行政院國家科學委員會輔助專題研究計劃成果報告

分散式資料庫系統可靠度分析之研究

The Study of Reliability Analysis in Distributed Database System

計劃類別：[x]**個別型計劃**　　[　]**整合型計劃**

計劃編號：NSC 89-2213-E-009-008

執行期間：88 **年** 8 **月** 1 **日至** 89 **年** 7 **月** 31 **日**

計劃主持人：**陳登吉 教授**

共同主持人：**張明桑　博士**

本成果報告包括以下應繳交之附件：

[　]**赴國外出差或研習心得報告一份**

[　]**赴大陸地區出差或研習心得報告一份**

[　]**出席國際學術會議心得報告及發表之論文各一份**

[　]**國際合作研究計劃國外研究報告書一份**

執行單位：**國立交通大學**

中　華　民　國 89 **年** 7 **月** 31 **日**

分散式資料庫系統可靠度分析之研究

The Study of Reliability Analysis in Distributed Database System

計畫編號：NSC 89-2213-E-009-008

執行期限：88 **年** 8 **月** 1 **日至** 89 **年** 7 **月** 31 **日**

主持人：陳登吉　交通大學資訊工程系教授

一、摘要

中文摘要

本計畫將對分散式資料庫系統（DDBS）之可靠度問題提出一完整之研究。分散示資料庫系統之可靠度，除了和通訊網路線的可靠度和網路平台的可靠度有關之外，並且和其資料庫的分散情形也有關係。我們首先對各種不同的網路結構做一個分類（例如：匯流排網路,環狀網路,樹狀網路,星狀網路等)之後，接著再分析建構在其上的分散式資料庫系統之可靠度問題的計算複雜度（computational complexity)，歸納出哪幾類的網路結構，其分散式資料庫的可靠度問題可以在 polynomial time 或 linear time 之內得到正確答案，又其相對應的演算法為何？此外，對於其它分散式資料庫系統可靠度問題屬於 NP-hard（即不存在 polynomial time 的演算法）的網路結構，我們也提出計算其可靠度估計值的逼近演算法，並且是在容許的誤差範圍之內。

關鍵字詞：

分散式資料庫系統(Distributed database system), **可靠度** (Reliability), **計算複雜度** (Computational complexity)

Abstract

In this project, we propose the research and analysis methodology for the reliability of distributed database systems. The reliability of a distributed database system (DDBS) not only depends on the reliability of its communication links and nodes but also on the distribution of databases. First, we will classify the networks into various types depended on their topologies (such as bus, ring, tree, and star topologies). Then, we will analyze the computational complexity of the reliability problem for the distributed database system built on each classified network. For the types of networks whose reliability of distributed database system are not NP-hard problem, we will propose the linear time or polynomial time algorithms to compute their reliability. For the other networks whose reliability of distributed database system are NP-hard problem, we will propose the approximate algorithms, to compute their approximate reliability.

Keywords：

Distributed database system, Reliability, Computational complexity

二、計畫緣由與目的

在分散式資料庫系統中有許多不同的研究主題，包括: 系統設計 (System Design)、程序管理 (Process Management)、負荷平衡 (Load Balance) 、檔案管理 (File Management) 、存取控制 (Access Control) 、分散式演算法 (Distributed Algorithm) 、可靠度(Reliability) 、及錯誤容忍度 (Fault Tolerance) 等等。這些研究主題對設計分散式資料庫系統都非常的重要，本研究計畫主要是針對分散式資料庫系統可靠度做探討。

自從 1990 年網際網路（Internet）開放商業服務後，加上 1994 年全球資訊網（WWW）在 PC 視窗用戶端瀏覽器（如 Netscape、Explorer 及 Mosaic 等）的成熟後，使得網際網路的使用者呈倍數成長，同時建構在全球資訊網上的資料庫系統更是急速的膨脹。這些資料庫系統為了提高其系統可靠度（Reliability）與資料的傳輸速率，大多採用分散的方式來儲存資料。例如在新一代的 MS-SQL 資料庫系統中，允許資料庫放在不同的網路平台上，並且可以複製（Replication）多份相同的資料庫，分別放在不同的網路平台上，以提高其整個資料庫系統之可靠度。雖然分散式資料庫系統（DDBS)目前已廣泛地應用在網際網路上，但是對於分析與計算分散式資料庫系統可靠度的工具卻還很缺乏。在 1994 年, D.J Chen 與 M.S. Lin 提出了 FREA 演算法[1]，用來分析計算分散式資料庫系統之可靠度，雖然 FREA 演算法對於建構在具有 parallel-serial reduction[2]特性網路上的分散式資料庫系統，能夠很快地分析出其可靠度，但是對於建構在其它網路結構上的分散式資料庫系統，卻還是有可能要花費冗長的時間來計算其可靠度。隨者全球資訊網上分散式資料庫系統的急速擴張，目前我們急需要有一套完整且快速的可靠度分析工具，以提供給資料庫設計與管理人員，作為規畫出高可靠度分散式資料庫系統之參考。

Kumar[3]於 1985 年, 首先將網路可靠度[4]（Network Reliability)的問題轉移到分散式系統上。對於分散式資料庫系統的可靠度分析，除了要考慮傳統網路可靠度問題上的通信網路線路故障（link failure）與平台故障(node failure)之外，同時還必須考慮到資料的複製、分散的方式與資料庫之間的關連性。因為不同的資料分散方式，系統會得到不同的可靠度，所以分散式資料庫系統上的可靠度問題較傳統網路的可靠度問題更為複雜。Chen 與 Huang[5]於 1992 年提出了一個新的計算分散式資料庫系統可靠度的演算法，他們先利用圖形演算法來求得檔案擴展樹（file Spanning Tree)，再利用所找到的檔案擴展樹求出此分散式資料庫系統的可靠度。接者 Chen 與 Lin[2]於 1993 年提出了化簡分散式網路的一般化法則，利用此圖形化簡方法，能夠大大的縮小網路圖形的搜尋空間，相對的也就能夠很快的計算出其可靠度。接者 Chen 與 Lin[1]於 1994 年提出了 FREA 演算法，此演算法結合上述的圖形化簡法則與一般化的 factoring theorem。FREA 演算法對於具有 Parallel-Serial Reduction 特性的分散式資料庫系統，能夠在極短的時間之內得到其可靠度，但是對於不具有 Parallel-Serial Reduction 特性的分散式資料庫系統，卻不一定有相同的效果。接著 Chen and Lin 提出了另一個演算法，用來分析當網站與通信網路線路都不是可靠的時候，其分散式資料庫系統可靠度的計算方法[6]。

雖然, 上述所提的研究對於具有某些特性（例如：parallel-serial reduction）的分散式資料庫系統能夠在短時間之內得到其可靠度，但是隨者網際網路上分散式資料庫的日益膨脹，相對的其關連性與網路結構也就日益的複雜，傳統的分散式資料庫可靠度分析方法已經無法應付，所以我們必須發展出另一套新的可靠度分析模式來分析這些較為複雜的分散式資料庫系

統。

本計畫的目的是希望對分散式資料庫系統的可靠度問題提出一完整的研究與探討。首先對各種不同的網路結構做一分類，接著再分析建構在各種不同網路結構上的分散式資料庫系統上的可靠度問題的計算複雜度（computational complexity)，歸納出那幾類的網路結構，其分散式資料庫的可靠度計算可以在 polynomial time 或 linear time 之內得到，方法為何？又哪些網路結構的可靠度計算問題是屬於 NP-hard 問題（即不存在 polynomial time 演算法)，對於屬於這類 NP-hard 的可靠度計算問題，我們也希望能夠提出一些求逼近值的演算法，且是在容許的誤差範圍內。最後，希望能夠將我們所提出的可靠度分析模式應用在實際的分散式資料庫系統中，並且提供給網路管理與資料庫規劃人員，使其能夠在極短的時間之內得到分散式資料庫的可靠度，以作為規劃出高可靠度分散式資料庫系統之參考，因而能夠降低因網路節點或通信網路線路故障所造成的無法讀取資料庫的風險。同時亦希望能夠將我們所提出的可靠度分析模式應用到現有的分散式資料庫系統中，評估現有的分散式資料庫系統可靠度如何，以做為未來分散式資料庫系統 upgrade 及系統架構以任何不同形式之改變參考，因而能使分散式資料庫系統具有極佳的可靠度。

三、研究方法

本計劃針對網路結構的不同, 來探討分散式資料庫系統可靠度的問題。

首先我們針對平面網路(Planar Network) 的可靠度問題做說明。在平面網路的可靠度評估方法之設計，Kumar 和 Ragharendra 曾提出 MFST 演算法, 此演算法先找出網路的所有最小檔案擴展樹（Minimal File Spanning Tree)，再去計算出分散式程式可靠度(Distributed Program Reliability)。基本上，此演算法是將分散式資料庫系統當成是一個圖形(graph) ，圖形中的節點視為電腦主機，邊則視為通訊連線，當找出圖形中的所有最小檔案擴展樹後，再將此所有最小檔案擴展樹利用 Terminal Reliability Evaluation Algorithm 去計算出分散式程式可靠度。此方法非常的直覺簡單，但當其找出圖形中的所有最小檔案擴展樹後，發現有相當多的重複樹，因此還須移除這些重複樹，故此步驟須執行二回，才可利用 Terminal Reliability Evaluation Algorithm 去計算出分散式程式可靠度，因此執行效率並不好。

Kumar 接著又提出只須執行一回就可以計算出可靠度的 FARE 演算法，此演算法是利用 Connection Matrix 及 Extended Conservation Policy 計算出分散式程式可靠度，此演算法雖然只須執行一回就可以計算出可靠度，但是卻只適用於節點是 Perfect 的情形, 故並不完全適用在實際的分散式資料庫系統。

藉由 Kumar 所提出計算可靠度方法的觀念上，我們構思出新的評估可靠度方法，此方法稱為快速可靠度評估演算法(Fast Reliability Evaluation Algorithm) [1]，此方法是利用 Factoring Theorem 及 Reliability Preserving Reduction 來化簡分散式資料庫系統的圖形, 藉由化簡後的圖形，可快速計算出可靠度，但此方法亦只適用於節點是 Perfect 的情形, 並不完全適用在實際的分散式資料庫系統。不過其執行效率卻較 Kumar 所提出 FARE 演算法來得好。

由於一開始我們均針對節點是 Perfect 的情形來設計出新的評估可靠度方法，但此並不完全適用在實際的分散式資料庫系統，因此我們亦考慮節點是 Imperfect 的情形。

其次我們針對環形網路(Ring Network)

的可靠度問題做說明。在構思環形網路的可靠度評估方法之前，我們先分析環形網路之特性，接著再依分散式資料庫系統的資料分佈，整理出兩種環形網路分散式程式可靠度分析的情形。第一種情形是程式及資料檔均只有一份，第二種情形是程式及資料檔可有多重備份。第一種情形和 K-Terminal Reliability 的分析方法相同，目前我們已經設計出一個方法，可以正確的計算出環形網路的分散式程式可靠度[26]。至於第二種情形，我們亦嘗試設計出 Linear Chain 分散式程式可靠度分析之演算法，並再結合 Factoring Theorem ，而提出求解環形網路的分散式程式可靠度之演算法。

然而，環形網路中的 Ring of Tree 拓樸，在第一種情形即程式及資料檔均只有一份的情形，目前我們已經構思出一個方法，可以正確的計算出環形網路的分散式程式可靠度。至於第二種情形即在程式及資料檔可有多重備份的情形，目前我們尚未找到求解 Ring of Tree 拓樸的分散式程式可靠度之 Polynomial Time 演算法，因此我們正嘗試在這方面努力的探討是否存在有 Polynomial Time 演算法，若無，我們亦要證明此種問題是否 NP-hard 問題。若證明此種問題是 NP-hard 問題，可否找出一有效率的演算法計算出分散式程式可靠度。

另外，星形網路(Star Network) 亦是相當普遍被使用的網路拓樸，在星形網路架構，我們亦嘗試在這方面探討是否存在有 Polynomial Time 演算法來評估分散式程式可靠度。若對程式和資料檔案的分佈做某種程度的限制，則有可能找到 Polynomial Time 演算法來評估星形網路分散式程式可靠度。故我們將先分析程式和資料檔案的分佈情形，以做為設計 Polynomial Time 演算法，用來評估星形網路分散式程式可靠度之基礎。

若是以上各種網路拓樸，均無法設計出 Polynomial Time 演算法來計算分散式程式可靠度之正確解，我們將嘗試設計 Polynomial Time 演算法來求其可靠度之逼近解。目前的想法是將問題以 Sum of Disjoint Product (SDP) 表現，再想辦法使 SDP 的每一項均為 Simple Product ，進而求得各種網路拓樸分散式程式可靠度之逼近解。

四、成果及結論

具體成果如下所述

在平面網路(Planar Network) ，針對節點是 Imperfect 的情形，我們設計了兩種方法，即為 SM (Symbolic Method) 及 FM (Factoring Method) [2]。SM 演算法和 Kumar 提出的 MFST 演算法類似，同樣是先找出網路的所有最小檔案擴展樹，再去計算出分散式程式可靠度。但我們提出的 SM 演算法保證找出的所有最小檔案擴展樹，不會有重複樹出現，因此不需要有移除這些重複樹的步驟。直接可由找出的所有最小檔案擴展樹去計算出分散式程式可靠度。是故 SM 演算法執行效率較 Kumar 所提出的 MFST 演算法來得好。至於 FM 演算法是利用 Factoring Theorem 及 Reliability Preserving Reduction 來化簡分散式資料庫系統的圖形，而求出可靠度。

在環形網路裡，我們主要探討 dual ring 及 ring of tree 這兩種拓樸。針對這兩種拓樸，依據程式及資料檔的分佈，歸納出兩種分散式程式可靠度分析的情形。第一種情形即程式及資料檔均只有一份的情形，第二種情形即在程式及資料檔可有多重備份的情形。

在 dual ring 拓樸，當程式及資料檔均只有一份時，我們提出 polynomial time 演算法去評估分散式程式可靠度[26]。又當程式及資料檔可有多重備份時，我們亦提出 polynomial time 演算法(Algorithm Reliability_Circular_DCS)去評估分散式程式可靠度[27]，此演算法是利用 recursive 方法先設計出分析 linear chain 的演

算法，再進而設計出 dual ring 的 polynomial time 演算法。

在 ring of tree 拓樸，當程式及資料檔均只有一份時，我們提出 polynomial time 演算法去評估分散式程式可靠度[26]。又當程式及資料檔可有多重備份時，我們證明這是 NP-Hard 問題。我們首先證明星形拓樸是 NP-Hard 問題，再進而證明樹狀拓樸亦是 NP-Hard 問題，最後證明 ring of tree 拓樸是 NP-Hard 問題。又針對 ring of tree 拓樸，利用 Factoring Theorem 及 Reduction Method ，我們提出 FREA 演算法以便可以在較快時間評估 ring of tree 拓樸分散式程式可靠度。

在星形網路(Star Network) ，我們只針對程式及資料檔可有多重備份時，評估分散式程式可靠度。由於在 general case 中，我們已證明星形拓樸是 NP-Hard 問題[28]，因此，必須對程式及資料檔的分佈做限制，才有可能設計出 Polynomial Time 演算法來計算分散式程式可靠度之正確解。我們在星形網路中，分析程式及資料檔的分佈，並定義 Consecutive File Distribution Property ，我們利用 Recursive 方法，先探討 Linear Chain 分散式程式可靠度，接著在星形網路中，當程式及資料檔的分佈滿足 Consecutive File Distribution Property 時，我們提出 polynomial time 演算法評估星形網路分散式程式可靠度[29]。

在星形網路中，我們亦定義 File Cutset ，再利用程式及資料檔的分佈限制，提出 另一 polynomial time 演算法去評估星形網路分散式程式可靠度[30] ，同時我們也提出並證明一些 Property 及 Theorem ，這些 Property 及 Theorem 可以幫我們很快分析星形網路中的程式及資料檔的分佈，是否滿足我們提出對星形網路中程式及資料檔的分佈限制之特性，而可利用我們提出 polynomial time 演算法去評估星形網路分散式程式可靠度[30]。

當分析出無法否滿足我們提出對星形網路中程式及資料檔的分佈限制之特性時，以上所提 polynomial time 演算法將不適用。我們設計出 Polynomial Time 演算法來求其可靠度之逼近解，我們將問題以 Sum of Disjoint Product (SDP) 方式表現，再設計出 polynomial time 演算法，且使 SDP 的每一項均為 Simple Product ，進而求得星形網路的分散式程式可靠度之逼近解，並將此逼近解和正確解比較[30]。

五、參考文獻

[1]D. J. Chen and M. S. Lin, "On Distributed Computing Systems Reliability Analysis Under Program Constraints," IEEE Transaction on Computers, Vol. 16, No. 1, January, 1994.

[2]M. S. Lin and D. J. Chen, "General Graph Reduction Methods for the Reliability Analysis of Distributed Systems," The Computer Journal, Vol. 36, No. 7, 1993, pp. 631-644.

[3]A. Kumar, S. Rai and D. P. Agrawal, "On Computer Communication Network Reliability Under Program Execution Constraints," IEEE JSAC, vol. 6, pp. 1393-1399, Oct. 1988.

[4]Kevin Wood, "Factoring Algorithms for Computing K-terminal Network Reliability", IEEE Trans. Reliability, Vol. R-35, pp.269-278, Aug. 1986.

[5]J. Chen, T. H. Huang, " Reliability Analysis of Distributed Systems Based on a Fast Reliability Algorithm," IEEE Trans. Parallel and Distributed System, Vol. 3, No. 2, pp. 139-153, Mar. 1992

[6]M.S. Lin and D. J. Chen, "Reliability Analysis of Distributed Computing System with Imperfect node," The 8th International Joint Workshop on Computer Communication (JWCC-8), Taipei,

Taiwan, Dec. 12-14, 1993, pp. C4-2-1 to C4-2-9.

[7]P. Enslow, "What is a distributed data processing system," Computer, vol. 11, Jan. 1978.

[8]T. C. K. Chou and J. A. Abraham, "Load redistribution under failure in distributed systems," IEEE Trans. Comput., vol. C-32, pp. 799-808, Sep. 1983.

[9]D. A. Rennels, "Distributed fault-tolerant computer systems," Computer, vol. 13, pp. 55-65, Mar. 1980.

[10]T. C. K. Chou and J. A. Abraham, "Load redistribution under failure in distributed systems," IEEE Trans. Comput., vol. C-32, pp. 799-808, Sep. 1983.

[11]J. A. Stankovic, "A perspective on distributed computer systems," IEEE Trans. Comput., vol. C-33, pp. 1102-1115, Dec. 1984.

[12]M. Sloman and J. Kramer, Distributed Systems and Computer Networks, Prentice Hall, 1987.

[13]J. Garcia-Molina, "Reliability Issues for Fully Replicated Distributed Databases," IEEE Computer, vol. 16, pp. 34-42, Sept. 1982.

[14]J. P. Ignizio, D. F. Palmer and C. M. Murphy, "A Multicriteria Approach to Supersystem Architecture Definition," IEEE Trans. Comput., vol. C-31, pp. 410-418, May1982.

[15]A. Satyanarayana and J. N. Hagstrom, "A New Algorithm for the Reliability Analysis of Multi-Terminal Networks," IEEE Trans. on Reliability, vol. R-30, pp. 325-334, Oct. 1981.

[16]A. P. Grnarov and M. Gerla, "Multiterminal Reliability Analysis of Distributed Processing System," in Proc. 1981 Int. Conf. Parallel Processing, Aug. 1981, pp. 79-86.

[17]R. Kevin Wood, "Factoring Algorithms for Computing K-terminal Network Reliability", IEEE Trans. Reliability, Vol. R-35, pp.269-278, Aug. 1986.

[18]L. G. Valiant, "The Complexity of Enumeration and Reliability Problems," SIAM J. Computing, vol. 8, pp. 410-421, 1979.

[19]M. Jerrum, "On the complexity of evaluating multivariate polynomials", Ph.D. thesis, Department of Computer Science, University of Edinburgh, 1981.

[20]O. R. Theologou and J. G. Carlier, "Factoring & Reductions for Networks with Imperfect Vertices," IEEE Trans. Reliability, Vol. 40, No. 2, pp. 210-217, Jun. 1991.

[21]M. R. Garey and D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-completeness, Freeman, San Francisco, 1979.

[22]M. O. Ball , J. S. Provan and D. R. Shier, "Reliability Covering Problems," Networks, vol. 21, pp. 345-357, 1991

[23]J. S. Provan and M. O. Ball, "The Complexity of Counting Cuts and of Computing the Probability that a Graph is Connected," SIAM J. Computing, vol. 12, no. 4, pp. 777-788, Nov. 1983.

[24]A. Satyanarayana and R. Kevin Wood, " A Linear-Time Algorithm for Computing K-terminal Reliability in Series-Parallel Networks," SIAM Journal of Computing, Vol. 14, No. 4, Nov. 1985, pp. 818-832.

[25]P. Winter, "Steiner Problem in Networks: A Survey," Networks, vol. 17, pp. 129-167, 1987.

[26]. D.J.Chen, M.S.Chang, C.L.Yang, K.L.Ku, "Multimedia Task Reliability Analysis Based on Token Ring Network," International Conference on Parallel and Distributed Systems, Tokyo,

Japan, Jun.2-7,1996, pp.265-272.

[27]. M.S.Chang , D.J.Chen, M.S.Lin, K.L.Ku, "Reliability Analysis of Distributed Computing System in Ring Networks" Journal of Communication and Networks, Vol.1, No.1, pp.68-77, March, 1999.

[28] M.S.Chang , D.J.Chen, M.S.Lin, K.L.Ku, "The Distributed Program Reliability Analysis on Star Topologies" Computer and Operations Research, Vol.27, No.2, pp.129-142, Feb. 2000

[29] M.S.Lin, M.S.Chang , D.J.Chen, K.L.Ku," Distributed Program Reliability Analysis: Complexity and Efficient Algorithms ", IEEE Transaction on Reliability, Vol.48,No.1, pp.87-95, March. 1999

[30] M.S.Chang , D.J.Chen, M.S.Lin, K.L.Ku," The Distributed Program Reliability Analysis on a Star Topology: Efficient Algorithms and Approximate Solution ", IEICE Trans. on Information and Systems, Vol.E82-D,No.6, pp.1020-1029, Jun. 1999