

行政院國家科學委員會專題研究計畫 期中進度報告

以模擬進行排序佳化的方法研究及其於生產排程之應用

(2/3)

計畫類別：個別型計畫

計畫編號：NSC92-2212-E-002-060-

執行期間：92年08月01日至93年07月31日

執行單位：國立臺灣大學電機工程學系暨研究所

計畫主持人：張時中

計畫參與人員：吳欣曄、趙圻軒、謝旻旻、林宗慶、林偉誠、陳俊宏教授、謝博偉博士

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 6 月 28 日

以模擬進行排序佳化的方法研究及其於生產排程之應用(2/3)

Simulation-based Ordinal Optimization Methods with Applications to Production Scheduling(2/3)

計畫編號：NSC 92-2212-E-002-060

主持人：張時中教授

計畫參與人員：吳欣曄、趙圻軒、謝旻旻、林宗慶、林偉誠
陳俊宏教授、謝博偉博士

執行期限：92年8月1日至93年7月31日

執行單位：國立台灣大學電機工程學系

一、中文摘要

本報告總結第二年在本計畫的支持下的研究進展，包含三項工作：(1) 結合排序佳化模擬與策略疊代求解派工問題馬可夫決策模型快速演算法研發；(2) 研究增強式學習應用在派工問題的馬可夫決策模型求解的可能性，著重在有設置時間之單一機台派工問題，及可調速率之生產機台問題；(3) 透過文獻回顧推估全球運籌資訊服務體系加入資訊服務網路可能帶來的潛在效益與成本降低，進而評估全球運籌資訊服務對電子業整體市場之潛在效益。

Abstract

The second year of research includes three tasks (1) development of fast simulation for solving the Markov decision process formulation of dispatching problem by combining ordinal optimization and policy iteration, (2) investigation of applicability of re-enforcement learning to Markov decision process formulation of dispatching problems, and (3) niche assessment by literature survey of logistic information application service provider for electronics industry.

關鍵詞：派工、馬可夫決策過程、增強式學習、時變、全球運籌資訊服務體系、資訊服務網路

二、緣由與目的

在現今的工廠中，由於具有多樣化的產品、複雜的製程、客戶導向且高度複雜性與高成本，和機台會當機與定時的維修等特性，所以如何決定將產品有效的派工(Dispatching)給機台，來達到準時交貨，富有彈性地製造不同產品，生產速度快，且降低製造成本仍然是一個具有相當挑戰性的研究課題。另外，環境的不確定性也造成了

不必要的花費與過長的製造時間。所以如何有效的排程派工來使得此產業提升競爭優勢便顯得非常重要。本研究將分別討論派工與機台配置的問題，在工廠中如何有效的派工與配置機台一直是個重要的課題。派工是根據現場的狀況，產生某一工作站的前面等候產品的加工順序，而機台配置在決定如何將機台產能配置給製程站，以達成準時交貨、節省生產週期和降低庫存。

在實際的工廠中，以往的派工決策是靠著工程師的經驗與協調來達成。因此，當工程師無法憑著經驗來作出決定時，這時候往往就會有不夠有效率的法則產生。因此結合學習與模擬的方法，可以模擬實際狀況，在藉由學習機制學出對應的好決定，可以用來預測當時狀況需要作的決策。

我們考慮有設置時間之單一機台派工問題，及可調速率之生產機台問題。我們研究採用增強式學習(Reinforcement Learning)來處理這兩類問題的可能性與效能。增強式學習是一種可以隨著環境不同而持續學習的方法，藉由定義合理的回饋(Reward)，估計價值函數(Value Function)來找到最適合的決策。1.)現今的派工知識多是依賴經驗法則，增強式學習正是學習經驗的方法。2.)現場環境多變，且是隨時間改變，必須持續不斷學習改進派工法則。3.)增強式學習有別於一般學習法，是以嘗試錯誤法來持續改進。4.)增強式學習為一種有理解力與自動化的目標導向式學習，以及決策與學習的計算式(Computational)的方法。5.)增強式學習結合模擬與學習的能力。

另一方面，策略疊代演算法是一種解決馬可夫決策過程(Markov decision process, MDPs)來求得最佳派工策略的方法。本體上該演算法包括了兩個部分。第一部份是策略評估(policy evaluation)：根據當時的策略來評估出一數值。第二部分是策略

改善(policy improvement)。此二步驟不斷的重複一直到某些終止需求達到為止。當MDP的階段成本(stagewise cost)很複雜，必須以模擬來評估時，就使用以模擬為基礎的策略疊代法則(simulation-based policy iteration, SBPI)。當問題的維度很大的時候，進行模擬的次數就會變的相當的龐大，需要快速演算法才有實用的可能。

除排程派工外，建立全球運籌資訊服務體系，為企業全球運籌管理電子化的核心工程。建立運籌資訊服務體系的主要優點，除了在於滿足企業對於運籌資訊的即時性與對運籌流程透明度掌握的要求，以提升企業在運籌鏈的管理績效，並增加客戶的滿意度之外，從網路經濟相關研究的觀點來看，亦能藉由運籌資訊應用服務業者所提供的資訊仲介功能，大幅減少企業與其交易夥伴之間繁複的文件及資料交換之處理與連結次數，進而降低整理運籌體系的成本。換言之，經由資訊仲介的功能，得以使企業與其交易夥伴在文件及資料交換上，從以往多對多的連結方式，轉而成為多對一的連結方式(參見圖八)，以此縮短資訊交換的流程與時間，並且可減少可觀的資料重覆輸入成本，與其衍生的資料輸入錯誤，造成交貨延遲或產生交易糾紛等問題。

然而欲成功地建立全球運籌資訊服務體系，使其能永續經營，除了軟硬體技術、資訊基礎建設、和政府行政上配合之外，仍有許多問題尚待克服。其中最關鍵的問題，是網路經濟理論曾明確指出的典型「chickens and eggs」困境(Shapiro and Varian, 1998)。因為網路的潛在參與者對於加入網路的意願高低，決定在加入該網路後可獲得之價值；而網路之價值本身，又取決於加入網路者的多寡。也就是說，一個網路的加入者多寡和網路價值呈互為因果之關係。

因此本研究之目的，即是透過文獻回顧推估加入此資訊服務網路可能帶來的潛在效益與成本降低，進而評估全球運籌資訊服務整體市場之潛在效益(market potentials)，降低參與者疑慮，並能夠共同享有加入此資訊服務平台後所獲得之網路外部性效益，以達成提升企業運籌管理能力，而增加國際競爭力。

三、結果與討論

I. 應用增強式學習設計派工法則

我們討論有設置時間的單一機台派工問題。我們的目標(objective)是使得產品的等候時間最小。之後我們設計了一個實驗，為兩個產品有設置時間的問題。做不同產品的時候必須花費設置時間(Setup Time)來轉換製程，如在半導體廠做不同產

品時，更換光阻液需要花費時間清洗，此時會花費時間與金錢。然而每個產品的價值又不一樣，價值較高的產品所佔的等候權重較高，所以如果我們希望使產品的平均權重等待時間小，達到單站最大產能，就必須減少轉換製程的次數，使得設置時間盡量不會影響到產品的等候時間。

我們為派工問題建立一個連續時間的馬可夫決策過程(Continuous Time Markov Decision Process)模型，利用法則疊代法(Policy Iteration)與Sarsa[RsA98]的增強式學習法求解，並比較得出的結果與最佳解。結果顯示出隨著學習次數增加，學出來的結果越接近最佳解(見圖二)。這意味著增強式學習可以用來解決這類問題。而與最佳價值函數的誤差也隨著學習次數增加而減少(見圖三)。

由於實際派工問題中，訂單的多寡影響到產品的到達率，而訂單的多寡並不是一種簡單的波松(Poisson)機率分佈而已，它往往受到一些外在的因素影響，比如景氣的好壞，產業的發展性等等。所以產品到達率通常是一種非平穩(Non-Stationary)的分佈，亦即會隨著時間變化而改變其機率特性，於是這種機率分佈變得很難預測，也無法使用法則疊代法來求得最佳法則。由於增強式學習具有持續不斷學習的能力，所以我們將增強式學習應用在時變的派工問題上，並與一個隨機的派工法則來做討論與比較，結果顯示，即使是在時變的環境中，增強式學習也能使平均等待時間趨向穩定(見圖四)，並且讓總切換次數減少(見圖五)，使得總產出增加(見圖六)。

其次，我們研究單一機台產能配置的問題，我們所看的是有當機可能及可以調整平均生產速率的機台，生產速率的控制策略代表產能的配置。當機台在正常運作下且機台前的等待工作總數超過生產速率切換值 k 時，機台以低的速率運作，反之，則採用高速率。我們模擬工廠機台的產能分配的問題。當等候區的產品達到某一程度的數目後，則將速度調高或調低，以達成我們所要的需求。因此我們可以藉由定義不同的回饋，由增強式學習得到在不同回饋下，所得出 k 值的控制策略(見表一)。我們利用已經學習出來的最佳 k 值，在改變到達率的情況下，嘗試是否能學出隨環境變化的 k 值，結果顯示：增強式學習在已有最佳法則後，隨著環境變遷學習的能力增加了(見表二)。這代表著，如有適當的已知法則，將使得增強式學習能夠快速學到最佳解。對於實際生產線的派工學習，則有待進一步的評估。

II: 結合排序佳化模擬與策略疊代求解派工馬可夫決策模型[6]

我們創新的結合排序佳化 (ordinal optimization, OO) 以及策略疊代 (policy iteration) 的概念, 發展了一個快速求解派工問題馬可夫決策模型的演算法 (OOBPI) 如圖七。該演算法包含兩個部分: 第一部份是 OO 的部分。我們利用排序佳化的優點, 在每一次的疊代中, 對於每一個狀態的決策選擇來進行排名以及作信心機率 (APCS) 的計算。當所排名最佳決策的 APCS 值比預設的信心程度來的高時, 就選擇該決策為最佳決策。第二個部分是策略疊代。我們將原本的策略跟改善後的策略加以比對, 當兩個策略完全一模一樣的時候, 我們停止所有的模擬, 然後宣稱最後的那個策略為最佳策略。初步模擬結果顯示這個方法較傳統模擬為基礎的策略疊代法的計算效能可快百倍。OOBPI 法實際的計算效能如何則有待我們更進一步的分析驗證。

III. 評估全球運籌資訊服務對電子業整體市場之潛在效益

在回顧舊有文獻的過程中, 我們發現了許多性質類似的資訊服務網路成功先例, 這些服務網路透過提昇使用者的管理績效, 使使用者成本降低等方式來增加其滿意度, 進而達到成功營運的目的。在這些成功的例子中, 我們舉出較著名的案例詳述, 並提出其中值得參考之處, 包括如何令使用者及其合作夥伴能共同協調一致地加入資訊服務的網路平台, 提高並共享此資訊服務所帶來的網路經濟效益等等。同時也將之與一些失敗的案例比較, 提出其中的不同點, 用以了解決定這些資訊服務成功與否的關鍵因素。

成功案例 -- Traxon 空運:

Traxon 所建構的網路系統是以提供下列服務為目標:

- (1) 使參與運籌體系的企業激能夠以電子資訊的方式, 藉由幾乎零時差的網路來傳遞訊息。
- (2) 簡化以使資訊的傳遞更有效率。
- (3) 使用產業標準規格來簡化資訊。
- (4) 加速資訊流在各企業之間的流動效率。

成功因素

對於 Traxon 的成功因素, 主要的關鍵因素應可歸納為以下兩項:

- (1) Traxon 的系統在各地都有據點, 這是它的優勢之一。主要創立的四家航空業者分別在香港, 日本, 法國以及德國, 此外在韓國等地也有業者參與合作。因此對於各地的市場情況都十分了解, 使得一開始的使用者群就散佈在世界各地, 造成接下來國際化很大的優勢。
- (2) 在 Traxon 的例子中, Traxon 以小心謹慎限制服務提供的項目來解決先前提到「chickens and eggs」困境的問題, 由於目標的客戶群, 也就是空運業者, 往往擔心利用了像網路這樣

透明化的機制, 會使得原本他們用以營利的技術受到侵害, 陷入無法繼續經營的危機。為了使使用者減低這樣的疑慮而樂於參與這項體系, Traxon 小心合理的限制所提供的服務範圍, 提供使用者認為便捷卻不遭受威脅的服務, 支持整個原有的運籌鍊, 而不試圖去改變運籌鍊的架構。因此在 Traxon 初期便有許多使用者願意加入, 而突破了這個困境。

失敗案例 -- Reuters (路透社):

失敗的原因探討

空運業者與其客戶之間的主要互動可以分成兩項: 運輸的空間與運輸的服務。但 Reuters 的創始者表示這兩項服務是可以分開被探討的, 在 Reuters 的原始概念中, 價格才是主要被探討的因素, 可以與服務完全分開, 因此在它提供的資訊裡, 僅僅只包含價格, 而不對服務的品質激以描述, 因此它的架構類似以下:

- (1) 僅僅提供價格的資訊而沒有任何產品或服務的訊息
- (2) 將空運的市場看作日用品的市場, 服務提供者之間的競爭以價格完全取代服務品質
- (3) 空運業者只被認為是運送空間的提供者

由於 Reuters 這樣的態度使得空運業者及航空業者不願意加入這項系統。除此外, 由於 Reuters 的原始創辦單位都是一些資訊媒體提供者, 與 Traxon 不同, 其創辦者是空運業者, 亦即運籌鍊中的一環, 且 Reuters 的創辦業者以提供訊息為主要的服務項目, 將整個運籌鍊中的資訊流完整的透明化, 造成目標客戶擔心自己的業務內容過分被公開而造成營運的威脅, 因而拒絕參與這項體系。

由表一看來, Traxon 所包含的資訊流內容僅涵蓋諮詢流程的 5, 7, 10 及 15 (參見圖九), 而 Reuters 則涵蓋了全部的範圍, 使的市場透明化的程度增加而令空運業者卻步, 造成失敗的主因。

在文獻中, 也有許多事關於使用者對 3PL 產業的調查報告, 藉由觀察國外所作的調查, 可以發現 3PL 產業已經是未來的趨勢, 企業對於這項資訊服務的依賴性也日益增長, 最重要的是, 在調查之中, 對於這項產業的評價普遍都是一片好評 (參見表三, 表四)。

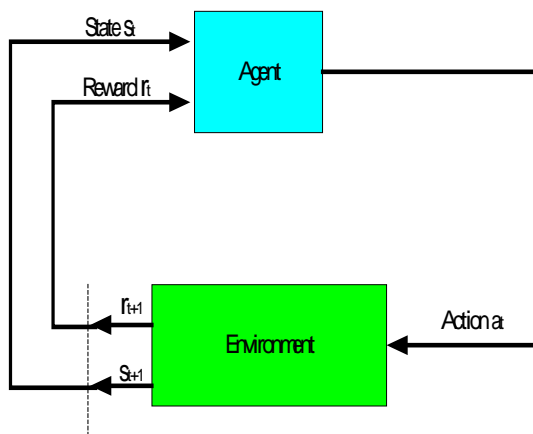
III. Publications

1. H.-Y. Wu, S.-C. Chang, "Design of Job and Machine Dispatching Policy by Reinforcement Learning" Master Thesis, Dept. of Electrical Engineering, National Taiwan University July 2004.
2. [HCC03] B.-W. Hsieh, S.-C. Chang, C.-H. Chen, "Efficient Selection of Scheduling Rule Combination by Combining Design of

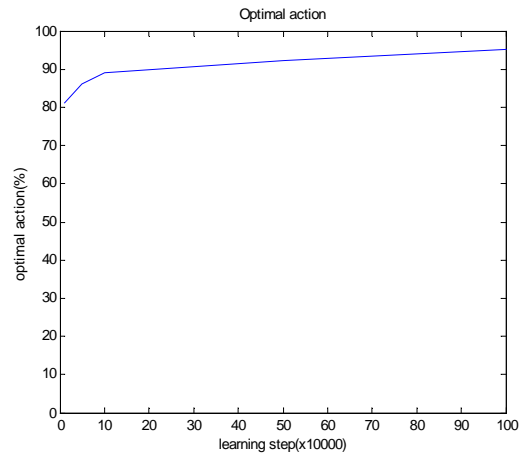
Experiment and Ordinal Optimization-based Simulation,” *Proceedings of ICRA2003*, Taipei, Sept., 2003; best automation paper award.

四、參考文獻

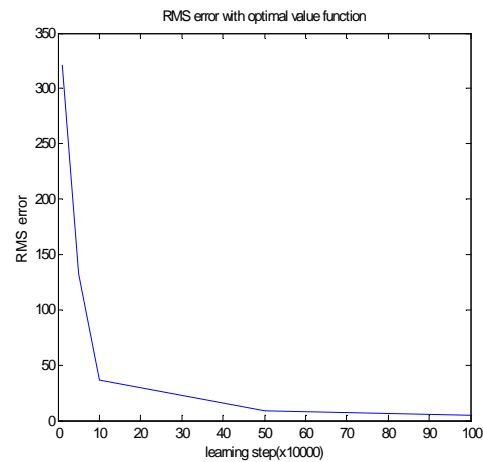
- [1] Richard S. Sutton and Andrew G. Barto, “Reinforcement Learning: An Introduction”, Cambridge, Mass. : MIT Press, c1998.
- [2] Chen I.C.,” Stochastic Production Flow Control: Single Station Case” NTUEE Master thesis 2000.
- [3] Y.Z. Chang”A Learning Agent for Supervisors of Semiconductor Tool Dispatching,” NTUEE Master Thesis June 2002.
- [4] E. Chirstiaane, “A Comparison of Electronic Infrastructures in the Air Cargo Industry in the Netherlands and Hong Kong SAR,” *Journal of Global Information Management*
- [5] J. Damsgaard, “Global Logistics System Asia Co .,Ltd , ” *Journal of Information Technology* (1999) 14 , 303-314 .
- [6] [HCC01] B.-W. Hsieh, C.-H. Chen, S.-C. Chang, “Scheduling Semiconductor Wafer Fabrication by Using Ordinal Optimization-based Simulation,” *IEEE Trans. on Robotics and Automation*, Vol. 17, Issue: 5 , Oct. 2001, pp. 599 -608.
- [7] B.-W. Hsieh, S.-C. Chang, C.-H. Chen, “Efficient Selection of Scheduling Rule Combination by Combining Design of Experiment and Ordinal Optimization-based Simulation,” *Proceedings of ICRA2003*, Taipei, Sept., 2003.



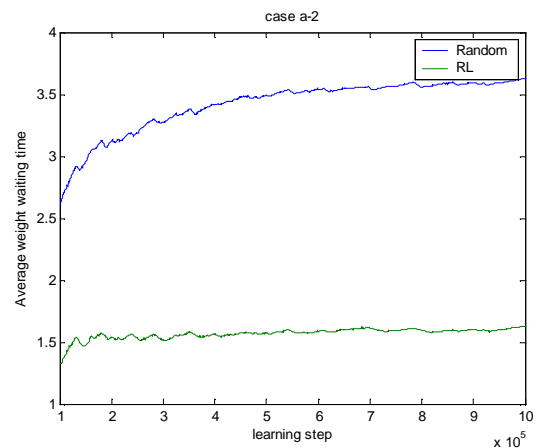
圖一：增強式學習示意圖



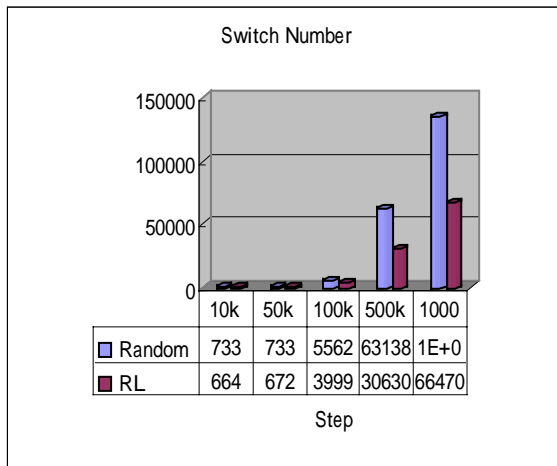
圖二：最佳決策與學習次數圖



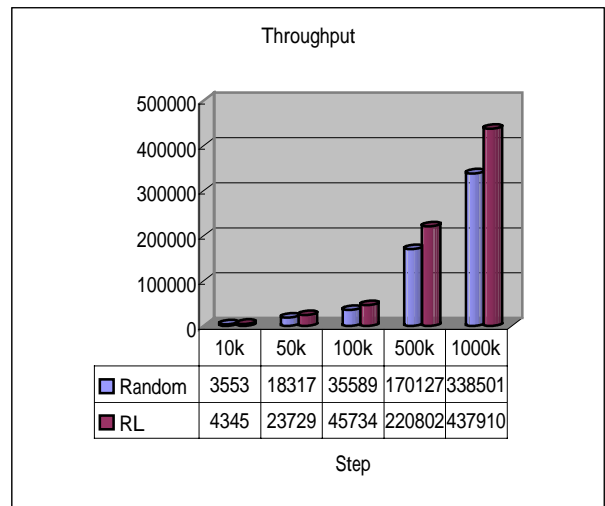
圖三：學習與最佳價值函數的誤差



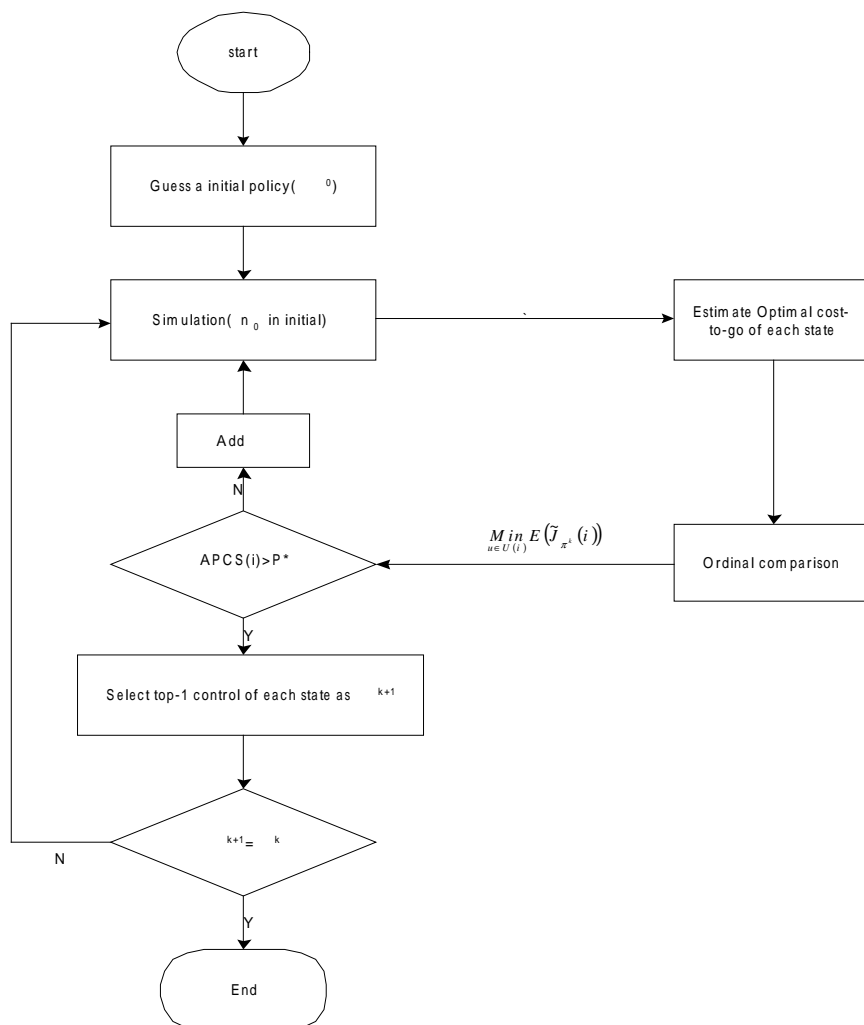
圖四：增強式學習與隨機法則之平均等候時間比較



圖五: 增強式學習與隨機法則設置次數比較



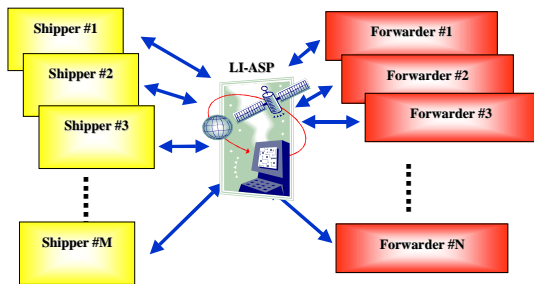
圖六: 增強學習式與隨機法則產量比較



圖七: OOBPI 流程圖

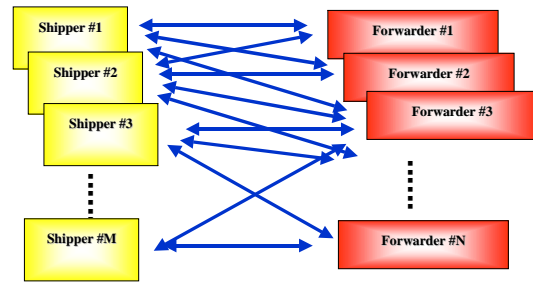
交易過程: 多對一的連結方式

多對一架構下的整體連結次數 = $M + N$

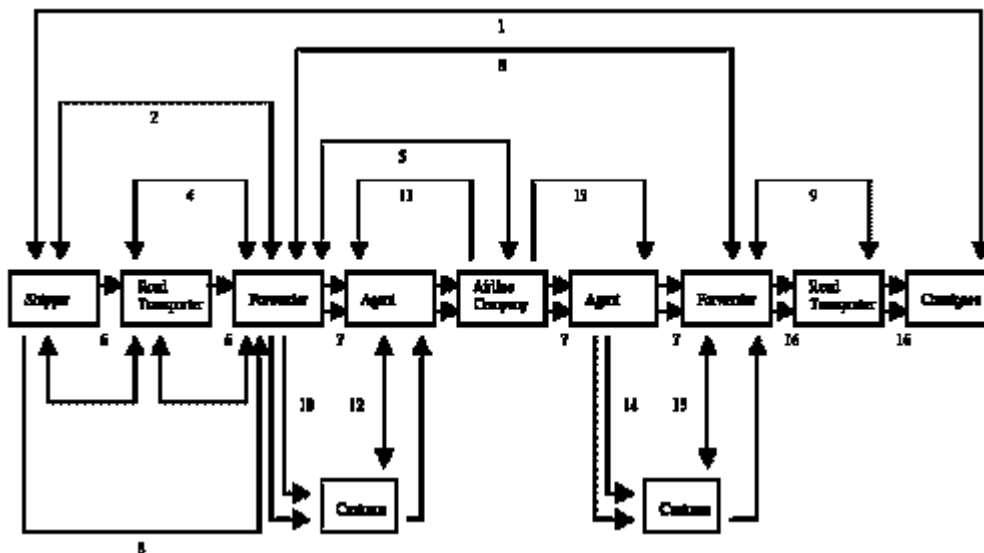


交易過程: 多對多的連結方式

多對多架構下的整體連結次數 = $M \times N$



圖八: Shipper 與 Forwarder 在不同資料連結架構下之連結次數比較



圖九. 傳統空運運籌鍊的資訊流

K^*	2	2	2	3	3	3	3	4	4	4
λ	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

表一: 最佳轉換點與到達率的關係

$\lambda = 1 \rightarrow 0.9$	
Learning step	RMS
1778	12.58
666	30.14
6756	34.052

1903	22.75
2809	19.44
922	33.33
1142	35.65
1514	37.31
951	40.08
1484	21.12
6194	4.17

表二：改變到達率在已有最佳法則下學習效果

	Traxon	Reuters
創辦者 & 所有人	四家主要航空業者	外部成員：新聞業者
創辦者的主要目的	促進空運市場更加有效率與合作	收集費用等消息
目標客戶	空運與航空業者	空運與航空業者
運籌鍊動態	保留現有的運籌鍊	預期消去空運的中介者
市場動態	保留市場的隱密性	使市場透明化
初期市場佔有率	四家主要航空業者	無
主要服務	定位與確認	價位比較
結果	獨占市場的地位	計劃終止

表三：Traxon 與 Reuters 的主要不同比較

Issue	Very Negative (%)	Negative (%)	No Impact (%)	Positive (%)	Very Positive (%)
Employee morale	3	30	44	18	5
Logistics costs		14	11	65	10
Customer satisfaction	2	11	23	59	5
Systems development/support	2	13	47	35	3
Logistics service levels		7	21	67	5

表四：使用 3PL 服務對企業的各项影響

Financial Category	Average Financial Improvement (%)	
	2000	1999
Sales/revenue improvement	4	7
Logistics expense reduction	12	10
Working capital improvement	4	8
Capital asset reduction/avoidance	18	20

表五：使用 3PL 服務造成的經濟影響，2000