

fat-tree におけるパケット 分布の可視化

須藤 稔宗† 横田 隆史† 大川 猛† 大津 金光† 馬場 敬信†

† 宇都宮大学工学部情報工学科

1 はじめに

高性能計算機は多数のノードを用いた並列構成が主流であり、ノード間における通信性能は計算機性能全体に影響する重要な課題の一つである。この課題を解決するため、多数のノード間を接続する形態としてメッシュ構造やトーラス構造、fat-treeといった多様なネットワークトポロジが使用されている。また Minimal Oblivious Routing だけでも fat-tree に適応させたものとトーラス構造に適応させたもの [1] があるなど、各トポロジに合わせた様々なルーティングアルゴリズムの開発も同様に進められている。

fat-tree は Leiserson により提案されたネットワークトポロジである。図 1 に 8 つのノードが接続された fat-tree を示す。図 1 のように fat-tree はリンクが多重化した tree となっている。これにより、tree の欠点であった根付近でのバンド幅の不足を補うことができる [2]。また、通信時のデッドロックの原因となる閉じた経路を作らないといった特徴も持つ。fat-tree は近年の高性能計算機のノード間の接続形態や On-Chip Network (OCN) で用いられており、今後も fat-tree を用いた研究および開発が進められると考えられる。開発において混雑状況などのネットワークの状況を把握することは問題の発見につながるため重要であるが、数値だけでネットワーク全体の状況を直感的に把握することは困難である。より直感的かつ容易にネットワークの状況を把握するために可視化を行うことは有効な手段である。そこで本稿では fat-tree におけるパケット分布を可視化することで fat-tree や fat-tree に対応したルーティングアルゴリズムの問題点を浮き彫りにし、その開発を支援する。

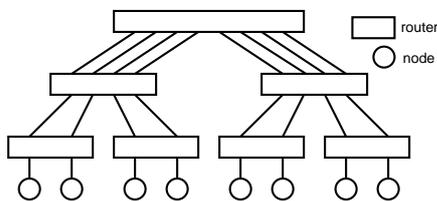


図 1: 階層数 3 ノード数 8 の fat-tree

2 ネットワークシミュレータ BookSim[1]

BookSim は Dally らによって作られた汎用ネットワークシミュレータで、様々なネットワークトポロジ

やルーティングアルゴリズムの開発に用いられている。ネットワークサイズやルーティングアルゴリズム、バッファサイズなど幅広いパラメータ設定が可能であり様々なネットワークのシミュレーションをサイクル単位で行うことができる。またモジュールごとに分けてソースコードが記述してあるため、用途に合わせた機能を容易に追加できる構造となっているなど開発に適したシミュレーションツールである。しかしネットワーク全体のレイテンシやスループットしか出力しない。このため、ネットワークのパケット分布といった混雑状況を把握することは困難である。本稿では BookSim へ必要な機能を追加し、fat-tree のパケット分布状況を可視化する。

3 BookSim を用いたパケット 分布の可視化

fat-tree の可視化を行うにあたり、本稿ではパケットの分布状況を可視化することとした。パケット分布状況を可視化することにより、どの程度の混雑が起こっているのかを直感的に把握できると考えたためである。どの位置でどの程度の混雑が発生しているのかをより正確にとらえられるよう、各ルータの入力ポートごとにバッファの使用量を計測する。バッファの使用量は BookSim を用いて任意のサイクル単位で計測し、ルータの位置情報、入力ポート番号、バッファの使用量等をファイルに出力させる。出力したデータをもとにバッファの使用量を色の濃淡で表し、カラー画像ファイルとして ppm 形式で出力する。

fat-tree の可視化を行うにあたり、ノード数が非常に多い場合に混雑している箇所の把握が難しくなるといった問題点がある。これに対し本稿では 2 つのアプローチによる解決を図る。1 つめは全てのバッファからバッファの使用量の多いものをリストアップし、その位置情報をファイルに書き出すものである。これによりノード数が多く見づらい画像となった場合でも、混雑が発生している位置の目安となる。2 つめは混雑が発生しているバッファとそのバッファを回避する経路のバッファのみを表示した画像ファイルを出力する方法である。fat-tree の特徴は多重化した tree であることを用いたトラフィックの分散が可能な点が挙げられる。図 1 のような fat-tree では 1 つのルータに対し複数の入力ポートが存在する。この複数の入力ポートに均等にパケットが分散されるルーティングが望ましい。そこで混雑が発生しているバッファとそのバッファを回避する経路のバッファのみを表示させる機能を実装する。これにより混雑している箇所とそれを回避する経路についての最低限の情報が得られると考える。

Visualization of Packet Distribution in Fat-tree

†Toshimune Suto, Takashi Yokota, Takeshi Ohkawa, Kanemitsu Ootsu and Takanobu Baba

Department of Information Science, Faculty of Engineering, Utsunomiya University (†)



図 2: oblivious routing でのパケット分布

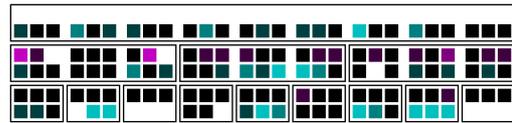


図 3: 適応ルーティングでのパケット分布

4 可視化プログラム

シミュレーションには BookSim を用いるが、BookSim は指定されたパラメータでのネットワーク全体のレイテンシやスループットしか出力しない。まず、パケットの分布を知るために各ルータの入力ポートごとにバッファの計測を行う機能を追加した。さらに fat-tree においてそのバッファがどの階層数の何番目のルータの何番目の入力ポートであるかといった、バッファの位置情報と使用量をファイルに出力する。また、計測と出力は指定されたサイクル間隔ごとに行い、ファイルに出力する。加えて開始サイクル数から終了サイクル数を指定し、そのサイクル間隔は 1 サイクル毎に計測する機能を追加した。これらの機能の追加により、どの位置でパケットの混雑が起きているかが詳細にわかる。

BookSim への機能の追加により各入力ポートごとのバッファ使用量がわかるようになったので、BookSim から出力されたファイルを用いて可視化プログラムで画像を作成する。出力されたバッファの位置情報を元に、バッファを表す正方形の箱を配置する。そのバッファの使用量は色の濃淡で表すことで直感的に使用量がわかるような表現にする。同一のルータ内にあるバッファは線で囲み、fat-tree のどの位置のルータのバッファで混雑が発生しているのかを容易に把握できるように表示する。さらに補助機能として、バッファ使用量の多いバッファの位置情報を別のファイルに書き出すことで混雑の発生している箇所の発見を支援する。

5 可視化の例

図 2 および図 3 はいずれも BookSim を用いて太さ 3, 階層数 3 と設定した fat-tree を用いている。この fat-tree に対して oblivious routing と適応ルーティングを用いてパケットを注入し、パケットの分布を見る。まず、この fat-tree にルーティングアルゴリズム nca (nearest common ancestor) という oblivious routing を適用した。これは最短経路で出発地点から目的地まで経路をランダムに選ぶルーティングアルゴリズムである。通信パターンはユニフォーム通信を用いた。なおバッファのサイズは 4 とした。今回は開始サイクル数を 1, 終了サイクル数を 200 とし、1 サイクル毎にバッファの使用量を出力させた。図 2 は出力されたデータを用いて可視化したものの一部である。線で囲まれた部分は同一ルータ内にあることを、塗りつぶされている部分はバッファをそれぞれ表している。上側にあるルータが上位の階層のルータ、下側にあるルータが下

位の階層のルータである。実際は最下位の階層のルータの下にノードが接続されている。黒いものはバッファにパケットが溜っていない状態で、色が明るいものほどバッファの使用量が多いことを表す。また、ルータの上部にあるバッファは上の階層からの入力、下部にあるバッファは下の階層からの入力であることをそれぞれ示す。今回はバッファサイズは 4 なのでバッファが空の状態から満たされるまで 5 段階の濃淡で表される。バッファが満たされると白色となる。

次に、図 3 に示したものは anca という上位階層への経路選択のみ適応ルーティングを用いた nca を改良したルーティングアルゴリズムによるルーティングを行ったものである。それ以外の設定は図 2 と同じ設定である。図 2 では最上位の階層のルータでパケット分布にばらつきがあることがわかる。しかし図 3 では図 2 と比較して最上位のルータのパケット分布のばらつきが抑えられていることがわかる。このことから適応ルーティングを用いた効果が出ていると考えられる。一方で下位の階層へのルーティングには適応ルーティングが適用されていないためにパケット分布に偏りが生じている。このように可視化を行うことで各ルーティングにおける問題点が浮き彫りなる。また改良したルーティングを適用した場合の効果を明らかにすることにも成功した。

6 おわりに

本稿では fat-tree におけるパケットの分布を提案し、可視化の際の要件を挙げた。さらに要件に基づき、入力ポートごとのバッファの使用量を計測できるように BookSim に機能を追加した。また、出力したデータを用いて画像を作成し、パケットの分布をより直感的にとらえられるような可視化を行った。今後の課題として、シミュレーション時に動的に可視化を行うなどの試みが挙げられる。

謝辞

本稿は、一部日本学術振興会科学研究費補助金 (基盤研究 (C)21500050, 同 (C)21500049) の援助による。

参考文献

- [1] William James Dally and Brian Towels: “Principles and Practices of Interconnection Networks”, Morgan Kaufmann Publishers, 2004.
- [2] Charles E. Leiserson: “Fat-trees: Universal Networks for Hardware Efficient Supercomputing”, IEEE transactions on Computers, pp.892–901, 1985.