(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 201017434 A1
(43)公開日： 中華民國 99 (2010) 年 05 月 01 日

(21)申請案號：098132061 (22)申請日： 中華民國 98 (2009) 年 09 月 23 日
(51)Int. Cl.： *G06F15/163 (2006.01)*
(30)優先權：2008/10/02 世界智慧財產權PCT/US08/78621
組織
(71)申請人：惠普研發公司 (美國) HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P. (US)
美國
(72)發明人：雷斯阿特瑞 葛雷格B LESARTRE, GREGG B. (US)；維那爾 克雷格 WARNER, CRAIG (US)；構斯汀 蓋瑞 GOSTIN, GARY (US)；柏克豪斯 約翰W BOCKHAUS, JOHN W. (US)
(74)代理人：惲軼群；陳文郎
申請實體審查：無 申請專利範圍項數：10 項 圖式數：4 共 29 頁

(54)名稱

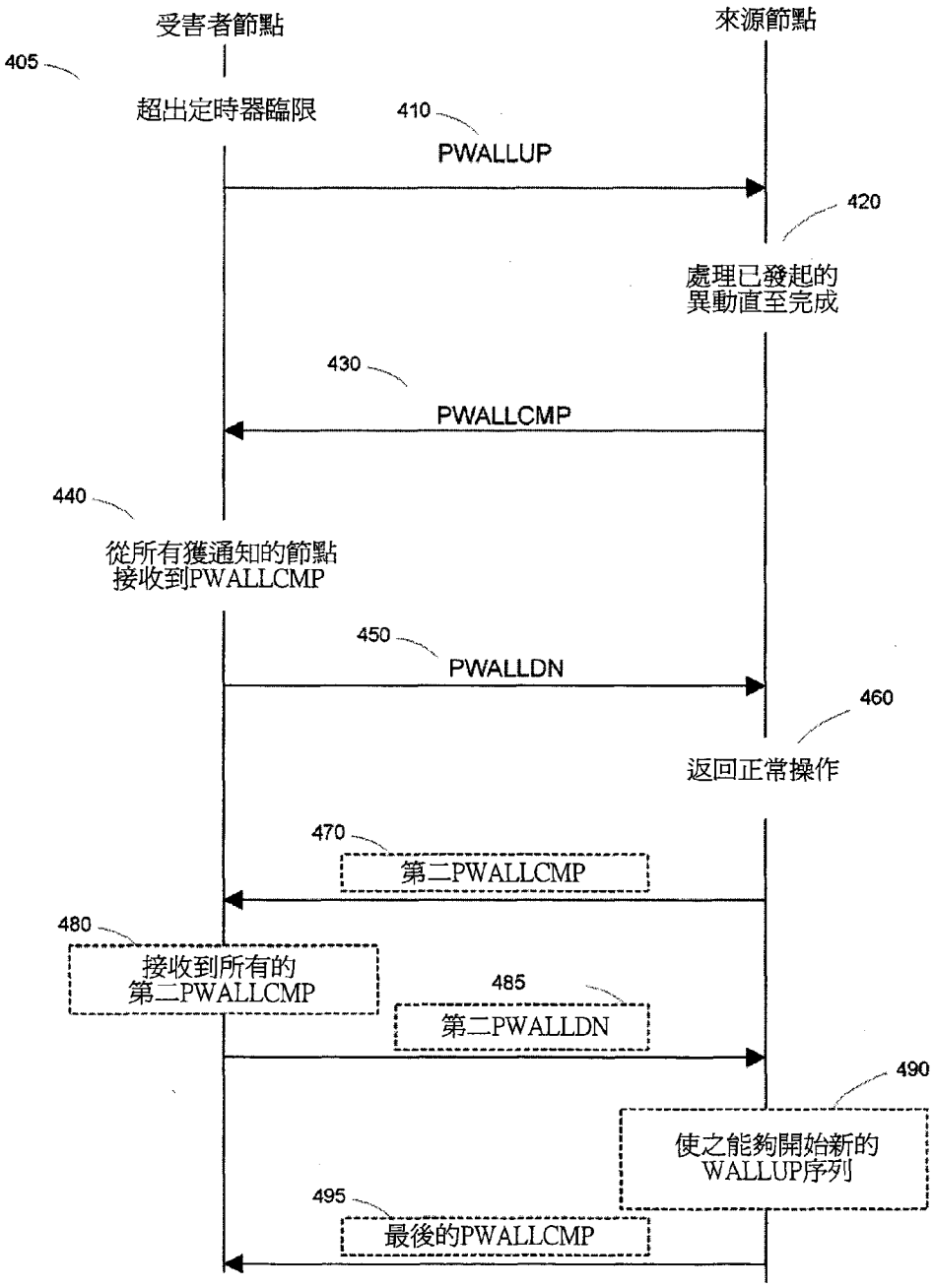
管理多處理器互連體中潛伏期的技術

MANAGING LATENCIES IN A MULTIPROCESSOR INTERCONNECT

(57)摘要

在一計算系統中具有發送異動到一交換結構之多個異動來源節點，一所獲服務不足的節點通知該系統中的來源節點其需要額外的系統頻寬以及時完成一正在進行的異動。該等獲通知的節點繼續處理已經開始的異動直至完成，但停止引入新的訊務到該結構中直到該所獲服務不足的節點指示其已經進展到一預選定點之時。

405~495：步驟


受害者節點
來源節點
405
超出定時器臨限
410
PWALLUP
420
處理已發起的
異動直至完成
430
PWALLCMP
440
從所有獲通知的節點
接收到PWALLCMP
450
PWALLDN
460
返回正常操作
470
第二PWALLCMP
480
接收到所有的
第二PWALLCMP
485
第二PWALLDN
490
使之能夠開始新的
WALLUP序列
495
最後的PWALLCMP

(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 201017434 A1
(43)公開日： 中華民國 99 (2010) 年 05 月 01 日

(21)申請案號：098132061 (22)申請日： 中華民國 98 (2009) 年 09 月 23 日
(51)Int. Cl.： *G06F15/163 (2006.01)*
(30)優先權：2008/10/02 世界智慧財產權PCT/US08/78621
組織
(71)申請人：惠普研發公司 (美國) HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P. (US)
美國
(72)發明人：雷斯阿特瑞 葛雷格B LESARTRE, GREGG B. (US)；維那爾 克雷格 WARNER, CRAIG (US)；構斯汀 蓋瑞 GOSTIN, GARY (US)；柏克豪斯 約翰W BOCKHAUS, JOHN W. (US)
(74)代理人：惲軼群；陳文郎
申請實體審查：無 申請專利範圍項數：10 項 圖式數：4 共 29 頁

(54)名稱

管理多處理器互連體中潛伏期的技術

MANAGING LATENCIES IN A MULTIPROCESSOR INTERCONNECT

(57)摘要

在一計算系統中具有發送異動到一交換結構之多個異動來源節點，一所獲服務不足的節點通知該系統中的來源節點其需要額外的系統頻寬以及時完成一正在進行的異動。該等獲通知的節點繼續處理已經開始的異動直至完成，但停止引入新的訊務到該結構中直到該所獲服務不足的節點指示其已經進展到一預選定點之時。

六、發明說明：

【發明所屬之技術領域】

本揭露關於用於使用在具有多個節點之電腦系統中的架構，每一節點包含一異動來源，其中該等節點透過一系統交換結構(switching fabric)互相通訊。

就性質而言，在一大的電腦系統中之計算節點未必平等地接取該系統中的所有其它節點。較接近一目標節點之節點往往比其它較遠之節點獲得此目標節點之頻寬之一較大部分。藉此，在該系統結構中有大量擁塞之情景下，較遠之節點可能經歷難以接受的長響應時間(潛伏期)。此過長的潛伏期可最終導致系統故障，因為該系統或作業系統(OS)中的元件放棄該等緩慢異動。

【先前技術】

在先前技術中，此問題之一個解決方案是限制總系統的大小。該問題也可藉由預分配可得頻寬而管理，但如果所有節點都不需要它們的分配則使得頻寬未獲使用，此情況可能出現在例如分區系統中。另一習知的解決方案是增加用於節點之間通訊之虛擬通道之數目。雖然這樣可緩解潛伏期問題，但其出現了一額外的費用，因為額外的虛擬通道需要該結構中之額外的緩衝及控制資源。

所需的是更好地管理包含多個異動來源節點之一電腦系統中的潛伏期之一方式，其改善現有實施之缺點。

【發明內容】

發明概要

一多處理器計算系統中一所獲服務不足的異動來源節點通知該系統中的其它節點其未接收到足以及時完成一正在進行的異動之系統頻寬。該系統中的其它節點繼續允許完成已經開始的異動所需的訊務，但停止產生新的訊務到該結構中，直到該所獲服務不足的節點指示其已取得可接受的進展之時。藉此，由於過長通訊潛伏期所產生的一系統故障之稀少但災難性的問題可被避免，而就系統區、電力或複雜度而言無需施加高額外費用。

應當了解的是，上述的大體描述及下面的詳細描述都是示範性及說明性的且意在提供對申請專利範圍中描述的方法及系統之進一步說明。

圖式簡單說明

附圖被包括以提供本揭露之一進一步理解且被併入並構成本說明書之一部分，其說明了揭露的實施例且與描述一起用於說明揭露的方法及系統之原理。

在該等圖式中：

第1圖是根據本文描述的系統及方法之一實施態樣之一示範性計算環境之一方塊圖；

第2圖是顯示一示範性資料通訊架構之示範性元件之協作之一方塊圖。

第3A圖及第3B圖是各種可分區計算系統之方塊圖，其中可使用本文描述之系統及方法之原理；及

第4圖是根據本文描述的系統及方法之一實施態樣，在管理通訊資料中的潛伏期時藉由一示範性資料通訊架構執

行的處理之一方塊圖。

【實施方式】

詳細描述

說明性的計算環境

第1圖描述了根據本文描述的系統及方法之諸如可形成一示範性多處理器計算環境之一部分的一示範性計算系統100。計算系統100能夠執行各種計算應用程式180。示範性計算系統100主要受電腦可讀指令控制，該等電腦可讀指令可以是儲存於諸如一硬式驅動機或記憶體之一電腦可讀媒體中的軟體形式。這樣的軟體可在中央處理單元(CPU)130內執行以使資料處理系統100運作。在很多習知的電腦伺服器中，工作站及個人電腦中央處理單元130藉由被稱為微處理器之一個或多個微電子晶片實施。共處理機140是一可取捨的處理器，其有別於主CPU 130，其執行額外的功能或協助CPU 130。一個一般類型之共處理器是浮點共處理器，也被稱為一數值或數學共處理器，其被設計以比該通用CPU 130更快且更好地執行數值計算。

應當明白的是，儘管說明性的計算環境被顯示為包含一單一CPU 130，但這樣的描述只是說明性的，計算環境100可以包含多個CPU 130。此外，計算環境100可透過諸如一交換結構(圖未示)之一通訊網路利用遠端CPU(圖未示)之資源。

在操作中，CPU 130提取、解碼並執行指令，且透過該電腦之主要資料傳送路徑125傳送資訊到其它資源及自其

它資源傳送資訊。資料傳送路徑125可包含一並列系統匯流排或一個或多個稱為線道(lane)之點對點串列鏈結(link)。在串列鏈結之情況下，一集線器(圖未示)可作為允許點對點裝置互連體即時被重新路由之一縱橫交換器(crossbar switch)。此動態點對點連接行為可使得系統裝置同時執行操作，因為多於一對裝置可同時互相通訊。多個這樣的線道可被群組化且協同共同工作以提供較大的頻寬。該資料傳送路徑125連接計算系統100中的元件且提供用於資料交換之媒體。資料傳送路徑125典型地包括用於發送資料之資料線道或通道、用於發送位址之位址線及用於發送中斷及控制訊息之控制線。

耦接於系統資料傳送路徑125之記憶體裝置包括隨機存取記憶體(RAM)110及唯讀記憶體(ROM)115。這樣的記憶體包括允許資訊被儲存及擷取之電路。ROM 115通常包含不能修改的儲存資料。儲存在RAM 110中的資料可藉由CPU 130或其它硬體裝置讀取或改變。存取RAM 110及/或ROM 115可受一記憶體控制器105控制。記憶體控制器105可提供當執行指令時將虛擬位址轉譯成實體位址之一位址轉譯功能。記憶體控制器105也可提供隔離該系統內之進程(process)之一記憶體保護功能。

此外，計算系統100可包含負責把指令從CPU 130傳遞到周邊設備(諸如列印機150、鍵盤155、滑鼠160及資料儲存裝置165)之周邊設備控制器145。

顯示器170，其受顯示器控制器175控制，其用以顯示

由計算系統100產生的視覺輸出。這樣的視覺輸出可包括文字、圖形、動畫圖形、視訊及類似之物。例如，顯示器170可藉由例如一基於CRT之視訊顯示器、一基於LCD之平板顯示器、基於氣體電漿之平板顯示器或一觸摸面板來實施。顯示器控制器175包括產生發送到該顯示器170之一視訊信號所需的電子元件。

而且，計算系統100可包含網路配接器120，其用以把計算系統100連接到一外部通訊網路185。通訊網路185可為電腦使用者提供透過電子方式通訊及傳送軟體及資訊之裝置。應當明白的是，所示的該網路及其它連接是示範性的且建立電腦及電腦元件之間的通訊連接之其它裝置可被使用。

係為本文描述的系統及方法可操作於其中的一計算環境之一部分的示範性電腦系統100只是說明性的，且不限制本文描述的系統及方法在具有不同元件及組態之計算系統中之實施態樣，因為本文描述的概念可在具有各種元件及組態之各種計算環境中實施。

資料通訊架構

第2圖描述了用在一示範性計算環境中的一說明性資料通訊架構200之一方塊圖。該說明性資料通訊架構可作為該計算環境之元件被實施且可使用串列器及解串列器(SERDES)元件。如第2圖所示，資料通訊架構200包含經由實體鏈結220協作傳遞資料230之節點205及210。節點205及210是資料異動源，諸如快取-處理器介面(CPIs)及輸入/輸出(I/O)介面之根聯合體(root complex,RC)。實體鏈結220經

由實體連接器225附接到節點205及210。

在操作中，該示範性計算環境(圖未示)與節點205及210協作，以在該等節點之間傳遞資料。在該說明性實施態樣中，該等節點可位於不同的位置，諸如該示範性計算環境(圖未示)內的不同的系統板或抽屜，或可位於示範性計算環境之系統板(圖未示)之一的一部分。如圖所示，資料可在該等節點之間以一特定方向被傳遞，如實體鏈結220及資料230上的箭頭所指示。而且，明顯的是，實體鏈結220被繪示為具有不同的線厚度以指示不同的實體鏈結220介質。

而且，如圖所示，虛線框215顯示了節點205、210之間的兩個通訊通道之建立。在提供的實施態樣中，虛線框215被顯示為包含一對操作以傳遞資料之發射-接收核心。特定地，資料藉由節點205之發射核心235處理以經由實體連接器225及實體鏈結220傳遞到節點210之接收核心245。類似地，資料藉由節點210之發射核心250處理以傳遞到節點205之接收核心240。該等通訊通道中的一個通道是一請求通道，透過該請求通道來請求資料，另一通道為一回應通道，透過該回應通道提供請求的資料。在一示範性實施態樣中，該發射-接收核心對可被對齊且被訓練以依據諸如8位元-10位元(8b10b)編碼之一已選定的串列編碼協定來處理資料。

視該系統之需要而定，每一節點可作為一請求者或一回應者而動作。而且，如第2圖中所示，資料230可包含多個微封包。特定地，資料230可包含一標頭部分及資料部分。

在操作中，資料異動及用於根據本文描述的方法及系統管理傳遞該等資料異動之潛伏期之控制訊息可作為資料230透過示範性資料通訊架構200來通訊。

可分區電腦系統

一多處理器計算系統可受組配為一單一操作環境，或可被分成多個獨立的操作環境。在此脈絡中，操作環境意味著獨立的硬體及軟體，其中每一分區被分配供其自己使用之記憶體、處理器及I/O資源且執行其自己的作業系統影像。分區可以是用於定界一單一系統內獨立的操作環境之實體的或邏輯的機制，或可以包含一單一操作環境內的多個獨立的操作環境。分區允許協調組配及管理大量計算資源，響應於需求的變動分配計算資源，最大化資源利用，且能夠使在一個分區中發生的破壞事件免於不利地影響其它分區。

參考第3A圖，一計算系統之一部分可包含多個異動來源節點304。在第3A圖中，只出現了四個節點304A到304D。該等節點可全部位於一單一操作環境中或它們可位於一可分區電腦系統之兩個或更多分區內。每一節點304可經由可路由資料封包之一路由裝置312(諸如一縱橫交換器)與其它節點通訊。該路由裝置312有助於把封包從一來源位址傳送到一目的位址。例如，對於節點304A發送一封包到節點304D來說，節點304A發送該封包到該路由裝置312，該路由裝置312轉而發送該封包到節點304D。在此脈絡中，該路由裝置可被稱為一交換結構。

在諸如第3B圖中顯示的系統之一較大的可分區電腦系統中，可以有不止一個路由裝置312。例如，第3B圖顯示了具有四個路由裝置312A、312B、312C及312D之一系統，雖然明顯的是可使用其它類型及/或數目之路由裝置。在此，該等路由裝置312可全體被稱為一交換結構。該等路由裝置312可彼此通訊且可與多個節點304通訊。例如，節點304A、304B、304C及304D可直接與該路由裝置312A通訊。節點304E、304F、304G及304H可直接與路由裝置312B通訊。節點304I、304J、304K及304L可直接與該路由裝置312C通訊。節點304M、304N、304O及304P可直接與路由裝置312D通訊。在這樣一組態中，每一路由裝置312及與該路由裝置312直接通訊之該等節點304可受組配以包含在一獨立的分區中，如虛線316所指示。如圖所示，在第3B圖中有四個分區316A、316B、316C及316D。如圖所示，每一分區包括四個節點；然而，任意個節點及節點之組合可被包括在一分區中。例如，分區316A及316B可被重新組配及組合以形成一個包含全部8個節點之分區。在另一範例中，一系統可受組配以包括節點304A、304B、304M及304N於一個分區中，且剩餘的節點在一個或多個其它分區中。而且，分區可響應於該系統之變化的需要而被創建、消除及/或動態地重新組配。

雖然以示範性組態來顯示，但節點及分區之組織不受限於這樣的組態中。更確切地說，顯示的組態只是說明，且依據申請專利範圍之元件之組態不是為了受所提供描述

之限制。

封閉(Wall-Up)逾時架構

計時器藉由係為發送到該系統結構的異動來源之一個或多個節點來維持。該等計時器被用以在該系統中建立一"封閉"模式，該模式可停止或放緩自所有來源到該結構之新的異動之發送，直到所有未處理的連貫異動被完成。該封閉模式減輕該結構上的訊務擁塞，藉此由於該擁塞而處於未完成之危險中的一異動可被完成。該異動完成後，該等節點返回正常操作。

第4圖顯示了在封閉架構中發送之訊息之一示意圖。在一示範性實施例中，把連貫異動引入到該結構上之該等來源節點(包括快取-處理器介面(CPI)及輸入/輸出(I/O)介面之根聯合體(RC))中的一些或全部節點為其所發起的每一未處理的連貫異動維持一"封閉"計時器。發起一特定連貫異動之一來源在此被稱為該異動之"發起者"，及該異動指向之節點在此被稱為目標。如果與該異動相關聯之封閉計時器達到一選定或預定的臨限，該異動被視為由於該結構上之擁塞而處於未完成之危險中。此異動之發起者這裡被稱為一"受害者"。為了及時完成該異動，決定該受害者需要提高對該目標之接取，諸如藉由獲得額外的系統頻寬接取該目標。為了獲得額外的系統頻寬，該受害者藉由發送一PWALLUP訊息(步驟410)到該等來源中之一個或多個且較佳地全部來源，諸如藉由多播該PWALLUP到該等來源，以喚起該封閉模式(升起該封壁(wall))。在一可選擇的實施

態樣中，視一個或多個選擇參數而定，該受害者可請求另一來源節點代表其升起該封壁。例如，該系統可受組配使得CPI能夠與所有可能的封閉參與者通訊，而RC不能。在這樣一組態中，係為一RC之一受害者可發送一訊息到一CPI以代表其升起該封壁。要明白的是，其它的組配是可能的。而且，如果該電腦系統是可分區的，且如果該目標與該受害者在同一分區中，該受害者可受組配以只向在該分區中的來源節點發送一PWALLUP訊息。該受害者記錄該PWALLUP訊息被發送到之該等節點。

回應於接收到該PWALLUP訊息，該等來源節點停止發起新的異動，但繼續處理正在進行的及新接收到的異動(步驟420)。此外，使該等來源節點無法開始它們自己的新的封閉序列。在一可選擇的實施態樣中，一個或多個來源節點可根據選擇參數繼續發起新的異動。例如，一個或多個RC可被選擇以繼續使其能夠發起新的異動，諸如達一選定的有限時間。在另一實施例中，一個或多個CPI可類似地被選擇以繼續使其能夠發起新的異動，諸如達為該等CPI選定的一有限時間內。在又一實施例中，包括一個或多個RC及CPI之來源節點之一子集可被選擇以繼續使其能夠發起新的異動，達相同的時間量內或分別為該等RC及該等CPI選定的不同的時間。應當明白的是，來源及各自參數之其它組合也是可能的。

該等CPI可藉由使它們相關聯的處理器(包括該發起者，如果它是一CPI的話)靜止而停止新的連貫訊務之持續

產生。RC藉由暫停對來自相關聯的I/O介面(再一次包括該發起者，如果它是一RC的話)之新的異動之接收而停止新的異動之持續發生。在這兩種情況下，該等來源節點繼續處理在升起該封壁之前已發起的或已經被該等來源節點接收到的該等異動。在一示範性實施態樣中，當諸如一RC之一來源成為一受害者(升起該封壁之該受害者或具有當該封壁升起時已達到一臨限之一計時器之一RC)時，該RC可繼續發起新的訊務，直到一安全計時器達到其臨限。一CPI可類似地動作。為了防止該封壁停留時間過長，可包括一與那些額外的已經接收到的異動之處理相關聯之安全計時器以確保它們被及時處理。

為了保證自該受害者到該目標之該異動之及時完成，必須決定的是，該異動已取得可接受的進展，例如，進展到一選定點，諸如進展到完成。這可藉由確認包括該受害者在內的所有來源節點已經完成它們各自的未處理異動來決定。在示範性實施例中，一旦接收到一PWALLUP之每一來源節點已經完成其所有的未處理連貫異動，其藉由發送一PWALLCMP訊息來回應該受害者(步驟430)。該受害者記錄從其等接收一PWALLCMP訊息之該等節點。一旦該受害者已接收到來自該受害者已發送一PWALLUP訊息到其之每一來源之一PWALCMP(步驟440)，且也已經完成其自己所有的未處理異動，該受害者發送一PWALLDN訊息到相同的來源節點(步驟450)。這些來源節點可接著返回到正常操作(步驟460)。可使該等來源節點能夠開始新的封閉序列。

可取捨地，(如第4圖中用虛線說明的隨後的步驟所示)，該等來源節點可發送新的異動到該結構，但仍不能啓動一新的封閉序列。在此情況下，當返回到正常操作之後該等來源節點藉由發送一第二PWALLCMP訊息來回應於該受害者(步驟470)。該封閉計時器尚不重置，且任何新接收到的PWALLUP應當被記住(例如，儲存於快取中)但尚不執行動作。

接收到針對其所有的未處理PWALLDN之該等第二PWALLCMP(步驟480)之後，該受害者可發送一第二序列之PWALLDN(步驟485)作為一機制以向該等來源節點指示它們現在可以重新致能對它們的異動檢查可能發送新的PWALLUP之時期。

一旦該第二PWALLDN已經遭接收到，接收到一PWALLUP之每一來源節點返回到全面正常操作(步驟490)。新的PWALLLUP被致能，即，該等來源可發送一新的PWALLUP或依據一新的PWALLUP而動作。該封閉計時器獲重置，且該等來源發送最後的PWALLCMP到該受害者(步驟495)。

可取捨地，該受害者可受組配以藉由利用該結構發送PWALLUP、PWALLCMP及PWALLDN訊息到其本身來經由該WALL算法管理其自身的進展。

可能的是，多個節點會實質上同時發送PWALLUP訊息。在此情況下，一個被指定為“主控器”。應當明白的是，這可以以各種方式完成，例如，藉由簡單比較發送該等

PWALLUP之該等來源節點之ID且選擇最低的一個作為該主控器。該系統中的接收來自不同受害者之多個PWALLUP之來源節點可藉由一PWALLCMP來回應每一受害者。在一實施態樣中，來源節點可受組配以及早回應不是發自該主控器之PWALLUP。

在該示範性實施例中，已經開始一WALLUP序列且又接收到來自具有一較高編號ID之另一節點之一PWALLUP訊息之一受害者應當藉由一PWALLCMP,M訊息回應那個較高編號節點以指示其(該較低編號受害者)將是該主控器。這確保了該較高編號節點在其可能決定已準備好發送一PWALLDN之前，察覺到該主控器之PWALLUP。

接收到一PWALLUP,M(其將來自具有一較低編號ID之一節點)之一受害者必須把該封閉模式之控制權讓給該主控器。這樣做，該受害者可繼續發送PWALLUP以完成它的封閉序列。如果這樣，該受害者還應當收集PWALLCMP，這樣它可辨識出什麼時候它的異動離開該結構，且可發送其所欠任何其它的非主控器受害者之PWALLCMP。然而，它將等到其首先接收到其自己的所有PWALLCMP時，發送一PWALLCMP到該主控器且其不發送任何PWALLDN訊息。在一時間只有一個主控器被認可。

在一實施態樣中，一系統狀態機可具有一“解封閉(wall-down)”計時器，當達到一選擇的臨限之後，降下該封壁(每一異動來源、發起者或目標)。這可用以在該封壁升起時，如果一作業系統崩潰，降下該封壁。此計時器之目的

是允許當崩潰時收集關於該系統狀態之資訊，因此該計時器應無關於該OS。在此實施態樣中，PWALLUP及PWALLDN訊息也應當被發送到該狀態機。該解封閉計時器在接收到一PWALLUP時開始計時且當接收到該第一PWALLDN時重置。

可對該等揭露的實施例做各種修改及變化而不脫離本發明之精神及範圍。因此，目的是，本揭露之修改及變化應受到保護，條件是它們在後附申請專利範圍及其等效物之範圍內。

【圖式簡單說明】

第1圖是根據本文描述的系統及方法之一實施態樣之一示範性計算環境之一方塊圖；

第2圖是顯示一示範性資料通訊架構之示範性元件之協作之一方塊圖。

第3A圖及第3B圖是各種可分區計算系統之方塊圖，其中可使用本文描述之系統及方法之原理；及

第4圖是根據本文描述的系統及方法之一實施態樣，在管理通訊資料中的潛伏期時藉由一示範性資料通訊架構執行的處理之一方塊圖。

【主要元件符號說明】

100...示範性計算系統、資料處理系統、計算環境、電腦系統

105...記憶體控制器

110...隨機存取記憶體 (RAM)

115...唯讀記憶體 (ROM)

120...網路配接器

125...資料傳送路徑、系統資料傳送路徑

130...中央處理單元、主要CPU、通用CPU

140...共處理器

145...周邊設備控制器

150...列印機

155...鍵盤

160...滑鼠

165...資料儲存裝置

170...顯示器

175...顯示器控制器

180...計算應用程式

185...外部通訊網路

200...說明性資料通訊結構

205、210、304A、304B、304C、304D、304E、304F、304G、304H、304I、304J、304K、304L、304M、304N、304O、304P...節點

215...虛線框

220...實體鏈結

225...實體連接器

230...資料

235、250...發射核心

240、245...接收核心

304...異動來源節點、節點

312、312A、312B、312C、312D...路由裝置

316...虛線

316A、316B、316C、316D...分區

405~495...步驟

201017434

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：98132061

※申請日：98-9-23　　※IPC 分類：G06F15/163 (2006.01)

一、發明名稱：(中文/英文)

管理多處理器互連體中潛伏期的技術
MANAGING LATENCIES IN A MULTIPROCESSOR INTERCONNECT

二、中文發明摘要：

在一計算系統中具有發送異動到一交換結構之多個異動來源節點，一所獲服務不足的節點通知該系統中的來源節點其需要額外的系統頻寬以及時完成一正在進行的異動。該等獲通知的節點繼續處理已經開始的異動直至完成，但停止引入新的訊務到該結構中直到該所獲服務不足的節點指示其已經進展到一預選定點之時。

三、英文發明摘要：

In a computing system having a plurality of transaction source nodes issuing transactions into a switching fabric, an underserviced node notifies source nodes in the system that it needs additional system bandwidth to timely complete an ongoing transaction. The notified nodes continue to process already started transactions to completion, but stop the introduction of new traffic into the fabric until such time as the underserviced node indicates that it has progressed to a preselected point.

七、申請專利範圍：

1. 一種在包含能夠透過一系統交換結構藉由引入異動到該結構上而互相通訊之多個來源節點的一電腦系統中管理節點間通訊之潛伏期之方法，該方法包含以下步驟：

發起從一發起者來源節點到一目標來源節點之一第一異動之通訊；

決定該發起者節點需要額外的系統頻寬至該目標節點以及時完成該第一異動；

通知多個來源節點該發起者節點需要額外的系統頻寬；

回應於接收到該通知，停止該等獲通知節點中的至少一些節點引入新的異動到該系統結構上；

繼續從該發起者節點到該目標節點之該第一異動之通訊；

決定該第一異動已進展到一預選定點；

向該等獲通知節點指示該發起者節點不再需要額外的系統頻寬；及

回應於接收到該指示，終止停止該等獲通知節點引入新的異動到該系統結構上，

藉此，從該發起者節點到該目標節點之該第一異動之該通訊之該潛伏期獲管理。

2. 如申請專利範圍第1項所述之方法，其進一步包含，與該發起通訊步驟一起開始一異動計時器，其中決定該發起者節點需要額外的系統頻寬至該目標節點之該步驟

包含，決定該異動計時器已經達到一選定臨限。

3. 如申請專利範圍第1項所述之方法，其中該通知步驟包含，該發起者節點發送一訊息(PWALLUP)到該等多個來源節點，指示它們應當停止引入新的異動到該系統結構上。

4. 如申請專利範圍第1項所述之方法，其中決定該第一異動已經進展到一選定點之該步驟進一步包含，決定所有該等獲通知節點之異動已經進展到各自的選定點。

5. 如申請專利範圍第4項所述之方法，其進一步包含，由該發起者接收來自各該獲通知節點之一訊息(PWALLCMP)，指示該等獲通知節點之異動已經進展到各自的選定點。

6. 如申請專利範圍第1項所述之方法，其中該電腦系統是一可分區電腦系統，

其中該通知步驟包含，該發起者節點多播一訊息(PWALLUP)到該分區之所有來源節點以指示該等節點應當停止引入新的異動到該系統結構上，及

其中決定該第一異動已經進展到一選定點之該步驟包含，接收來自各該獲通知節點之一訊息(PWALLCMP)，指示該等獲通知節點之異動已經進展到各自的選定點。

7. 如申請專利範圍第6項所述之方法，其中指示該發起者節點不再需要額外的頻寬之該步驟包含，多播一訊息(PWALLDN)到該分區中的所有該等來源節點，指示該發起者節點已經接收到來自各該來源節點之一

PWALLCMP。

8. 如申請專利範圍第7項所述之方法，其進一步包含，該等來源節點回應於接收到它們各自的PWALLDN而返回正常操作。

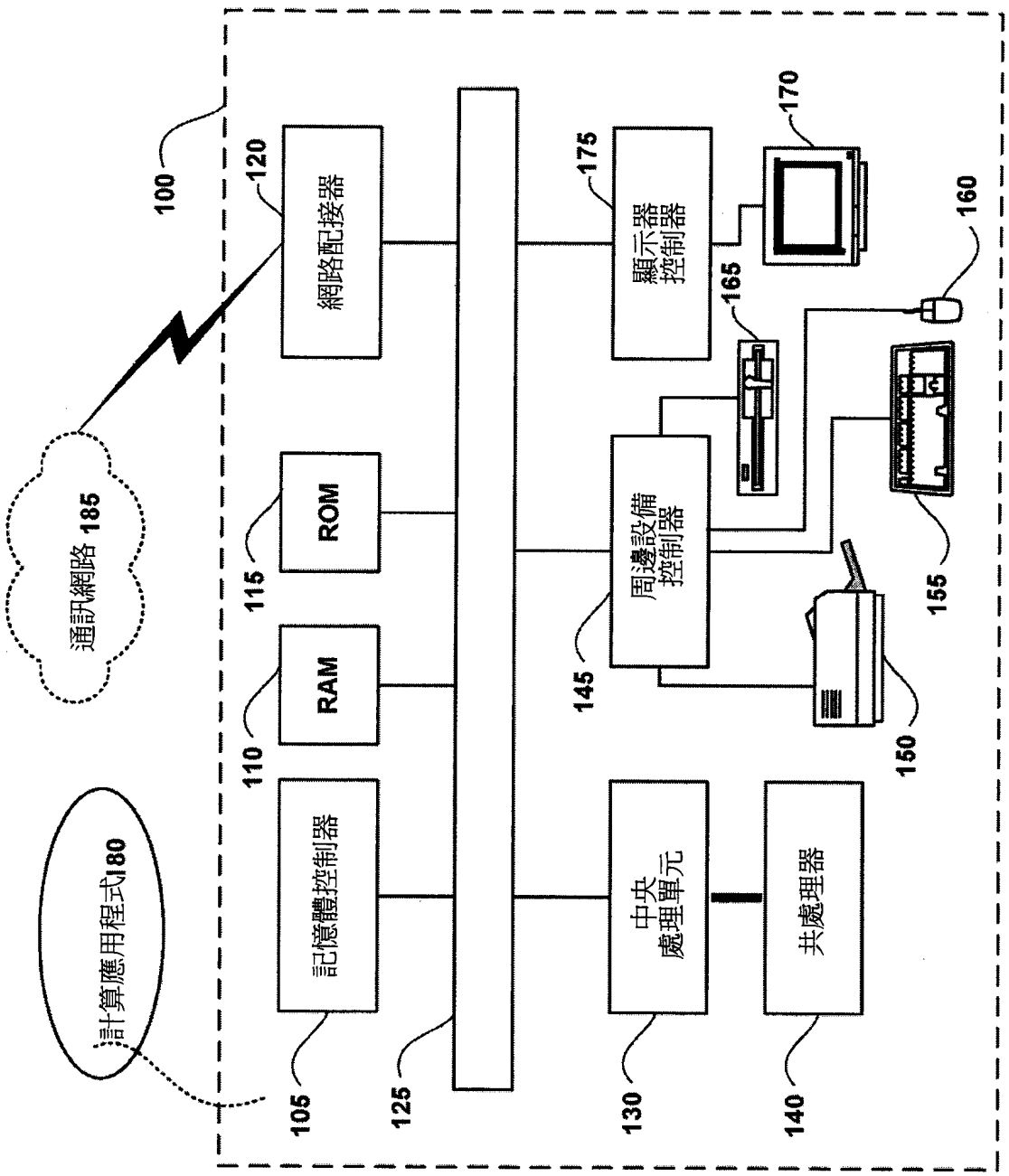
9. 如申請專利範圍第8項所述之方法，其進一步包含：

當該第一發起者多播該第一PWALLUP時，一第二發起者多播一第二PWALLUP到該分區中的所有該等來源節點；及

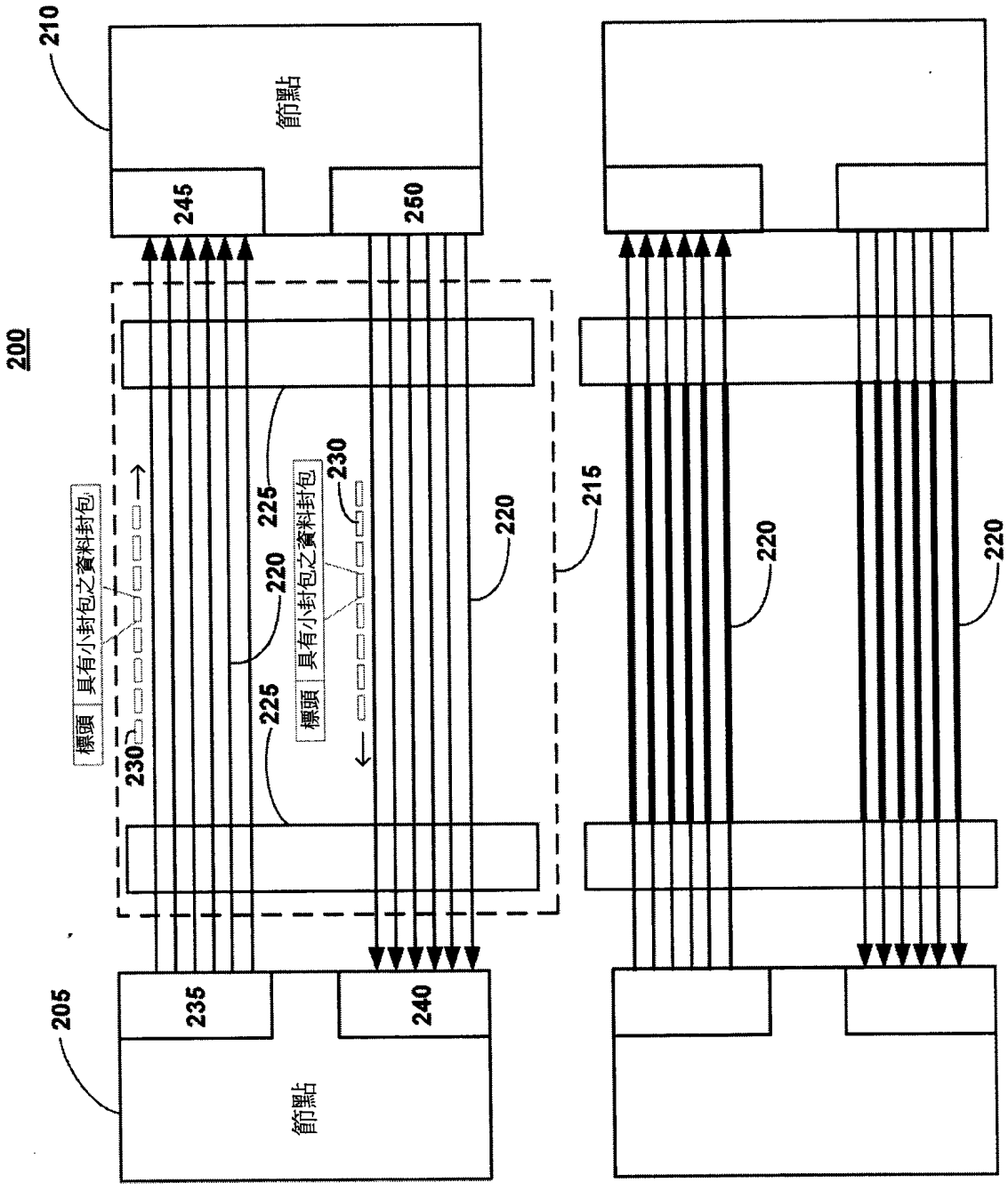
選擇該等發起者中的一個發起者作為主控器且將該選擇通知另一個、非主控器發起者，

其中該PWALLDN由該主控器發送，且該非主控器發起者不發送任何PWALLDN。

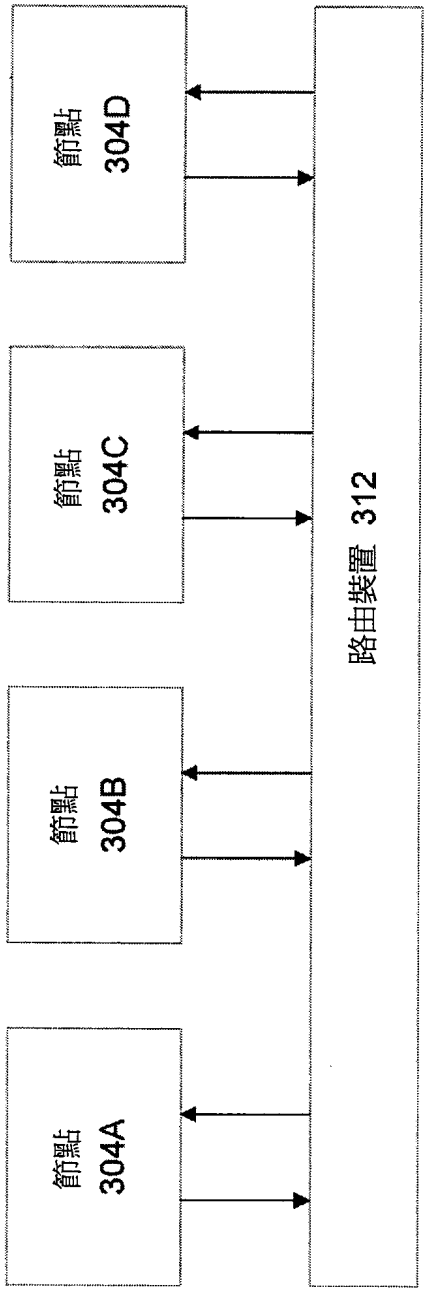
10. 如申請專利範圍第9項所述之方法，其進一步包含，該非主控器發起者等到其已接收到其自己的所有PWALLCMP時發送一PWALLCMP到該主控器。


100
120
網路配接器
通訊網路 185
計算應用程式 180
115
ROM
110
RAM
記憶體控制器
105
125
175
顯示器
控制器
170
165
160
145
周邊設備
控制器
155
150
130
中央
處理單元
140
共處理器

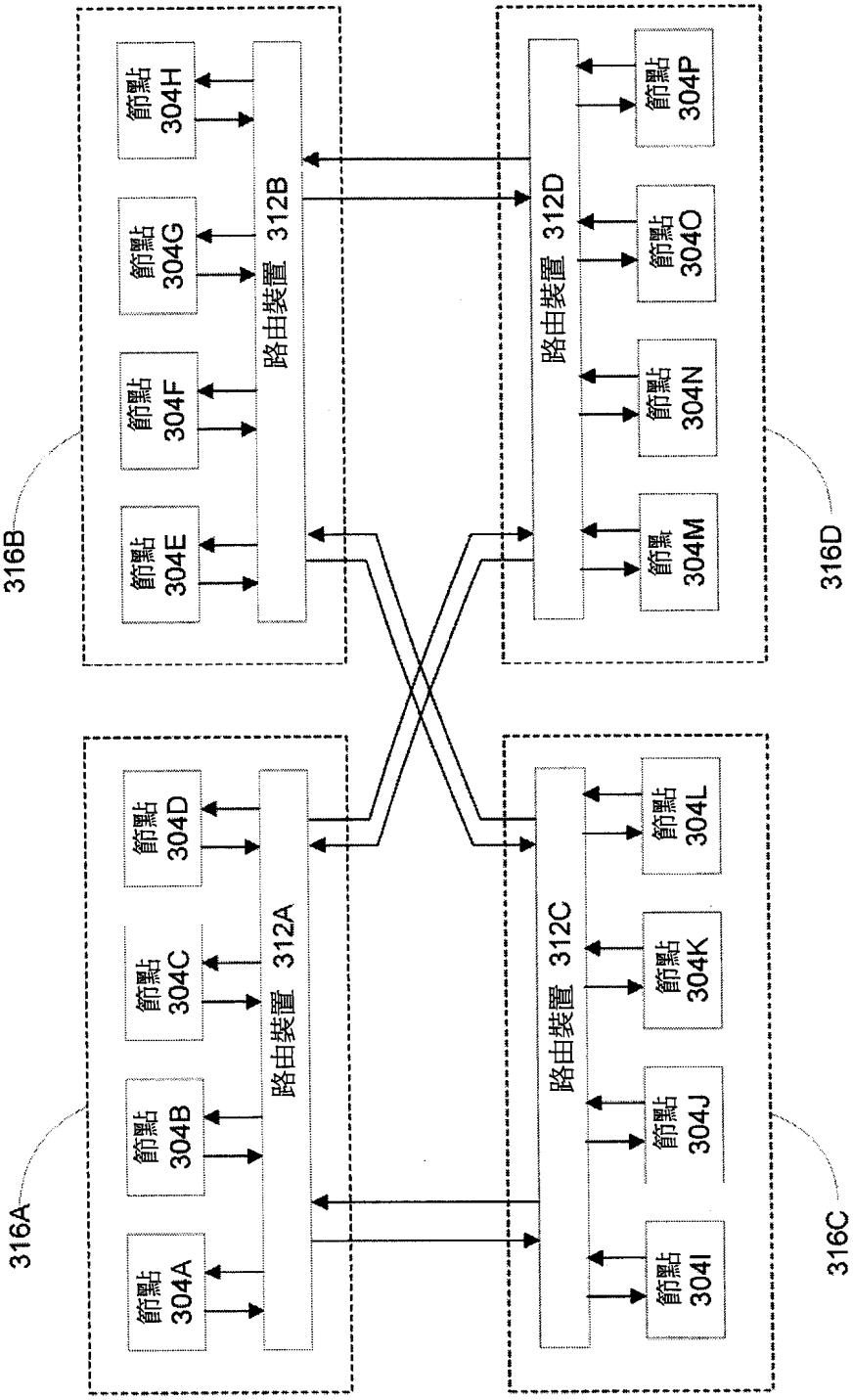
第1圖



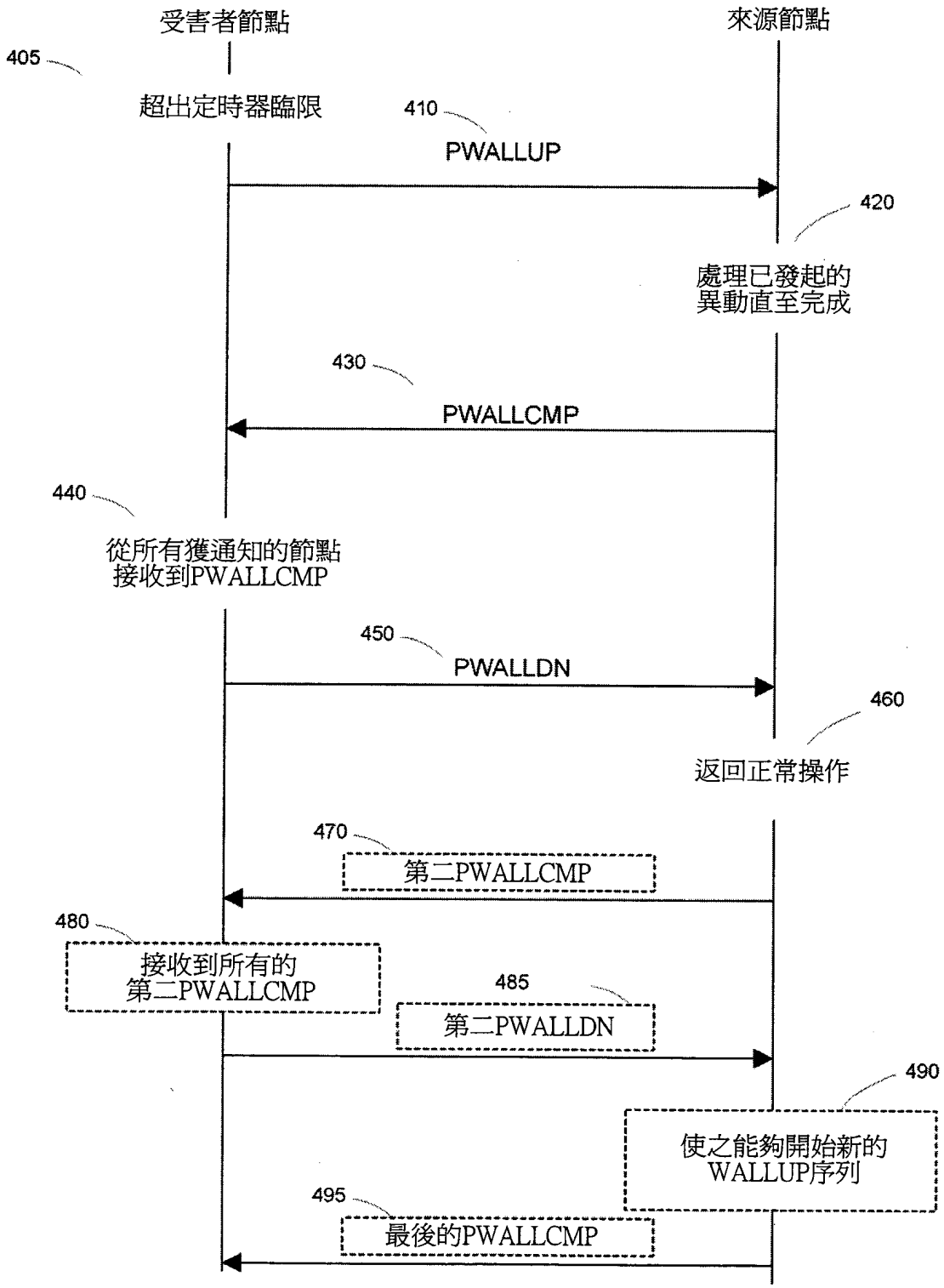
第2圖



第3A圖


316B
316D
316A
316C
路由裝置 312B
路由裝置 312D
路由裝置 312A
路由裝置 312C
節點 304H
節點 304G
節點 304F
節點 304E
節點 304P
節點 304O
節點 304N
節點 304M
節點 304D
節點 304C
節點 304B
節點 304A
節點 304L
節點 304K
節點 304J
節點 304I

第3B圖


受害者節點
來源節點
405
超出定時器臨限
410
PWALLUP
420
處理已發起的
異動直至完成
430
PWALLCMP
440
從所有獲通知的節點
接收到PWALLCMP
450
PWALLDN
460
返回正常操作
470
第二PWALLCMP
480
接收到所有的
第二PWALLCMP
485
第二PWALLDN
490
使之能夠開始新的
WALLUP序列
495
最後的PWALLCMP

第4圖

四、指定代表圖：

(一)本案指定代表圖為：第（4）圖。

(二)本代表圖之元件符號簡單說明：

405~495...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：