Tạp chí Tin học và Điều khiển học, T. 18, S. 4 (2002), 333–338

MINIMIZING THE AVERAGE DELAY TIME IN A QUEUEING NETWORK BY USING GENETIC ALGORITHMS

LUONG HONG KHANH, VU NGOC PHAN

Abstract. The present paper is dealing with the issue of minimizing the maximum average delay time of a queueing network by using a modified genetic algorithm.

Tóm tắt. Bài báo này đề cập đến vấn đề cực tiểu hóa thời gian trễ cực đại trong một mạng hàng đợi. Bài toán cực tiểu hóa được thực hiện nhờ thuật toán di truyền được cải tiến cho thích hợp với điều kiện ràng buộc đặc biệt của vấn đề đặt ra.

1. INTRODUCTION

The queueing theory initiated by Erlang has gained a wide applicability to communication system design and analysis ([1,3,4,6,7,8,13]). The major performance measures in a communication network are the delay time of calls, packets, or cells. In a previous paper, the authors investigated the mean value method to calculate the average arrival rates and the average delay time at queues of a communication queueing network ([10]). This study is based on the fact that there is a need for reducing the maximum delay time occurring in the network. Obviously, if one needs to reduce the average delay time at some queues, the service capacities at those queues have to be enlarged. However, during a detailed study of this issue we have found a very interesting phenomenon. While the service capacity is enlarged by increasing the average service rate, the maximum time delay in the network begins to decrease to a minimum value and then increases again, although the service capacity of the entire system increases continuously. Thus, a question may be asked here, how the average service rates have to be arranged at the queues such that the maximum delay time of the system will be minimum. In the present paper, the maximum delay time in a queueing network is minimized by using a genetic algorithm.

2. PRELIMINARY

The closed queueing network is a network of queues where the total number of customers inside the network is fixed. That means, there is no customer departing from the network and no customer is allowed to enter into the network. Let K denote the fixed number of customers in the network, x_i be the number of customers at queue i and N be the total number of queues. It is easily to recognise that

$$\sum_{i=1}^{N} x_i = K \tag{1}$$

Let $S(K,N)$ denote the state of the system, where

$$S(K,N) = (x_1, x_2, ..., x_N \Big| \sum_{i=1}^{N} x_i = K) \tag{2}$$

It can be imagined that the number of the network states is very large. For simplicity, the following assumptions are made.

- In the closed queueing network the first-come first-serve discipline is used.
- The service time at queue i follows the exponential distribution with rate λ_i

- Service rate of each queue is independent of its queue length.

Let w_{ij} denote the probability that after a customer is served at queue i it continuously joins the queue j like in the open queueing network. Here the condition

$$\sum_{j=1, j\neq i}^{N} w_{ij} = 1 \quad \text{for} \quad i \in \mathcal{N} \tag{3}$$

is hold. An essential difference between the open queueing network and the closed queueing network is that the total arrival rate at queue i can not be determined by Eq.(2). Consider the state of the network when there is no customer departing the network and because of that, there is no customer can enter into the network. For this situation, the total arrival at queue i is determined by

$$\Lambda_i = \sum_{j=1, i\neq j}^{N} \Lambda_j w_{ji} \quad \text{for} \quad i \in \mathcal{N} \tag{4}$$

or in the matrix form

$$\Lambda = W^T \Lambda \tag{5}$$

The equation (5) is singular and can not be solved. The transition from state $x = (x_1, x_2, ..., x_N)$ to state $x = (x_1, ..., x_i + 1, ..., x_j - 1, ..., x_N)$ corresponds to the situation in which a customer finishes his service at queue j and joins queue i. The following relation is hold.

$$\mu_i w_{ij} P_r(x_1, ..., x_i + 1, ..., x_j - 1, ..., x_N) = \mu_j w_{ji} P_r(x_1, ..., x_i, ..., x_j, ..., x_N) \tag{6}$$

or

$$P_r(x_1, ..., x_i + 1, ..., x_j - 1, ..., x_N) = \frac{q_i}{q_j} P_r(x_1, ..., x_i, ..., x_j, ..., x_N) \tag{7}$$

Here, it is supposed that $w_{ij} = w_{ji}$. The solution of Eq.(6) gets the form

$$P_r(x_1, ..., x_i, ..., x_N) = \frac{1}{G(K, N)} \prod_{i=1}^{N} q_i^{x_i} \tag{8}$$

where $G(K, N)$ is a normalization constant to guarantee that $P_r(x_1, ..., x_i, ..., x_N)$ is a proper probability distribution, that means:

$$\sum_{x\in S(K,N)} P_r(x) = \sum_{x\in S(K,N)} \frac{1}{G(K, N)} \prod_{i=1}^{N} q_i^{x_i} = 1 \tag{9}$$

From (9) it is easily to identify that

$$G(K, N) = \sum_{x\in S(K,N)} \times \prod_{i=1}^{N} q_i^{x_i} \tag{10}$$

By substituting $g_n(k) = G(k, n)$ the following recursive equation can be derived from Eq.(10)

$$g_n(k) = g_{n-1}(k) + q_n g_n(k - 1); \quad k = 0, 1, ..., K; \quad n = 1, 2, ..., N \tag{11}$$

$G(K, N)$ is equal to $g_n(k)$ when $n = N$ and $k = K$.

Given a closed queueing network with K packets and N nodes each of which represents a queue. The typical interesting values are the expected number of packets in the network and the expected delay time. It can be show that these expected values can be determined without knowing the normalization constant $G(K,N)$. The expected number of packets in a queue can be expressed by the utilization of queue as follows

$$E\{x_i(K)\} = \sum_{v=1}^{K} q_i^v \frac{G(K-v,N)}{G(K,N)} \tag{12}$$

From Eq.(12) it is easily to get the recursive equation

$$E\{x_i(K)\} = U_i(K)[1 + E\{x_i(K-1)\}]; \quad i = 1,2,...,N \tag{13}$$

where $U_i(K)$ is determined by

$$U_i(K) = q_i \frac{G(K-1,N)}{G(K,N)} \tag{14}$$

The first and second terms in Eq.(13) represent the average service time and waiting time of packets, respectively. Substitute

$$S_i(K) = \Lambda_i \frac{G(K-1,N)}{G(K,N)} \tag{15}$$

Let

$$E\{d_i(K)\} = \frac{E\{x_i(K)\}}{S_i(K)} \tag{16}$$

be the expected delay time at queue i. After substituting Eq.(14) and (15) in Eq.(16) it leads to

$$E\{d_i(K)\} = \frac{1}{\mu_i}(1 + E\{x_i(K-1)\}) = \frac{1}{\mu_i} + \frac{1}{\mu_i}E\{x_i(K-1)\} \tag{17}$$

Eq.(17) indicates that the expected delay time at a queue equals the sum of average service time and average waiting time.

3. PROBLEM STATEMENT

Let us consider a closed queueing network with the transition probability matrix W. Supposing, there are K customers at N queues of the network. The total service capacity of the network is constant, i.e.

$$\sum_{i=1}^{N} \mu_i = \text{const} \tag{18}$$

The problem is how the total service capacity is arranged for minimizing the maximum average delay time. This problem can mathematically expressed as following.

$$\underset{\mu_i}{\text{Min}}\underset{i}{\text{Max}} E\{d_i(K)\} \tag{19a}$$

subject to

$$\sum_{i=1}^{N} \mu_i = C = \text{const}, \tag{19b}$$

$$0 < \alpha_i \le \mu_i \le \beta_i; \quad I = 1,2,...,N \tag{19c}$$

It is easy to recognize that the problems (19a), (19b), (19c) belongs to the nonlinear programming issue, where the objective function can not be explicitly described. Therefore, the utilization of the traditional programming methods can not ensure a sensible result. In this case, the minimization problem will be solved by a modified genetic algorithm that will be described in the next section.

4. THE MODIFIED GENETIC ALGORITHM

Genetic algorithms are generalized random search methods simulating the evolution process of living bodies according to the natural selection principle ([2,5,9,11,12]). A genetic algorithm normally consists of three parts: the parameter encoding, the simulation of the evolution process and the parameter decoding for getting the final result. The evolution process can be simulated by three stages, namely the reproduction, the crossover and the mutation with different probabilities ([12]). The genetic algorithm used in the present paper is almost the same as that in ([11]) except some changes to adapt to the new circumstance. For encoding the search parameters we set

$$X = [\mu_1, \mu_2, ..., \mu_N] \tag{20}$$

as the individual. Then, X is encoded by a binary string of appropriate length. The number of bits required for encoding μ_i is determined as following.

$$l_i = \text{Int}\Big[\log_2 \frac{\beta_i - \alpha_i}{\varepsilon_i}\Big] \tag{21}$$

Here, Int$[y]$ is the smallest integer greater or equal y, ε_i is the change level of μ_i. The length of X is then determined by

$$L = \sum_{i=1}^{N} l_i \tag{22}$$

Because the delay time is always positive and we are occupying with the minimization problem, it is sensible to chose the reciprocal of the delay time as the fitness function for the evolution process i.e.

$$F = \frac{1}{\underset{\mu_i}{\text{Min}}\underset{i}{\text{Max}} E\{d_i(K)\}} \tag{23}$$

Let M denote the population size and F_i denote the fitness of the i-th individual, the reproduction process is simulated as follows.

- Set $V_1 = F_1$
  $V_2 = F_1 + F_2 = V_1 + F_2$
  ...
  $V_i = V_{i-1} + F_i; \quad (i = 3, 4, ..., M)$
- Create a random number between 0 and V_M, supposing V_h
- Chose the first individual that has the fitness value greater or equal V_h and introduce it to the new generation.
- Repeat the procedure until getting the new population with the desired population size.

The initial population is created in the way that the bits of X are randomly set by 0 or 1. The crossover process is simulated as following.

- Chose randomly two individuals of the current population, assuming X_1 and X_2.
- Create two random integers between 1 and L, assuming z_1 and z_2.
- Exchange the $z_1 - th$ bit of X_1 with the $z_1 - th$ bit of X_2 and the $z_2 - th$ bit of X_1 with the $z_2 - th$ bit of X_2.
- Enroll the two new individuals as the members of the population for the next evolution step.

The crossover probability can be chosen equal 0.18 due to [12]. The mutation process is simulated only at one allele as following.

- Chose randomly an individual of the current population, assuming X_1.

- Create a random integer between 1 and L, assuming z.
- Change the z-th bit value of X_1.
- Enroll the new individual as the member of the population for the next evolution step.

The mutation probability can be chosen due to [12] the value of 0.005.

The subject condition expressed in Eq. (19c) is automatically satisfied by the encoding method described above. For satisfying the subject condition expressed in Eq.(19b), the following procedure is utilized to every new individual.

- Create a random integer between 1 and N, assuming z.
- The parameter μ_z is determined depending on the other $N-1$ remainders

$$\mu_z = C - \sum_{i=1,\mu\neq z}^{N} \mu_i$$

5. SIMULATION RESULT

Consider a closed queueing network of $K = 6$ and $N = 8$. The probability matrix W is given by

$$W = \begin{bmatrix} 0.00 & 0.11 & 0.26 & 0.26 & 0.00 & 0.12 & 0.00 & 0.25 \\ 0.53 & 0.00 & 0.01 & 0.00 & 0.32 & 0.00 & 0.00 & 0.14 \\ 0.16 & 0.00 & 0.00 & 0.47 & 0.00 & 0.21 & 0.00 & 0.16 \\ 0.11 & 0.26 & 0.00 & 0.00 & 0.12 & 0.25 & 0.00 & 0.26 \\ 0.04 & 0.00 & 0.00 & 0.26 & 0.00 & 0.59 & 0.01 & 0.10 \\ 0.28 & 0.14 & 0.12 & 0.25 & 0.00 & 0.00 & 0.15 & 0.06 \\ 0.31 & 0.00 & 0.26 & 0.00 & 0.00 & 0.00 & 0.00 & 0.43 \\ 0.20 & 0.00 & 0.24 & 0.01 & 0.20 & 0.24 & 0.11 & 0.00 \end{bmatrix}$$

The total service capacity is $\sum_{i=1}^{8} \mu_i = C = 4$. The bounds for all μ_i are the same, i.e. $\alpha = 0.125$ and $\beta = 2.5$. The maximum average delay time of the network is 4.5263 with $\mu_1 = 0.6350$; $\mu_2 = 0.4250$; $\mu_3 = 0.4850$; $\mu_4 = 0.4850$; $\mu_5 = 0.6350$; $\mu_6 = 0.4550$; $\mu_7 = 0.3650$; $\mu_8 = 0.4850$.

6. CONCLUSION

The average delay time of calls, cells or packets in a queueing network plays a significant role for the network performance. In the present paper we have shown how the maximum average delay time of a queueing network is minimized due to the distribution of the total service capacity in the nodes by using a genetic algorithm. The approach has been demostrated for a close queueing network. However, it can be extended for open queueing networks.

REFERENCES

[1] Ash, G. R. *Dynamic Routing in Telecommunication Network* McGraw Hill 1997.

[2] Chin-Teng Lin George Lee *Neural Fuzzy Systems* Prentice-Hall International, Inc. 1996.

[3] Guizani, M Rayes, *A Designing ATM Switching Networks* McGraw Hill 1999.

[4] Hayes, J. F. *Modelling and Analysis of Computer Communication Networks,* Plenum Press, New York 1984.

[5] Holland, J.H. “Genetic Algorithms and Classifier Systems. Foundation and Future Directions” Gen. Alg. and Their Appl. Proc. 2nd Int. Conf., Cambridge (1987).

[6] Kleinrock, L. *Queueing Theory,* Wiley 1975.

[7] Minoli, D. "Broadband Network Analysis and Design. Bell Communications Research", Inc. New York University. Artech House, Boston, London (1993).

[8] Ross, K. W. *Multiservice Loss for Broadband Telecommunication Networks* Springer Verlag 1995.

[9] Simon, R. *Robust Encodings in Genetic Algorithms. Evolutionary Algorithms in Engineering Applications,* (Eds. Dasgupta and Michalewicz.) Springer Verlag Berlin 1997 (29–44).

[10] Luong Hong Khanh, Vu Ngoc Phan, Queueing Network theory AND its application to Communication Systems, *Journal of Computer Science and Cybernetics*, **18** (2) (2002) (1 – 6).

[11] Vu Ngoc Phan, Application of evolutionary algorithms to the multi-objective nonlinear optimization (in Vietnamese), *Journal of Computer Science and Cybernetics* **16** (3) (2000) (16–22).

[12] Vu Ngoc Phan, Evolutionary algorithms and application to automatic control (in Vietnamese), *Journal of Computer Science and Cybernetics* **44** (10) (1999) (1–14).

[13] Zbigniew, D. *ATM resource management,* McGraw Hill 1997.

Received on February 21 - 2002

Institute of Information Technology