行政院國家科學委員會專題研究計畫　成果報告

模擬最佳化演算法則的研擬與軟體製作(2/2)

計畫類別：個別型計畫
計畫編號：NSC94-2213-E-009-044-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：國立交通大學電機與控制工程學系(所)

計畫主持人：林心宇

計畫參與人員：洪士程，黃榮壽，林成梓，林啟新

報告類型：完整報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 95 年 9 月 28 日

國科會研究計劃成果期末報告

計劃名稱：模擬最佳化演算法則的研擬與軟體製作(2/2)

計劃執行起迄：2005.08.01 至 2006.07.31

計劃主持人：林心宇

計劃編號：NSC94-2213-E-009-044-

執行機關：國立交通大學電機與控制工程研究所

計畫中文摘要

模擬最佳化(simulation optimization)是最佳化方法領域中最新的發展，模擬最佳化問題主要是指對任一個 input variable setting 的 objective function 的 evaluation 都須利用模擬(simulation)的方法來求得，所以我們無法以傳統最佳化方法解決此類型的的問題。此類問題涵蓋的範圍甚廣，如一些具有廣大輸入變數空間(huge input-variable space)的隨機模擬最佳化問題(stochastic simulation optimization problem)，大型系統中具有決策變數與 discrete variable 的最佳化問題等等。

本計劃在上一年度中，已獲致兩項研究成果[15]，[16]。在這一年度中，我們更針對 polling system 使用 k-limited service discipline 的模擬最佳化問題作深入的研究，並以序的最佳化方法(ordinal optimization)來求得一個不錯的 service discipline。

Abstract

Simulation optimization is one of the most frontier research area in optimization. The main characteristics of simulation optimization problem is the evaluation of the objective function of an input-variable setting requires lengthy simulation. Therefore we cannot use the conventional optimization techniques to solve them. There are various types of simulation optimization problems such as stochastic optimization problems with huge input-variable space and large scale optimization problems with decision and discrete control variables.

In the first year of this project, We have obtained two research results. In this year , we have studied the simulation optimization problem on polling system using k-limited service discipline and use ordinal optimization method to obtain a good enough solution.

一、前言

在 1992 年時，Professor Azadivar 在[1]中給模擬最佳化問題(simulation optimization problem)下了一個簡單的定義：對任一個 input(variable)的 objective function 的 evaluation 都需利用模擬(simulation)才能得知其 objective value 的最佳化問題即是模擬最佳化問題。此類型問題涵蓋範圍甚廣，如在隨機模擬最佳化問題(stochastic simulation optimization)中需要以隨機模擬(stochastic simulation)來計算一個input variable 的 objective value，由於每一個隨機模擬都需相當長的計算時間，所以如果輸入變數空間(input-variable space)很龐大的話，那麼要得到最佳解所需要的計算量實在是無比的冗長。在上一年度計畫中，我們已完成兩項成果：

(i) 我們已將晶圓測試程序中減少晶粒誤宰及重測的問題 formulate 成一個模擬最佳化問題，並研擬出一個 two-level ordinal optimization 的 algorithm 成功地解決了此問題。此成果所撰寫成的論文[15]已被 IEEE Trans. on Systems, Man and Cybernetics Part A 期刊 accept.

(ii)對具離散控制變數的分散式最佳電力潮流問題上已初步研擬出一個連續變數的分散式最佳電力潮流演算法。此成果已於 2005 年 6 月 15 日至 17 日在 EuroPES 2005 於西班牙召開的會議中發表[16]。

在本年度中，我們更進一步地將所研擬的解模擬最佳化問題的序的最佳化演算法則應用到 polling system 上。

二、研究目的

在電腦網路、通訊網路、製造系統、運輸系統等眾多重要電子與傳統工業的領域中，其某些服務型態常以 polling model 的方式出現，而在 polling model 中如何設定 service discipline 便成為這些應用領域擴大其獲益的重要手段，所以我們這一年度的計畫內容及研究成果對工業的貢獻將相當大。

三、文獻探討

在 polling system 中典型的 service disciplines 計有 exhaustive，gated，limited，k-limited 及 time-limited，等多種，它們在[2]-[5]中皆有詳盡的探討，這些現有的分析方法，大都根據 queuing theory 來推演其結果，所以他們都必須做相當強的假設，例如 arrival process 是 Poisson，service time 是 exponential distribution 或甚至如 buffer size 是∞等。所以他們的結果並不切合實際狀況，而我們擬將 polling system with k-limited service discipline 形成一個模擬最佳化問題。此類型問題在[6]-[11]中有相當多的探討，但與他們不同的是，我們將採用序的最佳化方法[12]-[14]來解它。

四、研究方法

我們所考慮的 multi-queue polling model 是一個 G/G/1/K polling model，它假設 arrival process 是 general probability distribution，service time 也是 general probability distribution，這個 model 只有一個 cyclic server，但是有 J 個 queues 而每個 queue 的長度都是 K(有限的)。這樣的 polling model 便如圖 1 所示。

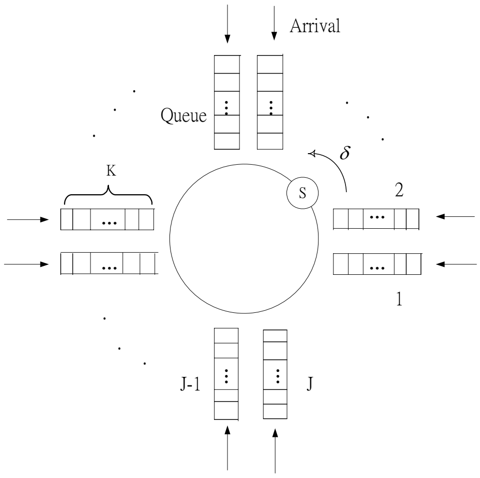

Figure.1.A G/G/1/K polling model

我們可以將如何設定 k-limited service discipline 以使 polling system 的平均等待時間(average waiting time)最小化的問題形成一個模擬最佳化的問題，其細節如下所述：

The k-limited service discipline, that is when the server attends the jth queue, it will serve for $m_j(\leq K)$ customers (jobs or packets) or until the queue becomes empty, whichever comes first. Thus, $(m_1, m_2, \cdots, m_K)$ is a decision vector of our k-limited service discipline. We assume that a customer that completing the service will depart from the system, and any customer being served but yet completed will not be interrupted by any reason. The switchover time of the cyclic server switching to next queue is assumed to be of normal distribution with mean δ and variance σ^2.

We denote the random variable W_j as the waiting time of a typical customer of the j th queue at steady state. The waiting time is defined as the time length from arrival instant until the beginning of service. Then, $E[W_j]$ represents the average waiting time of a customer in the j th queue at steady state.

We let τ_j denote the weighting coefficient of queue j, then $\frac{1}{J}\sum_{j=1}^{J}\tau_j E[W_j]$ denotes the weighting average waiting time of the G/G/1/K polling system.

Now, we can formulate our stochastic simulation optimization problem for the G/G/1/K polling system as

$$\min_{m_j, j=1,\ldots,J} \frac{1}{J}\sum_{j=1}^{J}\tau_j E[W_j] \quad (1)$$

subject to the G/G/1/K polling model.

In other words, we are looking for an optimal k-limited service discipline to minimize the weighting average waiting time of a G/G/1/K polling model. Suppose $K = 20$, then the size of the decision variable space of (2) will be J^{20} or 10^{20} provided that $J = 10$. Note that m_j is allowed to be more than K, because during the period when the server serves a customer, new customers may arrive; however, such an $m_j (> K)$ should not be a good choice unless the arrival rate is very high. Thus, problem (1) is a stochastic optimization problem with huge decision variable space as depicted in (1) and is especially suitable for the proposed ordinal optimization approach to solve for a good enough k-limited service discipline.

五、結果與討論

我們已研擬出一個序的最佳化演算法來求解一個不錯的 k-limited service discipline。我們也將所得到的結果與現有的 service disciplines 來比較並發現我們的結果好的非常多。我們已將這些研究發現撰寫成論文，如附件一，並發表於今年(2006) 6 月 26-28 日在希臘羅德島召開的 IASTED Simulation and Modeling 國際會議[17]。同時，我們也在文章中註明此研究成果係本國科會計畫所贊助，並將計劃編號明列其上。

六、參考文獻

[1] Azadivar, F.(1992) Tutorial on Simulation Optimization, In *Proceedings of the 1992 Winter Simulation Conference*, Arlington, ed. J. Swain, D. Glodsman, R. C. Crain, and J. R. Wilson, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 198-204.

[2] Takagi, H. (1989). Analysis of polling systems with a mixture of exhaustive and gated service disciplines. Journal of the operations research society of Japan, vol. 32, no. 3, pp. 450-461.

[3] Takagi, H. (2000). Analysis and application of polling model. In: Performance Evaluation: Origins and Directions, eds. G. Haring, C. Lindemann and M. Reiser, Lecture Notes in Computer Science, vol. 1769, pp. 423-442, Springer, Berlin.

[4] Fujian, L., R.-H. Sun, C.-Z. Wu and Z.-J. Liu (1997). Cyclic Service Systems with Server Time-outs. 2nd IEEE Symposium on Computers and Communications, Alexandria, Egypt, pp. 43-47.

[5] Hirayama, T., S. J. Hong and M. Krunz (2004). A New Approach to Analysis of Polling Systems. Queueing Systems, vol. 48(1-2), pp. 135-158.

[6] Carson, Y. Maria, A. (1997) Simulation Optimisation Methods And Applications, in *Proceedings of the 1997 Winter Sumulation Conference*, Georgia, ed. S. Andradottir, K. J. Healy, D. H. Withers, B. L. Nelson, Institute of Electrical and Electronics Engineers, Piscatatway, New Jersey, 118-126.

[7] Fu, M. C. (1994) A Tutorial Review of Techniques for Simulation Optimization, In Proceedings of the 1994 Winter Simulation Conference, Lake Buena Vista, ed. Jeffey D. Tew, S. Manivannan, D. A Sadowski, and A. F. Seila, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 149-156.

[8] Johnson, D. S., Aragon, C. R., McGeoch, L. A., and Schevon, C. (1989) Optimization by Simulated Annealing: An Experimental Evaluation; part I, Graph Partitioning, *Operations Research*, 37(6): 865-892.

[9] Lee, Y. H. and Iawate, K. (1991) Part Ordering through Simulation-Optimization in an FMS. *International Journal of Production Research*, 29(7): 1309-1323

[10] Paul, R. J. and Chanev, T. S. (1998) Simulation Optimization Using a Genetic Algorithm, *Simulation Practice and Theory*, 6(6):601-611

[11] Tompkins, G. and Azadiver, F. (1995) Genetic Algorithms in Optimizing Simulated Systems, In *Proceedings of the 1995 Winter Sumulation Conference*, ed. C. Alexopolous, K. Kang, W. R. Lilegdon, and D. Goldsman, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 757-762.

[12] Ho, Y.C., Cassandras, C.C., Chen, C.H., and Dai, L., “Ordinal optimization and simulation,” *Journal of Operation Research Society*, vol. 21, pp. 490-500, 2000.

[13] Lau, T.W.E. and Ho, Y.C., “Universal alignment probability and subset selection for ordinal optimization,” *JOTA*, vol. 39, no. 3, June 1997.

[14] Lin, S. and Ho, Y.C., “Universal alignment probability revisited,” *JOTA*, pp.299-407, May 2002.

[15] Shin-Yeu Lin and Shih-Cheng Horng, ”Application of an Ordinal Optimization Algorithm to the Wafer Testing Process”, to appear in the IEEE Transactions on Systems, Man and Cybernetics, Part A.

[16] Shin-Yeu Lin and Ch’i-Hsin Lin, ”A Distributed Optimal Power Flow Algorithm, ”Euro PES 2005, Benalmadena, Spain.

[17] Shin-Yeu Lin and Shih-Cheng Horng,” Ordinal Optimization Approach to Stochastic Simulation Optimization Problems and Applications”, Proceedings of the 15th IASTED International Conference on Applied Simulation and Modeling, pp. 274-279, Rhodes, Greece, June 26~28, 2006.

七、計劃成果自評

本計劃總計進行兩年，所獲成果堪稱豐碩，共計已發表了一篇知名國際期刊論文，以及兩篇國際會議論文，同時，我們也在這三篇文章中註明此研究成果係本國科會計畫所贊助，並將計畫編號明列其上。

附件一

Proceedings of the 15th IASTED International Conference
APPLIED SIMULATION AND MODELLING
June 26-28, 2006, Rhodes, Greece

ORDINAL OPTIMIZATION APPROACH TO STOCHASTIC SIMULATION OPTIMIZATION PROBLEMS AND APPLICATIONS

Shin-Yeu Lin[†], Shih-Cheng Horng[†,§]
[†] Department of Electrical & Control Engineering
National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
e-mail: sylin@cc.nctu.edu.tw
[§] Department of Electronic Engineering
Chin Min Institute of Technology, Miaoli, Taiwan, R.O.C.

ABSTRACT
In this paper, we propose an ordinal optimization approach to solve for a good enough solution of the stochastic simulation optimization problem with huge decision-variable space. We apply the proposed ordinal optimization algorithm to G/G/1/K polling systems to solve for a good enough number-limited service discipline to minimize the weighting average waiting time. We have compared our results with those obtained by the existing service disciplines and found that our approach outperforms the existing ones. We have also used the genetic algorithm and simulated annealing method to solve the same stochastic simulation optimization problem, and the results show that our approach is much more superior in the aspects of computational efficiency and the quality of obtained solution.

KEY WORDS
Ordinal optimization, stochastic simulation optimization, neural network, genetic algorithm, polling system, average waiting time.

1. Introduction

Simulation optimization problems could be viewed as optimization problems of a simulated system whose outputs can only be evaluated by simulations, which can be either a real simulation of the simulated system or simply a computer simulation [1]. Thus, the objective of simulation optimization is to find the optimal settings of the decision variables to the simulated system that makes the output variables at their best or optimal conditions. Various methods had been developed for this purpose such as the Gradient Search based methods [2], the Stochastic Approximation methods [3], the Response Surface method [4], and Heuristic methods. These methods had been thoroughly discussed in [5]. Among them, the Heuristic methods including the Genetic Algorithm (GA) [6], the Simulated Annealing (SA) method [7], and the Tabu Search (TS) method [8] are frequently used in simulation optimization [9]. Despite the success of several applications of the above heuristic methods [10], many technical hurdles and barriers to broader application remain as indicated in [11]. Chief among these is speed, because using the simulation to evaluate the output variables for a given setting of the decision variables is already computationally expensive not even mention the search of the best setting provided that the decision-variable space is huge. Furthermore, simulation often faces situations where variability is an integral part of the problem. Thus, stochastic noise further complicates the simulation optimization problem. The *purpose* of this paper is to resolve this challenging stochastic simulation optimization problem efficiently and effectively.

The considered stochastic simulation optimization problem is stated in the following

$$\min_{\theta\in\Theta} E[J(\theta)] \quad (1)$$

where Θ is a huge decision-variable space, $E[(\cdot)]$ denotes the expectation of $(\cdot)$, and $J(\cdot)$ denotes the output or a function of outputs of the simulated system, and $E[J(\theta)]$ represents the objective function. To cope with the computational complexity of this problem, we will employ the Ordinal Optimization (OO) theory based goal softening strategy [12]-[13], which efficiently seeks a good enough solution with high probability instead of searching the best for sure based on the observation that the performance order of the decision-variable settings is likely preserved even evaluated by a surrogate model. From here on, we will use the word *setting* to represent the setting of decision variables.

The basic idea of the OO theory based goal softening strategy is to reduce the searching space gradually, and its existing searching procedures can be summarized in the following [12]: (i) Uniformly select N, say 1000, settings from Θ. (ii) Evaluate and order the N settings using a surrogate model of the considered problem, then pick the top s, say 35, settings to form the Selected Subset (SS), which is the *estimated* Good Enough Subset (GS). (iii) Evaluate and order all the s settings in SS using the exact model, then pick the best setting among the s. The OO theory had shown that for $N = 1000$ in (i) and a surrogate model with moderate noise in (ii), the best setting selected from (iii) with $s \cong 35$ *must* belong to the GS with probability 0.95, where GS represents a collection of the top 5% *actually* good enough settings among the N. This means the best setting in SS selected from (iii) is among the actual top 5% of the N settings with probability 0.95. However, the good enough solution of problem (1) that we are searching for should be a good enough setting in Θ instead of the N settings unless Θ is as small as N [14]. As indicated in a recent paper by Lin and Ho [15], under a moderate modeling noise, the top 3.5% of the uniformly selected N settings will be among the top 5% settings of a huge Θ with a very high probability (≥ 0.99), and the best case can be among the top 3.5% settings of Θ provided that there is no modeling error. However, for Θ with a size of 10^{30}, a top 3.5% setting is a setting among the top 3.5×10^{28} ones. The solution among the top 3.5×10^{28} of the 10^{30} solutions is not convincing to be a good enough solution with high probability in the sense of practical applications. Therefore to apply the existing goal softening searching procedures, we need to develop a new scheme to select N *roughly good* settings from Θ to replace (i) so as to ensure the final selected-setting is an actually good enough solution of (1) with high probability.

Heuristic methods for obtaining N roughly good settings may depend on how well one's knowledge about the considered system. For instance in the optimal power flow problems with discrete control variables, Lin, *et al.* [16] proposed an algorithm based on the OO theory and engineering intuition to select N roughly good discrete control vectors. However, the engineering intuition may work only for specific systems. Thus, in this paper, we will propose an OO theory based systematic approach to select N roughly good settings from Θ and combine with the existing goal softening searching procedures to find a good enough solution of (1). The presentation of this OO theory based algorithm to solve (1) for a good enough solution is a *novel approach* in the area of stochastic simulation optimization and is the *contribution* of this paper. Application of the proposed algorithm to a stochastic simulation optimization problem of a G/G/1/K polling system, which will be introduced and formulated in Section 4, is another *contribution* of this paper.

We organize our paper in the following manner. In Section 2, we will describe our approach for finding N roughly good settings from Θ. In Section 3, we will present the proposed OO theory based algorithm to solve for a good enough solution of the stochastic simulation optimization problem. In Section 4, we will present the G/G/1/K polling model and describe the corresponding stochastic simulation optimization problem for minimizing the weighting average waiting time. In Section 5, we will compare the results obtained by our approach with those obtained by the existing service disciplines. In addition, we will demonstrate the computational efficiency and the quality of the obtained good enough solution of our approach by comparing with the genetic algorithm (GA) and simulated annealing (SA) method. Finally, we will make a conclusion in Section 6.

2. Fiding N Roughly Good Settings from Decision-Variable Space

As indicated in the OO theory [12]-[13], performance "order" of the settings is likely preserved even evaluated using a crude model. Thus, to select $N(=1000)$ roughly good settings from Θ without consuming much computation time, we need to construct a surrogate model that is computationally easy to estimate the objective value of (1) for a given setting θ, and use an efficient scheme to select N roughly good settings. Our surrogate model is constructed based on an ANN [17], and our selection scheme is GA [6].

2.1 The Artificial Neural Network (ANN) Based Model

Considering the inputs and outputs as the settings $\theta \in \Theta$ and the corresponding objective values $E[J(\theta)]$, respectively, we can use an ANN to implement the mapping from the inputs to the outputs [17]. First of all, we will select a representative subset of Θ by uniformly picking M, say 500, settings from Θ. Then we will evaluate $E[J(\theta)]$ of these M settings using an *approximate model*, which can be a stochastic simulation with moderate number of random test samples, say 1000 random test samples, as indicated in [14]. These collected M input-output pairs of $(\theta, E[J(\theta)])$ will be used to train the ANN to adjust its arc weights. Once this ANN is trained, we can input any setting θ to obtain an estimation of the corresponding $E[J(\theta)]$ from the output of the ANN; in this manner, we can avoid an accurate but lengthy stochastic simulation to evaluate $E[J(\theta)]$ for a given θ. This forms our surrogate model to estimate the objective value of (1) for a given setting θ roughly but efficiently. ANN is considered to be a universal function approximator [17] including the relationship between the input and output of the discrete event simulated systems as presented in [18] and [19], however, the approximation accuracy is closely related to the complicacy of the structure of the ANN. In other words, there is a tradeoff between the accuracy and training time. Since what we care here are the *relative* performance order of θ's rather than the values of $J(\theta)$'s, we can employ a simple two-layer feedforward ANN as our model. To speed up the convergence, we employ the scaled conjugate gradient algorithm [20] to train the ANN.

2.2 The Genetic Algorithm (GA)

By the aid of the above effective and efficient objective value evaluation model, we can efficiently search N roughly good settings from Θ using heuristic global searching techniques. Since GA improves a pool of populations from iteration to iteration, it should best fit our needs. The population in GA terminology represents a setting θ in our problem, and each setting is encoded by a string of 0s and 1s. We start from I, say 5000, randomly selected settings from Θ as our initial populations. The fitness of each setting is set to be the reciprocal of the corresponding objective value $E[J(\theta)]$ (provided that $E[J(\theta)] > 0$, $\forall \theta \in \Theta$) computed based on the ANN. The members in the mating pool are selected from the pool of populations using roulette wheel selection scheme based on the fitness values. We set the probability of selecting members in the mating pool to serve as parents for crossover, p_r, to be 0.7. We use a single point crossover scheme and assume the mutation probability to be 0.02. We stop the GA when the number of generations exceeds 20. After the applied GA converges, we rank the final I populations based on their fitness values and pick the top N populations, which are the N roughly good settings that we look for.

2.3 Searching the Good Enough Solution Among the N

Starting from the selected N roughly good settings, we will proceed directly with step (ii) of the existing goal softening searching procedures described in Section 1. In this step, we will evaluate the objective value of each setting using a more refined model than the ANN, that is a stochastic simulation with moderate number of random test samples. We will then order the N settings based on the estimated objective values and choose the top s settings to form the Selected Subset (SS). Subsequently, we will evaluate each of the s, say 35, settings using the exact model of the considered problem as indicated in step (iii) of the existing goal softening searching procedures. The exact model is a stochastic simulation with sufficiently large number of random test samples that makes the value estimation of $E[J(\theta)]$ for a given θ sufficiently stable. The setting associated with the smallest objective value of (1) among the s is the good enough solution that we seek.

3. The Ordinal Optimization (OO) Theory Based Algorithm

3.1 The Algorithm

Now, our OO theory based algorithm to solve for a good enough solution of (1) can be stated as follows.

Step 1: Uniformly select M θ's from Θ and use an approximate model to compute the corresponding $E[J(\theta)]$'s using 1000 random test samples. Train an ANN by adjusting its arc weights using the mapping between the given M input-output pairs, that are the M $(\theta, E[J(\theta)])$ pairs.

Step 2: Randomly select I settings from Θ as the initial populations. Apply a GA equipped with a simple roulette-wheel selection scheme, $p_r = 0.7$, a single-point crossover scheme and a 0.02 mutation probability to these populations by the aid of the efficient and effective fitness-value evaluation model based on the ANN trained in Step 1. After the GA evolves for 20 generations, we rank all the final I populations based on their fitness values and select the top N populations.

Step 3: Use a stochastic simulation with moderate number of random test samples, say L_m random test samples, to estimate the objective values of the N settings obtained in Step 2. Rank the N settings based on their estimated objective values and select the top s settings.

Step 4: Use the stochastic simulation with sufficiently large number of random test samples, say L_s random test samples, to compute the objective values of the s settings. The setting with the smallest objective value of (1) is the good enough solution.

4. Application to G/G/1/K Polling Systems

4.1 Introduction

A polling model represents a system of multiple queues served by a single server in a cyclic order [21]. The studies of polling models had lasted for more than half century, and various applications had been found such as the computer network, communication network, manufacturing systems, transportation systems, etc.. Typical *service disciplines* when server admits customers in the attended queue are *exhaustive* (the server continues to serve all customers at a queue until it empties), *gated* (the server continuously serves only those customers that are found at a queue when it is inspected), *limited* (at most one customer is served at a queue in a cycle) [21]-[22], and *time-limited* (the server dwell certain amount of time at a queue even if it is empty) [23]. Each service discipline represents a *decision strategy* to achieve a certain performance of the polling system, for example the *average waiting time*. Numerous analysis techniques [21]-[24] have been developed for computing the average waiting times in polling models of different service disciplines. For the sake of analysis, all these techniques assumed Poisson arrival processes and infinite queue length for each queue, which may not be valid in practice.

4.2 G/G/1/K Polling Model

The polling model considered here is a G/G/1/K polling model, which accounts for general arrival processes, general distribution of service time and finite queue length for each queue. Assuming there are J queues and each queue has length K, the G/G/1/K polling model is shown in Fig. 1. This polling model is more realistic, however it will cause vast difficulties for the existing service disciplines mentioned in Section 4.1 to analyze the system's performance. Since it is hardly to get any analytical formula for evaluating the system's performance using the existing service discipline, it would be more practical to design a service discipline that can obtain better system's performance for the G/G/1/K polling system.

The proposed service discipline is a number-limited service discipline, that is when the server attends the j th queue, it will serve for $m_j(\leq K)$ customers (jobs or packets) or until the queue becomes empty, whichever comes first. Thus, $(m_1, m_2, \cdots, m_K)$ is a decision vector of our number-limited service discipline. We assume that a customer that completing the service will depart from the system, and any customer being served but yet completed will not be interrupted by any reason. The switchover time of the cyclic server switching to next queue is assumed to be of normal distribution with mean δ and variance σ^2.

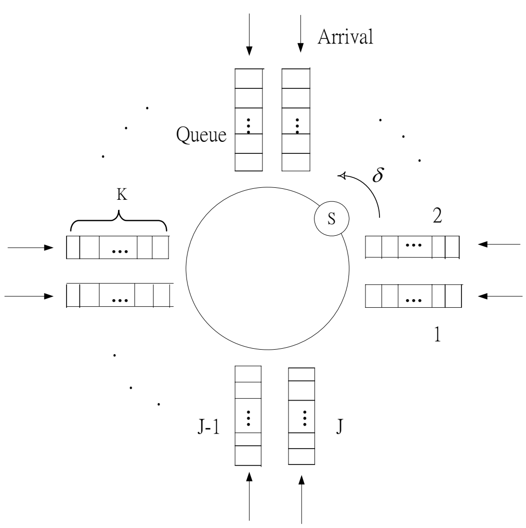

Fig. 1. G/G/1/K Polling Model.

4.3 The Stochastic Simulation Optimization Problem

We denote the random variable W_j as the waiting time of a typical customer of the j th queue at steady state. The waiting time is defined as the time length from arrival instant until the beginning of service. Then, $E[W_j]$ represents the average waiting time of a customer in the j th queue at steady state. We let τ_j denote the weighting coefficient of queue j, then $\frac{1}{J}\sum_{j=1}^{J}\tau_j E[W_j]$ denotes the weighting average waiting time of the G/G/1/K polling system.

Now, we can formulate our stochastic simulation optimization problem for the G/G/1/K polling system as

$$\min_{m_j, j=1,\ldots,J} \frac{1}{J}\sum_{j=1}^{J}\tau_j E[W_j] \quad (2)$$

subject to the G/G/1/K polling model.

In other words, we are looking for an optimal number-limited service discipline to minimize the weighting average waiting time of a G/G/1/K polling model. Suppose $K = 20$, then the size of the decision variable space of (2) will be K^J or 20^{10} provided that $J = 10$. Note that m_j is allowed to be more than K, because during the period when the server serves a customer, new customers may arrive; however, such an $m_j(> K)$ should not be a good choice unless the arrival rate is very high. Thus, problem (2) is a stochastic optimization problem with huge decision variable space as depicted in (1) and is especially suitable for the proposed ordinal optimization approach to solve for a good enough number-limited service discipline.

Remark: The four existing service disciplines stated in Section 1 can be viewed as special cases of the number-limited service discipline. For examples, $m_j >> K$ for all $j = 1, \cdots, J$ corresponds to the exhaustive service discipline; sufficiently large $m_j(\leq K)$ for all $j = 1, \cdots, J$ corresponds to the gated service discipline; $m_j = 1$ for all $j = 1, \cdots, J$ corresponds to the limited service discipline; the time-limited service discipline is an inefficient scheme from the number-limited service discipline viewpoint, because the server has to stay at the queue before time out expires even if the queue is empty.

5. Test Results and Comparisons

To test our approach, we set the parameters of the polling model as follows: $J = 10$; the arrival process of the j th queue is Poisson with arrival rate λ_j for $j = 1, \cdots, J$ as shown in Table 1; the service time is of exponential distribution with service rate $\mu = 20$; the mean and standard derivation of the switchover time of normal distribution are $\delta = 1/30$ sec and $\sigma = 0.01$, respectively;

the assumed weighting coefficients τ_j for the 10 queues are also shown in Table 1.

We set the following parameters in the OO theory based algorithm: $M = 500$ in Step 1, $I = 1000$ and $N = 1000$ in Step 2, $L_m = 1000$ and $s = 35$ in Step 3, and $L_s = 10000$ in Step 4.

The good enough decision vector of the number-limited service discipline and the corresponding weighting average waiting time obtained by our OO theory based algorithm are shown in Table 2. The CPU time consumed by our approach is only *3 minutes*, which will meet the real time application. We have also applied the exhaustive, gated, limited, and time-limited (=3 seconds) disciplines to the same polling system for the same number of customers used in the exact model of Step 4. The weighting average waiting time they obtained are shown in Table 3. We also show the percentage of the weighting average waiting time saved by our approach with respect to the existing service disciplines in the last row of Table 3. From this row, we see that our approach drastically outperforms the existing service disciplines.

Table 1 The Arrival Rates and Weighting Coefficients of the 10 Queues

j	1	2	3	4	5	6	7	8	9	10
λ_j	1	1	1	1	1	1	1	1	1	1
τ_j	1	1	1	10	1	50	1	1	1	1

Table 2 The Good Enough Decision Vector of the Number-Limited Service Discipline and the Corresponding Weighting Average Waiting Time Resulted by Our Approach, GA and SA Method

Method	Decision vector										WAWT† (secs.)	$\frac{\text{WAWT} - *^{\S}}{*} \cdot 100\%$
OO	7	12	13	1	18	2	16	9	16	17	27.6278	0%
SA	17	10	9	2	19	3	9	18	11	15	40.1982	45.5%
GA	18	13	11	4	15	4	11	14	17	19	54.6926	97.9%

† WAWT: weighting average waiting time.

§ $*$: the WAWT obtained by our approach.

Table 3 The Weighting Average Waiting Time Resulted by the Existing Service Disciplines

Discipline	Exhaustive	Gated	Limited	Time-out
WAWT†(secs.)	62.7187	80.1327	98.4026	126.3718
$\frac{\text{WAWT} - *^{\dagger}}{*} \cdot 100\%$	127.0%	190.0%	256.2%	357.4%

† WAWT, $*$: same as in Table 2.

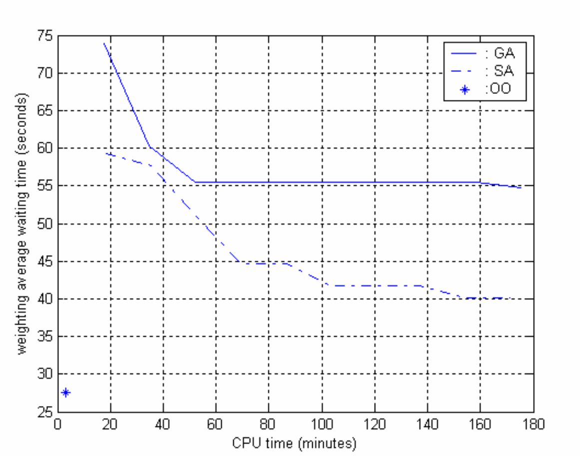

Fig. 2. Comparison of the Computational Efficiency and the Quality of the Obtained Solution of Our Approach, GA and SA Method.

We have also used the GA and SA method to solve (2). As we have indicated in Section 1 that these heuristic global searching techniques are very time consuming, we terminate the execution of these two methods when they consume 3 hours of CPU time, which is 60 times of the CPU time consumed by our approach. The resulting decision vector of the number-limited service discipline and the corresponding weighting average waiting time are also shown in Table 2. We see that even when they consume 60 times of the CPU time consumed by our approach, the best-so-far weighting average waiting time they obtained are still 45.5% (SA) and 97.9% (GA) more than ours. In the meantime, we also show the progress of the best-so-far weighting average waiting time versus the CPU times consumed by the GA and SA method in Fig. 2. From this figure, we can observe the sluggish improvement of the best-so-far weighting average waiting time of these two methods. Although the weighting average waiting time obtained by the GA and SA method are worse than that obtained by our approach, they are still better than those obtained by the existing service disciplines as can be observed from Tables 2 and 3. This shows the superiority of the number-limited service policy.

6. Conclusion

To cope with the computationally intractable stochastic simulation optimization problems, we have proposed an ordinal optimization approach to solve for a good enough solution using reasonable computation time. As for the performance of minimizing the weighting average waiting time, we have demonstrated that our approach drastically outperforms the existing service discipline. Regarding the computational efficiency and the quality of the obtained solution for solving a stochastic simulation optimization problem, we have demonstrated that our approach is much more superior than the GA and the SA method.

Acknowledgements

This research work is supported in part by the National Science Council in Taiwan, R.O.C., under grant NSC94-2213-E-009-044.

References

[1] F. Azadivar, Simulation optimization methodologies, *Proceedings of the 1999 Winter Simulation Conference,* Phoenix, AZ, 1999, 93-100.

[2] J.R. Swisher, P.D. Hyden, S.H. Jacobson, & L.W. Schruben, A survey of simulation optimization techniques and procedures, *Proceedings of the 2000 Winter Simulation Conference*, Orlando, FL, 2000, 119-128.

[3] S.M. Robinson, Analysis of sample-path optimisation, *Mathematics of Operations Research, 21*(3), 1996, 513-528.

[4] A.G. Greenwood, L.P. Rees, & F.C. Siochi, An investigation of the behavior of simulation response surfaces, *European Journal of Operational Research, 110*, 1998, 282-313.

[5] Y. Carson, & A. Maria, Simulation optimization: methods and applications, *Proceedings of the 1997 Winter Simulation Conference*, Atlanta, GA, 1997, 118-126.

[6] R.L. Haupt, & S.E. Haupt, *Practical genetic algorithms* (2nd edition, Hoboken, NJ:John Wiley, 2004).

[7] M. Locatelli, Simulated annealing algorithms for continuous global optimization: convergence conditions, *Journal of Optimization Theory and Applications, 104*(1), 2000, 121-133.

[8] D.T. Pham, & D. Karaboga, *Intelligent optimisation techniques* (London, Springer Verlag, 2000).

[9] O. Hajji, S. Brisset, & P. Brochet, Comparing stochastic optimization methods used in electrical engineering, *2002 IEEE International Conference on Systems, Man and Cybernetics, 7*, Hammamet, Tunisia, 2002.

[10] F. Glover, J.P. Kelly, & M. Laguna, New advances and applications of combining simulation and optimisation, *Proceedings of the 1996 Winter Simulation Conference*, Coronado, CA, 1996, 144-152.

[11] H. Pierreval, & J.-L. Paris, Distributed evolutionary algorithms for simulation optimisation, *IEEE Transactions on Systems, Man and Cybernetics, Part A*, *30*(1), 2000, 15–24.

[12] T.W.E. Lau, & Y.C. Ho, Universal alignment probability and subset selection for ordinal optimisation, *Journal of Optimization Theory and Applications, 39*(3), 1997, 455-489.

[13] Y.C. Ho, An explanation of ordinal optimization: Soft computing for hard problems, *Information Sciences, 113*(3-4), 1999, 169-192.

[14] C.-H. Chen, S.D. Wu, & L. Dai, Ordinal comparison of heuristic algorithms using stochastic optimisation, *IEEE Transactions on Robotics and Automation, 15*(1), 1999, 44–56.

[15] S.-Y. Lin, & Y.C. Ho, Universal alignment probability revisited, *Journal of Optimization Theory and Applications, 113*(2), 2002, 399-407.

[16] S.-Y. Lin, Y.C. Ho, & C.-H. Lin, An ordinal optimization theory based algorithm for solving the optimal power flow problem with discrete control variables, *IEEE Transactions on Power System, 19*(1), 2004, 276-286.

[17] K. Hornik, M. Stinchcombe, & H. White, Multilayer Feedforward Networks are Universal Approximators, *Neural Networks, 2*(5), 1989, 359-366.

[18] C.G. Panayiotou, C.G. Cassandras, & W.B. Gong, Model abstraction for discrete event systems using neural networks and sensitivity information**,** *Proceedings of the 2000 Winter Simulation Conference, 1*, Orlando, FL, 2000, 335-341.

[19] R.A. Kilmer, A.E. Smith, & L.J. Shuman, Computing confidence intervals for stochastic simulation using neural network metamodels, *Computers and Industrial Engineering, 36*(2), 1999, 391-407.

[20] M.F. Moller, A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks, 6*(4), 1993, 525-533.

[21] H. Takagi, Analysis and application of polling model. In: Performance Evaluation: Origins and Directions, eds. G. Haring, C. Lindemann and M. Reiser, *Lecture Notes in Computer Science* (Berlin, Springer, 2000), *1769*, 423-442.

[22] H. Takagi, Analysis of polling systems with a mixture of exhaustive and gated service disciplines, *Journal of the operations research society of Japan, 32*(3), 1989, 450-461.

[23] L. Fujian, R.-H. Sun, C.-Z. Wu, & Z.-J. Liu, Cyclic Service Systems with Server Time-outs, *2nd IEEE Symposium on Computers and Communications*, Alexandria, Egypt, 1997, 43-47.

[24] T. Hirayama, S.J. Hong, & M. Krunz, A New Approach to Analysis of Polling Systems, *Queueing Systems, 48*(1-2), 2004, 135-158.