様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月　 ８ 日現在

機関番号：２３２０１
研究種目：基盤研究(C)（一般）
研究期間：2019～2021
課題番号：１９Ｋ１１９２８

研究課題名（和文）分散リバランシング制御による非輻輳データセンタ・ネットワークの研究

研究課題名（英文）Non-congestion datacenter networks based on the distributed rebalancing algorithm

研究代表者

太田　聡（Ohta, Satoru）

富山県立大学・工学部・教授

研究者番号：８０４３８１６８

交付決定額（研究期間全体）：（直接経費）　　1,900,000円

研究成果の概要（和文）：制御上のオーバヘッドが少なく，かつ高いスループットを実現するデータセンタ・ネットワークを提供することを目的に，Clos網をトポロジーとして，輻輳を防ぐ分散リバランシング制御について研究した．まず，パケット・レベルの計算機シミュレーションを行うことで，分散リバランシング制御のスループット改善効果を明らかにした．次に，分散リバランシング制御をLinux PC上でソフトウェアとして実装し，実験系で動作確認し，実現性を示した．また実験により従来技術と比較評価し，分散リバランシング制御の優位性を確認した．同時に，Clos網やエラスティック光スイッチ網に関する基礎的な性質も明らかにした．

研究成果の学術的意義や社会的意義
本研究の成果，特に実装を行ったことで，Clos網に基づく現実のデータセンタ・ネットワークに対する，分散リバランシング制御が適用可能性について見通しがついた．また，従来技術よりも優れたスループット特性を達成可能なことを計算機シミュレーションと実験で確認したので，分散リバランシング制御はClos網の伝送能力を最大限引き出すことが可能であることを示した．これは，多くのITサービスがデータセンタを介して提供される現状では，安定した品質良いサービスの提供上，有意義である．特にビッグデータを利用するアプリケーションで要求される大きな伝送帯域の達成のため有益な結果である．

研究成果の概要（英文）：For the purpose of establishing non-congestion data center networks with low control overhead, this study investigated the distributed rebalancing algorithm, which evenly diffuses flows in a Clos network with employing locally obtainable information. The study first clarified the throughput improvement by the algorithm through packet-level computer simulation. Then, the algorithm was implemented by software. This implementation runs on Linux PCs, which emulate switches of a Clos network. The implementation confirms the feasibility of the algorithm. The implemented algorithm was tested on an experimental Clos network, and compared with an existing technique: random routing. The result shows that the rebalancing algorithm outperforms the random routing technique. Meanwhile, fundamental investigations were also carried out on Clos networks. These investigations include the problem to rearrangements needed for unblocking, and efficient method to control elastic optical switching networks.

研究分野：情報ネットワーク

キーワード：データセンタ　スイッチ網　ネットワーク　ルーティング　スループット

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

多くの IT サービスがデータセンタを介して提供される．ことにビッグデータを利用するアプリケーションでは，データセンタ・ネットワークの伝送帯域を最大限利用することが求められる．したがって，高性能なデータセンタ・ネットワークを提供することは今日の IT 技術では極めて重要な課題である．Clos 網は，入出力スイッチと中間スイッチをリンクで結線することで構成される．この構成で，パケットの転送経路が適切ならば，トラヒック・パタンによらず高いスループットを実現でき，データセンタ・ネットワークのトポロジーとして有力である．Clos 網では，特定のリンクに負荷が集中すれば輻輳を起こし，性能は著しく劣化する．Clos 網の伝送帯域を最大限に利用するには，特定のリンクに負荷を集中させないように，パケットの経路を適切に決定する必要がある．

パケットの経路決定手法としては，各入出力スイッチが分散処理によって経路を決定する手法が，制御に伴う処理のオーバヘッドや，負荷変動への追従性の点で有効である．また，制御で必要な情報としては，入出力スイッチで局所的に得られる情報を使えば，制御情報の通信が不要となり，システムを簡素に構成できる．したがって，Clos 網で入出力スイッチが局所的に取得可能な情報に基づいて分散処理を行い，パケットの転送経路を適切に決定し，輻輳を防ぐ手法を確立することは，高性能なデータセンタ・ネットワークの構築上有意義である．

２．研究の目的

本研究の目的は，制御上のオーバヘッドが少なく，かつ高いスループットを実現するデータセンタ・ネットワークを提供することである．研究対象は，Clos 網において，構成要素のスイッチが局所的に取得できる情報のみを使い，分散処理によりフロー・レベルのルーティング制御を行い，この制御によりリンクの負荷を一定値以下に抑えるアルゴリズム[1]である．本アルゴリズムを，分散リバランシング制御と呼ぶ．

分散リバランシング制御に関し，現実のデータセンタ・ネットワークへの適用性を評価するため，本研究では以下の点を明らかにする．

(1) 分散リバランシング制御に関し，従来理論的に明らかにしてきたのはフロー数で評価した負荷の均一性であった．この評価では，パケット・レベルの挙動を考慮していないので，フロー数の均一性がスループットに与える影響は分からない．そこで，パケット・レベルの計算機シミュレーションを行い，分散リバランシング制御の優位性をより精密に評価する．

(2) 分散リバランシング制御を実機上で実装し，実験を通じて実現可能性と性能を明らかにする．

(3) Clos 網が持つ本質的な性質を理論的に明らかにすると共に，データセンタ・ネットワークにおいて有力な手法であるエラスティック光ネットワーク技術に対応したスイッチ網についても研究を進める．

３．研究の方法

(1) Clos 網を分散リバランシング制御と従来技術で制御したときの挙動を，パケット・レベルの計算機シミュレーションにより明らかにする．このシミュレーションでは，輻輳の回避によるスループット改善の効果を比較評価し，分散リバランシング制御の優位性を明らかにする．

(2) 分散リバランシング制御をソフトウェアとして実装する．この実装では TCP のフローに対するルーティングを想定する．TCP フローの発生と終了の検出，アルゴリズムに従った経路決定，ルーティングへの反映を行う処理をプログラムとして実装し，分散リバランシング制御の実現性と優位性を示す．

(3) Clos 網に関する基礎的な研究として，ブロッキングの解除に必要な接続の再配置の回数の問題に取り組む．また，エラスティック光ネットワーク技術に対応したスイッチ網について，メタスロット手法の研究を行う．

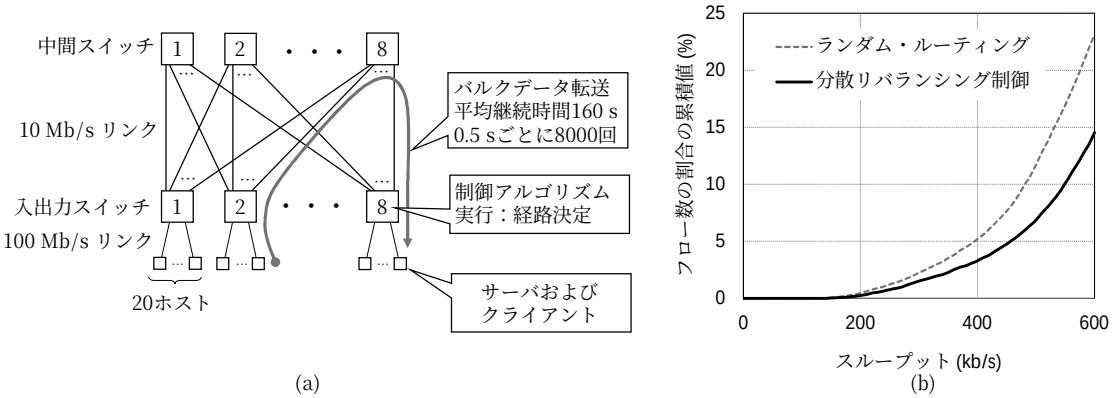
４．研究成果

(1) Clos 網に分散リバランシング制御と従来技術であるランダム・ルーティングを適用したときのスループット特性を，パケット・レベルの計算機シミュレーションにより明らかにした．

計算機シミュレーションのツールとして ns-3[2]を使用した．Clos 網にホストを接続したネットワーク・モデルを ns-3 によって作り，ホストには TCP 上で動作するバルクデータ転送のクライアントとサーバを設置した．Clos 網の入出力スイッチでは分散リバランシング制御，または比較対象であるランダム・ルーティングを実行する．この系において，ランダムに選んだクライアント・サーバ間に，ランダムに決定した時間継続するデータ転送を，多数回繰り返し発生させた．データ転送の都度，データの転送量を計測し，その値を転送時間で除算することでスループットを求めた．シミュレーション条件を図 1(a)に示す．

その結果，TCP フローのスループット平均値は，いずれの方法でも大きな違いは見られなかった．一方，スループットが基準値以下となる TCP フロー数は，分散リバランシング制御を適用することでランダム・ルーティングを用いた場合より少なくなることを明らかとした．図 1(b)は

そのことを示しており，横軸はスループット，縦軸はTCPフロー数の累積値を百分率で表している．同図は，スループットが600 kb/s以下となるフロー数が，従来技術では全フローの23 %に達するのに対し，分散リバランシング制御では15 %に抑えられることを示している．このことから，分散リバランシング制御では，ユーザが性能の劣化を経験する頻度を低減可能であり，優れていると結論できる．



(a) (b)

図1　計算機シミュレーションの(a)評価条件と(b)評価結果

(2) 分散リバランシング制御をソフトウェアにより実装し，実現性を明らかにした．Linux PCをスイッチとして，Clos 網を構築し，その入出力スイッチで実行するソフトウェアとして制御を実装した．本ソフトウェアはTCPフローを対象として動作する．その機能は，TCPフローの発生と終了の検出，フローが通過する中間スイッチの決定，ルーティングへの反映，で構成される．中間スイッチの決定は分散リバランシング制御のアルゴリズムにより行われる．ここで，アルゴリズム実行のトリガとなるのはフローの発生と終了であるが，フローの発生はキャプチャしたパケットのSYNフラグで検出し，フローの終了はパケットのFINフラグにより検出する．また，TCPフローごとに経路を設定するルーティングが必要となるが，これをLinux OSのルーティング機能を活用し，次の方法で実現した．通過する中間スイッチごとにルーティング・テーブルを用意し，パケットに付与したマークに応じてルーティング・テーブルを選択するように設定する．アルゴリズムがフローの通過する中間スイッチを決定したならば，そのフローのパケットに，中間スイッチに対応したマークを付与する．これによりパケットはアルゴリズムの決定した経路へ転送される．

実験系で，このソフトウェアを動作させ，フローが入出力スイッチで実行される分散制御により，リンクに均等に分配され，輻輳が抑制されることを確認した．比較のため，従来技術であるランダム・ルーティングについても評価した．評価は，図2の系で，ホストCでWebサーバを起動し，ホストAでWebサーバのベンチマーク・ツールであるhttperf[3]を実行し，TCPフローの平均継続時間を比較することで行った．ここで，リンクで輻輳が起きればスループットが低下し，フローの平均継続時間は増加する．測定条件としては，サーバ側には平均30MBの指数分布に従ったランダムなサイズのデータファイルを100個用意した．これらに対し平均0.09秒の指数分布に従ったランダムな時間間隔で，ファイルの取得命令を10000回発生させる測定を5回繰り返し，フローの平均継続時間の平均値を求めている．結果を図3に示す．同図が示すように分散リバランシング制御では，従来技術と比べフローの継続時間が短く，これは分散リバランシング制御の方がより均等に負荷を分配し，輻輳を抑制していることを示している．

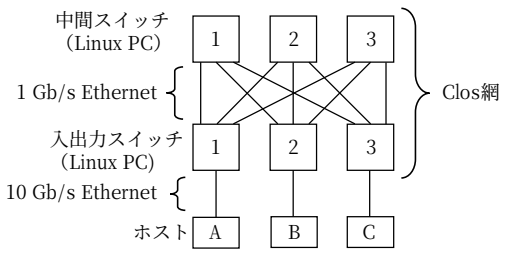


図2　実験系

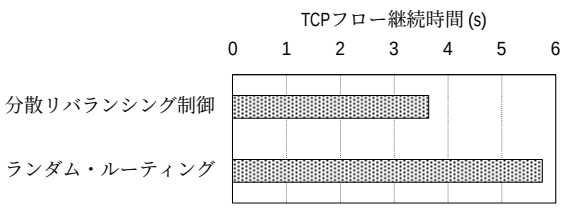


図3　TCPフロー継続時間による輻輳の評価結果

(3) Clos網に関する基礎的な研究として，再配置型Clos網において，ブロッキングを解消するために必要な既存の接続の再配置回数の問題に取り組んだ．本問題に関し，従来知られていない上界値を明らかにした．これは，本申請者が以前に提案した「拡張した接続の鎖」の概念[4]に基づくものである．本概念を利用して再配置を行うとき，まず中間スイッチを通過する接続の数 s を条件とする再配置回数を導き，さらに，s の上界を求めることで，s を含まない再配置回数の

上界値を明らかにした．この上界値は，従来知られている再配置回数の上界値よりも小さい．この結果は，拡張した接続の鎖の概念を用いることで再配置回数が低減することを理論的に保証するものである．

Clos 網に基づくデータセンタ・ネットワークに関しては，エラスティック光ネットワーク技術の導入も重要な研究分野と考えられている．そこで，波長（W）スイッチと空間（S）スイッチで構成され，Clos 網と等価である W-S-W スイッチ網において，エラスティック光ネットワークの特徴である連続した FSU（Frequency Slot Unit）で構成されるデータ流を,ブロッキングなく接続する手法を研究した．この問題に関し，1 つ以上の FSU の帯域から成る「メタスロット」を導入することにより，従来技術よりも少ないハードウェア量でノンブロッキングな接続が可能なことを明らかにした．

文献

[1] S. Ohta, “Flow diffusion algorithms for folded Clos networks,” IEEJ Transactions on Electronics, Information and Systems, vol. 139, no. 11, pp. 1224-1233, Nov. 2019.

[2] Ns developers, ns-3 | a discrete-event network simulator for internet systems. [Online]. Available: https://www.nsnam.org/.

[3] D. Mosberger and T. Jin, “httperf – A tool for measuring web server performance,” ACM SIGMETRICS Performance Evaluation Review, vol. 26, no. 3, pp. 31-37, 1998.

[4] 太田聡, 上田裕巳, “３段スイッチ網の再配置アルゴリズム,” 電子通信学会論文誌(B), J69-B, 2, pp.139-146, Feb. 1986.

５．主な発表論文等

〔雑誌論文〕　計6件（うち査読付論文　6件／うち国際共著　0件／うちオープンアクセス　4件）

１．著者名 Satoru Ohta	４．巻 10
２．論文標題 TCP Throughput Achieved by a Folded Clos Network Controlled by Different Flow Diffusion Algorithms	５．発行年 2020年
３．雑誌名 International Journal of Information and Electronics Engineering	６．最初と最後の頁 16-21
掲載論文のDOI（デジタルオブジェクト識別子） 10.18178/IJIEE.2020.10.1.714	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Satoru Ohta	４．巻 139
２．論文標題 Flow Diffusion Algorithms for Folded Clos Networks	５．発行年 2019年
３．雑誌名 IEEJ Transactions on Electronics, Information and Systems	６．最初と最後の頁 1224-1233
掲載論文のDOI（デジタルオブジェクト識別子） 10.1541/ieejeiss.139.1224	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Satoru Ohta	４．巻 12
２．論文標題 Techniques for Enhancing the Rebalancing Algorithm for Folded Clos Networks	５．発行年 2019年
３．雑誌名 International Journal On Advances in Networks and Services	６．最初と最後の頁 69-80
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Ohta Satoru	４．巻 814
２．論文標題 The number of rearrangements for Clos networks - new results	５．発行年 2020年
３．雑誌名 Theoretical Computer Science	６．最初と最後の頁 106～119
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.tcs.2020.01.018	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Satoru Ohta	4．巻 7
2．論文標題 Optimizing meta-slots for nonblocking elastic optical switching networks	5．発行年 2021年
3．雑誌名 Global Journal of Engineering and Technology Advances	6．最初と最後の頁 046～061
掲載論文のDOI（デジタルオブジェクト識別子） 10.30574/gjeta.2021.7.3.0078	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Ohta Satoru	4．巻 18
2．論文標題 The Number of Rearrangements Performed via Extended Connection Chains for Rearrangeable Clos Networks	5．発行年 2021年
3．雑誌名 WSEAS TRANSACTIONS ON INFORMATION SCIENCE AND APPLICATIONS	6．最初と最後の頁 57～67
掲載論文のDOI（デジタルオブジェクト識別子） 10.37394/23209.2021.18.8	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計8件（うち招待講演　0件／うち国際学会　2件）

1．発表者名 太田聡
2．発表標題 拡張した接続の鎖によるClos網の再配置回数上界値
3．学会等名 2020年電子情報通信学会ソサイエティ大会
4．発表年 2020年

1．発表者名 山田篤希，太田聡
2．発表標題 線形計画法による再配置型Clos 網の再配置回数最小化
3．学会等名 2020 年度電気・情報関係学会北陸支部連合大会
4．発表年 2020年

１．発表者名
Satoru Ohta

２．発表標題
Meta-Slot Schemes to Enhance Nonblocking Elastic Optical Switching Networks

３．学会等名
2019 International Conference on Advanced Technologies for Communications（国際学会）

４．発表年
2019年

１．発表者名
Satoru Ohta

２．発表標題
TCP Throughput Achieved by a Folded Clos Network Controlled by Different Flow Diffusion Algorithms

３．学会等名
2020 9th International Conference on Information and Electronics Engineering（国際学会）

４．発表年
2020年

１．発表者名
太田聡

２．発表標題
メタスロット手法を適用したエラスティック光スイッチ網

３．学会等名
電子情報通信学会フォトニックネットワーク研究会

４．発表年
2019年

１．発表者名
太田聡

２．発表標題
最短路モデルによるエラスティック光スイッチ網のメタスロット最適化

３．学会等名
2019年電子情報通信学会ソサイエティ大会

４．発表年
2019年

１．発表者名 前田健吾，太田聡
２．発表標題 データセンタ・ネットワークの負荷分散手法の評価
３．学会等名 2019年度電気・情報関係学会北陸支部連合大会
４．発表年 2019年

１．発表者名 宮本大地，太田聡
２．発表標題 Folded Clos網の省電力ルーティング制御
３．学会等名 2021年度電気・情報関係学会北陸支部連合大会
４．発表年 2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関