

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6197871号  
(P6197871)

(45) 発行日 平成29年9月20日 (2017.9.20)

(24) 登録日 平成29年9月1日 (2017.9.1)

(51) Int. Cl.

F I

G 0 6 F 9/50 (2006.01)

G 0 6 F 9/46 4 6 2 Z

G 0 6 F 9/46 (2006.01)

G 0 6 F 9/46 3 5 0

請求項の数 8 (全 24 頁)

(21) 出願番号 特願2015-524935 (P2015-524935)  
 (86) (22) 出願日 平成25年7月2日 (2013.7.2)  
 (86) 国際出願番号 PCT/JP2013/068129  
 (87) 国際公開番号 W02015/001615  
 (87) 国際公開日 平成27年1月8日 (2015.1.8)  
 審査請求日 平成28年2月26日 (2016.2.26)

(73) 特許権者 000005223  
 富士通株式会社  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号  
 (74) 代理人 100092978  
 弁理士 真田 有  
 (74) 代理人 100112678  
 弁理士 山本 雅久  
 (72) 発明者 土肥 実久  
 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  
 審査官 田中 幸雄

最終頁に続く

(54) 【発明の名称】 仮想マシン管理方法、仮想マシン管理装置、および仮想マシン管理プログラム

(57) 【特許請求の範囲】

【請求項 1】

第1の命令セットを用いて仮想マシンを制御する第1の情報処理装置へ、第2の命令セットを用いて仮想マシンを制御する第2の情報処理装置からの仮想マシンの移動を許容し、

前記第2の情報処理装置から前記第1の情報処理装置へ移動される仮想マシンに対して前記第1の命令セットに関する情報を通知し、

前記通知を受けた仮想マシンは、利用する命令セットを前記第2の命令セットから前記第1の命令セットに切り替え、

前記第1の情報処理装置は、移動された前記仮想マシンに対して、前記第1の命令セットを用いた制御を実行する、  
 ことを特徴とする、仮想マシン管理方法。

【請求項 2】

前記通知は、前記第1の命令セットの水準及び前記第2の命令セットの水準に基づいて、仮想マシンの移動前又は移動後に、前記移動される仮想マシンに対する、前記移動される仮想マシンの制御に用いられる命令セットを認識させる指示によって行なわれる、  
 ことを特徴とする、請求項1記載の仮想マシン管理方法。

【請求項 3】

前記通知において、前記第2の情報処理装置は、

前記移動される仮想マシンの制御に用いられる命令セットに関する情報を、前記第2の

10

20

命令セットに関する情報から前記第 1 の命令セットに関する前記情報に変更し、

前記移動される仮想マシンに対して、前記変更した命令セットに関する情報を認識させる前記指示を行なう、

ことを特徴とする、請求項 2 記載の仮想マシン管理方法。

【請求項 4】

前記通知において、前記第 1 の情報処理装置は、

前記移動された仮想マシンに対して、前記移動された仮想マシンの制御に用いられる命令セットに関する情報を認識させる前記指示を行なう、

ことを特徴とする、請求項 2 記載の仮想マシン管理方法。

【請求項 5】

前記第 1 の情報処理装置は、

前記移動された仮想マシンにより、前記第 2 の命令セットに含まれる命令であって前記第 1 の命令セットに含まれない命令が前記第 1 の情報処理装置において実行された場合、前記実行された命令のエミュレーションを行ない、

前記移動された仮想マシンに対して前記第 1 の命令セットに関する情報を再度通知すること、  
ことを特徴とする、請求項 1 ~ 4 のいずれか 1 項記載の仮想マシン管理方法。

【請求項 6】

前記第 1 の情報処理装置は、前記実行された命令について以前にエミュレーションを行なっている場合、前記実行された命令を、前記第 1 の情報処理装置が実行可能な命令に変換する、

ことを特徴とする、請求項 5 記載の仮想マシン管理方法。

【請求項 7】

第 1 の命令セットを用いて仮想マシンを制御する第 1 の仮想マシン管理装置であって、

第 2 の命令セットを用いて仮想マシンを制御する第 2 の仮想マシン管理装置からの仮想マシンの移動を許容する制御部をそなえ、

前記制御部は、前記第 2 の仮想マシン管理装置から前記第 1 の仮想マシン管理装置へ移動される仮想マシンに対して前記第 1 の命令セットに関する情報を通知する通知部をそなえ、

前記通知を受けた仮想マシンは、利用する命令セットを前記第 2 の命令セットから前記第 1 の命令セットに切り替え、

前記第 1 の仮想マシン管理装置は、移動された前記仮想マシンに対して、前記第 1 の命令セットを用いた制御を実行する処理部をさらにそなえる、

ことを特徴とする、仮想マシン管理装置。

【請求項 8】

第 1 の命令セットを用いて仮想マシンを制御する第 1 のコンピュータに、

第 2 の命令セットを用いて仮想マシンを制御する第 2 のコンピュータからの仮想マシンの移動を許容し、

前記第 2 のコンピュータから前記第 1 のコンピュータへ移動される仮想マシンに対して前記第 1 の命令セットに関する情報を通知する、

処理を前記第 1 のコンピュータに実行させ、

前記通知を受けた仮想マシンは、利用する命令セットを前記第 2 の命令セットから前記第 1 の命令セットに切り替え、

移動された前記仮想マシンに対して、前記第 1 の命令セットを用いた制御を実行する、処理を前記第 1 のコンピュータに実行させることを特徴とする、仮想マシン管理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本件は、仮想マシン管理方法、仮想マシン管理装置、および仮想マシン管理プログラム

10

20

30

40

50

に関する。

【背景技術】

【0002】

複数の仮想マシン (Virtual Machine ; VM) を制御する複数の物理サーバ (物理マシン, 情報処理装置) をそなえる情報処理システムでは、物理サーバの状態に応じて、物理サーバ間で仮想マシンを移動させるマイグレーションが行なわれることがある。なお、マイグレーションとしては、仮想マシンの動作を継続させたまま物理サーバ間で仮想マシンを移動するライブマイグレーションが挙げられる。

【0003】

マイグレーションを管理するサーバ又は管理者等 (以下、単にサーバ等という) は、物理サーバが異なる仕様を持つ装置である場合、仮想マシンの移動先を決定する際に、移動先が仮想マシンの動作に不都合のある物理サーバであるか否かを確認することがある。

例えば、サーバ等は、仮想マシンの移動前後の物理サーバがそなえる Central Processing Unit (CPU) の違いによる、命令 (拡張 CPU 命令セット) のサポート有無を確認する。サーバ等は、確認の結果、サポートする拡張 CPU 命令セットが異なる物理サーバ (CPU) 間では、仮想マシンのマイグレーションを行なわない。

【0004】

なお、拡張 CPU 命令セットとは、CPU が共通して持つ命令とは別に、所定の処理を実行させるために CPU に実装される命令群である。以下、拡張 CPU 命令セットを、単に拡張命令セット又は命令セットという場合がある。

CPU がサポートする拡張命令セットは、CPU の仕様等によってレベル (水準) が異なる。なお、古い世代の CPU は、新しい世代の CPU よりも拡張命令セットに含まれる命令が少ない、つまり低レベルであることが多い。

【0005】

ところで、サーバ等は、以下のように、サポートする拡張命令セットが異なる物理サーバ間で、強制的にマイグレーションを行なうこともできる。

サポートする拡張命令セットのレベルが低い物理サーバ (CPU) から高い物理サーバ (CPU) へ仮想マシンが移動される場合、仮想マシンは、移動前の CPU がサポートしない命令は移動後の CPU がサポートしていても利用しない。これは、仮想マシンが、起動したときの物理サーバ (CPU) でサポートされる拡張命令セットに従って動作するためである。

【0006】

一方、サポートする拡張命令セットのレベルが高い物理サーバ (CPU) から低い物理サーバ (CPU) へ仮想マシンが移動される場合、仮想マシンは、今まで使用していた命令を利用できないことがある。例えば、仮想マシンが、移動前に存在していたが移動後に存在しなくなった拡張命令セットを実行すると、“Illegal Instruction” が発生し、その後、仮想マシンは停止してしまう。

【0007】

このため、サーバ等は、マイグレーションによる移動先の物理サーバを決定する際には、移動先の CPU を確認し、移動可能か否かを判断する。そして、サーバ等は、少なくとも、移動前よりもサポートする拡張命令セットのレベルが低い物理サーバを移動先から除外する。

なお、仮想マシンが利用する拡張命令セットを、仮想マシンが実行され得る複数の物理サーバでサポートされる拡張命令セットのうちの、最低レベルの拡張命令セットに制限するハードウェアアシスト機構が知られている。ハードウェアアシスト機構によれば、仮想マシンはミニマムセットの拡張命令セットを利用するため、仮想マシンは、制限された結果のレベルと同レベルの命令セットをサポートする VM ホスト間において移動が可能になる。従って、VM 配置管理サーバ等は、移動前よりもサポートする拡張命令セットのレベルが低い物理サーバを移動先から除外せずに済む。なお、ハードウェアアシスト機構としては、例えば Intel (登録商標) の “FlexMigration” が挙げられる。

10

20

30

40

50

## 【 0 0 0 8 】

また、関連する技術として、マイグレーション前後でCPUが実装する命令セット間に共通な命令は、移動後のCPUに実行させ、移動後のCPUに存在しない命令は、命令エミュレータにエミュレーションさせる技術も知られている（例えば、特許文献1参照）。

さらに、関連する他の技術として、共有メモリに接続された第1及び第2スイッチングハイパーバイザ間で仮想マシンの実行を切り替える技術も知られている（例えば、特許文献2参照）。この技術では、仮想マシンは、仮想マシンに用いられるハードウェアアーキテクチャが実行の切り替え前と異なる場合、命令セットをシミュレートする。

## 【 0 0 0 9 】

また、関連する他の技術として、CPUが、ゲストOperating System（OS）が実行できない所定の命令を検出した場合に、仮想マシンモニタが、所定の命令のエミュレーションコードを生成して、仮想マシンへ追加して記憶させる技術も知られている（例えば、特許文献3参照）。

## 【 先行技術文献 】

## 【 特許文献 】

## 【 0 0 1 0 】

【 特許文献 1 】 特開 2 0 1 0 - 2 7 2 0 5 5 号 公 報

【 特許文献 2 】 特開 2 0 1 1 - 1 0 0 4 3 1 号 公 報

【 特許文献 3 】 特開 2 0 1 1 - 1 2 3 6 3 5 号 公 報

## 【 発明の概要 】

## 【 発明が解決しようとする課題 】

## 【 0 0 1 1 】

上述のように、サーバ等は、サポートする拡張命令セットに応じてマイグレーションの移動先の物理サーバを限定するため、システムの性能が制限されるという課題がある。例えば、サーバ等がマイグレーションを繰り返すと、サポートする拡張命令セットのレベルが低い物理サーバへ移動できる仮想マシンが順次減り、サポートする拡張命令セットのレベルが高い物理サーバへ、仮想マシンが集まってしまう場合がある。これにより、システムのリソースの利用効率が低下し、システムの性能が制限される。

## 【 0 0 1 2 】

また、サポートする拡張命令セットのレベルが低い物理サーバから高い物理サーバへ仮想マシンが移動される場合、移動先に存在しない拡張命令セットは、移動先の物理サーバで利用されない。同様に、上記ハードウェアアシスト機構が用いられる場合、制限された拡張命令セットは、移動先の物理サーバで利用されない。つまり、これらの場合、移動先の物理サーバは、移動してきた仮想マシンの実行において、本来持っている機能を発揮することができない場合があるという課題がある。

## 【 0 0 1 3 】

このように、物理サーバ間での仮想マシンの移動において、仮想マシンによるシステムの利用効率が低下する場合があるという課題がある。

なお、上述した関連する技術では、上記課題については考慮されていない。

1つの側面では、本発明は、情報処理装置間での仮想マシンの移動を許容する情報処理システムにおいて、仮想マシンによる情報処理システムの利用効率を向上させることを目的とする。

## 【 0 0 1 4 】

なお、前記目的に限らず、後述する発明を実施するための形態に示す各構成により導かれる作用効果であって、従来の技術によっては得られない作用効果を奏することも本発明の他の目的の1つとして位置付けることができる。

## 【 課題を解決するための手段 】

## 【 0 0 1 5 】

本件の仮想マシン管理方法は、第1の命令セットを用いて仮想マシンを制御する第1の情報処理装置へ、第2の命令セットを用いて仮想マシンを制御する第2の情報処理装置か

10

20

30

40

50

らの仮想マシンの移動を許容し、前記第2の情報処理装置から前記第1の情報処理装置へ移動される仮想マシンに対して前記第1の命令セットに関する情報を通知し、前記通知を受けた仮想マシンは、利用する命令セットを前記第2の命令セットから前記第1の命令セットに切り替え、前記第1の情報処理装置は、移動された前記仮想マシンに対して、前記第1の命令セットを用いた制御を実行する。

【発明の効果】

【0016】

一実施形態によれば、情報処理装置間での仮想マシンの移動を許容する情報処理システムにおいて、仮想マシンによる情報処理システムの利用効率を向上させることができる。

【図面の簡単な説明】

10

【0017】

【図1】一実施形態に係る情報処理システムの構成例を示す図である。

【図2】図1に示す物理サーバのハードウェア構成例を示す図である。

【図3】図1に示す物理サーバの構成例を示す図である。

【図4】図3に示す物理サーバによるマイグレーション処理の一例を説明するフローチャートである。

【図5】図3に示す物理サーバによるマイグレーション処理の一例を説明する図である。

【図6】図3に示す物理サーバによる例外発生命令への処理の一例を説明するフローチャートである。

【図7】図3に示す物理サーバによる例外発生命令への処理の一例を説明する図である。

20

【図8】図1に示す物理サーバの構成の第1変形例を示す図である。

【図9】図8に示す物理サーバによるマイグレーション処理の一例を説明する図である。

【図10】図1に示す物理サーバの構成の第2変形例を示す図である。

【図11】図10に示す物理サーバによる例外発生命令への処理の一例を説明する図である。

【発明を実施するための形態】

【0018】

以下、図面を参照して実施の形態を説明する。

〔1〕一実施形態

〔1-1〕情報処理システムの構成

30

以下、図1を参照して、一実施形態の一例としての情報処理システム1の構成について説明する。

【0019】

図1は、一実施形態に係る情報処理システム1の構成例を示す図である。

図1に示すように、一実施形態に係る情報処理システム1は、複数（例えば2つ）の物理サーバ2-1及び2-2（以下、物理サーバ2-1及び2-2を区別しない場合には、単に物理サーバ2という）、及び、VM配置管理サーバ3をそなえる。

情報処理システム1は、複数の物理サーバ2間で、仮想マシンのマイグレーション（例えばライブマイグレーション）を実行する。

【0020】

40

物理サーバ（仮想マシン管理装置）2は、仮想マシンの管理を行なうサーバであり、仮想マシンは、物理サーバ2のCPUやメモリ等のリソースを用いて動作する。物理サーバ2としては、Personal Computer（PC）やサーバ等のコンピュータ（情報処理装置）が挙げられる。

物理サーバ2-1は、Hypervisor（HV）4-1及び複数（例えば3つ）のVM5-1～5-3をそなえる。また、物理サーバ2-2は、HV4-2及び複数（例えば3つ）のVM5-4～5-6をそなえる。以下、HV4-1及び4-2を区別しない場合には、単にHV4といい、VM5-1～5-6を区別しない場合には、単にVM5という。

【0021】

HV（ハイパーバイザ、制御部）4は、VM5の制御を行なうVMホストとして動作す

50

るソフトウェアである。

また、HV 4は、拡張CPU命令セット（拡張命令セット，命令セット）を用いてVM 5を制御する他の物理サーバ2との間で、VM 5の移動（移送）を許容する。本実施形態においては、HV 4は、他の物理サーバ2のCPUが用いる拡張命令セットのレベル（命令レベル，利用レベル）が、自物理サーバ2のCPUが用いる拡張命令セットのレベルと異なる場合にも、VM 5の移動を許容することができる。

#### 【0022】

ここで、拡張命令セットとしては、Intel x86系CPUの場合、マルチメディア系命令の一例であるStreaming SIMD Extensions ( SSE ) が挙げられる。SSEは、SSE ~ SSE 4等の順に機能が強化されている。

10

また、Intel x86系CPUの拡張命令セットとして、暗号化系命令であるAdvanced Encryption Standard Instruction Set ( AES - NI ) 等も挙げられる。

#### 【0023】

以下の説明で用いる拡張命令セットのレベルとは、CPUが利用可能なこれらの拡張命令セットの水準をいう。例えば、拡張命令セットがSSEである場合、拡張命令セットのレベルとは、SSE ~ SSE 4等の水準をいう。

VM（仮想マシン）5は、HV 4による制御の下、OS（ゲストOS）の実行等の種々の処理を行なうソフトウェアである。

#### 【0024】

HV 4及びVM 5の詳細については、後述する。

20

VM配置管理サーバ3は、物理サーバ2間でのマイグレーションを管理するサーバである。VM配置管理サーバ3は、複数の物理サーバ2の情報と、VM 5の情報とを対応付けて管理し、情報処理システム1の管理者等からの指示に応じて、VM 5のマイグレーションを行なう。VM配置管理サーバ3としては、PCやサーバ等の情報処理装置が挙げられる。

#### 【0025】

また、VM配置管理サーバ3は、各物理サーバ2のCPUがサポートする拡張命令セットに関する情報（少なくとも、レベル）を把握している。例えば、VM配置管理サーバ3は、情報処理システム1が起動したときや構成が変更されたとき等に、各物理サーバ2の拡張命令セットに関する情報を、各物理サーバ2又は管理者等から取得する。なお、拡張命令セットに関する情報とは、拡張命令セットのレベルのほか、拡張命令セットを特定する情報であってもよい。

30

#### 【0026】

なお、VM配置管理サーバ3は、マイグレーションを行なう際に、VM 5の移動元の物理サーバ2（HV 4）へ、移動先の物理サーバ2で利用可能な拡張命令セットに関する情報を通知することが好ましい。また、VM配置管理サーバ3は、マイグレーションを行なう際に、VM 5の移動先の物理サーバ2（HV 4）へ、移動されるVM 5が移動元の物理サーバ2で利用する拡張命令セットに関する情報を通知することが好ましい。

#### 【0027】

〔1-2〕物理サーバのハードウェア構成

40

次に、図2を参照して、物理サーバ2のハードウェア構成について説明する。図2は、図1に示す物理サーバ2のハードウェア構成例を示す図である。

物理サーバ2は、図2に示すように、CPU 20a、メモリ20b、記憶部20c、インタフェース部20d、入出力部20e、記録媒体20f、及び読取部20gをそなえる。

#### 【0028】

CPU 20aは、メモリ20b、記憶部20c、インタフェース部20d、入出力部20e、記録媒体20f、及び読取部20gと接続され、種々の制御や演算を行なう演算処理装置（プロセッサ，処理部）である。CPU 20aは、メモリ20b、記憶部20c、記録媒体20f、読取部20gに接続又は挿入された記録媒体20h、又は図示しないRe

50

ad Only Memory ( R O M ) 等に格納されたプログラムを実行することにより、物理サーバ 2 における種々の機能 ( 例えば H V 4 及び複数の V M 5 ) を実現する。なお、C P U 2 0 a に限らず、プロセッサとしては、Micro Processing Unit ( M P U ) 等の電子回路が用いられてもよい。

【 0 0 2 9 】

メモリ 2 0 b は、種々のデータやプログラムを格納する記憶装置である。C P U 2 0 a は、プログラムを実行する際に、メモリ 2 0 b にデータやプログラムを格納し展開する。なお、メモリ 2 0 b としては、例えば Random Access Memory ( R A M ) 等の揮発性メモリが挙げられる。

記憶部 2 0 c は、例えば Hard Disk Drive ( H D D ) 等の磁気ディスク装置、Solid State Drive ( S S D ) 等の半導体ドライブ装置、又はフラッシュメモリ等の不揮発性メモリ等の、種々のデータやプログラム等を格納する 1 以上のハードウェアである。記憶部 2 0 c が有する記憶領域は、物理サーバ 2、H V 4、又は V M 5 等により用いられる。

【 0 0 3 0 】

インタフェース部 2 0 d は、図示しないルータ及びネットワークを介して接続されるホスト又はクライアント、V M 配置管理サーバ 3、又は、他の物理サーバ 2 との間の接続及び通信の制御を行なうコントローラである。インタフェース部 2 0 d としては、例えば、Local Area Network ( L A N )、ファイバチャネル ( Fibre Channel ; F C ) 等に準拠したインタフェースカードが挙げられる。

【 0 0 3 1 】

入出力部 2 0 e は、例えばマウスやキーボード等の入力装置及びディスプレイやプリンタ等の出力装置の少なくとも一方を含んでよい。例えば入出力部 2 0 e は、物理サーバ 2 ( 情報処理システム 1 ) の管理者や V M 5 のユーザ等による種々の作業に用いられる。

記録媒体 2 0 f は、フラッシュメモリや R O M 等の記憶装置であり、種々のデータやプログラムを記録する。読取部 2 0 g は、光ディスクや Universal Serial Bus ( U S B ) メモリ等のコンピュータ読取可能な記録媒体 2 0 h に記録されたデータやプログラムを読み出す装置である。

【 0 0 3 2 】

記録媒体 2 0 f、2 0 h の少なくとも一方には、一実施形態に係る物理サーバ 2 ( H V 4 及び V M 5 を含む ) の機能を実現する仮想マシン管理プログラムが格納されてもよい。すなわち、C P U 2 0 a は、記録媒体 2 0 f、又は読取部 2 0 g を介して記録媒体 2 0 h から出力された仮想マシン管理プログラムを、メモリ 2 0 b 等の記憶装置に展開して実行することにより、物理サーバ 2 の機能を実現する。

【 0 0 3 3 】

なお、上述した各ハードウェアは、互いにバスを介して通信可能に接続される。例えば、C P U 2 0 a、メモリ 2 0 b、及びインタフェース部 2 0 d は、システムバスに接続される。また、例えば、記憶部 2 0 c、入出力部 2 0 e、記録媒体 2 0 f、及び読取部 2 0 g は、Input/Output ( I / O ) インタフェース等を介してシステムバスに接続される。なお、記憶部 2 0 c は、ストレージバス ( ケーブル ) で、Disk Interface ( D I ) 等の I / O インタフェースに接続される。ストレージバスとしては、Small Computer System Interface ( S C S I )、Serial Attached SCSI ( S A S )、ファイバチャネル、Serial Advanced Technology Attachment ( S A T A ) 等に準拠したバスが挙げられる。

【 0 0 3 4 】

なお、物理サーバ 2 の上述したハードウェア構成は例示である。従って、物理サーバ 2 内でのハードウェアの増減や分割等は適宜行なわれてもよい。また、物理サーバ 2 間で異なる構成をそなえてもよい。

なお、V M 配置管理サーバ 3 についても、物理サーバ 2 と同様のハードウェア構成をそなえることができる。

【 0 0 3 5 】

[ 1 - 3 ] 物理サーバの機能構成

10

20

30

40

50

次に、図 3 を参照して、物理サーバ 2 の機能構成について説明する。図 3 は、図 1 に示す物理サーバ 2 の機能構成例を示す図である。

本実施形態に係る物理サーバ 2 は、CPU 20 a の拡張命令セットのレベルが異なる VM ホスト間でマイグレーションを実施することができる。なお、以下の説明では、物理サーバ 2 - 1 及び 2 - 2 の各々がそなえる CPU 20 a の拡張命令セットのレベルは、互いに異なるものとする。

#### 【0036】

物理サーバ 2 は、ソフトウェアの機能として、HV 4 に、再認識指示部 4 1、制限レベル管理部 4 2、及び拡張 CPU 命令セットエミュレーション部 4 3 をそなえ、VM 5 に、再認識部 5 1 をそなえる。なお、これらの再認識指示部 4 1、制限レベル管理部 4 2、拡張 CPU 命令セットエミュレーション部 4 3、及び再認識部 5 1 は、CPU 20 a が仮想マシン管理プログラムをメモリ 20 b に展開して実行することにより実現される。なお、再認識部 5 1 は、VM 5 が実行する OS の一機能として実装されてもよい。

10

#### 【0037】

また、物理サーバ 2 は、ハードウェアの機能として、CPU 20 a に、命令セット制限部 20 1 をそなえる。

なお、VM 5 の再認識部 5 1 は、VM 5 - 1 ~ 5 - 6 の各々がそなえることができ、HV 4、VM 5、及び CPU 20 a の各機能は、物理サーバ 2 - 1 及び 2 - 2 の各々がそなえることができる。

#### 【0038】

20

##### 〔1-3-1〕命令セット制限部の説明

命令セット制限部 20 1 は、VM 5 ごとに拡張命令セットの利用制限をかける機能であり、上述した“FlexMigration”等のハードウェアアシスト機構により実現される。

命令セット制限部 20 1 は、HV 4 からの指示に応じて、特定の VM 5 が認識する拡張命令セットを、所定のレベルに制限することができる。例えば、HV 4 から、特定の VM 5 に利用させる拡張命令セットをレベル 2 (SSE2) にする旨の指示を受けると、命令セット制限部 20 1 は、CPU 20 a がレベル 4 (SSE4) をサポートしていても、当該 VM 5 にはレベル 2 を通知する。つまり、命令セット制限部 20 1 は、特定の VM 5 からの利用可能な拡張命令セットの問い合わせ（例えば CPUID 命令）に対して、利用可能な拡張命令セットが SSE2 である旨の応答を行なう。

30

#### 【0039】

##### 〔1-3-2〕再認識部の説明

再認識部 5 1 は、自 VM 5 がマイグレーションの対象である場合に、HV 4 から、移動先の物理サーバ 2 で利用可能な拡張命令セットに関する情報を通知されることで、アーキテクチャを再認識する。

ここで、OS やアプリケーションは、利用可能な拡張命令セットを最大限使おうとすることが多い。例えば、OS やアプリケーションは、物理サーバや VM における実行が開始されると、利用可能な拡張命令セットを判定して、判定した拡張命令セットを利用するライブラリやコンパイラ出力モード等の選択を行なう。そして、OS やアプリケーションは、物理サーバや VM における実行が停止するまで、選択したライブラリやコンパイラ出力モード等を用いて動作する。

40

#### 【0040】

OS やアプリケーションによるこれらの処理は、物理サーバ上で動作する場合及び VM 上で動作する場合で共通である。つまり、拡張命令セットの判定は OS やアプリケーションの実行が開始されたときに行なわれるため、VM で実行される OS やアプリケーションは、ライブマイグレーションにより、実行開始後に途中で拡張命令セットが変わることを想定していない。

#### 【0041】

そこで、再認識部 5 1 は、HV 4 (再認識指示部 4 1) からの再認識指示に応じて、CPU 20 a の命令セット制限部 20 1 から、自 VM 5 が HV 4 で利用可能な拡張命令セッ

50



トに関する情報を取得する。例えば、再認識部 5 1 は、CPU 2 0 a がどの拡張命令セットをサポートするかを問い合わせる CPUID 命令を発行することにより、CPU 2 0 a から利用可能な拡張命令セットに関する情報を取得する。

#### 【0042】

VM 5 (OS) は、再認識部 5 1 により、HV 4 からの再認識指示に応じて再認識を行なうと、CPU 2 0 a へ再認識後に指示する処理 (コード) の切り替えを行なう。

例えば、VM 5 は、OS が直接ハンドリングできるコード (例えばカーネルやライブラリ等) については、OS により使用コードの切り替えを行なう。また、VM 5 は、OS が直接ハンドリングできないコード (例えばユーザコード等) については、OS によりユーザプログラム等へ使用コードの切替要求を発行する。

10

#### 【0043】

以上のように、再認識部 5 1 によれば、VM 5 (OS) は、実行が開始されたとき (例えば HV 4 上での起動後) のみでなく、HV 4 から任意のタイミングで発行された再認識指示を受けたときに、アーキテクチャの再認識を行なうことができる。また、VM 5 は、利用可能な拡張命令セットを再認識することで、使用コードの切り替えを行なうことができるため、再認識した拡張命令セットを可能な限り有効に活用することができる。

#### 【0044】

##### 〔1-3-3〕再認識指示部及び制限レベル管理部の説明

再認識指示部 (通知部) 4 1 は、マイグレーションにより移動される VM 5 に対して移動先の物理サーバ 2 で利用可能な拡張命令セットに関する情報を通知する。例えば、再認識指示部 4 1 は、VM 配置管理サーバ 3 からマイグレーションの移動対象の VM 5 を通知されると、通知された VM 5 に対して、アーキテクチャの再認識を指示することで、移動後に利用可能な拡張命令セットに関する情報を通知する。これにより、HV 4 は、VM 5 に対して、マイグレーションが行なわれる際に拡張命令セットを再認識するトリガを与えることができる。

20

#### 【0045】

具体的には、再認識指示部 4 1 は、ライブマイグレーションに際して、以下の (a) 及び (b) の少なくとも一方のタイミングで、VM 5 に再認識指示を発行する。

(a) VM 5 の移動後に再認識指示を発行し、拡張命令セットを切り替えさせる (ポスト切り替え方式)。

30

(b) VM 5 の移動前に再認識指示を発行し、事前に拡張命令セットを切り替えさせる (プレ切り替え方式)。

#### 【0046】

なお、上記 (a) の場合、再認識指示は、移動された (移動後の) VM 5 に対して、移動先の物理サーバ 2 の HV 4 (再認識指示部 4 1) により発行される。

一方、上記 (b) の場合、再認識指示は、移動される (移動前の) VM 5 に対して、移動元の物理サーバ 2 の HV 4 (再認識指示部 4 1) により発行される。なお、このとき、再認識指示部 4 1 は、命令セット制限部 2 0 1 への指示を通じて、事前に、当該 VM 5 が自物理サーバ 2 で利用する拡張命令セットに関する情報を、移動先で用いられる拡張命令セットに関する情報に変更する。ここで、拡張命令セットに関する情報とは、上述した拡張命令セットのレベルのほか、拡張命令セットを特定する情報であってもよい。

40

#### 【0047】

制限レベル管理部 4 2 は、自 HV 4 が実行する VM 5 のレベルを管理するものである。例えば、制限レベル管理部 4 2 は、自 HV 4 が実行する VM 5 と、命令セット制限部 2 0 1 により当該 VM 5 に設定された拡張命令セットのレベルとが対応付けられたテーブルを管理する。なお、このテーブルはメモリ 2 0 b 等に格納される。

制限レベル管理部 4 2 は、自 HV 4 の配下に VM 5 が追加されたときや他の HV 4 から移動してきたときに、当該 VM 5 が利用する利用レベルを命令セット制限部 2 0 1 から取得して、テーブルに追加することによりテーブルを更新する。また、制限レベル管理部 4 2 は、命令セット制限部 2 0 1 により VM 5 の利用レベルが変更されたとき、又は、自 H

50

V 4 から他の H V 4 へ V M 5 が移動したときに、テーブルの当該 V M 5 のエントリの利用レベルを変更、又は、当該エントリを削除する。

【 0 0 4 8 】

〔 1 - 3 - 3 - 1 〕再認識指示部による再認識指示の実行可否の判断手法

ここで、再認識指示部 4 1 が、移動される V M 5 への再認識指示の発行を、上記 ( a ) 及び ( b ) のいずれのタイミングで実行するかについて説明する。

再認識指示部 4 1 は、制限レベル管理部 4 2 が管理するテーブルと、V M 配置管理サーバ 3 から通知される情報とに基づいて、再認識指示の発行タイミングを決定する。

【 0 0 4 9 】

具体的には、移動元の再認識指示部 4 1 は、移動される V M 5 の利用レベルを、制限レベル管理部 4 2 から取得して確認する。

また、移動元の再認識指示部 4 1 は、上述のように V M 配置管理サーバ 3 から通知された、移動先の物理サーバ 2 で利用可能な拡張命令セットに関する情報 ( 例えば利用可能なレベル ; 利用可能レベル ) を確認する。

【 0 0 5 0 】

そして、移動元の再認識指示部 4 1 は、移動される V M 5 が、移動元での利用レベル及び移動先での利用可能レベルに基づいて、再認識指示の発行タイミングを決定する。

一例として、移動元の再認識指示部 4 1 は、移動元での利用レベルが、移動先での利用可能レベルよりも高いと判断した場合、V M 5 の移動前に再認識指示を発行すると判断する。

【 0 0 5 1 】

上述のように、サポートする拡張命令セットのレベルが高い物理サーバ 2 から低い物理サーバ 2 へ V M 5 が移動される場合、V M 5 は、今まで使用していた命令を利用できないことがある。そこで、移動元の再認識指示部 4 1 は、移動される V M 5 が利用する拡張命令セットがマイグレーションにより縮退されると判断すると、移動前に ( 事前に ) 、V M 5 の利用レベルを移動先での利用可能レベルまで引き下げるのである。

【 0 0 5 2 】

これにより、再認識指示部 4 1 は、V M 5 が移動前に存在していたが移動後に存在しなくなった拡張命令セットを実行して、“例外” ( 例えば “Illegal Instruction” ) を発生させてしまう確率を低減させることができる。

一方、移動元の再認識指示部 4 1 は、移動元での利用レベルが、移動先での利用可能レベルよりも低いと判断した場合、移動前には V M 5 への再認識指示を行わず、V M 5 の移動後に再認識指示を発行すると判断する。

【 0 0 5 3 】

つまり、移動される V M 5 が利用する拡張命令セットがマイグレーションにより拡張される場合、再認識指示部 4 1 が、移動前に V M 5 の利用レベルを移動先での利用可能レベルまで引き上げてしまうと、“例外”が発生する可能性がある。すなわち、移動前にレベルを引き上げた結果、V M 5 が、移動元には本来存在しない、移動先で利用可能な命令を実行してしまう可能性がある。

【 0 0 5 4 】

そこで、移動元の再認識指示部 4 1 は、マイグレーションにより、移動される V M 5 が利用する拡張命令セットが拡張されると判断すると、移動前には V M 5 への再認識指示は行なわない。この場合、移動元の再認識指示部 4 1 は、移動先の再認識指示部 4 1 へ、移動された V M 5 に対する再認識指示の発行を要求してもよく、或いは、移動される V M 5 の移動前後のレベルの比較結果を通知してもよい。

【 0 0 5 5 】

なお、移動元の再認識指示部 4 1 は、移動元での利用レベルが、移動先での利用可能レベルと同じである判断した場合、再認識指示の発行を行なわないと判断する。

以上のように、移動元の再認識指示部 4 1 は、再認識指示の発行タイミングを決定する。

10

20

30

40

50

また、移動先の再認識指示部 4 1 は、V M 配置管理サーバ 3 によってライブマイグレーションによる V M 5 の移動が実行されると、移動された V M 5 へ再認識指示を発行する可否かを判断する。この判断は、移動元の再認識指示部 4 1 から、再認識指示の発行要求を受けたか否かによって行なわれてもよいし、移動された V M 5 の移動前後のレベルの比較結果を受け取った場合には、当該比較結果に基づいて行なわれてもよい。

【 0 0 5 6 】

なお、移動先の再認識指示部 4 1 は、上述した移動元の再認識指示部 4 1 と同様に、移動された V M 5 の移動前後のレベルを比較してもよい。

例えば、移動先の再認識指示部 4 1 は、上述のように V M 配置管理サーバ 3 から通知された、移動された V M 5 が移動元の物理サーバ 2 で利用していた拡張命令セットに関する情報（例えば利用レベル）を確認する。また、移動先の再認識指示部 4 1 は、移動された V M 5 が自 H V 4 での利用可能レベルを、制限レベル管理部 4 2 から取得して確認する。そして、移動先の再認識指示部 4 1 は、移動された V M 5 の移動元での利用レベルと、自 H V 4 での利用可能レベルとを比較して、比較結果を得てもよい。

10

【 0 0 5 7 】

このように、移動先の再認識指示部 4 1 は、自 H V 4 で得たレベルの比較結果、又は、移動元から通知された比較結果に基づき、V M 5 の移動元での利用レベルが、自 H V 4 での利用可能レベルよりも低い（拡張されたか）否かを判断する。

そして、移動先の再認識指示部 4 1 は、拡張されたと判断した場合、又は、移動元から再認識指示の発行要求を受けた場合、移動された V M 5 へ再認識指示を発行する。

20

【 0 0 5 8 】

なお、移動先の物理サーバ 2 の C P U 2 0 a は、移動された V M 5 に対して、移動先の物理サーバ 2 で利用可能な拡張命令セットを用いた制御を実行する。

以上のように、移動先の再認識指示部 4 1 は、移動された V M 5 が移動元の H V 4 で再認識を行っていない場合に、移動された V M 5 へ再認識指示を発行する。

なお、移動元及び移動先の再認識指示部 4 1 は、移動される V M 5 の移動元での利用レベルと移動先での利用可能レベルとを、互いにネゴシエーションを行なって交換（確認）してもよい。

【 0 0 5 9 】

また、上述した例では、再認識指示部 4 1 は、移動される V M 5 の移動元での利用レベルと移動先の利用可能レベルとの大小の比較によって、再認識指示の発行タイミングを決定するものとして説明したが、これに限定されるものではない。

30

例えば、V M 5 は、上述のように、再認識指示を受けて再認識を行なうと、C P U 2 0 a へ再認識後に指示する処理（コード）の切り替えを行なう。このとき、V M 5 は、少なくとも再認識に伴い利用できなくなった命令について、当該命令を含むコンパイル済みのコードを無効化し、利用可能な命令により再度コンパイルする処理等を行なう。このような処理は、例えば移動元の H V 4 で実行される場合、V M ホスト（H V 4）の負荷を増大させ、マイグレーションを含む各処理を遅延させる要因となり得る。

【 0 0 6 0 】

そこで、移動元の再認識指示部 4 1 は、移動元での利用レベルが、移動先での利用可能レベルよりも高いと判断した場合であっても、V M 5 によるコード切替の処理負荷が高くなると推定される場合には、再認識指示の発行を行なわないようにしてもよい。この場合、移動元の再認識指示部 4 1 は、移動先の再認識指示部 4 1 へ、移動された V M 5 に対する再認識指示の発行を要求してもよく、移動元の再認識指示部 4 1 で再認識指示を発行しない旨（理由等）を通知してもよい。

40

【 0 0 6 1 】

なお、再認識指示部 4 1 は、例えば、移動元での利用レベルと移動先での利用可能レベルとの差が所定値よりも大きい場合に、V M 5 によるコード切替の処理負荷が高くなると推定することができる。これは、レベル差が大きい場合、移動先で利用できなくなる命令の数が多く、V M 5 により再コンパイルが行なわれるコードの数が増大すると考えられる

50

ためである。

#### 【 0 0 6 2 】

以上のように、再認識指示部 4 1 及び制限レベル管理部 4 2 によれば、移動される V M 5 への再認識指示を、“例外”が発生しないように、又は / 及びマイグレーションの処理を遅延させないように、移動前又は移動後に選択的に実行することができる。従って、情報処理システム 1 の安定性を向上させることができる。

〔 1 - 3 - 3 - 2 〕再認識指示部による再認識部への再認識指示手法

再認識指示部 4 1 は、V M 5 の再認識部 5 1 に対して、以下の ( A ) ~ ( D ) のいずれかの手法により、再認識指示を通知することができる。

・割り込みによる手法

( A ) 予め V M 5 側にアーキテクチャの再認識用の割り込みを設定 ( 規定 ) する。再認識指示部 4 1 ( H V 4 ) は、設定された割り込みを発生させる ( 割り込み注入 ( Interrupt Injection ) を行なう ) ことにより、V M 5 ( 再認識部 5 1 ) に対して割り込みが発生したと認識させる。

( B ) H V 4 は、割り込み元として、V M 5 ( O S ) に仮想ハードウェア及びこの仮想ハードウェアのドライバを設定する。このドライバは、割り込みが発生すると、V M 5 ( 再認識部 5 1 ) に対して再認識指示が発行されたと認識させるように設定される。再認識指示部 4 1 は、仮想ハードウェアから V M 5 ( O S ) のドライバに割り込みを発生させる。

( C ) H V 4 は、割り込み元として、V M 5 ( O S ) に、V M 5 ( 再認識部 5 1 ) に対して再認識指示が発行されたことを割り込み要因とする割り込みベクタを設定する。再認識指示部 4 1 は、割り込みベクタにより割り込みを発生させる。

・ハンドラ登録による手法

( D ) V M 5 は、O S にアーキテクチャの再認識用のハンドラを登録し、H V 4 へ通知する。再認識指示部 4 1 は、イベント ( 再認識指示の発行 ) に応じて、V M 5 から通知されたハンドラを実行し、V M 5 ( 再認識部 5 1 ) に対して再認識指示が発行されたと認識させる。

#### 【 0 0 6 3 】

なお、上記 ( A ) ~ ( D ) は例示であり、上述したものに限られず、再認識指示の手法としては、既知の種々の手法を用いることが可能である。

〔 1 - 3 - 4 〕拡張 C P U 命令セットエミュレーション部の説明

拡張 C P U 命令セットエミュレーション部 ( エミュレーション部 ) 4 3 は、移動された V M 5 により、移動元の拡張命令セットに含まれる命令であって移動先の拡張命令セットに含まれない命令が実行された場合、実行された命令のエミュレーションを行なう。

#### 【 0 0 6 4 】

つまり、拡張 C P U 命令セットエミュレーション部 4 3 は、再認識指示部 4 1 により、移動前又は移動後に、移動される V M 5 に対して再認識指示が発行されたにもかかわらず、効果が得られなかった場合 ( “例外” が発生した場合 ) に実行される。

V M 5 へ再認識指示が発行されたにもかかわらず、“例外”が発生する場合としては、V M 5 ( O S ) が再認識指示を無視した場合のほか、V M 5 ( O S ) 又はアプリケーション ( ユーザプログラム ) が再認識に対応できない場合が挙げられる。

#### 【 0 0 6 5 】

具体的には、拡張 C P U 命令セットエミュレーション部 4 3 は、V M 5 で実行される命令を監視し、或る V M 5 ( 例えば移動された V M 5 ) により発行された命令が “例外” を発生させたことを認識すると、当該命令をトラップする。

また、拡張 C P U 命令セットエミュレーション部 4 3 は、“例外” を発生させた命令 ( 例外発生命令 ) を実行した V M 5 の移動元での拡張命令セットの命令レベルを確認し、当該命令が移動元で利用可能な命令であるか否かを判断する。なお、拡張 C P U 命令セットエミュレーション部 4 3 は、例外発生命令を発行した V M 5 の移動元の H V 4 及び命令レベルを、例えば当該命令を発行した V M 5 に問い合わせることで特定してもよい。或いは、H V 4 は、マイグレーションのときに取得した情報をメモリ 2 0 b 等に保存しておき、

10

20

30

40

50

拡張CPU命令セットエミュレーション部43が、メモリ20b等から取得してもよい。

【0066】

拡張CPU命令セットエミュレーション部43は、例外発生命令が移動元で利用可能な命令ではないと判断した場合、当該命令による“例外”の発生は、VM5による再認識の未実行に起因するものではないと判断し、通常の命令例外処理に移行する。

一方、拡張CPU命令セットエミュレーション部43は、例外発生命令が移動元で利用可能な命令であると判断した場合、当該命令のエミュレーションを実行する。

【0067】

また、拡張CPU命令セットエミュレーション部43は、例外発生命令のエミュレーションを実行すると、VM5に対して、再度の再認識指示を発行するか否かを判断する。この判断は、例えば、例外発生命令による“例外”が所定回数以上発生したか否かにより行なわれてもよい。

10

例えば、拡張CPU命令セットエミュレーション部43は、例外発生命令による“例外”が所定回数以上発生し、再度の再認識指示を発行すると判断した場合、再認識指示部41へ、VM5に対する再度の再認識指示を要求する。

【0068】

なお、再度の再認識指示においては、再認識指示部41は、VM5に対して、重要度に応じて例外発生命令の発行元（例えばユーザプログラム）を停止させることを要求（問い合わせ）してもよい。

そして、拡張CPU命令セットエミュレーション部43は、再認識指示部41へ再度の再認識指示を要求すると、又は、再度の再認識指示を発行しないと判断すると、例外”からの復帰を行なう。

20

【0069】

以上のように、拡張CPU命令セットエミュレーション部43によれば、移動先のVMホスト（HV4）において実装されていない拡張命令セットの命令が発行された場合でも、“例外”の発生を抑止することができる。また、拡張CPU命令セットエミュレーション部43によれば、再認識指示部41へVM5に対する再度の再認識指示が要求されるため、例外発生命令の発生確率を低減させることができる。

【0070】

なお、拡張CPU命令セットエミュレーション部43は、情報処理システム1がそなえる全ての物理サーバ2（CPU20a）が実行する拡張命令セットをエミュレーションできるように、全てのCPU20aの命令に対応可能であることが好ましい。

30

〔1-4〕動作例

次に、図4～図7を参照して、上述の如く構成された一実施形態の一例としての情報処理システム1における物理サーバ2の動作例を説明する。

【0071】

図4は、図3に示す物理サーバ2によるマイグレーション処理の一例を説明するフローチャートであり、図5は、図3に示す物理サーバ2によるマイグレーション処理の一例を説明する図である。

図6は、図3に示す物理サーバ2による例外発生命令への処理の一例を説明するフローチャートであり、図7は、図3に示す物理サーバ2による例外発生命令への処理の一例を説明する図である。

40

【0072】

なお、図5は、VM5の移動元及び移動先の物理サーバ2の双方の処理の説明に共通して用いる。また、図7は、VM5の移動先の物理サーバ2の処理の説明に用いる。

〔1-4-1〕マイグレーション処理

はじめに、図4及び図5を参照して、物理サーバ2によるマイグレーション（ライブマイグレーション）処理について説明する。

【0073】

VM配置管理サーバ3からVM5の移動元（移送元）のHV4へ、マイグレーションの

50

実施（再認識指示の要求）が通知されると（図5の矢印（i）参照）、再認識指示部41により、移動対象（移送対象）のVM5の利用レベルが確認される（図4のステップS1）。このとき、再認識指示部41は、制限レベル管理部42から移動対象のVM5の利用レベルを取得する（図5の矢印（ii）参照）。

【0074】

次いで、図4に示すように、再認識指示部41により、移動先（移送先）のVMホスト（HV4）での利用可能レベルが確認される（ステップS2）。

そして、再認識指示部41により、移動前にVM5による拡張命令セットの利用レベルの再認識を実施するか否かが判断される（ステップS3）。再認識指示部41により、例えば、移動されるVM5が利用する拡張命令セットがマイグレーションにより縮退されると判断された場合、移動前の再認識を実施すると判断され（ステップS3のYesルート）、処理がステップS4に移行する。

10

【0075】

ステップS4では、再認識指示部41により、移動されるVM5に対して、移動前の再認識指示が発行され、処理がステップS5に移行する。なお、このとき、再認識指示部41は、CPU20aの命令セット制限部201へ、移動対象のVM5の利用レベルを移動先での利用可能レベルに変更させる指示を行なう（図5の矢印（iii）参照）。そして、再認識指示部41は、移動対象のVM5の再認識部51へ、変更された拡張命令セット（利用レベル）を再認識させるべく、再認識指示を発行する（図5の矢印（iv）参照）。

【0076】

20

一方、再認識指示部41により、例えば、移動されるVM5が利用する拡張命令セットがマイグレーションにより拡張される、又は、利用レベルの変更がないと判断された場合、移動前の再認識を実施しないと判断される（ステップS3のNoルート）。そして、処理がステップS5に移行する。

ステップS5では、VM配置管理サーバ3、並びに、移動元及び移動先の物理サーバ2により、ライブマイグレーションによるVM5の移動が実行される。

【0077】

ライブマイグレーションが完了し、VM配置管理サーバ3により完了が通知されると（図5の矢印（i）参照）、移動先の再認識指示部41により、移動後にVM5による拡張命令セットの利用レベルの再認識を実施するか否かが判断される（ステップS6）。なお、この判断は、上述の如く、移動元の再認識指示部41からの再認識指示の発行要求の有無、移動元から受信した移動前後のレベルの比較結果、又は、移動先の再認識指示部41による移動前後のレベルの比較結果に基づいて行なわれてもよい。

30

【0078】

移動先の再認識指示部41により、移動後の再認識を実施すると判断された場合（ステップS6のYesルート）、処理がステップS7に移行する。

ステップS7では、再認識指示部41により、移動されたVM5に対して、移動後の再認識指示が発行され、処理が終了する。なお、このとき、再認識指示部41は、移動対象のVM5の再認識部51へ、移動先のVMホストの拡張命令セット（利用レベル）を再認識させるべく、再認識指示を発行する（図5の矢印（iv）参照）。

40

【0079】

一方、再認識指示部41により、移動後の再認識を実施しないと判断された場合（ステップS6のNoルート）、処理が終了する。

なお、上述の如く、移動元及び移動先の再認識指示部41は、ステップS1～S3及びS6において、移動されるVM5の移動元での利用レベルと移動先での利用可能レベルとを、互いにネゴシエーションを行なって交換（確認）してもよい。

【0080】

以上のように、第1実施形態に係る物理サーバ2によれば、VM5が、移動前後でアーキテクチャの再認識を行なうことができるため、異なる拡張命令セットを適用する物理サーバ2にVM5を移動させることができる。また、VM5が移動先で適用される拡張命令

50

セットに対応することができるため、移動先の拡張命令セットを制限せずに済み、VM5は、各VMホスト(HV4)において拡張命令セットの機能を最大限利用することができる。

#### 【0081】

従って、物理サーバ2間でのVM5の移動を許容する情報処理システム1において、VM5による情報処理システム1の利用効率を向上させることができる。

#### 〔1-4-2〕例外発生命令への処理

次に、図6及び図7を参照して、物理サーバ2による例外発生命令への処理について説明する。

#### 【0082】

VM5により命令が発行されると(図7の矢印(v)参照)、拡張CPU命令セットエミュレーション部43により、発行された命令が監視される。VM5から発行された命令による“例外”が発生すると(図6のステップS11)、拡張CPU命令セットエミュレーション部43により、例外発生命令が確認される(ステップS12)。

また、拡張CPU命令セットエミュレーション部43により、例外発生命令を発行したVM5の移動元が特定され、移動元で利用可能な命令レベルが確認される(ステップS13)。

#### 【0083】

そして、拡張CPU命令セットエミュレーション部43により、例外発生命令が移動元の命令レベルで利用可能か否かが判定される(ステップS14)。例外発生命令が移動元の命令レベルで利用できない場合(ステップS14のNoルート)、拡張CPU命令セットエミュレーション部43により、通常の命令例外処理が実行され(ステップS15)、処理がステップS17に移行する。

#### 【0084】

一方、例外発生命令が移動元の命令レベルで利用可能な場合(ステップS14のYesルート)、拡張CPU命令セットエミュレーション部43により、当該命令のエミュレーションが実行され(ステップS16)、処理がステップS17に移行する。なお、ステップS16では、例外発生命令のエミュレーションがCPU20aにより実行される(図7の矢印(vi)参照)。

#### 【0085】

ステップS17では、拡張CPU命令セットエミュレーション部43により、VM5に対する再度の再認識を実行するか否かが判断される。なお、この判断は、例えば、例外発生命令による“例外”が所定回数以上発生したか否かに基づき行なわれる。再度の再認識を実行すると判断された場合(ステップS17のYesルート)、拡張CPU命令セットエミュレーション部43により、再認識指示部41へ、再度の再認識指示の発行が要求される(ステップS18、図7の矢印(vii)参照)。なお、再認識指示部41は、再度の再認識指示の発行要求を受けると、対象のVM5へ、再認識指示を発行する(図7の矢印(viii)参照)。そして、処理がステップS19へ移行する。

#### 【0086】

一方、ステップS17において、再度の再認識を実行しないと判断された場合(ステップS17のNoルート)、処理がステップS19へ移行する。

ステップS19では、拡張CPU命令セットエミュレーション部43により、例外発生命令について、“例外”からの復帰が行なわれ、処理が終了する。

#### 〔2〕第1変形例

次に、図8及び図9を参照して、一実施形態の第1変形例について説明する。

#### 【0087】

図8は、図1に示す物理サーバ2の構成の第1変形例を示す図であり、図9は、図8に示す物理サーバ2'によるマイグレーション処理の一例を説明する図である。

図8に示すように、第1変形例に係る物理サーバ2'は、図3に示す物理サーバ2と比較して、HV4'が命令セット制限部44をさらにそなえるとともに、CPU20a'が

10

20

30

40

50

命令セット制限部 201 をそなえていない点異なる。

【0088】

なお、図 8 において、図 3 に示す符号と同一の符号は、図 3 に示す構成と同一又は略同一のため、重複した説明は省略する。また、図 9 において、図 5 に示す符号と同一の符号は、図 5 に示す処理と同一又は略同一のため、重複した説明は省略する。

命令セット制限部 44 は、命令セット制限部 201 と同様の機能を、HV4' (ソフトウェア) により実現する。すなわち、命令セット制限部 44 は、再認識指示部 41 からの拡張命令セットの利用レベルの変更要求に応じて、VM5 ごとに拡張命令セットの利用制限をかけることができる (図 9 の矢印 (iii') 参照)。

【0089】

また、命令セット制限部 44 は、特定の VM5 (OS) からの利用可能な拡張命令セットの問い合わせ (例えば CPUID 命令) に対して、利用可能な拡張命令セットに関する情報の応答を行なう。

例えば、図 3 に示す HV4 は、VM5 の再認識部 51 から CPUID 命令等の拡張命令セットに関する情報の取得要求を受信すると、当該要求を CPU20a の命令セット制限部 201 へ渡し、命令セット制限部 201 からの応答を再認識部 51 へ返す。

【0090】

一方、図 8 に示す HV4' は、再認識部 51 から CPUID 命令等の拡張命令セットに関する情報の取得要求を受信すると、インターセプトを行ない、CPU20a' への転送を抑止し、当該取得要求を命令セット制限部 44 へ渡す。そして、HV4' は、命令セット制限部 44 からの応答を、CPU20a' からの応答として再認識部 51 へ返す。

このように、一実施形態の第 1 変形例に係る物理サーバ 2' によれば、一実施形態と同様の効果を奏することができる。

【0091】

また、物理サーバ 2' によれば、CPU20a が命令セット制限部 201 をそなえていない場合にも、HV4' (命令セット制限部 44) によるソフトウェアの処理によって、命令セット制限部 201 と同様の機能を実現することができ、利便性が高い。

〔3〕第 2 変形例

次に、図 10 及び図 11 を参照して、一実施形態の第 2 変形例について説明する。

【0092】

図 10 は、図 1 に示す物理サーバ 2 の構成の第 2 変形例を示す図であり、図 11 は、図 10 に示す物理サーバ 2' による例外発生命令への処理の一例を説明する図である。

図 10 に示すように、第 2 変形例に係る物理サーバ 2' は、図 3 に示す物理サーバ 2 と比較して、拡張 CPU 命令セットエミュレーション部 43 の代わりに、HV4' が拡張 CPU 命令セット変換部 45 をそなえる点異なる。

【0093】

なお、図 10 において、図 3 に示す符号と同一の符号は、図 3 に示す構成と同一又は略同一のため、重複した説明は省略する。また、図 11 において、図 7 に示す符号と同一の符号は、図 7 に示す処理と同一又は略同一のため、重複した説明は省略する。

拡張 CPU 命令セット変換部 (エミュレーション部) 45 は、拡張 CPU 命令セットエミュレーション部 43 と同様に、トラップした例外発生命令のエミュレーションを行なうことができる。

【0094】

また、拡張 CPU 命令セット変換部 45 は、VM5 により実行された例外発生命令について以前にエミュレーションを行なっている場合、当該命令を、CPU20a が実行可能な命令に変換する (コードを書き換える)。

例えば、HV4' による命令のエミュレーションの処理時間は、CPU20a による命令の処理時間の 10 倍以上となるため、エミュレーションの実行頻度の増加は、処理遅延の要因となる。また、HV4' によりトラップが行なわれると CPU20a の動作モードが変わるため、トラップの実行頻度の増加も、処理遅延の要因となる。

10

20

30

40

50



## 【 0 0 9 5 】

そこで、拡張CPU命令セット変換部45は、以前にエミュレーションを行なった例外発生命令については、エミュレーションに代えて、例外発生命令の変換を行なうことにより、エミュレーション及びトラップの実行頻度を低下させる。

具体的には、拡張CPU命令セット変換部45は、例外発生命令を確認し（図11の矢印（v'）参照）、以前にエミュレーションを行なっていると判断すると、以下の（I）～（III）のいずれかの手法により、例外発生命令の変換を行なう。

（I）拡張CPU命令セット変換部45は、例外発生命令を、CPU20aが実行可能なコードに置き換える（バイナリパッチをあてる）。

（II）拡張CPU命令セット変換部45は、例外発生命令を、代替りの処理へのジャンプに置き換え、ジャンプ先に、CPU20aが実行可能な処理とジャンプ元へのリターンとを記述する。

（III）拡張CPU命令セット変換部45は、メモリ20b上に展開された、例外発生命令を含むプログラムの命令内容をスキャンし、禁止命令をCPU20aが実行可能な命令に書き換えるコンパイル（バイナリコンパイル）を行なう。

## 【 0 0 9 6 】

そして、拡張CPU命令セット変換部45は、上記（I）～（III）のいずれかの手法により変換した命令をCPU20aへ渡す（図11の矢印（vi'）参照）。

なお、上記（I）～（III）は例示であり、上述したものに限られず、例外発生命令の変換処理としては、既知の種々の手法を用いることが可能である。

このように、一実施形態の第2変形例に係る物理サーバ2"によれば、一実施形態と同様の効果を奏することができる。

## 【 0 0 9 7 】

また、物理サーバ2"によれば、エミュレーション及びトラップの実行頻度を低下させることができるため、物理サーバ2"での処理遅延を緩和させることができ、情報処理システム1の性能を向上させることができる。

## 〔 4 〕 その他

以上、本発明の好ましい実施形態について詳述したが、本発明は、係る特定の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲内において、種々の変形、変更して実施することができる。

## 【 0 0 9 8 】

例えば、一実施形態、並びに、第1及び第2変形例に係る情報処理システム1は、2つの物理サーバ2～2"、及び、3つのVM5をそなえるものとして説明したが、これに限定されるものではなく、任意の数の物理サーバ2～2"及びVM5をそなえてよい。

また、上述した一実施形態、並びに、第1及び第2変形例に係る物理サーバ2～2"の機能は、任意に組み合わせてもよい。

## 【 0 0 9 9 】

なお、一実施形態、並びに、第1及び第2変形例に係る物理サーバ2～2"の各種機能の全部もしくは一部は、コンピュータ（CPU、情報処理装置、各種端末を含む）が所定のプログラムを実行することによって実現されてもよい。

そのプログラムは、例えばフレキシブルディスク、CD、DVD、ブルーレイディスク等のコンピュータ読取可能な記録媒体（例えば図2に示す記録媒体20h）に記録された形態で提供される。なお、CDとしては、CD-ROM、CD-R、CD-RW等が挙げられる。また、DVDとしては、DVD-ROM、DVD-RAM、DVD-R、DVD-RW、DVD+R、DVD+RW等が挙げられる。この場合、コンピュータはその記録媒体からプログラムを読み取って内部記憶装置または外部記憶装置に転送し格納して用いる。

## 【 0 1 0 0 】

ここで、コンピュータとは、ハードウェアとOSとを含む概念であり、OSの制御の下で動作するハードウェアを意味している。また、OSが不要でアプリケーションプログラ

10

20

30

40

50

ム単独でハードウェアを動作させるような場合には、そのハードウェア自体がコンピュータに相当する。ハードウェアは、少なくとも、CPU等のマイクロプロセッサと、記録媒体に記録されたコンピュータプログラムを読み取る手段とをそなえている。上記プログラムは、上述のようなコンピュータに、一実施形態、並びに、第1及び第2変形例に係る物理サーバ2～2”の各種機能を実現させるプログラムコードを含んでいる。また、その機能の一部は、アプリケーションプログラムではなくOSによって実現されてもよい。

〔5〕付記

以上の実施形態、並びに、第1及び第2変形例に関し、更に以下の付記を開示する。

（付記1）

第1の命令セットを用いて仮想マシンを制御する第1の情報処理装置へ、第2の命令セットを用いて仮想マシンを制御する第2の情報処理装置からの仮想マシンの移動を許す

10

、  
前記第2の情報処理装置から前記第1の情報処理装置へ移動される仮想マシンに対して前記第1の命令セットに関する情報を通知し、

前記第1の情報処理装置は、移動された前記仮想マシンに対して、前記第1の命令セットを用いた制御を実行する、

ことを特徴とする、仮想マシン管理方法。

（付記2）

前記通知は、前記第1の命令セットの水準及び前記第2の命令セットの水準に基づいて、仮想マシンの移動前又は移動後に、前記移動される仮想マシンに対する、前記移動される仮想マシンの制御に用いられる命令セットを認識させる指示によって行なわれる、  
ことを特徴とする、付記1記載の仮想マシン管理方法。

20

（付記3）

前記通知において、前記第2の情報処理装置は、

前記移動される仮想マシンの制御に用いられる命令セットに関する情報を、前記第2の命令セットに関する情報から前記第1の命令セットに関する前記情報に変更し、

前記移動される仮想マシンに対して、前記変更した命令セットに関する情報を認識させる前記指示を行なう、

ことを特徴とする、付記2記載の仮想マシン管理方法。

（付記4）

前記通知において、前記第1の情報処理装置は、

前記移動された仮想マシンに対して、前記移動された仮想マシンの制御に用いられる命令セットに関する情報を認識させる前記指示を行なう、

ことを特徴とする、付記2記載の仮想マシン管理方法。

30

（付記5）

前記通知は、前記第1の情報処理装置で動作する仮想マシンを制御する第1の制御部又は前記第2の情報処理装置で動作する仮想マシンを制御する第2の制御部が、前記移動される仮想マシンに対して割り込みを発生させて前記指示を行なうこと、並びに、前記第1又は第2の制御部が、前記移動される仮想マシンにより登録された前記指示を示すハンドラを実行すること、の少なくとも一方により行なわれる、

ことを特徴とする、付記2～4のいずれか1項記載の仮想マシン管理方法。

40

（付記6）

前記第1の情報処理装置は、

前記移動された仮想マシンにより、前記第2の命令セットに含まれる命令であって前記第1の命令セットに含まれない命令が前記第1の情報処理装置において実行された場合、前記実行された命令のエミュレーションを行ない、

前記移動された仮想マシンに対して前記第1の命令セットに関する情報を再度通知する

、  
ことを特徴とする、付記1～5のいずれか1項記載の仮想マシン管理方法。

（付記7）

50

前記第 1 の情報処理装置は、前記実行された命令について以前にエミュレーションを行なっている場合、前記実行された命令を、前記第 1 の情報処理装置が実行可能な命令に変換する、

ことを特徴とする、付記 6 記載の仮想マシン管理方法。

(付記 8)

第 1 の命令セットを用いて仮想マシンを制御する第 1 の仮想マシン管理装置であって、第 2 の命令セットを用いて仮想マシンを制御する第 2 の仮想マシン管理装置からの仮想マシンの移動を許容する制御部をそなえ、

前記制御部は、前記第 2 の仮想マシン管理装置から前記第 1 の仮想マシン管理装置へ移動される仮想マシンに対して前記第 1 の命令セットに関する情報を通知する通知部をそなえ

10

るとともに、移動された前記仮想マシンに対して、前記第 1 の命令セットを用いた制御を実行する処理部をさらにそなえる、

ことを特徴とする、仮想マシン管理装置。

(付記 9)

前記通知部は、前記第 1 の命令セットの水準及び前記第 2 の命令セットの水準に基づいて、仮想マシンの移動前又は移動後に、前記移動される仮想マシンに対して、前記移動される仮想マシンの制御に用いられる命令セットを認識させる指示を行なう、

ことを特徴とする、付記 8 記載の仮想マシン管理装置。

(付記 10)

20

前記制御部は、前記第 1 の仮想マシン管理装置から前記第 2 の仮想マシン管理装置への仮想マシンの移動を許容し、

前記第 1 の仮想マシン管理装置から前記第 2 の仮想マシン管理装置へ移動される仮想マシンの制御に用いられる命令セットに関する情報を、前記第 1 の命令セットに関する情報から前記第 2 の命令セットに関する前記情報に変更する命令セット制限部をさらにそなえ、

前記通知部は、前記第 1 の仮想マシン管理装置から前記第 2 の仮想マシン管理装置へ移動される仮想マシンに対して、前記変更した命令セットに関する情報を認識させる前記指示を行なう、

ことを特徴とする、付記 9 記載の仮想マシン管理装置。

30

(付記 11)

前記通知部は、前記移動された仮想マシンに対して、前記移動された仮想マシンの制御に用いられる命令セットに関する情報を認識させる前記指示を行なう、

ことを特徴とする、付記 9 記載の仮想マシン管理装置。

(付記 12)

前記移動された仮想マシンにより、前記第 2 の命令セットに含まれる命令であって前記第 1 の命令セットに含まれない命令が前記第 1 の仮想マシン管理装置において実行された場合、前記実行された命令のエミュレーションを行なうエミュレーション部をさらにそなえ、

前記通知部は、前記移動された仮想マシンに対して前記第 1 の命令セットに関する情報を再度通知する、

ことを特徴とする、付記 8 ~ 11 のいずれか 1 項記載の仮想マシン管理装置。

40

(付記 13)

前記エミュレーション部は、前記実行された命令について以前にエミュレーションを行なっている場合、前記実行された命令を、前記第 1 の仮想マシン管理装置が実行可能な命令に変換する、

ことを特徴とする、付記 12 記載の仮想マシン管理装置。

(付記 14)

第 1 の命令セットを用いて仮想マシンを制御する第 1 のコンピュータに、

第 2 の命令セットを用いて仮想マシンを制御する第 2 のコンピュータからの仮想マシン

50

の移動を許容し、

前記第 2 のコンピュータから前記第 1 のコンピュータへ移動される仮想マシンに対して前記第 1 の命令セットに関する情報を通知し、

移動された前記仮想マシンに対して、前記第 1 の命令セットを用いた制御を実行する、処理を前記第 1 のコンピュータに実行させることを特徴とする、仮想マシン管理プログラム。

(付記 15)

前記通知は、前記第 1 の命令セットの水準及び前記第 2 の命令セットの水準に基づいて、仮想マシンの移動前又は移動後に、前記移動される仮想マシンに対する、前記移動される仮想マシンの制御に用いられる命令セットを認識させる指示によって行なわれる、  
ことを特徴とする、付記 14 記載の仮想マシン管理プログラム。

10

(付記 16)

前記第 1 のコンピュータから前記第 2 のコンピュータへの仮想マシンの移動を許容し、前記第 1 のコンピュータから前記第 2 のコンピュータへ移動される仮想マシンの制御に用いられる命令セットに関する情報を、前記第 1 の命令セットに関する情報から前記第 2 の命令セットに関する前記情報に変更する、  
処理を前記第 1 のコンピュータに実行させ、

前記通知は、前記第 1 のコンピュータから前記第 2 のコンピュータへ移動される仮想マシンに対する、前記変更した命令セットに関する情報を認識させる前記指示によって行なわれる、  
ことを特徴とする、付記 15 記載の仮想マシン管理プログラム。

20

(付記 17)

前記通知は、前記移動された仮想マシンに対する、前記移動された仮想マシンの制御に用いられる命令セットに関する情報を認識させる前記指示によって行なわれる、  
ことを特徴とする、付記 15 記載の仮想マシン管理プログラム。

(付記 18)

前記通知は、前記第 1 のコンピュータで動作する仮想マシンを制御する制御部により、前記移動される仮想マシンに対して割り込みを発生させて前記指示を行なうこと、並びに、前記制御部により、前記移動される仮想マシンにより登録された前記指示を示すハンドラを実行すること、の少なくとも一方により実行される、  
ことを特徴とする、付記 15 ~ 17 のいずれか 1 項記載の仮想マシン管理プログラム。

30

(付記 19)

前記移動された仮想マシンにより、前記第 2 の命令セットに含まれる命令であって前記第 1 の命令セットに含まれない命令が前記第 1 のコンピュータにおいて実行された場合、前記実行された命令のエミュレーションを行ない、

前記移動された仮想マシンに対して前記第 1 の命令セットに関する情報を再度通知する、  
処理を前記第 1 のコンピュータに実行させることを特徴とする、付記 14 ~ 18 のいずれか 1 項記載の仮想マシン管理プログラム。

(付記 20)

前記実行された命令について以前にエミュレーションを行なっている場合、前記第 1 のコンピュータで動作する仮想マシンを制御する制御部により、前記実行された命令を、前記第 1 のコンピュータが実行可能な命令に変換する、  
処理を前記第 1 のコンピュータに実行させることを特徴とする、付記 19 記載の仮想マシン管理プログラム。

40

【符号の説明】

【0101】

1 情報処理システム

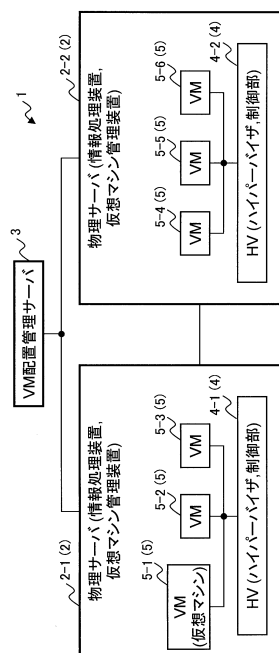
2, 2', 2'', 2-1, 2-2 物理サーバ(情報処理装置, 仮想マシン管理装置)

50

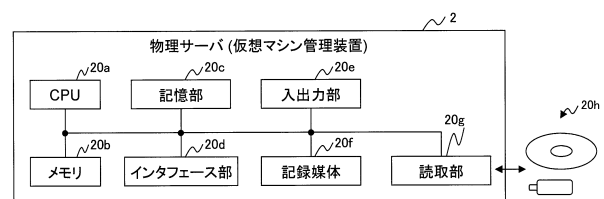
- 3 VM配置管理サーバ  
 4, 4', 4'', 4-1, 4-2 HV (ハイパーバイザ, 制御部)  
 5, 5-1 ~ 5-6 VM (仮想マシン)  
 20a, 20a' CPU (プロセッサ, 処理部)  
 20b メモリ  
 20c 記憶部  
 20d インタフェース部  
 20e 入出力部  
 20f, 20h 記録媒体  
 20g 読取部  
 41 再認識指示部  
 42 制限レベル管理部  
 43 拡張CPU命令セットエミュレーション部 (エミュレーション部)  
 44, 201 命令セット制限部  
 45 拡張CPU命令セット変換部 (エミュレーション部)  
 51 再認識部

10

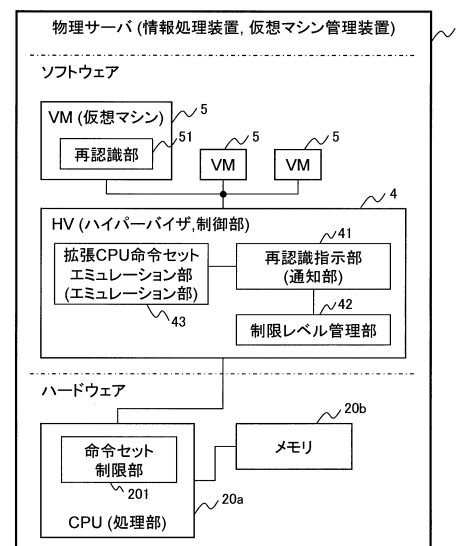
【図1】



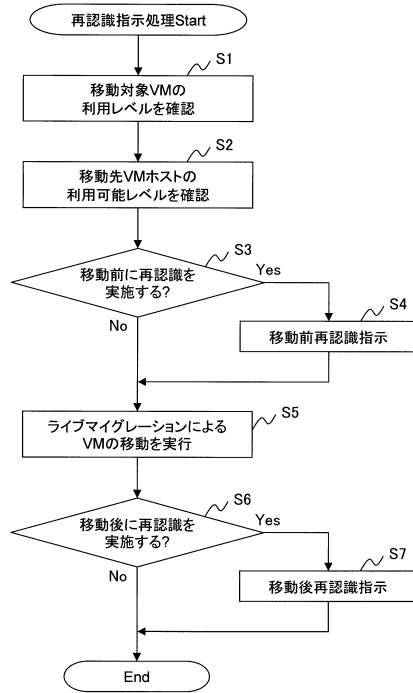
【図2】



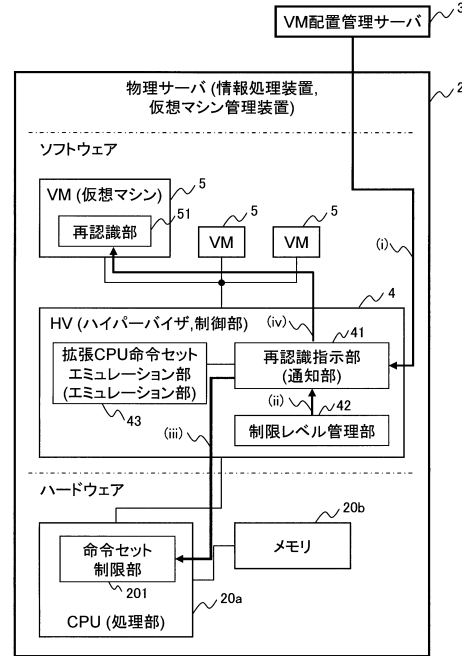
【図3】



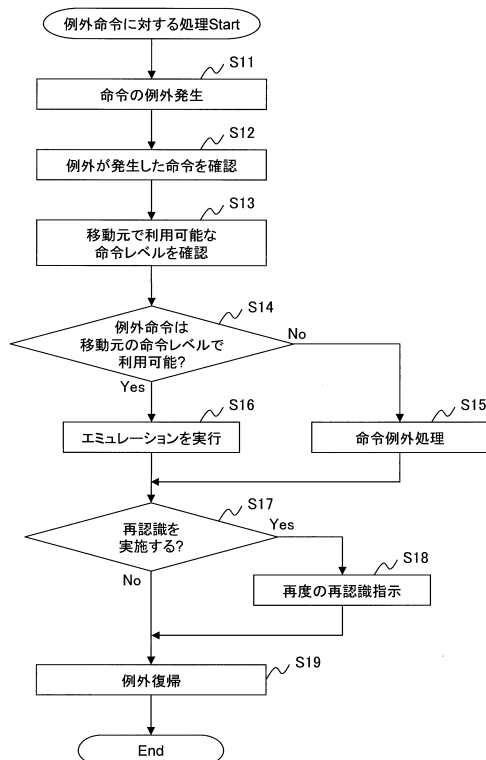
【図 4】



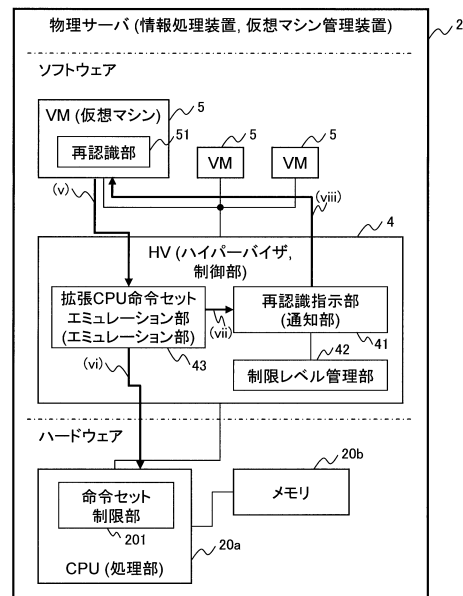
【図 5】



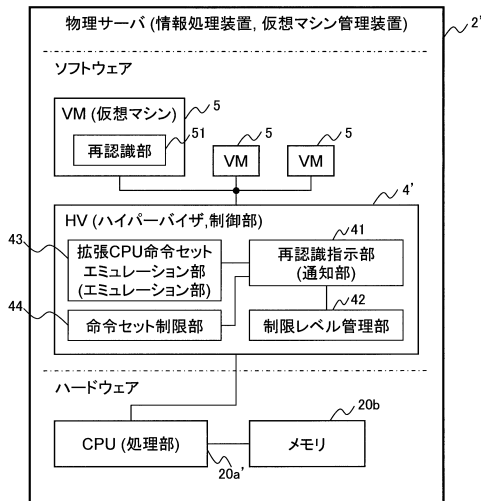
【図 6】



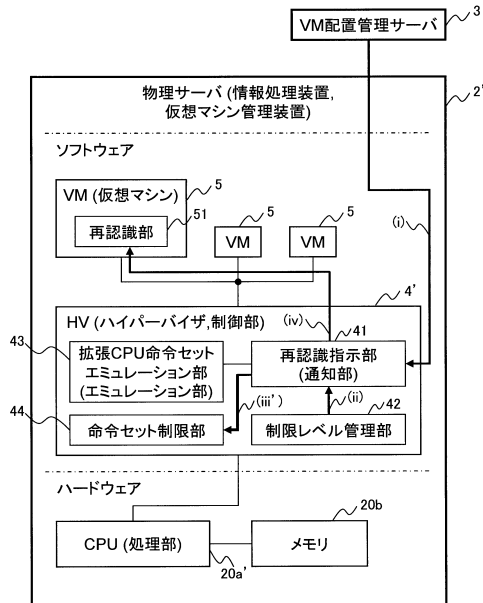
【図 7】



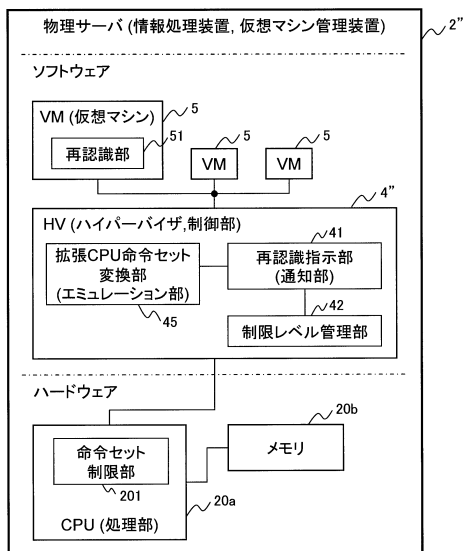
【図 8】



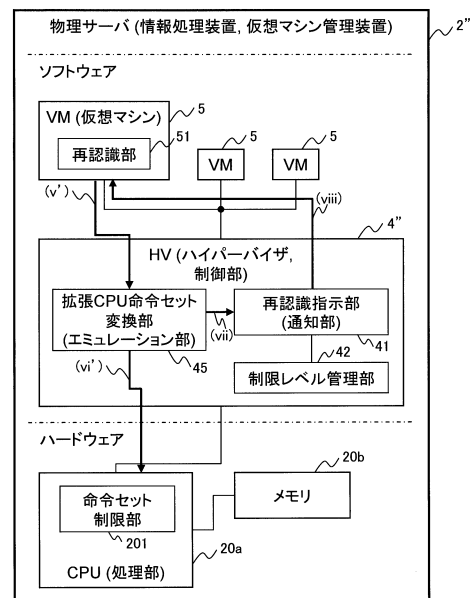
【図 9】



【図 10】



【図 11】



---

フロントページの続き

(56)参考文献 特開 2 0 1 0 - 2 7 2 0 5 5 ( J P , A )

(58)調査した分野(Int.Cl. , D B 名)

G 0 6 F            9 / 5 0

G 0 6 F            9 / 4 6