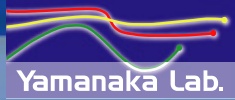




HOLST: タイムスロット型切り替えを行う 超高速光スイッチを用いた低電力DCN



Yamanaka Laboratory, Keio University, Japan

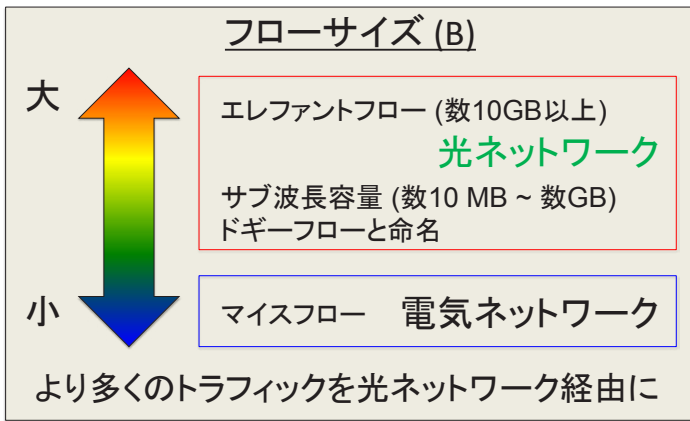
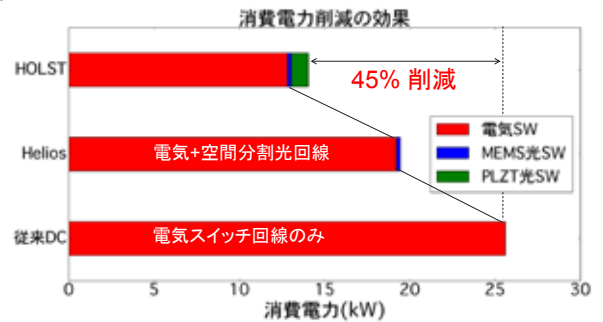


HOLST

HOLST: High-speed optical layer 1 switch system for time slot switching based optical data center networks
組曲『惑星』の作者Holstに由来

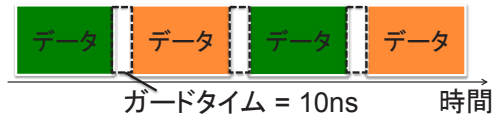
- ◆ ナノ(10^{-9})秒級の高速切替が可能な光スイッチを導入した、データセンターネットワーク
高速切替光スイッチ導入の効果
- 従来の空間分割に加えて時分割多重(TDM)光回線を利用可能に
- ◆ サブ波長容量のフローをTDMによって光回線収容
従来は、電気スイッチ回線へ収容
→ **電気スイッチを削減し、一層の省電力化**
- ◆ 大容量フローを発生させるデータセンターアプリケーション基盤と連携したネットワーク制御により
光ネットワークを有効活用

TDM: Time Division Multiplexing

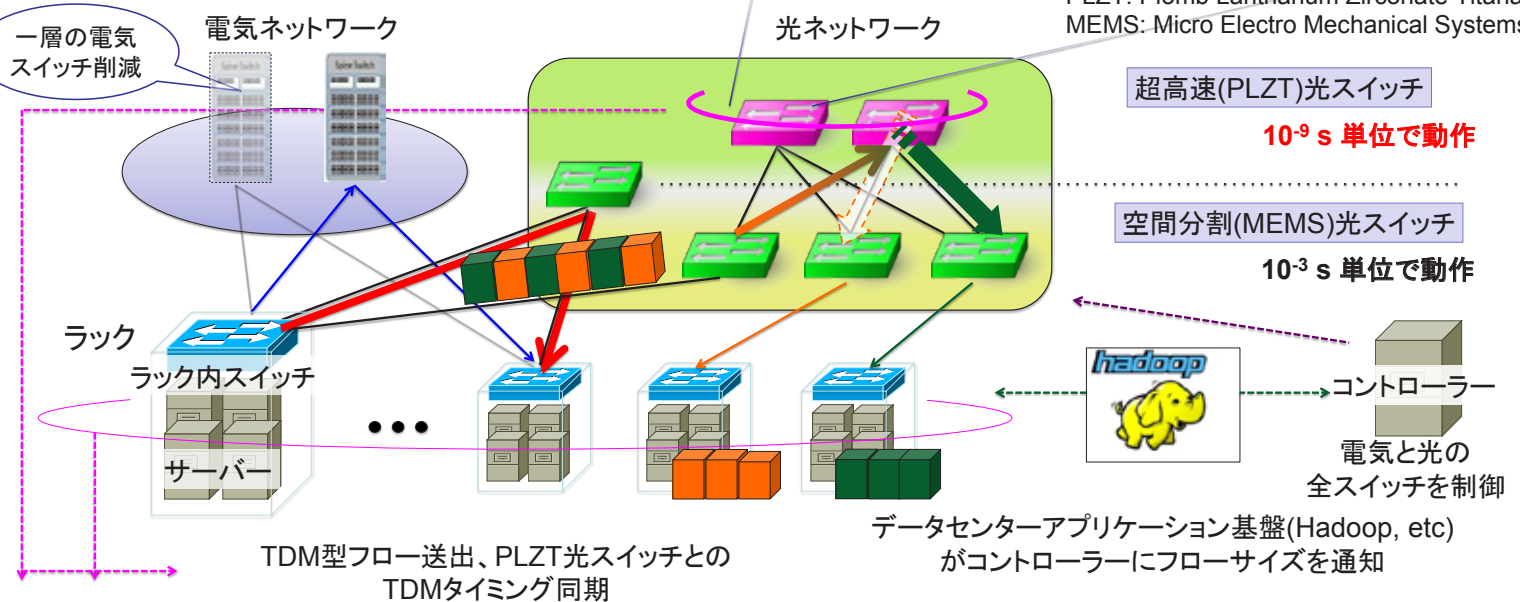


高速切替スイッチの利点:
サブ波長容量でも、ガードタイムの影響を受けずに、通信容量を最大限活用

(サブ波長を束ねてフル波長容量を活用)



PLZT: Plomb Lanthanum Zirconate Titanate
MEMS: Micro Electro Mechanical Systems



研究者名

理工学部 情報工学科 教授 山中 直明 (Yamanaka Naoaki)

お問合せ先

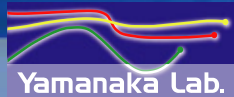
Mail : yamanaka@ics.keio.ac.jp

URL : <http://www.yamanaka.ics.keio.ac.jp>

This work is supported by New Energy and Industrial Technology Development Organization (NEDO) of Japan.



HOLSTにおける TDM通信設定で行う同期手法

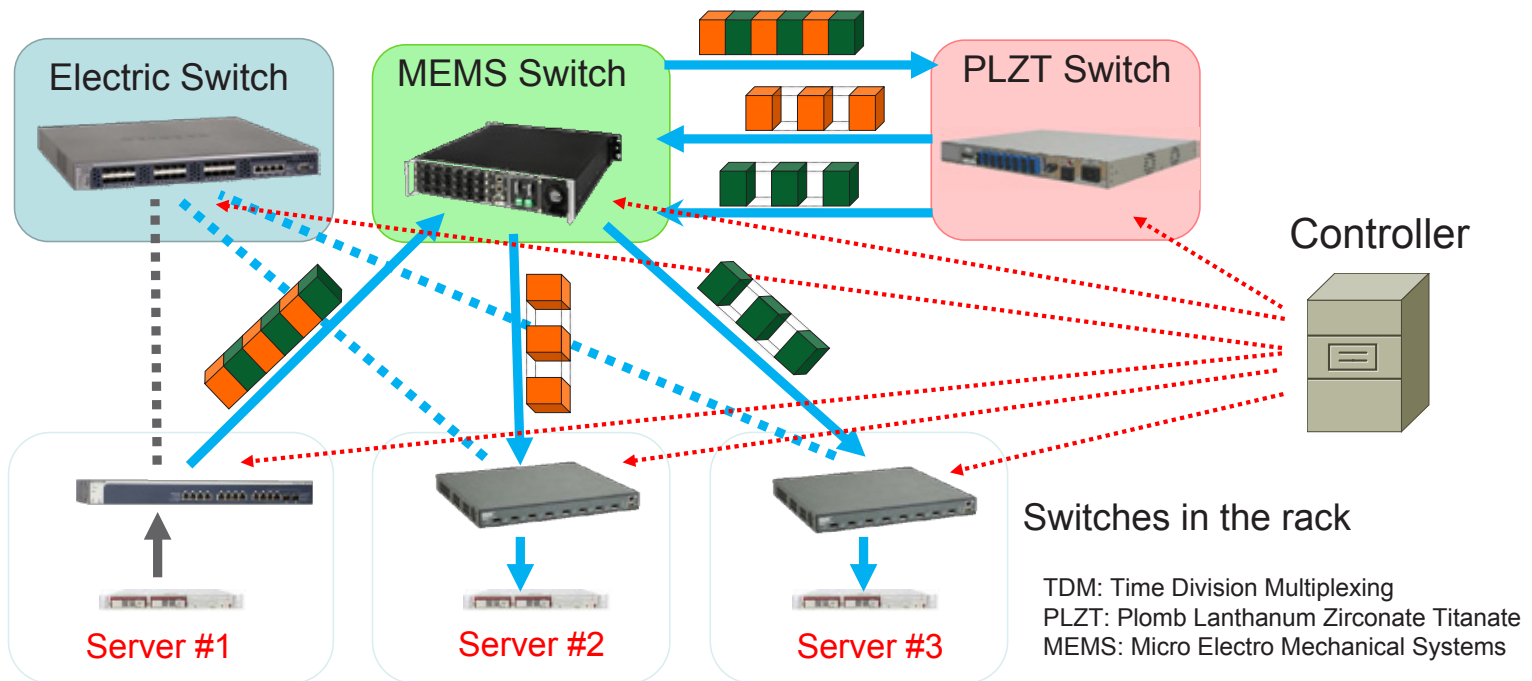


Yamanaka Laboratory, Keio University, Japan

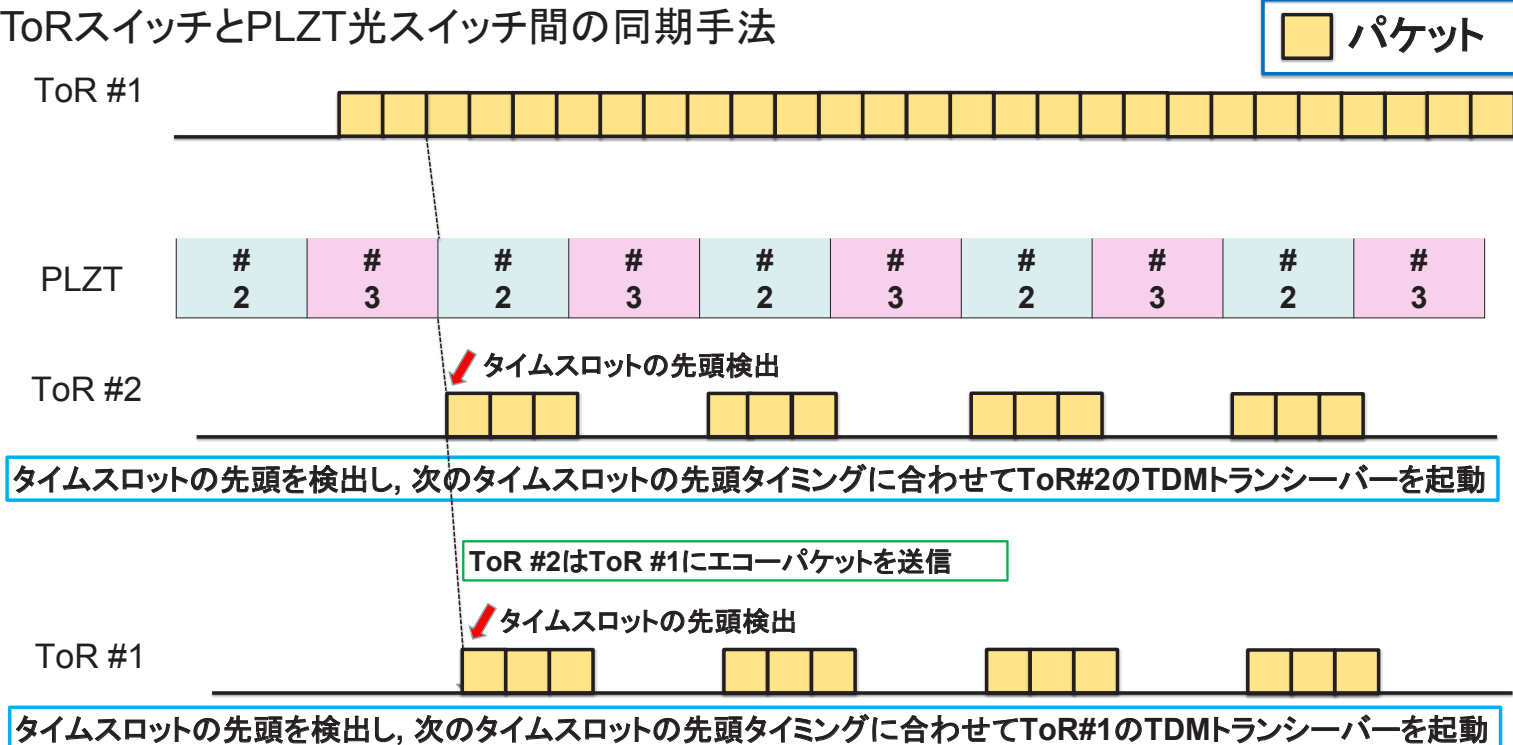


超高速光L1スイッチ(PLZT光スイッチ)は, HOLSTでTDM通信を行うことを可能にしたが,
TDM通信を行う際, 各ToRスイッチとPLZT光スイッチ間で同期を行うことが必要

◆複数の宛先に向けた, TDMパケット送信



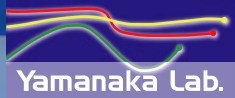
◆ToRスイッチとPLZT光スイッチ間の同期手法



この方法により, ToRスイッチとPLZT光スイッチ間の
同期を行った上で, 双方向のTDM通信が可能



トラヒックパターンを考慮した データセンター指向フロー判定手法

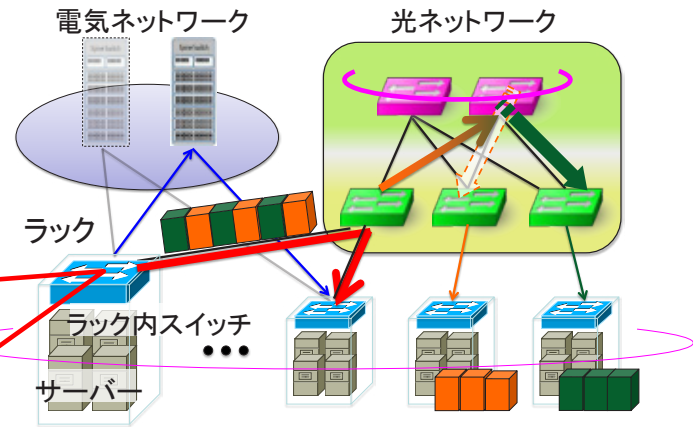
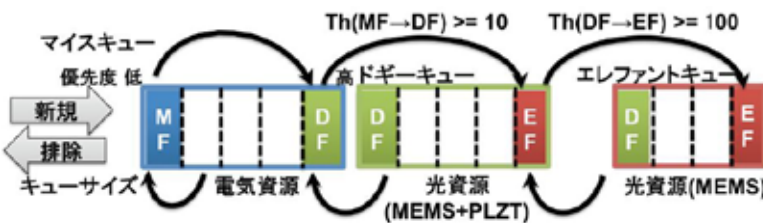


Yamanaka Laboratory, Keio University, Japan



HOLSTでは電気スイッチ・MEMS光スイッチ・PLZT光スイッチを組み合わせて利用しており、フローの種類(マイスフロー・ドギーフロー・エレファントフロー)によって利用するスイッチが異なる
また、データセンターネットワークでは高頻度にフローの発着が発生するため高速なフロー判定が必要

階層型キューを用いたフロー判定手法



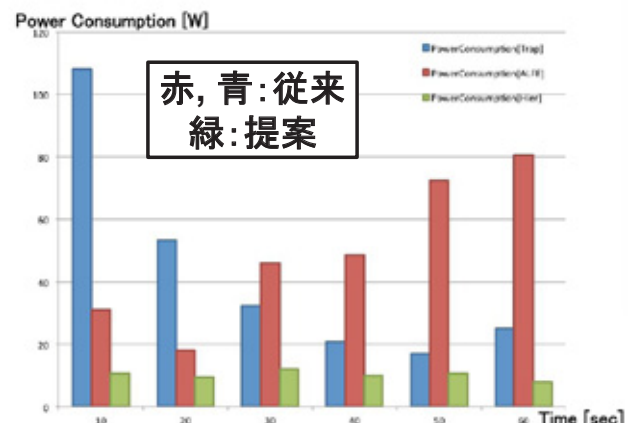
ToR(Top of Rack)スイッチ内で判定

◆フロー判定手順

- ① パケット到達時にFlow IDを取得し、該当Flow IDがLRUキュー内に存在するか確認
- ② ①の結果に応じて以下の操作を実行
 - I. 新規のFlow IDの場合
マイスキューの先頭に挿入
 - II. 既存のFlow IDの場合
そのFlow IDの参照回数(カウンタ)の値を1増やす
カウンタの値が閾値を超えた場合、上位のキューへ遷移
- ③ キュー内から溢れたFlow IDは下位のLRUキューの先頭に遷移
マイスキューから追い出されたFlow IDは、排除される

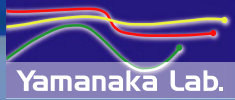
※LRU: Least Recently Used

既存のフロー判定手法を利用した場合と比較して高い省電力効果を得ることに成功





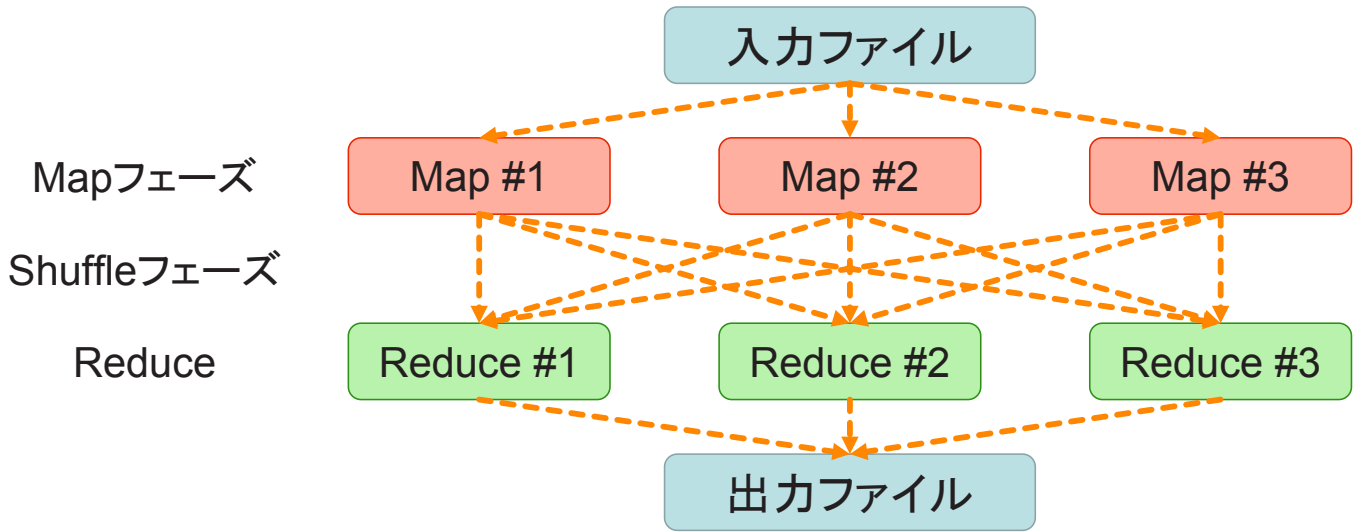
HOLSTにおけるHadoopを用いた Application Drivenな光回線設定



Yamanaka Laboratory, Keio University, Japan



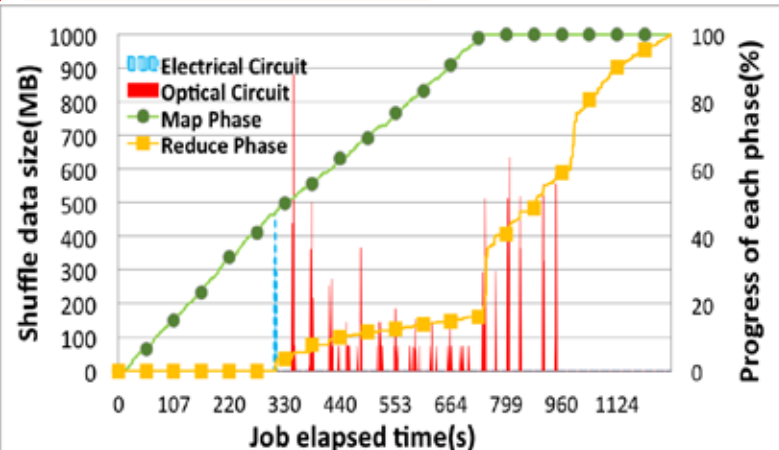
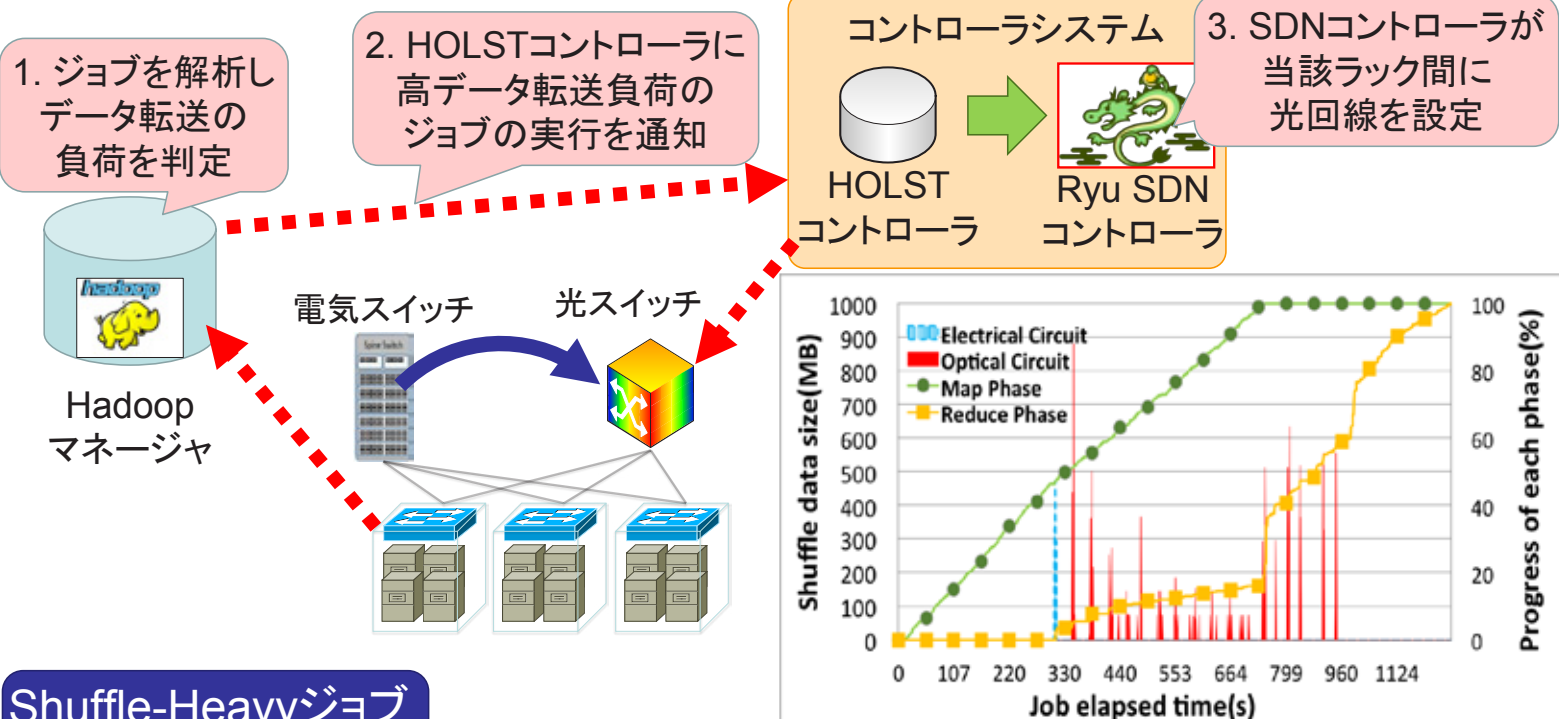
Hadoop : MapReduceプログラミングモデルを用いた並列分散処理用ミドルウェア



MapタスクとReduceタスクを各サーバで分散して実行

HOLSTにおけるApplication Drivenな光回線設定

ジョブの種類によって、Shuffleフェーズに多量のデータ転送が行われるものが存在 (Shuffle-Heavyジョブ)
 → Shuffle-Heavyジョブの実行時には、HOLSTコントローラがラック間を光回線で接続



TeraSort (Shuffle-Heavy) 実行時の電気/光スイッチ経由データ量

Shuffle-Heavyジョブ (ex. Terasort)