

## マルチプロセッサ通信処理システムにおけるボード間通信方式

4M-12

飯島 良行<sup>†</sup> 井戸上 彰<sup>‡</sup> 加藤 聰彦<sup>‡</sup> 鈴木 健二<sup>‡</sup><sup>†</sup> 電気通信大学 <sup>‡</sup> KDD 研究所

### 1. はじめに

筆者らは、CPUを持つ通信ボードを、複数枚、PCIバスを有するパソコンに搭載し、高機能かつ高性能な通信処理を実現するマルチプロセッサ通信処理システムの検討を行っている<sup>[1]</sup>。本システムでは、各通信ボードが必要に応じてバスの制御権を獲得し、他のボードのメモリに直接アクセスすることを可能としている。そこで、通信ボードを制御するカーネルでは、このようなハードウェア機能に基づいて、オーバヘッドの小さい、ボードにまたがるタスク間通信機能を提供する必要がある。本稿では、その設計について述べる。

### 2. 基本方針

本システムでは、ボードにまたがるタスク間通信方式（以下、ボード間タスク通信と呼ぶ）に対して、以下の方針を採用する。

- ① ボード内のタスク間通信と同様に、キューを介したメッセージ転送を用いる。その実現のために、カーネルは、バスを介してメッセージ本体とカーネル間の制御情報の転送を行う。
- ② 高速なボード間タスク通信を行うため、バス占有時間を必要最小限にする。
- ③ ボード内に閉じたカーネルおよび通信プログラムの処理に対するボード間通信の影響を最小限とする。
- ④ ボード数の増加が、ボード間タスク通信に大きな影響を与えないようにする

### 3. 方式検討

ボード間でメッセージを転送する方法は、送信側が受信側ボード上メモリにメッセージ本体を書き込む方法と、受信側が送信側ボード上メモリのメッセージ本体を読み出す方法に大別される。このうち受信側がメッセージを読み出す方式は、送信側がメッセージの転送を受信側に対して依頼する必要があるため、ボード間でやり取りする制御情報の数が増加するため好ましくない。一方、送

信側が書き込む方式では、メッセージの転送を要求された時点で書き込むことができ、受信側からの応答を待つ等の処理を必要としない。よって本システムでは、送信側が受信側ボード上メモリにメッセージ本体を書き込む方式を採用する。

送信側がメッセージを書き込む方式としては、書き込むための受信側におけるメッセージ用のバッファの割り当て方法により、

- (a) 受信側メモリ管理情報を、送信側が直接変更してバッファを動的に確保する方式
  - (b) 受信側が前もって割り当てておいたバッファを利用する方式
  - (c) バッファ確保を受信側に依頼する方式
- の3つが考えられる。

受信側のメモリ管理情報を直接変更する方式では、ボード間での排他制御を必要とする。このため、受信側のローカルな処理に影響を与えると同時に、ボード枚数の増加により、オーバヘッドが増大する。受信側にバッファ確保を依頼する方式では、送信側は割り当てが完了するまで待たされる。受信側が前もってバッファを確保しておく方式では、バッファが取得できない場合の影響を考慮しなければならないが、ボード間タスク通信のための時間ならびに制御情報の数が最小になる。

以上の比較により、受信側が前もって確保した通信用バッファに対し、送信側がメッセージ本体を書き込む方式を採用することとする。

### 4. 詳細設計

ボード間タスク通信方式の詳細を以下に示す（図1）。

#### (1) 実装方針

- ① キューを介したメッセージ本体の転送と制御情報の転送により実現する。
- ② 各ボード上には、メッセージ本体用バッファキューと制御情報用キューを、通信相手ボード毎に用意する。

Inter-Board Communication Mechanism for Multi-Processor Communication System

Yoshiyuki IJIMA<sup>†</sup>, Akira IDOUE<sup>‡</sup>, Toshihiko KATO<sup>‡</sup> and Kenji SUZUKI<sup>‡</sup>

<sup>†</sup> Univ. of Electro-Communications / <sup>‡</sup> KDD R&D Laboratories

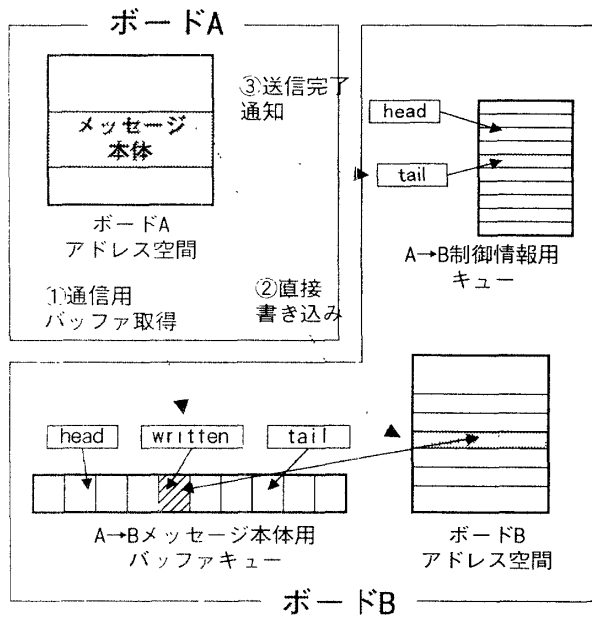


図1 ボード間タスク通信方式の詳細

- ③ 上記のキューに対して固定長の配列を用いることにより、ボード間での排他制御を行わせない。
- ④ メッセージ本体用バッファとして、固定長バッファを用いる方法と、ページング機能を利用した可変長バッファを用いる方法を実現する。

(2) メッセージ本体用バッファキュー

このキューは、受信側が用意したメッセージ本体用の空きバッファの送信側への通知、並びに、送信側がバッファに書き込んだメッセージ本体の受信側への通知の2つの目的で使用される。このため、次の3つの制御変数を使用している。

- tail: 受信側が確保した最新の空きバッファのアドレスを代入する要素を示す。受信側のみが変更する。
- head: 受信側が用意した先頭のバッファに対応する要素を示す。このバッファには送信側が書き込んだメッセージ本体が含まれ、受信側はそれを読み込む。受信側のみが変更する。
- written: 送信側がメッセージ本体を書き込んだ最新のバッファに対応する要素を示す。送信側のみが変更する。

(3) 制御情報用キュー

制御情報用キューは、受信側が制御情報を読み取る要素(head)、送信側が制御情報を書き込む要素(tail)を持つ。

(4) カーネル内の処理手順

- ① 送信側タスクから、メッセージへのポインタがカーネルに渡される。

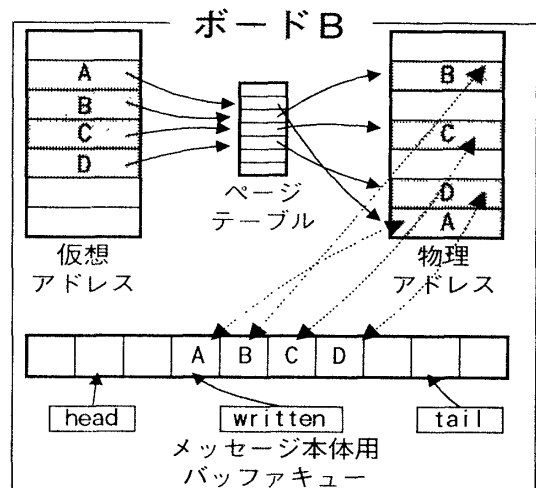


図2 ページングによるボード間タスク通信

- ② 受信側のメッセージ本体用バッファキューから、バッファを取り出す。取り出せるバッファが無い場合、受信側に対しバッファ割り当て要求を送り、バッファが利用可能になるまで待つ。
- ③ メッセージ本体を受信側バッファに書き込む
- ④ 受信側に対し、どのバッファにメッセージ本体を格納したかを制御情報によって通知する。
- ⑤ 受信側は、制御情報を受け取り、データが格納されたバッファポイントを確認する。
- ⑥ 受信側タスクにメッセージを渡す。

(5) ページング機能の利用

次のような方式により、ページング機能を用いて、大きさの異なるメッセージに柔軟に対応するボード間タスク通信を実現できる(図2参照)。

- ① 受信側ボードにおいて、バッファとして物理ページを用意し、メッセージ本体用バッファキューにつなげる。
- ② 送信側は必要なページを取り出し、メッセージ本体を書き込み、上記の(2)(3)(4)と同様な方法で受信側に通知する。
- ③ 受信側では、受信した物理ページを、仮想アドレス空間上の連続した領域にマッピングする。

5. おわりに

本稿では、マルチプロセッサ通信処理システムにおけるボード間通信方式について述べた。現在、本方式に基づいてボード間通信機能の実装を進めており、今後、これらの評価を行う予定である。

参考文献

- [1] 井戸上, 飯島, 加藤, 鈴木, “PCI バスを用いたマルチプロセッサ通信処理システムの設計,” 本大会予稿