



(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 201333839 A1

(43)公開日：中華民國 102 (2013) 年 08 月 16 日

(21)申請案號：101103616

(22)申請日：中華民國 101 (2012) 年 02 月 03 日

(51)Int. Cl.：

G06Q10/00 (2012.01)

G06F1/32 (2006.01)

G06F15/16 (2006.01)

(71)申請人：國立交通大學(中華民國) NATIONAL CHIAO TUNG UNIVERSITY (TW)

新竹市大學路 1001 號

(72)發明人：林盈達 LIN, YING DAR (TW)；黃霆鈞 HUANG, TING JUN (TW)；賴源正 LAI,

YUAN CHENG (TW)；朱宗賢 CHU, TSUNG HSIEN (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：12 項 圖式數：3 共 27 頁

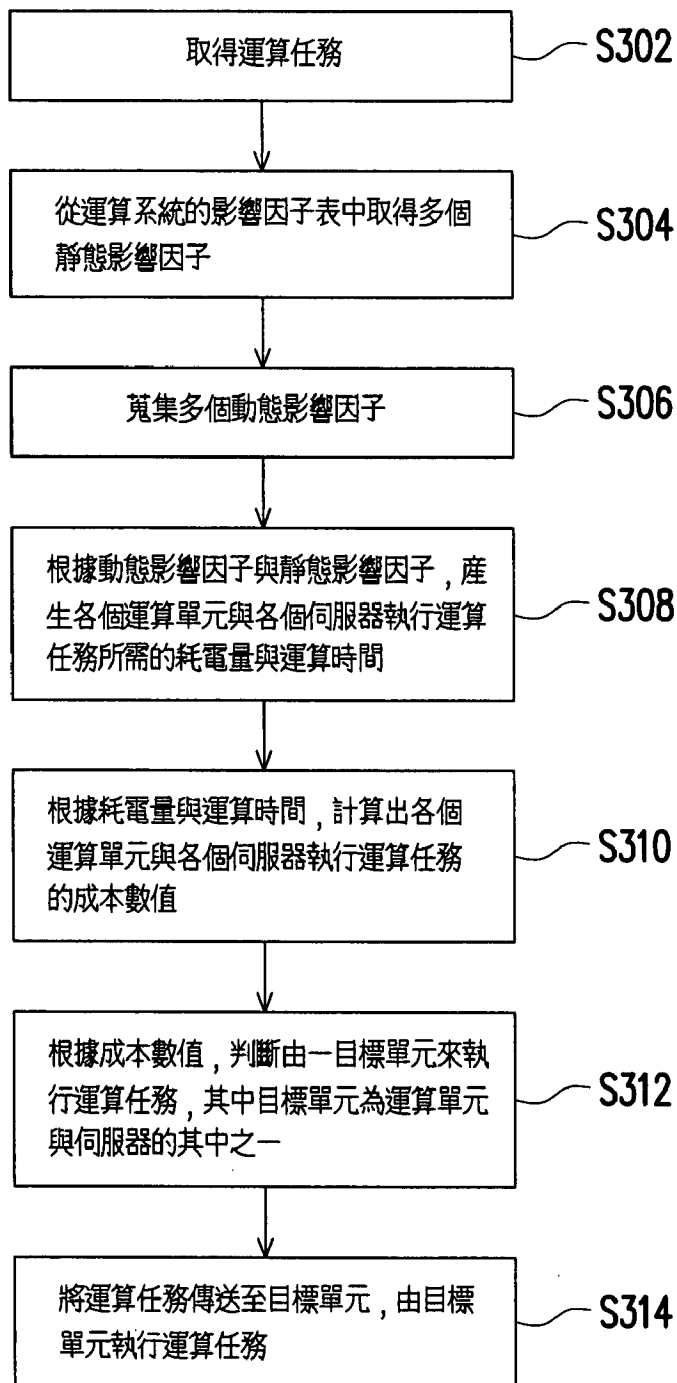
(54)名稱

考慮耗時與耗電的卸載運算量的決策方法與運算系統

DECISION METHOD CONSIDERING TIME AND POWER CONSUMPTION FOR OFFLOADING
COMPUTATION AND COMPUTING SYSTEM

(57)摘要

一種考慮耗時與耗電的卸載運算量的決策方法。此方法包括：取得運算任務；從影響因子表中取得多個靜態影響因子；蒐集多個動態影響因子。此方法還包括：根據動態影響因子與靜態影響因子，產生各個運算單元與伺服器執行運算任務所需的耗電量與運算時間；根據耗電量與運算時間來計算出執行運算任務的成本數值；根據成本數值，判斷由一目標單元來執行運算任務，此目標單元是一運算單元或一伺服器；由目標單元執行運算任務。藉此，可同時考慮耗電與運算時間，適當地將運算量分配至運算單元與伺服器的其中之一。



S302：卸載運算量的決策方法的各步驟

S304：卸載運算量的決策方法的各步驟

S306：卸載運算量的決策方法的各步驟

S308：卸載運算量的決策方法的各步驟

S310：卸載運算量的決策方法的各步驟

S312：卸載運算量的決策方法的各步驟

S314：卸載運算量的決策方法的各步驟

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：101103616

G06Q 10/00 (2006.01)

※申請日：101.2.03

※IPC 分類：

G06F 1/32 (2006.01)

一、發明名稱：

G06F 15/16 (2006.01)

考慮耗時與耗電的卸載運算量的決策方法與運算系統 / DECISION METHOD CONSIDERING TIME AND POWER CONSUMPTION FOR OFFLOADING COMPUTATION AND COMPUTING SYSTEM

二、中文發明摘要：

一種考慮耗時與耗電的卸載運算量的決策方法。此方法包括：取得運算任務；從影響因子表中取得多個靜態影響因子；蒐集多個動態影響因子。此方法還包括：根據動態影響因子與靜態影響因子，產生各個運算單元與伺服器執行運算任務所需的耗電量與運算時間；根據耗電量與運算時間來計算出執行運算任務的成本數值；根據成本數值，判斷由一目標單元來執行運算任務，此目標單元是一運算單元或一伺服器；由目標單元執行運算任務。藉此，可同時考慮耗電與運算時間，適當地將運算量分配至運算單元與伺服器的其中之一。

三、英文發明摘要：

A decision method considering time and power consumption for offloading computations is provided. The method includes: obtaining a computing mission; obtaining a plurality of static effect factors from an effect factor table; collecting a plurality of dynamic effect factors. The method also includes: generating power consumptions and computing time that each of the computing unit and each of the server needs according to the dynamic effect factors and the static effect factors; calculating cost values of executing the computing mission according to the power consumptions and the computing time; determining a target unit to execute the computing mission according to the cost values, in which the target unit is a computing unit or a server; executing the computing mission by the target unit. Therefore, the power consumptions and the computing time are simultaneously considered, and computations are adequately allocated to one of the computing unit and the servers.

四、指定代表圖：

(一) 本案之指定代表圖：圖 3

(二) 本代表圖之元件符號簡單說明：

S302、S304、S306、S308、S310、S312、S314：卸載
運算量的決策方法的各步驟

**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卸載運算量的技術，且特別是有關於一種同時考慮耗時與耗電的卸載運算量的決策方法與使用此方法的系統。

【先前技術】

隨著網路速度越來越快，以及行動裝置越來越普及，雲端運算的技術也越來越重要。雲端運算的主要目的就是要將一裝置上的運算量透過網路卸載至其他運算能力較強的伺服器，如此一來可能可以減少此裝置上的運算時間與耗電。然而，並不是每個運算任務都適合被卸載至網路上的一伺服器。這是由於要將運算任務卸載至伺服器，還必須消耗其他的電力與運算時間來傳輸此運算任務。另一方面，由於一般運算系統或是行動裝置上會配置多個運算單元，因此另一種卸載的方式是將運算任務卸載至相同裝置的其他運算單元。

一般來說，適合被卸載至伺服器的運算任務需要有幾個特性，即運算量大且資料量小。例如，若在行動裝置上執行一個下棋的遊戲，要卸載至一伺服器所需要傳輸的資料量僅是描述整個棋盤的資料，但所需要的運算量卻可能非常大。因此，此下棋的遊戲便適合卸載至一遠端的伺服器。而適合卸載至相同裝置上其他運算單元的運算任務，卻可能需要具備不同的特性。因此，如何能夠同時考慮耗

電與運算時間，並將運算任務適當的卸載到一伺服器或一運算單元，為此領域研究人員所關心的問題。

【發明內容】

本發明一實施例提出了一種卸載運算量的決策方法與使用此方法的運算系統，可以同時考慮耗電與耗時來決定是否要卸載一運算任務。

本發明提出一種考慮耗時與耗電的卸載運算量的決策方法，用於一運算系統。此運算系統包括多個運算單元且耦接至至少一伺服器。此卸載運算量的決策方法包括：取得一運算任務；從運算系統的一影響因子表中取得多個靜態影響因子；以及蒐集多個動態影響因子。此卸載運算量的決策方法還包括：根據這些動態影響因子與靜態影響因子，產生各個運算單元與各個伺服器執行運算任務所需的耗電量與運算時間；根據這些耗電量與運算時間，計算出各個運算單元與各個伺服器執行運算任務的成本數值；根據這些成本數值，判斷由一目標單元來執行運算任務，此目標單元是上述運算單元與伺服器的其中之一；以及，將運算任務傳送至所決定出的目標單元，由目標單元執行運算任務。

在本發明的一實施例中，上述從運算系統的影響因子表中取得靜態影響因子的步驟還包括：判斷運算系統是否已產生影響因子表；以及，若運算系統尚未產生影響因子表，則建立影響因子表。

在本發明的一實施例中，上述將運算任務傳送至目標單元，由目標單元執行運算任務的步驟之後還包括：根據目標單元執行運算任務的結果，更新影響因子表。

在本發明的一實施例中，上述根據耗電量與運算時間，計算出各個運算單元與各個伺服器執行運算任務的成本數值的步驟包括：根據方程式(1)、(2)與(3)來計算成本數值，

$$\varepsilon_T^p = (\hat{T}_p - \hat{T}_{cpu}) / \hat{T}_{cpu} \dots\dots\dots(1)$$

$$\varepsilon_E^p = (\hat{E}_p - \hat{E}_{cpu}) / \hat{E}_{cpu} \dots\dots\dots(2)$$

$$C_p = \alpha \cdot \varepsilon_T^p + (1 - \alpha) \cdot \varepsilon_E^p \dots\dots\dots(3)$$

其中，cpu 為運算單元中的一主要運算單元。p 為一比較運算單元，此比較運算單元是運算單元的其中之一或者是伺服器的其中之一。C_p 為比較運算單元的成本數值。 \hat{T}_p 為比較運算單元執行運算任務所需的運算時間。 \hat{E}_p 為比較運算單元執行運算任務所需的耗電量。 \hat{T}_{cpu} 為主要運算單元執行運算任務所需的運算時間。 \hat{E}_{cpu} 為主要運算單元執行運算任務所需的耗電量。 α 為一個大於等於 0 且小於等於 1 的實數；以及，將成本數值最小的比較運算單元設定為目標單元。

以另外一個角度來說，本發明還提出一種運算系統。此運算系統包括網路介面，記憶體，多個運算單元，一卸載決策單元。其中，網路介面是用以連接至至少一個伺服器。記憶體中儲存了一個影響因子表。卸載決策單元是耦接至網路介面、記憶體與運算單元，用以接收一運算任務。其中，卸載決策單元會從影響因子表中取得多個靜態影響因子，並且蒐集多

個動態影響因子。卸載決策單元會根據所取得的動態影響因子與靜態影響因子，產生各個運算單元與各個伺服器執行運算任務所需的耗電量與運算時間。卸載決策單元會根據耗電量與運算時間，計算出各個運算單元與各個伺服器執行運算任務的成本數值。卸載決策單元還會根據所計算出的成本數值，判斷由一目標單元來執行運算任務，此目標單元為運算單元與伺服器的其中之一。卸載決策單元會將運算任務傳送至目標單元，由目標單元執行運算任務。

在本發明的一實施例中，上述卸載決策單元還用以判斷運算系統是否已產生影響因子表，若運算系統尚未產生影響因子表，則卸載決策單元會建立影響因子表。

在本發明的一實施例中，上述動態影響因子包括運算任務的運算量、各個伺服器的一伺服器運算速度以及網路速度。

在本發明的一實施例中，上述靜態影響因子包括記憶體速度與各個運算單元的一運算單元運算速度。

在本發明的一實施例中，上述卸載決策單元還用以根據目標單元執行運算任務的結果，更新影響因子表。

在本發明的一實施例中，上述卸載決策單元還用以根據上述的方程式(1)、(2)與(3)來計算成本數值，並將成本數值最小的比較運算單元設定為目標單元。

基於上述，本發明所提出的運算系統與卸載運算量的決策方法，可以同時考慮耗電與運算時間，並根據一個變數來將耗電與運算時間產生出成本數值。使用者可以調整

此變數來計算出不同的成本數值。根據此成本數值，可以適當的決定是否要卸載運算任務。

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】

圖 1 是根據一實施例說明運算系統的方塊圖。

請參照圖 1。運算系統 100 是耦接至伺服器 140，且接收運算任務 120，並決定運算任務 120 是否要卸載至伺服器 140。或者是，運算系統 100 中包括了多個運算單元，運算系統 100 是用以決定運算任務 120 是要由哪一個運算單元來執行。特別的是，運算系統 100 可以同時考慮伺服器 140 或運算系統中的多個運算單元，用以決定運算任務 120 是否要卸載到伺服器 140 或是上述多個運算單元的其中之一。在本實施例中，運算系統 100 是一個行動裝置，例如為智慧型手機。然而，在其他實施例中，運算系統 100 也可以是桌上型電腦或是筆記型電腦，本發明應不在此限。

運算任務 120 例如為一個矩陣相乘運算或是三維影像繪圖的運算，本發明並不限制運算任務 120 的內容。

而伺服器 140 例如為一個大型的運算裝置，擁有比運算系統 100 好的運算能力。然而，發明並不限制伺服器 140 的規模與運算能力。另一方面，在其他實施例中，運算系統 100 也可耦接至數目更多的伺服器，運算系統 100 可以決定運算任務 120 是要卸載是哪一個伺服器。本發明並不

限制運算系統 100 耦接至的伺服器的數量。

運算系統 100 包括運算單元 102、運算單元 104、網路介面 106、卸載決策單元 108 以及記憶體 160。

運算單元 102 例如為中央處理器(central processing unit, CPU)，為運算系統 100 主要的運算單元。而運算單元 104 例如為圖形處理器(Graphics Processing Unit, GPU)，用以處理特定的運算。然而，本發明也可以將運算單元 104 設定為主要的運算單元，且運算單元 104 也可以是一微處理器(microprocessor)。在其他實施例中，運算單元 102 與運算單元 104 也可以是同一個處理器的兩個核心，本發明並不限制運算單元 102 與運算單元 104 的種類。另一方面，在本實施例中運算系統 100 包括兩個運算單元，然而在其他實施例中運算系統 100 也可以包括數目更多的運算單元，運算系統 100 可以決定要將運算任務 120 分配給哪一個運算單元來執行，本發明並不限制運算單元的個數。

網路介面 106 是用以耦接至伺服器 140。例如，網路介面 106 是符合 WiFi 的網路介面卡。然而，網路介面 106 也可以是符合其他網路協定或其他通訊協定的裝置，本發明並不在此限。

記憶體 160 是用以儲存運算系統 100 的資料，例如為動態隨機存取記憶體(dynamic random access memory, DRAM)。然而，記憶體 160 也可以是靜態隨機存取記憶體，硬碟，非揮發性的快閃記憶體，本發明並不在此限。特別的是，記憶體 160 中儲存了影響因子表 162。影響因

子表 162 是用以儲存多個靜態影響因子。這些靜態影響因子是不容易隨著時間或環境而改變，但會影響計算運算任務 120 所需的耗電與運算時間的變數。例如，靜態影響因子為運算單元 102、104 的運算單元運算速度(例如，運算單元 102 的時脈頻率(clock frequency))。或者，靜態影響因子也可以是記憶體 160 的記憶體速度，例如，記憶體 160 的頻寬(bandwidth)。運算系統 100 可以根據影響因子表 162 中的靜態影響因子，來決定運算任務 120 要卸載至哪一個運算單元或者是伺服器。

卸載決策單元 108 是耦接至運算單元 102、運算單元 104、記憶體 160 以及網路介面 106。例如，卸載決策單元 108 是一個特殊應用積體電路 (Application-specific integrated circuit, ASIC)。然而，卸載決策單元 108 也可以是微處理器，本發明並不限制卸載決策單元的種類。卸載決策單元 108 是用以接收運算任務 120，並決定運算任務 120 是要由哪一個運算單元或是伺服器來執行。在另一個實施例中，卸載決策單元 108 所執行的功能也可以被實作為程式碼，由一運算單元來執行。

具體來說，卸載決策單元 108 會先從影響因子表 162 中取得靜態影響因子。例如，卸載決策單元 108 會取得運算單元 102、104 的運算速度，與記憶體 160 的記憶體速度。由於這些靜態影像因子並不會經常改變，例如，記憶體 160 的頻寬為固定值，因此運算系統 100 是將這些靜態影響因子儲存在記憶體 160 的影響因子表中。另一方面，卸載決

策單元 108 還會判斷運算系統 100 是否以建立了影響因子表 162。若卸載決策單元 108 判斷影響因子表 162 尚未被建立，則會建立一個新的影響因子表 162，並蒐集靜態影響因子來儲存至影響因子表 162 中。

卸載決策單元 108 還會蒐集多個動態影響因子，這些動態影響因子是可能會隨著時間或是環境而改變的變數。例如，動態影響因子為伺服器 140 的伺服器運算速度、網路速度與運算任務 120 的運算量。由於網路上的各種狀態變化非常快，因此電腦裝置 100 連線至網路的速度可能隨時都在改變。例如，運算系統 100 所送出封包的路徑可能經過一段時間之後就會改變。或者，在一實施例中，運算系統 100 是透過無線網路連線至伺服器 140，而無線網路的通訊品質也會不斷的改變。因此，卸載決策單元 108 會先取得目前的網路速度。另一方面，伺服器 140 的運算速度也可能不斷改變。例如，在某些時間有許多使用者都將運算卸載到伺服器 140，使得伺服器 140 的負載增加，藉此使得運算系統 100 所能分配到的運算資源減少。因此，卸載決策單元 108 還會取得伺服器 140 目前的運算速度。最後，運算任務 120 的運算量也可能不是固定。例如，運算任務 120 是一個矩陣相乘的運算，然而所相乘的矩陣大小卻不固定，使得做矩陣相乘所需的運算量也不相同。也就是說，卸載決策單元 108 可以根據目前運算系統 100 的狀況，取得多個會影響計算的動態影響因子。

在取得上述的靜態影響因子與動態影響因子以後，卸

載決策單元 108 會根據這些動態影響因子與靜態影響因子來產生運算單元 102、104 與伺服器 140 執行運算任務 120 所需的耗電量與運算時間。舉例來說，當運算單元 102 的運算單元運算速度較低時，運算單元 102 執行運算任務 120 所需的運算時間便較長。當目前的網路速度較慢時，要卸載運算任務 120 至伺服器 140 所需的耗電量與運算時間都會增加。或者是，當運算任務 120 的運算量較大時，如果將運算任務 120 卸載至伺服器 140，運算系統 100 便不需要消耗用以執行運算任務 120 的電力。然而，如果將運算任務 120 卸載至伺服器 140，運算系統 100 便需要額外消耗電力來傳送運算任務 120。卸載決策單元 108 會產生每個運算單元與伺服器執行運算任務 120 所需的耗電量與運算時間，藉此判斷由哪一個運算單元或伺服器來執行運算任務 120。

接著，卸載決策單元 108 會根據所產生的耗電與運算時間，計算出各個運算單元 102、104 與伺服器 140 執行運算任務 120 的成本數值。此成本數值是用以表示運算單元或伺服器執行運算任務 120 時所付出的代價為多少。

卸載決策單元 108 會將運算系統 100 中的一個運算單元設定為主要的運算單元，例如，運算單元 102。接著，卸載決策單元 108 會以運算單元 102 執行運算任務 120 所需的耗電與運算時間為基準，來計算出其他運算單元或伺服器的成本數值。詳細來說，卸載決策單元 108 是根據下列方程式(1)、(2)與(3)來計算出各個運算單元與伺服器的

成本數值。

$$\varepsilon_T^P = (\hat{T}_p - \hat{T}_{cpu}) / \hat{T}_{cpu} \dots\dots\dots(1)$$

$$\varepsilon_E^P = (\hat{E}_p - \hat{E}_{cpu}) / \hat{E}_{cpu} \dots\dots\dots(2)$$

$$C_p = \alpha \cdot \varepsilon_T^P + (1 - \alpha) \cdot \varepsilon_E^P \dots\dots\dots(3)$$

其中，cpu 是運算系統 100 的一個主要運算單元，在本實施例中為運算單元 102。p 為一比較運算單元，此比較運算單元是運算單元 102 或運算單元 103 的其中之一，或者是伺服器 140。也就是說，比較運算單元是用來與主要運算單元 102 比較，並計算出比較運算單元的成本數值。C_p 即是此比較運算單元的成本數值。 \hat{T}_p 為比較運算單元執行運算任務 120 所需的運算時間。 \hat{E}_p 為比較運算單元執行運算任務所需的耗電量。 \hat{T}_{cpu} 為主要運算單元執行運算任務 120 所需的運算時間。 \hat{E}_{cpu} 為主要運算單元執行運算任務 120 所需的耗電量。最後， α 為一個大於等於 0 且小於等於 1 的實數，可以由使用者來調整。

方程式(1)所代表的意義是計算出比較運算單元執行運算任務 120 所需的運算時間與主要運算單元的運算時間之間的差距。方程式(2)所代表的意義是計算出比較運算單元執行運算任務 120 所需的耗電與主要運算單元的耗電之間的差距。結合方程式(1)、(2)與(3)，其所代表的意思是，當一個比較運算單元執行運算任務 120 所消耗的運算時間與耗電越少時，所計算出的成本數值便越低。並且，成本數值是運算時間與耗電分別乘上 α 與 $1 - \alpha$ 之後的相加。因此，當使用者覺得運算時間比較重要時，便可以調整 α 為較大的值。當使用者覺得耗電比較重要時，便可以調整 α 為較

小的數值。舉例來說，若運算系統 100 是一個行動裝置，由於行動裝置上的電池容量有限，此時便可以將 α 調整為較小的數值，此時便比較有機會將運算任務 120 卸載給耗電較小的比較運算單元(例如，伺服器 140)來執行。在另一實施例中，運算系統 100 是一個桌上型電腦，擁有比行動裝置更好的運算能力。此時使用者在意的是運算時間，因此可以將 α 調整為較大的數值。然而， α 可以是隨著需要而調整的數值，本發明並不限制 α 的數值。

根據方程式(1)、(2)與(3)，卸載決策單元 108 會選擇成本數值最小的一個比較運算單元(即，運算單元 102、104 或伺服器 140 的其中之一)作為一目標單元。此目標單元便是要用以執行運算任務 120 的運算單元或是伺服器。因此，卸載決策單元 108 會進一步的將運算任務 120 傳送至所決定出的目標單元，由目標單元執行運算任務 120。若目標單元是伺服器 140，卸載決策單元 108 便會透過網路介面 106 將運算任務傳送至伺服器 140。

當目標單元執行完運算任務 120 以後，卸載決策單元 108 會根據目標單元執行運算任務 120 的結果來更新影響因子表 162。例如，卸載決策單元 108 會更新影響因子表 162 中的靜態影響因子，並且由於靜態影響因子可能只有一部份有改變，因此卸載決策單元 108 只需要更新影響因子表 162 的一部份資料，不需要將運算結果全部重新寫入。

圖 2 是根據一實施例說明卸載決策單元的運作流程圖。

請參照圖 2。在步驟 S202 中，卸載決策單元 108 會判斷影響因子表是否存在。若是則進行步驟 S208，否則進行步驟 S204。

步驟 S204 中，卸載決策單元 108 會建立影響因子表。接著在步驟 S206 中，卸載決策單元 108 會將運算任務傳送至主要運算單元，由主要運算單元執行運算任務。

在步驟 S208 中，卸載決策單元 108 會取得動態影響因子與靜態影響因子。接著在步驟 S210 中，卸載決策單元 108 會根據動態影響因子與靜態影響因子來產生各個運算單元與伺服器的成本數值。步驟 S212 中，卸載決策單元 108 根據成本數值決定出目標單元。步驟 S214 中，卸載決策單元 108 會判斷目標單元是否為主要運算單元，若是則進行步驟 S206，否則進行步驟 S216。

步驟 S216 中，卸載決策單元 108 會將運算任務卸載至主要運算單元以外的運算單元或伺服器以執行運算任務。在步驟 S218 中，卸載決策單元 108 會取得目標單元執行運算任務的結果，並根據運算結果來更新影響因子表。最後，會結束圖 2 的流程。

然而，圖 2 中各步驟以詳細說明如上，在此便不再贅述。

另一方面，本發明還提出一種卸載運算量的決策方法。圖 3 是根據一實施例說明卸載運算量的決策方法的流程圖。

請參照圖 3。在步驟 S302 中，取得一運算任務。在步

驟 S304 中，從運算系統的一影響因子表中取得多個靜態影響因子。在步驟 S306 中，蒐集多個動態影響因子。在步驟 S308 中，根據動態影響因子與靜態影響因子，產生各個運算單元與各個伺服器執行運算任務所需的耗電量與運算時間。在步驟 S310 中，根據耗電量與運算時間，計算出各個運算單元與各個伺服器執行運算任務的成本數值。在步驟 S312 中，根據成本數值，判斷由一目標單元來執行運算任務，其中目標單元為運算單元與伺服器的其中之一。在步驟 S314 中，將運算任務傳送至目標單元，由目標單元執行運算任務。

值得注意的是，步驟 S304 與步驟 S306 的執行順序可以互換，本發明並不在此限。並且，圖 3 中各步驟已詳細說明如上，在此便不再贅述。

綜上所述，本發明實施例所提出的運算系統與卸載運算量的決策方法，可以根據各個運算單元與伺服器執行運算任務的耗電量與運算時間來決定出一個成本數值。並且，使用者可以調整一個變數來控制耗電量與運算時間之間的重要性。根據此變數與成本數值，可以根據不同的情況，適當的決定是否要將運算任務卸載至其他運算單元或伺服器。

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 是根據一實施例說明運算系統的方塊圖。

圖 2 是根據一實施例說明卸載決策單元的運作流程圖。

圖 3 是根據一實施例說明卸載運算量的決策方法的流程圖。

【主要元件符號說明】

100：運算系統

102：運算單元

104：運算單元

106：網路介面

108：卸載運算單元

120：運算任務

140：伺服器

160：記憶體

162：影響因子表

S202、S204、S206、S208、S210、S212、S214、S216、
S218：卸載運算單元運作的各步驟

S302、S304、S306、S308、S310、S312、S314：卸載
運算量的決策方法的各步驟

七、申請專利範圍：

1. 一種考慮耗時與耗電的卸載運算量的決策方法，用於一運算系統，其中該運算系統包括多個運算單元且耦接至至少一伺服器，該決策方法包括：

取得一運算任務；

從該運算系統的一影響因子表中取得多個靜態影響因子；

蒐集多個動態影響因子；

根據該些動態影響因子與該些靜態影響因子，產生各該些運算單元與各該些伺服器執行該運算任務所需的一耗電量與一運算時間；

根據該些耗電量與該些運算時間，計算出各該些運算單元與各該些伺服器執行該運算任務的一成本數值；

根據該些成本數值，判斷由一目標單元來執行該運算任務，其中該目標單元為該些運算單元與該些伺服器的其中之一；以及

將該運算任務傳送至該目標單元，由該目標單元執行該運算任務。

2. 如申請專利範圍第 1 項所述之卸載運算量的決策方法，其中從該運算系統的該影響因子表中該些靜態影響因子的步驟還包括：

判斷該運算系統是否已產生該影響因子表；以及

若該運算系統尚未產生該影響因子表，則建立該影響因子表。

3. 如申請專利範圍第 1 項所述之卸載運算量的決策方法，其中該些動態影響因子包括該運算任務的一運算量、各該些伺服器的一伺服器運算速度以及至少一網路速度。

4. 如申請專利範圍第 1 項所述之卸載運算量的決策方法，其中各該些靜態影響因子包括一記憶體速度與各該些運算單元的一運算單元運算速度。

5. 如申請專利範圍第 1 項所述之卸載運算量的決策方法，其中將該運算任務傳送至該目標單元，由該目標單元執行該運算任務的步驟之後還包括：

根據該目標單元執行該運算任務的結果，更新該影響因子表。

6. 如申請專利範圍第 1 項所述之卸載運算量的決策方法，其中根據該些耗電量與該些運算時間，計算出各該些運算單元與各該些伺服器執行該運算任務的該成本數值的步驟包括：

根據方程式(1)、(2)與(3)來計算該成本數值，

$$\varepsilon_T^P = (\hat{T}_p - \hat{T}_{cpu}) / \hat{T}_{cpu} \dots\dots\dots(1)$$

$$\varepsilon_E^P = (\hat{E}_p - \hat{E}_{cpu}) / \hat{E}_{cpu} \dots\dots\dots(2)$$

$$C_p = \alpha \cdot \varepsilon_T^P + (1 - \alpha) \cdot \varepsilon_E^P \dots\dots\dots(3)$$

其中，cpu 為該些運算單元中的一主要運算單元，p 為一比較運算單元，該比較運算單元是該些運算單元的其中之一或者是該些伺服器的其中之一， C_p 為該比較運算單元的該成本數值， \hat{T}_p 為該比較運算單元執行該運算任務所需的該運算時

間， \hat{E}_p 為該比較運算單元執行該運算任務所需的該耗電量， \hat{t}_{cpu} 為該主要運算單元執行該運算任務所需的該運算時間， \hat{E}_{cpu} 為該主要運算單元執行該運算任務所需的該耗電量， α 為大於等於 0 且小於等於 1 的實數；以及

將該成本數值最小的該比較運算單元設定為該目標單元。

7. 一種運算系統，包括：

一網路介面，用以連接至至少一伺服器；

一記憶體，其中儲存一影響因子表；

多個運算單元；以及

一卸載決策單元，耦接至該網路介面、該記憶體與該些運算單元，用以接收一運算任務，

其中該卸載決策單元從該影響因子表中取得多個靜態影響因子，

其中該卸載決策單元蒐集多個動態影響因子，

其中該卸載決策單元根據該些動態影響因子與該些靜態影響因子，產生各該些運算單元與各該些伺服器執行該運算任務所需的一耗電量與一運算時間，

其中該卸載決策單元根據該些耗電量與該些運算時間，計算出各該些運算單元與各該些伺服器執行該運算任務的一成本數值，

其中該卸載決策單元根據該些成本數值，判斷由一目標單元來執行該運算任務，其中該目標單元為該些運算單元與該些伺服器的其中之一，

其中該卸載決策單元將該運算任務傳送至該目標單元，由該目標單元執行該運算任務。

8. 如申請專利範圍第 7 項所述之運算系統，其中該卸載決策單元還用以判斷該運算系統是否已產生該影響因子表，若該運算系統尚未產生該影響因子表，則該卸載決策單元建立該影響因子表。

9. 如申請專利範圍第 7 項所述之運算系統，其中該些動態影響因子包括該運算任務的一運算量、各該些伺服器的一伺服器運算速度以及至少一網路速度。

10. 如申請專利範圍第 7 項所述之運算系統，其中各該些靜態影響因子包括一記憶體速度與各該些運算單元的一運算單元運算速度。

11. 如申請專利範圍第 7 項所述之運算系統，其中該卸載決策單元還用以根據該目標單元執行該運算任務的結果，更新該影響因子表。

12. 如申請專利範圍第 7 項所述之運算系統，其中該卸載決策單元還用以根據方程式(1)、(2)與(3)來計算該成本數值，

$$\varepsilon_T^p = (\hat{T}_p - \hat{T}_{cpu}) / \hat{T}_{cpu} \dots\dots\dots(1)$$

$$\varepsilon_E^p = (\hat{E}_p - \hat{E}_{cpu}) / \hat{E}_{cpu} \dots\dots\dots(2)$$

$$C_p = \alpha \cdot \varepsilon_T^p + (1 - \alpha) \cdot \varepsilon_E^p \dots\dots\dots(3)$$

其中，cpu 為該些運算單元中的一主要運算單元，p 為一比較運算單元，該比較運算單元是該些運算單元的其中之一或者是該些伺服器的其中之一， C_p 為該比較運算單元的該成本

數值， \hat{t}_p 為該比較運算單元執行該運算任務所需的該運算時間， \hat{e}_p 為該比較運算單元執行該運算任務所需的該耗電量， \hat{t}_{cpu} 為該主要運算單元執行該運算任務所需的該運算時間， \hat{e}_{cpu} 為該主要運算單元執行該運算任務所需的該耗電量， α 為大於等於 0 且小於等於 1 的實數，

該卸載決策單元還用以將該成本數值最小的該比較運算單元設定為該目標單元。

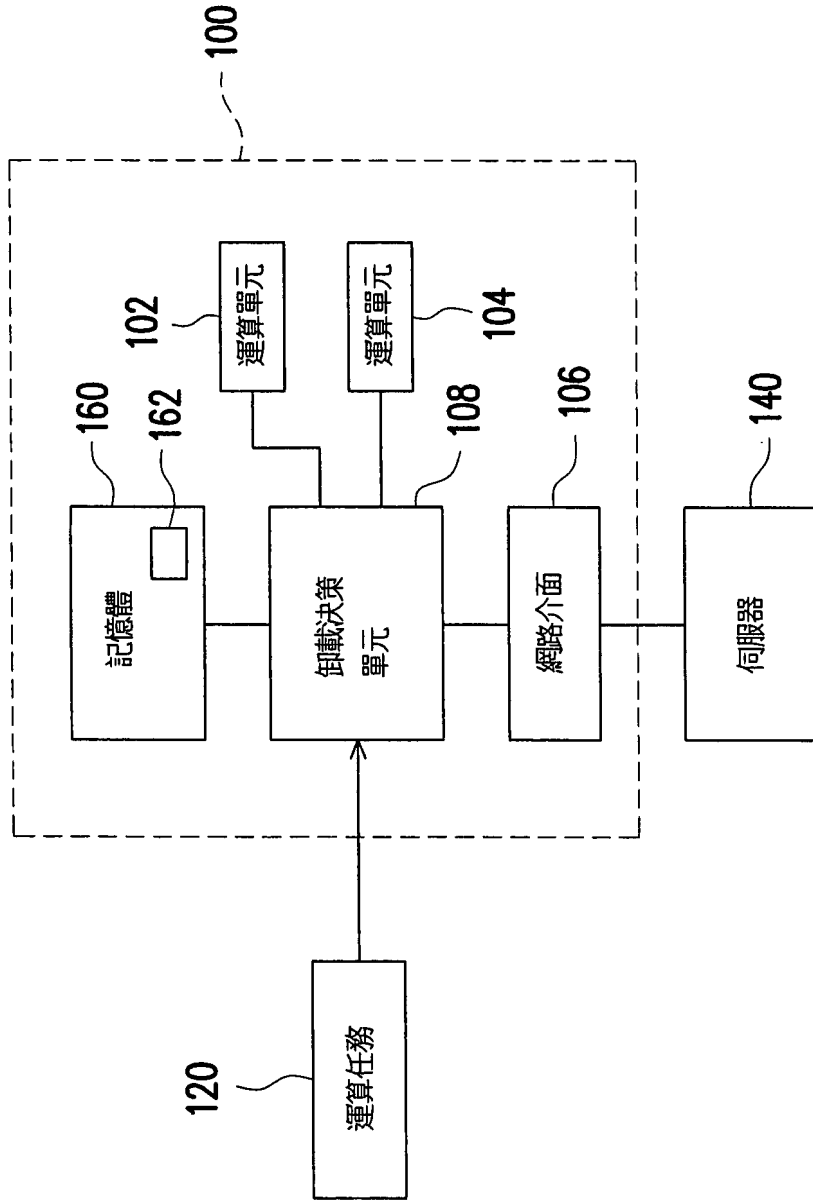


圖1

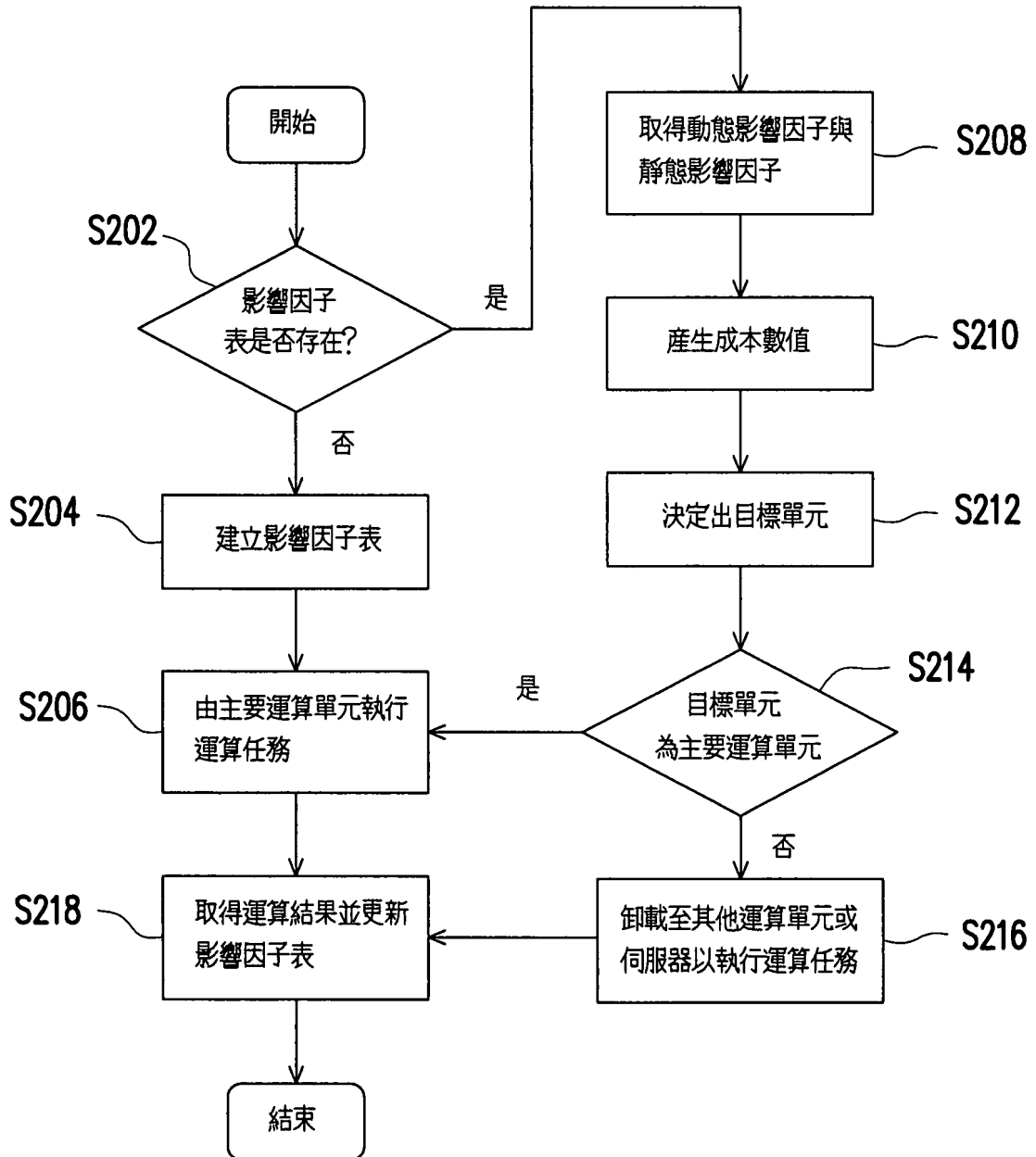


圖 2

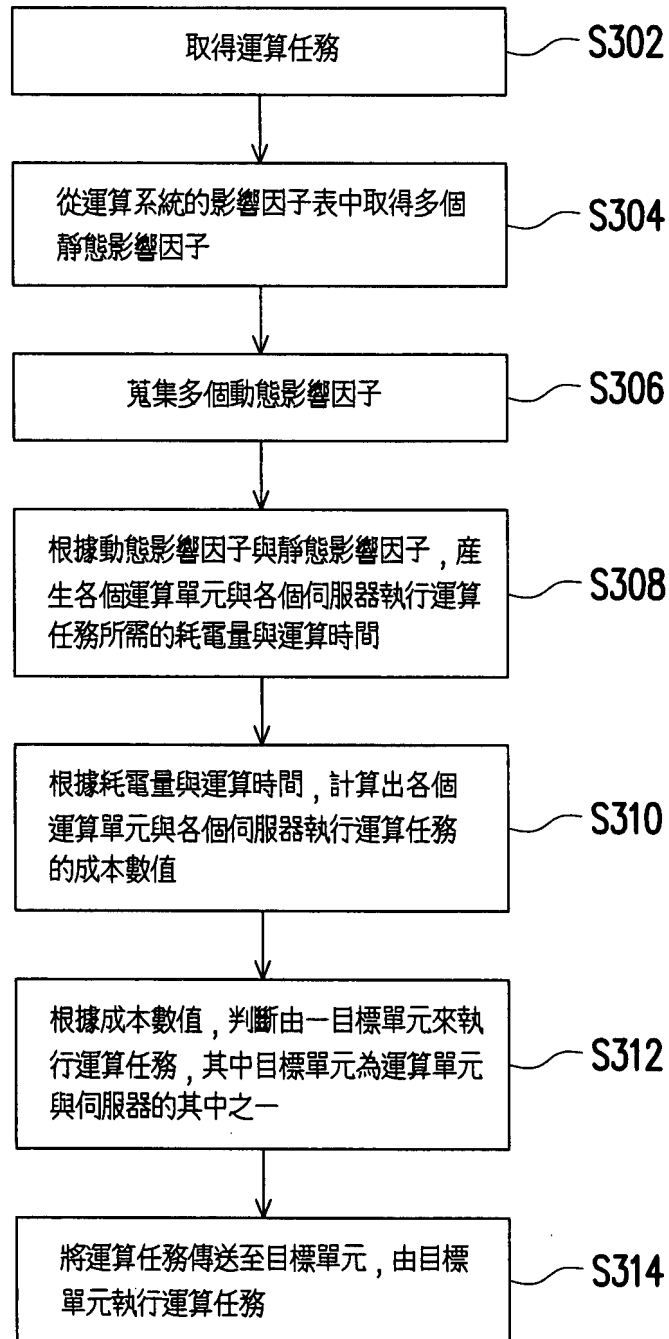


圖 3