

(1)

九、發明說明

【發明所屬之技術領域】

本發明有關於封包聚合的方法與系統。

【先前技術】

網路使電腦與其他裝置能夠通訊。舉例來說，網路能夠運送表示視訊、聲訊、電子郵件等之資料。典型來說，傳送於網路上之資料係藉由稱為封包的較小訊息（message）所承載。以此類推，封包幾乎就像被丟入信箱中之信封。封包典型包括”酬載（payload）”與”標頭”。封包的”酬載”類似於信封內的信。封包的”標頭”幾乎就像寫在信封本身上的資訊。標頭能包括幫助網路裝置適當地處理封包的資訊。

一些網路協定（例如”協定堆疊（a protocol stack）”）合作，以處理複雜的網路通訊。舉例來說，被稱為傳輸控制協定（TCP）的傳送協定提供具有簡單機制應用程式（application），以在網路上建立流（flow）與傳輸資料。事實上，TCP 透通地處理種種的通訊事項，例如資料再傳輸、適應網路流量壅塞等。

為了提供這些服務，TCP 以被稱為段（segment）的封包為單位運作。一般而言，TCP 段行進經過網路，該 TCP 段係在例如網際網路通訊協定（IP）資料報（datagram）的較大封包內所”包住（encapsulated）”。經常地，舉例來說，在區域網路（LAN）中，IP 資料報進一

(2)

步由更大的封包（例如乙太網路幀（frame））所包住。TCP 段的酬載承載為應用程式所運送於網路上的資料串流（stream）之一部份。接收器能藉由重新組合所接收的段，而恢復原始的資料串流。為了准許所接收的資料之重組（reassembly）與應答（ACK）返回傳送端，TCP 將序列號碼相關至各個酬載位元組。

【發明內容及實施方式】

許多應用程式接收與處理大量的網路資料。桌上型電腦的應用程式之範例包括網頁瀏覽器、串流媒體播放器、以及網路檔案分享應用程式。伺服器應用程式包括網頁伺服器、檔案伺服器、儲存伺服器、電子郵件伺服器、以及資料庫後端（back-ends）。典型來說，內涵的（underlying）協定堆疊（例如 TCP/IP 堆疊）接收許多封包並個別地處理，即使部分或全部的這些封包是相同之流的部分。與每個封包的處理相關的是一些處理損耗（overhead），舉例來說，由於分析（parsing）標頭、識別（identifying）與更新流狀態資訊、產生 ACK 訊息等。

第 1A 至 1C 圖係顯示將用於給定流之多封包聚合成單一封包的技術之樣品實施法。顯示於第 1A 至 1C 圖之樣品系統包括處理器 104 與記憶體 102。該系統亦包括網路介面控制器（NIC）（稱為網路介面卡）100，該 NIC 從網路接收封包。除了將所接收的每一封包寫入記憶體 102 內作後續處理外，該控制器 100 以聚合封包的邏輯 112 為特

(3)

徵。此邏輯 112 結合屬於相同流之不同的封包之 TCP 酬載並對於所結合的 TCP 酬載準備單一 TCP 標頭與單一 IP 標頭。IP 標頭、TCP 標頭、以及所結合的 TCP 酬載之結合形成單一聚合的封包。該協定堆疊能因此執行較少但較大的封包的接收處理，降低每個封包所發生之處理懲罰（penalty）。

爲了說明聚合，第 1A 圖顯示由該網路介面控制器 100 所接收之具有 TCP 與 IP 標頭 106a 以及 TCP 酬載 106b 之封包 106。該控制器 100 可執行許多任務，包括將該封包 106 從幀中拆封（de-encapsulating）、驗證幀核對和、以及其他鏈結層操作。

如圖所示，該封包屬於流（第 1A 圖中任意標爲”1”）。封包流能藉由控制器 100 以標頭中的資料所指明。舉例來說，TCP/IP 流能藉由元組（tuple）來指明，該元組係藉由該 TCP 標頭中的 IP 來源與目的地地址以及來源與目的地埠號碼之結合所形成。元組可不包括所有的這些標頭欄且可包括其他資訊（例如協定識別符（identifier））。

於第 1A 圖中，該控制器 100 儲存該所接收的封包 106 之標頭 106a 與酬載 106b，以用來與後來所接收的封包之作可能的聚合。舉例來說，如圖所示，該控制器 100 可儲存該封包 106 之酬載 106b 於記憶體 102（例如透過一個或多個直接記憶體存取（DMA）操作）並儲存該標頭 106a 於控制器 100 表。該表亦可包括其他用於聚合處理的

(4)

資訊。記憶體中用以寫入酬載資料 106b 的位置可由描述符所指定，該描述符係被運作於處理器 104 上之驅動程式所傳送至該控制器 100 上。該描述符亦可包括其他欄（例如用以儲存封包標頭的位置之記憶體位址），以支援標頭切開（header splitting）。

於第 1B 圖中，該控制器 100 接收第二封包 108，該第二封包屬於與第 1A 圖中所接收的封包之相同流（“1”）。除了將該封包簡單寫入記憶體 102 外，該控制器 100 將封包 106 與 108 之兩個酬載 106b、108b 結合在一起成整體的（monolithic）酬載 110b。結合可涉及實際將酬載位元 106b、108b 鄰接地儲存在一起。或者，結合可涉及例如將酬載 106b、108b 相關為在鏈結名單中之節點。此酬載之結合可持續用於此流，所接收之其他封包。

除了收集不同的酬載以外，該控制器 100 亦準備單一 IP 標頭與單一 TCP 標頭 110a 給該聚合的封包 110，該聚合封包 110 反映該結合的 TCP 酬載 110b。舉例來說，該控制器 100 可查詢有關該流之 TCP/IP 標頭 106a 並修改該 IP 標頭的長度欄，以反映該結合酬載之長度。該控制器 100 亦可修正該 TCP 標頭的核對和。此外，該控制器 100 可改變該 TCP 標頭的 ACK 序列號碼，以聚合進來的（incoming）ACK 訊息。此更新可當各個酬載結合時執行。或者，該更新可例如被延遲一段時間。

最後（例如在聚合窗結束後），如第 1C 圖所示，該控制器 100 可將該聚合的封包之標頭 110a 與該流的描述

(5)

符寫入記憶體 102。該控制器 100 接著可發出中斷訊號，以開始該聚合的封包 100 之接收處理（例如網路與/或傳輸層處理）。舉例來說，TCP 接收處理能包括重新組合、重新排序、ACK 的產生、導航（navigating）於流中之 TCP 狀態機等。

聚合之封包的數量與/或用以聚合封包的時間可為可設定的。舉例來說，典型地，網路介面控制器使用一種稱為中斷緩和（interrupt moderation）的技術，以將在一些時間窗（window of time）中所接收的封包以一次一批的方式發出訊號。該控制器能使用中斷緩和窗，以儘可能聚合流之越多封包。為了允許聚合損耗（overhead）（例如標頭準備），該控制器 100 可使用少於中斷緩和窗之時間窗（聚合窗），以聚合封包。在聚合窗期間，該控制器 100 獲得用於流之描述符（該流在聚合視窗期間接收資料）（例如藉由佇列移除（dequeue）由控制器 100 裝置驅動器所提供之描述符），並且大致上保留該描述符直到該聚合窗過期（expire）或該控制器 100 接收不符合聚合標準（將說明於後文）之流封包，或是該酬載之尺寸超過該描述符所指明之封包緩衝器中可用的空間為止。在該聚合窗過期後，該控制器 100 準備標頭、將描述符寫入記憶體、在中斷緩和時間結束時發出中斷訊號、以及清除在前面的窗期間用以聚合封包的資料。該聚合處理接著開始新的一個處理。

為了簡明說明的目的，顯示於第 1A 至 1C 圖的系統不

(7)

206 可以可程式化引擎為特徵，該引擎執行指令以處理封包。如圖所示，該電路 206 經由 DMA 控制器 210 介接至主系統。

該聚合電路 206 可以用許多方式來實施聚合。舉例來說，如第 3 圖所示，該電路 206 可建立一表 212，該表 212 可追蹤正在進行的 (on-going) 聚合。如圖所示，此表 212 可將一流 ID (例如 TCP/IP 元組或 TCP/IP 元組的雜湊 (hash)) 與封包的開始位元組序列號碼、酬載位元組的號碼、封包描述符的位址、酬載緩衝的位址、以及標頭緩衝的位址相關。該表 212 可儲存其他資料 (未圖示)，例如用於該流的標頭欄。舉例來說，該表 212 可儲存 IP 來源、IP 目的地、IP 識別碼與版本、IPv6 流 ID 與優先權、TCO 來源埠、TCP 目的地埠、TCP 序列號碼、TCP ACK 號碼、TCP 核對和、與 / 或 TCP 時間戳 (timestamp)。該表 212 亦可記錄對於流進行聚合，以稍後將此資訊傳遞至 TCP/IP 堆疊的封包之數量 (例如經由描述符中的一欄)、聚合的 ACK 段之數量，並可儲存老化 (aging) 計數器，以支援 "描述符老化 (descriptor aging)" (將說明於後文)，其被用以在聚合窗結束之前關閉閒置的描述符。

用於給定流的該表 212 資料被修改為聚合程序。舉例來說，可調整位元組的數量，以反映最近結合的酬載之其他位元組。同樣地，聚合酬載之數量可隨每個其他 TCP 酬載之結合而增加。該表 212 資料能用以準備用於聚合的封

(8)

包之標頭與準備相對應的描述符。同樣地，該表 212 資料可在例如聚合窗結束之後清除。

該控制器可包括其他元件（未圖示）。舉例來說，該控制器可包括暫存器，該暫存器致能（enable）例如驅動器，以致能或去能（disable）該聚合。

第 4 圖係顯示聚合封包之過程流程圖。如圖所示，該過程在 256 結合相同流中之封包的酬載，並在 258 準備單一 TCP 段標頭與單一 IP 標頭給結合的酬載。藉由 TCP/IP 堆疊，接著可產生中斷，以起始該聚合的封包之處理。

如圖所示，在 254 某些封包不考慮聚合。舉例來說，封包需要滿足一個或多個標準。舉例來說，聚合可僅執行於 TCP 段具有有效的核對和時。此外，根據標頭資訊（例如指明該段為控制段（例如 RST、FIN、SYN、SYN-ACK、URG 旗標設定）之資訊），甚至有效 TCP 段亦可排除與前一接收封包聚合。在這些情況中，對於此流的先前之正在進行的聚合可終止（例如可準備 IP 與 TCP 標頭並寫入記憶體給任何先前結合的流酬載與寫入的相對應之描述符資料）。

某些情況中，TCP/IP 封包可能不按照次序接收（亦即接收封包的序列號碼不匹配流之下一個連續的序列號碼）。在此情況中，可開始新的聚合封包（例如獲得的描述符與寫入的表項目（table entry））。亦即，給定的流可在流的位元組序列中之多個點具有進行中的（in-progress）聚合。此後，對於給定流，流封包的酬載可根

(9)

據接收的封包序列號碼而增加至許多進行聚合之封包的其中一個。或者，爲了簡明起見，對於流的先前之正在進行的封包聚合可在封包不按照次序接收之後終止。

其他配置也可影響封包聚合。舉例來說，若封包的 TCP 標頭表示 "PUSH" 旗幟已設定，則對於此流的聚合可在聚合接收的封包之後完成，而對於此流之後續封包將使用新的描述符加以聚合。同樣地，若進來的封包之酬載的聚合超過指定緩衝器中之可用空間，則控制器能終止（例如產生單一 TCP 與單一 IP 標頭並寫入相對應的描述符）目前之正在進行的聚合並對於該流重新開始新的聚合（例如寫入新的表項目並獲得新的描述符）。

第 5 圖係顯示封包聚合的樣品實施法。於所示之實施中，若封包 300 在 302 爲 IP 資料報（例如 IPv4 或 IPv6 資料報）或包住 IP 資料報之幀，則在 304 檢驗該 IP 標頭是否有標頭選項（option）與/或分離（fragmentation）。若有任一種情況存在，則在 308 聚合不會發生，而該封包可照慣例般處理（例如獲得描述符、寫回、以及將該封包作 DMA 入記憶體）。否則，該處理在 306 試圖證實該 TCP 段在該 IP 封包中是否有效（例如藉由決定是否該 TCP 段標頭核對和爲有效的）。若該 TCP 段不是有效的，同樣地，在 308 沒有聚合發生於該封包。

對於有效的 TCP 段，該處理在 310 決定流 ID，例如根據該封包的 TCP/IP 元組。若在 312 該 TCP 段爲資料段（例如 `IPheader.total_len - IPheader.header_len -`

(10)

TCPheader.Data_Offset>0)，則在 314、316 檢驗該 TCP 段標頭是否有時間戳選項以外的選項、以及是否有 ACK 與 / 或 PSH 以外的選項。若在 312、314、316 有這些情況的任何一種情況存在，則在 308 沒有聚合發生。此外，若聚合已經開始於該流，則在 332 藉由產生該 TCP 與 IP 標頭、停止用以對於該流聚合封包之該描述符、以及使該流的表項目無效，而使現存的聚合暫停 (halt)。

假設滿足情況 302、304、306、312、314、316，該處理在 320 決定是否該流已經執行聚合。若否，且該 TCP PSH 旗標未被設定，則該處理能（在 326 表空間允許）對於該流初始化 (initialize) 表輸入、讀取描述符、以及在 330 開始對於該流聚合目前的封包。若在 326 中該表不存在足夠的空間以用於額外的項目，則可犧牲先前寫入的項目（未圖示），舉例來說，使用最近最少使用 (Least Recently Used) 演算法，以選擇一項目來刪除以及關閉相關的描述符。

若在 320 聚合已經建立於此流，則該處理能根據其序列號碼而在 324 決定是否該 TCP 段按照次序接收。若在 324 該段沒有按照次序接收，則在 332 可終止對於該流之正在進行的聚合。若在 324 該段按照次序取回且在 334 該酬載緩衝具有足夠的空間以用於額外的 TCP 酬載，則該處理能藉由在 336 以決定的偏移量 (offset) 在 328 複製該酬載資料至由該流的描述符所指定的酬載緩衝器中以及更新用於該流的項目資料（例如更新聚合的封包之數量、下

(11)

一個預期要發生的序列號碼、酬載位元組的數量等），而結合該接收的 TCP 段之酬載與先前接收的 TCP 段之酬載於該流中。若在 338 設定用於目前的段之該 PSH 旗標，則在 342 聚合可在這些操作後終止。

若在 334 中用以儲存酬載之緩衝器沒有足夠的空間，以加入目前的封包之 TCP 酬載，則可在 340 準備該 TCP 與 IP 標頭並關閉流描述符。於此情況中，若在 346 設定該 PSH 旗標，則該封包按照慣例來處理。否則，獲得新的讀取描述符以用於該流，且在 348 再一次開始聚合此封包。

在處理該封包後，若聚合窗過期，則準備封包描述符與標頭並寫入記憶體，且清出（flush）表內容。否則，重複第 5 圖所示之處理以用於另一個封包。

可實施如第 5 圖所示之許多不同的變化之樣品處理。舉例來說，為了防止未完成的描述符阻礙較晚的描述符（例如若 NIC 驅動器依序地讀取描述符），即使該聚合窗還沒有過期，該處理能在若干固定時間後關閉老化的描述符，而沒有接收其他之後續封包。或者，當較晚的描述符完成時，可關閉較早的描述符。

儘管第 1 至 5 圖與對應的本文說明簡單的實施，許多的其他實施可使用一個或多個的上述之技術。舉例來說，代替聚合該封包於記憶體，該控制器可在傳輸封包至記憶體之前，聚合封包於其擁有的內部緩衝中。此外，該技術可用以實施其他傳輸層協定、於網路協定堆疊中之其他層

(12)

中的協定、TCP 與 IP 以外的協定，且用以處理其他協定資料單元。舉例來說，代替乙太網路幀，該封包可由 HDLC 或 PPP 幀運送。此外，專業用語”IP”包含 IPv4 與 IPv6 兩者的實施。

此處所使用的專業用語”電路”包括硬體連接電路、數位電路、類比電路、可程式化電路等。該可程式化電路可運作於配置在製造物品上之可執行指令（例如非揮發性記憶體，如唯讀記憶體）。

其他實施例係在後述之申請專利範圍之範疇內。

【圖式簡單說明】

第 1A 至 1C 圖係顯示封包聚合的範例；

第 2 圖係網路介面控制器的示意圖；

第 3 圖係由網路介面控制器所使用之用以聚合封包的表；以及

第 4 與第 5 圖係顯示封包聚合的流程圖。

【主要元件符號說明】

100：控制器

102：記憶體

104：處理器

106：封包

106a：TCP 與 IP 標頭

106b：酬載

(13)

108：第二封包

108b：酬載

110：聚合的封包

110a：單一 IP 標頭與單一 TCP 標頭

110b：整體的酬載

112：邏輯

200：網路介面控制器

202：實體層裝置

204：媒體存取控制器

206：聚合電路

210：DMA 控制器

212：表

五、中文發明摘要

發明之名稱：封包聚合的方法與系統

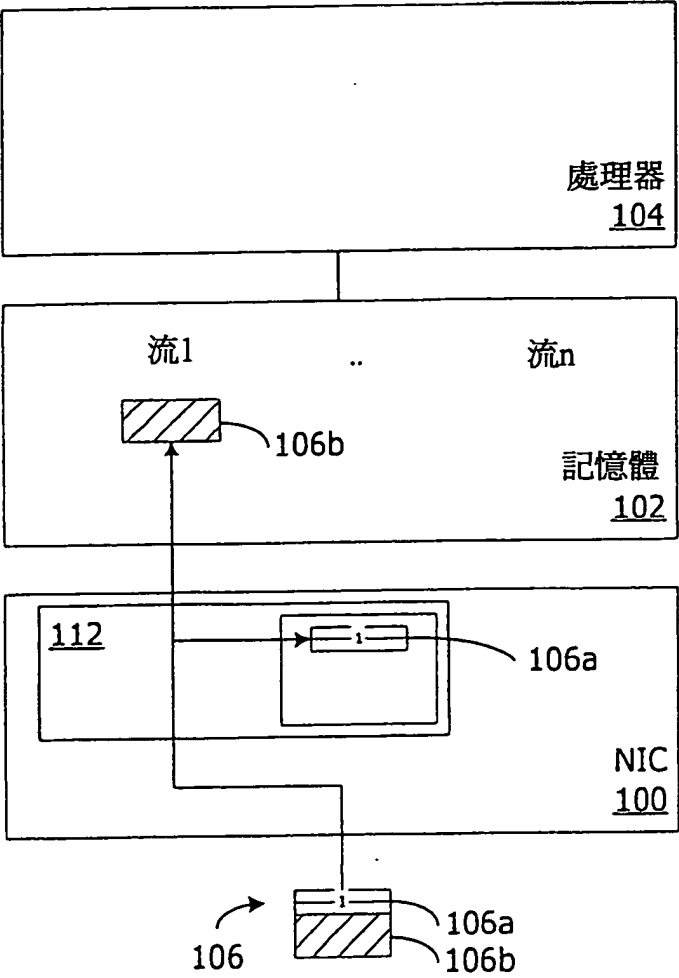
一般說來，以一個觀點來看，本發明揭露一種方法，包括接收多進入網際網路協定封包的步驟，各多進入網際網路協定封包具有網際網路協定標頭與具有傳輸控制協定標頭與傳輸控制協定酬載之傳輸控制協定段，其中該等多封包屬於相同的傳輸控制協定/網際網路協定流的步驟。該方法亦包括準備具有單一網際網路協定標頭之網際網路協定封包與具有單一傳輸控制協定標頭與單一酬載之單一傳輸控制協定段的步驟，該單一酬載係藉由結合該等多網際網路協定封包之傳輸控制協定段酬載而形成。該方法更包含產生訊號的步驟，該訊號引起該網際網路協定封包之接收處理。

六、英文發明摘要

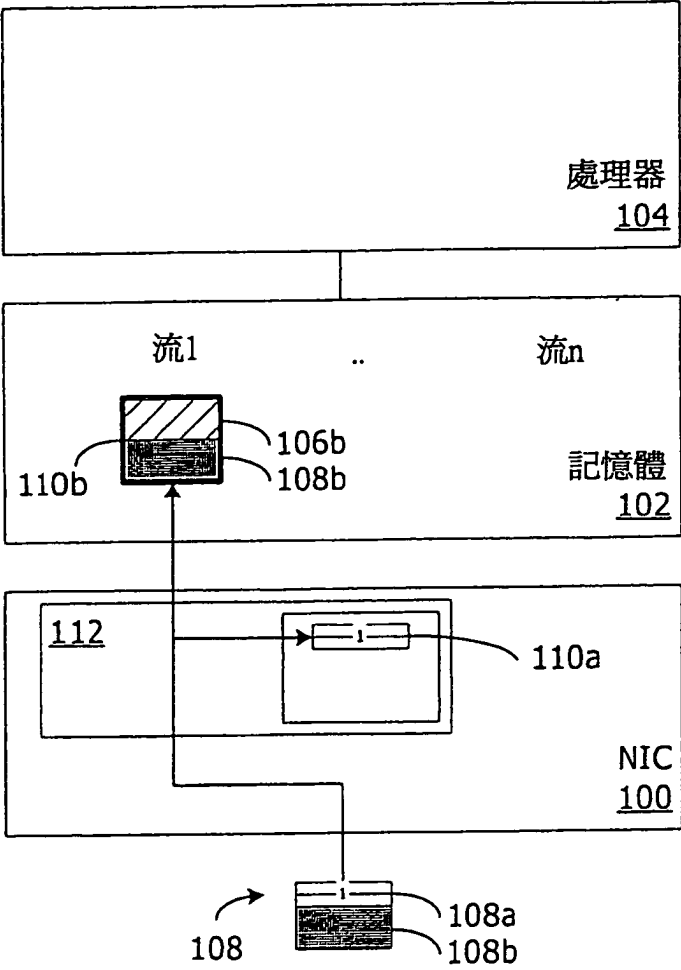
發明之名稱： Packet coalescing

In general, in one aspect, the disclosures describes a method that includes receiving multiple ingress Internet Protocol packets, each of the multiple ingress Internet Protocol packets having an Internet Protocol header and a Transmission Control Protocol segment having a Transmission Control Protocol header and a Transmission Control Protocol payload, where the multiple packets belonging to a same Transmission Control Protocol/Internet Protocol flow. The method also includes preparing an Internet Protocol packet having a single Internet Protocol header and a single Transmission Control Protocol segment having a single Transmission Control Protocol header and a single payload formed by a combination of the Transmission Control Protocol segment payloads of the multiple Internet Protocol packets. The method further includes generating a signal that causes receive processing of the Internet Protocol packet.

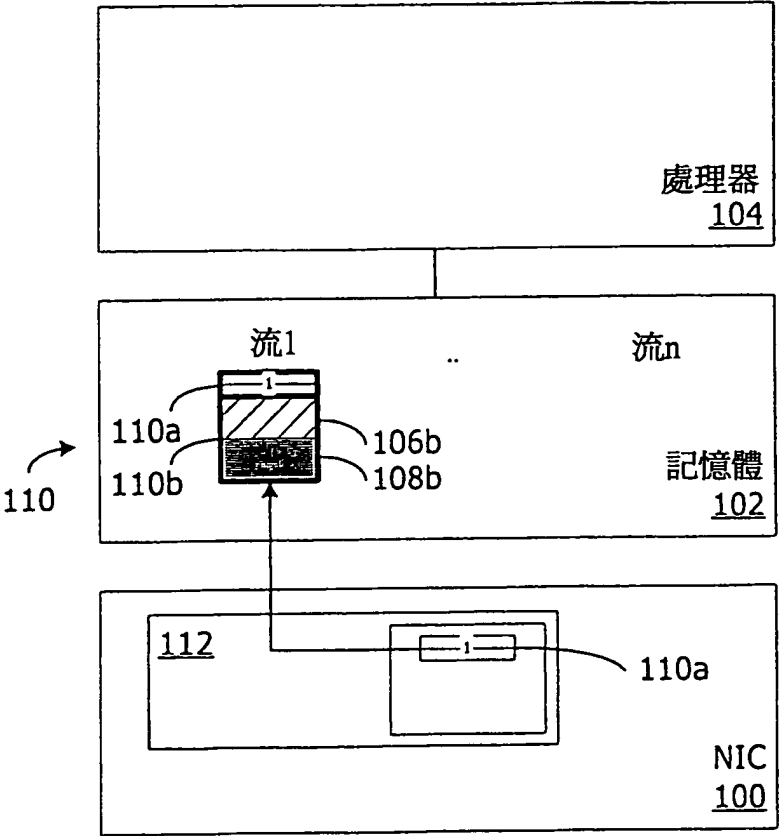
第1A圖



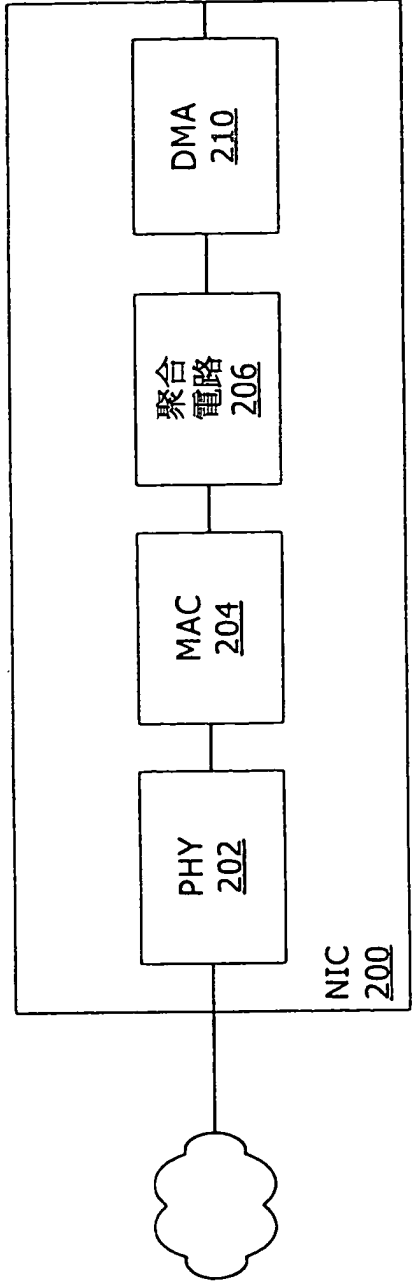
第1B圖



第1C圖



第2圖

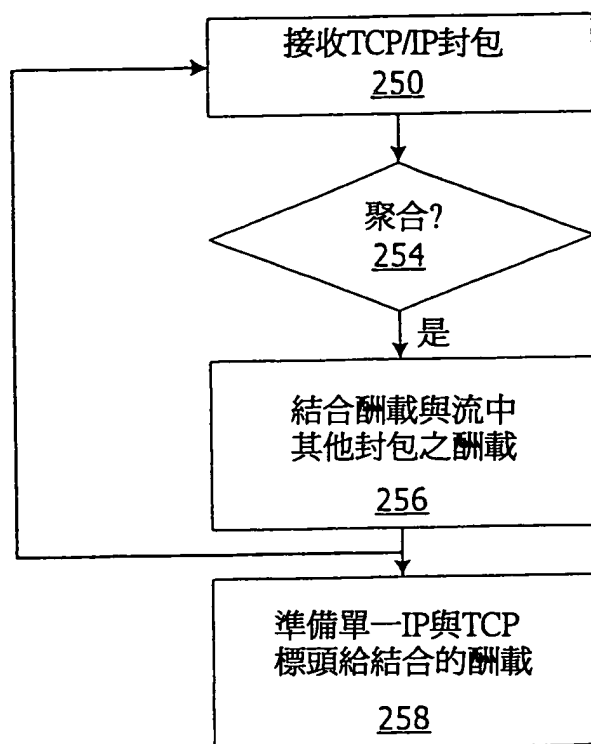


第3圖

流	開始序列號碼#	聚合的位元組	描述符位址	標頭位址	酬載位址

212

第4圖



七、指定代表圖：

(一)、本案指定代表圖為：第(4)圖

(二)、本代表圖之元件代表符號簡單說明：無

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：無

(此處由本局於收
文時黏貼條碼)

民國 96 年 10 月 30 日呈

年 月 日修(更)正替換頁

848928

發明專利說明書

(本申請書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

公 告 本

※申請案號：94139145

※申請日期：94 年 11 月 08 日

※IPC 分類：H04L 29/06
12/56

一、發明名稱：

(中) 封包聚合的方法與系統

(英) Packet coalescing

二、申請人：(共 1 人)

1. 姓 名：(中) 英特爾股份有限公司

(英) INTEL CORPORATION

代表人：(中) 1. 大衛 賽門

(英) 1. SIMON, DAVID

地 址：(中) 美國加州聖大克拉瑞密遜學院路二二〇〇號

(英) 2200 Mission College Blvd., Santa Clara, CA 95052, USA

國籍：(中英) 美國 U.S.A.

三、發明人：(共 6 人)

1. 姓 名：(中) 斯里哈利 馬奇奈尼

(英) MAKINENI, SRIHARI

國 籍：(中) 印度

(英) INDIA

2. 姓 名：(中) 洛維許安卡 艾爾

(英) IYER, RAVISHANKAR

國 籍：(中) 印度

(英) INDIA

3. 姓 名：(中) 大衛 明特恩

(英) MINTURN, DAVID

國 籍：(中) 美國

(英) U.S.A.

4. 姓 名：(中) 單淑久

(英) SEN, SUJOY

國 籍：(中) 印度

(英) INDIA

5. 姓 名：(中) 唐納德 奈威
(英) NEWELL, DONALD
國 籍：(中) 美國
(英) U.S.A.

6. 姓 名：(中) 趙禮
(英) ZHAO, LI
國 籍：(中) 大陸地區
(英) CHINA

四、聲明事項：

◎本案申請前已向下列國家（地區）申請專利 ☐ 主張國際優先權：

【格式請依：受理國家（地區）；申請日；申請案號數 順序註記】

1. 美國 ; 2004/11/16 ; 10/991,239 ☒ 有主張優先權

(英) INDIA

5. 姓 名：(中) 唐納德 奈威
(英) NEWELL, DONALD
國 籍：(中) 美國
(英) U.S.A.

6. 姓 名：(中) 趙禮
(英) ZHAO, LI
國 籍：(中) 大陸地區
(英) CHINA

四、聲明事項：

◎本案申請前已向下列國家（地區）申請專利 ☐ 主張國際優先權：

【格式請依：受理國家（地區）；申請日；申請案號數 順序註記】

1. 美國 ; 2004/11/16 ; 10/991,239 ☒ 有主張優先權

包括許多典型平台的習知元件（例如晶片組與/或將處理器 104、記憶體 102、以及 NIC100 互相連接的 I/O 控制集線器）。此外，顯示於第 1A 至 1C 圖的架構於不同的系統中可作相當大變化。舉例來說，給定系統可以多處理器（分離的處理器與/或處理器核心集成於相同晶粒中）、多 NIC、與/或各種記憶體裝置（例如單一、雙埠、或四埠記憶體）為特徵。同樣地，該控制器 100 可集成於處理器 104、晶片組（未圖示）、或其他電路中。此外，該系統可包括 TCP/IP 卸載引擎（TCP/IP offload engine，TOE），該 TOE 能執行上述之任務，如由該 NIC100 或處理器 104 所處理。

第 2 圖係顯示更詳細的網路介面控制器 200 之樣品架構。雖然所示為處理從網路來的進入（ingress）封包，但該控制器 200 亦可處理至網路之外出（egress）封包。

如圖所示，該控制器 200 能包括實體層裝置（PHY）202，該實體層裝置介接（interface）至通訊媒體（例如纜線或無線電設備）。該 PHY202 能轉換於通訊媒體的類比訊號與用以處理封包的數位位元之間。如圖所示，媒體存取控制器（MAC）204 收集由 PHY202 所輸出的位元（例如經由 FIFO 佇列）。MAC204 能執行許多的鏈結-層（link-layer）操作（例如驗證乙太網路核對和等）。聚合電路 206 運作於 MAC204 所輸出的封包（如第 1A 至 1C 圖所示）。聚合電路 206 可為”硬體連接（hard-wired）”電路，例如特殊應用積體電路（ASIC）。或者，該電路

附件 3A：第094139145號申請專利範圍修正本

民國 100 年 9 月 27 日修正

十、申請專利範圍

1.一種用於封包聚合之方法，包括下列步驟：

接收多進入網際網路協定封包，各個多進入網際網路協定封包具有網際網路協定標頭與具有傳輸控制協定標頭與傳輸控制協定酬載之傳輸控制協定段，該等多網際網路封包屬於相同的傳輸控制協定/網際網路協定流；

準備網際網路協定封包，其具有單一網際網路協定標頭與具有單一傳輸控制協定標頭與單一酬載之單一傳輸控制協定段，該單一酬載係藉由結合該等多網際網路協定封包之傳輸控制協定段酬載而形成；以及

根據由以下群組選出之至少之一，由該多網際網路協定封包中，排除屬於相同流的至少一封包：

該至少一封包，包含無效傳輸控制協定核對和；

該至少一封包，不包含傳輸控制協定資料段；

該至少一封包，包含失序接收的傳輸控制協定段；

該至少一封包，包含傳輸控制協定緊急旗標組；

產生訊號，該訊號造成該網際網路協定封包之接收處理；

開始至少一個直接記憶體存取，以儲存該單一傳輸控制協定標頭、該單一網際網路協定標頭、以及該等傳輸控制協定段之酬載至記憶體；及

140# 9.27 修正
補充

經由描述符表示所結合的傳輸控制協定段酬載之數量，多數段具有個別酬載；

其中該準備包含：

決定用於該單一傳輸控制協定標頭的傳輸控制協定核對和；

決定用於該單一網際網路協定標頭的位元組長度；及

決定該單一傳輸控制協定標頭的認可序列號碼；

其中該流包含一至少部份為接收網際網路協定封包的網際網路源及目的地地址所指定的流，及該傳輸控制協定段的傳輸控制協定源及目的地埠數係包含在該網際網路協定封包內；及

其中該接收、準備、及產生包含為網路介面控制器所接收、準備、及產生。

2.如申請專利範圍第 1 項之方法，更包含下列步驟：
鄰接地儲存該等酬載至該記憶體。

3.如申請專利範圍第 1 項之方法，其中，該準備步驟包含：調整所接收的該等多封包之其中一個的至少一個標頭的至少一個值。

4.如申請專利範圍第 1 項之方法，更包含下列步驟：
保持識別不同流之表，該表將各別的流與下列之群組所選擇的至少一個相關聯：序列號碼、酬載位元組編號、封包描述符之位址、酬載之位址、以及標頭之位址。

5.如申請專利範圍第 1 項之方法，其中，該準備步驟

100* 9.27

包括：基於視窗時間少於中斷緩和視窗時間而準備。

6.如申請專利範圍第 1 項之方法，更包含：經由該描述符表示所聚合的 ACKs 之編號。

7.一種網路介面控制器，該控制器包括：

至少一介面，連接至通訊媒體；

至少一媒體存取控制器；以及

電路，該電路係用以：

接收多進入網際網路協定封包，各個多進入網際網路協定封包具有網際網路協定標頭與具有傳輸控制協定標頭與傳輸控制協定酬載之傳輸控制協定段，該等多網際網路協定封包屬於相同的傳輸控制協定/網際網路協定流；

準備網際網路協定封包，其具有單一網際網路協定標頭與具有單一傳輸控制協定標頭與一酬載之單一傳輸控制協定段，該酬載係藉由結合該等多網際網路協定封包之傳輸控制協定段酬載而形成；

根據以下群組選出之至少之一，由該多數網路網路協定封包排除屬於相同流的至少一封包：

該至少一封包，包含無效傳輸控制協定核對和；

該至少一封包，不包含傳輸控制協定資料段；

該至少一封包，包含失序接收的傳輸控制協定段；

該至少一封包，包含傳輸控制協定緊急旗標組；

產生訊號，該訊號造成該網際網路協定封包之接收處

100 9.27

理；

開始至少一個直接記憶體存取，以儲存該單一傳輸控制協定標頭、該單一網際網路協定標頭、以及該等傳輸控制協定段之單一酬載至記憶體；及

經由描述符表示所結合的傳輸控制協定段酬載之數量，多數段具有個別酬載；

其中該準備電路包含電路用以：

決定用於該單一傳輸控制協定標頭的傳輸控制協定核對和；

決定用於該單一網際網路協定標頭的位元組長度；及

決定該單一傳輸控制協定標頭的認可序列號碼；

其中該流包含一至少部份為接收網際網路協定封包的網際網路源及目的地地址所指明的流，及該傳輸控制協定段的傳輸控制協定源及目的地埠號碼係包含在該網際網路協定封包內。

8.如申請專利範圍第 7 項之控制器，其中該準備電路包含用以調整所接收的該等多封包之其中一個的至少一個標頭的至少一個值。

9.如申請專利範圍第 7 項之控制器，其中，該等所接收的進入封包包括在少於網路介面控制器之中斷緩和時段之時段內所接收之封包。

10.如申請專利範圍第 7 項之控制器，其中，該至少一個媒體存取控制器包括乙太網路媒體存取控制器。

11. 一種用於聚合封包成單一封包之系統，包括：

至少一個處理器；

記憶體；

至少一個介面，連接至通訊媒體；以及

網路介面控制器電路，該電路係用以：

接收多進入網際網路協定封包，各個多進入網際網路協定封包具有網際網路協定標頭與具有傳輸控制協定標頭與傳輸控制協定酬載之傳輸控制協定段，該等多封包屬於相同的傳輸控制協定/網際網路協定流；

準備網際網路協定封包，其具有單一網際網路協定標頭與具有單一傳輸控制協定標頭與一酬載之單一傳輸控制協定段，該酬載係藉由結合該等多網際網路協定封包之傳輸控制協定段酬載而形成；

根據由以下群組選出之至少之一，由該多數網際網路協定封包排除屬於相同流的至少一封包：

該至少一封包，包含無效傳輸控制協定核對和；

該至少一封包，不包含傳輸控制協定資料段；

該至少一封包，包含失序接收的傳輸控制協定段；

該至少一封包，包含傳輸控制協定緊急旗標組；

產生訊號，該訊號造成該網際網路協定封包之接收處理；

開始至少一個直接記憶體存取，以儲存該單一傳輸控制協定標頭、該單一網際網路協定標頭、以及該等傳輸控

100-927

制協定段之酬載至記憶體；及

經由描述符表示所結合的傳輸控制協定段酬載之數目，多數段具有個別酬載；

其中該準備路包含電路用以：

決定用於該單一傳輸控制協定標頭的傳輸控制協定核對和；

決定用於該單一網際網路協定標頭的位元組長度；及

決定該單一傳輸控制協定標頭的認可序列數；及

其中該流包含一至少部份為該接收網際網路協定封包的網際網路源及目的地地址所指定的流，及該傳輸控制協定段的傳輸控制協定源及目的地埠數係包含在該網際網路協定封包內。

12.如申請專利範圍第 11 項之系統，其中，該用以準備的電路包括：電路，係用以調整所接收的該等多封包之其中一個的至少一個標頭的至少一個值。

13.如申請專利範圍第 11 項之系統，其中，該等所接收的進入封包包括在少於網路介面控制器之中斷緩和時段之時段內所接收之封包。

14.如申請專利範圍第 11 項之系統，其中，該電路包括該處理器之一部份。

15.一種用於聚合封包成單一封包之系統，包括：

至少一個處理器；

至少一個雙埠記憶體；

100.927

至少一個介面，連接至通訊媒體；以及

網路介面控制器電路，用以：

接收多進入網際網路協定封包，各個多進入網際網路協定封包具有網際網路協定標頭與具有傳輸控制協定標頭與傳輸控制協定酬載之傳輸控制協定段，該等多封包屬於相同的傳輸控制協定/網際網路協定流；

準備網際網路協定封包，其具有單一網際網路協定標頭與具有單一傳輸控制協定標頭與一酬載之單一傳輸控制協定段，該酬載係藉由結合該等多網際網路協定封包之傳輸控制協定段酬載而形成；

根據由以下之群組所選出之至少之一，由該多數網際網路協定封包排除屬於相同流的至少一封包：

該至少一封包，包含無效傳輸控制協定核對和；

該至少一封包，不包含傳輸控制協定資料段；

該至少一封包，包含失序接收的傳輸控制協定段；

該至少一封包，包含傳輸控制協定緊急旗標組；

產生訊號，該訊號造成該網際網路協定封包之接收處理；

開始至少一個直接記憶體存取，以儲存該單一傳輸控制協定標頭、該單一網際網路協定標頭、以及該等傳輸控制協定段之酬載至記憶體；及

經由描述符表示所結合的傳輸控制協定段酬載之數量，多數段具有個別酬載；

100年7月27日 補充

其中該準備電路包含電路用以：

決定用於該單一傳輸控制協定標頭的傳輸控制協定核對和；

決定用於該單一網際網路協定標頭的位元組長度；及

決定該單一傳輸控制協定標頭的認可序列編號；

其中該流包含一至少部份為接收網際網路協定封包的網際網路源及目的地地址所指明的流，及該傳輸控制協定段的傳輸控制協定源及目的地埠數係包含在該網際網路協定封包內。

16.如申請專利範圍第 15 項之系統，其中，該處理器包括多核心處理器。

17.如申請專利範圍第 15 項之系統，其中，該網際網路協定封包包括 IPv4 封包與 IPv6 封包之其中一種。

18.一種用以封包聚合的方法，包含：

接收具有一網際網路協定標頭的進入網際網路協定封包；及具有傳輸控制協定標頭及傳輸控制協定酬載的輸控制協定區；

決定是否所接收之進入網際網路協定封包具有流識別符匹配具有相同流識別符的一組網際網路協定封包的流識別符，該予以組合至單一網際網路協定封包的該組網際網路封包具有單一網際網路協定封包標頭及單一傳輸控制協定標頭及單一傳輸控制協定酬載，該流識別符至少部份根據該網際網路協定源、網際網路協定目的地、傳輸控制協

100 927

定源埠、及該接收的進入網際網路協定封包的傳輸控制協定目的地埠，該組網際網路協定封包具有相關描述符；

至少在決定該接收進入網際網路協定封包的該傳輸控制協定段的傳輸控制協定順序為用於該流的下一期待傳輸控制協定序列號及所接收進入網際網路協定封包的該傳輸控制協定標頭並沒有任一 RST、SYN、FIN、或 URG 旗標組之後，

將該網際網路協定封包加入至該組網際網路協定封包；及

增量在該描述符中相關於該組網際網路協定封包的封包計數；

決定組合該組的網際網路封包的單一網際網路協定封包，該單一網際網路協定封包具有單一網際網路協定標頭及單一傳輸控制協定標頭及由接收進入網際網路協定封包的傳輸控制協定酬載的該傳輸控制協定酬載與超出一個以上之先前接收網際網路協定封包的傳輸控制酬載的組合所形成單一傳輸控制酬載，該決定單一網際網路協定封包包含：

設定該單一網際網路協定標的該網際網路協定長度欄以對應於該組合組的網際網路協定封包，

設定該單一傳輸控制協定標頭的 ACK 順序號，以對應於該組合組的傳輸控制協定段；及

提供該決定的單一網際網路協定封包至主 TCP/IP 協定堆疊，以替代該組網際網路協定封包。

100 9 7

19.如申請專利範圍第 18 項所述之方法，

其中該決定該單一網際網路協定標頭更包含設定該網際網路協定核對和，以對應於該組合組的網際網路協定封包。

20.如申請專利範圍第 18 項所述之方法，更包含：

至少在決定該所接收進入網際網路協定封包的該傳輸控制協定段的該傳輸控制協定順序不是用於該流的下一期待傳輸控制協定序列號或該所接收進入網際網路協定封包的該傳輸控制協定標頭具有任一 RST、SYN、FIN、或 URG 旗標組後，

提供該所接收進入網際網路協定封包至該主 TCP/IP 協定堆疊；及

提供由該組網際網路協定封包的組合所形成的單一網際網路協定封包。

21.如申請專利範圍第 18 項所述之方法，

其中該加入該網際網路協定封包至該組網際網路協定封包具有相同流識別碼予以組合至單一網際網路協定封包包含：將一輸入項加至一鏈結名單，其中在該鏈結名單中之輸入項對應於在該組中之不同封包。

第5圖

