

通信予測機構を用いた低遅延ネットワークの構成方法と評価

村上 弘和[†] 吉永 努[†] 鯉渕 道紘^{††}

[†] 電気通信大学 情報システム学研究所 ^{††} 国立情報学研究所

我々は、並列アプリケーションとルーティングアルゴリズムのもつ通信の規則性を利用する予測スイッチング方式を提案している。本論文では、予測アルゴリズムとして反辞書法を導入し、3Dネットワークへの適用について考察する。この動的通信予測においては、予測ミスが生じるとネットワーク内に予測ミスパケットが発生する。そこで、予測ミスパケットを抑えるヒントビットを用いたチェック機構を導入し、予測ミスパケットのオーバーヘッドを削減する。その結果、受信スループットが最大で38%向上することを確認した。

Lowering Network Latency : Utilizing a Communication Prediction Mechanism and Its Evaluation

Hirokazu MURAKAMI[†] Tsutomu YOSHINAGA[†] Michihiro KOIBUCHI^{††}

[†] Graduate School of Information Systems, University of Electro-Communication

^{††} National Institute of Informatics

We introduce a dynamic communication prediction mechanism that utilizes communication regularity inherent in parallel applications and underlying routing algorithms. Our approach takes advantage of an anti-dictionary method for the prediction algorithm and investigates its application to 3D-torus networks. The predictive switching has a drawback that prediction misses introduce mispredicted packets. Therefore, we propose a misprediction check scheme that utilizes hint-bits to reduce the misprediction packets. Our evaluation results show that the mispredicted packets are considerably reduced by employing the proposed scheme, and network throughput is improved by up to 38%.

1. はじめに

並列計算機ネットワークの重要な性能指標として、スループットと通信遅延時間がある。大規模なネットワークではメッセージの通信距離が長くなり、通信遅延時間が通信距離の影響を大きく受ける。さらに、ネットワークを構成するルータのパイプライン段数も多段化しており通信遅延の増加に繋がる。通信遅延を削減するためには、ルータごとにかかる通信処理時間を削減することが重要となる。そこで、我々は通信の規則性を利用する動的通信予測機構を提案している[1]。我々は、これまでに2D トーラスネットワークにおける予測スイッチングの有効性を報告しているが、本論文ではさらに3D トーラスネットワークに予測スイッチングを適用する。予測スイッチングにおいては、予

測精度が重要となる。そこで、新たな予測アルゴリズムとして反辞書法[2] を用い評価を行う。さらに、予測ミスを検出するためのヒントビットを導入し、予測ミスの発生を抑制する。加えて、予測ミスにより生じる誤った方向へのパケット(以下、予測ミスパケットと呼ぶ)を破棄する手法を提案する。そして、予測ミスパケットが通信性能に与える影響を考察する。

本論文の構成は以下の通りである。まず2章で、我々が提案する動的予測機構を備えるルータアーキテクチャについて説明する。次に3章において、本稿で用いる予測アルゴリズムの説明を行う。4章で、予測ミスへの対処法について述べる。5章では、予測スイッチングの性能評価を示す。6章で関連する研究を紹介し、最後に7章で本論文をまとめる。

2. 動的通信予測ルータ

図1に我々が提案する動的通信予測機構を有するルータアーキテクチャ構成を示す。ルータは、予測器、ルーティングロジック、3次元の場合東西南北と上下方向の計6つの入出力ポートとPE(Processing Element)用の注入/排出ポート、クロスバスイッチ、入力/注入ポートに取り付ける通信履歴保存用メモリから成る。このメモリは、それぞれの入力/注入ポートが先頭 phit(Physical Transfer Unit) を転送する際、その出力ポート番号をメモリの末尾に保存する。通信履歴数がメモリの許容量を超えると、最も古い履歴に新しい履歴を上書きする。各入力ポートは仮想チャネル(VC)を1つ以上所持し、それを入力VCと呼ぶ。

通常、ルータのパイプラインステージは、入力VCでのバッファリング(IB)、ルーティング計算(RC)、出力VC割り当て(VA)、クロスバスイッチ設定(SA)、スイッチ転送(ST)の順で実行する。予測器は通信履歴から次の出力ポートを予測し、IBと同時にSAを実行してRCとVAの遅延無しにSTを実行する。このSTを予測スイッチング(PST)と呼ぶ。

PSTは予測に基づいて実行するため、予測が外れた場合には正しい通信経路を獲得する必要がある。そこで、パケットヘッダを受信しIBが完了すると、PSTと平行して通常のRCを実行する。通常のRCによって獲得した出力ポートの候補の中に、予測出力ポートがある場合、予測は成功となるので後続のパケットに対しSAとPSTのみを行う。予測が外れた場合は、PSTを中止し通常のVA、SA、STステージを実行する。そして、通常の正しい転送による出力履歴をバッファに保存する。ここで、PSTにより予測を行い獲得した出力VC(隣接ルータの入力VC)は仮予約の状態であるので、すでに他の入力ポートに到着しているパケットがその出力VCを要求した場合は、仮予約をキャンセルし実際に転送するパケットを優先する。また、予測スイッチングの適用一形態として考えられる高速なシリアル転送では予測ミスパケットを単一ルータ内で

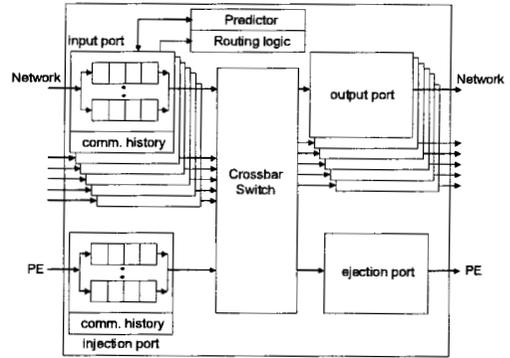


図1: ルータアーキテクチャ構成

破棄することは難しく、誤ってネットワークに送出してしまう。このような予測ミスパケットへの対策は4章で説明する。

3. 反辞書法を用いた予測アルゴリズム

本章では、予測アルゴリズムとして用いる反辞書法について説明を行う。また、5章で用いる他の予測アルゴリズム(パターンマッチング予測機構 SPM, 静的直進予測機構 SS, 直前ポート予測機構 LP)の説明は文献 [1] を参照されたい。

太田ら[2]は、反辞書木と呼ばれる構造を用いて、パターンマッチングに相当する処理を線形時間内で行う手法を提案した(以下、この手法を反辞書法と呼ぶ)。また、この手法を分岐予測に応用することでパターンマッチと並ぶような性能を発揮することが報告されている[3]。

そこで、この反辞書法を各ポートの通信履歴からそのポートに対応する予測出力ポートを算出するよう改造し予測アルゴリズムとして用いる。反辞書法の簡単な動作例を示す。通信履歴として{1, 2, 2, 1, 3, 1}という系列が与えられたとする。最近の履歴1に続く記号を予測する。そこで、反辞書法のために入力系列の接尾辞(入力系列の先頭から1文字ずつ記号を取り除いた残りの部分記号列全体の集合){122131, 22131, 2131, 131, 31, 1}を作成する。そして、入力系列の反辞書(過去に出現していない系列群)を作成する。こ

ここで、直近の系列 { 31 } は初めて出現したもののなので、{ 31 } に 1 記号を付け加えた系列は反辞書には加えない。すると、{ 11, 23, 32, 33, 121, 212 } という反辞書を得る。反辞書の中から予測対象である 1 や 1 より長い接尾辞 (31 など) に続く系列を探すと、{ 11 } という系列しか存在しない (31 より長い接尾辞を含む系列は反辞書には無い) ため 1 を次に出現しない記号であると決定する。残りの予測候補 2 と 3 から、各ノードの頻度カウンタの大きなものを予測値として選択する。このように反辞書法とは、反辞書により予測候補を絞り込んで予測を行うアルゴリズムである。

4. 予測ミスへの対策

本章では、予測ミスを検出し予測ミスパケットを削減する手法と、ネットワーク中に出てしまった予測ミスパケットを破棄する手法について述べる。

4.1 ヒントビットの導入

まず、予測ミスを検出する手法について述べる。本稿では、次元順ルーティング(DOR)を用いる。DORは、3次元の場合 X, Y, Z 次元の順にルーティングを行うアルゴリズムである。DORのような最短経路ソース(または固定)ルーティングでは、パケットの出力方向を容易に知ることができる。そこで、この特徴を用いて予測ミス削減のためのヒントビットを導入する。具体的には、パケットヘッダにパケットの出力方向を示すヒントビットを加え、入力/注入ポートで予測を行う際にその予測値とヒントビットとを比較し、予測値がヒントビットの出力候補に含まれている場合は予測が正しいと判断し PST ステージに移る。このヒントビットを用いた予測ミス検出機構により、明らかに誤りである予測を検出することができる。

4.2 予測ミスパケットの破棄

動的予測機構はパケットヘッダが複数の phit に分割される場合や、RC に複数サイクルかかる場合などに大きな効果が期待できる。しかし、そのような時に予測ミスが発生すると誤った方向へ予測ミスパケットの転送が行われる。予測ミスパケットはネットワーク

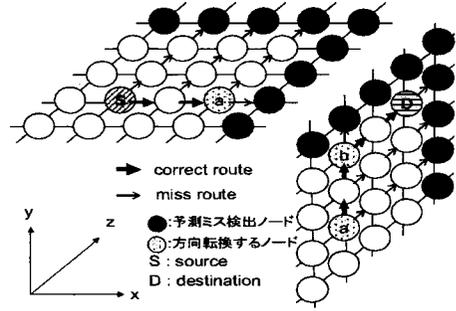


図 2: 予測ミスパケット検出ポート配置イメージ図

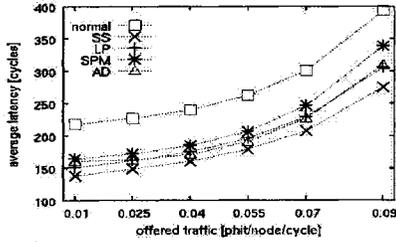
内に蔓延し通信を妨げる。そこで、予測ミスパケットを適正に取り除くための手法について考える。

図 2 に 3 次元トラスネットワーク(k-ary 3-cube)をイメージした予測ミスパケットの流れと、それを検出するポートの位置関係を示す。黒で描かれたノードは予測を行わないポートを有している。それぞれ各次元のアドレス 0 及び k/m (m= 2, 3, 4) 番ノードにおいて、対応する各次元 X, Y, Z の入力ポートでは予測を行わないルータを置く。このポートで通常の RC を実行し、最短経路の範囲内にあるかどうかを調べることで予測ミスパケットを検出する。ここで、図のように S から D に向けてパケットを送信する場合を考える。S を出発したパケットは、a と b ノードで方向転換を行い D に至る。その途中で図の細い矢印のように予測ミスにより経路を外れたパケットは、黒で描かれた予測ミス検出ポートに到着しそこで破棄される。

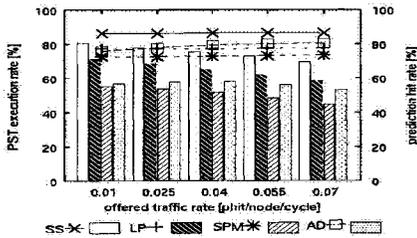
5. シミュレーションによる評価

5.1 シミュレーション条件

3 章で述べた各予測機構に対し、ネットワークシミュレータを用いて実験を行った。使用するネットワークシミュレータは、Book Simulator [4] をベースとして予測機構に対応させたものである。シミュレーション条件として、ネットワークは 16-ary 3-cube, 32-ary 3-cube を使用し、通信パターンはユニフォームランダム通信、ビット列逆順通信を用いる。パケット長は 16-phit 長とし、X-Y-Z 次元順ルーティング(DOR)を



(a) 平均遅延



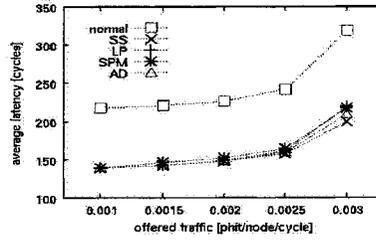
(b) 予測ヒット率, PST 実行率

図 3 : 32-ary 3-cube での uniform 通信の結果

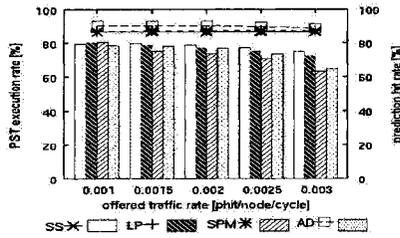
用いる。物理チャンネル当たり 2 本の VC (各 16-phit の FIFO を含む) を所持し、通信履歴を 512 ビットの循環キュー構造メモリで保存する。通常のルータパイプラインは 5 段、予測スイッチングの場合は 2 段で動作する。ただし、RC は 3 サイクル、他の段は 1 サイクルで動作する。予測器はルータ当たり 1 つとし、入力または注入ポートが通信履歴を更新してから予測出力ポートを決定するまでに 4 サイクルかかると仮定する。ウォームアップを十分に行った後、16-ary 3-cube は約 30,000 パケット、32-ary 3-cube は約 60,000 パケットを受信するまでの受信スループット、平均遅延、予測ヒット率、PST 実行率を測定する。

5.2 ユニフォームランダム通信に対する評価

図 3 に 32-ary 3-cube でのユニフォームランダム通信の結果を示す。図 3(a) は、横軸にネットワークへのパケットの注入負荷をとり、その負荷での平均遅延を縦軸に表す。図 3(b) は、横軸に同じくパケットの注入負荷をとり、それに対する受信パケットの全ホップ数に対する PST 実行率と予測のヒット率をそれぞれ縦軸左右に表す。



(a) 平均遅延



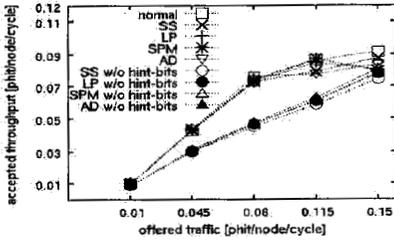
(b) 予測ヒット率, PST 実行率

図 4 : 32-ary 3-cube での bit-reversal 通信の結果

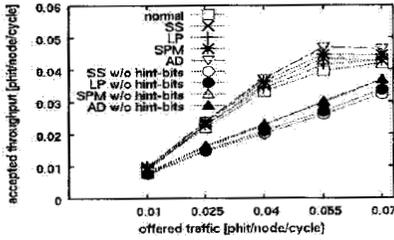
図 3(a) から、PST を実行すると PST を行わない normal の場合と比べて低遅延となることがわかる。予測ヒット率は SS が約 86% と最も高く、次いで反辞書法 (以下、AD) が約 80% 程度である。低遅延化効果はヒット率の高い SS が最も高く、最大で約 30% 遅延を削減する。AD も最大で約 21% 削減する。SS が最も良い性能を示す理由は、ユニフォーム通信は規則性を有さないため、SPM や AD、LP といった過去の履歴を用いる予測アルゴリズムより静的に直進を予測する SS の方が DOR を用いる通信に適しているからである。そのため、ネットワークが大きくなり転送距離が伸びるほど SS に有利であると考えられる。また、負荷が高くなるにつれて PST 実行率が低下しているのは、ネットワークが混雑するとブロッキングにより PST に使用できない出力ポートが増加するためである。

5.3 ビット列逆順通信に対する評価

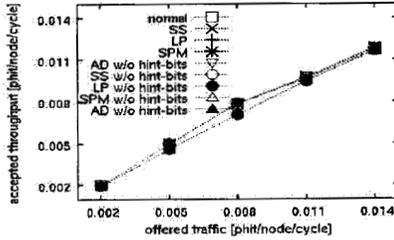
図 4 に 32-ary 3-cube でのビット列逆順通信の実験結果を示す。この通信パターンは、2 進数で表現した送信元ノードアドレス $(b_{n-1}, b_{n-2}, \dots, b_0)$ を逆順にしたノードアドレス $(b_0, b_1, \dots, b_{n-1})$ へ送信を送り返す通信



(a) 16-ary 3-cube uniform 通信



(b) 32-ary 3-cube uniform 通信

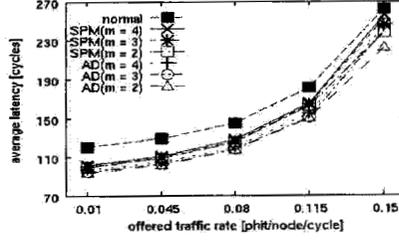


(c) 16-ary 3-cube bit-reversal 通信

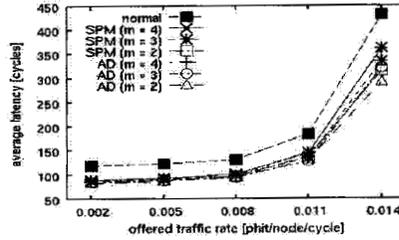
図 5 : 予測ミスパケット削減の効果

パターンである。

ビット列逆順通信は規則性を持つため、SPM や AD、LP の予測ヒット率がユニフォーム通信の時よりも上昇し AD で約 91%、SPM で約 88% となり、normal に対する平均遅延の削減効果も AD が最大で 34%、SPM で約 32% と高くなっている。通常、ビット列逆順通信のような規則性の強い通信パターンでは過去の履歴を用いる予測アルゴリズムが有利であるが、32-ary 3-cube といった大きなネットワークになると平均ホップ数が約 24 ホップと多く、直進する頻度が増加するため SS に有利となる。そのため、図 4(a) のように SS が最大で約 37% の遅延削減を達成する。また、SS は常に直進を予測するため他の入力ポートとの競合が



(a) uniform 通信での平均遅延



(b) bit-reversal 通信での平均遅延

図 6 : 予測ミス検出ポートの配置の影響

少ない。そのため、高負荷時でも PST 実行率が高い。

以上から、2D トーラスばかりではなく入出力ポートが増加した 3D トーラスに対しても予測スイッチングが有効に機能することが確認できた。

5.4 予測ミスパケット削減の有効性

図 5 に、4.1 で述べたヒントビットの有無で性能にどれほどの影響が出るかを、受信スループットを例に挙げて示す。ユニフォームランダム通信では、予測精度がビット列逆順通信よりも低い。そのため、ヒントビットを有効にしないと予測ミスパケットがより多くネットワークを流れる。そのため、多くの予測ミスパケットがネットワーク資源を浪費し、受信スループットに非常に大きな影響を及ぼす。図 5(a) を見ると、16-ary 3-cube の AD で最大 38%、図 5(b) を見ると 32-ary 3-cube の AD で約 33% 受信スループットが低下する。図 5(c) に示すビット列逆順通信は、予測精度が高くユニフォーム通信ほど予測ミスパケットが多くないためユニフォーム通信ほどの受信スループットの低下は見られないが、AD において最大約 10% 受信スループットが低下する。高負荷時に受信スループット

がヒントビットの有無に関わらず同程度になるのは、PST 実行率が低下するためである。

5.5 予測ミスパケット検出ポートの配置

図 6 に、4.2 で述べた予測を実行しないポートの配置が平均遅延に与える影響を示す。ここでは、16-ary 3-cube の結果を用いた。予測を実行しないポートを、各次元に等間隔で 2 列、3 列、4 列 (図中 m) 存在するように増加させて実験を行った。図 6 より、予測をしないポートの数を増やすと平均遅延が AD において最大約 19% 増加する。この結果から、予測の精度がある程度高い場合は予測しないポートを増やして予測ミスパケットの検出や破棄を優先するより、PST を多く実行する方が低遅延化に繋がることわかる。

6. 関連研究

Express VC は、数ホップ離れたルータ間に仮想的なバイパス経路を確保し、それを用いて通信を行うことでその間のルータでの所要パイプライン遅延を削減する[5]。局所的な通信パターンに対する効果は期待できず、非常に広い物理チャンネルを想定しているなどネットワークオンチップに特化した設計となっている。動的予測機構は、局所的な通信においても低遅延化効果が期待でき、ヘッダが複数に分割されるような狭い物理チャンネルにも対応できる点が異なる。

Pre-configured Router は、オンチップネットワークでマッドポストマンスイッチングを利用するために考えられた技術である[6]。通常、マッドポストマンスイッチングは、X 方向から Y 方向へターンするときに必ず予測が外れてしまう。しかし、Pre-configured Router では特殊なパケットを用いて、動的に遅延無しでパケットを転送できる Preferred Path を入出力ポート間に構成し通信経路を柔軟に設定する。したがって、回線交換方式とも類似した技術である。ただし、動的な通信予測は行わない。

7. まとめと今後の課題

本論文では、動的通信予測機構を用いた低遅延ネットワークの構成方法を示した。そして、3次元トラス

ネットワークを例に挙げ、予測機構によって低遅延化効果が得られることを示した。特に、通信の規則性が強い通信パターンに対し、高い低遅延化効果を得られることを確認した。さらに、予測ミス削減のための手法と予測ミスパケットを破棄するための手法を導入し、その評価を行うことで予測ミスパケットが通信に与える影響を考察した。

今後の課題としては、低コスト化を意識した通信予測機構の検討などが挙げられる。

謝辞 本研究は、一部科学研究費補助金 基盤研究 (B) 課題番号 17360178, 基盤研究 (C) 課題番号 19500040 及び NII 共同研究 (提案型) 予測機構を持つルータに関する研究の援助による。また、予測器に関して貴重なご意見を頂いた東工大・吉瀬謙二講師、長野県工科短大・太田隆博講師に深く感謝いたします。

参考文献

- [1] 吉永, 村上, 鯉淵: “2-D トーラスネットワークにおける動的通信予測の効果”, SACSIS '07 論文集, pp.219-226 (2007).
- [2] T. Ota, H. Morita: “On the on-line arithmetic coding based on antictionaries with linear complexity”, Proc. of ISIT'07, pp.86-90 (2007).
- [3] 西新, 森田, 太田: “反辞書木を用いた分岐予測手法”, IEICE Technical Reports CAS2007-38, vol. 107, No. 264 pp. 21-24 (2007).
- [4] W.J. Dally and B. Towles: “Principles and Practices of Interconnection Networks”, Morgan Kaufman Publishers, P.550 (2003).
- [5] A. Kumar, L.-S. Peh, P. Kundu and N. K. Jha: “Express Virtual Channels: Towards the Ideal Interconnection Fabric”, Proc. of the HPCA, pp. 255-266 (2007).
- [6] G. Micheliogiannakis, D. Pnevmaticatos and M. Katevenis: “Approaching Ideal NoC Latency with Pre-Configured Routes”, Proc. of NOCS '07, pp. 153-162 (2007).