

組織架構下的格網資源管理系統建置

黃智威、楊明正

德明財經科技大學資訊科技系

William Huang, Ming-Jeng Yang

Department of Information Technology, Takming University of Science and Technology

E-Mail: {huang.grid@gmail.com;mjyang@takming.edu.tw}

摘要

在格網系統中資源是分散的，這些資源透過網路和資源管理器(Resource Management System, RMS)相互連接，每一個資源可以提供計算服務的能力。但是當他們有計算服務的需求時，也可以向RMS請求計算的服務。RMS會將一個計算服務需求分割成多個子需求，分配給能夠提供計算服務的資源去執行計算，然而RMS會爲了提高計算服務的可靠度以及防止資源在執行計算服務的時候產生錯誤，所以RMS會將一個子需求任務分配給多個資源去執行計算，但是當RMS本身發生錯誤的時候，則這個格網系統將無法提供分配計算需求的服務。本文提出一個格網資源管理系統的容錯架構新設計，並且提出分析數據結果。結果亦顯示分層式RMS架構的合理性。

1. 介紹

格網[1][2]是一個非常龐大的分散式計算系統也是資源分享的系統，它在網際網路和多種平台介面上運作。爲了有效的提昇格網支援服務能力，所以資源管理系統(RMS)被提出，它主要功能是在提供容錯能力及計算服務穩定性。RMS 主要是管理所有可利用的資源的處理行程、網路頻寬及資料儲存等。

在 Chung 等學者[4] 所提出的文章中提到，在格網系統架構中他們建議使用星型結構(圖一)來建置，這個設計是將 RMS 透過單一通道直接的连接到每一個資源，但這樣似乎不夠準確，即使它簡化了分析和計算。

在 Dai 等學者[5]所提出的文章中提到，他們提出一套更適合虛擬架構的樹狀結構。在這個樹狀結構中，樹根代表是 RMS，葉子則是代表資源，而樹枝代

表通訊通道，它將樹根和葉子連結在一起，這一些通道是被很多資源共同使用的。雖然這是相當適合格網系統架構的模型，但是當 RMS 不能正常運作的時候，即無法提供分配計算的服務。另一個考慮是當一個不是很龐大計算需求服務向 RMS 提出的時候，可能它只需要一個小型區域格網系統就可以完成這個服務。



圖 1. 格網系統 RMS 的星狀結構

在這篇文章中，我們提出一套二層式格網資源管理系統架構，由一個中央 RMS 和一些區域 RMS 組成。這個格網系統有一些區域型格網系統，在每一個區域格網系統中擁有一個區域 RMS，在多個區域格網系統之中會有一個中央 RMS。依照服務的計算任務大小分別交由區域 RMS 及中央 RMS 處理。

在第二章節我們描述是如何設計二層式資源管理系統的架構。在第三章節提出數據結果。在第四章節則是結論。

2. 系統架構

在這章節中，爲了格網計算服務我們提出一套二層式 RMS 的結構。在一個小型區域網路內會放置一台資源管理器，這台機器我們稱之爲區域 RMS，結合多個小型網路所形成的一個大型網路中會再放置一台資源管理器，這台機器我們稱之爲遠端 RMS。一套二層式 RMS 結構在圖二。在每一個小型區域網路中包含數個可利用資源及一台小型的資源管理系統(區域 RMS)，在多個小型區域網路透過網路與遠端一台大型中央資源管理系統(遠端 RMS)連結在一起。例子 1：在圖二裡，當 1 號區域 RMS 發生故障無法提供服務

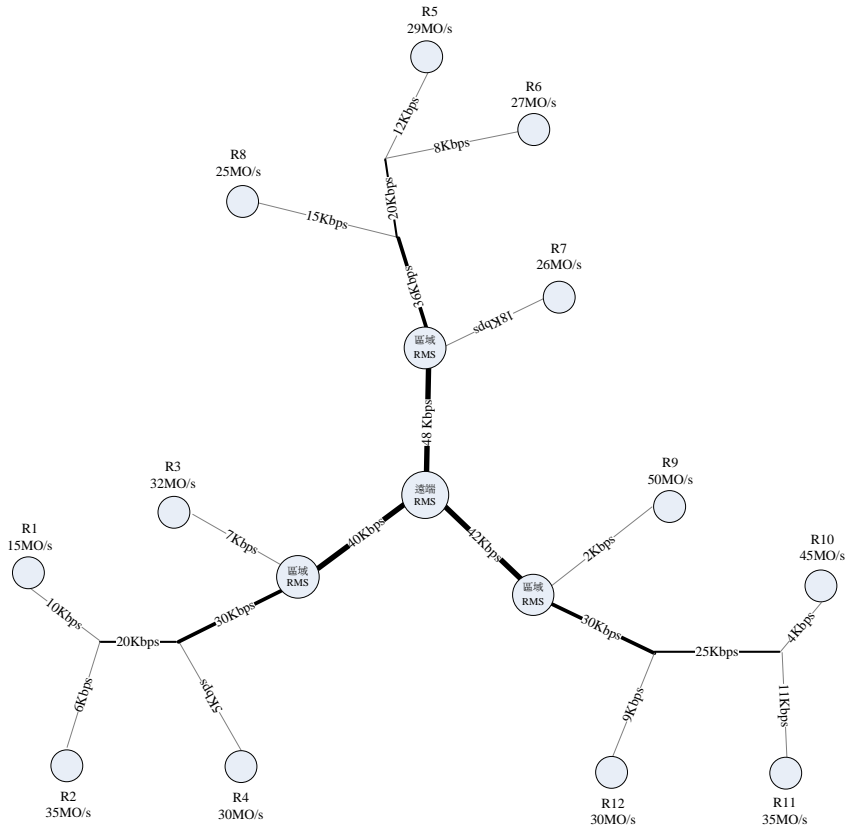


圖 2. 二層式格網資源管理系統架構

的時候，此時遠端 RMS 就會取代 1 號區域 RMS 的工作來提供分配計算服務。例子 2：當 R5 向 2 號區域 RMS 請求計算服務時，區域 RMS 在收到請求之後會先判斷這個計算請求任務的複雜度和計算量有沒有超過此區域內所有可利用資源能提供的服務能力，如果沒有超過的時候，區域 RMS 就會直接將任務分割成多個子任務，分配給多個區域內的資源去執行計算。假如計算任務超過區域內的服務能力，此時區域 RMS 會向遠端的遠端 RMS 請求計算服務的協助，並將計算任務傳送到遠端 RMS，遠端 RMS 則會將此計算任務分割成多個子任務，再分配給所有的區域 RMS 去執行二次分配，區域 RMS 接收到分配任務之後會再將任務分配給區域內的資源去執行計算服務。當資源們完成服務之後會回傳結果給區域 RMS/遠端 RMS，區域 RMS/遠端 RMS 整合所有回傳的結果時會將整個任務結果傳回。

3. 數據分析

在這個章節，我們提出二層式資源管理模型的數據分析(服務時間)，接下來我們介紹數據分析所使用的名稱：

C ：整個任務複雜度 (MO—Mega Operations)

c_j ：子任務複雜度 (MO)

R_p ：資源處理速度 (MO/s)

D ：資源到區域資源管理器的資料量(Kb)

B ：區域頻寬 (Kbps) – 資源在區域內的頻寬

D' ：從遠端 RMS 到區域 RMS 的資料傳輸量 (Kb)

B' ：從遠端 RMS 到區域 RMS 的頻寬 (Kbps)

t_p ：資源處理任務所耗費的時間 (s)

t_l ：區域內資料傳輸時間(s)

t_r ：遠端資料傳輸時間(s)

t_{C-RMS} ：需求傳送至遠端 RMS 花費時間 (s)

T_{c_local} ：區域任務完成時間(s)

T_{c_remote} : 遠端任務完成時間(s)

Γ_{s_local} : 區域服務時間(s)

Γ_{s_remote} : 遠端服務時間(s)

M : 區域格網系統數量

m : 區域中可利用資源數

λ : 資料率.

在我們二層式 RMS 格網系統中，我們考慮兩種區域及遠端所有資源服務管理的情況，第一種情況是假設將一個計算任務交給區域格網系統執行，區域的 RMS 將任務分配給區域內可利用的資源執行需要的時間。第二種情況是假設有一個計算任務超過區域格網所能提供的服務能力，區域 RMS 將這個任務轉送到遠端的 RMS，由遠端 RMS 來分配整個任務到所有可利用的資源執行。

在第二種情況，假如有一個計算任務超過區域格網系統所能提供服務能力時，區域 RMS 會將這個

任務轉送到遠端 RMS，遠端 RMS 會重新將整個任務分配給所有的區域格網系統去執行，不過這樣將任務從區域 RMS 轉送到遠端 RMS，會增加資料在傳輸上花費的時間。

我們針對二層式 RMS 系統，考慮二個分析的例子，分別在區域 RMS 以及在遠端 RMS 執行之後的數據結果。這個系統有三個區域型格網系統，分別是區域格網系統 I，II，和 III，每一個區域格網系統擁有一個區域 RMS，在這三個區域格網系統中包含一個遠端 RMS。整個資源服務系統架構圖、資源處理能力及各個通訊通道頻寬數值在圖二。在第一個例子中，整個任務計算複雜度是 $C=300,000$ MO(mega operations)。簡化例子，我們假設子任務資料傳輸量與計算複雜度是成比例成長，所以資料率 $\lambda=0.00005$ 。這個例子每一個資源完成子任務時間在圖三。

在第二個例子整個任務計算複雜度是 $C=1,000$ MO(mega operations)，資料率 $\lambda=0.0005$ 。這個例子每一個資源完成子任務時間在圖四。

| | Resource | c_j | $R_p(\text{MO})$ | t_p | D (Kb) | B (Kbps) | t_i | D' (Kb) | B' (Kbps) | t_r | t_{c-RMS} | Completion Time | Θ_{max} |
|-------------------------------------|----------|-----------|------------------|----------|----------|----------|--------|---------|-----------|-------|-------------|-----------------|----------------|
| 區域 C M=1 m=4 c_j= λ= | R1 | 75,000.00 | 15 | 5,000.00 | 3,750.00 | 10 | 375.00 | | NULL | | | 5,375.00 | 5,375.00 |
| | R2 | 75,000.00 | 35 | 2,142.86 | 3,750.00 | 6 | 625.00 | | NULL | | | 2,767.86 | |
| | R3 | 75,000.00 | 32 | 2,343.75 | 3,750.00 | 7 | 535.71 | | NULL | | | 2,879.46 | |
| | R4 | 75,000.00 | 30 | 2,500.00 | 3,750.00 | 5 | 750.00 | | NULL | | | 3,250.00 | |
| | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | |
|-----------------------------------|-----|-----------|----|----------|----------|----|--------|----------|--|----|--------|----------|----------|----------|
| 遠端 C M= m= c_j= λ= | R1 | 25,000.00 | 15 | 1,666.67 | 1,250.00 | 10 | 125.00 | | | | | 2,291.67 | 2,291.67 | |
| | R2 | 25,000.00 | 35 | 714.29 | 1,250.00 | 6 | 208.33 | 5,000.00 | | 40 | 125.00 | 1,422.62 | | |
| | R3 | 25,000.00 | 32 | 781.25 | 1,250.00 | 7 | 178.57 | | | | | 1,459.82 | | |
| | R4 | 25,000.00 | 30 | 833.33 | 1,250.00 | 5 | 250.00 | | | | | 1,583.33 | | |
| | R5 | 25,000.00 | 29 | 862.07 | 1,250.00 | 12 | 104.17 | | | | | 1,466.24 | | |
| | R6 | 25,000.00 | 27 | 925.93 | 1,250.00 | 8 | 156.25 | 5,000.00 | | 48 | 104.17 | 375 | | 1,582.18 |
| | R7 | 25,000.00 | 26 | 961.54 | 1,250.00 | 18 | 69.44 | | | | | 1,530.98 | | |
| | R8 | 25,000.00 | 25 | 1,000.00 | 1,250.00 | 15 | 83.33 | | | | | 1,583.33 | | |
| | R9 | 25,000.00 | 50 | 500.00 | 1,250.00 | 2 | 625.00 | | | | | 1,625.00 | | |
| | R10 | 25,000.00 | 45 | 555.56 | 1,250.00 | 4 | 312.50 | 5,000.00 | | 42 | 119.05 | | | 1,368.06 |
| | R11 | 25,000.00 | 35 | 714.29 | 1,250.00 | 11 | 113.64 | | | | | 1,327.92 | | |
| | R12 | 25,000.00 | 30 | 833.33 | 1,250.00 | 9 | 138.89 | | | | | 1,472.22 | | |

C=整個任務計算複雜度, c_j =子任務計算複雜度, λ =資料率, M=區域格網數目, m=區域格網內資源數目, Θ =任務完成時間

表格 I

| | Resource | c_j | $R_p(MO)$ | t_b | D (Kb) | B (Kbps) | t_i | D' (Kb) | B' (Kbps) | t_r | t_{c-RMS} | Completion Time | Θ_{max} |
|--|----------|--------|-----------|-------|--------|----------|-------|---------|-----------|-------|-------------|-----------------|----------------|
| 區域 C M=1 m=4 c _j =250 λ=0.0005 | R1 | 250.00 | 15 | 16.67 | 125.00 | 10 | 12.50 | | NULL | | | 29.17 | 33.33 |
| | R2 | 250.00 | 35 | 7.14 | 125.00 | 6 | 20.83 | | NULL | | | 27.98 | |
| | R3 | 250.00 | 32 | 7.81 | 125.00 | 7 | 17.86 | | NULL | | | 25.67 | |
| | R4 | 250.00 | 30 | 8.33 | 125.00 | 5 | 25.00 | | NULL | | | 33.33 | |
| | | | | | | | | | | | | | |

| | | | | | | | | | | | | | |
|---|-----|-------|----|------|-------|----|-------|--------|----|------|----|-------|-------|
| 遠端 C M=3 m=4 c _j =83 λ=0.0005 | R1 | 83.33 | 15 | 5.56 | 41.67 | 10 | 4.17 | 166.67 | 40 | 4.17 | 13 | 26.39 | 39.17 |
| | R2 | 83.33 | 35 | 2.38 | 41.67 | 6 | 6.94 | | | | | 25.99 | |
| | R3 | 83.33 | 32 | 2.60 | 41.67 | 7 | 5.95 | | | | | 25.22 | |
| | R4 | 83.33 | 30 | 2.78 | 41.67 | 5 | 8.33 | | | | | 27.78 | |
| | R5 | 83.33 | 29 | 2.87 | 41.67 | 12 | 3.47 | 166.67 | 48 | 3.47 | | 23.01 | |
| | R6 | 83.33 | 27 | 3.09 | 41.67 | 8 | 5.21 | | | | | 24.96 | |
| | R7 | 83.33 | 26 | 3.21 | 41.67 | 18 | 2.31 | | | | | 22.19 | |
| | R8 | 83.33 | 25 | 3.33 | 41.67 | 15 | 2.78 | | | | | 22.78 | |
| | R9 | 83.33 | 50 | 1.67 | 41.67 | 2 | 20.83 | 166.67 | 42 | 3.97 | | 39.17 | |
| | R10 | 83.33 | 45 | 1.85 | 41.67 | 4 | 10.42 | | | | | 28.94 | |
| | R11 | 83.33 | 35 | 2.38 | 41.67 | 11 | 3.79 | | | | | 22.84 | |
| | R12 | 83.33 | 30 | 2.78 | 41.67 | 9 | 4.63 | | | | | 24.07 | |

C=整個任務計算複雜度, c_j =子任務計算複雜度, λ =資料率, M=區域格網數目, m=區域格網內資源數目, Θ =任務完成時間

表格 II

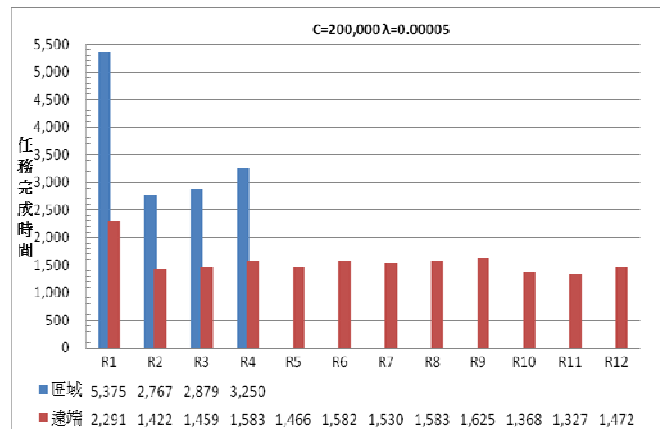


圖 3. 第一個例子在區域 RMS 執行及在遠端 RMS 執行結果比較圖

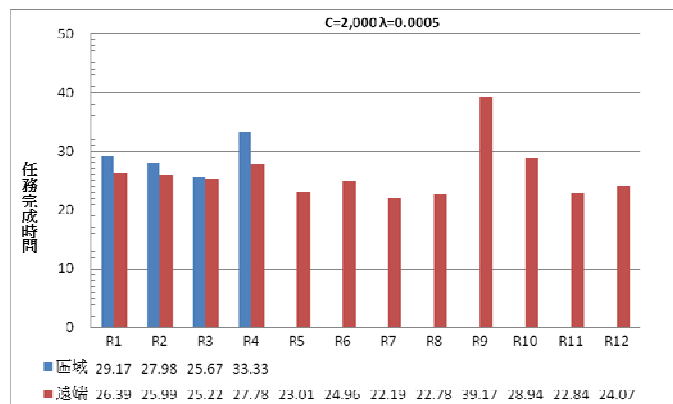


圖 4. 第二個例子在區域 RMS 執行及在遠端 RMS 執行結果比較圖

4. 結論

在這篇文章中，我們主要設計出一套二層式資源管理系統，針對不同計算複雜度的服務任務，將任務分配到不同服務能力的格網系統中和提供RMS容錯的能力。當區域RMS獲得一個使用者計算服務的時候，分配給區域內可利用的資源去執行計算服務。假如這個計算服務太龐大，超過區域格網系統所能提供計算能力的時候，此時區域RMS會將這個計算服務轉送到遠端RMS去，遠端RMS收到這個計算服務要求之後，會將任務分配給所有可利用的資源執行計算。未來我們將進一步測試尋找出計算任務複雜度的臨界值，用來判斷區域RMS是否要將計算任務轉送到遠端RMS。

參考文獻

[1] Foster I, Kesselman C (eds.). *The Grid: Blueprint for a New*

Computing Infrastructure. Morgan Kaufmann: San Francisco, CA, 1999.

[2] I. Foster and C.Kesselman, *The Grid 2: Blueprint for a New Computing Infrastructure*. : Morgan-Kaufmann, 2003.

[3] Sinha PK. *Distributed Operating Systems: Concepts and Design*. IEEE Press: New York, NY, 1997.

[4] M. S. Chang, D. J. Chen, and M. S. Lin, "The distributed program reliability analysis on a star topology," *Computers and Operations Research*, vol. 27, no. 2, pp. 129–142, 2000.

[5] Yuan-Shun Dai, and Gregory Levitin, "Reliability and Performance of Tree-Structured Grid Services", *IEEE Transactions on Reliability*, Vol. 55, No. 2, June 2006.

[6] K. Krauter, R. Buyyaj, and M. Maheswaran, "A taxonomy and survey of grid resource management systems for distributed computing," *Software—Practice and Experience*, vol. 32, no. 2, pp. 135–164, 2002.

[7] J. Nabrzyski, J. M. Schopf, and J. Weglarz, *Grid Resource Management*. : Kluwer Publishing, 2003.