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

平成　３０　年　　６　月　２６　日現在

機関番号：　８２６３６
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ００１４４

研究課題名（和文）理論限界に迫る高効率な相互結合網

研究課題名（英文）Interconnection networks approaching the theoretical limit

研究代表者

藤原　一毅（Fujiwara, Ikki）

国立研究開発法人情報通信研究機構・ユニバーサルコミュニケーション研究所データ駆動知能システム研究センター・主任研究員

研究者番号：９０６４８０２３

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：本研究は、次世代のスーパーコンピュータにおいてノード間の通信遅延を理論的下限に到達させることを目指し、それを実現する高効率なネットワーク構成法を追究するため、下記3つのアプローチで研究を実施した。(1)単方向リンクを用いたネットワーク構成法を提案し、従来の双方向リンクに比べてアプリケーション性能が平均30%向上することを明らかにした。(2)小直径グラフ探索コンペ「Graph Golf」を開催し、低遅延ネットワークの設計に直結する小直径グラフとその構成法を広く一般から募集した。(3)グラフデータベース「Graph Bank」を創設し、有用なグラフを蓄積・検索可能として研究成果の社会還元を図った。

研究成果の概要（英文）：In this research, we aim to reach the theoretical lower bound of communication latency between nodes in a next-generation supercomputer. We conducted research using the following three approaches to pursue highly efficient networks. (1) We propose a network construction method using unidirectional links and clarified that the application performance is improved by 30% on average as compared with a conventional network that uses bidirectional links. (2) We held an open online competition for small-diameter graphs, named "Graph Golf", and collected a wide variety of graph instances and graph composition methods from the public. (3) We established a graph database, named "Graph Bank", that return the research results to the society by helping researchers and engineers who search for valuable graphs.

研究分野：　計算機ネットワーク

キーワード：　ネットワーク

１．研究開始当初の背景

スーパーコンピュータの大規模化が進むにつれ、通信遅延がアプリケーション性能向上の足かせとなっている。2019 年ごろのエクサスケール計算機システム上で実行される並列アプリケーションは、300 ナノ秒程度の非常に小さい通信遅延を要求することが予測されている。通信遅延は経路上のスイッチ数に支配されるため、少ないホップ数で多くのノードを結合できる、ランダム性を持つネットワークトポロジを大規模計算機システムに応用しようとする研究が 2012 年ごろから盛んに行われている。しかし、最新の高次数・低遅延スイッチを用いても、10 万ノード規模のシステムの通信遅延は 1000 ナノ秒程度までしか小さくできないと見られている。一方、グラフ理論分野では、与えられた次数（スイッチポート数）と直径（最大ホップ数）をもつグラフのなかでノード数が最大のものを探す Degree/diameter 問題が古くから研究されており、既知のグラフがカタログ化されている。また、次数とノード数が同じならば、無向グラフ（双方向リンク）よりも有向グラフ（単方向リンク）の方が直径を小さくでき、任意の次数・ノード数をもつ直径準最小の(直径が理論下限より最大１だけ大きい)有向グラフ構成法を今瀬らが明らかにしている[引用 1]。我々の試算では、これら最善のトポロジ構成と将来の超低遅延スイッチを用いることで、目標とする通信遅延 300 ナノ秒に肉薄できることが示唆されている（図 1)。図１からわかるように、既知のトポロジや直径準最小トポロジを計算機ネットワークにうまく利用できれば、現在最先端の研究成果であるランダムトポロジよりも通信遅延を小さくでき、場合によっては理論的下限に到達できる可能性がある。しかし、現行の計算機ネットワークが双方向リンクを前提として設計されているのに対し、直径準最小トポロジは単方向リンクでしか成立しないため、そのまま計算機ネットワークに適用することはできない。また、計算機システムのノード数が性能要件から任意に定められるのに対し、既知のトポロジは飛び飛びのノード数でしか発見されていないため、こちらもそのままでは適用できない。

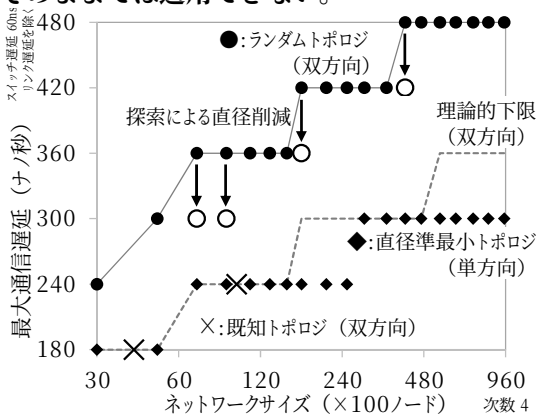

図 1: ネットワークサイズと最大通信遅延

２．研究の目的

上述のような理論的背景を踏まえ、本研究は、エクサスケール計算機システムの相互結合網において、ノード間の通信遅延を理論的下限に到達させることを目指し、それを実現する高効率なネットワーク構成法を追究する。具体的には (1)単方向リンクに基づくネットワーク設計および (2)探索に基づくグラフの直径削減という 2 つのアプローチで通信遅延の理論的下限に迫る。グラフ理論上の知見とネットワーク設計上の制約との間にある溝を本研究が埋めることで、新たなブレイクスルーを達成する。

３．研究の方法

(1) 単方向リンクの利用

有向グラフについては、任意のノード数・次数で準最小の直径を得る構成法が存在する。Ethernet の光ネットワークは１個のスイッチポートに 2 本の単方向ケーブルを挿す形になっており、両ケーブルをそれぞれ異なるノードに接続すれば、配線上は、有向グラフをそのままネットワークトポロジとして実装できる。しかし、単方向リンクではフロー制御のための信号を受信側から送信側へ伝えることができないため、広帯域ネットワークに欠かせないワームホールルーティングなどの高度なルーティング法を実装できない。本研究では、この問題を解決するため、フロー制御信号を必要としない、単方向リンク上で動作する広帯域ルーティング法を開発する。性能評価手段として、連携研究者である鯉渕が保有するフリットレベル・ネットワークシミュレータを利用し、任意のルーティング法を用いた数百ノード規模の精密な性能評価を行う。また、研究代表者らが改良に参加しているフローレベル・ネットワークシミュレータである SimGrid を利用し、数万ノード規模のシステムにおける実アプリケーションの性能評価も行う。

(2) 小直径グラフの探索

無向グラフについては、小直径グラフは飛び飛びのノード数・次数でしか発見されていない。当初の計画では、それら既知のグラフまたはランダムグラフを出発点として、任意のノード数・次数で小直径グラフを発見するアルゴリズムを我々の手で開発する予定だった。しかし、当初方針ではアイディアの幅が限られると考え、この問題を Order/Degree 問題として再定義し、オープンサイエンスの俎上に載せる方針に転換した。この新方針に基づき、小直径グラフ探索コンペ「Graph Golf」を開催したところ、予想を上回る多分野の参加者から多様なアイディアを得ることができた。グラフ探索による直径削減の面で予想以上の成果を得たことから、本研究は同コンペの継続開催と募集したグラフのデータベース化に重点を移すこととした。

４．研究成果

(1) 単方向リンクの利用

我々はまず、単方向リンクではパケットの受信側から送信側へ制御信号を送り返す方法によるリンクレベルのフロー制御が実施できないという問題に対し、制御信号を用いない代わりに最短経路での伝送が保証されない hot-potato ルーティング手法の適用を検討した。パケットレベル・シミュレーションによる性能評価の結果、hot-potato ルーティングを用いても、従来の最短経路ルーティングに比べてスループットは大きく悪化せず、実用上十分な性能が得られることが明らかになった。次に、単方向ネットワークと双方向ネットワークでアプリケーション性能に差異が出るかどうかを検証した。ネットワークトポロジとして、Degree/Diameter 問題の解である既知のグラフを元に、所望のノード数を持つよう頂点を追加／削除したグラフである MDDP グラフを用いた。256 ノード・次数 6 の MDDP グラフに基づくネットワーク上における各種並列アプリケーションの性能をフローレベル・シミュレーションによって比較した結果、単方向ネットワークは双方向ネットワークに対し、アプリケーション性能が平均約 30%、最大約 80%高くなることが明らかになった（図 2）。この成果は並列計算機分野の国際会議 CLUSTER 2017 で発表した[論文 1]。

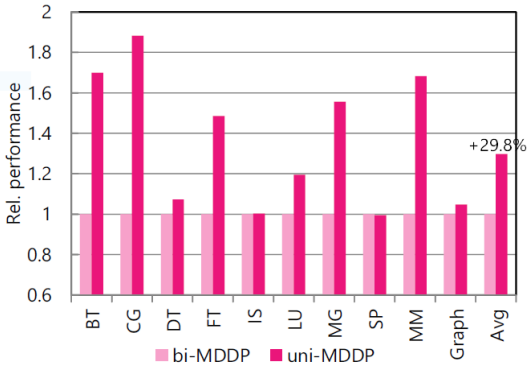

図 2: 単方向ネットワーク(uni-MDDP)上でのアプリケーション性能。双方向ネットワーク(bi-MDDP)での性能を 1 とした相対値

(2) 小直径グラフの探索

与えられたノード数・次数で直径の小さい無向グラフを探す問題は Order/Degree 問題と呼ばれるが、その効率的な解法についてはグラフ理論家による研究が進んでいなかった。これに対し我々は、小直径グラフ探索コンペ「Graph Golf」を開催した（図 3）。このコンペは、主催者が指定したノード数・次数をもつグラフを一般から募集し、その直径・平均距離の小ささを競うものである[学会 3]。開催期間中，参加者は発見したグラフを随時ウェブ経由で投稿し、主催者は投稿された解を毎週 1 回公開する。参加者は自分の順位を確認したり，他の参加者の解を改良したりできる。究極の目標は、あらゆる頂点数・次数の組合せに対し最小の直径・平均距離をもつグラフのカタログを作り、相互結合網の設計者に提供することである。初回である 2015 年は 6 月 1 日から 10 月 15 日まで投稿を受け付け、最終的に 284 件の有効投稿があった。このうち、優れた解の作者 3 チーム 7 名を招待して、2015 年 12 月 10 日、札幌市産業振興センターにおいて Graph Golf ワークショップを開催した(国際会議 CANDAR'15 に併設)。計 22 名の参加者を前にコンペの最終結果を発表し、最多数の最善解を発見した東工大チームに Widest Improvement Award を、理論下限に等しい最適解を 1 個以上発見した東工大チームと京大チームに Deepest Improvement Award を、それぞれ授与した。東工大チームは直径 3 のグラフに着目し、ランダムグラフを初期解として近傍探索および Simulated Annealing を行った。熊本大チームは良い初期解を選ぶことに注力し、東工大チームが苦手とした密なグラフで良い解を発見した。京大チームは直径 2 のグラフに着目し、コンピュータによる探索を行わず、複数のグラフを効率よく連結することで理論下限に到達した。

第 1 回 Graph Golf の成功を踏まえ、2 件の講演会を 2016 年 9 月に企画した。チップ内ネットワークに関する国際会議 NOCS 2016 にて開催した「Small-Diameter Graphs for Low-Latency NoCs」と、第 15 回情報科学技術フォーラム FIT2016 にて開催した「小直径グラフの追究　～グラフ理論の未解決問題とインターコネクトの未来～」である。前述した 3 チームを両イベントに招待してご講演いただき、成果の周知と研究コミュニティの拡大に努めた。

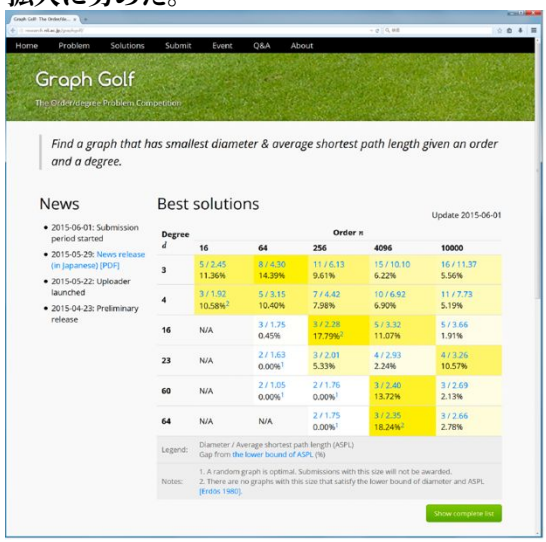

図 3: 第 1 回 Graph Golf ウェブサイト

Graph Golf は以後毎年実施している。毎年新しい小直径グラフが発見される中でも、2017 年の第 3 回コンペで出題した Cray 社製スーパーコンピュータ“Cori”に即した問題において、直径・平均距離とも理論下限に等しいグラフが発見されたことが特筆される。また、高校生・高専生対象のプログラミング

大会である SuperCon2016 では Graph Golf から着想を得た問題が出題され、我々が取り組む問題を研究者以外のコミュニティにも広めることとなった。Graph Golf は本研究終了後の 2018 年以降も継続開催する予定である。

Graph Golf 以外の取り組みとして、グラフ探索において 2 つの派生的な方向性をもった研究を行い、それぞれに成果を得た。ひとつは、ネットワーク実装上の制約を逆手に取り、マシンルーム内のラック配置とケーブル長の上限を陽に考慮したトポロジ最適化手法を提案したものである。マシンルーム内でケーブルが格子状に配線され、各ケーブルの長さには上限があるという現実的な条件において、理論限界に近い低遅延ネットワークを構築できる見通しを得た。この成果を並列計算分野の国際会議で発表した[論文 2][論文 3]。もうひとつは、典型的なネットワーク設計の前提である正則性を覆すことで通信遅延のさらなる削減を追究したものである。正則性とは、各ネットワークスイッチに接続されるケーブル数が同じという条件である。この前提を崩し、スイッチごとに異なる数のケーブルを接続することで、同一規模（ノード数）のネットワークであっても直径をさらに小さくできることを明らかにした。この知見を踏まえ、非正則グラフに基づく相互結合網の構成手法を提案した[学会 1][学会 2]。

本研究は全体として当初の想定を上回る速さで進展したことから、最終年度は成果の継続的な社会還元に重点を移した。具体策として、グラフデータベース「Graph Bank」を創設した。「Graph Golf」では Order/Degree 問題の解を収集したが、この他にも、工学的に有用であっても理論的研究例がないグラフ問題が存在する。このような知の空隙を埋めるため、さまざまなグラフを網羅的に蓄積し、性質や特徴量を用いて有用なグラフを検索できるサービスとして「Graph Bank」を創設することとした。第 1 回「Graph Golf」優勝チームの協力を得て、Order/Degree 問題の解のデータベース化を手始めに、幅広い分野の研究者や実務家が活用するグラフのコーパスとして世界で唯一の存在になることを目指す。これにより、本研究を通じて得られた知見を幅広い分野の理論家・実務家に提供するだけでなく、本研究の終了後も知的基盤として存続するものである。

引用文献

[1] 伊藤正樹，今瀬真，吉田靖之，"準最小な直径を持つ正則グラフ構成法"，電子通信学会論文誌 A，vol.66，no.1，pp.48-55，1983 年 1 月

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 5 件）

[1] Michihiro Koibuchi, Tomohiro Totoki, Hiroki Matsutani, Hideharu Amano, Fabien Chaix, Ikki Fujiwara, Henri Casanova: "A Case for Uni-Directional Network Topologies in Large-Scale Clusters", Proceedings of the 2017 IEEE International Conference on Cluster Computing (Cluster 2017) pp.178-187, 2017 年 9 月. [査読あり]
DOI: 10.1109/CLUSTER.2017.33
[2] Satoshi Fujita, Koji Nakano, Michihiro Koibuchi, Ikki Fujiwara: "Deterministic Construction of Regular Geometric Graphs with Short Average Distance and Limited Edge Length", Proceedings of the 16th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2016) pp.295-309, 2016 年 12 月. [査読あり]
DOI: 10.1007/978-3-319-49583-5_23
[3] Koji Nakano, Daisuke Takafuji, Satoshi Fujita, Hiroki Matsutani, Ikki Fujiwara, Michihiro Koibuchi: "Randomly Optimized Grid Graph for Low-Latency Interconnection Networks", Proceedings of the 45th International Conference on Parallel Processing (ICPP 2016) pp.340-349, 2016 年 8 月. [査読あり]
DOI: 10.1109/ICPP.2016.46

（他 2 件）

〔学会発表〕（計 12 件）

[1] 安戸僚汰，藤原一毅，鯉渕道紘，松谷宏紀，天野英晴，中村維男："可変次数列を持つ相互結合網の構成法"，第 15 回情報科学技術フォーラム（FIT2016），2016 年 9 月 8 日
[2] 安戸僚汰，藤原一毅，鯉渕道紘，松谷宏紀，天野英晴，中村維男："非正則グラフによる低遅延相互結合網の検討"，2016 年並列／分散／協調処理に関する『松本』サマー・ワークショップ（SWoPP2016），2016 年 8 月 10 日
[3] 藤原一毅，藤田聡，中野浩嗣，井上武，鯉渕道紘："みんなで Order/Degree 問題を解いて究極の低遅延相互結合網をつくろう"，2015 年並列／分散／協調処理に関する『別府』サマー・ワークショップ（SWoPP 別府 2015），2015 年 8 月 6 日

（他 9 件）

〔その他〕

ホームページ等

[1] Graph Golf http://research.nii.ac.jp/graphgolf/
[2] Graph Bank http://graphbank.org/

ニュースリリース

[3] https://www.nii.ac.jp/news/release/2015/0528.html
[4] https://www.nii.ac.jp/news/release/2015/1210.html
[5] http://www.nii.ac.jp/news/2016/1122
[6] http://www.nii.ac.jp/news/2016/0306

[7] https://www.nii.ac.jp/news/release/2017/0306.html
[8] https://www.nii.ac.jp/news/release/2017/1122.html
報道等
[9] 日刊工業新聞 2015 年 6 月 9 日 「単純構成のグラフ競技／情報学研が参加募集」
[10] 日刊工業新聞 2016 年 6 月 17 日「情報学研、スパコンのネットワーク設計でコンペ／より単純化モデルを選定」

６．研究組織

(1)研究代表者

藤原　一毅（Ikki Fujiwara）
国立研究開発法人情報通信研究機構・ユニバーサルコミュニケーション研究所データ駆動知能システム研究センター・主任研究員
研究者番号：90648023

(2)研究分担者

なし

(3)連携研究者

鯉渕　道紘（Michihiro Koibuchi）
国立情報学研究所・アーキテクチャ科学研究系・准教授
研究者番号：40413926

河原林　健一（Ken-ichi Kawarabayashi）
国立情報学研究所・情報学プリンシプル研究系・教授
研究者番号：40361159

(4)研究協力者

藤田　聡（Satoshi Fujita）
中野　浩嗣（Koji Nakano）
高藤　大輔（Daisuke Takafuji）
松谷　宏紀（Hiroki Matsutani）
安戸　僚汰（Ryota Yasudo）