



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 201314433 A1

(43)公開日：中華民國 102 (2013) 年 04 月 01 日

---

(21)申請案號：100135035

(22)申請日：中華民國 100 (2011) 年 09 月 28 日

(51)Int. Cl. : **G06F1/32 (2006.01)**

(71)申請人：英業達股份有限公司 (中華民國) INVENTEC CORPORATION (TW)

臺北市士林區後港街 66 號

(72)發明人：盧盈志 LU, YING CHIH (TW)

(74)代理人：詹銘文；葉璟宗

申請實體審查：有 申請專利範圍項數：10 項 圖式數：6 共 32 頁

---

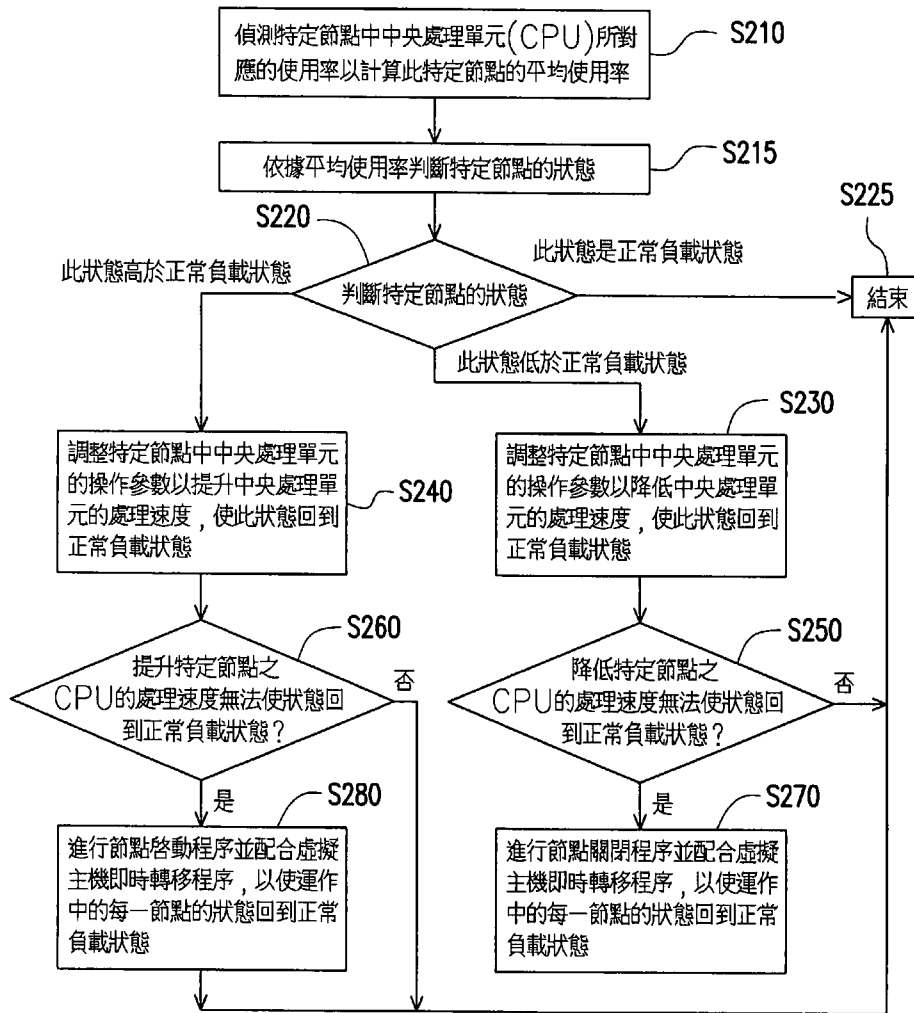
(54)名稱

伺服器系統及其電源管理方法

SERVER SYSTEM AND POWER MANAGING METHOD DATA THEREOF

(57)摘要

一種伺服器系統及其電源管理方法。伺服器系統的電源管理方法包括下列步驟：偵測一特定節點中至少一中央處理單元(CPU)所對應的一使用率以計算特定節點的平均使用率。依據平均使用率判斷此特定節點的狀態。當此狀態低於正常負載狀態時，調整 CPU 的操作參數以降低 CPU 的處理速度。以及，當此狀態高於正常負載狀態時，調整 CPU 的操作參數以提升 CPU 的處理速度，藉以使上述狀態回到正常負載狀態。本電源管理方法依據節點的中央處理單元(CPU)使用率來動態調整 CPU 之運算效能及工作週期，以在不影響虛擬主機的工作效率之下節省電源消耗。



## 發明專利說明書

(本說明書格式、順序，請勿任意更動，※記號部分請勿填寫)

※ 申請案號：100/35035

※ 申請日：

100.9.29

※IPC 分類：G06F 1/32

(2006.01)

### 一、發明名稱：

伺服器系統及其電源管理方法 / SERVER SYSTEM  
AND POWER MANAGING METHOD DATA THEREOF

### 二、中文發明摘要：

一種伺服器系統及其電源管理方法。伺服器系統的電源管理方法包括下列步驟：偵測一特定節點中至少一中央處理單元(CPU)所對應的一使用率以計算特定節點的平均使用率。依據平均使用率判斷此特定節點的狀態。當此狀態低於正常負載狀態時，調整 CPU 的操作參數以降低 CPU 的處理速度。以及，當此狀態高於正常負載狀態時，調整 CPU 的操作參數以提升 CPU 的處理速度，藉以使上述狀態回到正常負載狀態。本電源管理方法依據節點的中央處理單元(CPU)使用率來動態調整 CPU 之運算效能及工作週期，以在不影響虛擬主機的工作效率之下節省電源消耗。

### 三、英文發明摘要：

A server system and a power managing method thereof are provided. The power managing method of the server system includes the following steps. A utilization of at least

one central processing unit (CPU) in a specific node is detected to calculate an average utilization of the specific node. A state of the specific node is determined by the average utilization. When the state is lower than a normal loading state, operating parameters of the CPU is adjusted to lower an operating speed of the CPU. And, when the state is higher than the normal loading state, operating parameters of the CPU is adjusted to raise the operating speed of the CPU so that the state is returned to the normal loading state. The power managing method adjusts dynamically an operating performance and a clock duty cycle of the CPU according to the utilization of the CPU, so that power consumption is reduced without affecting efficiency of virtual machines.

#### 四、指定代表圖：

(一) 本案之指定代表圖：圖 2

(二) 本代表圖之元件符號簡單說明：

S210~S280：步驟

#### 五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無

one central processing unit (CPU) in a specific node is detected to calculate an average utilization of the specific node. A state of the specific node is determined by the average utilization. When the state is lower than a normal loading state, operating parameters of the CPU is adjusted to lower an operating speed of the CPU. And, when the state is higher than the normal loading state, operating parameters of the CPU is adjusted to raise the operating speed of the CPU so that the state is returned to the normal loading state. The power managing method adjusts dynamically an operating performance and a clock duty cycle of the CPU according to the utilization of the CPU, so that power consumption is reduced without affecting efficiency of virtual machines.

#### 四、指定代表圖：

(一) 本案之指定代表圖：圖 2

(二) 本代表圖之元件符號簡單說明：

S210~S280：步驟

#### 五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無

## 六、發明說明：

### 【發明所屬之技術領域】

本發明是有關於一種電源管理機制，且特別是有關於一種伺服器系統及其電源管理方法。

### 【先前技術】

雲端運算(cloud computing)技術是透過網際網路(Internet)結合大量的伺服器以形成超高運算與超大儲存能力。雲端運算技術可將這些伺服器的軟硬體進行共享或分工，並透過網路及瀏覽器來操作這些服務的網頁，使用上面的介面進行各種運算和工作。由於這些伺服器的數量龐大，由伺服器組成的伺服器機群(server group)通常會採用電源管理機制以降低電源消耗。

圖 1 所示，圖 1 是一種伺服器系統 100 的示意圖。在伺服器系統 100 中具有大量的節點 PM (Physical Machine)，每個節點 PM 便是伺服器系統 100 中的實體主機，而這些節點 PM 皆與網路 110 相連。並且，每個節點 PM 上皆架設有一至多個虛擬主機 VM (virtual machine)，以透過網路 110 執行雲端運算功能。管理伺服器 120 透過網路 110 來對這些節點 PM 執行電源管理。於部分實施例中，管理伺服器 120 可以是伺服器系統 100 中某個節點 PM，其可執行虛擬主機中的特定功能以進行電源管理機制。

以往所知的電源管理機制會將部份節點 PM 中負載較重的虛擬主機 VM 轉移至總量負載較輕的節點 PM，以平衡每個節點 PM 的負載，此動作亦可稱為虛擬主機 VM 的即時轉移(live migration)，藉以實現每個節點的負載平衡(load balance)。

進行即時轉移後，如果每個節點 PM 的平均總量負載仍然偏低時，便可將某些節點 PM 中的虛擬主機 VM 全數轉移至其他節點 PM 中，並關閉沒有虛擬主機 VM 在其中的節點 PM，便可減少運作中節點 PM 的數量，亦可使運作中的節點 PM 能夠在最有效的電源使用效率中執行所有的虛擬主機，上述動作可稱為節點關閉(node power-off)。相對地，如果每個節點 PM 的平均總量負載偏高而難以負荷時，便會開啟/再啟動未運作的節點 PM，並進行虛擬主機 VM 的即時轉移，以維持電源使用效率，上述動作亦可稱為可稱為節點啟動(node power-on)。

雖然上述的電源管理機制在理論上可節省較多的電源，但虛擬主機 VM 在其轉移過程中將會耗費大量的網路頻寬，因而降低整體雲端運算的效能。每個虛擬主機 VM 的負載大小也會因為使用者數量、應用軟體的種類等因素而隨時產生變化，因而難以估計這些變化的發生時點。因此，如果節點 PM 會經常性進行虛擬主機 VM 的即時轉移、節點 PM 的電源關閉與再啟動等電源管理機制時，將會大幅度降低雲端運算的效率。因此，在雲端運算技術中，如

何在具有眾多節點的伺服器系統 100 採用較佳的電源管理機制便是目前欲待深入研究的方向。

### 【發明內容】

本發明提供一種伺服器系統及其電源管理方法，其依據節點的中央處理單元(CPU)使用率來動態調整 CPU 之運算效能及工作週期的狀態，以在不影響虛擬主機的工作效率之下節省電源消耗。

本發明提供一種伺服器系統的電源管理方法，上述伺服器系統包括有多個節點。伺服器系統的電源管理方法包括下列步驟。偵測每個節點(例如，一特定節點)內中央處理單元所對應的使用率以計算此特定節點的 CPU 平均使用率。依據此平均使用率以判斷此特定節點的狀態。當此狀態低於正常負載狀態時，便調整此特定節點中的 CPU 操作參數，以降低 CPU 的處理速度，並使此狀態回到正常負載狀態而節省電源消耗。類似地，當此狀態高於正常負載狀態時，便調整此特定節點中的 CPU 操作參數以提升 CPU 的處理速度，並使此 CPU 的狀態回到正常負載狀態。

在本發明之一實施例中，伺服器系統的電源管理方法更包括下列步驟。當調升上述 CPU 的操作參數仍然無法使上述狀態回到正常負載狀態時(例如，當 CPU 的狀態已是滿載負載狀態且 CPU 的運算效能已調整至最大值並且 CPU 的工作週期已調整至最大值)，則進行一虛擬主機即時轉移程序，以平衡這些節點所承受的負載。



在本發明之一實施例中，伺服器系統的電源管理方法更包括下列步驟。當降低 CPU 的處理速度仍然無法使上述狀態回到正常負載狀態時，便進行一節點關閉程序並配合上述之虛擬主機即時轉移程序，以使運作中之每個節點的狀態回到正常負載狀態。相對地，當提升 CPU 的處理速度仍然無法使上述狀態回到正常負載狀態時，便進行一節點啟動程序並配合上述之虛擬主機即時轉移程序，以使運作中之每個節點的狀態回到正常負載狀態。

在本發明之一實施例中，判斷上述特定節點的狀態包括有下列步驟。當平均使用率高於一最大 CPU 門檻值時，將上述狀態設定為滿載負載狀態。當平均使用率高於一高度 CPU 門檻值且低於最大 CPU 門檻值時，則將上述狀態設定為高度負載狀態。當平均使用率高於一低度 CPU 門檻值且低於高度 CPU 門檻值時，則將上述狀態設定為正常負載狀態。以及，當平均使用率低於上述低度 CPU 門檻值時，將上述狀態設定為低度負載狀態。此外，當平均使用率低於一閒置 CPU 門檻值時，將上述狀態設定為閒置負載狀態，其中，上述閒置 CPU 門檻值應低於低度 CPU 門檻值。

另一方面，本發明提供一種伺服器系統，包括有多個節點及一管理伺服器。每個節點可以視作為一個實體主機，因此每個節點皆包括至少一中央處理單元(CPU)。每個節點及管理伺服器皆耦接至網路。管理伺服器可偵測每個節點(例如，一特定節點)之 CPU 的使用率以計算此特定

節點的平均使用率，並依據平均使用率來判斷出特定節點的狀態。當上述狀態低於正常負載狀態時，管理伺服器便調整特定節點中 CPU 的操作參數以降低 CPU 的處理速度。並且，當上述狀態高於正常負載狀態時，管理伺服器便調整特定節點中的 CPU 操作參數以提升 CPU 的處理速度，以使上述狀態回到正常負載狀態。此外，伺服器系統的其餘實施細節請參照上述說明，在此不加贅述。

基於上述，由於以往的電源管理機制在執行過程中會佔據較多系統資源、也耗費較多的處理時間。藉此，本發明實施例可依據每個節點的整體負載(或是，各節點之中央處理單元(CPU)的平均使用率)來動態調整中央處理單元(CPU)的運轉速度，使得各個節點能夠在正常負載狀態中進行運轉，以在不影響各個節點的工作效率、不佔據系統資源之下節省電源消耗。

藉此，本發明實施例可避免經常性地執行虛擬主機的即時轉移及節點關閉/節點啟動，免除了雲端運算效率的降低。當調整 CPU 的運作頻率(P)及工作週期(T)仍無法使伺服器系統運作在較佳的耗電效率時，再執行以往的電源管理機制，便可達到最佳的電源管理功效。

為讓本發明之上述特徵和優點能更明顯易懂，下文特舉實施例，並配合所附圖式作詳細說明如下。

## 【實施方式】

現將詳細參考本發明之示範性實施例，在附圖中說明所述示範性實施例之實例。另外，凡可能之處，在圖式及實施方式中使用相同標號的元件/構件/符號代表相同或類似部分。

請參照圖 2，圖 2 是根據本發明一實施例說明伺服器系統 300 的電源管理方法的流程圖。圖 3 是根據本發明一實施例說明伺服器系統 300 的示意圖。於本實施例中，在此以圖 3 所示的伺服器系統 300 來執行圖 2 繪示的伺服器系統 300 電源管理方法。

伺服器系統 300 具有大量的節點 PM，每個節點 PM 可以是伺服器系統 300 中的實體主機。這些節點 PM 及管理伺服器 320 皆與網路 310 相連。每個節點 PM 上皆具備有一至多個虛擬主機(virtual machine)VM、多個中央處理單元(CPU)以及基本輸入輸出系統(BIOS) 330，以透過網路 310 執行雲端運算功能。管理伺服器 320 則可透過網路 310 來對這些節點 PM 執行電源管理。於部分實施例中，管理伺服器 320 可以是伺服器系統 300 中某個節點 PM，或是在節點 PM 中其中一個虛擬主機 VM 的電源管理機制。

在此說明的是，此處所指的實體主機 PM 可以有多個 CPU 插槽，因此可以架設許多具有多核心(core)CPU，例如在實體主機 PM 中架設多個四核心 CPU。每個 CPU 當中的每個核心也可以包含有一個或多個執行緒(thread)。因此，本實施例將多核心 CPU 稱為實體 CPU，而將每個實體 CPU 中的每一個核心稱為核心 CPU。並且，每個核心 CPU 中

也可以包括有多個執行緒，而每一個 CPU 執行緒稱為邏輯 CPU。藉此，在本實施例以及圖 1 及圖 3 中所提及的 CPU 可以是實體 CPU，可以是核心 CPU，也可以是邏輯 CPU。

在此說明伺服器系統 300 的電源管理方法，於本實施例中，此電源管理方法可以每經過一段時間(例如，60 秒)便執行一次。為了簡化說明，在此將其中一個節點 PM 稱為是一特定節點。因此，於步驟 S210 中，管理伺服器 320 偵測特定節點內 CPU 所對應的使用率，並藉此來計算此特定節點的平均使用率。CPU 所對應的使用率係指每個 CPU 當中的執行緒在執行指令的負載程度，因此每個 CPU 使用率則可以將這個 CPU 中每個執行緒的負載程度進行平均以計算求得。最高的 CPU 使用率便是 100%(此時 CPU 中所有的執行緒應處於滿載狀態)，而最低的 CPU 使用率則為 0%(此時 CPU 中所有的執行緒幾乎處於閒置狀態)。

如同上述，由於每個實體節點 PM 上皆會包括有多個邏輯 CPU 或邏輯 CPU。因此，在計算特定節點中 CPU 的平均使用率時，便會先行偵測出特定節點中每個 CPU 所對應的使用率。接著，將這些 CPU 的使用率進行加總並除以特定節點上 CPU 的數量，以計算得到特定節點上的 CPU 平均使用率。

特別提及的是，於本實施例中，管理伺服器 320 對於各個節點 PM 的 CPU 使用率偵測、CPU 操作參數偵測及控制，皆可以採用基本輸入輸出系統(BIOS)的軟體(Software ; SW)系統管理中斷(System management

interrupt；SMI)機制配合各個節點 PM 中的晶片組進行管理，而不需要透過基板管理控制器(BMC)來進行此電源管理方法，因此可降低節點 PM 的架設成本。當然，於其他實施例中，管理伺服器 320 也可以利用節點 PM 中的基板管理控制器來進行 CPU 使用率的偵測、操作參數的偵測、管理與控制，本發明並不限制於此。

在此詳細說明本實施例如何利用 BIOS 中的 SMI 處理程序，藉以取得節點 PM 中各個 CPU 的使用率、CPU 操作參數。圖 4 為特定節點 PM 中 BIOS 330、晶片組 410 與其中一個實體 CPU 的示意圖。為了方便說明，本實施例中所述的實體 CPU 例如是具有 N 個執行緒的邏輯 CPU，N 為一正整數。每個執行緒 CT1~CTN 都內含一組暫存器，為便於說明，在此以 CT1~CTN 來分別表示位在執行緒 CT1~CTN 內的該組暫存器。於本實施例中，每個節點 PM 可透過其中的晶片組(chipset) 410 來產生 SMI 訊號。

舉例來說，晶片組 410 包括有表單指向暫存器 412、SMI 指令輸出/輸入埠(SMI command I/O Port) 414 及 SMI 資料輸出/輸入埠(SMI Data I/O Port)415。於本實施例中，SMI 指令輸出/輸入埠 414 及 SMI 資料輸出/輸入埠 415 是透過暫存器來實現。SMI 指令輸出/輸入埠 414 及 SMI 資料輸出/輸入埠 415 用以進行 SW SMI 機制的 SMI 指令及資料存取，而表單指向暫存器 412 儲存有一記憶體位址，藉以指向至記憶體(memory buffer)413 中諸多表單的存放處，讓 BIOS 330 之 SMI 處理程式(SMI handler)得以參考

下述之 CPU 目前使用率狀態表、CPU 之運算效能 P 及工作週期 T 範圍表、CPU 之運算效能 P 及工作週期 T 操作參數表等表單，藉以提供操作之所需之輸入、輸出參數。在實作中，假設 SMI 指令輸出/輸入埠 414 的位址為 B2 且 SMI 資料輸出/輸入埠 415 之位址為 B3，且定義 F0 值為 SW SMI 機制之 CPU 執行緒的管理命令。

因此，請參照圖 4，當晶片組 410 將 F0 值寫入 SMI 指令輸出/輸入埠 414 的位址 B2 時，B2 將產生一個 SMI 訊號至實體 CPU，實體 CPU 將會立刻跳至 BIOS 330 中對應的 SMI 處理程序(SMI handler)以處理此 SMI 訊號。詳言之，實體 CPU 將會讀取 SMI 指令輸出/輸入埠 414 的位址 B2 中的值而得 F0 值，此時實體 CPU 將會執行 CPU 執行緒的管理程式，此程式會進一步讀取 SMI 資料輸出/輸入埠 415 的位址 B3 中的值以便進一步得知要做何工作。於此處，本實施例定義 SMI 資料輸出/輸入埠 415 的位址 B3 中的值如下：

0--從 CPU 暫存器所取得的目前的 CPU 使用率。

1--從 CPU 暫存器所取得的 CPU 運算效能 P、工作週期 T 的最大值與最小值。

2--從 CPU 暫存器所取得目前的運算效能 P、工作週期 T 之值。

3--對 CPU 暫存器所設定的運算效能 P、工作週期 T 之值。

因此，CPU 執行緒的管理程式將會有如上之工作要進行，但是這些工作所需之輸入、輸出參數可由表單指向暫存器 412 所指向之記憶體(memory buffer)413 中諸多表單取得。

當晶片組 410 發出 SMI 訊號至 CPU 後，CPU 中的執行緒都會收到此 SMI 訊號，然後每個執行緒便會到 BIOS 330 中讀取各自對應的 SMI 處理程序來執行，藉以取得對應 CPU 的使用率，並依照操作參數控制對應的 CPU。

為了便於說明，在此以實體 CPU 中一執行緒對應的一組暫存器 CT1 作為舉例，而圖 4 省略了其他暫存器組(例如，暫存器組 CT2~CTN)當中的架構，因為其架構與暫存器組 CT1 應為相同。於本實施例中，每個暫存器組 CT1~CTN 包括有系統管理模式基底暫存器(system management mode base register) 450；用以控制及暫存運算效能 P 之現狀、P 最大值及 P 最小值的 P 現狀暫存器 421、P 最大值暫存器 422 及 P 最小值暫存器 423；以及，用以控制及暫存工作週期 T 之現狀、T 最大值及 T 最小值的 T 現狀暫存器 431、T 最大值暫存器 432 及 T 最小值暫存器 433。SMM(system management mode)基底暫存器 450 用以暫存 BIOS 330 中的位址起始點，藉以儲存各個執行緒中所需執行之 SMI 處理程序的進入點(entry)。此外，暫存器組 CT1~CTN 更包括可暫存此執行緒使用率的 CPU 使用率暫存器 460。

透過上述架構，CPU 中的各個執行緒在接收到 SMI 訊號後，便利用 SMM(system management mode)基底暫存器 450 中所儲存的位址起始點以跳至 SMI 處理程序的進入點，藉以開始執行各個執行緒對應的 SMI 處理程序，並將在 SMI 處理程序中所取得的相關資訊暫存及控制相對應的暫存器 421~460 中。以目前的實施例流程而言，是利用 CPU 當中其中一個稱作系統啟動帶狀處理器(system bootstrap processor；SBSP)的執行緒會執行正常的 SMI 處理程序，其他執行緒(稱作應用處理器(application processor))透過訊號旗幟(semaphore)方式來實現與 SBSP 之間的同步動作，再藉由執行 RSM 指令來一起結束 SMI 處理程序，並使 CPU 離開 System Management 模式(SMM)。此 CPU 的 SMI 機制也就是上述如何利用 BIOS 中的 SMI 介面取得節點 PM 中各個 CPU 的使用率、CPU 操作參數及使用操作參數控制 CPU 之 P 及 T 之操作的相關動作。

另一方面，上述之 SMI 處理程序的流程略述如下。首先，此 SMI 處理程序會先分析由晶片組 410 傳送來的 SMI 訊號，以判斷所需執行的動作。舉例而言，本實施例將 SMI 訊號分為以下四種：

指令 0--取得 CPU 目前使用率。

指令 1--取得 CPU 運算效能 P、工作週期 T 之最大值及最小值的狀態資訊。

指令 2--取得 CPU 目前的運算效能 P、工作週期 T 的狀態資訊。



指令 3--設定 CPU 現在的運算效能 P、工作週期 T。

在分析完 SMI 訊號之後，SMI 處理程序會透過晶片組 410 中表單指向暫存器 412 所指出的各種表單，並依據上述指令來執行相對應的動作。例如，透過這些表單來得知其指令之輸出及輸入參數為何。

回到圖 2 及圖 3，於步驟 S215 中，管理伺服器 320 接著便依據由步驟 S210 計算得到的平均使用率來判斷此特定節點的負載狀態。於本實施例中，管理伺服器 320 可利用內建的 CPU 平均使用率狀態表來判斷每一個節點目前的負載狀態，而此 CPU 平均使用率狀態表亦可由使用者進行調整與設定。

圖 5 為管理伺服器 320 中的 CPU 平均使用率狀態表。如圖 5 所示，由於 CPU 使用率最高便是 100%，因此當特定節點的平均使用率高於最大 CPU 門檻值(例如，95%)且小於等於 100%時，管理伺服器 320 便將特定節點的狀態設定為滿載負載狀態，換句話說，將特定節點的狀態碼設定為 5。當平均使用率高於高度 CPU 門檻值(例如，80%)且低於最大 CPU 門檻值(95%)時，則將此特定節點的狀態設定為高度負載狀態(設定特定節點的狀態碼為 4)。

依此類推，當平均使用率高於低度 CPU 門檻值(例如，40%)且低於高度 CPU 門檻值(80%)時，表示此特定節點位於較佳的電源消耗功率中，而特定節點上的虛擬主機 VM 皆可在不影響其效能下正常執行，因此則將此特定節點的狀態設定為正常負載狀態(設定特定節點的狀態碼為

3)。以及，當平均使用率分別低於低度 CPU 門檻值(40%)或甚至低於閒置 CPU 門檻值(10%)時，則將此特定節點的狀態分別設定為低度負載狀態(設定特定節點的狀態碼為 2)或是閒置負載狀態(設定特定節點的狀態碼為 1)。

回到圖 2，接著進入步驟 S220，管理伺服器 320 分別判斷每個節點 PM 的負載狀態。如果每個節點 PM 的負載狀態皆為正常負載狀態，便可直接進入步驟 S225 以結束本次的電源管理方法。當特定節點的負載狀態低於正常負載狀態(狀態碼為 3)的時候(例如，當特定節點位於低度負載狀態(狀態碼為 2)或閒置負載狀態(狀態碼為 1))，便由步驟 S220 進入步驟 S230，管理伺服器 320 便調整此特定節點中的 CPU 操作參數(例如，CPU 的運算效能 P 及工作週期(duty cycle)T)以降低 CPU 的處理速度，並期望使此狀態回到正常負載狀態，藉以節省此特定節點中的電源消耗。

相對而言，當特定節點的負載狀態高於正常負載狀態(狀態碼為 3)的時候(例如，當特定節點位於滿載負載狀態(狀態碼為 5)或高度負載狀態(狀態碼為 4))，便由步驟 S220 進入步驟 S240，管理伺服器 320 依此調整特定節點中的 CPU 操作參數(例如，CPU 的運算效能 P 或工作週期(duty cycle)T)以提高 CPU 的處理速度，並期望使此狀態回到正常負載狀態。

在此說明本實施例如何利用調整特定節點中的 CPU 操作參數來降低/提高 CPU 處理速度。於本實施例中，CPU 處理速度的調整是管理伺服器 320 利用 BIOS 的 SMI 介

面，並對每個 CPU 的運算效能 P 及工作週期 T 利用一 CPU 操作參數表來進行調整，圖 6 為本發明實施例所述之 CPU 操作參數表。

於圖 6 中，CPU 操作參數碼用以設定每個 CPU 的運算效能 P 及工作週期 T。運算效能 P 及工作週期 T 在圖 4 中則是以百分比作為舉例，例如，當某一 CPU 的操作參數碼被設定為 1 時，表示此 CPU 的運算效能 P 被設定為最大值(100%)，而此 CPU 的工作週期 T 亦被設定為最大值(100%)，以使 CPU 能夠以最大額度的處理速度來處理指令。相對而言，當某一 CPU 的操作參數碼被設定為 8 時，表示此 CPU 的運算效能 P 則被設定為最大值的 25%，而此 CPU 的工作週期 T 亦被設定為最大值的 12.5%，以使 CPU 能夠以最低限度的處理速度來處理指令。在此特別說明的是，應用本實施例者可依據其需求來調整圖 6 之 CPU 操作參數表中的數據，本發明不應以此為限。

在此利用圖 6 並舉例說明步驟 S240。假設特定節點的負載狀態為高度負載狀態(狀態碼為 4)，且特定節點中每一個 CPU 的操作參數碼皆設定為 4，亦即特定節點中每一個 CPU 的運算效能 P 設定為最大值的 70%，其工作週期 T 亦被設定為最大值的 62.5%。因為特定節點的負載狀態(高度負載狀態的狀態碼為 4)高於正常負載狀態(狀態碼為 3)，因此於步驟 S240 中，管理伺服器 320 首先取得特定節點中先前設定的 CPU 操作參數碼(4)，並且減少特定節點

中 CPU 操作參數碼的數值(例如將 CPU 操作參數碼由 4 減少至 3)，以提升 CPU 的處理速度。

相對來說，在此利用圖 6 並舉例說明步驟 S230。假設特定節點的負載狀態為低度負載狀態(狀態碼為 2)，且特定節點中每一個 CPU 的操作參數碼亦設定為 4。因為特定節點的負載狀態(低度負載狀態的狀態碼為 2)低於正常負載狀態(狀態碼為 3)，因此於步驟 S230 中，管理伺服器 320 首先取得特定節點中先前設定的 CPU 操作參數碼(4)，並且降低特定節點中 CPU 操作參數碼的數值(例如將 CPU 操作參數碼由 4 提升至 5)，以降低 CPU 的處理速度，並期望特定節點的負載狀態能夠藉此回到正常負載狀態。

有鑒於此，本發明實施例的電源管理方法可基於時間相依(time locality)原則，依據此次在每個節點中偵測得到的平均 CPU 使用率來預測出下次每個節點的負載狀態，進而對 CPU 的操作參數進行調整，以期望下次每個節點的平均使用率皆能回到正常負載狀態中。

但是，如果特定節點的負載狀態已經為滿載負載狀態(狀態碼為 5)，且此特定節點中每一個 CPU 的操作參數碼已被設定為 1，亦即每個 CPU 皆位於滿載狀態，此時上述的電源管理方法已無法進行更多的調整來使每個節點回到正常負載狀態。或是，如果特定節點的負載狀態已經為閒置負載狀態(狀態碼為 1)，且此特定節點中每一個 CPU 的操作參數碼已被設定為 8，亦即特定節點中每個 CPU 幾乎位於閒置狀態，此時上述的電源管理方法已無法進行更多

的調整來使每個節點回到正常負載狀態。因此，本發明實施例便結合以往的電源管理機制來完善此電源管理方法。

請繼續參照圖 2，於步驟 S250 中，當特定節點的負載狀態已經為閒置負載狀態(狀態碼為 1)，且此特定節點中每一個 CPU 的處理速度已降至最低時(亦即，CPU 的工作參數碼已被設定為 8)，則表示降低特定節點之 CPU 的處理速度已經無法使特定節點的負載狀態回到正常負載狀態時，則進入步驟 S270，管理伺服器 320 進行虛擬主機 VM 的即時轉移程序，以平衡這些節點 PM 所承受的負載。

如果虛擬主機 VM 的即時轉移程序仍然無法使特定節點的負載狀態回到正常負載狀態時，則進行節點關閉程序，配合虛擬主機 VM 的即時轉移程序，將某些節點 PM 中的虛擬主機 VM 全數轉移至其他節點 PM 中，並關閉已沒有虛擬主機 VM 在其中的節點 PM，以減少運作中節點 PM 的數量。相對而言，在步驟 S250 中，如果降低特定節點之 CPU 的處理速度可以使狀態回到正常負載狀態時，便可從步驟 S250 回到步驟 S225 以結束本次電源管理方法。而在執行步驟 S270 完畢之後，亦回到步驟 S225 以結束本次電源管理方法。

請繼續參考圖 2，於步驟 S260 中，當特定節點的負載狀態已經為滿載負載狀態(狀態碼為 5)，且此特定節點中每一個 CPU 的處理速度已提升至最高時(亦即，CPU 的工作參數碼已被設定為 1)，則進入步驟 S280，管理伺服器 320

便進行虛擬主機 VM 的即時轉移程序，以平衡這些節點 PM 所承受的負載。

如果虛擬主機 VM 的即時轉移程序仍然無法使特定節點的負載狀態回到正常負載狀態時，則進行節點啟動程序以開啟/再啟動未運作的節點 PM，並配合虛擬主機 VM 的即時轉移程序將虛擬主機 VM 依據其負載程度進行節點 PM 之間的轉移，以在不影響虛擬主機 VM 的工作效率之下來節省電源。並且，類似於步驟 S250 及 S270，在步驟 S260 中，如果降低特定節點之 CPU 的處理速度可以使狀態回到正常負載狀態時，便可從步驟 S260 回到步驟 S225 以結束本次電源管理方法。而在執行步驟 S280 完畢之後，亦回到步驟 S225 以結束本次電源管理方法。

綜上所述，由於以往的電源管理機制在執行過程中會佔據較多系統資源、也耗費較多的處理時間。藉此，本發明實施例可依據每個節點的整體負載(亦即各節點之中央處理單元(CPU)的使用率)來動態調整中央處理單元(CPU)的運轉速度，使得各個 CPU 能夠在正常負載狀態中進行運轉，以在不影響虛擬主機的工作效率、不佔據系統資源之下節省電源消耗。

藉此，本發明實施例可避免經常性地執行虛擬主機的即時轉移及節點關閉/節點啟動，免除了雲端運算效率的降低。當調整 CPU 轉速仍無法使伺服器系統運作在較佳的耗電效率時，再透過以往的電源管理機制以達到最佳的電源管理功效。

雖然本發明已以實施例揭露如上，然其並非用以限定本發明，任何所屬技術領域中具有通常知識者，在不脫離本發明之精神和範圍內，當可作些許之更動與潤飾，故本發明之保護範圍當視後附之申請專利範圍所界定者為準。

### 【圖式簡單說明】

圖 1 是一種伺服器系統的示意圖。

圖 2 是根據本發明一實施例說明伺服器系統的電源管理方法的流程圖。

圖 3 是根據本發明一實施例說明伺服器系統的示意圖。

圖 4 為特定節點中 BIOS、晶片組與其中一個實體 CPU 的示意圖。

圖 5 為管理伺服器中的 CPU 平均使用率狀態表。

圖 6 為本發明實施例所述之 CPU 操作參數表。

### 【主要元件符號說明】

100、300：伺服器系統

110、310：網路

120、320：管理伺服器

330：基本輸入輸出系統(BIOS)

410：晶片組

412、414~415、421~460：暫存器

413：記憶體

CT1~CTN：CPU 執行緒內含的該組暫存器

CPU：中央處理單元

PM：節點/實體主機

VM：虛擬主機

S210~S280：步驟

P：CPU 的運算效能

T：CPU 的工作週期



## 七、申請專利範圍：

1. 一種伺服器系統的電源管理方法，其中該伺服器系統包括多個節點，該伺服器系統的電源管理方法包括下列步驟：

偵測一特定節點中至少一中央處理單元(CPU)所對應的至少一使用率以計算該特定節點的一平均使用率，其中該特定節點是該些節點之其一；

依據該平均使用率判斷該特定節點的一狀態；

當該狀態低於一正常負載狀態時，調整該特定節點中該中央處理單元的一操作參數以降低該中央處理單元的處理速度，以使該狀態回到該正常負載狀態；以及

當該狀態高於該正常負載狀態時，調整該操作參數以提升該中央處理單元的處理速度，以使該狀態回到該正常負載狀態。

2. 如申請專利範圍第 1 項所述之伺服器系統的電源管理方法，更包括下列步驟：

當調整該操作參數無法使該狀態回到該正常負載狀態時，進行一虛擬主機即時轉移程序，以平衡該些節點的負載。

3. 如申請專利範圍第 2 項所述之伺服器系統的電源管理方法，更包括下列步驟：

當降低該中央處理單元的處理速度無法使該狀態回到該正常負載狀態時，進行一節點關閉程序並配合該虛擬主

機即時轉移程序，以使運作中的每一節點的該狀態回到該正常負載狀態。

4. 如申請專利範圍第 2 項所述之伺服器系統的電源管理方法，更包括下列步驟：

當提升該中央處理單元的處理速度無法使該狀態回到該正常負載狀態時，進行一節點啟動程序並配合該虛擬主機即時轉移程序，以使運作中的每一節點的該狀態回到該正常負載狀態。

5. 如申請專利範圍第 1 項所述之伺服器系統的電源管理方法，判斷該特定節點的該狀態包括下列步驟：

當該平均使用率高於一最大 CPU 門檻值時，將該狀態設定為一滿載負載狀態；

當該平均使用率高於一高度 CPU 門檻值且低於該最大 CPU 門檻值時，將該狀態設定為一高度負載狀態；

當該平均使用率高於一低度 CPU 門檻值且低於該高度 CPU 門檻值時，將該狀態設定為該正常負載狀態；

當該平均使用率低於該低度 CPU 門檻值時，將該 CPU 的該狀態設定為該低度負載狀態，

其中，該滿載負載狀態及該高度負載狀態高於該正常負載狀態，且該低度負載狀態低於該正常負載狀態。

6. 如申請專利範圍第 5 項所述之伺服器系統的電源管理方法，判斷該特定節點的該狀態更包括下列步驟：

當該平均使用率低於一閒置 CPU 門檻值時，將該狀態設定為該閒置負載狀態，其中該閒置 CPU 門檻值低於該低度 CPU 門檻值，且該閒置負載狀態低於該正常負載狀態。

7. 如申請專利範圍第 1 項所述之伺服器系統的電源管理方法，其中該中央處理單元的該操作參數為該中央處理單元的一運算效能或一工作週期。

8. 一種伺服器系統，包括：

多個節點，每一節點耦接至一網路，且每一節點包括至少一中央處理單元；

一管理伺服器，耦接至該網路，該管理伺服器偵測一特定節點中該中央處理單元所對應的一使用率以計算該特定節點的一平均使用率，依據該平均使用率判斷該特定節點的一狀態，當該狀態低於一正常負載狀態時，調整該特定節點中該中央處理單元的一操作參數以降低該中央處理單元的處理速度，且當該狀態高於該正常負載狀態時，調整該操作參數以提升該中央處理單元的處理速度，以使該狀態回到該正常負載狀態，其中該特定節點為該些節點之一。

9. 如申請專利範圍第 8 項所述之伺服器系統，當降低該中央處理單元的處理速度無法使該狀態回到該正常負載狀態時，該管理伺服器進行一虛擬主機即時轉移程序及一節點關閉程序，以使運作中的每一節點的該狀態回到該正常負載狀態。

10. 如申請專利範圍第 8 項所述之伺服器系統，當提升該中央處理單元的處理速度無法使該狀態回到該正常負載狀態時，該管理伺服器進行一節點啟動程序並配合一虛擬主機即時轉移程序，以使運作中的每一節點的該狀態回到該正常負載狀態。

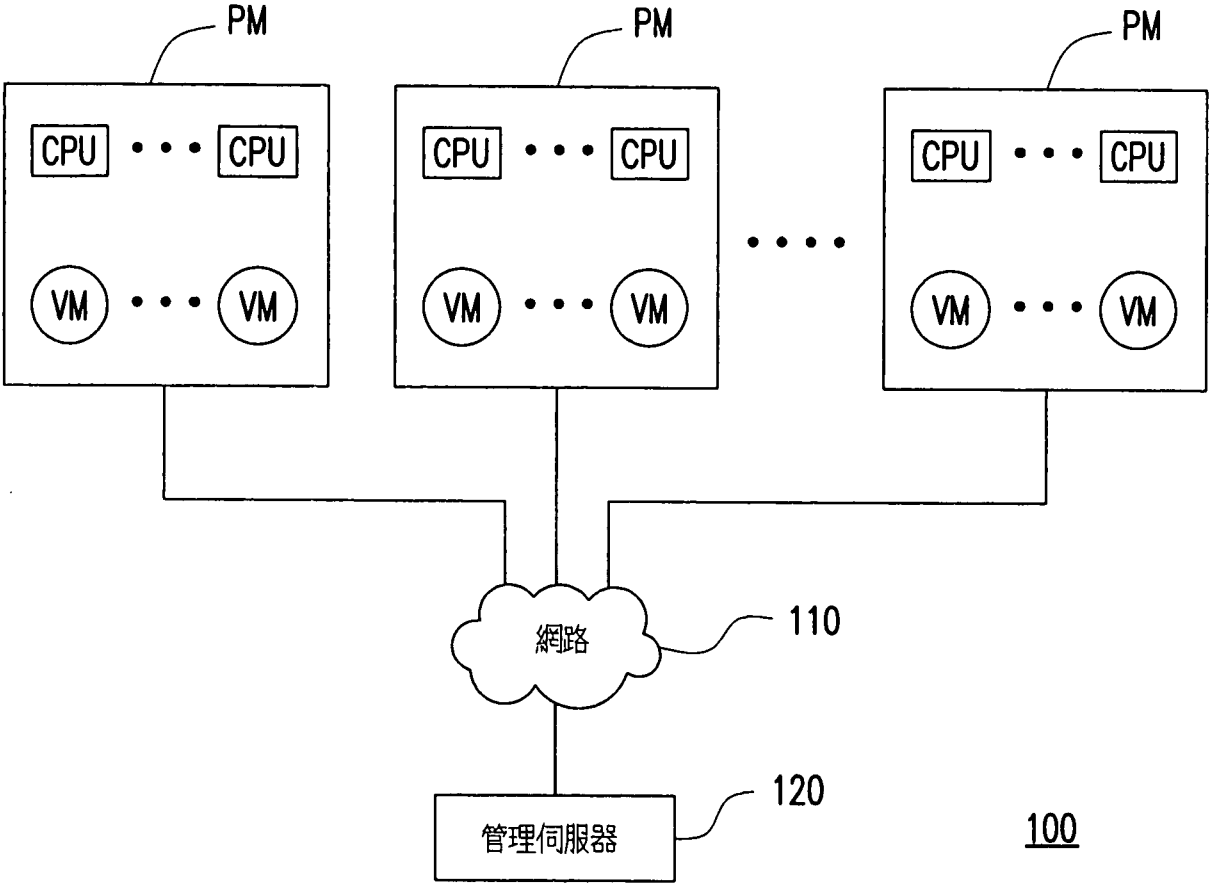


圖 1

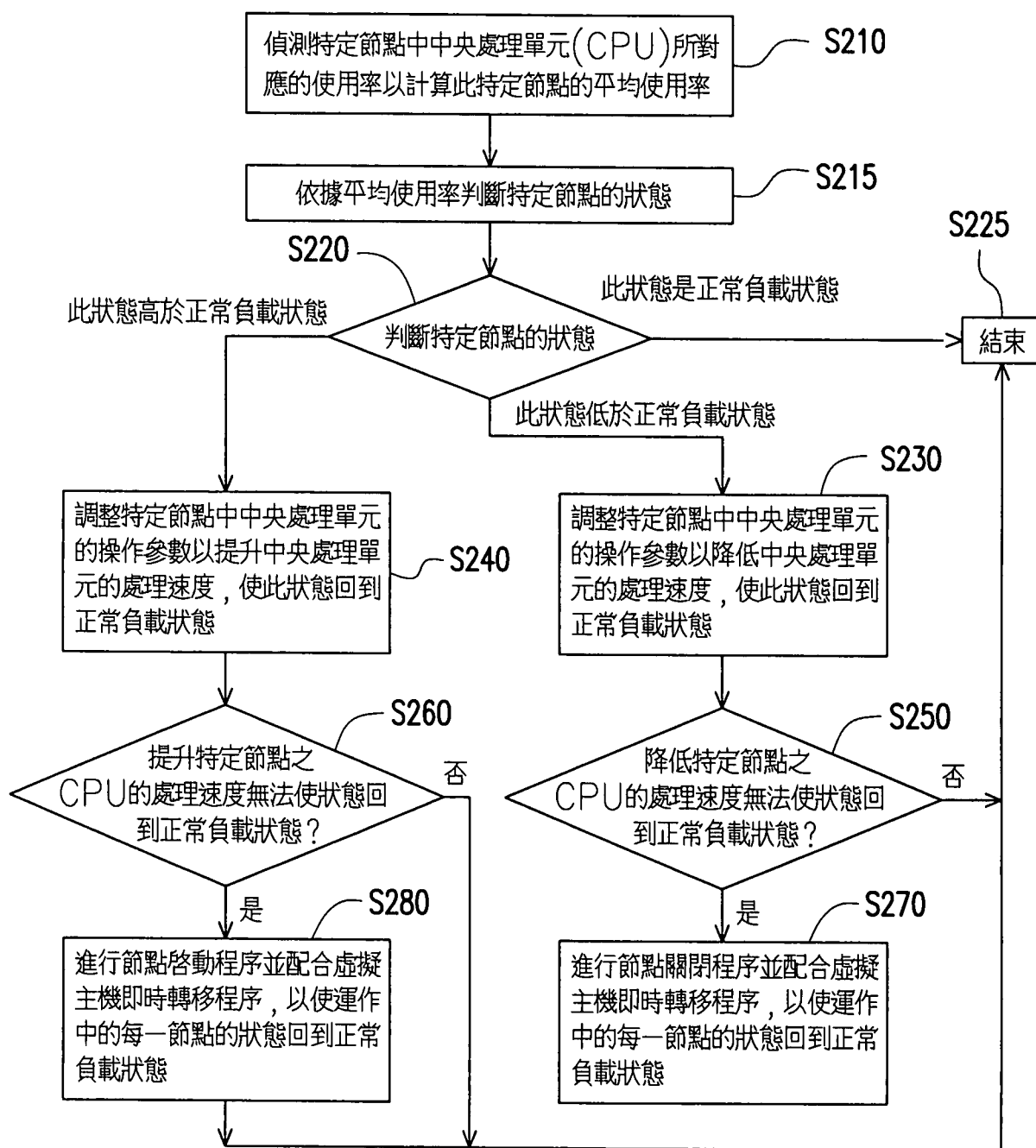


圖 2

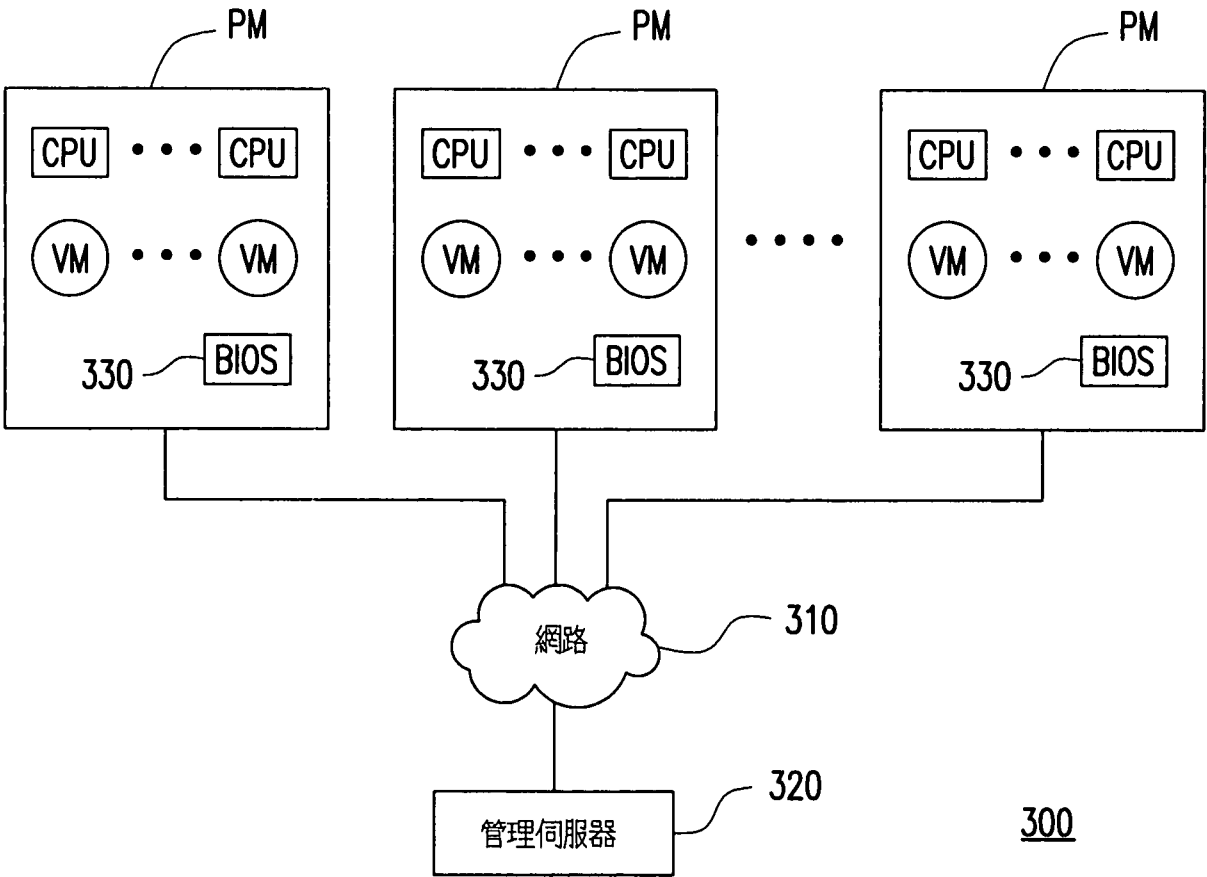


圖 3

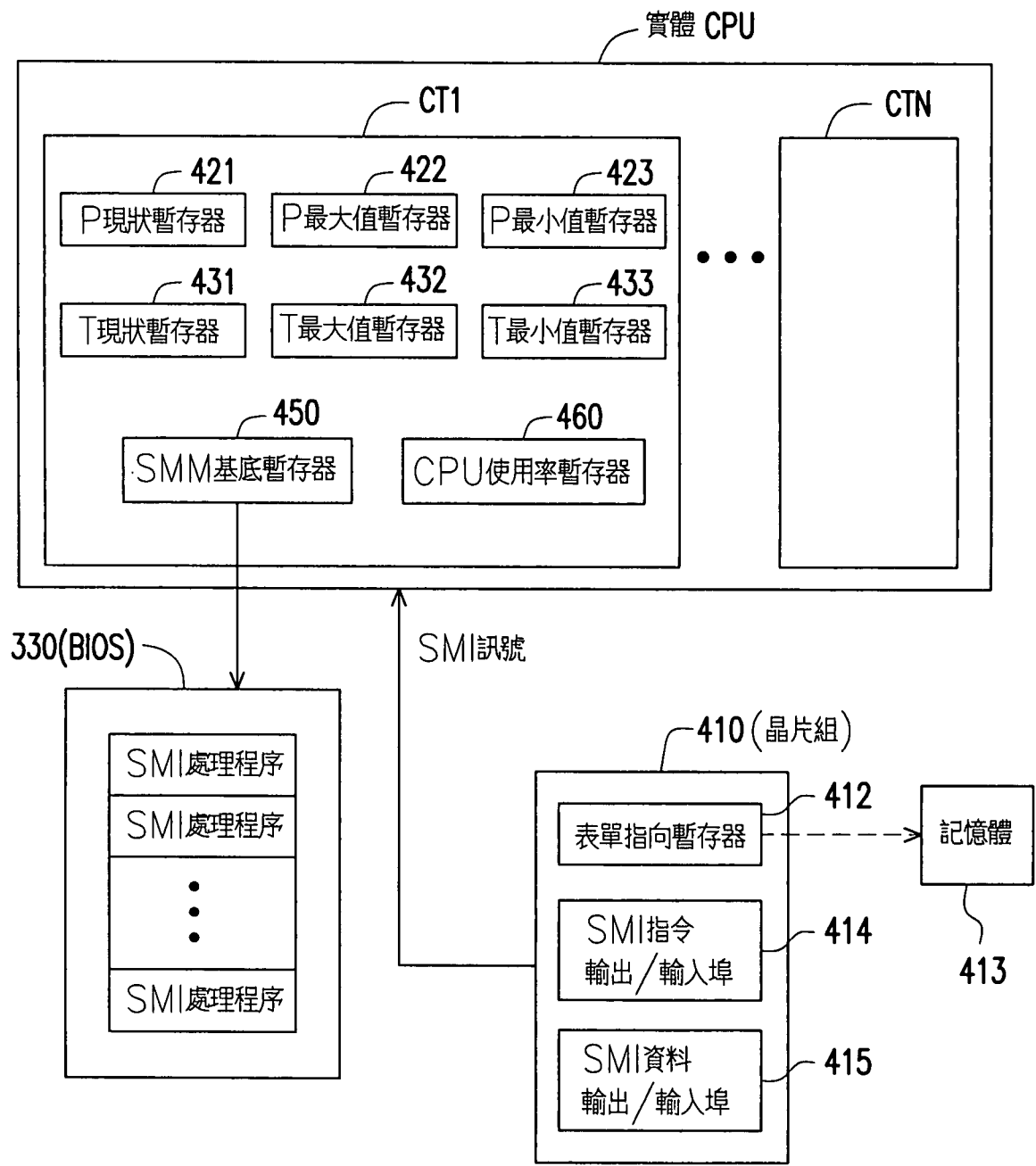


圖 4



| 平均使用率      | 特定節點的負載狀態 | 狀態碼 |
|------------|-----------|-----|
| 100% ~ 95% | 滿載負載狀態    | 5   |
| 95% ~ 80%  | 高度負載狀態    | 4   |
| 80% ~ 40%  | 正常負載狀態    | 3   |
| 40% ~ 10%  | 低度負載狀態    | 2   |
| 10% ~ 0%   | 閒置負載狀態    | 1   |

圖 5

| 操作參數碼 | 運算頻率 P | 工作週期 T |
|-------|--------|--------|
| 1     | 100%   | 100%   |
| 2     | 90%    | 87.5%  |
| 3     | 80%    | 75%    |
| 4     | 70%    | 62.5%  |
| 5     | 60%    | 50%    |
| 6     | 50%    | 37.5%  |
| 7     | 30%    | 25%    |
| 8     | 25%    | 12.5%  |

圖 6