

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5146174号
(P5146174)

(45) 発行日 平成25年2月20日 (2013. 2. 20)

(24) 登録日 平成24年12月7日 (2012.12.7)

(51) Int. Cl.		F I			
G 0 6 F	12/02	(2006.01)	G O 6 F	12/02	5 1 O M
G 0 6 F	12/10	(2006.01)	G O 6 F	12/10	5 5 1 B
G 0 6 F	9/46	(2006.01)	G O 6 F	9/46	3 5 O
			G O 6 F	12/10	5 5 3 Z

請求項の数 6 (全 16 頁)

(21) 出願番号	特願2008-192965 (P2008-192965)	(73) 特許権者	000005223 富士通株式会社
(22) 出願日	平成20年7月28日 (2008. 7. 28)		神奈川県川崎市中原区上小田中4丁目1番1号
(65) 公開番号	特開2010-33206 (P2010-33206A)	(74) 代理人	100119161 弁理士 重久 啓子
(43) 公開日	平成22年2月12日 (2010. 2. 12)	(74) 代理人	100111822 弁理士 渡部 章彦
審査請求日	平成23年4月18日 (2011. 4. 18)	(72) 発明者	岩松 昇 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		(72) 発明者	西口 直樹 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 仮想マシンモニタ装置およびプログラム、ならびに仮想マシン間メモリ共有管理方法

(57) 【特許請求の範囲】

【請求項1】

物理メモリと仮想的に多重化された複数の仮想メモリとの対応を管理して複数の仮想マシンを動作させるコンピュータシステムの仮想マシンモニタ装置であって、

仮想マシンを一意に識別する仮想マシンIDと、前記仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを一意に識別するオブジェクトIDと、前記オブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報とを保存するマッピング情報データベースと、

前記仮想マシンIDと前記オブジェクトIDと前記仮想ページマッピング情報と、前記仮想ページがマッピングされる前記物理メモリのページを示すマシンページマッピング情報と、同一のオブジェクトIDの存在を示す存在フラグとを対応づける仮想ページ・オブジェクト対応テーブルと、

前記オブジェクトIDと、前記マシンページマッピング情報と、前記マシンページマッピング情報が示すページを参照する仮想マシン数を示す参照仮想マシン数とを対応づけるオブジェクト・マシンページ対応テーブルと、

仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを読み込み、読み込んだオブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報を生成し、当該オブジェクトファイルのオブジェクトIDと前記オブジェクトファイルが動作する仮想マシンの仮想マシンIDと前記生成した仮想ページマッピング情報とを前記マッピング情報データベースに登録するオブジ

エクトファイルマッピング情報生成部と、

前記マッピング情報データベースを参照して、前記仮想マシンの起動に対応して前記仮想ページ・オブジェクト対応テーブルに該当する仮想マシンのデータを追加し、前記仮想マシンの終了に対応して前記仮想ページ・オブジェクト対応テーブルに該当する仮想マシンのデータを削除し、前記仮想ページ・オブジェクト対応テーブルに同一のオブジェクトIDが複数存在する場合に、複数存在するオブジェクトIDの存在フラグに有効を設定し、前記存在フラグが有効となっている仮想ページに対応する前記物理メモリのページの割り当てまたは変更があった場合に、前記仮想ページ・オブジェクト対応テーブルの該当するマシンページマッピング情報を更新する仮想ページ・オブジェクト対応管理部と、

前記オブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページから対応するハードウェアのページへマッピングしている仮想マシンを検出して前記オブジェクト・マシンページ対応テーブルの参照仮想マシン数を更新するオブジェクト・マシンページ対応管理部と、

前記仮想マシンの仮想メモリのページのアドレスと自装置内の物理メモリのページのアドレスとを変換する仮想ページ・マシンページ変換テーブルを備えて、当該仮想メモリのページに前記物理メモリのページを割り当てるページマッピング部と、

前記仮想メモリへのアクセス時に、アクセスする仮想マシンIDと仮想メモリのアクセス先ページとをキーとして前記仮想ページ・オブジェクト対応テーブルを参照し、前記アクセス先ページに対応するオブジェクトIDの存在フラグが有効であるかを判定し、さらに、当該判定において存在フラグが有効である場合に、前記対応するオブジェクトIDをキーとして前記オブジェクト・マシンページ対応テーブルを参照し、当該オブジェクトIDのアクセス先ページに対応する前記物理メモリのページが存在するかを判定する共有可能マシンページ検出部とを備えて、

前記ページマッピング部は、前記共有可能マシンページ検出部が、前記アクセス先ページに対応する前記物理メモリのページが存在する場合に、当該アクセス先の仮想ページに対して前記判定した物理メモリのページを参照するように前記仮想ページ・マシンページ変換テーブルを更新し、

前記オブジェクト・マシンページ対応管理部は、前記オブジェクト・マシンページ対応テーブルの該当するオブジェクトの参照仮想マシン数を更新する

ことを特徴とする仮想マシンモニタ装置。

【請求項2】

前記ページマッピング部は、前記物理メモリのページをコピーオンライトで管理し、前記仮想マシン間で共有される前記物理メモリのページに対する前記仮想マシンからの書き込み要求があった場合に、コピーオンライトによって前記物理メモリの新しいページを割り当てる

ことを特徴とする請求項1記載の仮想マシンモニタ装置。

【請求項3】

前記オブジェクトファイルマッピング情報生成部は、前記仮想マシン上で動作するゲストOSを構成するオブジェクトファイルの変更を検出して、変更を検出した仮想マシンの仮想メモリマッピング情報を再生成し、

前記仮想ページ・オブジェクト対応管理部は、前記再生成された仮想メモリマッピング情報をもとに前記仮想ページ・オブジェクト対応テーブルを更新する

ことを特徴とする請求項1または請求項2に記載の仮想マシンモニタ装置。

【請求項4】

前記オブジェクトファイルマッピング情報生成部は、前記仮想マシンのゲストOSを構成するオブジェクトファイルのプログラム実行部がマッピングされる仮想ページの候補を複数用意し、前記仮想ページの候補に対する仮想ページマッピング情報を生成し、前記マッピング情報データベースに保存する

ことを特徴とする請求項1ないし請求項3のいずれか一項に記載の仮想マシンモニタ装置。

10

20

30

40

50

【請求項5】

物理メモリと仮想的に多重化された複数の仮想メモリとの対応を管理して複数の仮想マシンを動作させるコンピュータを、

仮想マシンを一意に識別する仮想マシンIDと、前記仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを一意に識別するオブジェクトIDと、前記オブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報とを保存するマッピング情報データベースと、

前記仮想マシンIDと前記オブジェクトIDと前記仮想ページマッピング情報と、前記仮想ページがマッピングされる前記物理メモリのページを示すマシンページマッピング情報と、同一のオブジェクトIDの存在を示す存在フラグとを対応づける仮想ページ・オブジェクト対応テーブルと、

10

前記オブジェクトIDと、前記マシンページマッピング情報と、前記マシンページマッピング情報が示すページを参照する仮想マシン数を示す参照仮想マシン数とを対応づけるオブジェクト・マシンページ対応テーブルと、

仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを読み込み、読み込んだオブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報を生成し、当該オブジェクトファイルのオブジェクトIDと前記オブジェクトファイルが動作する仮想マシンの仮想マシンIDと前記生成した仮想ページマッピング情報とを前記マッピング情報データベースに登録するオブジェクトファイルマッピング情報生成部と、

20

前記マッピング情報データベースを参照して、前記仮想マシンの起動に対応して前記仮想ページ・オブジェクト対応テーブルに該当する仮想マシンのデータを追加し、前記仮想マシンの終了に対応して前記仮想ページ・オブジェクト対応テーブルに該当する仮想マシンのデータを削除し、前記仮想ページ・オブジェクト対応テーブルに同一のオブジェクトIDが複数存在する場合に、複数存在するオブジェクトIDの存在フラグに有効を設定し、前記存在フラグが有効となっている仮想ページに対応する前記物理メモリのページの割り当てまたは変更があった場合に、前記仮想ページ・オブジェクト対応テーブルの該当するマシンページマッピング情報を更新する仮想ページ・オブジェクト対応管理部と、

前記オブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページから対応するハードウェアのページへマッピングしている仮想マシンを検出して前記オブジェクト・マシンページ対応テーブルの参照仮想マシン数を更新するオブジェクト・マシンページ対応管理部と、

30

前記仮想マシンの仮想メモリのページのアドレスと自装置内の物理メモリのページのアドレスとを変換する仮想ページ・マシンページ変換テーブルを備えて、当該仮想メモリのページに前記物理メモリのページを割り当てるページマッピング部と、

前記仮想メモリへのアクセス時に、アクセスする仮想マシンIDと仮想メモリのアクセス先ページとをキーとして前記仮想ページ・オブジェクト対応テーブルを参照し、前記アクセス先ページに対応するオブジェクトIDの存在フラグが有効であるかを判定し、さらに、当該判定において存在フラグが有効である場合に、前記対応するオブジェクトIDをキーとして前記オブジェクト・マシンページ対応テーブルを参照し、当該オブジェクトIDのアクセス先ページに対応する前記物理メモリのページが存在するかを判定する共有可能マシンページ検出部とを備えて、

40

前記ページマッピング部は、前記共有可能マシンページ検出部が、前記アクセス先ページに対応する前記物理メモリのページが存在する場合に、当該アクセス先の仮想ページに対して前記判定した物理メモリのページを参照するように前記仮想ページ・マシンページ変換テーブルを更新し、

前記オブジェクト・マシンページ対応管理部は、前記オブジェクト・マシンページ対応テーブルの該当するオブジェクトの参照仮想マシン数を更新する仮想マシンモニタ装置として機能させる

ことを特徴とする仮想マシンモニタプログラム。

50

【請求項6】

物理メモリと仮想的に多重化された複数の仮想メモリとの対応を管理して複数の仮想マシンを動作させるコンピュータシステムの仮想マシンモニタ装置が実行する仮想マシン間メモリ共有管理方法であって、

仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを読み込むオブジェクトファイル読み込む処理と、

仮想マシンを一意に識別する仮想マシンIDと、前記仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを一意に識別するオブジェクトIDと、前記オブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報とを保存するマッピング情報データベースを備えて、前記読み込んだオブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報を生成し、当該オブジェクトファイルのオブジェクトIDと前記オブジェクトファイルが動作する仮想マシンの仮想マシンIDと前記生成した仮想ページマッピング情報とを前記マッピング情報データベースに登録する処理過程と、

10

前記仮想マシンIDと前記オブジェクトIDと前記仮想ページマッピング情報と、前記仮想ページがマッピングされる前記物理メモリのページを示すマシンページマッピング情報と、同一のオブジェクトIDの存在を示す存在フラグとを対応づける仮想ページ・オブジェクト対応テーブルを備えて、前記マッピング情報データベースを参照して、前記仮想マシンの起動または終了に対応して前記仮想ページ・オブジェクト対応テーブルに該当する仮想マシンのデータを追加または削除する処理過程と、

20

前記仮想ページ・オブジェクト対応テーブルに同一のオブジェクトIDが複数存在する場合に、前記仮想ページ・オブジェクト対応テーブルの複数存在するオブジェクトIDの存在フラグに有効を設定する処理過程と、

前記存在フラグが有効となっている仮想ページに対応する前記物理メモリのページの割り当てまたは変更があった場合に、前記仮想ページ・オブジェクト対応テーブルの該当するマシンページマッピング情報を更新する処理過程と、

前記オブジェクトIDと、前記マシンページマッピング情報と、前記マシンページマッピング情報が示すページを参照する仮想マシン数を示す参照仮想マシン数とを対応づけるオブジェクト・マシンページ対応テーブルを備えて、前記オブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページから対応するハードウェアのページへマッピングしている仮想マシンを検出した場合に、前記オブジェクト・マシンページ対応テーブルの参照仮想マシン数を更新する処理過程と、

30

前記仮想マシンの仮想メモリのページのアドレスと自装置内の物理メモリのページのアドレスとを変換する仮想ページ・マシンページ変換テーブルを備えて、当該仮想メモリのページに前記物理メモリのページを割り当てる処理過程と、

前記仮想メモリへのアクセス時に、アクセスする仮想マシンIDと仮想メモリのアクセス先ページとをキーとして前記仮想ページ・オブジェクト対応テーブルを参照し、前記アクセス先ページに対応するオブジェクトIDの存在フラグが有効であるかを判定する処理過程と、

40

前記判定において存在フラグが有効である場合に、前記対応するオブジェクトIDをキーとして前記オブジェクト・マシンページ対応テーブルを参照し、当該オブジェクトIDのアクセス先ページに対応する前記物理メモリのページが存在するかを判定する処理過程とを備えて、

前記アクセス先ページに対応する前記物理メモリのページが存在する場合に、当該アクセス先の仮想ページに対して前記判定した物理メモリのページを参照するように前記仮想ページ・マシンページ変換テーブルを更新する処理過程とを備える

ことを特徴とする仮想マシン間メモリ共有管理方法。

50

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、1台の実マシン上で複数の仮想マシンを動作させる仮想化技術において、仮想マシン間のメモリ共有を可能にする仮想マシンモニタ装置およびプログラム、ならびに仮想マシンモニタ装置が実行する仮想マシン間メモリ共有管理方法に関する。

【0002】

より詳しくは、仮想マシンに割り当てられるメモリ空間を実マシンのメモリ空間にマッピングする際に、一部の実マシンのメモリ空間を複数の仮想マシンのメモリ空間から参照させることにより、仮想マシン間でメモリを共有可能とするメモリ管理機能を備える仮想マシンモニタ技術に関する。

【背景技術】**【0003】**

仮想化技術では、1台の物理計算機上に複数の仮想マシンを作成し、仮想マシンはその上で動作するOS（ゲストOSと呼ぶ）に対し、あたかも独立したハードウェア上で動作しているかのように見せかける。複数の独立したゲストOSを1台のハードウェアに集約して実行し、CPU・ストレージなどのハードウェア資源を効率的に運用できるなどのメリットがある。

【0004】

一方で、1つのハードウェア・OS上で複数のサーバアプリケーションを動作させる場合と比較した仮想化のデメリットとしては、メモリ利用効率の悪さが挙げられる。例えば、1つのOS上でA、Bという2つのサーバアプリケーションを動かす場合と、1つのOS上でそれぞれA、Bを動かす2つの独立した仮想マシンとを比較すると、後者では各仮想マシンがそれぞれ個別にOSなどのプログラム実行コードをメモリ上に持つことになり、その分仮想マシン全体で消費するメモリ量が大きくなる。

【0005】

仮想化においてメモリ利用効率を改善するには、仮想マシンに割り当てられているページで複数の仮想マシン間で同一内容のものを共通で参照すればよい。

【0006】

仮想マシンに割り当てられているページから同一内容のページを探しだす従来技術として、仮想化ソフトウェアであるVMware ESX Serverが行っている方法がある。このVMware ESX Serverでは、ゲストOSのページ領域をスキャンし、実行属性または読み取り属性のページ（CPUから確認可能で一般にプログラムの実行コードが置かれるページ）のハッシュ値をインデックス化し、そのハッシュテーブルを用いてハッシュ値が一致するページを検出し、検出したページのページ内容を比較することによって共有可能なページを特定し、メモリの共有を行っている（例えば、特許文献1参照）。

【0007】

また、別の従来技術として、1台の計算機上で動作する複数のOS間で、OSの構成モジュールを共有する方法がある。この方法では、OSのマイクロカーネルをロードする際に、カーネルローダが、マイクロカーネルのオブジェクトを複数OSで参照可能な共通領域に配置し、複数OSで参照可能なオブジェクトのリスト管理とページテーブルへの設定を行う（例えば、特許文献2参照）。

【特許文献1】 米国特許US 6789156号公報

【特許文献2】 特開2007-35066号公報

【発明の開示】**【発明が解決しようとする課題】**

10

20

30

40

50

【 0 0 0 8 】

仮想マシンに割り当てられるページから同一内容のページを探しだす従来技術では、基本的にページ内容を比較するため、ハードウェアの搭載メモリ量、仮想マシン数等が増加するにつれて、計算負荷が大きくなるという問題があった。

【 0 0 0 9 】

従来技術により、仮想マシン間でメモリを共有しようとする場合に、仮想マシン実行中に仮想マシンの使用する全ページ、例えば、通常のページサイズ4KBの場合には1GB当たり262144ページを参照し、計算を行う必要があり、仮想マシン数や使用メモリ量に応じて計算量がリニアに増加していく。

【 0 0 1 0 】

また、従来のOSの構成モジュールを共有する方法では、共有可能なページがマイクロカーネルモジュールに限定されるうえ、共有の設定もOS初期化時のみに行われる。そのため、メモリ共有対象が仮想マシン起動時に初期化されるOSのマイクロカーネルモジュールに制限され、共有化の適用範囲が非常に狭くなる。

【 0 0 1 1 】

本発明は、かかる問題に鑑みてなされたものであり、仮想マシン間で同一内容のページを少ない計算負荷で検出して、仮想マシン上のOSを構成する実行ファイル全般に広く適用可能な、仮想マシン間でのメモリ共有を可能とする仮想マシンモニタ装置およびプログラム、ならび仮想マシンモニタ装置が実行する仮想マシン間メモリ共有管理方法を提供することを目的とする。

【 課題を解決するための手段 】

【 0 0 1 2 】

本発明に仮想マシンモニタ装置は、仮想マシンを一意に識別する仮想マシンID、仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを一意に識別するオブジェクトID、およびオブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報を保存するマッピング情報データベースと、仮想マシンID、オブジェクトID、仮想ページマッピング情報、仮想ページがマッピングされる物理メモリのページを示すマシンページマッピング情報、および、同一のオブジェクトIDの存在を示す存在フラグを対応づける仮想ページ・オブジェクト対応テーブルと、オブジェクトID、マシンページマッピング情報、およびマシンページマッピング情報が示すページを参照する仮想マシン数を示す参照仮想マシン数を対応づけるオブジェクト・マシンページ対応テーブルとを備える。

【 0 0 1 3 】

そして、仮想マシンモニタ装置は、仮想マシン上で動作するゲストOSを構成するオブジェクトファイルを読み込むオブジェクトファイル読み込み、読み込んだオブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページを示す仮想ページマッピング情報を生成し、このオブジェクトファイルのオブジェクトIDとオブジェクトファイルが動作する仮想マシンの仮想マシンIDと生成した仮想ページマッピング情報とをマッピング情報データベースに登録する。また、マッピング情報データベースを参照して、仮想マシンの起動または終了に対応して、仮想ページ・オブジェクト対応テーブルに該当する仮想マシンのデータを追加または削除する。

【 0 0 1 4 】

さらに、仮想ページ・オブジェクト対応テーブルに同一のオブジェクトIDが複数存在する場合に、仮想ページ・オブジェクト対応テーブルに複数存在するオブジェクトIDの存在フラグに有効を設定し、存在フラグが有効となっている仮想ページに対応する物理メモリのページの割り当てまたは変更があった場合に、仮想ページ・オブジェクト対応テーブルの該当するマシンページマッピング情報を更新し、オブジェクトファイル中のプログラム実行コード部がマッピングされる仮想メモリのページから対応するハードウェアのページへマッピングしている仮想マシンを検出した場合に、オブジェクト・マシンページ対応テーブルの参照仮想マシン数を更新する。

10

20

30

40

50

【 0 0 1 5 】

また、仮想マシンモニタ装置は、仮想マシンの仮想メモリのページのアドレスと自装置内の物理メモリのページのアドレスとを変換する仮想ページ・マシンページ変換テーブルを備えて、仮想メモリのページに前記物理メモリのページを割り当て、仮想メモリへのアクセス時に、アクセスする仮想マシンIDと仮想メモリのアクセス先ページとをキーとして仮想ページ・オブジェクト対応テーブルを参照し、アクセス先ページに対応するオブジェクトIDの存在フラグが有効であるかを判定して、この判定において存在フラグが有効である場合に、対応するオブジェクトIDをキーとしてオブジェクト・マシンページ対応テーブルを参照し、このオブジェクトIDのアクセス先ページに対応する物理メモリのページが存在するかを判定する。

10

【 0 0 1 6 】

そして、アクセス先ページに対応する物理メモリのページが存在する場合に、このアクセス先の仮想ページに対して判定した物理メモリのページを参照するように仮想ページ・マシンページ変換テーブルを更新する。

【 0 0 1 7 】

これにより、仮想マシン間で同一の内容のページを少ない計算負荷で検出し、仮想マシン上のゲストOSを構成する事項ファイル全般に広く適用可能なメモリ共有を実現することができる。

【 0 0 1 8 】

また、開示する装置のページマッピング部は、物理メモリのページをコピーオンライトで管理し、仮想マシン間で共有される物理メモリのページに対する仮想マシンからの書き込み要求があった場合に、コピーオンライトによって物理メモリの新しいページを割り当てることができる。

20

【 0 0 1 9 】

これにより、共有中の物理メモリのページが実行中の仮想マシン（ゲストOS）によって書き換えられようとする場合に、共有中の当該ページを保持したまま、書き換えを行うゲストOS用に新たな物理メモリのページを割り当てることができる。

【 0 0 2 0 】

また、開示する装置のオブジェクトファイルマッピング情報生成部は、仮想マシン上で動作するゲストOSを構成するオブジェクトファイルの変更を検出して、変更を検出した仮想マシンの仮想メモリマッピング情報を再生成し、仮想ページ・オブジェクト対応管理部は、前記再生成された仮想メモリマッピング情報をもとに前記仮想ページ・オブジェクト対応テーブルを更新することができる。

30

【 0 0 2 1 】

これにより、仮想マシンのゲストOSを構成するオブジェクトファイルが更新される場合に、この更新を検知して、該当するオブジェクトファイルの仮想ページ・オブジェクト対応テーブルを更新することができる。

【 0 0 2 2 】

また、開示する装置のオブジェクトファイルマッピング情報生成部は、仮想マシンのゲストOSを構成するオブジェクトファイルのプログラム実行部がマッピングされる仮想ページの候補を複数用意し、前記仮想ページの候補に対する仮想ページマッピング情報を生成し、前記マッピング情報データベースに保存することができる。

40

【 0 0 2 3 】

これにより、オブジェクトファイルのマッピングされる仮想ページが重複しているために再配置される可能性がある場合、マッピングされるアドレスに予め複数の仮想ページの候補があることが分かっている場合等にも対応することができる。

【 発明の効果 】

【 0 0 2 4 】

本発明によれば、物理メモリと仮想的に多重化された複数の仮想メモリとを有して複数の仮想マシンを動作させるコンピュータシステムにおいて、計算負荷が小さくかつ適用範

50

囲が広い仮想マシン間のメモリ共有を実現することができる。

【発明を実施するための最良の形態】

【0025】

図1は、本発明にかかる仮想マシンモニタ装置を備えたコンピュータのシステム構成例を示す図である。

【0026】

仮想マシンモニタ装置(VMM)1は、物理メモリと仮想的に多重化された複数の仮想メモリとを管理して複数の仮想マシン(VM1, ..., VMn)を動作させるコンピュータに構成される。

【0027】

図1において、コンピュータのハードウェア資源(H/W)として、中央演算ユニット(CPU)、メモリ管理ユニット(MMU)、外部記憶装置(HDD)、メモリ(MEM)からなる。メモリ管理ユニットは通常、中央演算ユニットに内蔵されている。

【0028】

仮想マシンモニタ装置1は、1台のコンピュータを複数のコンピュータとして仮想的に利用するための処理手段であり、コンピュータのハードウェア資源を区分けして、複数の仮想マシンVMに割り当てて利用させるように管理する。

【0029】

仮想マシンVM上のゲストOS(Guest OS:オペレーティングシステム)のメモリアクセスは、仮想マシンモニタ装置1のメモリマッピング部2によって、実際の物理メモリMEMへのアクセスに変換される。

【0030】

仮想マシン(VM1, ..., VMn)は、仮想マシンモニタ装置1により、仮想的にあたかも独立したコンピュータのように動作する。仮想マシンの中央演算ユニット(vCPU)、仮想マシンのメモリ管理ユニット(vMMU)、仮想マシンの外部記憶装置(vHDD)、仮想マシンのメモリ(vMEM)等で構成される仮想ハードウェア資源(vH/W)上でゲストOSを実行し、ゲストOSによりアプリケーションAPPを実行する。

【0031】

ところで、多くのOSは仮想メモリと呼ばれる実際の物理メモリ以上のメモリがあるかのようにふるまう機能を持ち、プロセス毎に独立した仮想メモリ空間を割り当て、プロセスの切り替え毎に仮想メモリ空間も切り替えている。仮想メモリ空間にはOSの実行コードとプロセスの実行コードが配置されている。

【0032】

メモリ管理ユニットMMUは、「ページング」と呼ばれる手法によって、アドレスをページ単位に分割して管理する。メモリ管理ユニットMMUは、仮想メモリ空間のページ(以下「仮想ページ」と呼ぶ)を実際の物理メモリ空間のページ(以下「マシンページ」と呼ぶ)に対応づけるために、ページテーブルという変換テーブルを用いてメモリ管理を行う。

【0033】

OSは、常に全ての仮想ページを物理ページにマッピングしているわけではない。アクセス先となる仮想ページだけを物理メモリにマッピングする「デマンドページング」と呼ばれる手法を用いて、物理メモリを効率的に利用する。

【0034】

図1に示すような仮想化されたコンピュータシステムにおいて、実際の物理メモリMEMは、仮想マシンモニタ装置1が管理している。仮想マシンモニタ装置1は、仮想マシンVMにあたかもその仮想マシンVMの物理メモリがあるかのように見せかけている。

【0035】

仮想ページへのアクセスは、仮想マシンVMの仮想メモリ管理ユニットvMMUと、仮想マシンモニタ装置1が管理するメモリ管理ユニットMMUの二重のアドレス変換を通して、マシンページへのアクセスに変換される。

10

20

30

40

50

【 0 0 3 6 】

仮想マシンVMのゲストOSを構成するオブジェクトファイルは、実行形式ファイル（拡張子が“exe”のファイル）、ライブラリシステムのファイル（拡張子が“dll”のファイル）等のプログラムコードを含む広義のオブジェクトファイルである。

【 0 0 3 7 】

オブジェクトファイルは、ヘッダ部分に、アーキテクチャ・バージョンなどの情報を持つファイルヘッダ、ファイルをそのオフセットでセグメントに区分しセグメントの属性と実行時にマッピングされる仮想メモリのアドレス情報を持つプログラムヘッダ等を持つ。さらに、実行時にプログラム実行コードとしてそのままメモリ上にマッピングされるテキストセグメント、プログラムの使用するデータとしてメモリ上にマッピングされるデータセグメント等を持つ。

10

【 0 0 3 8 】

オブジェクトファイルは仮想マシンVMの仮想外部記憶装置vHDD上のファイルであるが、一般に仮想外部記憶装置vHDDは、仮想マシンモニタ装置（VMM）1が管理する実際の外部記憶装置HDD上のパーティションの一部やファイルであり、仮想マシンモニタ装置（VMM）1に管理される。

【 0 0 3 9 】

まず、本発明にかかる仮想マシンモニタ装置の動作原理を説明する。

【 0 0 4 0 】

仮想マシンモニタ装置1は、仮想マシンVM1のゲストOSを構成するオブジェクトファイルのヘッダから、オブジェクトファイルのプログラム実行コード部であるテキストセグメントの仮想ページへのマッピング情報を仮想マシン起動前に生成しておく。

20

【 0 0 4 1 】

そして、仮想マシン起動時に、オブジェクトファイルとその仮想ページとを対応づけて管理する。

【 0 0 4 2 】

さらに、メモリマッピング時に、オブジェクトファイルと、その仮想ページがマッピングされたマシンメモリのページとを対応づけて管理する。

【 0 0 4 3 】

ここで、他の仮想マシンVM2に同一のオブジェクトファイルがある場合に、そのオブジェクトファイルのテキストセグメントのメモリマッピング時に、同一オブジェクトファイルのテキストセグメントがマッピングされている同一内容のページを探し出す。そして、仮想マシンVM2の仮想ページに探し出したマシンページをマッピングすることによって、仮想マシンVM1、VM2間でページを共有することができる。

30

【 0 0 4 4 】

以下に、本発明の実施の形態を説明する。

【 0 0 4 5 】

図2は、仮想マシンモニタ装置の構成例を示す図である。

【 0 0 4 6 】

仮想マシンモニタ装置1は、オブジェクトファイルマッピング情報生成部11、マッピング情報データベース12、仮想ページ・オブジェクト対応管理部13、ページマッピング部15、共有可能マシンページ検出部16、およびオブジェクト・マシンページ対応管理部17を備える。

40

【 0 0 4 7 】

オブジェクトファイルマッピング情報生成部11は、オブジェクトファイル4を読み込み、オブジェクトファイル4のヘッダ情報を解析して、オブジェクトファイル中のテキストセグメントがマッピングされる仮想アドレスの情報を、マッピング情報データベース12に登録する。

【 0 0 4 8 】

オブジェクトファイルマッピング情報生成部11は、仮想マシンVMのゲストOSを構

50

成するオブジェクトファイル 4 について、図 3 に示されるように、仮想マシンを一意に識別する仮想マシン ID、オブジェクトファイルのファイル名、オブジェクトファイルのバージョン、オブジェクトファイルの更新日時、オブジェクトファイルのサイズ等をリストとして保持し、管理する。

【 0 0 4 9 】

オブジェクトファイルマッピング情報生成部 1 1 は、オブジェクトファイル 4 を一意に識別するオブジェクト ID を生成する。オブジェクト ID は、図 3 に示すリストの各フィールドを組み合わせた文字列、または、その文字列やファイル内容から生成したハッシュ値などを用いてもよい。本実施例では説明を簡単にするため、“ <ファイル名> - <バージョン> の文字列 ” を、オブジェクト ID として用いる。

10

【 0 0 5 0 】

オブジェクトファイルマッピング情報生成部 1 1 は、オブジェクトファイル 4 のヘッダ情報を解析して得た仮想ページマッピング情報を、仮想マシン ID とオブジェクト ID と対応づけて、マッピング情報データベース 1 2 に登録する。

【 0 0 5 1 】

図 4 に、マッピング情報データベース 1 2 に登録されるマッピング情報のデータ構成例を示す。マッピング情報は、仮想マシン ID、オブジェクト ID、オブジェクトファイル 4 のテキストセグメントがマッピングされる仮想メモリのページを示す仮想ページマッピング情報で構成される。仮想ページマッピング情報は、マッピングされる仮想アドレスの先頭アドレスおよびマッピングされるサイズの組である。

20

【 0 0 5 2 】

また、オブジェクトファイルマッピング情報生成部 1 1 は、仮想マシン VM 上で動作するゲスト OS を構成するオブジェクトファイル 4 の変更が検出された場合に、変更が検出された仮想マシン VM のマッピング情報を再生成し、マッピング情報データベース 1 2 を更新する。

【 0 0 5 3 】

また、オブジェクトファイルマッピング情報生成部 1 1 は、仮想マシン VM のゲスト OS を構成するオブジェクトファイルのテキストセグメントがマッピングされる仮想ページの候補が複数ある場合に、複数の仮想ページそれぞれに対する仮想ページマッピング情報を生成して、マッピング情報データベース 1 2 に保存する。

30

【 0 0 5 4 】

仮想ページ・オブジェクト対応管理部 1 3 は、仮想マシン VM の起動時に、マッピング情報データベース 1 2 から、該当する仮想マシン VM のマッピング情報を読み出し、仮想ページ・オブジェクト対応テーブル 3 1 に登録する（テーブル管理処理）。

【 0 0 5 5 】

図 5 に、仮想ページ・オブジェクト対応テーブル 3 1 のデータ構成例を示す。仮想ページ・オブジェクト対応テーブル 3 1 は、オブジェクト ID と、仮想マシン ID と、仮想ページを示す仮想ページ番号と、仮想ページ番号に対応するマシンページを示すマシンページ番号と、同一オブジェクト ID の存在の有無を表す同一オブジェクトの存在フラグとをフィールドとして持つ。

40

【 0 0 5 6 】

仮想ページ・オブジェクト対応管理部 1 3 は、新規に仮想ページ・オブジェクト対応テーブル 3 1 が追加される際に、既に存在する仮想ページ・オブジェクト対応テーブル 3 1 に同一オブジェクト ID が存在する場合には、該当する同一オブジェクト ID の「存在フラグ」に“有効 (t r u e) ” を設定する。

【 0 0 5 7 】

また、仮想ページに対応するマシンページが割り当て、または、変更された場合には、仮想ページ・オブジェクト対応テーブル 3 1 のマシンページ番号を更新する（マシンページ番号更新処理）。

【 0 0 5 8 】

50

ページマッピング部 15 と共有可能マシンページ検出部 16 とは、メモリマッピング部 2 の内部に、または、メモリマッピング部 2 と協働可能なように構成される。メモリマッピング部 2 は、仮想ページ・マシンページ変換テーブル 32 を用いてメモリマッピングを行う。

【0059】

仮想マシン VM の物理メモリ vMEM は、仮想マシンモニタ装置 1 が用意した仮想的なものであり、実際の物理メモリにアクセスするために、仮想マシンモニタ装置 1 は、仮想マシン VM の物理アドレスを実際の物理アドレスに変換する必要がある。

【0060】

メモリマッピング部 2 では、仮想ページ・マシンページ変換テーブル 32 を用いて、仮想マシン VM のゲスト OS から仮想ページ・マシンページ変換テーブル 32 を参照することによって、ゲスト OS の仮想メモリのアドレスを、直接、仮想マシンモニタ装置 1 の物理メモリ (マシンメモリ) にアドレス変換させている。

10

【0061】

「デマンドページング」によれば、仮想ページに対応するマシンページは、常に割り当てられているわけではなく、初期状態において、仮想ページ・マシンページ変換テーブル 32 の物理ページのアドレスは“空”である。

【0062】

メモリマッピング部 2 は、最初の仮想ページのアクセスの際には、仮想ページ・マシンページ変換テーブル 32 において対応する物理ページのアドレスが“空”であるため、ページフォルトが発生する。このページフォルトが検知されて、アクセスに物理ページを割り当て、仮想ページ・マシンページ変換テーブル 32 を更新し、物理ページへのアクセスが可能となる。

20

【0063】

ページマッピング部 15 は、仮想ページのアクセスの際に、仮想ページ・マシンページ変換テーブル 32 によりマシンページの割り当てを行う (マシンページ割り当て処理)。

【0064】

また、ページマッピング部 15 は、仮想ページ・マシンページ変換テーブル 31 の更新の要求に応じて、仮想ページ・オブジェクト対応管理部 13 に働きかけて、仮想ページ・オブジェクト対応テーブル 33 のマシンページのフィールドを更新する (テーブル管理処理)。

30

【0065】

また、ページマッピング部 15 は、マシンページをコピーオンライトで管理し、仮想マシン VM 間で共有されるマシンページに対する仮想マシン VM からの書き込み要求があった場合に、コピーオンライトによって新しいマシンページを割り当てる。

【0066】

共有可能マシンページ検出部 16 は、仮想ページ・オブジェクト対応テーブル 31 を参照し、アクセス先の仮想ページに該当するオブジェクトが、既に他の仮想マシン VM の仮想ページ上に存在しているか否かを判定する (オブジェクト判定処理)。

【0067】

40

さらに、オブジェクト・マシンページ対応テーブル 33 を参照して、共有可能なマシンページがあるかを検出し、共有可能なマシンページがあれば、ページマッピング部 15 に働きかけて、仮想マシン VM 間で同一オブジェクトがマッピングされたマシンページを参照するように仮想ページ・マシンページ変換テーブル 32 を変更させる (共有可能ページ判定処理)。

【0068】

オブジェクト・マシンページ対応管理部 17 は、オブジェクトとマシンページとを対応づけたオブジェクト・マシンページ対応テーブル 33 を持ち、オブジェクト・マシンページ対応テーブル 33 を管理する (テーブル管理処理)。

【0069】

50

オブジェクト・マシンページ対応テーブル 33 は、図 6 に示されるように、オブジェクト ID、そのオブジェクトファイルのテキストセグメントがマッピングされているマシンページ番号、マシンページを参照している仮想マシン数からなる。

【 0070 】

また、オブジェクト・マシンページ対応管理部 17 は、マシンページを共有する仮想マシン数に変動があれば、オブジェクト・マシンページ対応テーブル 33 の参照仮想マシン数を更新する（参照仮想マシン数更新処理）。

【 0071 】

図 7 は、仮想マシンモニタ装置 1 の処理を概説する図である。

【 0072 】

仮想マシンモニタ装置 1 のオブジェクトファイルマッピング情報生成部 11 は、仮想マシン VM の起動前に、仮想マシン VM の仮想外部記憶装置 vHDD に保存されているオブジェクトファイル 4 のファイルヘッダおよびプログラムヘッダを解析し、オブジェクトファイル 4 のテキストセグメントがゲスト OS の仮想メモリへマッピングされる情報（仮想ページマッピング情報を生成し、マッピング情報データベース 12 へ保存しておく。

【 0073 】

その後、仮想マシン VM が起動時に、仮想ページ・オブジェクト対応管理部 13 は、マッピング情報データベース 12 を参照して、実行されるオブジェクトファイル 4 とその仮想ページとの対応づけ情報を、仮想ページ・オブジェクト対応テーブル 31 に登録する。

【 0074 】

ゲスト OS の仮想ページは、仮想メモリ管理ユニット vMMU、メモリマッピング部 2、メモリ管理ユニット MMU を介して、仮想マシンモニタ装置 1 のマシンメモリにマッピングされ、仮想マシンのゲスト OS は、仮想ページのアクセスで、仮想マシンモニタ装置 1 の該当するマシンページを直接参照する。

【 0075 】

仮想ページ・オブジェクト対応管理部 13 は、メモリマッピング時に仮想ページのオブジェクトとそのマッピングされたマシンページとの対応づけを管理し、他の仮想マシン VM に同一のオブジェクトファイルがある場合に、そのメモリマッピング時に、同一のテキストセグメントがマッピングされている同一内容のページが仮想マシンモニタ装置 1 内のマシンメモリにあるかを判定する。マシンメモリ内に同一内容のページがあれば、共有可能なマシンページを検出したとして、そのマシンページをマッピングし、仮想マシン間でページ共有を実現する。

【 0076 】

図 8 に、仮想マシン VM の仮想メモリのアクセス時の処理フローを示す。

【 0077 】

仮想マシン VM のゲスト OS から仮想メモリ vMEM へのアクセスがあった場合に、メモリマッピング部 2 は、仮想ページ・マシンページ変換テーブル 32 を参照して、仮想マシン VM のアクセス先の仮想ページに対応する仮想マシンモニタ装置 1 のマシンページがあるかを判定する（ステップ S1）。仮想マシン VM のアクセス先の仮想ページに対応する仮想マシンモニタ装置 1 のマシンページがある場合は（ステップ S1 の YES）、そのままマシンメモリへアクセスして、処理が終了する。

【 0078 】

一方、仮想マシン VM のアクセス先の仮想ページに対応するマシンページがない場合、例えば、マシンページが未割り当ての場合、初回アクセス時の場合等は（ステップ S1 の NO）、共有可能マシンページ検出部 16 が、仮想ページ・オブジェクト対応テーブル 31 を参照し、アクセス先の仮想ページに該当するオブジェクトが、既に他の仮想マシン VM の仮想ページ上に存在しているかを判定する（ステップ S2）。

【 0079 】

他の仮想マシン VM の仮想ページ上に同一のオブジェクトが存在する場合は（ステップ S2 の YES）、共有可能マシンページ検出部 16 は、オブジェクト・マシンページ対応

10

20

30

40

50

テーブル 33 を参照し、該当するオブジェクトがマシンページに存在するか否かを確認する（ステップ S3）。

【0080】

該当するオブジェクトがマシンページに存在する場合は（ステップ S3 の YES）、メモリ共有が可能であると判定され、ページマッピング部 15 は、アクセス先の仮想ページを検出したマシンページに変換するように、仮想ページ・マシンページ変換テーブル 32 を更新し、検出したマシンページを割り当てるマッピング（共有マッピング）を行う（ステップ S4）。

【0081】

さらに、ページマッピング部 15 がオブジェクト・マシンページ対応管理部 17 に働きかけ、オブジェクト・マシンページ対応管理部 17 が、オブジェクト・マシンページ対応テーブル 33 の参照仮想マシン数を更新する（ステップ S5）。

10

【0082】

また、ページマッピング部 15 が仮想ページ・オブジェクト対応管理部 13 に働きかけ、仮想ページ・オブジェクト対応管理部 13 が、仮想ページ・オブジェクト対応テーブル 31 の該当する仮想ページの割り当てられたマシンページ番号を更新する（ステップ S6）。

【0083】

一方、該当するオブジェクトがマシンページに存在しない場合は（ステップ S3 の NO）、ページマッピング部 15 が、マシンページの割り当てを行う（ステップ S7）。さらに、オブジェクト・マシンページ対応管理部 17 が、オブジェクト・マシンページ対応テーブル 33 に登録する（ステップ S8）。

20

【0084】

また、ステップ S2 の処理で、他の仮想マシン VM の仮想ページ上に同一のオブジェクトが存在しない場合は（ステップ S2 の NO）、コンピュータシステムにおいて、該当オブジェクトがマシンページに存在しないため、ページマッピング部 15 が、マシンページの割り当てを行い（ステップ S7）、オブジェクト・マシンページ対応管理部 17 が、オブジェクト・マシンページ対応テーブル 33 へオブジェクトと割り当てられたマシンページとを登録する（ステップ S8）。

【0085】

以上、本発明をその実施の形態により説明したが、本発明はその主旨の範囲において種々の変形が可能であることは当然である。

30

【図面の簡単な説明】

【0086】

【図 1】本発明にかかる仮想マシンモニタ装置を備えたコンピュータのシステム構成例を示す図である。

【図 2】仮想マシンモニタ装置の構成例を示す図である。

【図 3】オブジェクトファイルから抽出される情報例を示す図である。

【図 4】マッピング情報データベースに登録される情報のデータ構成例を示す図である。

【図 5】仮想ページ・オブジェクト対応テーブルのデータ構成例を示す図である。

40

【図 6】オブジェクト・マシンページ対応テーブルのデータ構成例を示す図である。

【図 7】仮想マシンモニタ装置の処理を概説する図である。

【図 8】仮想マシン VM の仮想メモリのアクセス時の仮想マシンモニタ装置の処理フローを示す図である。

【符号の説明】

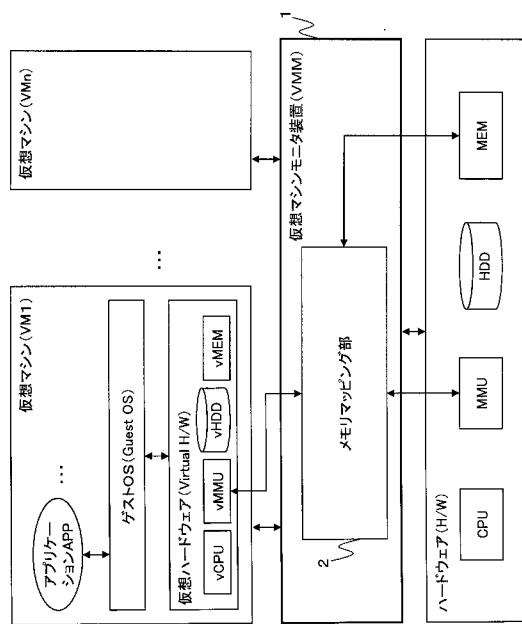
【0087】

- 1 仮想マシンモニタ装置（VMM）
- 2 メモリマッピング部
 - 11 オブジェクトファイルマッピング情報生成部
 - 12 マッピング情報データベース

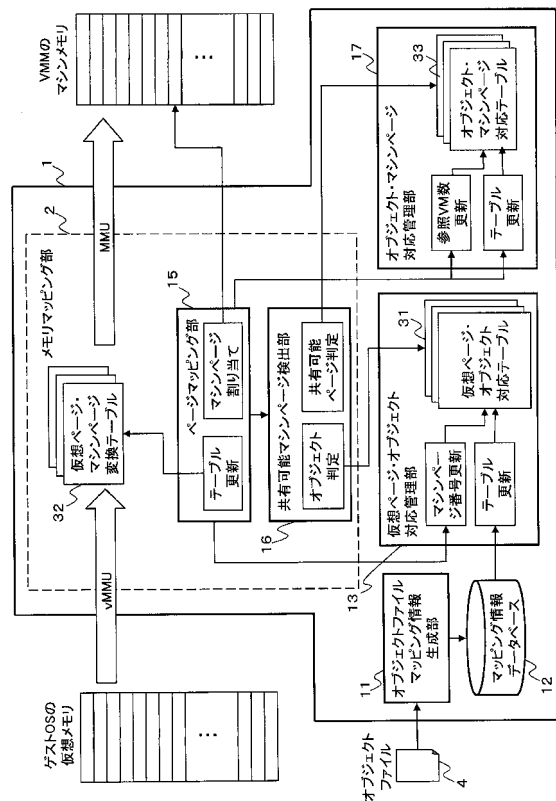
50

- 1 3 仮想ページ・オブジェクト対応テーブル
- 1 5 ページマッピング部
- 1 6 共有可能マシンページ検出部
- 1 7 オブジェクト・マシンページ対応管理部
- 3 1 仮想ページ・オブジェクト対応テーブル
- 3 2 仮想ページ・マシンページ変換テーブル
- 3 3 オブジェクト・マシンページ対応テーブル
- VM 仮想マシン
- Virtual H/W 仮想マシンのハードウェア資源
- vMMU 仮想マシンのメモリ管理ユニット
- vHDD 仮想マシンの外部記憶装置
- vMEM 仮想マシンのメモリ
- Guest OS ゲストOS (オペレーティングシステム)
- H/W ハードウェア資源
- MMU メモリ管理ユニット
- HDD 外部記憶装置
- MEM メモリ

【図1】



【図2】



フロントページの続き

審査官 渡部 博樹

(56)参考文献 特開2008-225620(JP,A)
特開平09-319643(JP,A)
特開平05-012125(JP,A)
特開2008-097173(JP,A)

(58)調査した分野(Int.Cl., DB名)
G06F12/00 - G06F12/06
G06F12/08 - G06F12/12
G06F13/16 - G06F13/18
G06F 9/46