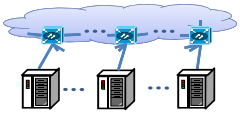
光技術と電気技術の融合によるデータセンターネットワーク

大阪大学 大学院情報科学研究科
〇大下 裕一
西島 孝通
小泉 佑揮
村田 正幸

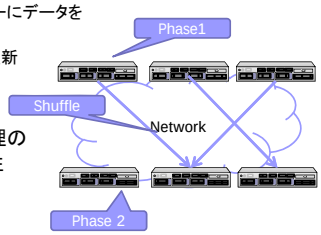
データセンターネットワーク

- 多数のサーバーとサーバー間のネットワークで構成
  - サーバーで連携をとることにより、多量のデータを処理
  - 一か所のデータセンターで数万台のサーバーを接続するものも設置
- ネットワークが性能に大きな影響を与える
  - ネットワークの遅延や帯域不足がサーバー間の連携を阻害

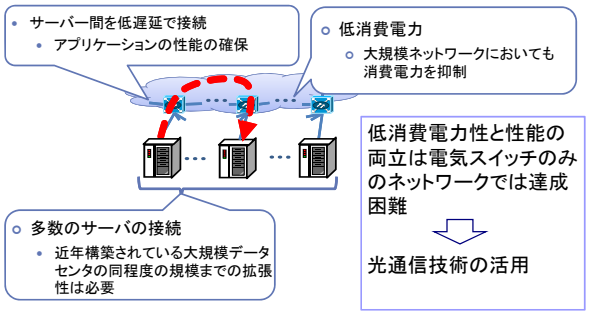


データセンター上のアプリケーションの一例

- データセンター内では、サーバー間の連携によって多量のデータを処理
  - 例：サーチエンジンのバックグラウンド
  - Phase 1:　収集したWEBのキーワードを識別
  - Shuffle: 対応するサーバーにデータを送る
  - Phase 2:データベースを更新
- サーバー間の転送が処理のボトルネックになる可能性



データセンターネットワークの要求

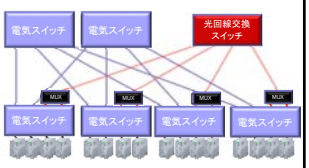


光通信技術を用いたデータセンターネットワーク

- 光スイッチを用いることにより、以下を達成
  - 低消費電力：
    - 電気の処理を行うよりも、低消費電力なネットワークを構成可能
  - 広帯域・低遅延：
    - 光スイッチの広帯域・低遅延性を生かした接続
- 構成
  - 光回線交換スイッチを用いた構成
  - 光パケットスイッチを用いた構成

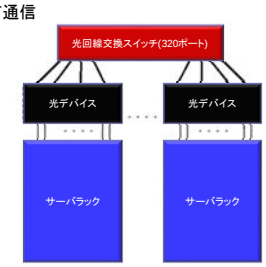
光回線交換スイッチを用いた構成(1)　Helios

- 構成
  - コアスイッチに電気パケットスイッチと光回線交換スイッチを配置
  - トラヒックが多い地点間を光パスで接続
  - それ以外の地点間のトラヒックはコアの電気パケットスイッチを経由して転送
- メリット
  - 少ないコストで広帯域・低遅延の通信経路を確保
- デメリット
  - パスの切り替え時間
    - 頻繁にトラヒック状況が変わる場合に対応不可



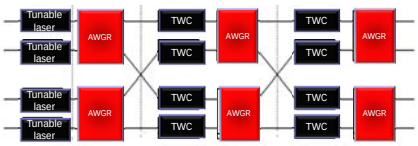
光回線交換スイッチを用いた構成(2)　Proteus

- 構成
  - 全サーバラックを大きな光回線交換スイッチに接続
  - 光パスで論理トポロジを構築
  - 全通信は論理トポロジを経由して通信
- メリット
  - 少ないコストで広帯域・低遅延の通信経路を確保
- デメリット
  - パスの切り替え時間
    - 頻繁にトラヒック状況が変わる場合に対応不可



光パケットスイッチを用いた構成

- データセンター向け大規模光パケットスイッチを開発
  - 構成
    - AWGRを多段で構成
    - AWGR：入射する波長によって出口ポートが決まる
    - TWC：入力波長を指定した波長に変換→変換することにより、経路を変えて制御
    - 内部にバッファはなく、集中制御により衝突を防止
  - デメリット：
    - ネットワーク規模が大規模化・トラヒック量が増大すると集中制御が困難に
    - バッファはToRスイッチ任せ

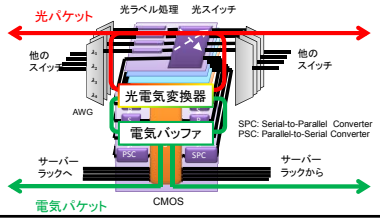


データセンターネットワークへの光技術の適用

- 光回線交換スイッチ＋電気パケットスイッチ
  - 広帯域の通信を光パスで収容
  - 回線交換スイッチの切り替え時間が問題となる場合も
- 全光パケットスイッチ
  - 中でも、バッファーレスなスイッチの研究が盛ん
    - ToRスイッチでのバッファを活用
  - ただし、タイミングの集中制御が必要
- 光電子融合型パケットスイッチ
  - 光スイッチ・電気バッファの組み合わせ
  - パケットの衝突がなければ、光信号のまま中継
  - パケットの衝突が起きたら、電気バッファを活用

光電子融合型パケットスイッチ

- 光通信技術と電気技術を融合
  - 光ポートと電気ポートを持つ
    - 光ポートを用いて他のスイッチと接続
    - 電気ポートを用いてサーバラック内の電気スイッチと接続
  - 光パケットと電気パケットを変換するための変換器・電気バッファを持つ

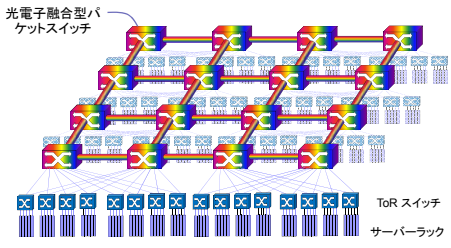


光電子融合型パケットスイッチの特徴

- 光通信技術の高性能・低消費電力を維持したまま、光通信技術における制御や実現性の問題を解決
  - パケットの衝突が発生しない場合、光/電気変換が不要で、光パケットをそのまま中継可能
    - ⇒ 低遅延・低消費電力で通信可能
  - パケットの衝突が発生する場合、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
    - ⇒ 大容量光バッファの実現やパケットの衝突回避の制御が不要

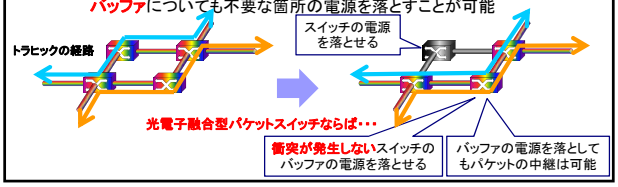
光電子融合型パケットスイッチを用いたデータセンターネットワーク

- データセンター内のコアネットワーク
  - 各光電子融合型パケットスイッチが多数のサーバラックからの通信を束ねて転送するネットワーク構造



トラヒック経路選択による低消費電力化制御

- 従来のネットワークの低消費電力化制御
  - 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
    - IP スイッチの電源や、NIC のポート、光電気変換器など
- 光電子融合型パケットスイッチネットワークの低消費電力化制御
  - 発生したトラヒックの経路を選択する際、スイッチの電源のみではなく、**バッファ**についても不要な箇所の電源を落とすことが可能



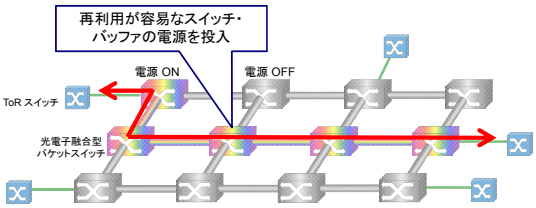
データセンター内経路制御

トラヒック経路選択により
遅延の性能要件を満たした上で消費電力を削減

- 課題
  - 必ずしもバッファの電源を落とすことが遅延および消費電力の削減に繋がるとは限らない
- アプローチ
  - 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**ことで、既に電源が投入されているスイッチおよびバッファを効率的に利用
  - **再利用が容易な**スイッチ・バッファの電源を優先的に投入

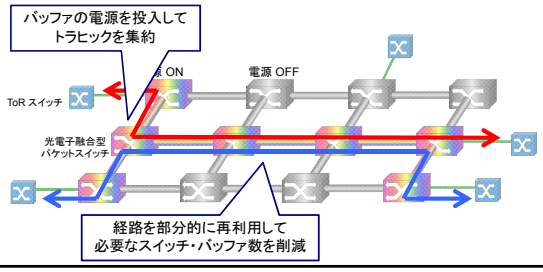
低消費電力なトラヒック経路選択手法のアイデア

- **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、電源の投入が必要なスイッチ・バッファの総消費電力を最小化
  - ➡ 多くのトラヒックで利用可能なスイッチおよびバッファのみ電源を投入



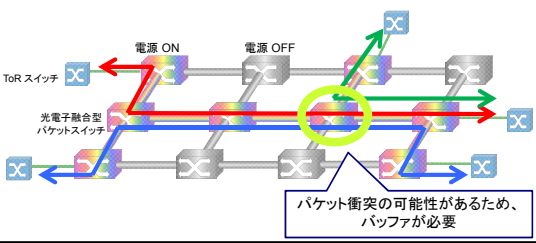
低消費電力なトラヒック経路選択手法のアイデア

- **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、電源の投入が必要なスイッチ・バッファの総消費電力を最小化
  - ➡ 多くのトラヒックで利用可能なスイッチおよびバッファのみ電源を投入



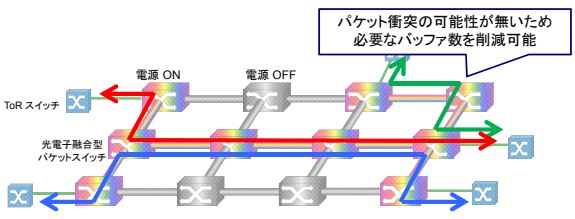
低消費電力なトラヒック経路選択手法のアイデア

- **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、電源の投入が必要なスイッチ・バッファの総消費電力を最小化
  - ➡ 多くのトラヒックで利用可能なスイッチおよびバッファのみ電源を投入



低消費電力なトラヒック経路選択手法のアイデア

- **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、電源の投入が必要なスイッチ・バッファの総消費電力を最小化
  - ➡ 多くのトラヒックで利用可能なスイッチおよびバッファのみ電源を投入



スイッチ・バッファの利用される可能性を調べる指標

- 再利用容易性

ノード n を通るトラヒック量の期待値に相当

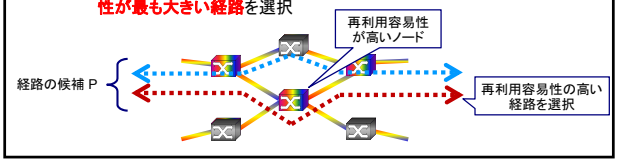
$$\text{ノード } n \text{ の再利用容易性} = \sum_{s,d} \frac{N_{s,n,d}}{N_{s,d}} \lambda_{s,d}$$

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

- 再利用容易性が高いノード程、他のトラヒックの経路でも利用されやすい

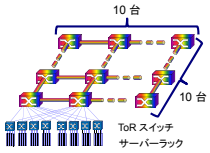
再利用容易性を用いた低消費電力なトラヒック経路選択手法

- 物理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
  - 短時間で経路を決めるため、最適化問題は用いない
- エンド間のトラヒックの経路決定の手順
  1. 性能要件を満たすエンド間の経路の候補 P を取得
  2. 経路の候補 P の内、新たに電源の投入の必要があるスイッチ・バッファの総消費電力が最小の候補を選択
  3. 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



評価方法

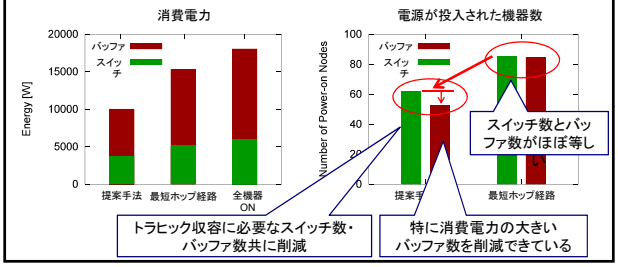
- 評価対象
  - 提案手法
  - 最短ホップ経路にトラヒックを収容し、不要な機器の電源を落とす手法
  - すべての機器の電源を投入する手法
- 評価条件
  - ネットワーク
    - 10 スイッチ×10 スイッチのグリッド
    - リンク帯域 ： 10 Gbps
  - トラヒックパタン
    - 総トラヒック量 ： 240 Gbps
    - 全体の 30 % の ToR 間でパレート分布に従うトラヒックが発生
  - 性能要件
    - ToR 間 27 ホップ以下（最大ホップ数の 1.5 倍までを許容）



消費電力の削減効果

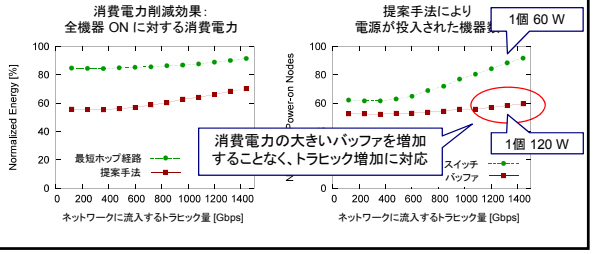
スイッチの消費電力
- バッファの電源を落とした場合 ：60 W
- バッファの電源を投入した場合 ：180 W（追加で 120 W 消費）

- 最短ホップ経路と比較して、34 % 程度の消費電力削減
  - 提案手法を用いた場合、スイッチやバッファを再利用するために最短ホップ経路以外の経路も選択するため、平均ホップ数は増加
  - ただし、最大ホップ数は遅延の制約条件を満足



ネットワークに流入するトラヒック量の影響

- トラヒック量によらず消費電力を削減
  - トラヒック量が大きくなるにつれ削減効果が低下
  - 消費電力削減効果を高めるためには、トラヒック量が大きいエンド間を同一サーバラックに配置するなどのトラヒック量を削減する工夫が有効



まとめと今後の課題

- データセンタネットワークにおいて、高通信性能と低消費電力の両立の要望
- 光電子融合型パケットスイッチネットワーク上の低消費電力化制御を提案
  - 遅延の性能要件を満たした上で消費電力を削減するようにトラヒック経路を選択
  - 簡単な評価により、提案手法が遅延の制約条件を満たした上で消費電力を最大 34 % 削減可能なことを確認
- 今後の課題
  - 多様なデータセンター環境における性能評価