

TOP-1 オペレーティング・システム

(3) メモリ管理

3P-6

河内谷清久仁 山崎秘砂 白鳥敏幸 森山孝男 穂積元一

日本アイ・ビー・エム株式会社 東京基礎研究所

1. はじめに

TOP-1^[1] オペレーティング・システム(以下、TOP-1 OSと略す)は、IBM AIX PS/2²をベースに、マルチ・プロセッサを管理するための機構を追加したOSである^[2]。AIX PS/2は、80386のページ変換機構を利用したメモリ管理を行なっているが、MP対応化にあたってはいくつかの手直しが必要となる。本稿では、MP化にあたって行なったメモリ管理コードの変更について説明する。

2. メモリ管理機構の概要

TOP-1 OSのベースである AIX PS/2は、80386のページ変換機構を利用したメモリ管理を行なっている。図1は80386のページ変換機構を示したものである^[3]。ページ変換はPD(ページ・ディレクトリ)とPT(ページ・テーブル)の2つの表を参照して行なわれる。

AIX PS/2は、デマンド・ページングとスワッピングの両方のメモリ管理機構をサポートしており、ページングではページ・フレームのイン/アウト、スワッピングではPD, PTを含めた全体のイン/アウトが行なわれる。TOP-1 OSでは、PD, PT, ページ・フレームの一部を除いて共有メモリ上に置かれており、カーネルPU, ユーザーPUの双方からアクセスされる(図2参照)。

3. MP化にあたっての変更点

メモリ管理機構をMP対応に変更するにあたって、TOP-1 OSでは次の4点について考慮した。

1. PDE, PTEのロック
2. TLBの整合性の維持
3. スワップ・アウト候補の決定
4. ページ・フォールト時の時間差

以下、これらについて問題となった点と解決法を順に説明する。

3.1. PDE, PTEのロック

マルチ・プロセッサ環境では複数のプロセッサが同時に動くため、1つのデータに対して同時にアクセスが起こらないようにする必要がある。TOP-1 OSでは、カーネル・コードを実行するプロセッサを1つ(カーネルPU)

に固定することで、大部分のカーネル・データに関しては同時にアクセスされないことが保証されている。しかし、PD, PTの各エントリ内のAccessedビットやDirtyビットはユーザーPUからも書き換えられるため、問題が起こる場合がある。図3aはその例である。ユーザーPUが立てたDirtyビットがカーネル・コードによって誤まって落とされてしまっている。これを避けるためには、このような変更を不可分に行なう必要がある。

TOP-1 OSでは、カーネル・コード内でPDE, PTEを変更している部分には明示的にバス・ロックをかけて、ユーザーPUからのアクセスを禁止している(図3b参照)。

※ 80386がAccessedビットやDirtyビットを立てる時には自動的にバス・ロックが行なわれる。

3.2. TLBの整合性の維持

80386はページ変換を高速に行なうために、過去の交換結果をTLB(Translation Lookaside Buffer)内にキャッシングしている。このため、カーネル・コードがページ・アウトなどのためにPDE, PTEを書き換える際には、TLBの内容も更新しなければならない。

TOP-1 OSではメッセージ・パッシングを用いてTLBの整合性を保っている。図4に具体的なページ・アウトの手順を示す。カーネルPUはまず、ページ・アウトするページのPresentビットを落とし、ユーザーPUがこのページを新たに使い始められないようにする。続いて、TOP-1のEverybody-broadcastメッセージを用いて全てのユーザーPUに割り込みをかける。ユーザーPUはこの割り込みを受け付けると、自分のTLBをフラッシュし、共有変数を通じてカーネルPUにフラッシュ完了を知らせる。カーネルPUは全てのプロセッサでTLBフラッシュが終わったのを確認してから実際にそのページをページ・アウトし解放する。

3.3. スワップ・アウト候補の決定

シングル・プロセッサの場合、カーネル・コードを実行中は他のユーザー・プロセスは停止していることが保証されており、AIX PS/2のカーネルにはそれを仮定して書かれている部分が存在する。このため、MP化にあたってはその部分を手直しする必要がある。メモリ管理に関し

TOP-1 Operating System (3) Memory Management

Kiyokuni Kawachiya, Hisa Yamasaki, Toshiyuki Shiratori, Takao Moriyama, and Motokazu Hozumi

IBM Research, Tokyo Research Laboratory, IBM Japan, Ltd.

ては、ページ・イン/アウト時にPresentビットを変更するタイミングの問題があり、3.2で示したような注意が必要である。その他に問題になるのは、スワップ・アウト候補のプロセスを選ぶ部分である。これについてTOP-1 OSでは、ユーザーPUで実行中のプロセスはスワップ・アウト候補から外すように変更を行なっている。

3.4. ページ・フォルト時の時間差

TOP-1 OS特有の問題としては、プロセスがユーザー・モードからカーネル・モードに移行する際のタイムラグがあげられる。メモリ管理に関しては、ユーザー・プロセスがページ・フォルトを起こしてから、カーネルPUで実際にページ・フォルト・ハンドラが動きだすまでの時間差が問題となる。例えば、この間にフォルトを起こしたページがページ・インされてしまう可能性があるため、ページ・フォルト・ハンドラでは、フォルトを起こしたページの状況が変わっていないかどうか調べる必要がある。

4. おわりに

本稿ではMP化にあたってのメモリ管理機構の変更について述べた。TOP-1 OSを例にして説明したが、ここで示した方法の大部分は一般のUNIX^{*2}のMP化の際にも適用可能である。

*1 AIXはIBM Corp.の商標、PS/2はIBM Corp.の登録商標です。

*2 UNIXはAT&Tおよびベル研究所が開発し、AT&TがライセンスしているOSの名称です。

【参考文献】

- [1] 鈴木他, 「高速並列処理ワークステーション(TOP-1)」, 情報処理学会第37回(昭和63年後期)全国大会論文集(1988)
- [2] 穂積他, 「TOP-1オペレーティング・システム」, 情報処理学会第39回(平成元年後期)全国大会論文集(1989)
- [3] "80386 Programmer's Reference Manual," Intel Corporation (1986)

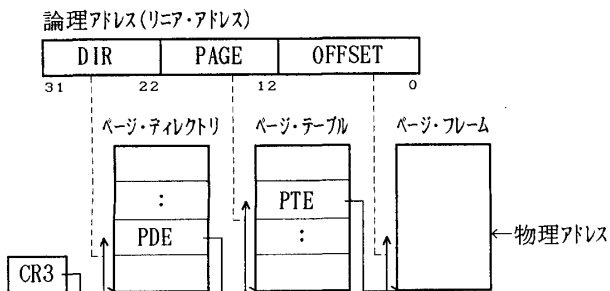


図1. 80386のページ変換機構

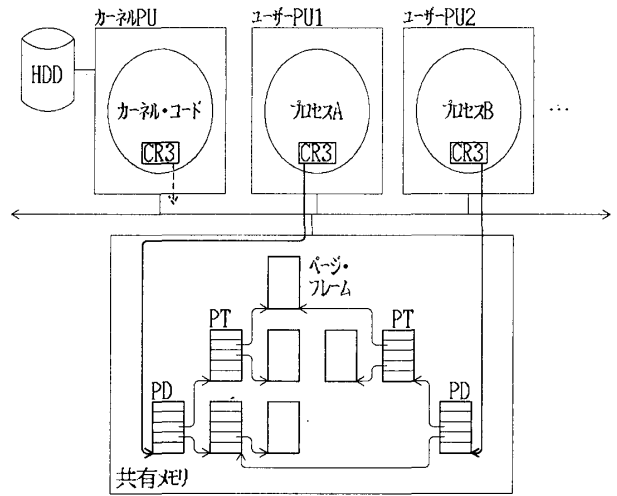
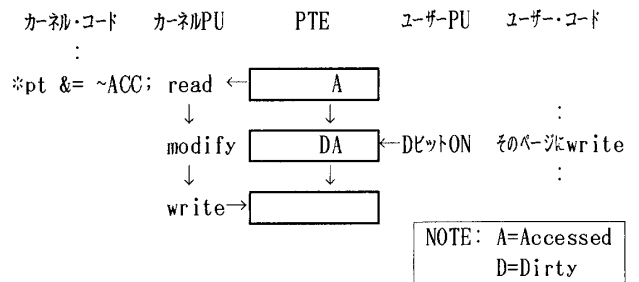


図2. TOP-1 OSのメモリ管理構造



a. カーネル、ユーザーPUから同時にアクセスが起こった例

lock_and(pt, ~ACC)

```
lock_and: movl 4(%esp),%eax / pt
          movl 8(%esp),%ecx / ~ACC
          lock andl %ecx, (%eax) / バス・ロックしてAND
          ret
```

b. 変更後のカーネルコード

図3. PTEのロック

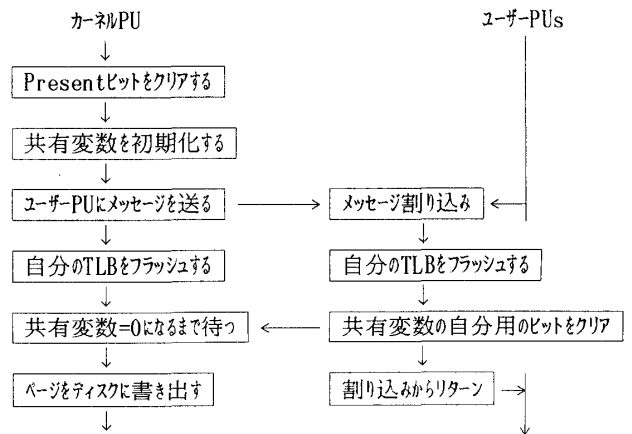


図4. ページ・アウトの手順