TECHNICAL NOTE

OPTIMIZATION OF M/G/1 QUEUE WITH VACATION

R. P. Ghimire

Department of Mathematics, S. N. Multiple Campus
Mahendra Nagar, Nepal

M. Jain

Department of Mathematics, Institute of Basic Science
Khandari, Agra, 282002, India, madhuj@yahoo.com

(Received: June 26, 2001 – Accepted in Final From: May 5, 2002)

ABSTRACT This paper reports on the minimization of the average waiting time of the customers in the M/G/1 queue with vacation. Explicit formula for the unknown service parameter of a particular customer has been obtained by considering the exhaustive service discipline in the case of multi-user with unlimited service system. Moreover, results in case of partially gated and gated service disciplines under limited/unlimited service systems have been provided. Some particular cases such as M/M/1 and M/D/1 models with and without vacation have also been discussed.

Key Words M/G/1 Queue, Optimization, Average Waiting Time, Kuhn-Tuker Condition, Vacation, Limited/Unlimited Partially Gated

چکیده این مقاله به حداقل کردن متوسط زمان انتظار مشتریان در صف تعطیلات M/G/l می پردازد. فرمول روشنی برای پارامتر ناشناخته خدمت به یک مشتری خاص از طریق در نظر گرفتن روش خدمتگزاری برای چند کاربر با سیستم خدمت نا محدود بدست آمده است. همچنین نتایج مربوط به روشهای دروازه ای جزئی و خدمت دروازه ای ارائه شده است. حالتهای خاصی همچون مدل M/M/l و M/D/l با و بدون تعطیل بحث شده است.

1. INTRODUCTION

Vacation models having multi queues attended by a single sever have been extensively studied for the last two decades due to its several applications in the performance evaluation of computers, communication and production systems. A rich literature exists on vacation in different disciplines (e.g. exhaustive, gated, limited etc.).

In the exhaustive vacation model, the server serves customers in a queue until the queue is empty while in gated vacation model only those customers in a queue who arrived before the server's visit are served continuously. Under limited service discipline, a maximum number of customers (say L) are served in a visit of the server. Doshi [1,2] surveyed excellently vacation systems under different service disciplines. Takagi [3-6] discussed several applications of polling models and reposted an extensive set of results for the M/G/1 queue with vacation.

DeMorases [7] studied message delay analysis of polling schemes with limited transmission times. Takine et al. [8,9] proposed the exact analysis of asymmetric polling system with a single buffer and round robin scheduling of services. DeMorasas and Furmann [10] established approximations for the mean delay in polling systems with batch Poisson input. Takine and Hasegawa [11] considered time dependent behavior of the M/G/1 queue with multiple vacations and gated service discipline. Levy and Kleinrock [12] analyzed the polling system with zero switchover periods (ZSOP) and used a general method for analyzing the expected delay.

Shimogawa and Takahashi [13] obtained the inter-departure time-distributions of an M/G/1 queuing model with server vacations. Leung and Lucantoni [14] discussed the vacation models for token ring networks where timers control the service. The optimal choice of service parameter for different types of customers in a queuing system is of great importance. Rousseau and Laporte [15] gave a practical optimization technique for the M/M/1 queuing system.

Belu and Shama [16] investigated the problems of scheduling the total time of a service facility between different types of customers on priority basis for the M/D/1 model. Jain et al. [17] employed the optimization technique for a more general queuing system with pareto service time distribution and obtained a formula for the service time parameter of arbitrary customers explicitly by taking particular cases for the M/M/1 and M/D/1 model.

Sharma et al. [18] presented a non-linear programming approach for the single server bulk arrival queuing model and obtained the unknown service parameters for different types of customers in order to minimize the total time of the service facility. Leung [19] developed the scheduling policy for serving an additional job in priority queue under the vacation queuing model.

It is worth noting that despite the fact that there exists a considerable literature for vacation models, the effort for the scheduling of available time of service facility in case of such models according to different types of demands for service is rare. The present investigation is an attempt in this direction.

The M/G/1 model with vacation under the exhaustive partially gated and gated disciplines for unlimited/limited service systems are considered. The total time T of a single server is distributed among m users in such way that the average queuing time of an individual customer is minimized. Using a non-linear programming approach has derived the results for unknown service parameters.

2. MATHEMATICAL MODEL AND ANALYSIS

Consider the M/G/1 queuing model with vacation by assuming that m types of customers arrive in Poisson fashion with the mean arrival rate of λ_i (i=1,2,…,m) to the service facility having a single server. The server can provide service according to the need of the customers.

Let for the customers of class i (i = 1, 2, ..., m), μ_i and c_s^2 be the mean service rate and square coefficient of variation of service time respectively. We denote $\overline{v}_i$ and $\overline{v}_i^2$ respectively, the first two moments of the vacation (or reservation) intervals of customers of class i. The average queuing time for ith class of customers for exhaustive unlimited service discipline is given by (see Bertsekas and Gallanger [20]).

$$W_i = \frac{\lambda_i (c_0^2 + 1)}{2\mu_i^2 (1-\rho_i)} + \frac{(m-\rho_i)\overline{v}}{2(1-\rho_i)} + \frac{\sigma_v^2}{2\overline{v}} \quad (1)$$

Where $\overline{v}$ and σ_v^2 are the mean and variance to the vacation intervals, respectively average over all customers and are given by

$$\overline{v} = \frac{1}{m}\sum_{i=1}^{m}\overline{v}_i \quad (2a)$$

and

$$\sigma_v^2 = \frac{1}{m}\sum_{i=1}^{m}(v_i^2 - v_i^2) \quad (2b)$$

Let $\mu_i - \lambda_i = X_i$ so that substituting the value $\mu_i = \lambda_i + X_i$ Equation 1, we get

$$W_i = \frac{\lambda_i}{2}\left\{\frac{(c_0^2+1)(x_i+\lambda_i)(m-1)\overline{v}}{x_i}\right\} + \frac{m\overline{v}}{2} + \frac{\sigma_v^2}{2\overline{v}} \quad (3)$$

Our main aim is to minimize the average queuing time between different types of customers so that we have the following optimization problem:

$$\text{Min} \sum_{i=1}^{m} e_i \lambda_i w_i \quad (4)$$

such that

$$E = \sum_{i=1}^{m} \sigma_i x_i \quad (5)$$

E being the excess time, e_i and σ_i are the weights reflecting the relative priority and service time of the customers of ith class respectively. Now putting the value of w_i from Equation 3 in Equation 4 and applying Kuhn-Tucker condition [21], we have

$$\frac{-\lambda_i e_i}{2x_i^2}\left[\frac{(c_0^2+1)(1+\frac{2x_i}{\lambda_i})}{(1+\frac{x_i}{\lambda_i})^2}+\lambda_i(m-1)\bar{v}\right]+\sigma_i=0 \qquad (6)$$

where z is an unknown real number.

We note that denominator of Equation 6 can be approximated as

$$1+(x_i/\lambda_i)^2 \approx 1+2x_i/\lambda_i \qquad (7)$$

Thus (6) reduces to

$$\sigma_i z \approx \frac{\lambda_i e_i}{2x_i^2}\left[(c_0^2+1)+\lambda_i(m-1)\bar{v}\right] \qquad (8)$$

On simplification Equation 8 gives

$$x_i=\left\{\frac{\lambda_i e_i(c_0^2+1)+\lambda_i(m-1)\bar{v}}{2\sigma_i z}\right\}^{\frac{1}{2}} \qquad (9)$$

Using Equations 5 and 9, we get

$$E=\sum_{i=1}^{m}\left\{\frac{\sigma_i\lambda_i e_i(c_0^2+1)+\lambda_i(m-1)\bar{v}}{2z}\right\}^{\frac{1}{2}} \qquad (10)$$

so that

$$z'=\sqrt{2z}=\frac{1}{E}\sum_{i=1}^{m}\left[\sigma_i\lambda_i e_i\left\{(c_0^2+1)+\lambda_i(m-1)\bar{v}\right\}\right]^{\frac{1}{2}} \qquad (11)$$

Now from Equations 9 and 11, we obtain

$$x_i=\frac{\left[\lambda_i e_i\left\{(c_0^2+1)+\lambda_i(m-1)\bar{v}\right\}\right]^{\frac{1}{2}}}{z'\sigma_i} \qquad (12)$$

which determines the unknown service parameter μ_i, (i = 1, 2,...,m) for different classes of customers.

The expected queuing time, for other types of vacation models, are given by Bertsekas and Gallanger [20]. Defining $\beta_i=1/(1-\lambda_i\bar{v})$ and

$$r_i=\left\{\frac{\lambda_i e_i}{\sigma_i}\right\}^{\frac{1}{2}}.$$

$$W_i=\begin{cases}\frac{(1+c_0^2)\rho_i}{2\mu_i(1-\rho_i)}+\frac{(m+\rho_i)\bar{v}}{2(1-\rho_i)}+\frac{c_0^2}{2\bar{v}} & \text{unlimited partiially gated (UPG)}\\ \frac{(1-c_0^2)\rho_i}{2\mu_i(1-\rho_i)}+\frac{(m+2-\rho_i)\bar{v}}{2(1-\rho_i)}+\frac{\sigma_0^2}{2\bar{v}} & \text{unlimited gated (UG)}\\ \frac{(1+c_0^2)\rho_i}{2\mu_i(1-\rho_i-\lambda_i)}+\frac{(m+\rho_i)\bar{v}}{2(1-\rho_i-\lambda_i)}+\frac{\sigma_0^2(1-\rho_i)}{2\bar{v}(1-\rho_i-\lambda_i)} & \text{unlimited partiially gated (LPG)}\\ \frac{(1+c_0^2)\rho_i}{2\mu_i(1-\rho_i-\lambda_i\bar{v})}+\frac{(m-2-\rho_i-2\lambda\bar{v})\bar{v}}{2(1-\rho_i-\lambda_i\bar{v})}+\frac{\sigma_0^2(1-\rho_i)}{2\bar{v}(1-\rho_i-\lambda_i\bar{v})} & \text{unlimited gated (LG)}\end{cases} \qquad (13)$$

The unknown parameter x_i (i = l, 2, ..., m) for these models can be obtained in a manner similar to the one discussed earlier for unlimited exhaustive cases. Thus

$$x_i=\begin{cases}\frac{r_i}{x'}\left[(c_0^2+1)+\lambda_i(m+1)\bar{v}\right]^{\frac{1}{2}} & \text{(LPG)}\\ \frac{r_i}{z'}\left\{(c_0^2+1)+\lambda_i(m+1)\bar{v}\right\}^{\frac{1}{2}} & \text{(UG)}\\ \frac{r_i}{z'}\left[\beta_i(c_0^2+1)+\lambda_i\frac{\left\{(m+\beta_i)\bar{v}^2+\sigma_v^2(1-\beta_i)\right\}}{\beta_i\bar{v}}\right]^{\frac{1}{2}} & \text{(LPG)}\\ \frac{r_i}{z'}\left[\beta_i(c_0^2+1)+\lambda_i\frac{\left[\left\{m+2-(2\bar{v}+\beta_i)\lambda_i\right\}\bar{v}^2+\sigma_v^2(1-\beta_i)\right]}{\beta_i\bar{v}}\right]^{\frac{1}{2}} & \text{(LG)}\end{cases} \qquad (14)$$

3. PARTICULAR CASES

Case 1 - M/M/1 Model with Vacation In this case $c_0^2 = 1$ so that the unknown parameter x_i for several service disciplines reduces to

$$x_i=\begin{cases}\frac{r_i}{x'}\left[2+\lambda_i(m-1)\bar{v}\right]^{\frac{1}{2}} & \text{(UE)}\\ \frac{r_i}{z'}\left\{2+\lambda_i(m+1)\bar{v}\right\}^{\frac{1}{2}} & \text{(UPG)}\\ \frac{r_i}{z'}\left\{2+\lambda_i(m+1)\bar{v}\right\}^{\frac{1}{2}} & \text{(UG)}\\ \frac{r_i}{z'}\left[2\beta_i+\frac{\left\{\lambda_i(m+\beta_i)\bar{v}^2+\sigma_v^2(1-\beta_i)\right\}}{\beta_i\bar{v}}\right]^{\frac{1}{2}} & \text{(LPG)}\\ \frac{r_i}{z'}\left[2\beta_i+\frac{[\lambda_i(m+2-(2\bar{v}+\beta_i)\lambda_i]\bar{v}^2+\sigma_v^2(1-\beta_i)}{\beta_i\bar{v}}\right]^{\frac{1}{2}} & \text{(LG)}\end{cases} \qquad (15)$$

Case 2 - M/D/1 Model with Vacation For this model $c_s^2 = 0$, so that Equation 14 reduces to

$$x_i = \begin{cases} \frac{r_i}{x'}\left[1+\lambda_i(m-1)\bar{v}\right]^{\frac{1}{2}} & \text{(UE)} \\ \frac{r_i}{z'}\left\{1+\lambda_i(m+1)\bar{v}\right\}^{\frac{1}{2}} & \text{(UPG)} \\ \frac{r_i}{z'}\left\{1+\lambda_i(m+1)\bar{v}\right\}^{\frac{1}{2}} & \text{(UG)} \\ \frac{r_i}{z'}\left[\beta_i + \frac{\left\{\lambda_i(m+\beta_i)\bar{v}^2 + \sigma_v^2(1-\beta_i)\right\}}{\beta_i \bar{v}}\right]^{\frac{1}{2}} & \text{(LPG)} \\ \frac{r_i}{z'}\left[\beta_i + \lambda_i \frac{[(m+2-(2\bar{v}+\beta_i)\lambda_i]\bar{v}^2 + \sigma_v^2(1-\beta_i)}{\beta_i \bar{v}}\right]^{\frac{1}{2}} & \text{(LG)} \end{cases} \qquad (16)$$

Case 3 - M/M/1 and M/D/1 Model without Server Vacations Substituting $\bar{v} = 0$ and $c_v^2 = 0$ in Case 1 and Case 2. We get

$$x_i = \begin{cases} \frac{1}{z'}\left[\frac{2\lambda_i e_i}{\sigma_i}\right]^{\frac{1}{2}} & \text{M/M/1 model} \\ \frac{1}{z'}\left[\frac{\lambda_i e_i}{\sigma_i}\right]^{\frac{1}{2}} & \text{M/D/1 model} \end{cases} \qquad (17)$$

which agrees with results obtained by Rousseau and Laporte [15] and Belu and Shrama [16] for M/M/1 models respectively.

4. DISCUSSION

In this investigation, we have suggested a non-linear programming technique in order to minimize total queuing time of a service facility with vacation. The explicit formulae obtained for the optimal service parameter for multi-class users of the service systems may be helpful in scheduling the total time of an organization. Since vacation models have widespread applicability in local area network (LAN), computer communication systems, manufacturing system etc, the results obtained would have useful applications in real life situations.

5. REFERENCES

1. Doshi, B. T., “Queuing Systems with Vacations: A Survey Queuing System”, Vol. 29, (1986).
2. Doshi, B. T., “Single Server Queues with Vacations”, ***J. Stochastic Analysis of Computer and Communication Systems***, Elsevier Science Pub., B.V., (1990), 217-265.
3. Takagi, M., “Queuing Analysis of Vacation Models”, Part-IV, M/G/1/N, TR 87-0043. IBM, Tokyo Research Lab., (1987).
4. Takagi, H., “Queuing Analysis, Vacations Models Part I. M/G/1 and Part II. M/G/1 with Vacations”, TR 87-0032 IBM Tokyo Research Lab., (1987).
5. Takagi, H., “Queuing Analysis of Vacations Models, Part III, M/G/1 with Priorities”, TR87-0038, IBM Tokyo Research Lab., (1987).
6. Takagi, H., “Queuing Analysis of Polling Models”, ***ACM, Computing Surveys***, Vol. 20, (1988), 5.
7. De Moraes, L. F. M., “Message Delay Analysis of Polling Schemes with Limited Transmission Times”, IBM Zurich Res. Lab. Research Report, (1989), R-1843.
8. Takine, T., Takahashi, Y., and Hasegawa, T., “Exact Analysis of Asymmetric Polling Systems with Single Buffers”, ***IEEE Trans. Commun.***, Vol. 36, No. 10, (1988), 119-127.
9. Takine, T., Takahashi, Y. and Hasegawa, T., “Average Message Delay of an Asymmetric Single Buffer Polling System with Round Robin Scheduling of Services”, ***Proc. 4th Int. Conf. Modeling Tech. and Tools for Comp. Perf. Eval,*** Palmade Mallorca, Spain, (1988), 233-243.
10. De Moraes, L. F. M. and Furman, S., “Mean Delay Approximation for Polling Systems with Batch Poisson Input”, ***Perf. Eval.***, Vol. 12, (1991), 147-156.
11. Takine, T. and Hasegawa, T., “On the M/G/1 Queue with Multiple Vacations and Gated Service Discipline”, ***Oper. Res. Soc. Japan***, Vol. 35, No. 3, (1992), 217-235.
12. Levy, H. and Klienrock, L., “Polling Systems with Zero Switch Over Periods: General Method for Analyzing the Expected Delay”, ***Perf. Eval.***, Vol. l3, (1991), 97-l 07.
13. Shimogawa, S. and Takahashi, Y., “On a Set of Interdeparture Time Distributions of the M/G/1 Queue with Server Vacation”, ***J. Oper. Res. Soc. of Japan***, Vol. 36, No. 4, (1993), 206-219.
14. Leung, K. K. and Lucantoni, D. M, “Two Vacation Models for Token-Ring Networks where Service is Controlled by Times”, ***Perf. Eval.***, Vol. 20, No. (1-3), (1994), 165-184.
15. Rousseau, J. M. and Laporte, G., “A Practical Optimization Technique for a Queuing System”, ***Opl. Res. Quart.***, Vol. 28, No. l, (1977), 33-35.
16. Belu and Sharma, G. C., “An Optimization Technique for M/D/1 Model Using Approximations”, ***M.A.C.T.***, Vol. 14, (1981), l5-18.
17. Jain, M., Belu and Sharma, G. C., “A Practical Optimization Technique for a Queuing System with Pareto Service Time Distribution, ***M.A.C.T.***, Vol. 18, (1985), 23-27.
18. Sharma, C., Jain, M. and Sharma, G. C., “A Non-Linear Programming Approach for M/G/1 Queuing Model”, ***Math. Ed.***, Vol. 23, No. 3, (1989), 99-104.
19. Leung, K. K., “An Execution/Sleep Scheduling Policy for Serving an Additional Job in Priority Queuing Systems”, ***J. ACM***, Vol. 40, No. 12, (1993), 394-417.
20. Bertsekas, D. and Gallenger, R., “Data Network”, Prentice Hall of India.
21. Kuhn and Tucker, “Non-Linear Programming”, ***Proc. Second Berkeley Symp. On Mathematical Statistic and Probability***, (Edited by J. Heyman), Univ. of California, Berkeley, (1951).