

具效率且容錯之資訊內容傳遞格網

楊明正

德明科技大學資訊科技系

Email: mjyang@takming.edu.tw

摘要—在即時內容服務網路系統中,如果採用集中式的設計,則內容伺服器可能產生負載分散、負載平衡、即時串流回應、系統容錯等問題,另外亦可能產生分派伺服器錯誤或停止運轉時使用者端無法取得需求內容位置資訊,因此無法即時取得內容。本文提出一個具即時效率的客戶需求內容傳遞格網(Content Delivery Grid)新架構,新架構設計中考量負載分散,將儲存內容透過網路複製並分佈到接近用戶端的區域伺服器群,減輕單一伺服器的大量需求負載,另外也考慮了伺服器的負載平衡,避免高運量負載對伺服器造成可能潰堤的危機,減輕網路頻寬壅塞問題。同時新架構設計也達成位置資訊的分散存放及擷取,以達到位置資訊被存取的容錯功能,避免因集中式存放位置資訊而導致該伺服器潰堤時,使用者程序無法找到內容存放位置資訊,導致無法即時取用資訊內容。

關鍵詞—內容傳遞格網、容錯、即時內容服務、負載平衡。

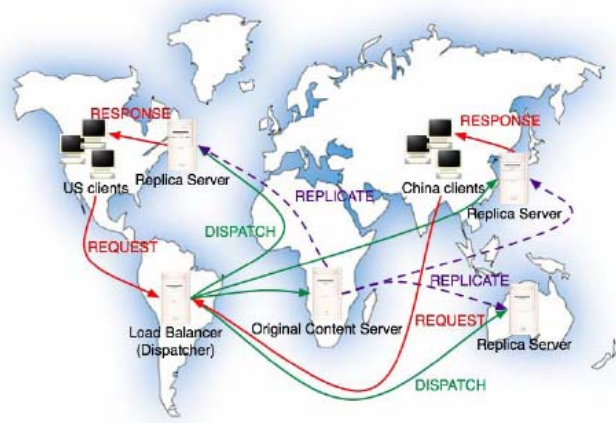
一、相關背景介紹

Luk 等人提到一個 Scenario of the On-Demand Content Delivery Grid [1], Luk 提出的架構劇本(圖一)是使用者端向遠端負載平衡暨分派器提出資訊內容的要求,分派器轉向通知原始內容提供伺服器,並將原始內容複製數份到各地區複製伺服器內,再由各地複製伺服器將資訊內容回應給提出需求的使用者。Baert 等人亦提出一個 Grid Delivery Network's Architecture [2], 架構如圖二所示,其中有一中央伺服器,負責管理整個系統及其他的 Hosts, 且假設此中央伺服器永遠不停止運作,並接受所有使用者提出內容需求,然後回應一個計畫需求(亦即,如何取得資訊內容)給使用者。使用者端在取得計畫內容後可以向 Hosts 或中央

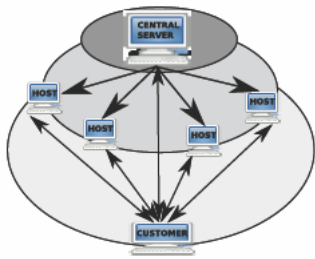
伺服器取得資訊內容。

相對於將資訊內容儲存於單一伺服器,當資訊內容透過網路被分佈或複製到接近使用者端時,此時使用者端可以即時或快速的存取到所需的資訊內容,內容傳遞格網的設計需考慮將內容分散到區域的數個伺服器儲存,以作為分散需求負載,並且亦需考量內容儲存的負載平衡。另外容錯的考量亦需融入系統的設計,如何擷取分散的資訊內容亦是個關鍵設計技術,尤其是設計出分散且達到容錯的效果。

資訊內容傳遞格網系統如果是一個集中式的設計,則將可能產生一些議題值得去探討及解決:例如,(1)單一伺服器承受高負載的流量、(2)多用戶端同時要求同一內容時導致串流速度降低或被拒絕要求、(3)集中式單一伺服器的設備高標準要求及高成本。



圖一：Luk等人所提出的 Scenario of the On-Demand Content Delivery Grid



圖二：Baert等人所提出的 Grid Delivery Network's Architecture

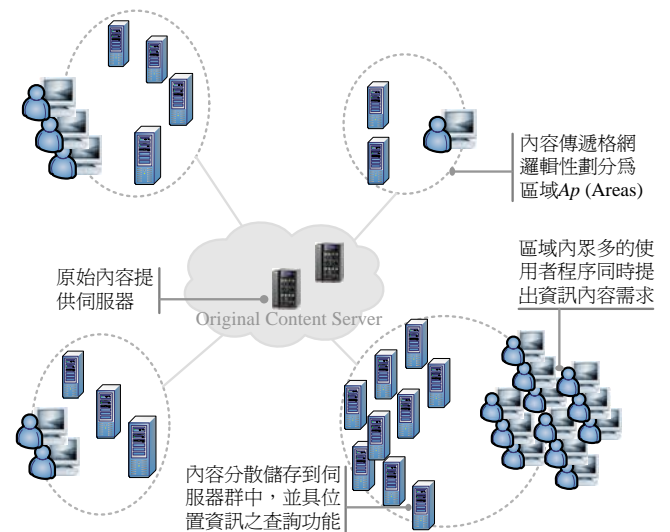
接下來本篇論文的章節架構為：第二章提出本文設計的內容傳遞格網新架構，並解釋系統架構圖。第三章探討區域內容複製及分佈，並同時考量內容負載分散及負載平衡。第四章提出了位置資訊擷取及容錯。第五章為本文提出結論。第六章為參考文獻。

二、系統架構

在前一節介紹相關背景中已經提及的目前被提出之內容傳遞格網設計架構，在這些架構中仍然可能存在一些問題，例如負載分散、負載平衡、即時串流回應、系統容錯等。本文提出一個具即時效率的客戶需求內容傳遞格網新架構，新架構設計中考量負載分散，將儲存內容分散到多個伺服器中，減輕單一伺服器的大量需求負載，另外也考慮了伺服器的負載平衡，避免高運量負載對伺服器造成可能潰堤的危機，同時減輕網路頻寬壅塞問題。架構中對於服務需求的提出不再透過遠端負載平衡分派伺服器，而由區域伺服器聯合完成資訊服務需求，由這些伺服器提供如何取得、向哪些伺服器取得之計畫資訊，因此服務需求不再是中央式，也避免了中央式分派伺服器錯誤或停止運轉時使用者無法取得計畫需求資訊，而無法取得所需求之即時內容服務。同時此架構設計亦提供了容錯功能，部分伺服器的暫時失效並不影響使用者需求內容的取得。

接下來介紹本文所提出內容傳遞格網系統架構構成要素，圖三中，系統架構將被分成數個邏輯性區域，每個區域配備有區域伺服器

群，資訊內容將被複製並分散儲存到伺服器群中，同時伺服器群具有位置資訊查詢功能，伺服器的多寡視該區域內使用者端的需求容量而定，整個內容傳遞格網系統中，還有一個或多個原始內容提供伺服器，這些伺服器是資訊內容產生者將其產生的內容置放的儲存庫，其目的是提供給合法的各地使用者。



圖三：容傳遞格網系統架構

三、區域複製及負載平衡

本章節討論有關具即時效率的客戶需求內容傳遞格網(Content Delivery Grid)架構新設計。新架構設計中將儲存內容透過網路複製並分佈到接近使用者端的區域伺服器群，並同時考量內容負載分散及負載平衡。

(一) 複製資訊內容

為了讓使用者端可以即時或快速的存取到所需的資訊內容，本系統設計區域複製策略如下：

- 使用者端在要求一個資訊內容時，首先由使用者端程序向區域內伺服器提出需求，在確認無此需求資訊時，再轉由向原始內容提供伺服器提出要求。
- 如果是區域內第一次有使用者端提出此項內容服務，原使內容提供伺服器會將需

求內容複製一份並切割成數個區塊 (Blocks)到此區域內的伺服器群，也因此區域內的伺服器群快取(複製一份)了此份資訊內容。以後再有其他使用者端要求此項資訊內容時，就近在區域內向伺服器群提出需求即可，以達成即時或快速存取的目的。

- 區域內伺服器群不需儲存所有的內容資訊，僅是在該區域有使用者提出過的需求內容，才會將該內容複製及分佈到該區域的伺服器群。亦即，某些資訊內容不曾被該區域任何使用者提出過需求時，則該資訊內容仍然存放在原始提供伺服器內。
- 如果某一資訊內容之區塊(Blocks)儲存在區域內之伺服器群，而當其中有部分伺服器無法正常運作時，則使用者端程序在一預設的時間內將無法取得這部分的區塊內容，此時程序將從區域內轉向原始提供伺服器向其提出此部分的區塊內容。

以上的區域複製設計策略目的在於(1)分散原始提供伺服器提供眾多使用者所需的頻寬需求、(2)緩和太多的使用者同時對原始提供伺服器提出需求的暴衝資料流並避免該伺服器的潰堤、(3)使用者端可以從區域內的伺服器群即時或快速的要求到所需的資訊內容、(4)同時提供容錯的機制，使得儲存部分內容區塊的伺服器在無法正常提供需求時，亦可獲得解決。

(二)負載平衡機制

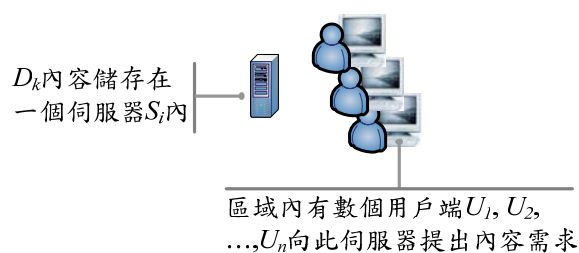
前一小節中，區域複製的策略已經被提及，當儲存內容透過網路複製並分佈到接近使用者端的區域伺服器群時，即應考慮如何分配負載，為了達到負載平衡的原則，本文設計採用了一個簡單的機制。機制的設計如下：

- 假設在內容傳遞格網中邏輯性劃分為 R 個區域 A_p (Areas)，其中 $1 \leq p \leq R$
- 邏輯性的劃分區域內所有的伺服器，並邏輯性的編號為 S_i ，其中 $i=0, 1, 2, 3, \dots, n-1$

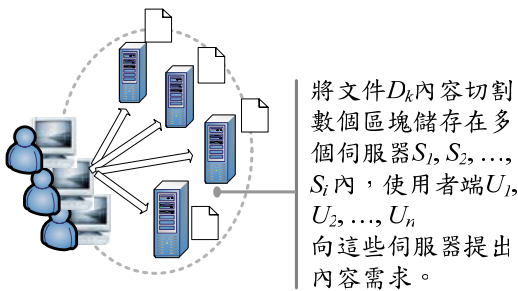
- 假設每個內容 D_k ，其中 $1 \leq k \leq N$ ，切割成 M_k 個固定的區塊 B_k^j (Block)，其中 $1 \leq j \leq M_k$ ，大小為 Z 。
- 原始內容提供伺服器在接受區域內 A_p 使用者端第一次的資訊內容文件需求提出時，此內容 D_k 被切割成 M_k 個固定的區塊 (Block)，系統將隨機選出此區域 A_p 內 M_k 個伺服器(這些伺服器有可能被重複被選出)，將每個區塊 B_k^j 存入被選出的伺服器 S_i 儲存庫內，亦即將內容 D_k 平均負載分散儲存在此區域 M_k 個伺服器儲存庫內。

(三)區域多伺服器分散需求負載

在本文提出之系統負載平衡機制下，區域 A_p 內所有伺服器將平均儲存了所有的資訊內容，因此使用者端的資訊內容需求負載將被分散到數個儲存此資訊內容的伺服器，也因此分散了頻寬消耗的需求。在圖四(a)，原本將文件 D_k 內容儲存在一個伺服器 S_i 內，區域內有數個用戶端 U_1, U_2, \dots, U_n 向此伺服器提出內容需求，在圖四(b)，將文件 D_k 內容切割數個區塊儲存在多個伺服器 S_1, S_2, \dots, S_i 內，使用者端 U_1, U_2, \dots, U_n 向這些伺服器提出內容需求。這些使用者端的總頻寬需求因此分散到多個伺服器 S_1, S_2, \dots, S_i ，也提供了單一伺服器頻寬的容錯。同時避免某些熱門需求內容所導致的需求過量集中於一個或某些伺服器。



圖四(a)：內容儲存在一個伺服器內



圖四(b)：內容切割數個區塊儲存在多個伺服器內

四、位置資訊擷取及容錯

在上一節中討論了資訊內容如何被複製及分散儲存，並達到負載平衡及分散需求。當資訊內容被妥善的分散儲存之後，接下來一個重要的議題就是如何知道這些分散儲存的資訊內容其各個位置資訊 (*Location Information of Content*)，也就是說，只有在需求者程序知道這些內容的位置之後，透過這個位置資訊的索引，才有辦法存取到資訊的內容。

在本文系統設計中，設計目標是要達成位置資訊的分散式存放及擷取，以達成位置資訊被存取的容錯功能，避免因集中式存放位置資訊，而當該伺服器無法正常運作時，使用者程序無法找到內容存放位置資訊，導致無法即時取用資訊內容。設計的方向是利用現有的區域內的部分伺服器作為位置資訊索引的存放及擷取，亦即不存放在單一伺服器內、亦不廣播給區域內的所有伺服器。避免及集中式的缺點，亦不造成大量分散儲存的增加成本。接下來幾個小章節分別探討如何儲存及擷取位置資訊的理論架構及方式。

(一) 位置資訊容錯架構理論

本文的位置資訊容錯架構理論採用了我們自己的先前相關研究理論[3]，在此篇論文中定義了一個 *Location Service Structure (LSS)*。根據此理論架構，我們可以設計出位置資訊之儲存集合 *D-set* 以及位置資訊之查詢集合 *Q-set*。

定義一：在一位置資訊系統中，我們定義依邏輯伺服器編號所組成的儲存集合 *D-set* 以及查詢集合 *Q-set* 如下：

$$D\text{-set} = \{ \{n, (n+d) \bmod N, (n+2d) \bmod N, \dots, (n+kd) \bmod N\} \mid 0 \leq n \leq N-1 \}$$

$$Q\text{-set} = \{ \{n, (n+1) \bmod N, (n+2) \bmod N, \dots, (n+d-1) \bmod N\} \mid 0 \leq n \leq N-1 \}$$

其中 $k = \lfloor (N-1)/d \rfloor$, $d = \lceil \sqrt{N} \rceil$; n, k , 及 N 為整數。 n 為伺服器邏輯編號，而 N 為伺服器總個數。

依據 *Location Service Structure (LSS)* 的理論，上面定義一所定義出的集合 *D-set* 內的任一元素和 *Q-set* 內的任一元素會有至少一個交集點。此特性在我們的先前研究中有詳細的驗證及說明[3]。利用此交集的特性，本文設計將位置資訊分散儲存到部分伺服器內，再由查詢部分伺服器，從交集的伺服器取得位置資訊，進而獲得使用者端想要的內容，同時又可達到容錯功能。

(二) 位置資訊之分散儲存

某個區域第一次存取一特定內容時，原始的內容提供伺服器會將此資訊內容複製一份到該區域的伺服器群。然後需要將資訊內容的儲存位置 (*Location Information of Content*) 做紀錄，例如表一，並儲存在從 *D-set* 隨機選取的儲存集伺服器內，以作為使用者透過這個位置資訊的索引，去存取到資訊的內容。其中，分散儲存位置資訊的程序策略如下：

- 執行程序將位置資訊做成索引表，並將此位置資訊索引表加上時間印記傳送到從 *D-set* 隨機選取的一個儲存集內的伺服器。
- 在儲存集合內的伺服器收到此位置資訊後，將此資訊存入資料庫內，並回傳已經收到的回覆訊息給執行程序。
- 如果執行程序在一預設的時間內沒有收到所有該回覆訊息的伺服器之回覆，則執行程序將再重新從 *D-set* 隨機挑選另一組儲存集內的伺服器，並將位置資訊索引表加上時間印記傳送重新傳送給這些伺服器，然後繼續執行前一步驟，否則執行程序結束執行。

表一：資訊內容的儲存位置(Location Information of Content)紀錄表。

伺服器	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7
區塊存放號碼	10 14	06 09 15	08	01 07	04 11	02 12	05 13	03 16

(三) 位置資訊之查詢

接下來另一個重要的問題就是如何知道這些分散儲存的資訊內容其各個位置資訊(Location Information of Content)，也就是說，只有在使用者程序知道這些內容的位置之後，透過這個位置資訊的索引，才有辦法存取到資訊的內容。這個分散式的位置資訊查之詢的程序策略如下：

- 使用者端程序以關鍵字提出內容需求，並將此需求傳送到從 Q -set 隨機選出的一個查詢集內的伺服器。
- 當這些伺服器接收到此查詢需求時，進行核對資料庫，如有符合之內容索引，則將此內容存放位置資訊連同時間印記回覆給使用者端執行程序。否則，伺服器回覆查無符合需求訊息。
- 在一預設的時間內，如果使用者端程序沒有收到查詢集內的所有伺服器的全部回覆，則使用者端程序從新由 Q -set 隨機選出另一個查詢集內的伺服器，並重新將需求傳送給這些伺服器，然後繼續執行前一步驟，否則執行程序執行下一步驟。
- 當使用者端程序收到查詢集內的所有伺服器的回覆，則此程序將依時間印記挑選出所需要之位置資訊，然後根據此位置資訊的索引，去存取到資訊的內容。

在本章節所設計之位置資訊之存放及查詢具有容錯的機制，針對第(二)及第(三)小節的程序策略設計，我們現在舉一個存放及查詢例子。假設此區域內總計有九個伺服器，將其邏輯性的編號為 $S_0, S_1, S_2, S_3, \dots, S_8$ ，當此區域內有使用者端第一次存取一特定資訊內容 D_k 時，原始的內容提供伺服器會將此內容複製一份分散到此區域的部分(或依內容大小有可能是全部的)

伺服器，然後執行程序將資訊內容的儲存位置(Location Information of Content)做成位置資訊索引表，並將此位置資訊索引表加上時間印記(Time Stamp)傳送到從 D -set 隨機選取的一個儲存集，例如 $\{S_5, S_8, S_2\}$ 內的伺服器，因此 S_5 、 S_8 及 S_2 具有 D_k 的內容存放位置資訊索引，現在假設有一使用者端 U_i 以關鍵字尋找內容 D_k ，此時由使用者端執行程序從 Q -set 隨機選取一個查詢集，例如 $\{S_8, S_0, S_1\}$ ，並向這些伺服器以關鍵字查詢內容 D_k 的位置資訊。依據 Location Service Structure (LSS) 結構理論所定義出來的儲存集合 D -set 以及查詢集合 Q -set 特性是：任意二個元素，分別屬於 D -set 及 Q -set 必定有交集點。因此上述例子中隨機選出的儲存集 $\{S_5, S_8, S_2\}$ 及查詢集 $\{S_8, S_0, S_1\}$ 就有一個交集點為 S_8 伺服器。透過 S_8 伺服器使用者端程序可以取得以關鍵字查詢內容 D_k 之存放位置資訊，根據此位置資訊，使用者端程序即可即時取得 D_k 之內容。

假設如果伺服 S_8 器剛好停止服務，則使用者端程序可能就無法取得內容 D_k 之位置資訊，此時使用者端程序可以立即隨機從 Q -set 選取另一個查詢集，例如 $\{S_5, S_6, S_7\}$ ，透過交集點 S_7 之伺服器可以重新取得 D_k 之位置資訊。

五、結論

本論文提出了一個新的內容傳遞格網設計架構，這架構解決可能存在一些問題，例如負載分散、負載平衡、即時串流回應、系統容錯等。新架構設計中將儲存內容分散到多個伺服器中，減輕單一伺服器的大量需求負載，也考慮了伺服器的負載平衡，同時減輕網路頻寬壅塞問題。架構中對於服務需求的提出亦不再透過遠端負載平衡分派伺服器，而由區域伺服器群聯合完成，因此服務需求不再是集中式，也避免了集中式分派伺服器錯誤或停止運轉時使用者無法取得需求資訊，同時此架構設計亦提供了容錯功能，部分伺服器的暫時失效亦不影響使用者的需求內容的取得。

六、參考文獻

- [1] Vincent Wing-Hei Luk, Jacky Chan, Kam-Wing Ng, and Albert Kai-Sun Wong, “On-Demand Content Delivery Grid ”, ICITA2008, 2008.
- [2] A.-E. Baert, V. Boudet, and A. Jean-Marie, “Performance analysis of Data replication in Grid Delivery Network”, IEEE International Conference on Complex, Intelligent and Software Intensive Systems, 2008.
- [3] M.J. Yang, “Replica Location Service in Data Grid,” The Third Workshop on Grid Technologies and Applications (WoGTA' 06), 2006.