

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-171293

(P2008-171293A)

(43) 公開日 平成20年7月24日(2008.7.24)

(51) Int. Cl.		F I		テーマコード (参考)	
G 0 6 F	9/50	(2006.01)	G 0 6 F	9/46	4 6 2 A
G 0 6 F	9/46	(2006.01)	G 0 6 F	9/46	3 5 0
G 0 6 F	9/48	(2006.01)	G 0 6 F	9/46	3 1 5 Z

審査請求 未請求 請求項の数 3 O L (全 8 頁)

(21) 出願番号	特願2007-5350 (P2007-5350)	(71) 出願人	000005108
(22) 出願日	平成19年1月15日 (2007. 1. 15)		株式会社日立製作所
			東京都千代田区丸の内一丁目6番6号
		(74) 代理人	100080001
			弁理士 筒井 大和
		(72) 発明者	井上 裕功
			神奈川県秦野市堀山下1番地 株式会社日
			立製作所エンタープライズサーバ事業部内

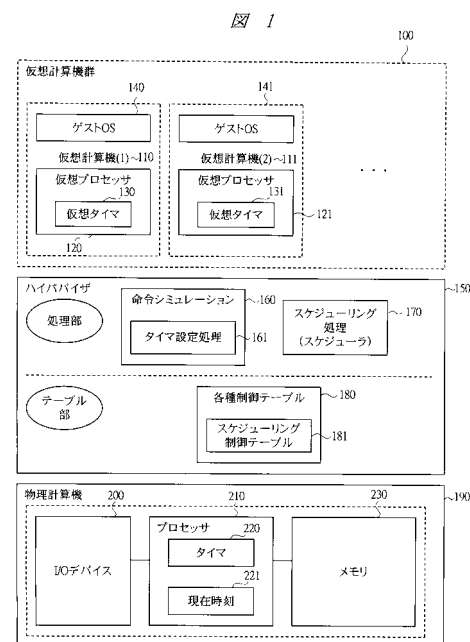
(54) 【発明の名称】 仮想計算機システムのスケジューリング方法

(57) 【要約】

【課題】仮想計算機におけるスケジューリング処理の3つの基本要請事項である、「(1) 割込みの誤差をなるべく少なく」、「(2) 各仮想計算機のプロセッサ使用時間の累積を均等に」、「(3) プロセッサの空き時間を有効に利用する」、が同時に満足されるスケジューリングの方法を提供し、特に(1)については、タイマ割込み時の仮想計算機への報告遅延が0となるような機能を実現する。

【解決手段】物理計算機におけるタイマ機構を、物理タイマをタイマ源とする仮想タイマという形で、各仮想計算機に専用に持たせるようにする。ハイパバイザ150は仮想計算機へのスケジューリング処理の際、各仮想計算機140、141の「仮想タイマ値」や「プロセッサ使用時間の累積」などの情報を利用して、優先的にディスパッチすべき仮想計算機とそのディスパッチ時間を算出してディスパッチを行う。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

物理計算機を論理的に分割して複数の仮想計算機として使用する仮想計算機システムにおいて、優先度に従って各仮想計算機にプロセッサを割当てるスケジューリング方法であって、

前記各仮想計算機のタイマ割込みの起こる時刻を格納する仮想タイマ値、プロセッサ使用時間の累積、タイムスライス値、および現在時刻を設け、前記仮想タイマ値、前記プロセッサ使用時間の累積、前記タイムスライス値、および前記現在時刻を前記各仮想計算機のプロセッサ割当てアルゴリズムに利用して、前記各仮想計算機上で動作するプログラムが設定するタイマによる割込みを発生させることを特徴とする仮想計算機システムのスケジューリング方法。

10

【請求項 2】

請求項 1 記載の仮想計算機システムのスケジューリング方法において、

前記プロセッサ割当てアルゴリズムは、前記各仮想計算機の仮想タイマ値を利用して、他の仮想計算機のプロセッサ割当て時間を、最も近い将来の仮想タイマ値までに短縮することでタイマによる割込みを発生させることを特徴とする仮想計算機システムのスケジューリング方法。

【請求項 3】

請求項 1 記載の仮想計算機システムのスケジューリング方法において、

前記プロセッサ割当てアルゴリズムは、前記各仮想計算機のプロセッサ使用時間の累積を利用して、均等なサービス率を実現するために累積時間が最も少ない仮想計算機にプロセッサを割当てることを特徴とする仮想計算機システムのスケジューリング方法。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、仮想計算機システム、特に物理計算機を論理的に分割して複数の仮想計算機として使用する仮想計算機システムにおける、精度の高いタイマ割込みを実現するスケジューリング方法に適用して有効な技術に関するものである。

【背景技術】

【0002】

30

仮想計算機システムのスケジューリング方法の一つとして、タイムスライス方式が知られている。タイムスライス方式は、物理計算機の命令プロセッサの稼動時間をタイムスライスと呼ばれる時間間隔で分割して利用する技術である。仮想計算機ごとにあらかじめ処理能力を定義しておき、その定義に従って各仮想計算機に対して物理計算機のプロセッサ割当てをタイムスライス単位で行う。これにより、複数の仮想計算機で物理計算機の処理能力を分割利用できる。

【0003】

この種のタイムスライス方式による仮想計算機のスケジューリング方法としては、例えば特許文献 1 に、タイムスライス時間を可変的な値とすることによりスケジュールパターンの偏りの発生を防止するスケジューリング方法が開示されている。仮想計算機のサービス率によって、一回に割当てるタイムスライス時間を可変的な値とし、サービス率により異なっていた一定時間におけるスケジューリング回数が同一となるようにスケジューリングする技術である。

40

【0004】

一方、各仮想計算機のプライオリティの変化に応じて、ハイパバイザが仮想計算機に対するプロセッサの割当て対象・性能・時間を算出して割当てるスケジューリング方法が、例えば特許文献 2 に開示されている。各仮想計算機上にプライオリティ設定部と監視部を設け、ハイパバイザは「各仮想計算機のプロセッサ割当て時間の累積」や「前単位時間のプロセッサ資源の過不足」といったプライオリティの変化通知を受付けて、それに応じてプロセッサ資源の割当てを行う技術である。

50

【特許文献1】特開2003-177928号公報

【特許文献2】特開2005-18560号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

ところで、上記特許文献1で述べられている、従来技術のタイムスライス方式のスケジューリング方法には次の問題が存在する。仮想計算機に対する割り込みが発生した場合、その割り込みを受付ける仮想計算機が物理計算機の命令プロセッサを使用できるようになるまで、その割り込みを受付けられないため、割り込み処理の遅延が発生し、仮想計算機の処理性能の低下を招くという点である。

10

【0006】

また、上記特許文献2ではそれを解決するため、各仮想計算機のプライオリティの変化に応じてプロセッサ資源（特に割当て時間）を変えて、割り込み処理の遅延を改善している。これにより、低コストで安定運用ができるプロセッサ割当てスケジューリング方法が実現される。

【0007】

しかし、割り込みの中でも特にタイマ割り込みのように、ハイパバイザが割り込みを発生させるべき時刻が既知である割り込み処理に対して、その仮想計算機への報告遅延を0にするような方法は開示されていない。

【0008】

そこで、本発明では、仮想計算機におけるスケジューリング処理の3つの基本要請事項である、「（1）割り込みの誤差をなるべく少なく」、「（2）各仮想計算機のプロセッサ使用時間の累積を均等に」、「（3）プロセッサの空き時間を有効に利用する」、が同時に満足されるスケジューリングの方法を提供し、特に（1）については、タイマ割り込み時の仮想計算機への報告遅延が0となるような機能を実現することを目的とするものである。

20

【0009】

本発明の前記ならびにその他の目的と新規な特徴は、本明細書の記述および添付図面から明らかになるであろう。

【課題を解決するための手段】

30

【0010】

本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、次のとおりである。

【0011】

本発明は、上記の目的を達成するため、以下のような特徴を有するものである。物理計算機におけるタイマ機構を、物理タイマをタイマ源とする仮想タイマという形で、各仮想計算機に専用に持たせるようにする。ハイパバイザ（プロセッサ割当てアルゴリズム利用）は仮想計算機へのスケジューリング処理の際、各仮想計算機の「仮想タイマ値」や「プロセッサ使用時間の累積」などの情報を利用して、優先的にディスパッチ（プロセッサの割当て）すべき仮想計算機とそのディスパッチ時間を算出してディスパッチを行うものである。

40

【発明の効果】

【0012】

本願において開示される発明のうち、代表的なものによって得られる効果を簡単に説明すれば以下のとおりである。

【0013】

本発明によれば、ハイパバイザのスケジューリング処理に、「仮想タイマ値」を考慮に入れることにより基本要請事項（1）が、「プロセッサ使用時間の累積」を考慮に入れることにより基本要請事項（2）が解決され、さらに休止状態のプロセスをディスパッチの対象としないことで基本要請事項（3）も解決される。これら3つの機能の実現により、

50

プロセッサリソースの効率的な運用が可能となり、仮想計算機の処理性能の向上が期待できる。

【発明を実施するための最良の形態】

【0014】

以下、本発明の実施の形態を図面に基づいて詳細に説明する。なお、実施の形態を説明するための全図において、同一の部材には原則として同一の符号を付し、その繰り返しの説明は省略する。

【0015】

図1は、本発明の一実施の形態における仮想計算機システムの構成例を示す図である。本実施の形態の仮想計算機システムにおいて、物理計算機190は、I/Oデバイス200、プロセッサ210、およびメモリ230を有する。プロセッサ210は、タイマ220と現在時刻221を内部に有する。

【0016】

ハイパバイザ150は、物理計算機190上で動作するソフトウェアであり、プロセッサ210などの物理計算機資源を論理的に分割して仮想計算機110, 111, ...を構築し、それらを管理、制御する機能を持つ。これらの仮想計算機110, 111, ...は、仮想計算機群100を構成する。

【0017】

仮想計算機110, 111, ...内の仮想プロセッサ120, 121, ...は、物理計算機190のプロセッサ210におけるタイマ220と同じ仕様の仮想タイマ130, 131, ...を有する。この仮想タイマ130, 131, ...は、物理計算機190のタイマ220をタイマ源とする論理的タイマである。

【0018】

これらの機能により、仮想計算機110, 111, ...の上では、ゲストオペレーティングシステム(OS)140, 141, ...を動作させることができる。

【0019】

プロセッサ210の分割は、ある一定期間(タイムスライス値)ごとにプロセッサ210を各仮想計算機110, 111, ...に割り当てることにより実現する。このプロセッサ210の時分割方式によりハイパバイザ150は、各仮想計算機110, 111, ...に対するタイムスライスのスケジューリングを実行したり、各仮想計算機110, 111, ...に物理計算機190から上がってくる割り込みを発生させたりする。この処理を実現するのが、命令シミュレーション160内のタイマ設定処理161、スケジューリング処理170、および各種制御テーブル180内の、スケジューリング処理を行う際の情報源となるスケジューリング制御テーブル181である。

【0020】

図2は、本実施の形態の仮想計算機システムにおいて、2つの仮想計算機(1)110, (2)111上でそれぞれ動作するゲストOSにおけるプログラム動作例を示す図である。

【0021】

仮想計算機(1)110上を走行するプログラムによりタイマがセットされると、仮想計算機(1)110はハイパバイザ150へその制御を移行する。ここで、ハイパバイザ150はそのタイマ値を記憶し、仮想計算機(1)110へ制御を返す。

【0022】

仮想計算機(1)110へ制御が戻った後、プロセッサ休止要求299が呼び出されたとする。このプロセッサ休止要求299はハイパバイザ150への介入となるが、一般的に効率化のため、ハイパバイザ150は休止状態のプロセスをディスパッチの対象としない。すなわちこの場合、ハイパバイザ150は仮想計算機(2)111をディスパッチする。

【0023】

その際、ハイパバイザ150は、仮想計算機(1)110へのタイマ割り込みの報告遅延

10

20

30

40

50

が発生しないよう、タイマ設定時に記憶した仮想タイマの値などから、仮想計算機(2) 111をディスパッチすべき時間を算出して割当てて。仮想計算機(2) 111がこのディスパッチ時間を使い切ると、制御が再びハイパバイザ150に戻り、ハイパバイザ150は仮想計算機(1) 110へのタイマ割込みを起こすが、仮想計算機(2) 111へのディスパッチ時間を調整したことにより、タイマの設定時刻に正確に割込みが起こり、タイマ割込みの報告遅延が発生することはない。

【0024】

図3は、ハイパバイザ150の内部における制御動作の詳細例を示す図である。仮想計算機群100の仮想計算機110, 111, ...の処理からハイパバイザ150へ戻る契機は2つある。一つはシステム全体を制御する命令(特権命令)が発行されたとき、もう一つは仮想計算機110, 111, ...に与えられたタイムスライス時間が満了したときである。

【0025】

前者の場合は命令シミュレーション160で必要なシミュレーションを行って、ハイパバイザ150のスケジューリング処理170へ制御を渡す。後者の場合は、ハイパバイザ150のスケジューリング処理170に直接制御が渡されるので、そのままスケジューリング処理を実行する。ハイパバイザ150のスケジューリング処理170は制御を受け取ると、各種制御テーブル180内のスケジューリング制御テーブル181により次にディスパッチすべき仮想計算機とそのディスパッチ時間を直ちに算出して、その仮想計算機をディスパッチする。

【0026】

以上が、ある仮想計算機(例えば仮想計算機(1) 110)からハイパバイザ150に制御が戻ってから、次の仮想計算機(例えば仮想計算機(2) 111)をディスパッチするまでの基本的な動作であるが、命令シミュレーション160においてタイマ設定処理161が発生した場合は、上記に仮想タイマ値の更新という処理が加わる。命令シミュレーション160でタイマ設定処理161が働くと、ハイパバイザ150はスケジューリング制御テーブル181内の、該仮想計算機の仮想タイマ値を更新してから、スケジューリング処理170へ制御を渡す。ハイパバイザ150のスケジューリング処理170は制御を受け取ると、更新された仮想タイマ値を含む、スケジューリング制御テーブル181の情報をもとに次にディスパッチすべき仮想計算機とそのディスパッチ時間を算出して、その仮想計算機へディスパッチする。

【0027】

図4は、スケジューリング制御テーブル181の構成例を示す図である。ハイパバイザ150のスケジューリング処理170は、このスケジューリング制御テーブル181をもとに行われる。スケジューリング制御テーブル181には、各仮想計算機110, 111, ...に次回タイマ割込みが起こる時刻である仮想タイマ値400, 401, ...、各仮想計算機のプロセッサ使用時間の累積410, 411, ...、仮想計算機のタイムスライス値420、および物理プロセッサ内の現在時刻221から取得した現在時刻430が格納されている。

【0028】

図5は、ハイパバイザ150のスケジューリング処理170が、スケジューリング制御テーブル181を利用して、優先的にディスパッチする仮想計算機とそのディスパッチ時間を決定するアルゴリズムの一例を示すフローチャートである。ただし前提として、図2の例におけるプロセッサの休止要求が呼び出された仮想計算機のように、ディスパッチ不可能な仮想計算機はディスパッチ対象から外して処理を進めるものとする。

【0029】

処理フローは、ディスパッチ対象の決定(500, 501, 502)と、ディスパッチ時間の決定(503)の2つに分かれる。スケジューリング処理170は制御を受け取ると、まずスケジューリング制御テーブル181の各仮想計算機の仮想タイマ値400, 401, ...と現在時刻430を比較し、即時、タイマ割込みを発生させる仮想計算機が

10

20

30

40

50

あるかどうかをチェックする（５００）。すなわち、〔仮想タイマ値〕＝〔現在時刻〕である仮想計算機があるか否かのチェックを行う。

【００３０】

即時、タイマ割込みを発生させる仮想計算機がある場合（ＹＥＳ）は、その仮想計算機をディスパッチ対象とし（５０１）、ない場合（ＮＯ）はスケジューリング制御テーブル１８１の、各仮想計算機のプロセッサ使用時間の累積同士を比較し、累積時間が最も少ない仮想計算機をディスパッチ対象とする（５０２）。このように累積時間の少ない仮想計算機の優先度を上げることで、各仮想計算機へのプロセッサリソースの均等な割当てを実現することができる。

【００３１】

続いて、ディスパッチ時間の決定フローであるが、最も近い将来に発生させるタイマ割込み時刻までの現在時刻からの時間と、スケジューリング制御テーブル１８１のタイムスライス値４２０とを比較し、短い方をディスパッチ時間とする（５０３）。これにより、直近（現在時刻からタイムスライス時間以内）でタイマ割込みを発生させる仮想計算機がある場合、その一つ前にディスパッチする仮想計算機へのディスパッチ時間がタイマ割込みの発生時刻直前までとなり、タイマ割込みの前にハイパバイザへ制御が戻る。ハイパバイザはスケジューリング処理により、即時、タイマ割込みを発生させる仮想計算機を優先的にディスパッチの対象とするので、タイマの設定時刻にタイマ割込みが発生し、割込みの報告遅延を防ぐことができる。

【００３２】

以上の方式により、本実施の形態の仮想計算機システムによれば、スケジューリング処理の３つの基本要請事項である、「（１）割込みの誤差をなるべく少なく」、「（２）各仮想計算機のプロセッサ使用時間の累積を均等に」、「（３）プロセッサの空き時間を有効に使う」、が同時に満足されるスケジューリングが実現される。特に、（１）に関してはタイマ割り込みの誤差が０に、（２）に関しては累積時間の最大・最小の差が常にタイムスライス時間以内に抑えられ、（３）に関しても休止状態のプロセスをディスパッチの対象としないことで、仮想計算機へのスケジューリング処理が連続して行なわれるため、プロセッサに空き時間が発生することがない。よって、プロセッサリソースの効率的な運用が可能となり、仮想計算機の処理性能の向上が期待できる。

【００３３】

以上、本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【産業上の利用可能性】

【００３４】

本発明は、物理計算機を論理的に分割して複数の仮想計算機として使用する仮想計算機システムに適用され、特に精度の高いタイマ割込みを実現することにより、例えば時間に対して高い精度が必要なソフトウェアなど、仮想計算機システムの適応対象の拡大が可能となる。

【図面の簡単な説明】

【００３５】

【図１】本発明の一実施の形態における仮想計算機システムの構成例を示す図である。

【図２】本発明の一実施の形態における仮想計算機システムにおいて、２つの仮想計算機上でそれぞれ動作するゲストＯＳにおけるプログラム動作例を示す図である。

【図３】本発明の一実施の形態における仮想計算機システムにおいて、ハイパバイザの内部における制御動作の詳細例を示す図である。

【図４】本発明の一実施の形態における仮想計算機システムにおいて、スケジューリング制御テーブルの構成例を示す図である。

【図５】本発明の一実施の形態における仮想計算機システムにおいて、スケジューリング処理の制御動作例（優先的にディスパッチする仮想計算機とそのディスパッチ時間を決定

10

20

30

40

50

するアルゴリズム)を示すフローチャート図である。

【符号の説明】

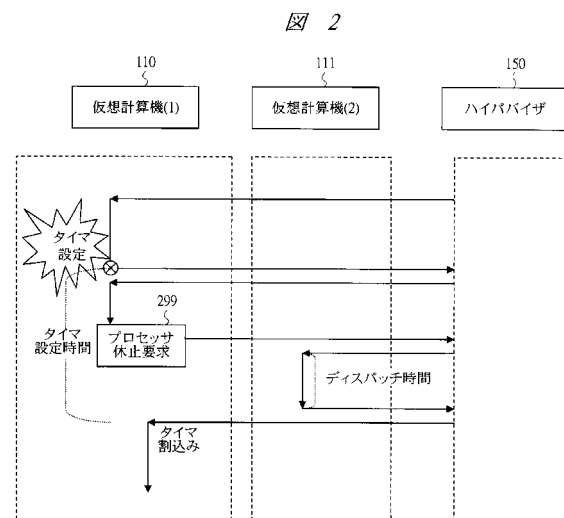
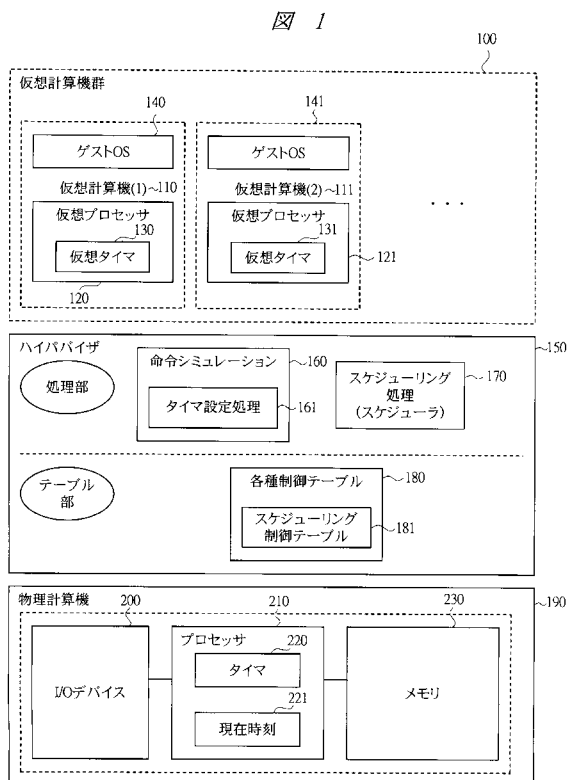
【 0 0 3 6 】

100…仮想計算機群、110、111…仮想計算機、120、121…仮想プロセッサ、130、131…仮想タイマ、140、141…ゲストオペレーティングシステム（OS）、150…ハイパバイザ、160…命令シミュレーション、161…タイマ設定処理、170…スケジューリング処理、180…各種制御テーブル、181…スケジューリング制御テーブル、190…物理計算機、200…I/Oデバイス、210…プロセッサ、220…タイマ、221…現在時刻、230…メモリ、299…プロセッサ休止要求、400、401…仮想タイマ値、410、411…プロセッサ使用時間の累積、420…タイムスライス値、430…現在時刻。

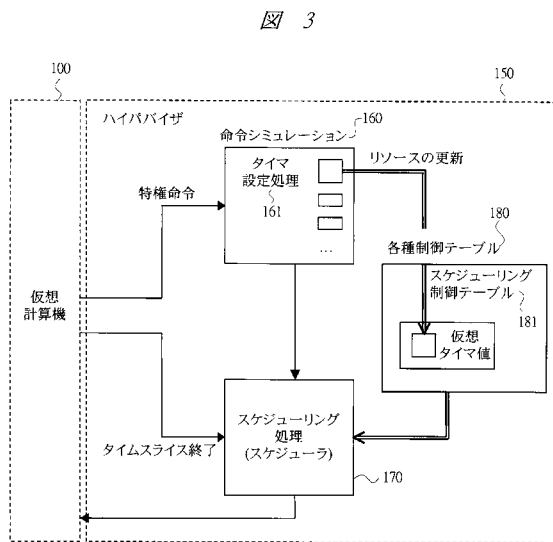
10

【 図 1 】

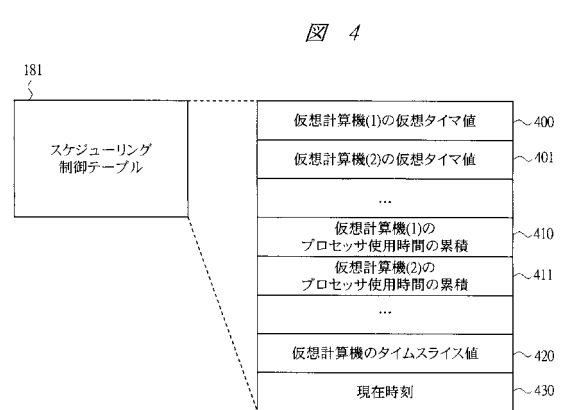
【図 2】



【図 3】



【図 4】



【図 5】

