ハイパクロスバ・ネットワークのスループットの理論的解析*

2B－7

山根幸治、三島健、朴泰祐、中村宏、中澤喜三郎†
筑波大学 電子・情報工学系‡
{yamane,mishima,taisuke,nakamura,nakazawa}@arch.is.tsukuba.ac.jp

1. はじめに

超並列計算機向きのプロセッサ間結合ネットワークの一つであるハイパクロスバ・ネットワーク(HXB)は、特にランダム転送において、他のネットワークに比べて高いスループットを実現できることが知られている[1]。

これまでHXBに於けるランダム転送性能は、主に計算機シミュレーションによって評価されてきた。また、理論解析によるHXBの転送性能は、メッセージのブロックを無視した確率モデルによってのみ評価されている[2]。そこで、本研究では、ネットワーク上でのメッセージのブロックを考慮した確率モデルに基づき、HXBのスループットを理論的に解析する。

2. ハイパクロスバ・ネットワーク

HXBは、比較的小規模なクロスバ・スイッチを組み合わせた多段の間接網であり、拡張性に優れている。また、並列処理において現れる各種パターンの転送を高速に処理する、高い柔軟性を持つ。3次元HXBの例を図1に示す。各次元方向に並ぶプロセッサ(PU)はクロスバ・スイッチによって完全結合されている。2次元方向以上の転送では、メッセージはエクスチェンジャ(EX)と呼ばれるルータ・スイッチを使ってクロスバ(XB)を乗り換える。

HXBはmultistage interconnection network(MIN)を拡張したものと捉えることができる[2]。そこで、以下では、メッセージのブロックを考慮したMINに対する解析方法[3]をHXBに適用した確率モデルを提案する。

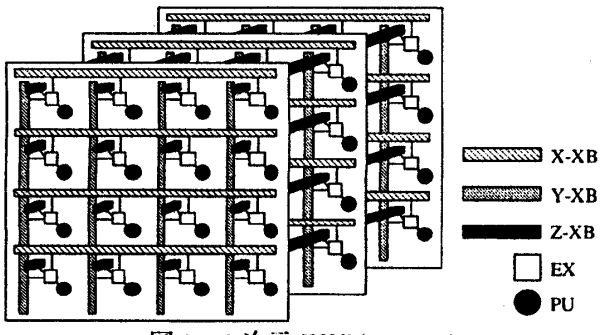

図1：3次元HXB(4×4×3)

3. 解析の仮定

解析を行なう上での仮定を以下に示す。

1. メッセージ長は固定で、1flitとする。
2. メッセージの発生確率は各PUで同じとする。
3. メッセージの転送先PUはランダムに決定される。
4. 次元オーダーの固定ルーティング方式である。
5. XBとEXの各inputには1つのbufferがあり、1メッセージを扱える。
6. メッセージの衝突はランダムに解決される。
7. XBにあるメッセージは、XB内での衝突に勝って出力リンクを獲得し、かつ、次のEXのinput bufferが空である(または、そのbufferにあったメッセージが次に進む)場合に、次のEXに進める(EXにあるメッセージについても同様である)。
8. XB内でブロックされたメッセージは、次のクロックでランダムにXBのoutputを決める(転送先PUが変わる可能性がある)。

4. 確率モデルのパラメータ

確率モデルで用いるパラメータを以下に示す。

- $s_x \times s_y \times s_z$：HXBのサイズ
- il：network input load (PUからネットワークへの、メッセージの送出要求率)
- XBにあるメッセージの状態(クロックt)
  - $P_1(k,t)$：k-XBのinput bufferにメッセージがある確率($= 1 - P_0(k,t)$)
  - $R(k,t)$：k-XBのinput bufferにあるメッセージが、次のEXに進める確率
- EXにあるメッセージの状態(クロックt)
  - $p_{1p}(t)$：EXのPU側のbufferにメッセージがある確率($= 1 - p_{0p}(t)$)
  - $p_{1pk}(t)$：k-XBに向かうメッセージがEXのPU側のbufferにある確率
    ($p_{1px}(t) + p_{1py}(t) + p_{1pz}(t) = p_{1p}(t)$)
  - $r_p(t)$：EXのPU側のbufferにあるメッセージが、先のXBに進める確率
  - (EXのk-XB側のbufferにあるメッセージについても、同様なパラメータを定義する。)
- $q(k,t)$：EXのk-XB側のbufferをメッセージが要求する(k-XBの出力リンクをメッセージが獲得する)確率
- $h(k,t)$：k-XBのinput bufferをメッセージが要求する確率

5. 解析の方法

まず、MINの解析方法[3]について簡単に述べ、その後にHXBのスループットの解析方法について述べる。

[MIN] メッセージ(1flit長)は、次に進むか、ブロックされることによって1クロックを消費する。ブロックされたメッセージは、今いるswitching element(SE)にとどまる。まず、$t = 0$において、すべてのbufferを空にしておく。クロック数を増やしていくと、ネットワーク

*Theoretical Performance Analysis of Throughput of Hyper-Crossbar Network

†Koji YAMANE, Takeshi MISHIMA, Taisuke BOKU, Hiroshi NAKAMURA, Kisaburo NAKAZAWA
‡Institute of Information Sciences and Electronics, University of Tsukuba

にメッセージが溜っていく。クロック t における、あるSE中のメッセージの存在確率は、前のクロックでのそのSEのinput bufferに対する要求率に依存する。また、あるSE中のメッセージが次のSEに進めるかどうかは、次のSE中のメッセージの状態に依存する。この方法では、定常状態になるまでクロック数を増やし、定常状態における最終ステージのSEの出力リンクの利用率を、ネットワーク全体のスループットとして求めている。

[HXB] HXBでは、ある次元を飛び越す転送をEXによって可能にしている。このEXの部分を考え、MINのモデルを拡張すれば、HXBの確率モデルを実現できる。以下に、HXBのスループットの解析方法を述べる。

(1) クロック t において k-XBにメッセージがないのは、前のクロックで k-XBのinput bufferに要求がなく、かつ、前のクロックで k-XBにメッセージがない場合か、メッセージがあってもクロック t では次のEXに進んでいる場合だから、
$P_0(k,t) = (1-h(k,t-1))\times(P_0(k,t-1)+R(k,t-1))$
となる。また、EXの部分も同様に考えると
$p_{0p}(t) = (1-il)\times(p_{0p}(t-1)+r_p(t-1))$
$p_{0k}(t) = (1-q(k,t-1))\times(p_{0k}(t-1)+r_k(t-1))$
となる。

(2) HXBでは、座標が等しい次元方向への転送は行なわれないので、k-XBにある少なくとも1つのメッセージが次のEXのinput bufferを要求する確率は $q(k,t) = 1-\left(1-\frac{P_1(k,t)}{s_k-1}\right)^{s_k-1}$ となる。
k-XBにおいて、あるinput buffer中のメッセージは、ある1本の出力リンクを $\frac{q(k,t)}{s_k-1}$ の確率で獲得する。また、出力リンクの選び方は (s_k-1) 通りある。よって、k-XBにおいて、あるinput buffer中のメッセージが出力リンクを獲得する確率も $\frac{q(k,t)}{s_k-1}\times(s_k-1) = q(k,t)$ である。

(3) HXBの形状から、メッセージがどのルートで転送されるかという確率を求めることができる。例えば、Y-XBを通過してEXに到着したメッセージが、次にZ-XBに向かう確率は $\frac{s_z-1}{s_z}$ で、次にPUに向かう確率は $\frac{1}{s_z}$ である。よって
$p_{1yz}(t) = p_{1y}(t)\times\frac{s_z-1}{s_z}$, $p_{1yp}(t) = p_{1y}(t)\times\frac{1}{s_z}$
となる。
$p_{1px}(t)$, $p_{1py}(t)$, $p_{1pz}(t)$, $p_{1xy}(t)$, $p_{1xz}(t)$, $p_{1xp}(t)$, $p_{1zp}(t)$ も同様に考えて求められる。

(4) $h(k,t)$ は、(3)で求めた値を用いて、確率の加法定理より求められる。

(5) EXのZ-XB側のbufferにあるメッセージは、EX内での衝突に勝ってPUに向かう出力リンクを獲得すると、そのクロック・サイクル中に必ずデスティネーションPUに進めるから、
$r_z(t) = p_{1zp}(t)\times(1-\frac{1}{2}\times(p_{1xp}(t)+p_{1yp}(t))+\frac{1}{3}\times(p_{1xp}(t)\times p_{1yp}(t)))$ となる。また、仮定7より
$R(z,t) = q(z,t)\times(p_{0z}(t)+r_z(t))$ となる。
$r_y(t)$, $R(y,t)$, $r_x(t)$, $R(x,t)$, $r_p(t)$ も同様に考えて求められる。

定常状態 ($t = t_{steady}$) になった後、tp_x=$p_{1xp}(t_{steady})$, tp_y=$p_{1yp}(t_{steady})$, tp_z=$p_{1zp}(t_{steady})$ とおくと、HXB全体のスループットは、確率の加法定理を用いて以下の式で求められる。
$tp_{total} = 1-(1-tp_x)\times(1-tp_y)\times(1-tp_z)$

6. シミュレーション結果との比較

HXBの構成が8×8×16の場合の解析値をシミュレーション値と共に図2に示す。シミュレーションは、理論解析と同じ仮定の下で行なった。

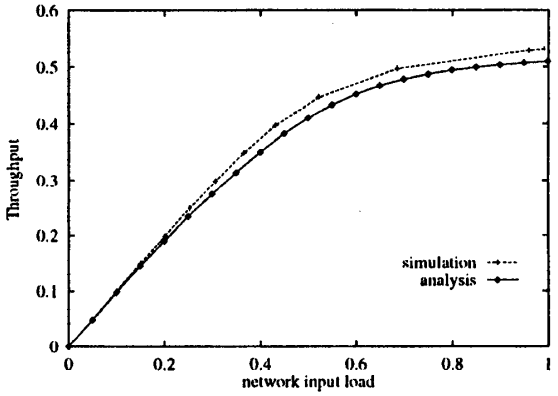

図2：解析値とシミュレーション結果

図2において、解析値とシミュレーション値の誤差(%)は、最大で4%である。また、ここには図に示していないが、同じ1024PUの構成で、4×8×32の場合の誤差が最大で5%、32×32の場合の誤差が最大で2%であった。

7. おわりに

本研究では、確率モデルを用いて理論的にHXBのスループットを解析した。その結果、単純な仮定の下ではシミュレーション値に近い値を得ることができた。

しかし、ここで述べた解析では、ブロックされたメッセージのリトライについて考えていない。この確率モデルでは、ブロックされたメッセージが次のクロックでランダムに出力リンクを選ぶので、空いているリンクの利用率が高くなる。そのために、実際のシステムよりもスループットが高くなっている。今後は、これらの点を考慮した解析を行ない、より現実的なモデルに近付ける。

謝辞

本研究に関し貴重な御意見を頂いた筑波大学西川博昭助教授並びに中澤研究室諸氏に深く感謝します。なお、本研究の一部は文部省創成的基礎研究(06NP0601)の補助による。

参考文献

[1] 朴 泰祐 他、「ハイパクロスバ・ネットワークの性能評価」、信学技報 CPSY93-40, pp.41-48,1993年
[2] 三島 健 他、「ハイパクロスバ・ネットワークの転送性能の解析」、情処全大49回 論文集(6), pp.57-58
[3] H.S.Yoon *et al.*, "Performance analysis of multi-buffered packet-switching networks in multiprocessor systems," *IEEE Trans. Comput.*, vol. c-39, pp.319-327, Mar. 1990.