3T-8 ネットワーク仮想記憶システム：NET－VMS〔1〕システムアーキテクチャ

八星　禮剛　　陣崎　明

(株)富士通研究所

1. はじめに

マルチプロセッサソフトウェアの開発を容易にするためにはプロセス間通信がソフトウェアからみて充分に単純なインタフェースであることが重要である。我々はこの点を目標とした共有メモリ型のプロセッサ結合方式としてネットワーク仮想記憶システム(NET-VMS:NETworked Virtual Memory System)を検討している[1]。本稿はNET-VMSのプロセス間通信について述べる。

2. 共有メモリでのプロセス間通信

共有メモリシステムでプロセス間通信を行うためには共有メモリ上に通信領域を設け，送信プロセスが領域にデータを書く操作と受信プロセスが領域を読む操作を相互排他的に同期して行う必要がある。従来この排他・同期制御はセマフォやメッセージ通信によって実現されているが，共にソフトウェア制御のためプロセッサの処理能力を低下させ，ソフトウェアの構造を複雑化させる等の問題点を抱えている。

以上に対し，NET-VMSの基本的なアイデアはプロセス間通信機能，すなわちアクセス競合の解決，排他・同期制御を単一階層記憶(Single Level Storage)化して実現することによりソフトウェアから見えるプロセス間通信を単なるメモリインタフェースとするものである。

3. ＮＥＴ－ＶＭＳの構成

NET-VMSはプロセッサエレメント(PE)それぞれに設けられたデマンドページング方式の仮想記憶を直接ネットワーク結合し，全体を一個の仮想記憶システムとして構成する(図１)。従って共有メモリは論理的にこの仮想記憶空間となり，物理的にはPEの実メモリに分散して存在する。

この構成で，プロセッサは自PEの仮想記憶に存在するメモリページに対してのみアクセスできるので，アクセスしたいページが自PEに存在しない時は他PEからページを獲得(ページイン)する必要がある。また同一ページが複数のPEに存在する場合もあるので，ページの変更に対してページ内容の一致を保証する必要がある。

NET-VMSはページイン，ページ内容の一致，先述の排他・同期制御を各PEの仮想記憶にページ毎に設けたアクセスキー(表１)を用いて完全分散制御で実現する。

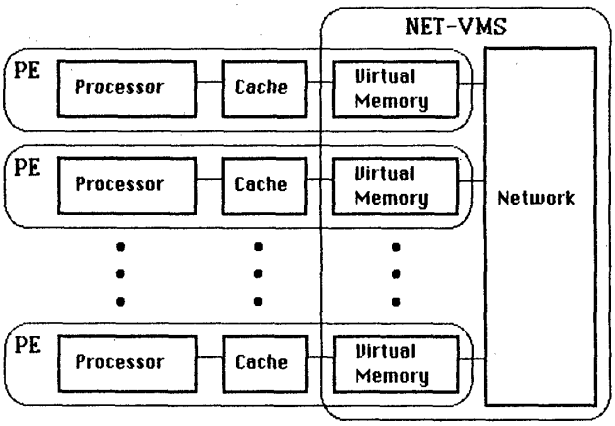

図１　ＮＥＴ－ＶＭＳの構成

表１　ＮＥＴ－ＶＭＳアクセスキー

アクセスキー	意　味
ＶＡＬＩＤ	ページデータが自ＰＥの実メモリに存在することを示すフラグ。
ＣＯＰＹ	ページデータが他の一つ以上のＰＥに存在することを示すフラグ。
ＬＯＣＫ	自ＰＥの実メモリに存在するページデータを他ＰＥがアクセスすることを禁止するフラグ。
ＳＹＮＣ	自ＰＥの実メモリに存在するページデータを他ＰＥがアクセスするまで，このページデータに対する自ＰＥのアクセスを禁止するフラグ。

Networked Virtual Memory System:NET-VMS (1)System Architecture

Reigo Yatsuboshi, Akira Jinzaki

Fujitsu Laboratories LTD.

4. プロセッサ間通信

NET-VMS は各PEに分散配置したアクセスキーの状態に従い，全体として矛盾のない様にページの複写(Copy 処理) とページの単一化(Unify処理) を行う。ここではCopy処理の様子をプロセッサ１(P_1)からプロセッサ２(P_2)へ共有メモリページ(Page)を介して通信する例で説明する（図２）。

(1) まずページはP_1の実メモリにあり，P_1のみライトアクセス可能である。この状態ではLOCKキーによってページを他へ移動することが禁止されているためNET-VMS はページをP_2の実メモリに移動できない。

(2) P_1はアクセスを終了するとLOCKをリセット，SYNCをセットする。この結果P_1はページにアクセスできなくなり，NET-VMS はページの移動(Copy 処理) 可能となる。

(3) Copy処理はページをP_2の実メモリに複写すると共にP_1のCOPYとP_2のVALID,COPYをセット，P_1のSYNCをリセットする。この結果P_1，P_2共にページに対してリードアクセス可能となる。P_2がライトアクセスする場合はこれに連続して Unify処理を行う。

5. ネットワーク

NET-VMS の通信処理であるCopy/Unify処理の性能はプロセス間通信性能を決定する最大の要因となる。NET-VMS では通信処理高速化のために次の方式を採用する。

① Copy/Unify処理要求をブロードキャスト通信により他PEに同時に伝える。

② Copy/Unify処理要求を受信したPEは処理要求されているページアドレスを仮想記憶のアドレス変換機構を用いて 100ns程度で検査し，直ちに応答を返す。

以上の方式によれば，Copy/Unify処理を一度のブロードキャスト通信（要求送信と応答）で行えるので，一般の通信プロトコルを用いた場合より格段に高速な通信を実現できる。ブロードキャストは，例えばバスやリングによって効率的に実現でき，特に光ファイバリングは数百Mbpsの伝送速度が実現可能である。

この方式によって得られる性能をページイン（Copy処理）の場合について待行列モデルによる解析[2] で評価した結果を図３に示す。200Mbps の伝送路を用いて，16台のPEが 256バイトのページをそれぞれ2000

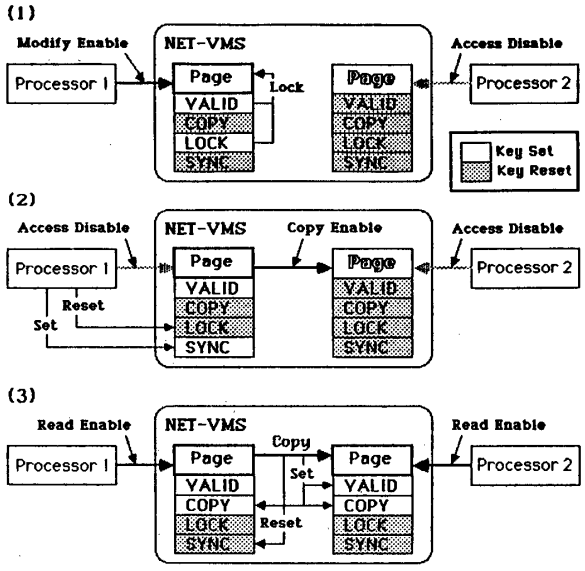

図２ ＮＥＴ－ＶＭＳの通信（Ｃｏｐｙ処理）

ページ／秒の割合でページインした場合のページイン時間は平均50μsとなる。

6. おわりに

NET-VMS はプロセス間通信を単一階層記憶化することにより，通信制御をプロセッサからメモリシステムにオフロードする。同時にブロードキャストと仮想記憶機構を融合し，ネットワーク結合による柔軟で大規模なシステムで数十MB／秒オーダの通信性能を実現する見通しを得た。今後はハードウェアプロトタイプ，ソフトウェアシステムの開発，評価を進める予定である。

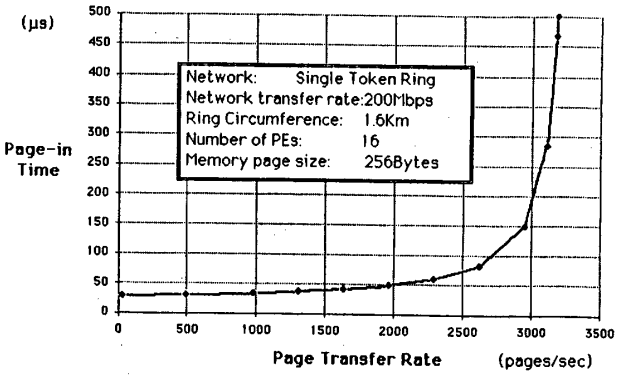

図３ 評価結果

〔参考文献〕

(1)陣崎他：ブロードキャストネットワークによる分散型単一階層仮想記憶システム，信学会研究会，CPSY 86-20，1986年７月

(2)W. Bux:Local-Area Subnetworks: A Performance Comparison, IEEE COM-29-10, Oct., 1981