高性能計算システムを支える プログラマブルネットワーク

産業技術総合研究所　高野了成

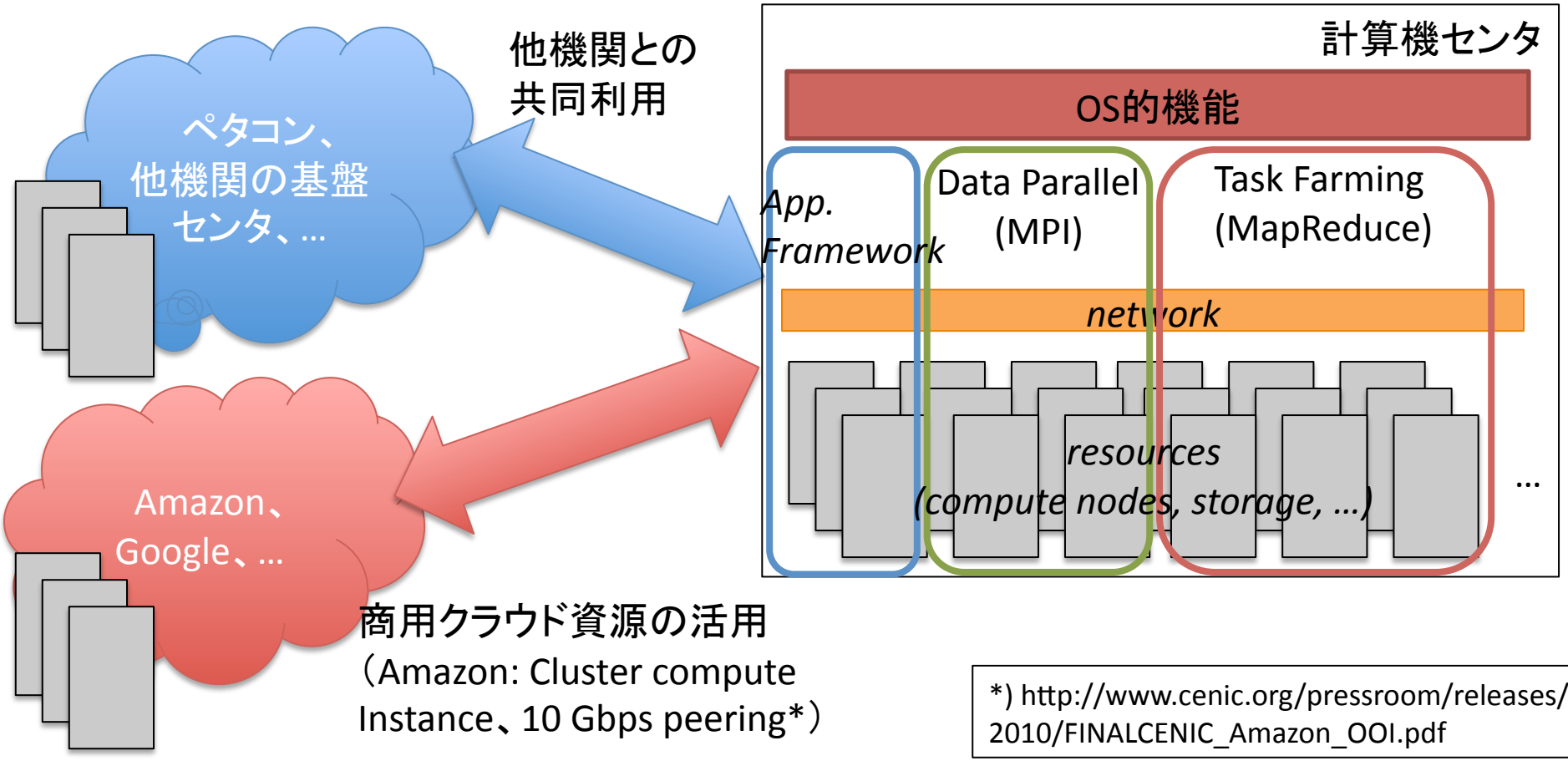
戦略的高性能計算システム開発に関するワークショップ
2010年8月2日＠金沢

自己紹介

- 名前：高野　了成
- 所属：産総研　情報技術研究部門
- 経歴：
  - 2003年　株式会社アックス 入社
  - 2005年　東京農工大学大学院博士後期課程満期退
  - 2008年　東京農工大学大学院 博士（工学）
  - 2008年　産業技術総合研究所 入所
- 主な研究：
  - ストリーミングサーバ向けキャッシュ技術、グリッド環境向け並列計算ライブラリ、広域TCP/IP通信の高速化、パスプロビジョニングの相互運用技術

計算機センタ as a Computer

- センタ全体の資源を有効利用するOS的機能（資源管理、スケジューリング、セキュリティ）が必要



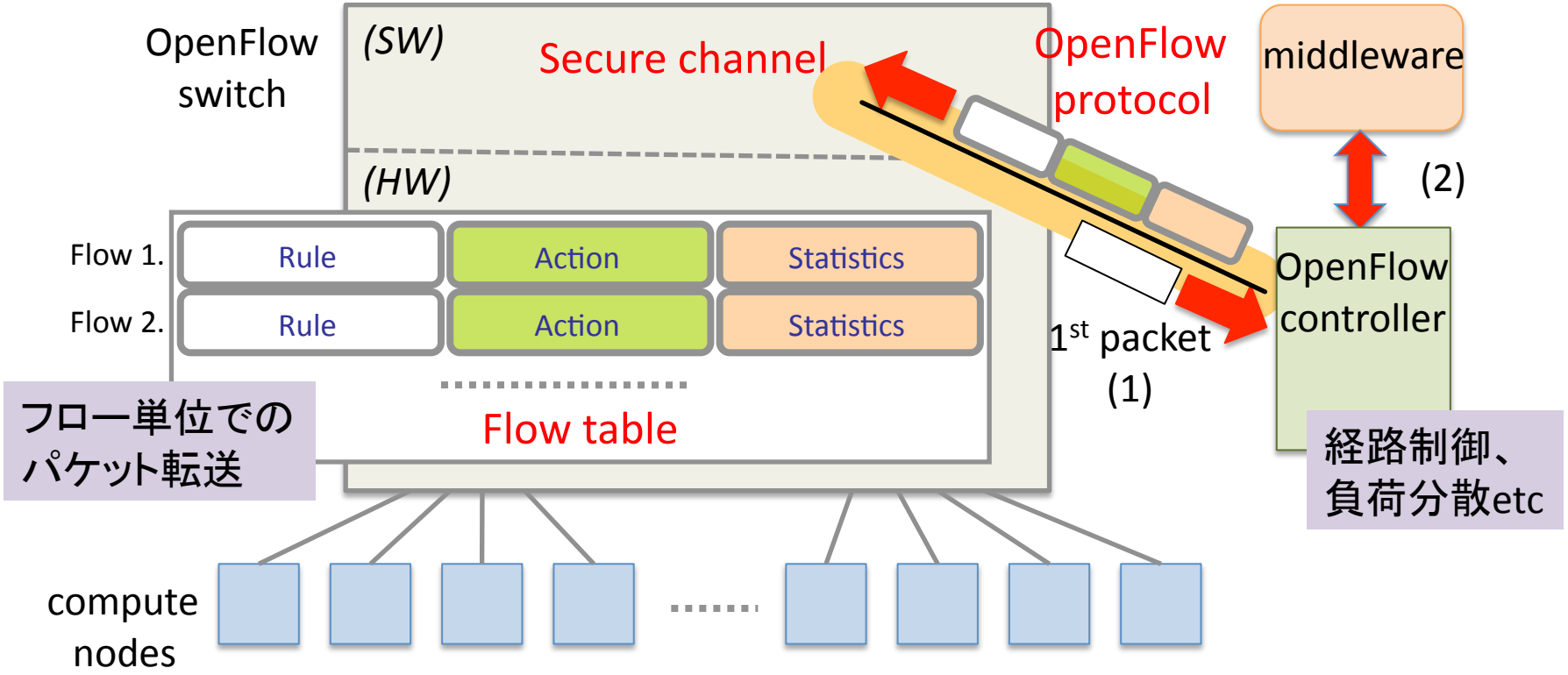
*) http://www.cenic.org/pressroom/releases/2010/FINALCENIC_Amazon_OOI.pdf

計算機センタにおけるNW仮想化

- 計算機センタとデータセンタの違いはソフトウェア
  - コモディティ部品を使えば高コスト性能比
    - 例）CEE（Converged Enhanced Ethernet）
  - システムソフトウェアスタック、ネットワーク構成
- ジョブ毎に最適化されたネットワークを構築
  - ジョブのトラフィック性質に合わせた最適化
    - 例）VLANルーティング法によるトポロジルーティング
  - ジョブやテナントごとの性能隔離
    - トラフィック調停：TCP incast（輻輳）問題の回避
- 管理コストの軽減
  - スイッチの管理設定の集中化
  - さらにルールに基づいた自律的制御

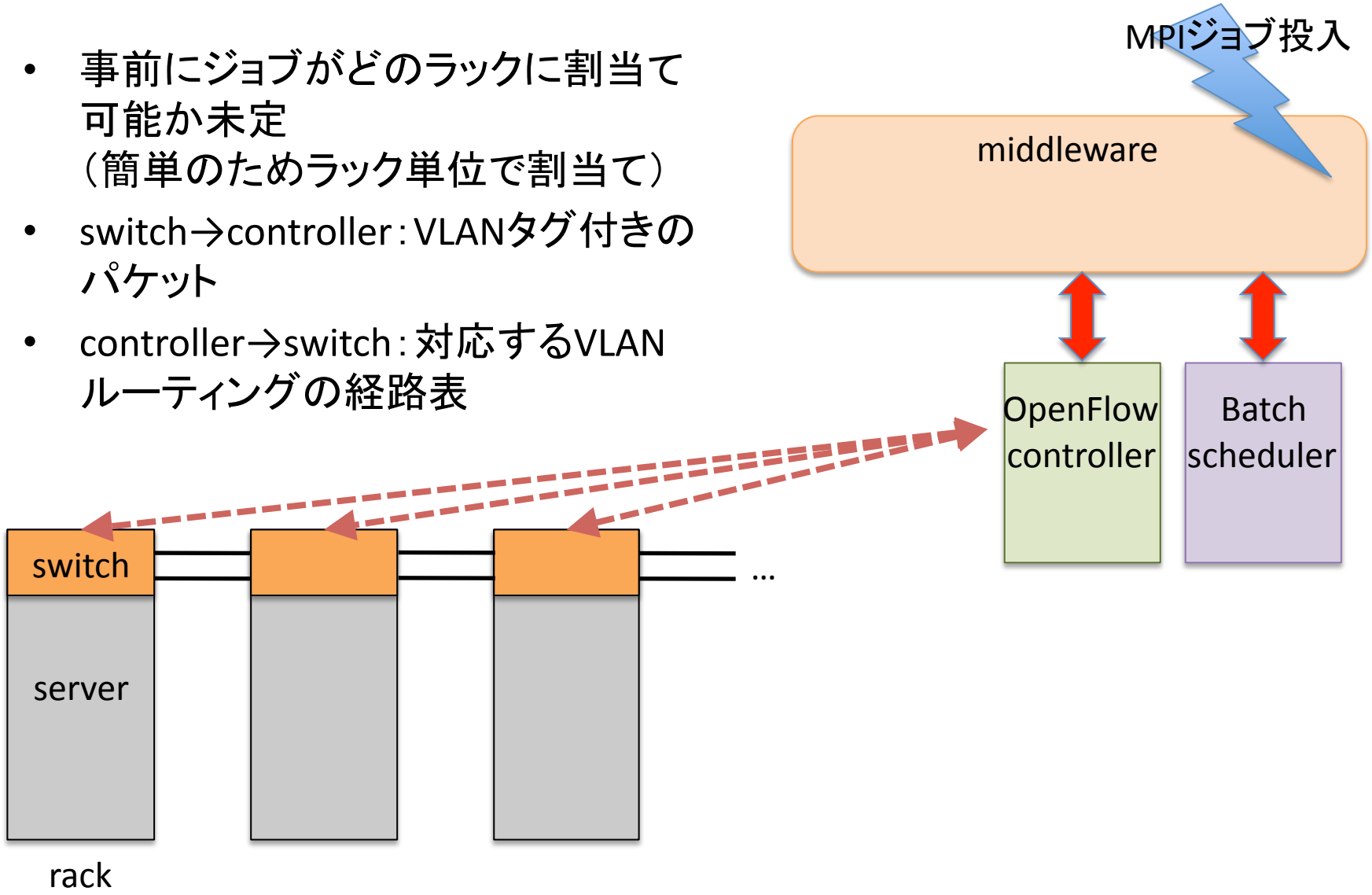
OpenFlow

- フロースイッチング機能（データプレーン）と経路制御機能（コントロールプレーン）をOpenFlowプロトコルにより分離
- コントローラが、(1)スイッチからの未定義フローの通知、または(2)連携ミドルウェアからの経路変更要求をトリガに、フローテーブルを設定



例）VLANルーティング法

- 事前にジョブがどのラックに割当て可能か未定
（簡単のためラック単位で割当て）
- switch→controller：VLANタグ付きのパケット
- controller→switch：対応するVLANルーティングの経路表



課題

- ルール＆アクションのテンプレート化、ノウハウの蓄積、共有が重要

- 大規模化にはコントローラがボトルネックに
- 10/40/100 Gigabit Ethernet、Infiniband対応
  - (GtrcNET-10上にOpenFlowを実装済み)

まとめ

- ユーザ、ジョブ毎に最適化されたネットワークを動的に構築し、管理コストも軽減したい
- センタ規模でもOpenFlowなどのプログラマブルネットワーク技術は有効ではないか

- もちろんネットワークと計算機、ストレージの連携も重要＝センタOS？