

## Virtual Network Control to Achieve Low Energy Consumption in Large-scale Optical Networks

大阪大学 大学院情報科学研究科  
情報ネットワーク学専攻  
村田研究室  
樽谷 優弥

Advanced Network Architecture Research Group  
<http://www.anarg.jp/>

### 研究背景

- ネットワークの大規模化と通信帯域の増大
- 消費電力が増大

#### 消費電力削減の方法

- 通信需要が時間によって変動
- ネットワークトポロジーを通信需要に応じて変更
  - 最小限の機器でネットワークを運用できるように変更

2014/06/24 2

### トポロジーを変更可能なネットワークアーキテクチャ

- 構成
  - コアネットワーク：光スイッチと光ファイバで構成
  - 電気スイッチが光スイッチに接続
- 光スイッチの設定により、電気スイッチ同士を接続する光パスを構築可能
- 光パスを多数構築することにより、仮想ネットワークを構築可能
- 光スイッチの設定変更により、仮想ネットワークのトポロジーを変更可能

2014/06/24 3

### 仮想ネットワーク制御の方法

- 制御サーバーが現在のトラフィック量を収集
  - 全てのスイッチから接続するリンクのトラフィック量を収集
- 収集したトラフィック量を用いて、適切なネットワークトポロジーを計算
  - 指標：リンクの輻輳、通信遅延、電力
- 制御メッセージを送信し、光パスの張替えを行う

2014/06/24 4

### 仮想ネットワーク制御の課題

- 制御サーバーが問い合わせるスイッチ数の増加によるトラフィック情報の収集負荷の増加
- ネットワークトポロジーの大規模化による適切なネットワークトポロジーを得るための計算時間の増加
  - 適切なネットワークトポロジーへの移行に時間がかかり、トラフィックの変化に対応できない

2014/06/24 5

### 研究目的

省電力を達成する仮想ネットワーク制御

仮想ネットワーク制御に必要な情報の収集負荷の増大へ対応

短い計算時間で省電力な仮想ネットワークを得る方法の提案

#### 研究目的

- 制御サーバーの情報収集負荷削減する方法の提案(Chapter2)
- 省電力を達成する仮想ネットワークトポロジー制御の提案(Chapter3)
- 耐故障性を保証した省電力な仮想マシンの配置の提案(Chapter4)

2014/06/24 6

### 博士論文の構成

- Chapter 1: Introduction
- Chapter 2: Virtual Network Reconfiguration with Link load Estimation for Large-Scale Optical Network  
制御サーバーの情報収集負荷削減方法の提案
- Chapter 3: Virtual Network Reconfiguration for Reducing Energy Consumption in Optical Data Centers  
省電力を達成する仮想ネットワークポロジ制御の提案
- Chapter 4: Virtual Machine Placement Considering the Redundancy of Data in Multi tenant Data Centers  
耐故障性を保証した省電力な仮想マシンの配置方法の提案
- Chapter 5: Conclusion

### Chapter 2:

### Virtual Network Reconfiguration with Link load Estimation for Large-Scale Optical Network

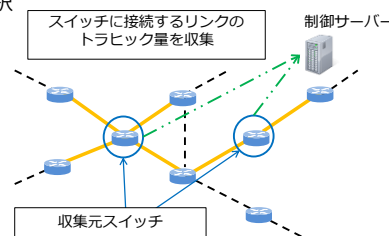
- Yuya Tarutani, Yuichi Ohsita, Shin'ichi Arakawa, and Masayuki Murata, "Optical-Layer Traffic Engineering with Link Load Estimation for Large-Scale Optical Networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 1, pp. 38-52, January 2012.
- Yuya Tarutani, Yuichi Ohsita, Shin'ichi Arakawa, and Masayuki Murata, "Estimation of traffic amounts on all links by using the information from a subset of nodes," in *Proceedings of The Second International Conference on Emerging Network Intelligence (EMERGING2010)*, pp. 18-23, October 2010.
- Yuya Tarutani, Yuichi Ohsita, Shin'ichi Arakawa, and Masayuki Murata, "Selecting monitored links for estimating all link utilizations," *Student Workshop on IEICE Technical Committee on Photonic Network*, March 2010 (in Japanese)
- Yuya Tarutani, Yuichi Ohsita, Shin'ichi Arakawa, and Masayuki Murata, "Estimation of Traffic Amounts on All Links by Using the Information From a Subset of Nodes," *Technical Report of IEICE (PN2010-27)*, vol. 110, no. 264, pp. 19-24, November 2010 (in Japanese)
- Yuya Tarutani, Yuichi Ohsita, Shin'ichi Arakawa, and Masayuki Murata, "Optical layer Traffic Engineering using the traffic information from a subset of Nodes," *Student Workshop on IEICE Technical Committee on Photonic Network*, August 2011 (in Japanese)

### Chapter 2 の背景と目的

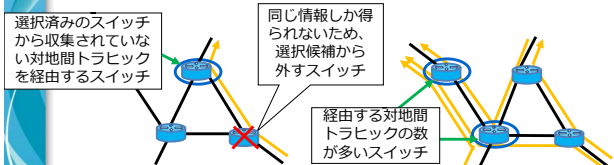
- 背景
  - 仮想ネットワーク制御では、各リンクの輻輳の有無を把握することが必要
  - ネットワークが大規模化すると、リンクのトラフィック量を問合わせるスイッチ数が増加
    - 制御サーバーの情報収集による負荷が増大
- 目的
  - 制御サーバーがリンクのトラフィック量を問合わせるスイッチの数を削減する方法の提案

### 制御サーバーの情報収集負荷削減方法

- アプローチ
  - 一部のスイッチから収集したリンクのトラフィック量から全リンクのトラフィック量を推定することでリンクの輻輳を判別
  - リンクのトラフィック量の推定誤差が少なくなる収集元スイッチを選択



### トラフィック量の収集元スイッチ選択



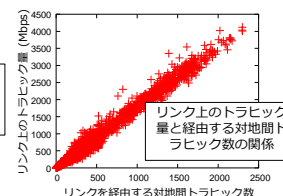
- 選択手法の目標：推定に用いる対地間トラフィックの情報の最大化
- 手順
  - 選択済みのスイッチが収集できない対地間トラフィックが経由するスイッチを選択
    - 全く収集されない対地間トラフィックが存在しないようにするため
  - 経由する対地間トラフィックの数が多きスイッチを選択
    - 多くの対地間トラフィックの情報を取得するため

### 収集していないリンク上のトラフィック量の推定

- リンク上のトラフィック量と経由する対地間トラフィックの数の関係

シミュレーションによって調査

シミュレーション環境  
トポロジ：AT&T トポロジ (523 ノード 1304 リンク)  
トラフィックマトリクス：Gravity モデルと対数正規分布 ( $\sigma = 4.3, \mu = 1.04$ ) に従って生成



シミュレーション結果より関係をモデル化

$$U_i = \alpha Z_i + \beta$$

$U_i$ : リンク  $i$  上の推定されたトラフィック量  
 $Z_i$ : リンク  $i$  を經由する対地間トラフィックの数

- 推定方法
  - 収集したトラフィック量から  $\alpha, \beta$  を算出
  - $Z_i$  に対地間トラフィックの数を代入し、トラフィック量を推定

### リンクのトラフィック量の補正

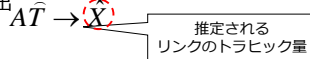
- リンク上のトラフィック量は各リンクを経由する対地間トラフィック量の和から求まるため、以下の関係が成立

$$X = AT$$

各リンク上のトラフィック量  $X$ 、経路情報、トラフィックマトリクス  $T$

- リンクのトラフィック量の補正方法

- トラフィックマトリクスの推定
  - $\|X' - A\hat{T}\|$  を最小化するような  $\hat{T}$  を取得
  - $X'$ : Step. 2 で得られた推定されたトラフィック量と収集したトラフィック量を含む
- 推定されたトラフィックマトリクスと経路情報からリンク上のトラフィック量を算出

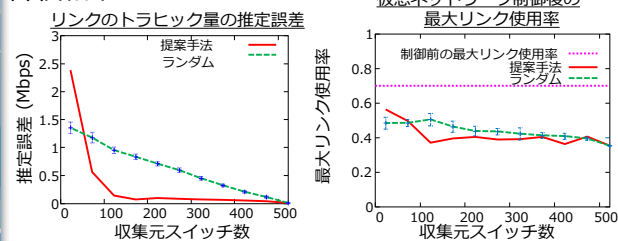


2014/06/24

13

### 評価結果

シミュレーション環境  
トポロジ: AT&T トポロジ (523 ノード 1304 リンク)  
トラフィック: Gravityモデルと対数正規分布( $\sigma = 4.3, \mu = 1.04$ )に従って生成



- 30%の収集元スイッチから接続するリンクのトラフィック量を収集することで、以下が可能なることを確認
  - 少ない推定誤差でリンクのトラフィック量を推定可能
  - 仮想ネットワーク制御 [1] によって最大リンク使用率を抑えることが可能

[1] A. Gencata and B. Mukherjee, "Virtual-topology adaptation for WDM mesh networks under dynamic traffic," *IEEE/ACM Transactions on networking*, vol. 11, pp. 236-247, Oct. 2003. 14

2014/06/24

### 本章のまとめ

- 制御サーバーがリンクのトラフィック量を問い合わせるスイッチの数を削減する方法を提案・評価
  - 一部の収集元スイッチを選択し、全リンクのトラフィック量を推定する手法を提案
  - シミュレーション結果より、提案手法によって制御サーバーが問い合わせる必要があるスイッチ数を70%削減することが可能であることを確認

2014/06/24

15

### 博士論文の構成

- Chapter 1: Introduction
- Chapter 2: Virtual Network Reconfiguration with Link load Estimation for Large-Scale Optical Network  
制御サーバーの情報収集負荷削減方法の提案
- Chapter 3: Virtual Network Reconfiguration for Reducing Energy Consumption in Optical Data Centers  
省電力を達成する仮想ネットワークトポロジー制御の提案
- Chapter 4: Virtual Machine Placement Considering the Redundancy of Data in Multi tenant Data Centers  
耐故障性を保証した省電力な仮想マシンの配置方法の提案
- Chapter 5: Conclusion

2014/06/24

16

## Chapter 3: Virtual Network Reconfiguration for Reducing Energy Consumption in Optical Data Centers

- Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "Virtual Network Reconfiguration for Reducing Energy Consumption in Optical Data Centers," submitted to *IEEE/OSA Journal of Optical Communications and Networking* [conditional accept], January 2014.
- Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "A Virtual Network to Achieve Low Energy Consumption in Optical Large-scale Datacenter," in *Proceedings of IEEE International Conference on Communication Systems (ICCS 2012)*, pp. 45-49, November 2012.
- Yuta Shimotsuma, Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "Evaluation of data center network structures considering routing methods," in *Proceedings of The Ninth Inter-national Conference on Networking and Services (ICNS 2013)*, pp. 146-152, March 2013.
- Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "A Virtual Network to Achieve Low Energy Consumption in Large-scale Datacenter," *Technical Report of IEICE (PN2012-98)*, vol. 111, no. 475, pp. 91-96, March 2012 (in Japanese)
- Yuta Shimotsuma, Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "Evaluation of data center network structures considering routing methods," *Technical Report of IEICE (IN2012-31)*, vol. 112, no. 88, pp. 43-48, June 2012 (in Japanese)
- Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "Virtual Network Configuration Method for Low Energy Consumption in Data Centers," *IEICE General Conference*, March 2014 (in Japanese)

2014/06/24

17

### Chapter 3 の研究背景

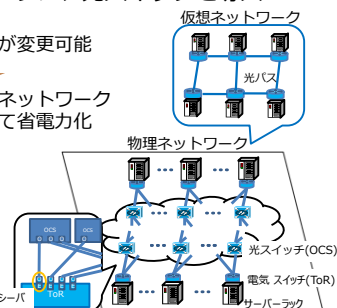
- データセンターにおいてネットワークの省電力化が必要

- データセンターネットワークに、光スイッチを導入

- トラフィック需要に応じて、仮想ネットワークトポロジーが変更可能
- 仮想ネットワークを少ないネットワーク機器で構築することによって省電力化

- 課題

- 機器数が多く、最適な仮想ネットワークを算出するための計算時間が膨大
- データセンター内で短時間で発生するトラフィック変動に対応が困難



### Chapter 3 の目的とアプローチ

- 目的
  - データセンターネットワークに適した省電力を達成する仮想ネットワーク制御を提案
    - 省電力なネットワークポロジ：
      - 稼働するサーバーラックのみを接続
      - 電気スイッチのポート数を最小化
- アプローチ
  - パラメーターによって構造を決定するネットワークポロジを使用し、パラメーターの調節を行うことで短時間で適切なネットワークポロジを算出
  - トラフィック変動は負荷分散手法とネットワークポロジの再構築の2つの方法によって対応

2014/06/24

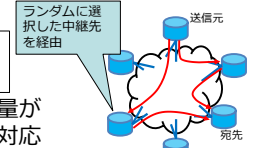
19

### トラフィック変動への対応

- 短時間で発生するトラフィック変動への対応
  - 負荷分散手法を用いたトポロジ構造を構築
    - VLB (Valiant Load Balancing) [2]
      - 宛先によらずランダムに経路する中継先を選択することにより負荷分散を行い、細かい変動に対応
      - ToR 間を流れるトラフィック量は以下で与えられる

$$T_{VLB} = (V_{out} + V_{in}) / E$$

$E$  : ToR 数  
 $V_{out}$  : ToR が送信するトラフィック量  
 $V_{in}$  : ToR が受信するトラフィック量  
 $T_{VLB}$  : VLB を用いた場合における各 ToR 間のトラフィック量



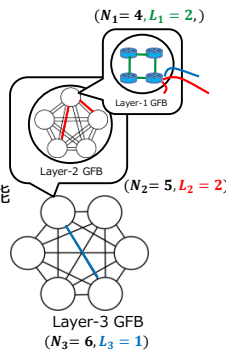
- データセンター内の総トラフィック量が変わるようなトラフィック変動への対応
  - 仮想ネットワークを再構築して対応

[2] M. Kodialam, T. Lakshman, and S. Sengupta, "Efficient and robust routing of highly variable traffic," in Proceedings of HotNets, Nov. 2004. 20

2014/06/24

### GFB (Generalized Flattened Butterfly)

- 構築パラメーターの調整によって、様々なネットワーク構造が構築可能
  - 構築パラメーター
    - $K_{max}$  : 階層数
    - $N_k$  : 階層  $k$  ( $K_{max} \geq k \geq 1$ ) で接続する  $(k-1)$  層の GFB の数
      - $N_1$  は ToR の数
    - $L_k$  : 階層  $k$  で利用する各 ToR あたりのリンク数
  - 構築パラメーターから以下が計算可能
    - ToR 間の最大ホップ数
    - 各リンクのフロー数



2014/06/24

21

### GFB の構築パラメーターの決定方法

- 制約条件を満たす GFB のパラメーターのうち仮想リンク数が最小のパラメーターを算出
  - 目的: ToR の使用ポート数の最小化
    - 各階層で使用する ToR のポート数の和

$$\min \sum_{i=1}^{K_{max}} L_i$$

- サーバーラック間で発生するトラフィックを輻辳なく収容
  - 負荷分散手法を前提として細やかな変動に対応
  - 負荷分散を考慮に入れたパラメーターを計算

$$X_i^k(k, N_k, L_k) \leq \bar{B} / T$$

階層  $k$  のリンクのフロー数

リンクの帯域

ToR 間のトラフィック量 (負荷分散を考慮した値)

- サーバーラック間の通信遅延を目標値以下に抑える
  - 最大ホップ数が目標値以下となるパラメーターを計算

$$H_k(k, N_k, L_k) \leq H_{max}$$

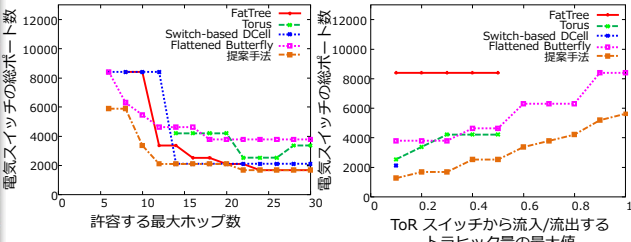
階層  $k$  の ToR 間の最大ホップ数

最大ホップ数の目標値

2014/06/24

22

### 既存のデータセンターネットワークポロジとの比較



- 性能要求を達成するために必要な電気スイッチの総ポート数を評価
  - 提案手法は既存のデータセンターネットワークポロジよりも少ない電気スイッチのポート数で性能要求を達成可能

評価環境

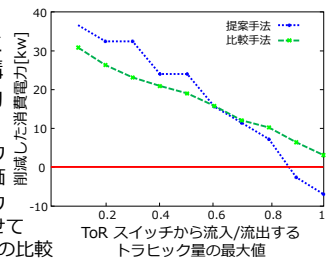
- ToR 数 : 420 台
- 各ネットワークポロジは電気スイッチの総ポート数が最小となるように設定

2014/06/24

23

### 省電力化の効果

- 仮想ネットワークをトラフィック量の変化に応じて再構築することの省電力化の効果を評価
  - 電気スイッチのみのネットワークとの消費電力の差を評価
  - 電気スイッチのみのネットワークからトラフィックにあわせてリンクを削減する手法[3]との比較
- 仮想ネットワークをトラフィック量の変化に応じて、再構築することにより省電力化が可能



評価環境

- ToR 数 : 420 台
- 電気スイッチのみのネットワークは最大時のトラフィックが収容可能な GFB を電気スイッチのみで構築

2014/06/24

24

## 本章のまとめ

- データセンターネットワークに適した省電力を達成する仮想ネットワーク制御を提案
  - 要求されるネットワークポロジの性能に対して、短い計算時間でネットワークポロジを算出
  - データセンター内で発生するトラフィック変動に対応可能
- 提案手法の評価
  - 既存のデータセンターネットワークポロジよりも少ない電気スイッチのポート数で要求される性能を達成可能
  - トラフィック量の変化に応じて、仮想ネットワークを制御することで省電力化が可能

2014/06/24

25

## 博士論文の構成

- Chapter 1: Introduction
- Chapter 2: Virtual Network Reconfiguration with Link load Estimation for Large-Scale Optical Network  
制御サーバーの情報収集負荷削減方法の提案
- Chapter 3: Virtual Network Reconfiguration for Reducing Energy Consumption in Optical Data Centers  
省電力を達成する仮想ネットワークポロジ制御の提案
- Chapter 4: Virtual Machine Placement Considering the Redundancy of Data in Multi tenant Data Centers  
耐故障性を保証した省電力な仮想マシンの配置方法の提案
- Chapter 5: Conclusion

2014/06/24

26

## Chapter 4: Virtual Machine Placement Considering the Redundancy of Data in Multi tenant Data Centers

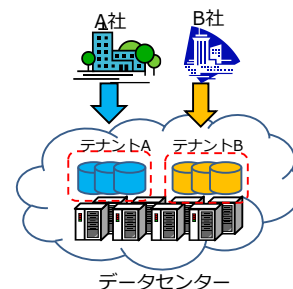
- Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "Virtual Machine Placement Considering the Redundancy of Data in Multi tenant Data Centers," submitted to *IEICE Transactions on Communications*, June 2014.
- Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata, "Virtual Network Reconfiguration with Fault-tolerance and Low Energy Consumption for Multi-Tenant Data Centers," *Technical Report of IEICE (IN2013-193)*, vol. 113, no. 473, pp. 293-298, March 2014 (in Japanese)

2014/06/24

27

## Chapter 4 の研究背景

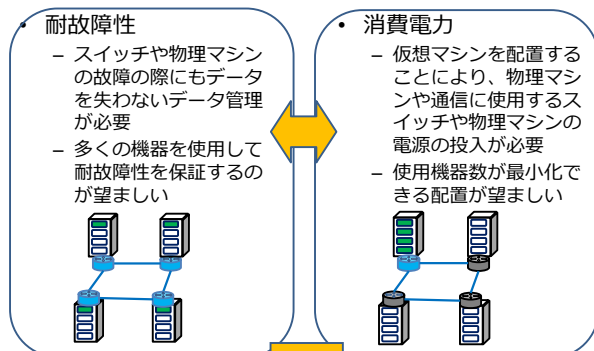
- マルチテナント方式
  - 複数のサービスを1つのデータセンター内に収容
    - ネットワークや計算資源を仮想化し、サービス毎に提供
    - 使用資源や運用コストを低減
  - 複数の仮想マシンを用いてサービスを提供
    - 物理マシンに仮想マシンを割り当て



2014/06/24

28

## 仮想マシンの配置の影響



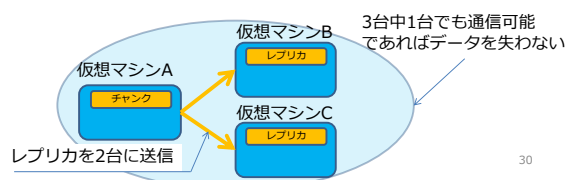
耐故障性を保証しつつ省電力な仮想マシンの配置方法が必要

2014/06/24

29

## レプリカを使用した耐故障性の保証

- データを複数のチャンクに分割し、複数台の仮想マシンで保持するアプリケーション
  - $k$  台の仮想マシンにレプリカを送信、保持
  - 故障によって  $k$  台の仮想マシンとの通信ができない状況でもデータを失わない
- 機器の故障が影響する仮想マシン間の通信数が  $k$  台以下となるように仮想マシンを配置することで耐故障性を保証



30

### 提案手法の目的とアプローチ

- 耐故障性を保証しつつ省電力な仮想マシンの配置方法の提案
- アプローチ
  - 機器の故障により通信ができなくなる仮想マシンがチャンクのレプリカ数  $k$  台以下となる条件化で消費電力を最小化
  - 故障発生時に通信できなくなる仮想マシンの数を調査する方法を提案

2014/06/24

31

### 仮想マシンの配置

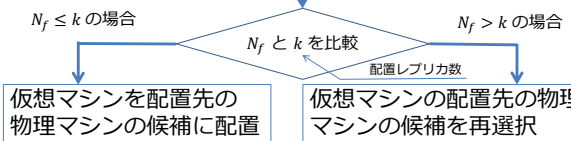
消費電力が最小となるように、仮想マシン配置先の物理マシンの候補を選択

- 消費電力は以下であたえられる

$$E_m + E_s$$

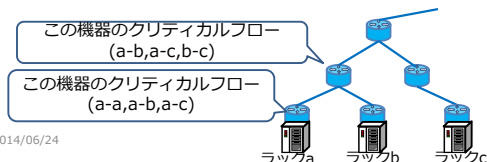
電源の投入が必要な物理マシンの消費電力  $E_m$  + 仮想マシン間の通信のために電源の投入が必要なスイッチの消費電力  $E_s$

故障によって通信できなくなる仮想マシンの数  $N_f$  を計算



### 故障時に通信不可能な仮想マシン数の計算方法

- 当該機器が故障した際に通信できなくなるフロー（クリティカルフロー）をもとに計算
  - 手順：
    - 各機器のクリティカルフローをあらかじめ計算
      - 各機器間の独立な経路数から計算可能
    - 配置済みの仮想マシンと新たに配置する仮想マシンの通信経路を計算
    - 各機器のクリティカルフローの経路数を計算、その最大値が故障時に通信不可能な仮想マシン数  $N_f$  となる

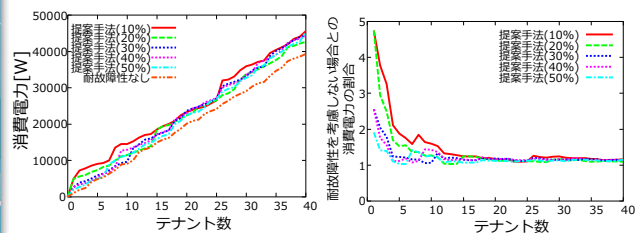


2014/06/24

33

### 提案手法の評価

- 評価環境
- ネットワーク：コアスイッチ36台の FatTree
  - 各テナントは必要な仮想マシン数をランダムに与える
  - レプリカは必要な仮想マシンの10%~50%にランダムに送信



- 収容テナント数に対する消費電力を評価
  - テナント数が増えると少ないレプリカの数でも、省電力な仮想マシンの配置が可能であることを確認
  - 耐故障性を考慮することによる消費電力への影響が小さいことを確認

2014/06/24

34

### 本章のまとめ

- 耐故障性を保証した省電力な仮想マシンの配置を提案
  - 物理機器の故障の影響を受ける物理マシン間の検出方法を提案
- 少ない消費電力の増加で耐故障性を保証した仮想マシンの配置が可能であることを確認

2014/06/24

35

### 本論文のまとめ

- 大規模光ネットワーク用仮想ネットワーク制御を提案
  - 制御サーバーが問い合わせるノード数の削減方法の提案
  - データセンターに適した省電力な仮想ネットワーク制御を提案
  - 耐故障性を保証する省電力な仮想マシンの配置方法を提案
- 今後の課題
  - 情報収集を行う周期を考慮した情報収集負荷の削減
    - 定期的に情報収集を行うのではなく、収集したトラフィック情報から情報収集を行う周期を考慮することによって、情報収集負荷を削減
  - 収容するアプリケーションの特性を考慮した、省電力な仮想ネットワーク制御
    - アプリケーションによって異なる、遅延や通信帯域の性能要求を考慮して、仮想ネットワークを構築

2014/06/24

36