Fat-Treeの評価とその実現方式について

1T－7

廣田勝久　岩本邦生　高橋義造

徳島大学工学部　知能情報工学科

1．はじめに

Fat-Tree（FT）は、在来Treeの欠点であった、ルート付近での回線容量の不足を補うようなトポロジを有しており、また、そのトポロジがacyclicであるため、メッセージ通信時のデッドロックの原因となる閉じたパスがない等の優れた特徴を持つ。本論文では、このFat-Tree用のメッセージルータを製作する際の問題点について考察し、さらにFPGA(Field Programmable Gate Array)によりFat-Tree用メッセージルータを実現した結果について述べる。Fig. 2.4に4-ary Fat-Tree（４進Fat-Tree）を示す。このFat-Treeは無閉塞通信経路を持つ。

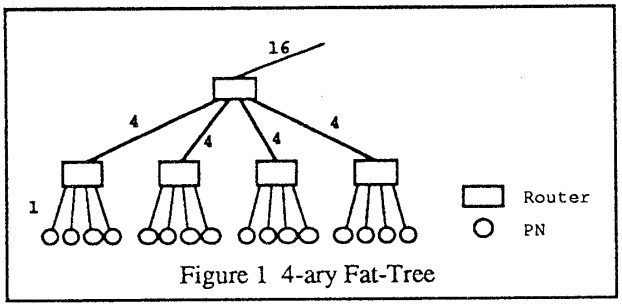

Figure 1 4-ary Fat-Tree

2．所要ハードウェア量

Fat-Treeは無閉塞を実現できるが、そのハードウェア量がかさむという問題点がある。4-ary Fat-Treeのコストをクロスバ網と比較した結果について述べ、次に、ネットワークを構築するのに必要となるチップ数、またそれらの間を接続するのに必要となる配線のコストについて考察する。

2．1　配線数

実際のFat-Treeの構成は、Fig.2.1.1のようになる。この様な構成のFat-Treeについて、ルータ間を接続している配線数について考える。

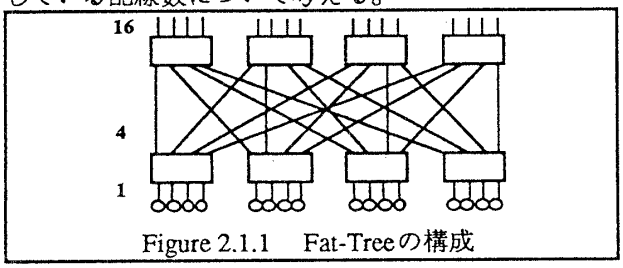

Figure 2.1.1　Fat-Treeの構成

グラフのエッジを１本の配線と数えると、基数k、レベルLのk-ary Fat-Treeの配線数は次式で表される。

$$Wire_{FT} = k^L L = PN \log_k PN \quad \text{... (2.1.1)}$$

ただし、

$PN = k^L$　　PN:プロセッシングノード数

$L = \log_k PN$

である。比較のために、k次ハイパーキューブHC(k)の配線数と比較してみる。

HC(k)の配線数は次式で表される。

$$Wire_{HC(k)} = \frac{(k-1)\{\log_k PN\}PN}{2} \quad \text{... (2.1.2)}$$

(2.1.1)式と(2.1.2)式より、Fat-Treeのほうが配線数の少くなるkを求めるとk>3となり、Fat-Treeは３進木以上でHC(k)より配線数が少なくなることが分かる。

2．2　接点数

Fat-Treeは無閉塞であるということから、同じ非閉塞網であるクロスバ網とコストの比較を行なってみる。コストを、ネットワークの総スイッチ接点数と定義すると、4-ary Fat-Tree網、クロスバ網のコストは、木のレベルをLとした場合、

$$Cost_{FT} = 48L\,4^{L-1} \quad \text{... (2.2.1)}$$

$$Cost_{CB} = 4^L(4^L - 1) \quad \text{... (2.2.2)}$$

となる。Fig.2.2.1に、式(2.2.1)と(2.2.2)を比較したグラフを示す。PN数６４以上で、4-ary Fat-Treeのほうがが Cross barより良くなる。

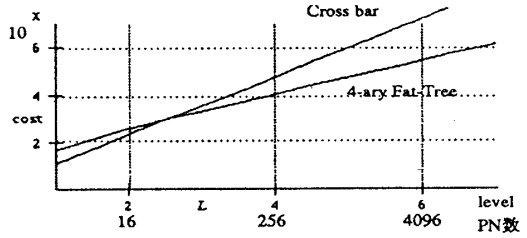

Figure 2.2.1　4-ary Fat-Tree網とクロスバ網のコスト比較

2．3　チップ数

Fat-Treeを構成する各ルータが、１チップで構成されていると仮定すると、k-ary Fat-Tree 全体のチップ数$Chip_{FT}$は、以下のような式で表される。

$$Chip_{FT} = k^{L-1} L \quad \text{... (2.3.1)}$$

ここで、kはTreeの基数、LはTreeのレベル数である。k=4のときについて、これを表にするとTable 2.3.1のようになる。

Evaluation and Impremention of Fat-Tree

Katsuhisa Hirota, Kunio Iwamoto, Yoshizo Takahashi
Tokushima University

Table 2.3.1 FTを構成するのに必要になるチップ数（k=4）

L	PN	Chip
1	4	1
2	16	8
3	64	48
4	256	256
5	1024	1280
6	4096	6144

ルータを１チップで構成する場合、PN数が増加してもルータの数の増加はゆるやかであることが分かる。

3．FPGAによる１チップメッセージルータの実現法

3.1 実装方式の検討

ルータをインプリメントするには、TTLとPALによるもの、FPGAによるもの、ゲートアレイ(GA)によるものが考えられる。これらの比較をTable 3.1.1に示す。ルータの回路は大きなものになることが予想される。TTLとPALでルータを作ると実装面積がかなり大きくなることが考えられる。

ゲートアレイは、集積度、スピードなど性能面では、３つのうちで最も良い。しかし製作に費用がかかり、また作り直しが容易に出来ない。

FPGAは、低コストで一個からでも製作でき、また回路の修正が簡単に出来るというメリットがある。また、FPGAは、ゲートアレイに比べてゲート規模は小さいが、最近はかなりのゲート規模を持つものも登場してきている[4]。これらの点をふまえ、今回のルータはFPGAを用いて実装することにした。

Table 3.1.1 実装方式の比較

	ゲート規模	費用	実装面積	設計変更
TTL + PAL	△	◎	×	△
FPGA	○	◎	◎	◎
Gate Array	◎	×	◎	×

Xilinx社のFPGAには、3000シリーズの上位バージョンである4000シリーズ[4]があるが、4000シリーズを用いてもルータが１チップに収まりきらない可能性がある。そこで、より回路規模の小さい2-ary Fat-Tree用のルータを実験的にインプリメントし、どの位の規模になるか実験した。

3.2 2-ary Fat-Tree用メッセージルータ

FPGAによる１チップメッセージルータの実現可能性を調べる目的で、実験的に2-ary Fat-Treeを構成するための１チップルータをFPGA上に試作した。使用したデバイスは、3000シリーズ中、最もサイズの大きい3090である。使用可能なCLB数は320である。以下に、試作したルータの概要を示す。

・フロー制御方式：WH方式
・サポートするFat-Treeのレベル数：４
（PN＝16台まで）
・データバス幅：８bit
・固定パケット長　（データ256byte+ヘッダ1byte）

フロー制御方式は、FPGAで使用できるロジック数の制限からWH（ワームホール）方式とした。１フリットは１byteである。ただし入力部のFIFOには3byteの容量を持たせてある。またパケット長も、内部の回路を小さくするため固定としている。Fig.3.2.1にパケットフォーマットを示す。

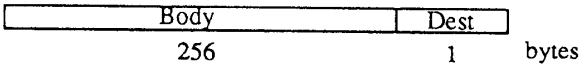

Figure 3.2.1 パケットフォーマット

Destは宛先アドレスで1byteのうち4bitを使用している。Bodyはデータ部分で、256byteの固定長である。

上記のような仕様のルータの回路をコンパイルしたところ、使用CLB数248、未配線ネットなしで正常にコンパイルが終了した。コンパイル時間は、386マシン（386SX 20MHz)を用いて約8時間であった。この実験により、FPGAによるルータの実現が可能であることが分かった。

4． まとめ

FPGAによる、Fat-Tree用のメッセージルータの実現方式について報告した。今回FPGA上にインプリメントしたルータは2-ary Fat-Tree用であり、これで実際のネットワークを構築するにはＰＮ数に対しルータの個数が多く必要になりコスト性能比が悪い。

今後は,よりコスト性能比の良い4-ary Fat-Tree用のルータを、ゲート規模の大きい4000シリーズを用いることにより実現する予定である。

＜ 参考文献 ＞

[1] Charles E. Leiserson, "Fat-Trees: Universal Networks for Hardware-Efficient Supercomputing" IEEE Trans. on Computers, Vol. C-34, No.10, pp. 892-901 ,(1985).

[2] Charles E. Leiserson et. al. ,"The Network Architecture of the Connection Machine CM-5", SPAA '92 pp. 272-285, (1992).

[3] フィールド プログラマブル ゲートアレイデータブック 1992年版 Xilinx Inc., (1992).

[4] THE XC4000 DATA BOOK Xilinx Inc. , (1992).

[5] Lionel M. Ni and Philip K. McKinley, "A Survey of Routing Techniques in Wormole Networks", IEEE Computer, Vol. 26 No.2, pp. 62-76 (1993).