

可変次数列を持つ相互結合網の構成法

A Design Method for Interconnection Networks with Variable Degree Sequence

安戸 僚汰†
Ryota Yasudo

藤原 一毅‡
Ikki Fujiwara
天野 英晴†
Hideharu Amano

鯉淵 道紘‡
Michihiro Koibuchi
中村 維男†
Tadao Nakamura

松谷 宏紀†
Hiroki Matsutani

1 はじめに

並列計算機システムの大規模化に伴い、相互結合網の性能が重要になっている。特にスパコンなど性能を重視する場合、ノード間平均距離の小さいトポロジを作ることが求められる。その要求を満たすため、近年の研究では平均距離を削減する目的で相互結合網を無向グラフとしてモデル化し、Order/degree 問題を解くアプローチが提案されている。この問題は、与えられた order(頂点数), degree(次数) を用いて直径・平均最短経路長 (ASPL) の小さいグラフを求める最適化問題である。先行研究において、逐次改善法によって直径・ASPL の下界に数%まで迫ることができ、ベンチマーク性能も有意に向上することが示されている [1]。

従来の相互結合網のモデル化では、次の二点を前提としている。第一に、次数は全ノードで一律、すなわち正則グラフである。したがって、ランダムグラフ、スケールフリーネットワークなどの非正則グラフは対象とはならない。第二に、スイッチに接続されるホスト数が一定である。これは、従来の相互結合網のモデル化がスイッチ数とスイッチの次数(スイッチ間のリンク数)に基づいており、ホストとスイッチの接続関係が考慮されていないためである。したがって Fat-tree などの間接網は構成できず、スイッチごとにホストの数を変えることも不可能である。

そこで、本研究では従来の相互結合網のモデルを拡張し、スイッチに接続するホスト数、スイッチへのリンク数を可変とするモデル化を行うことによって、グラフ理論的アプローチの応用範囲を広げる。その結果、ホスト間 ASPL の点で従来のモデルに基づいた相互結合網よりも優れた相互結合網を構成できることを示す。また、スイッチとホストの数の比率に関する新しい知見を述べる。

2 本研究の課題

上述した二点の制限を外した場合の課題を設定する。それぞれについて順に考えていく。第一の制限に関して、非正則グラフを用いデータセンターのネットワークを作る先行研究が存在するが、これらはポート数の

表 1: 記号の定義

	グラフ的意味	相互結合網における意味
n	頂点数	スイッチ数
k_i	頂点 i の次数	スイッチ i のスイッチへのリンク数
$\langle k \rangle$	平均次数	スイッチへの平均リンク数
h_i	頂点の重み	スイッチ i のホストへのリンク数
r	最大次数	スイッチの基数 (ポート数)
$l(s, t)$	頂点 s, t 間最短経路長	スイッチ s, t 間最短経路長

異なる複数種類のスイッチを用いており、予算の見積もりが困難である。ポート数の多いスイッチを使えばネットワークの性能が大きくなることは自明である。一方、すべて同種のスイッチを用いた場合に非正則グラフを構成した場合の議論はなされていない。第二の制限に関しては、規則的直接網と規則的間接網の両極端の構成法しか議論されていない(ここでいう規則的とは、次数が規則的ということの意味する)。ホストの付け方を自由にした場合のネットワーク構成法およびその性能は明らかになっていない。以上より、本研究の課題は、次数列を可変とした同種スイッチによるネットワークの構成およびその評価とする。

2.1 定義

スイッチを頂点集合 V の元、スイッチ間のリンクを辺集合 E の元とみなして単純無向グラフ $G = (V, E)$ を構成する(すなわち、辺に向きはなく、多重辺・ループは存在しない)。以下、表 1 に示す記号を用いて問題を定義する。

従来の Order/degree 問題は、 k_i がすべて一定の値 k であり、 n, k が与えられたときに頂点間の直径・平均最短経路長 (ASPL) が最小となるグラフを求める問題であった。本研究の問題はこれを拡張したものであり、 $n, \langle k \rangle, r$ が与えられたときにホスト間 ASPL が最小となるグラフを求める問題である。ここで $r = k_i + h_i$ とする。すなわち、スイッチのポート使用率は 100% であり、他のスイッチかホストかのいずれかに接続する。この問題を言い換えれば、ポート数が r のスイッチを n 個使って $n \cdot (r - \langle k \rangle)$ 個のホストをつなぐ最も低遅延な相互結合網を構成する問題である。

従来はホストの接続関係が無視されていたため、スイッチ間の ASPL が最小ならばホスト間の ASPL も最小であった。よって、スイッチ間の ASPL が目的関数

† 慶應義塾大学, Keio University ‡ 国立情報学研究所, National Institute of Informatics

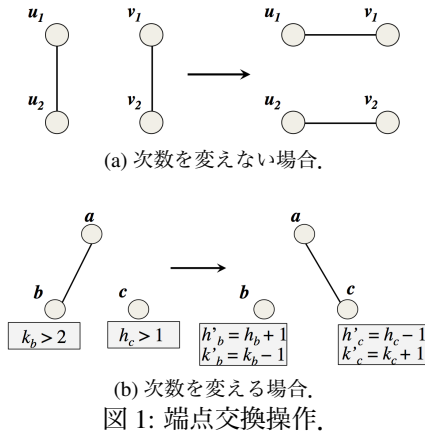


図1: 端点交換操作.

とされていた。しかしホストの接続関係を可変とした場合はそうではない。よってホスト間ASPLを目的関数として、それを最小にしなければならない。そこで、次のように定義されるホスト間ASPL, L_h を目的関数とする。

$$L_h := \frac{\sum_{i=0}^{n-1} ((\sum_{(i,s) \in V} l(i,s) \cdot h_s) \cdot h_i)}{(n \cdot (r - \langle k \rangle))^2 - n \cdot (r - \langle k \rangle)} \quad (1)$$

3 グラフ構成法

3.1 $\langle k \rangle$ -正則グラフ

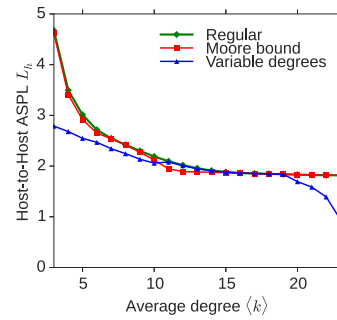
まず比較対象及び初期解としてASPLの小さい $\langle k \rangle$ -正則グラフの生成を行う。方法は、ランダムに生成した $\langle k \rangle$ -正則グラフを初期解として、焼きなまし法(SA)による逐次改善法を行うという単純なものである。このとき、図1aに示すように、2つの辺 (u_1, u_2) , (v_1, v_2) をランダムに選択し、端点を交換して (u_1, v_1) , (u_2, v_2) とする操作を行ったグラフを近傍とする。目的関数は、隣接リストを用いた幅優先探索(BFS)により求めた頂点間ASPLとする。正則グラフの場合はホストを考慮する必要がない。

3.2 可変次数グラフ

3.1節で生成した $\langle k \rangle$ -正則グラフを初期解とし、さらにSAによる逐次改善法を行うことによって L_h を削減する。正則グラフにおける逐次改善法と異なる点は次の二点である。第一に、近傍を図1bに示す操作によって求める。この操作は、 $k_b > 2$, $h_c > 1$ であるような辺 (a, b) 、頂点 c をランダムに選択し、端点を交換して辺 (a, c) に変える操作である。このとき、 h_b, h_c, k_b, k_c はそれぞれ図中に示す h'_b, h'_c, k'_b, k'_c に変化する。このように辺の付け替えと次数の変更を同時に行うため、この操作によって非正則グラフを探索することができる。第二に、目的関数を L_h とする。これは正則グラフのときと同様にBFSで求める。

4 評価

n, r の組に対して $\langle k \rangle$ を3から $r-1$ まで変化させて逐次改善法を行った正則グラフと可変次数グラフを出

図2: $n = 128, r = 24$ における $\langle k \rangle$ に対する L_h の変化。

力し、 L_h の変化を観察する。図2に $n = 128, r = 24$ のときの結果を示す。 $\langle k \rangle$ が増加するほどスイッチ間のリンク数が増加し、ホスト数が減少する。Moore bound (MB) [2]は、 $\langle k \rangle$ -正則グラフに対する L_h の下界である。

図2の結果より、正則グラフにおいては逐次改善法により L_h はMBの100%から108%の範囲まで減少している。次数が小さいときには L_h の変化が大きく、次数が大きくなるにしたがって変化が小さくなっている。一方、可変次数列のグラフでは、 $\langle k \rangle < 11, \langle k \rangle > 19$ の範囲においてMBを下回っている。この結果より、次数を可変とすれば正則グラフより L_h の小さいグラフが構成できることが明らかになった。また、正則グラフに比べて可変次数グラフの L_h が小さいときは、間接網に近い次数分布になっていることがわかった。

図の横軸はスイッチ数とホスト数の比率と言い換えられる。その観点から次のことが言える。まず、スイッチコストを抑えてホストをたくさん接続したい場合には、正則グラフでは L_h が大きくなってしまいうため、可変次数グラフを用いて L_h を抑えることが効果的である。逆に、スイッチコストを気にせずに低遅延性を得たい場合、正則グラフでは次数の増加に対して L_h の減少がゆるやかであり効果が小さいが、可変次数グラフを用いると L_h を劇的に下げることが可能である。

5 おわりに

本稿では、次数列が可変の相互結合網の構成法を述べた。この方法を用いると、ホスト間ASPLが正則グラフの場合の下界を下回ることができるという興味深い現象が現れる。また、スイッチとホストの数の比率によって最適な次数分布が異なることを示した。

謝辞

本研究の一部は総務省SCOPE(#152103004)、科研費(#15K00144, #16H02816)の支援による。

参考文献

- [1] K. Nakano, D. Takafuji, S. Fujita, H. Matsutani, I. Fujiwara, and M. Koibuchi. Randomly optimized grid graph for low-latency interconnection networks. In *Proc. of the Int'l Conf. on Parallel Processing*, Aug 2016.
- [2] M. Miller and J. Širáň. Moore graphs and beyond: A survey of the degree/diameter problem. *Electronic Journal of Combinatorics*, Vol. DS14, pp. 1–61, electronic only, 2005.