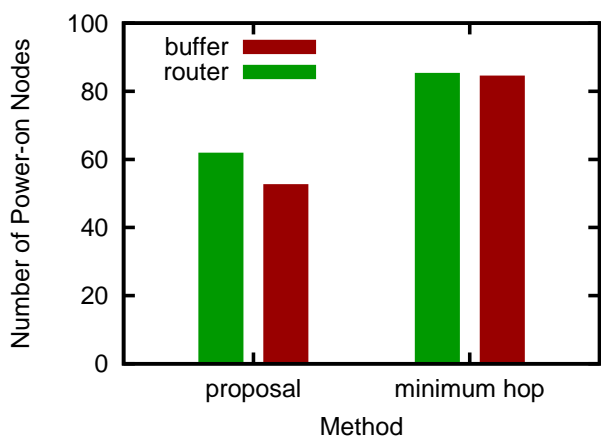
図3(a)に各手法を用いた場合の消費電力およびその内訳(ルータの消費電力およびバッファの消費電力)を、図3(b)に電源が投入されたルータおよびバッファの数を示す。このとき、ルータの消費電力は、バッファの電源が投入されたことによる消費電力の増加分は含まない。図3(a)より、提案手法の消費電力を考慮したトラヒック経路選択により、データセンターネットワークの消費電力を削減できることがわかる。提案手法を用



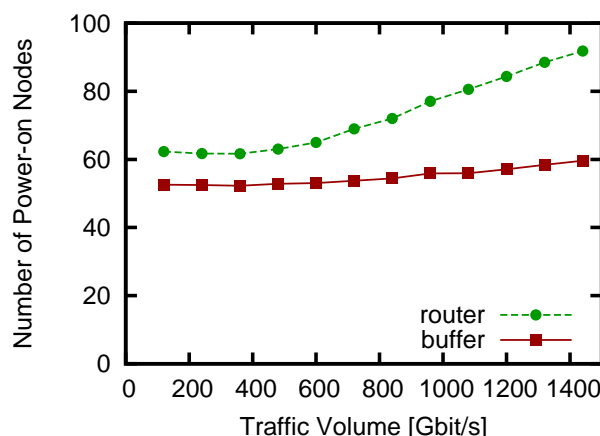
(a) 消費電力



(a) 消費電力削減効果



(b) 電源が投入されたルータ数およびバッファ数



(b) 提案手法により電源が投入されたルータ数およびバッファ数

図3 消費電力を考慮したトラフィック経路選択による消費電力および電源が投入されたルータ数・バッファ数

図4 総トラフィック量が消費電力削減効果および電源が投入されたルータ数・バッファ数に与える影響

表1 ToR ペア間の平均ホップ数および最大ホップ数

	平均ホップ数	最大ホップ数
提案手法	12.77	27
最短ホップ	7.83	18

いることにより、全機器の電源を投入した場合と比べて44%程度、最短ホップ経路を用いて不要な機器の電源を落とした場合と比べて34%程度の消費電力削減が可能である。消費電力が削減できている理由は、図3(b)より、トラフィック収容に必要なルータおよびバッファの数が共に削減できていることで説明できる。特に、最短ホップ経路を用いた場合に必要なバッファ数とルータ数がほぼ等しいのに対し、提案手法を用いた場合に必要なバッファ数が必要なルータ数と比較して少ない。提案手法により光パケットの衝突を回避するようにトラフィック経路が決定しているため、光パケットの衝突が発生しない中継用のルータが多数存在することを示している。その結果、消費電力が大きいバッファの数を削減でき、大きな消費電力の削減を可能にしている。

次に、表1に各手法を用いた場合のトラフィックのホップ数を示す。提案手法はトラフィック収容時に既に電源が投入されているルータやバッファを再利用するために最短ホップ以外の経路

も選択する。そのため、当然の結果であるが、提案手法の方がホップ数が大きい。ただし、提案手法を用いた場合であっても、ネットワークの性能要件(27ホップ)は満たしている。

4.3 トラフィックパタンの影響

トラフィックパターンにより電源の投入が必要なルータやバッファの数が変化することが予想される。そのため、さまざまなトラフィックパターンにおける提案手法の有効性を明らかにする。

そこで、総トラフィック量が消費電力削減に与える影響を評価する。トラフィックパタンの変動により、常に通信頻度の高い・通信量の大きいサーバを同一サーバラックに格納できるとは限らない。そのため、サーバがネットワークを介して大きなトラフィックの通信している状況を想定した評価を行なう。図4(a)に、総トラフィック量を変化させた場合の消費電力の削減効果を示す。消費電力の削減効果は、全機器の電源を投入した場合と比較した消費電力を意味する。なお、総トラフィック量によらず、ネットワークを介して通信しているToRペア数は全体の30%とした。

結果より、総トラフィック量が増加するにつれ、消費電力の削減効果が低下することが分かる。これは、トラフィック量が増加したことにより複数のトラフィックを同一経路上に収容すること

が困難になり、必要なルータ数が増加したことが原因である。参考として、図 4(b) に、総トラフィック量を変化させたときの電源を投入されたルータおよびバッファの数を示す。図より、総トラフィック量が増加するにつれ、必要なルータ数が増加していることが分かる。一方、総トラフィック量が増加したとしても、バッファ数の増加は抑えられており、提案手法によるバッファ数削減が有効に働いていることが分かる。

このことから、トラフィック量が増加したとしても消費電力削減効果を得ることはできるが、より消費電力を抑えるために、通信頻度の高い・通信量の大きいサーバを同一サーバラックまたは同一 ToR スイッチ下に格納し、総トラフィック量を抑えることが望ましいことが分かる。

5 まとめと今後の課題

本稿では、光通信技術と電気技術を融合した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における低消費電力化制御を提案した。簡単な評価の結果、グリッドトポロジのネットワークにおいて、提案手法により、不要な機器の電源を落とさない場合と比較して最大 44 % 程度、最短ホップ経路にトラフィックを収容して不要な機器の電源を落とした場合と比較して最大 34 % 程度消費電力の削減が可能であることを示した。

今後の課題として、多様なデータセンター環境における提案手法の有効性を明らかにし、さらに大規模なネットワークにも適用可能なように手法の改善を行う予定である。

謝 辞

本研究は情報通信研究機構 (NICT) の委託研究「高機能光電子融合型パケットルータ基盤技術の研究開発」の成果による。

文 献

- [1] 大下裕一, 村田正幸, “[招待講演] データセンターネットワークの研究動向,” 電子情報通信学会技術研究報告 (IN2012-28), pp. 25–30, June 2012.
- [2] N. Farrington, et al., “Helios: a hybrid electrical/optical switch architecture for modular data centers,” in *Proceedings of ACM SIGCOMM Computer Communication Review*, pp. 339–350, Oct. 2010.
- [3] H. J. Chao and K. Xi, “Bufferless optical cros switches for data centers,” in *Proceedings of OFC*, Mar. 2011.
- [4] “低消費電力・低遅延高機能光電子融合型パケットルータに必要な基盤技術の研究開発及び低消費電力・低遅延高機能光電子融合型パケットルータの応用技術の研究開発.” http://www.nict.go.jp/collabo/commission/k_151ai.html.
- [5] M. Zhang et al., “GreenTE: Power-aware traffic engineering,” in *Proceedings of the The 18th IEEE International Conference on Network Protocols (ICNP’ 10)*, pp. 21–30, Oct. 2010.
- [6] F. Vismara et al., “On the energy efficiency of IP-over-WDM networks,” in *proceedings of IEEE Latin-American Conference on Communications (LATINCOM)*, pp. 1–6, Sept. 2010.
- [7] G. Shen et al., “Energy-minimized design for IP over WDM networks,” *IEEE/OSA Journal of Optical Communications and Networking*, pp. 176–186, June 2009.
- [8] F. Idzikowski et al., “Saving energy in IP-over-WDM networks by switching off line cards in low-demand scenarios,” in *Proceedings of the 14th Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–6, Feb. 2010.