

連続メディア処理向きマイクロカーネルの開発（4）

5 F - 7

—入出力方式の設計と評価—

中野隆裕、岩崎正明、中原雅彦、竹内 理、芹沢 一
 （株）日立製作所 システム開発研究所

1.はじめに

VOD (Video On Demand) サーバを始めとする連続メディアアプリケーションでは、従来のOLTP等と比べ桁違いの入出力スループットが必要となる。また、予測不能なイベント（ネットワークからのパケット到着等）が発生しても、ビデオデータの再生が途切れない等のQoS (Quality of Service) 保証を実現する必要がある。

しかしながら、従来OSでは、以下の様な課題を抱えている。

- 1) 入出力処理オーバーヘッドが大きく、ハードウェア性能に見合ったスループットが実現できない
- 2) 割り込み制御機構、排他制御機構、スケジューリング機構等の構造がQoS保証に適さない

本稿では、上記の問題を解決する連続メディア処理向きマイクロカーネル (HiTactix) の入出力インタフェースの設計と実装、及び、実測結果について述べる。

2.設計方針

本稿では、上記で述べた課題のうち入出力スループットの向上について述べる。入出力スループットの向上を阻む主要な要因は以下の2点である。

- ・ 入出力データのメモリコピー
- ・ 頻繁に発生するコンテキストスイッチ

HiTactixでは、これらの入出力オーバーヘッドを以下の方式により低減する。

- a) **ダイレクトバッファマッピング**
 ユーザ空間に直接DMAバッファを割り当て、入出力データのメモリコピーをなくす。
- b) **コマンドチェーン型インタフェース**
 複数の入出力コマンドのキューを引数にとり、入出力処理を一括化する。

3.機能概要

ダイレクトバッファマッピングは、HiTactixのもつ物理メモリ管理機能[1]を用いている。この機能は、ユーザアプリケーションに物理ページ操作インタフェースを提供する。即ち、連続した物理ページを確保したり、確保した物理ページを直接デバイスのDMAバッファに割り当てる機能（ダイレクトバッファマッピング）を提供する。

A Micro-kernel for Continuous Media Processing,
 Design and Evaluation of I/O System.
 Takahiro NAKANO, Masaaki IWASAKI,
 Masahiko NAKAHARA, Tadashi TAKEUCHI,
 Kazuyoshi SERIZAWA
 Systems Development Laboratory, Hitachi, Ltd.

これによって、デバイスとのデータ入出力を直接ユーザデータ領域に確保した物理ページにDMA転送で行ない、メモリコピーを不要としている。また、連続物理ページの確保機能によって、大容量のDMAバッファが利用可能である。（図1）

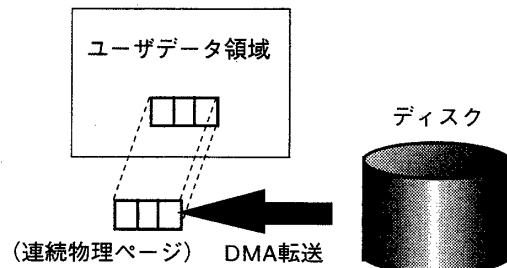


図1 ダイレクトバッファマッピング

コマンドチェーン型インタフェースは、従来OSのread,writeにかわる高性能入出力インタフェースである。また、UDP/IPプロトコルスタック[3]のインタフェースとしてmultisendインタフェースを提供し、入出力処理の一括化をはかっている。

本稿では、multisendインタフェースの概要をVODサーバアプリケーション[2]を例に、以下に示す。

図2は、VODシステムの例を示している。VODサーバは、ビデオデータをディスクから読み出し、ネットワークを用いて各クライアントへ一定レートで送信する。

従来OSでVODサーバを実現すると、送信パケット数と同じ回数sendtoシステムコールが必要となる。（図3）

multisendインタフェースでは、パケットの宛先情報をキューに繋ぎ、そのキューを引数とするmultisendシステムコールの発行1回で、全パケットを送信することができる。（図4）

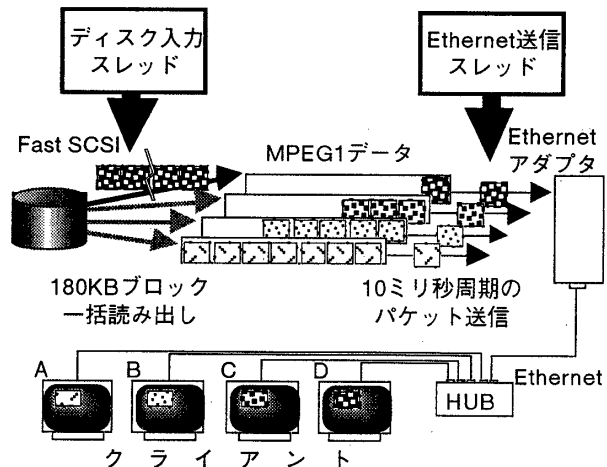


図2 VODシステムの例

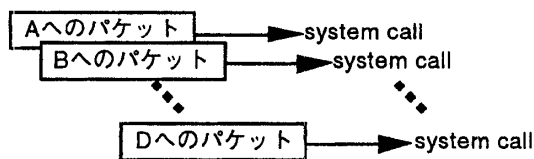


図3 従来OSの送信処理

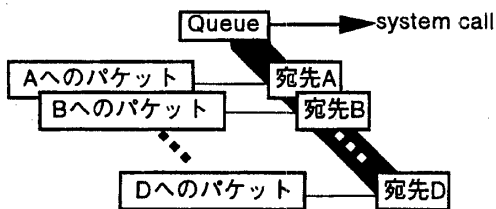


図4 multisendを用いた送信処理

さらにHiTactixでは、高性能デバイスコントローラの機能を利用している。今回使用したDEC[®]社製DC21040は、高性能PCI接続型Ethernetコントローラチップであり、ヘッドギヤザ機能や送信処理の一括化機能をもつ。

multisend方式と高性能デバイスコントローラを組み合わせることで、送信処理をプロトコルスタック層からハードウェア層まで一括化が可能となる。これにより、システムコールやコンテキストスイッチの回数を削減し、入出力性能を向上できる。

4.性能評価

上記の入出力方式を採用したHiTactixは、現在Pentium (90MHz) を搭載するPC-AT互換機で稼働している。本入出力方式による高速化の効果を以下の実験a)、b)により評価した。

実験a) Ethernetを用いた入出力スループットと、CPU利用率の実測を、以下の条件で行なった。

1. 2台のPC-AT互換機をEthernetでPeer-to-peer接続
2. 送信側は128個パケットをmultisendを用い一括送信、これを1024回繰り返した場合と
3. sendtoを用い、パケット送信を128×1024回繰り返した場合を比較する
4. 1パケットのデータサイズが1500, 1280, 1024, 768, 512, 256, 128, 64バイトの各値について、スループットとCPU利用率を計測

スループットの実測結果を図5、CPU利用率の実測結果を図6に示す。multisendを用いる場合は、sendtoを用いる場合に比べ、高スループット、低CPU利用率となっている。

実験b) HiTactix上で実現したVODサーバを用い、ディスクとネットワーク並行動作時のCPU利用率の実測を、以下の条件で行なった。

1. PC-AT互換機のHiTactix上でVODサーバアプリケーションを実行
2. VODサーバは、各クライアント毎に異なるMPEG1データ (180KB/秒) をEthernetで送信
3. VODサーバのCPU利用率を計測

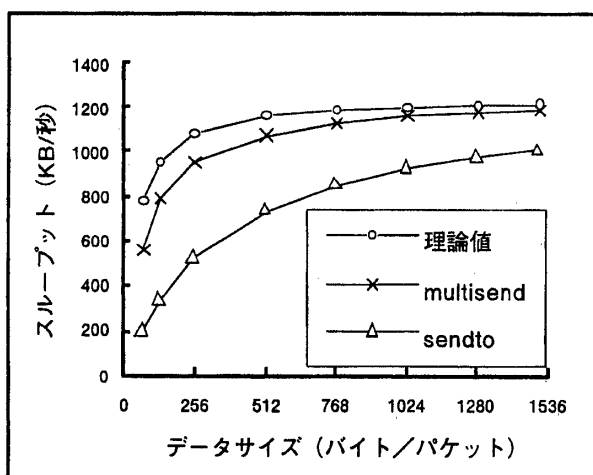


図5 スループット

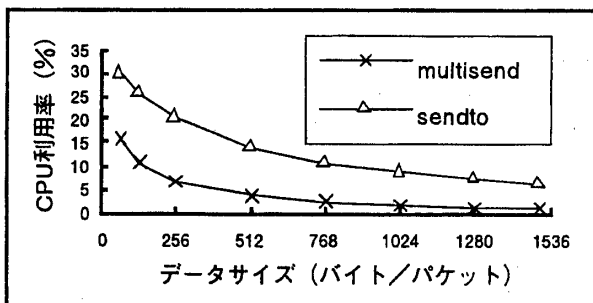


図6 CPU利用率

実測の結果、4クライアント同時稼働時の最大CPU利用率が10%以下、1クライアントあたりのCPU利用率の増加が約0.75%であった。この結果から、理論上クライアント数を100程度まで増加させることが可能である。

5.おわりに

HiTactixの連続メディア処理に必要な入出力方式の設計と評価を行なった。ダイレクトバッファマッピングや、入出力処理の一括化により、充分高いスループットが得られることが確認できた。現在普及しつつある、高速ネットワークの発展により、高スループット化の要求は高まることが予想される。今後、本入出力方式の高速デバイスへの適用に向けて研究を進める予定である。

参考文献

[1]中原他, "連続メディア処理向けマイクロカーネルの設計 (3) —メモリ管理の開発—", 情報処理学会第53回全国大会予稿集, 1996.
 [2]中村他, "連続メディア処理向けマイクロカーネルの設計 (5) —ビデオサーバ動画配信高速化方式の開発—", 情報処理学会第53回全国大会予稿集, 1996.
 [3]D.E.Comer etc, "Internetworking With TCP/IP Vol II: Design, Implementation, and Internals", Prentice-Hall, 1991.

^{注1} Digital Equipment Corporation