

1

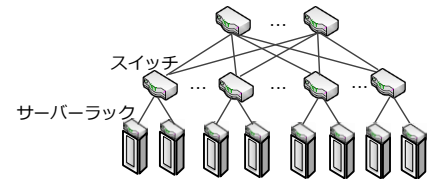
データセンターにおける負荷分散を実現するネットワーク構成手法および経路制御手法の評価

大阪大学基礎工学部情報科学科
村田研究室
下間 雄太

2

データセンター

- 多数のサーバーと、サーバー間を結ぶネットワークで構成
- サーバー間の連携により多量のデータを処理
 - 分散ファイルシステム
 - 分散コンピューティング
- ネットワークが処理性能に与える影響大
 - サーバー間の帯域の不足により、他サーバーとのデータの連携にかかる時間が増大



3

データセンターネットワークの課題

- データセンターネットワークへの要求
 - サーバー間に十分な通信帯域を確保
 - 短い間隔で発生する通信需要の変化への対応
 - 故障発生時も性能を維持
- 要求への対応策
 - ネットワーク構成
 - 通信帯域が確保可能・耐故障性が高い構成
 - 経路制御
 - 環境変動に対応しつつ、負荷分散によりネットワーク資源を有効に利用可能

従来研究では個別に検討

ネットワーク構成が経路制御の性能に与える影響大
局所的なトラフィック情報のみの経路制御で負荷分散が可能な構成
ネットワーク全体のトラフィック情報が負荷分散を行う際に必要な構成

4

研究目的と方法

研究目的

- データセンターに適した経路制御と経路制御を考慮した適切なネットワーク構成を明確化

研究方法

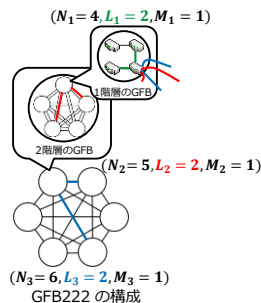
- 以下のネットワーク構成と経路制御の組み合わせを評価
 - ネットワーク構成
 - GFB (Generalized Flattened Butterfly) [1] のパラメーターの設定により生成した複数のネットワーク構成
 - 従来型データセンターネットワーク構成
 - 経路制御
 - 用いるトラフィック情報が異なる複数の経路制御手法
 - ランダム制御 (トラフィック情報なし)
 - 局所制御 (局所的なトラフィック情報)
 - 全体制御 (全体のトラフィック情報)

[1] Y. Tarutani, "Virtual Network Topology Control to Achieve Low Energy Consumption in an Optical Data Center Network," Master's thesis, Osaka University, 2012.

5

ネットワーク構成

- GFB のパラメーターの設定により様々なネットワーク構成を構築
 - 同一のスイッチ数のネットワークでも、リンクの張り方を変えた様々なネットワークを生成可能



パラメーター

k : 階層数
 N_k : 階層 k で接続する $k-1$ 層の GFB の数
 L_k : 階層 k で利用する各ノードあたりのリンク数
 M_k : 階層 k で同一の $k-1$ 層の GFB 間の接続に用いるリンクの最小本数

6

経路制御手法

- 転送先候補の中からトラフィック情報に基づいて適切な転送先を選択
 - 転送先候補
 - 最短ホップ経路または最短ホップ + 1 以内の経路

制御に用いるトラフィック情報

- トラフィック情報なし (ランダム制御)
 - 各スイッチが転送先候補の中から、ランダムに転送先を選択
- 局所的なトラフィック情報 (局所制御)
 - スイッチが自身が接続するリンクのトラフィック量を把握
 - 転送先候補の中から、最も負荷が低い転送先を選択
- ネットワーク全体のトラフィック情報 (全体制御)
 - ネットワーク全体のトラフィック量を収集
 - 収集したトラフィック量から最適な経路を計算

リアルタイムにトラフィック変動に対応可能

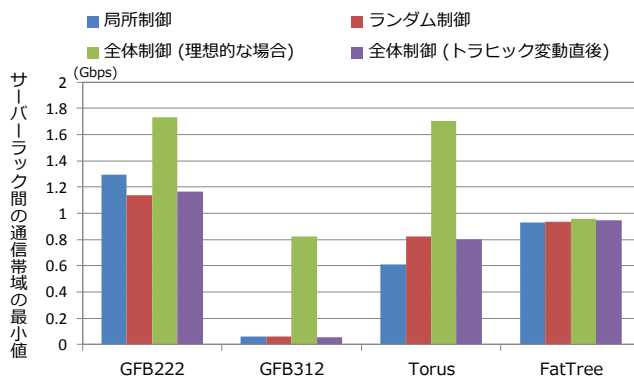
周期的な最適化のみ

評価方法

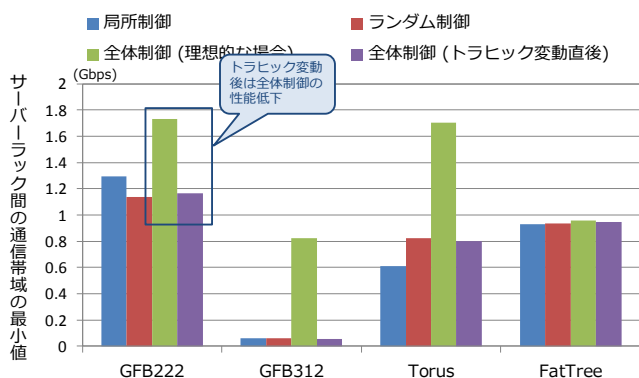
- 評価環境
 - ネットワーク構成
 - 各サーバーラックの全サーバーは1台のスイッチを経由してデータセンターネットワークに接続
 - 6ポートのスイッチ120台を用いた GFB で構築可能なネットワーク
 - 6ポートのスイッチを用いた4階層の FatTree [2]
 - 子ノードの数を均等に配置した構造 (全スイッチ数 189台)
 - トラフィック
 - サーバー間で交換する小さなメッセージ
 - 全サーバーラック間で発生させ、トラフィック量はランダムに決定
 - サーバー間で交換するファイルなどの大きなデータ
 - ランダムに選択したサーバーラック間では、空き帯域がある限り大きなトラフィックを送信
- 評価指標
 - サーバーラック間の通信帯域の最小値

[2] M. Al-Fares, A. Loukissas, and A. Vahdat, "A Scalable, Commodity Data Center Network Architecture," *ACM SIGCOMM 2008 Computer Communication Review*, vol. 38, pp. 63-74, Oct. 2008.

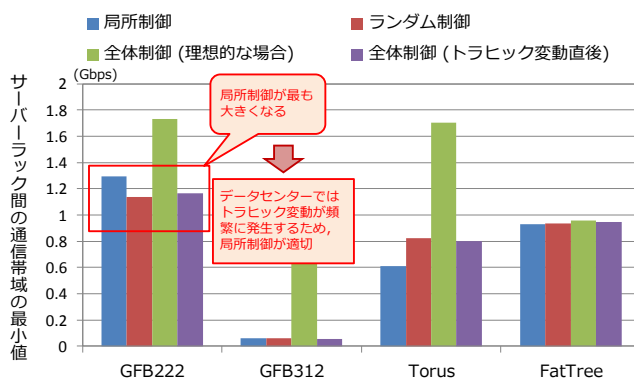
サーバーラック間の通信帯域の最小値の評価



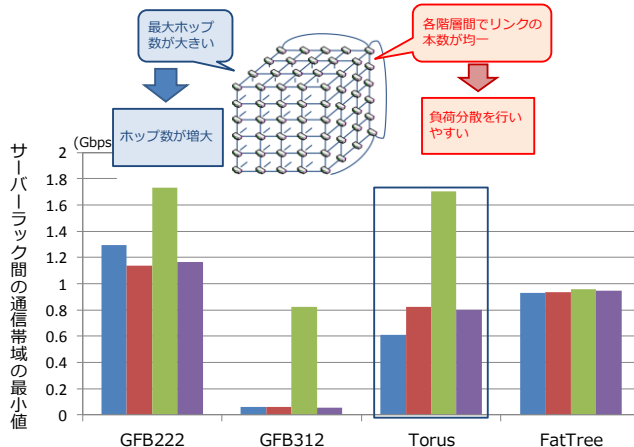
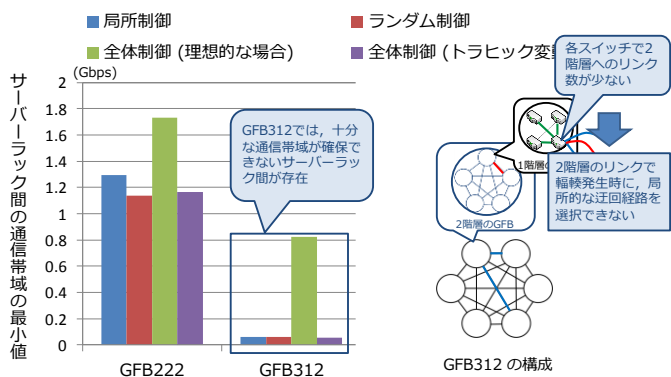
サーバーラック間の通信帯域の最小値の評価



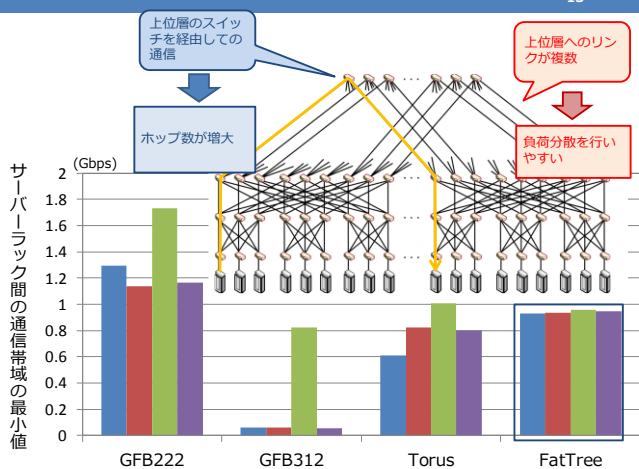
サーバーラック間の通信帯域の最小値の評価



サーバーラック間の通信帯域の最小値の評価

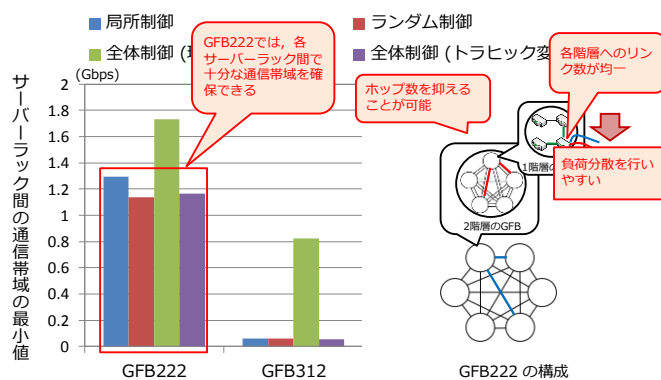


13



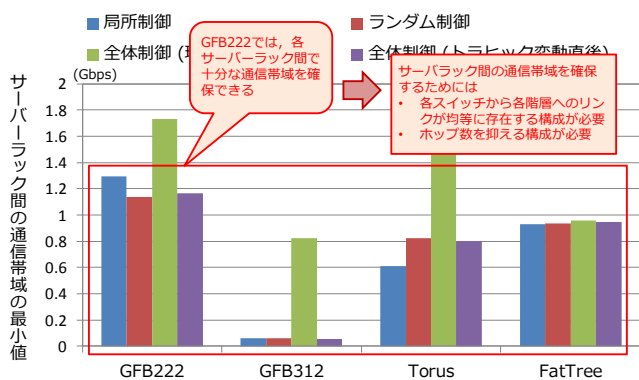
14

サーバーラック間の通信帯域の最小値の評価



15

サーバーラック間の通信帯域の最小値の評価



16

まとめと今後の課題

まとめ

- 経路制御とネットワーク構成の組み合わせの評価
 - トラフィック変動による経路制御の影響の調査
 - 負荷分散により多量のトラフィックを収容可能なネットワーク構成の調査
 - ポート数が6または8のスイッチを120台または220台用いることで構築可能なネットワーク構成の評価
 - 故障に対応可能なネットワーク構成と経路制御手法の調査
 - リンク故障およびスイッチ故障の影響の評価

- 環境変動への対応には、局所的な情報を用いた経路制御が必須
- 局所的な経路制御を動作させるためには、各スイッチからいずれの階層のリンクも複数本必要

今後の課題

- 低遅延の通信が可能な経路制御・ネットワーク構成の明確化