1. 講義資料:メモリ管理の基礎と仮想メモリ

第1章 メモリ管理の導入

● 1-1. なぜメモリ管理が必要か?

- コンピュータに搭載されるメモリ容量は増え続けていますが、それでもメモリは有限のリソースです。
- 一方で、ソフトウェアは大規模化し(プロセスの大型化)、小型デバイスでも高度なプログラムが実行されるようになっています。
- OSは、この限られたメモリ資源を複数のプロセスに効率よく、かつ安全に割り当てる役割を担います。これがメモリ管理です。

● 1-2. メモリの抽象化: アドレス空間

- o 最も単純な方法は、一度に一つのプログラムだけをメモリに配置することです(抽象化なし)。
- o しかし、複数のプログラムを同時にメモリ上で実行するためには、互いに干渉しないようにする 「保護 (protection)」 と、プログラムをメモリ上の任意の位置に配置可能にする 「再配置(relocation)」 という2つの課題を解決する必要があります。
- この解決策が**「アドレス空間」**という抽象概念です。OSは各プロセスに独立したメモリ空間(アドレス 空間)があるように見せかけます。
- ハードウェアの支援として、プログラムの開始アドレスを保持するベースレジスタと、そのサイズを示す リミットレジスタが用いられ、これにより保護と再配置が実現されます。

第2章 古典的なメモリ管理:スワッピング

• 2-1. スワッピングとは?

- 物理メモリに入りきらないプロセスを実行するため、プロセスのメモリ全体をディスクとの間でまるごと 移動させる手法です。
- これにより、メモリ上に一時的に存在しないプロセスも並行して実行できます。

● 2-2. 空き領域(ホール)の管理と割り当て

- スワッピングを繰り返すと、メモリ上に使用されていない領域(ホール)が点在するようになります。この管理方法として、ビットマップや連結リストなどがあります。
- 新しいプロセスをロードする際、どのホールを選ぶか決定するアルゴリズムが存在します。
 - First fit: 最初に見つかった十分な大きさのホールを割り当てる。高速です。
 - Next fit: First fitの変種で、前回割り当てた場所から探索を開始する。
 - **Best fit**: プロセスにちょうど良い、最小のホールを探す。全探索が必要で遅く、非常に小さな無駄なホールを生みやすいです。
 - Worst fit: 最も大きなホールを割り当てる。これも良い結果は出ないとされています。
- **考察**: スワッピング環境では、これらのアルゴリズムの速度よりも、**桁違いに遅いディスクI/Oがボトル ネック**となるため、いかにホールの発生を抑えるかが重要になります。

第3章 現代のメモリ管理:仮想メモリとページング

● 3-1. 仮想メモリの概念

- 物理メモリのサイズより大きなプロセスを実行するための仕組みです。
- プロセスのうち、**実行に必要な部分だけを物理メモリにロード**し、不要になった部分はディスクに追い出します。
- このアドレスの検出はハードウェア(MMU)が、データの転送はOSが自動的に行うため、プログラマは物理メモリのサイズを意識する必要がありません。

● 3-2. ページングによる仮想メモリの実現

○ 仮想メモリを実現する最も一般的な方法がページングです。

- 仮想アドレス空間と物理メモリを 「ページ」と呼ばれる固定長のブロックに分割して管理します。
- ハードウェアであるMMU (Memory Management Unit) が、プログラムが発行する仮想アドレスを、ページテーブルを参照して物理アドレスに変換します。

● 3-3. ページテーブルの課題と高速化

- **課題**: アドレス空間が大きくなると、ページテーブル自体のサイズも非常に大きくなり、高速なメモリに収まらなくなります。
- 解決策:
 - 1. **マルチレベルページテーブル**: ページテーブルを階層化し、参照の局所性を利用して、頻繁に使う部分だけをメモリに置きます。
 - 2. **TLB (Translation Look-aside Buffer)**: ページテーブルのエントリをキャッシュするための高速な 連想メモリです。アドレス変換の大半をTLBで処理することで、メモリアクセスの回数を減らしま す。
 - 3. **逆引きページテーブル**: 64ビットのような巨大なアドレス空間に対応するため、物理ページフレームを基準にテーブルを作成する方式です。

第4章 ページ置換アルゴリズム

● 4-1. ページフォールトとページ置換

- o プロセスがアクセスしようとしたページが物理メモリ上に存在しない場合、 「ページフォールト」 という例外が発生します。
- o OSはディスクから該当ページをメモリに読み込みますが、空きページフレームがない場合は、メモリ上のいずれかのページをディスクに追い出す(置換する)必要があります。
- どのページを追い出すかを決めるのが**ページ置換アルゴリズム**です。

● 4-2. 代表的なアルゴリズム

- 最適(Optimal): 将来最も参照されないページを置換する。実現不可能ですが、他のアルゴリズムの性能評価の基準となります。
- o FIFO (First-In, First-Out): 最も古くメモリにロードされたページを置換する。単純ですが、頻繁に使われるページを追い出してしまう可能性があります。
- セカンドチャンス / クロック: FIFOの改良版。参照されたことがあるページ(参照ビットR=1)にはもう一度チャンスを与えます。
- LRU (Least Recently Used): 最も長い間参照されていないページを置換する。性能は良いですが、完全な実装はコストが高いです。
- **ワーキングセットモデル**: プロセスが現在頻繁に使用しているページの集合(ワーキングセット)をメモリ上に維持しようとします。**スラッシング**(ページフォールトが多発し性能が低下する現象)を防ぐのに役立ちます。
- WSClock: ワーキングセットとクロックアルゴリズムを組み合わせた、現実に広く利用されている効率的なアルゴリズムです。

2. 重要用語の抽出と説明

• アドレス空間 (Address Space)

o OSが各プロセスに提供する、独立したメモリの抽象的なビューです。これにより、プロセスは他のプロセスやOSのメモリを意識することなく動作できます。

• スワッピング (Swapping)

○ プロセスのメモリイメージ全体を、主メモリとディスクの間で移動させるメモリ管理手法です。

● 仮想メモリ (Virtual Memory)

○ 物理メモリの大きさに制限されずに、より大きなプロセスを実行可能にする技術です。プログラムの必要 な部分だけを物理メモリに配置して動作させます。

• ページング (Paging)

○ 仮想メモリを実現する主要な手法で、アドレス空間を「ページ」という固定長の単位に分割して管理しま

す。

• MMU (Memory Management Unit)

○ 仮想アドレスを物理アドレスに変換する役割を担うハードウェアです。仮想メモリに不可欠な要素です。

• ページテーブル (Page Table)

○ 仮想ページ番号と、それが格納されている物理ページフレーム番号との対応関係を記録したデータ構造で す。アドレス変換に用いられます。

• ページフォールト (Page Fault)

○ アクセスしようとしたページが物理メモリ上に存在しない場合に発生するハードウェア割り込み(例外)です。これをきっかけにOSが必要なページをディスクからロードします。

• TLB (Translation Look-aside Buffer)

o ページテーブルのエントリを保持する高速なキャッシュメモリです。アドレス変換の速度を大幅に向上させます。

• ページ置換アルゴリズム (Page Replacement Algorithm)

o ページフォールトが発生し、物理メモリに空きがない場合に、どのページをディスクに追い出すかを決定するためのアルゴリズムです。

• ワーキングセット (Working Set)

• ある特定の時間において、プロセスが効率的に実行するためにメモリ上に保持しておくべきページの集合のことです。「参照の局所性」というプログラムの性質に基づいています。

• スラッシング (Thrashing)

ワーキングセットが物理メモリに収まりきらないなどの理由でページフォールトが頻発し、システムの処理能力が極端に低下する現象です。 (注: ソースでは "slashing" と記載されていますが、一般的には "thrashing" と呼ばれます)

3. 追加で学習すべき項目

提供された資料はメモリ管理の基本から仮想メモリ、ページ置換アルゴリズムまでを幅広く網羅していますが、さら に理解を深めるために以下の項目について学習することをお勧めします。

● セグメンテーションとの比較と組み合わせ

o 説明: 資料の目次にはセグメンテーションの項目がありますが、詳細な説明は省略されています。ページングがメモリを物理的な固定長ブロックで管理するのに対し、セグメンテーションはコード、データ、スタックといった論理的な単位で可変長のメモリ領域を管理します。両者の長所・短所を比較し、Intel x86アーキテクチャのように両者を組み合わせたハイブリッドなメモリ管理(セグメント方式による論理的な保護と、ページング方式による物理メモリ管理)がどのように動作するのかを学ぶことで、メモリ保護や共有の仕組みについてより深い知見が得られます。

● 64ビットアーキテクチャにおける巨大アドレス空間の管理

。 説明: 資料では主に32ビットアドレス空間を前提にページテーブルのサイズ問題が議論されています。現代の主流である64ビットアーキテクチャでは仮想アドレス空間が天文学的に広大になるため、マルチレベルページテーブルも4レベルや5レベルといった、より多段の構成が一般的です。資料で触れられている逆引きページテーブルも一つのアプローチですが、現代のCPU(例: x86-64)が実際にどのようにしてこの巨大なアドレス空間を効率的に管理しているのか、その具体的なページテーブル構造やアドレッシングモードを学ぶことは、現代のコンピュータシステムの動作を理解する上で非常に重要です。

● OSカーネル自身のメモリ管理(物理メモリの割り当て)

。 説明: 資料は主にユーザープロセスのための仮想メモリ管理に焦点を当てています。しかし、OSカーネル自身も動作のために物理メモリを動的に確保・解放する必要があります。特に、デバイスドライバなどは物理的に連続したメモリ領域を要求することがありますが、ページングによって物理メモリが断片化していると、このような要求に応えるのは困難です。この問題を解決するためにカーネル内部で用いられるバディシステム (Buddy System) や スラブアロケータ (Slab Allocator) といった物理メモリ管理アルゴリズムについて学習することで、OSのより低レベルな動作原理への理解が深まります。