C-008

InfiniBand™ ネットワークを用いた並列処理装置のリアルタイム性向上
Improvement of Real-Time Response for Parallel Processing System with an InfiniBand Network

桜井 祐市† 鳥羽 忠信† 平 重喜† 伊藤 建志† 安本 英樹‡ 鈴木 康祐‡ 飯泉 謙‡
Yuichi Sakurai Tadanobu Toba Shigeki Taira Takeshi Ito
Hideki Yasumoto Kosuke Suzuki Ken Iizumi

1. はじめに

マニピュレータ等制御機器を搭載した製造装置や分析・解析装置などに代表される産業用の大容量信号処理においては，処理完了時間を一定に収めるリアルタイム処理が要求される[1]．これら装置では，高精度な制御を目的に，センサの性能向上や，扱う物理量の多様化から，処理データが大容量化している．大容量データ処理の分野では，現在，汎用ネットワークと汎用コンピュータを複数並べた並列処理装置が多く用いられている．今回，我々は InfiniBand ネットワークを用いた汎用の並列処理装置において，産業向け製造装置のデータ処理系に求められるリアルタイム性と低レイテンシを実現する技術を開発した．

本稿では，InfiniBand を用いた並列処理装置にて，大容量データをリアルタイムに処理可能とする伝送制御技術について述べる．

2. 大容量データのリアルタイム並列処理

今回対象とするのは，大容量データを複数のコンピュータでリアルタイムに処理する並列処理装置である．構成は，大容量データ生成部，各コンピュータへのデータ分配制御部，および複数のコンピュータをツリー型のネットワークトポロジで接続した並列処理部である．この装置では，入力したデータを一定サイズに分割，並列の各コンピュータへ順次分配し，処理を行う．ここで，各プロセッサへ順次分配するレイテンシにばらつきが生じた場合，並列処理装置の各プロセッサにおいて処理データの到着タイミングにずれが生じる．このずれが蓄積され，データ処理の一貫性や，制御シーケンスの破綻を招く．このように，並列処理装置にて大容量データをリアルタイム処理するには，各コンピュータへのデータ分配レイテンシの時間管理が重要となっている．十分なリアルタイム性を持つ並列処理装置の実現には，データ分配レイテンシの揺れをなるべく小さくする事が必要である。

3. 既製の InfiniBand を用いた並列処理装置

InfiniBand アーキテクチャ[2]は，データ伝送容量増加に対するスケーラビリティと，チャネルアーキテクチャをベースとしたスイッチファブリックによるネットワークトポロジの柔軟性を持ち，大規模サーバに適した高速相互接続の各種要件を満たすように設計されている．大容量データを取り扱う並列処理装置においては，InfiniBand ネットワークが多用されている．InfiniBand の規格団体である The Open Fabrics Alliance が想定するネットワーク構成を用いた並列処理装置を図 1に示す．アービタとなるワークステーションは，Windows，Linux 等のオペレーティングシステム(OS)とドライバソフトウェアが搭載される．アービタは，一定サイズのデータをメモリ内にバッファリングした後，伝送制御ソフトウェアを起動し，処理データを各コンピュータへ InfiniBand ネットワークを通じ順次分配する．

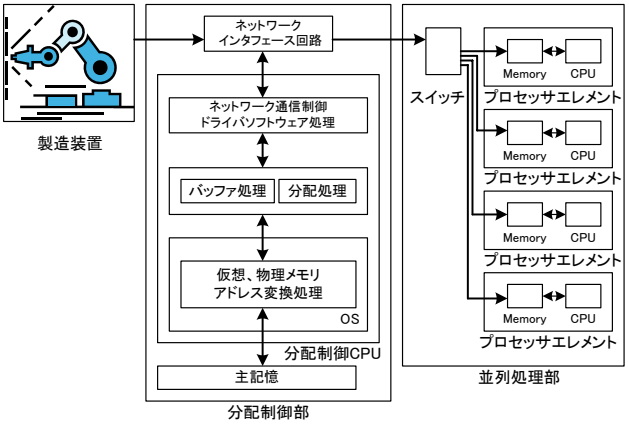

図 1 既製構成採用の並列処理装置構成

図 1 に示す構成では，単一 CPU 上の OS で分配処理とバッファメモリ制御を行うので，以下に示す 2 つの競合が発生し分配レイテンシが変動する．1 つ目の競合は，バッファメモリ制御に関する．OS の管理下では，メモリアクセスの際，仮想メモリアドレスと物理メモリアドレスの変換処理が必要となるが，分配制御 CPU においてはバッファ処理と分配処理でメモリアクセスが発生するため，競合状態となる．分配処理中に，バッファ処理によるアドレス変換要求割込みが発生した場合，分配処理が停止し，分配レイテンシが変動する．2 つ目の競合は，分配制御とバッファリング処理に関する．アービタにおいて，各プロセッサへデータを順次分配中，同時に処理データをバッファリングするので，ネットワーク通信制御ドライバソフトウェアが CPU に割り込みをかけ，タスクをバッファ処理へ変更する．このため，データ分配が停止し，レイテンシが変化する．このように，単一 CPU 上の OS で動作するタスクとして，分配制御とバッファリング処理が競合するので，割込みによる分配レイテンシの揺れという課題を解決する伝送制御技術を開発した．

4. リアルタイム性を向上する伝送制御方式

並列処理装置においてリアルタイム性を改良するため，アービタからのデータ分配レイテンシの揺れ（ばらつき）を抑える競合状態とならない分配制御方式について検討を行った．

先ず，バッファメモリ制御と分配処理の競合を解決すべく，データバスの経路を分離し，メモリバス調停回路を持つバス構成とした．また，通信パケットを解析し物理メモ

† 株式会社 日立製作所 Hitachi, ltd.
‡ 株式会社 日立ハイテクノロジーズ
Hitachi High-Technologies Corporation

リアドレスに直接アクセスするダイレクトメモリアクセス回路を開発した．これにより，バッファメモリ制御と分配処理からの主記憶アクセスに CPU，OS を不要とし，独立制御を可能として競合を解決した．

次に，分配制御とバッファリング処理の競合による，分配処理レイテンシの変化を抑止するため，分配制御テーブルによる分配タイミング制御を行うとともに，InfiniBand ネットワーク通信処理をシーケンス制御回路で構成する事とした．分配制御テーブルは，伝送データサイズ，分配先プロセッサ番号などを持ち，データ入力時にあらかじめこれらパラメタをデータに付加することで，分配制御処理の負荷を軽減する．また，データ量を監視し，一定間隔で分配処理を起動する分配タイミング制御回路と連携するネットワーク通信処理シーケンス制御回路は，CPU と OS を用いない独自開発のハードウェアである．InfiniBand ネットワークへの分配制御時間が予測可能であり，途中の割り込み等による処理中断も発生しないことから，レイテンシ一定化の要件を満たす．本提案技術を用いた分配制御部構成を図 2に示す．

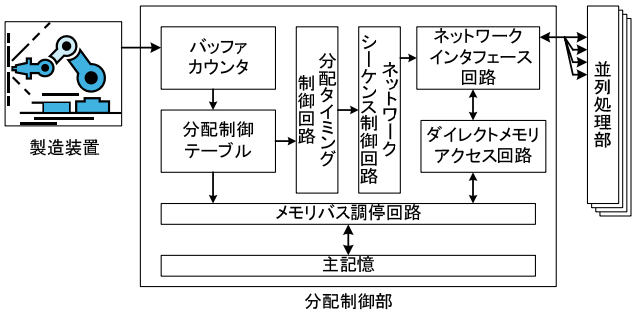

図 2 提案技術を用いた分配制御部構成

以上から，既製の InfiniBand 制御を用いた場合に存在した競合を解決し，データ分配レイテンシの時間管理が可能となった．データ分配レイテンシの揺れを抑える事で，並列処理装置のリアルタイム処理を実現する．

5. 性能比較

Linux サーバを用いた既製技術（OS 処理方式）と，提案した FPGA による伝送制御技術（ハードウェア処理方式）を用いた伝送系の分配レイテンシを比較した．実験環境は，データ入出力 PC と分配制御部から構成する．PC から分配制御部に対しデータを入力した状態で，分配制御部からのデータ分配レイテンシを測定した．図 1に示した既製構成の分配制御部は，Linux OS とデバイスドライバを搭載した PC を用いる．これに対して提案する分配制御部は，データのバッファリング，分配制御，InfiniBand ネットワーク制御機能を，FPGA を用いて試作した．FPGA と試作基板を図 3に示す．次に，それぞれの分配制御部を用いデータ分配レイテンシを測定した結果を図 4，表 1 に示す．図より，OS 処理方式は分配レイテンシが 700～1200ms あるが，ハードウェア処理方式ではこれが 200ms に抑えられ，約 1/4 の低レイテンシ化を達成した．さらにハードウェア処理方式は，データ入力されている状態においてレイテンシの揺れ幅が 1ms 以下であり，OS 処理方式に対してレイテンシの揺れを約 1/1000 に抑え，十分なリアルタイム性を持つ並列処理装置を実現可能とした．以上のように，InfiniBand 通信制御のハードウェア化により，分配制御を CPU と OS を用いずに FPGA 化でき，処理スループット向上に必要な低いレイテンシを，ばらつきが無く実現可能とした．また，リアルタイム性向上により，並列制御部のバッファメモリサイズを小さくすることで，装置の小型化を実現可能とした．

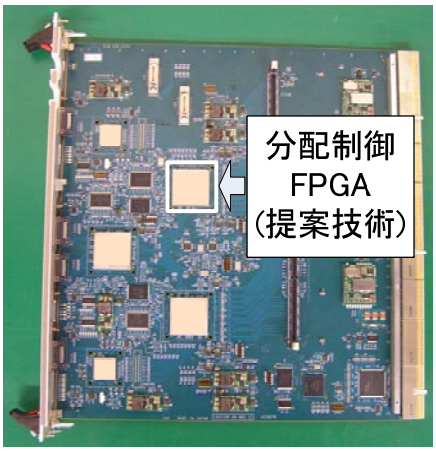

図 3 試作基板

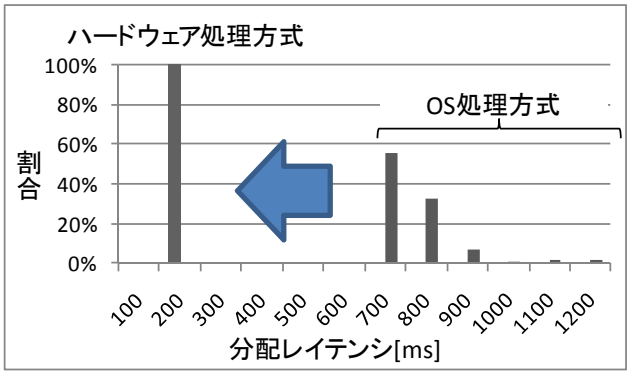

図 4 InfiniBand 分配レイテンシ(512MBytes 伝送時)

表 1 実験結果

	OS 処理方式[ms]	ハードウェア処理方式[ms]
レイテンシの揺れ(Max-Min)	514.88	0.32
平均レイテンシ	731.73	172.55

6. まとめ

本報告は，InfiniBand ネットワークを用いた並列処理装置のリアルタイム性向上を達成する制御技術開発についてのものである．既製の OS 処理方式に存在する 2 つの競合（バッファメモリ制御，分配処理タスクの競合）を解決すべく，ダイレクトメモリ制御回路とネットワーク通信処理シーケンス制御回路を開発し，データ分配を低いレイテンシでばらつき無く実現できた．本開発により，汎用並列処理装置を産業向け製造装置や分析・解析装置等に利用可能となり，高スループットな大容量リアルタイム処理を低コストに実現可能となった．

参考文献

[1]猿渡 俊介, 森川 博之, “ユビキタスセンサネットワーク”, 日本ロボット学会誌 Vol. 28 No. 3, pp.1～4 (2010).
[2]InfiniBand Trade Association. InfiniBand Trade Association. http://www.infinibandta.com.