Journal of the Korean Institute of Industrial Engineers
Vol. 33, No. 4, pp. 419-423, December 2007.

(p, T)-정책을 갖는 M/G/1 대기행렬 시스템의 분석

이두호† · 채경철

한국과학기술원 산업공학과

Analysis of the M/G/1 Queueing System with Randomized Control of T-Policy

Doo Ho Lee · Kyung C. Chae

Department of Industrial Engineering, KAIST, Daejeon 305-701, Korea

In this paper, we consider the $M/G/1$ queueing system with randomized control of T-policy. Whenever the busy period ends, the server is turned off and takes multiple vacations whose interval is fixed time T with probability p or stays on and waits for arriving customers with probability $1-p$. We introduce the cost function and determine the optimal combination of $(p,\ T)$ to minimize the average cost per unit time.

Keywords: $M/G/1$ queue, $(p,\ T)$-policy, cost function

1. 서 론

휴가형 대기행렬 시스템 중에서 대표적인 두 가지는 복수휴가 시스템과 단수휴가 시스템이다. 복수휴가 시스템은 다음과 같이 진행된다. 서비스할 고객이 없으면 서버(server)는 시스템을 떠나서 V_1시간 후에 돌아온다. 돌아왔을 때 서비스 할 고객이 없으면 길이 V_2의 휴가를 떠난다. 이와 같이 반복하여 휴가에서 돌아왔을 때 한명 이상의 고객이 있으면 바쁜기간(busy period)이 시작된다. 반면에, 단수휴가 시스템은 다음과 같이 진행된다. 서비스할 고객이 없으면 서버는 길이 V의 휴가를 떠난다. 휴가에서 돌아왔을 때 서비스 할 고객이 있으면 바쁜기간이 시작되고, 없으면 기다렸다가 첫 고객의 도착 즉시 바쁜기간이 시작된다(Lee, 2006).

서버의 재가동 여부를 제어하는 대표적인 서비스정책으로서는 Heyman(1977)에 의해 소개된 T-정책이 있다. T-정책은 복수휴가 시스템에서 휴가의 길이 $V_1,\ V_2,\cdots$가 확률변수가 아닌 고정된 T의 값인 경우이다. Heyman(1977)은 T-정책을 따르는 $M/G/1$ 대기행렬 시스템(이하 $M/G/1/T$)의 평균 고객수를 구하고, 비용함수를 통해 단위 시간당 비용을 최소로 하는 최적 T값을 제시하였다.

최근에 Kim and Moon(2006)은 바쁜기간이 끝나면 서버는 p의 확률로 T-정책을 따르거나 $q(=1-p)$의 확률로 휴가를 떠나지 않고 고객을 기다리는 $M/G/1$ 대기행렬 시스템(이하 $M/G/1/(p,\ T)$)의 평균 고객수를 분석하였다. 그리고, 평균 고객수를 제약(constraint)으로 하는 제약문제(constrained problem) 형태의 비용함수를 통해 단위 시간당 비용을 최소로 하는 최적 $(p,\ T)$ 조합을 제시하였다.

본 논문의 목적은 두 가지이다. 첫째로, Kim and Moon (2006)은 $M/G/1/(p,\ T)$ 대기행렬 시스템의 평균 고객수를 제시하였으나, 그 결과에 오류가 있으므로 이를 수정한다. 본 논문에서는 오류가 없는 비용함수를 이용하여 수치 예제를 통해 최적 $(p,\ T)$ 조합을 제시한다. 본 논문의 두 번째 목적은 $M/G/1/(p,\ T)$ 대기행렬 시스템에 대해서 Kim and Moon(2006)이 구하지 않은 기타 주요 성능척도를 구하는 것이다.

본 논문은 다음과 같이 구성된다. 제 2장에서 본 연구에 필요한 기본 가정과 기호를 정의한다. 제 3장에서 $M/G/1/(p,\ T)$의 주요 성능 지표를 PGF(probability generating function) 및 LST(Laplace Stieltjes transform)의 변환형태로 표현하고 이에 대한 미분을 통해 평균값을 얻는다. 제 4장에서는 비용함수를 소개하고 수치예제를 통해 최적 $(p,\ T)$ 조합을 구한다.

† 연락저자 : 이두호, 305-701 대전시 유성구 구성동 373-1 한국과학기술원 산업공학과, Fax : 042-869-3110, E-mail : enjdh@kaist.ac.kr
2007년 10월 접수; 2007년 11월 수정본 접수; 2007년 11월 게재 확정.

2. 모델 및 기호 설명

$M/G/1/(p,\ T)$ 대기행렬 시스템에 대한 정의 및 가정은 다음과 같다. 서버가 한 명인 시스템에 고객은 도착률이 λ인 포아송과정(Poisson process)으로 도착한다. 그리고 도착한 고객은 한명씩 서비스를 받는데, 서비스 시간 $S_1, S_2, \cdots$는 일반분포를 따르는 iid(independent and identically distributed) 확률변수이고 도착과정과 독립이다. 서비스가 하나 끝났을 때 차례를 기다리고 있는 고객이 없으면 바쁜기간이 종료된다. 바쁜기간이 끝나면 서버는 p의 확률로 T-정책을 따르거나 q의 확률로 휴가를 떠나지 않고 고객을 기다린다.

위의 가정 및 추가 사항을 다음의 기호로 정의한다.

$I,\ I^*(\theta)$: 유휴기간(idle period), I의 LST
$B,\ B^*(\theta)$: 바쁜기간, B의 LST
r : 유휴기간에 도착하는 임의고객이 기다리지 않고 바로 서비스 받을 확률
N^* : 유휴기간동안 도착하는 고객수
$R(z)=E[z^{N^*}]$: N^*의 PGF
$I_E(z)=\dfrac{1-R(z)}{(1-z)E[N^*]}$: 유휴기간 중의 안정상태 고객수 PGF
$\rho=\lambda E[S]$: 임의 시점에 서버가 바쁠 확률
$P(z)$: 안정상태(steady-state)에서 임의시점에 관찰된 고객수 PGF
$P^D(z)$: 안정상태에서 이탈하는 임의고객이 남기는 고객수 PGF
$P^A(z)$: 안정상태에서 도착하는 임의고객이 보는 시스템에 있는 고객수 PGF
$W,\ W^*(\theta)$: 체재시간, W의 LST
$W_Q,\ W_Q^*(\theta)$: 대기시간, W_Q의 LST
L : 임의시점 시스템 내 평균 고객수
$Cost$: 단위 시간당 평균 비용

3. $M/G/1/(p,\ T)$ 대기행렬의 분석

3.1 $M/G/1/(p,\ T)$ 대기행렬의 평균 고객수와 평균대기시간 분석

일반휴가(generalized vacation)형 대기행렬이란 고객이 존재함에도 불구하고 서버가 어떤 이유로(정해진 규칙에 따라) 서비스를 제공하지 않는 기간이 존재하는 모든 대기행렬을 의미한다. 일반휴가형 $M/G/1$ 대기행렬의 고객수 PGF에 대한 분해속성(decomposition property)은 다음과 같다(Lee(2006)). 안정상태 고객수의 PGF는 두 가지를 곱한 꼴인데, 하나는 유휴기간 중의 안정상태 고객수 PGF이고, 다른 하나는 휴가가 없는 $M/G/1$ 대기행렬의 안정상태 고객수 PGF이다. 후자는 이미 알려져 있으므로 전자만 구하면 되는데, 전자는 특정 휴가규칙에 따라 달라진다. 일반휴가형 $M/G/1$ 대기행렬의 고객수 PGF는 다음과 같다.

$$P(z)=I_E(z)\frac{(1-\rho)(1-z)S^*(\lambda-\lambda z)}{S^*(\lambda-\lambda z)-z} \tag{1}$$

여기서 $S^*(\lambda-\lambda z)$는 하나의 서비스 시간동안 도착한 고객수 PGF이다. 그러면 $I_E(z)$를 구해보자. 우선 r을 구하면 재생보상정리(renewal reward theorem)에 의해

$$\begin{aligned} r&=\lim_{t\to\infty}\frac{t\text{동안 도착하는 고객 중에서 기다리지 않는 고객수}}{t\text{동안 유휴기간에 도착하는 고객의 수}} \\ &=\frac{q\times 1+p\times 0}{E[N^*]}=\frac{q}{E[N^*]} \end{aligned} \tag{2}$$

이다. $I_E(z)$를 구하기 위해서는 유휴기간동안 도착한 고객수의 PGF인 $R(z)$가 필요하다. 바쁜기간이 종료되면 p의 확률로 복수휴가를 가거나 q의 확률로 휴가를 가지 않고 고객을 기다리므로

$$\begin{aligned} R(z)&=p\times R_{M/G/1/T}(z)+q\times R_{M/G/1}(z) \\ &=p\frac{e^{-(\lambda-\lambda z)T}-e^{-\lambda T}}{1-e^{-\lambda T}}+qz \end{aligned} \tag{3}$$

인데, 해석을 하면 $M/G/1/(p,\ T)$의 $R(z)$는 p의 확률로 $M/G/1/T$ 대기행렬의 $R(z)$가 되고, q의 확률로 $M/G/1$대기행렬의 $R(z)$가 된다. 식 (3)을 z에 대해 미분하고 z에 1을 대입하면

$$E[N^*]=\frac{d}{dz}R(z)\Big|_{z=1}=p\frac{\lambda T}{1-e^{-\lambda T}}+q \tag{4}$$

를 얻는다. 식 (4)를 식 (2)에 대입하면

$$r=\frac{q(1-e^{-\lambda T})}{q(1-e^{-\lambda T})+p\lambda T} \tag{5}$$

를 얻는다. Kim and Moon(2006)은 r값에 대해 오류를 범하였다. 그들이 제시한 r값은 $q/(q+p\lambda T)$인데, 분자, 분모에 $1-e^{-\lambda T}$값이 누락되었다.

$I_E(z)$를 구하기 위한 모든 값은 구해졌다. 위 식 (3), 식 (4)를 이용하여 $I_E(z)$를 구하면

$$\begin{aligned} I_E(z)&=\frac{q(1-z)(1-e^{-\lambda T})+p\{1-e^{-(\lambda-\lambda z)T}\}}{(1-z)\{q(1-e^{-\lambda T})+p\lambda T\}} \\ &=rz^0+(1-r)\frac{1-e^{-(\lambda-\lambda z)T}}{(\lambda-\lambda z)T} \end{aligned} \tag{6}$$

인데, 해석을 하면 유휴기간에 도착했는데 바로 서비스를 받는다면 $M/G/1$ 대기행렬의 $I_E(z)$값을 갖고 유휴기간에 도착했는데 바로 서비스를 받지 못한다면 $M/G/1/T$ 대기행렬의 $I_E(z)$값을 갖게 된다.

위에서 구한 식 (6)을 식 (1)에 대입하면 $P(z)$를 구할 수 있고 그 식은 다음과 같다.

$$P(z)=\frac{(1-\rho)(1-z)S^*(\lambda-\lambda z)}{S^*(\lambda-\lambda z)-z} \times\frac{q(1-z)(1-e^{-\lambda T})+p\{1-e^{-(\lambda-\lambda z)T}\}}{(1-z)\{q(1-e^{-\lambda T})+p\lambda T\}} \quad (7)$$

체재시간 분포의 LST를 구하기 위해 다음의 성질을 이용한다. 시험고객(test customer)의 대기시간은 뒤에 도착하는 사람들의 도착과정에 의해 영향을 받지 않고, 이탈하는 고객이 남기는 고객수는 시험고객이 시스템에 머무는 시간동안 포아송과정으로 도착한 고객이다. 그러므로 다음식이 성립한다.

$$P^D(z)=W^*(\lambda-\lambda z) \quad (8)$$

여기서, Burke의 정리에 의해 $P^D(z)=P^A(z)$이다. 그리고, 고객의 도착과정이 포아송과정이므로 PASTA(Poisson Arrival See Time Average) 속성에 의해 $P^A(z)=P(z)$이다(이상 Lee (2006) 참조). 즉, $P^D(z)=P^A(z)=P(z)$이므로 식 (8)에서 z 대신 $1-\theta/\lambda$를 대입하여 체재시간의 LST를 구하면

$$W^*(\theta)=\frac{\theta(1-\rho)S^*(\theta)}{\theta-\lambda+\lambda S^*(\theta)}\times\frac{q\theta(1-e^{-\lambda T})+p\lambda(1-e^{-\theta T})}{\theta\{q(1-e^{-\lambda T})+p\lambda T\}}$$

이다. 그리고, $W^*(\theta)=W_Q^*(\theta)\,S^*(\theta)$이므로 $W_Q^*(\theta)$는 다음과 같다.

$$W_Q^*(\theta)=\frac{\theta(1-\rho)}{\theta-\lambda+\lambda S^*(\theta)}\times\frac{q\theta(1-e^{-\lambda T})+p\lambda(1-e^{-\theta T})}{\theta\{q(1-e^{-\lambda T})+p\lambda T\}} \quad (9)$$

식 (9)를 θ에 대하여 미분하여 θ에 0을 대입하고 양변에 -1을 곱하면

$$E[W_Q]=-\frac{d}{d\theta}W_Q^*(\theta)\bigg|_{\theta=0}=\frac{\lambda E[S^2]}{2(1-\rho)}+r\times 0+\frac{(1-r)T}{2}$$

인데, 해석을 하면 우변의 첫째항은 $M/G/1$의 평균 대기시간이고 둘째항과 세째항의 합은 유휴기간에 도착하는 고객이 보는 잔여 유휴기간의 기대치가 된다. 즉, 유휴기간에 도착하는 고객이 기다리지 않고 즉시 서비스를 받는다면 잔여 유휴기간은 0임이 분명하고, 즉시 서비스를 받지 못하고 기다려야 한다면 잔여 휴가기간 만큼 더 기다려야 한다.

임의시점 시스템 내 평균 고객수 L은 Little의 법칙에 의해 $L=\lambda\{E[W_Q]+E[S]\}$이다. 그러므로 L은 다음과 같다.

$$L=\frac{\lambda^2E[S^2]}{2(1-\rho)}+\rho+\frac{p\lambda^2T^2}{2\{q(1-e^{-\lambda T})+p\lambda T\}}$$

3.2 $M/G/1/(p,\ T)$ 대기행렬의 Cycle 분석

Cycle은 각자 정의하기 나름이지만 본 논문에서는 유휴기간과 바쁜기간의 합으로 표시한다. 즉, Cycle을 기호상 C로 정의하면 $C=I+B$이다.

우선 유휴기간의 분포의 LST를 구해보자.

$M/G/1/(p,\ T)$의 유휴기간은 바쁜기간 종료 시 p의 확률로 $M/G/1/T$ 대기행렬의 유휴기간이 되고, q의 확률로 $M/G/1$ 대기행렬의 유휴기간이 된다. 그러므로

$$I^*(\theta)=p\times I^*_{M/G/1/T}(\theta)+q\times I^*_{M/G/1}(\theta) =p\frac{e^{-\theta T}-e^{-(\theta+\lambda)T}}{1-e^{-(\theta+\lambda)T}}+q\frac{\lambda}{\lambda+\theta}$$

이고, 기대치는

$$E[I]=p\frac{T}{1-e^{-\lambda T}}+\frac{q}{\lambda}$$

이다. 같은 방법으로 바쁜 기간에 대해서 풀면

$$B^*(\theta)=p\times B^*_{M/G/1/T}(\theta)+q\times B^*_{M/G/1}(\theta) =p\frac{e^{-(\lambda-\lambda B^*_{M/G/1}(\theta))T}-e^{-\lambda T}}{1-e^{-\lambda T}}+qB^*_{M/G/1}(\theta)$$

인데, 여기서 $B^*_{M/G/1}(\theta)$는 휴가가 없는 $M/G/1$ 대기행렬의 바쁜 기간의 LST이다. 기대치를 구하면

$$E[B]=p\frac{\rho T}{(1-\rho)(1-e^{-\lambda T})}+q\frac{E[S]}{1-\rho}$$

이다. $E[C]=E[I]+E[B]$이므로

$$E[C]=p\frac{T}{(1-\rho)(1-e^{-\lambda T})}+q\frac{1}{\lambda(1-\rho)}$$

이다. 재생보상 정리에 의하면 임의시점에 서버가 바쁠 확률은 $E[B]/E[C]=\lambda E[S]=\rho$ 임을 확인할 수 있다.

4. 수치 예제를 통한 $(p,\ T)$-정책의 최적 설계

Heyman(1977)은 $M/G/1/T$ 대기행렬의 비용함수가 볼록함

수(convex function)임을 증명하고 유일한 최적 T(unique minimum T)를 제시하였다. 본 절에서는 Heyman(1977)이 제시한 비용함수를 $M/G/1/(p,\ T)$ 대기행렬에 적용하고 수치 예제를 통해 최적$(p,\ T)$ 조합을 결정한다. 비용함수를 결정하기 위해 다음과 같은 비용을 정의한다.

C_s : Cycle당 준비비용(set-up cost)
C_h : 단위시간당 고객유지비용(holding cost)

위의 비용하에서 $Cost$는 다음과 같이 결정된다.

$$
\begin{aligned}
Cost &= LC_h + \frac{C_s}{E[C]} \\
&= \left\{ \frac{\lambda^2 E[S^2]}{2(1-\rho)} + \rho \right\} C_h \\
&\quad + \frac{p\lambda^2 T^2 C_h\, 2\lambda(1-\rho)(1-e^{-\lambda T})\, C_s}{2\{q(1-e^{-\lambda T}) + p\lambda T\}}
\end{aligned}
\tag{10}
$$

$Cost$ 함수가 p와 T에 대하여 어떠한 함수형태인지 판단하기 어려우므로 수치 예제를 통해 분석을 시행한다. $\lambda=2$, $\rho=0.25$, $C_h=10$, $C_s=20$인 $M/D/1/(p,\ T)$ 대기행렬의 $Cost$ 함수는 다음 그림과 같다.

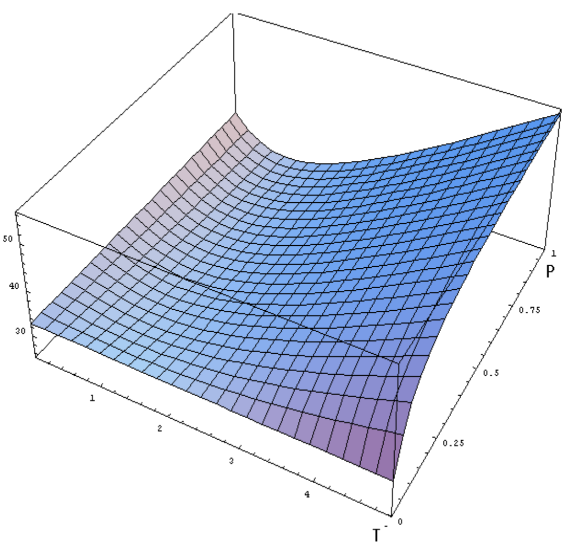

Figure 1. Graph of $Cost$

위의 그래프에서 볼 수 있듯이 $Cost$ 함수는 T에 대해서는 p에 상관없이 볼록함수이지만, p에 대해서는 T가 작으면 볼록함수이고 T가 크면 오목함수로 변한다. 많은 수치 예제 결과 모두 위의 현상들이 나타났다. 그러므로 $Cost$ 함수가 p와 T에 대하여 볼록함수인지 오목함수인지 결론지을 수 없다.

여러 수치 예제를 통해서 최적$(p,\ T)$ 조합을 얻는 과정에서 최적 p는 0 또는 1로 수렴함을 알 수 있었다. 위의 수치 예제에서 최적$(p,\ T)$는 $(1.0,\ 0.8953)$이고 $Cost$는 25.8282이다. 이외에 여러 수치 예제의 결과는 다음 표와 같다.

Table 1. Optimal$(p,\ T)$ combination with $\lambda=2$ and $C_h=10$

C_s	ρ	p	T	$Cost$
5	0.2	0	0	10.2500
	0.5	0	0	12.5000
	0.8	0	0	26.0000
10	0.2	1	0.3639	17.2544
	0.5	0	0	17.5000
	0.8	0	0	28.0000
20	0.2	1	0.9535	26.0730
	0.5	1	0.5458	25.1295
	0.8	0	0	32.0000

Table 2. Optimal$(p,\ T)$ combination with $\lambda=4$ and $C_h=10$

C_s	ρ	p	T	$Cost$
5	0.2	1	0.1819	18.0054
	0.5	0	0	25.0000
	0.8	0	0	76.0000
10	0.2	1	0.4768	26.8230
	0.5	1	0.2729	32.6295
	0.8	0	0	80.0000
20	0.2	1	0.8190	38.1780
	0.5	1	0.5807	42.1470
	0.8	1	0.1819	87.0054

5. 요약 및 결론

본 논문에서는 $(p,\ T)$-정책을 따르는 $M/G/1$ 모형을 분석하였다.

본 논문과 Kim and Moon(2006)의 차이점은 다음과 같다. Kim and Moon(2006)은 Heyman(1977)이 분석한 결과와 그들이 구한 r을 이용하여 평균 고객수를 구하였고, 이를 비용함수를 유도하는데 이용하였다. 그러나 본 논문에서는 시스템의 성능척도들을 변환형태로 제시하였다. 휴가형 $M/G/1$ 대기행렬에서 성립하는 분해속성을 이용하여 고객 분포의 PGF를 제시하였고, 고객 이탈시 시스템 내 고객수와 체제시간 사이에 성립하는 관계를 이용하여 체제시간 및 대기시간의 LST를 제시하였다. 그리고 미분을 통해 Kim and Moon(2006)의 평균 고객수를 수정하고, Cycle 분석의 결과를 이용하여 단위시간당 비용함수를 구하였다. 그러나 단위 시간당 비용함수의 볼록성(convexity)을 분석적으로(analytical) 규명하기 어려워 비용을 최소로 하는 최적$(p,\ T)$ 조합을 수치 예제를 통해 실험적으로 구하였다.

수치 예제를 통해 최적$(p,\ T)$ 조합을 얻는 과정에서 최적 p

는 0 또는 1값을 갖게 됨을 알 수 있었는데, 이는 (p, T)-정책의 최적 설계 문제는 두단계의 문제로 나누어 진다고 사료된다. 1단계 문제는 바쁜기간 종료 시 서버는 T-정책을 따를 것인가의 문제이고, 2단계 문제는 T-정책을 따른다면 그때의 최적 T 값은 무엇인가의 문제이다.

추후연구 과제는 다음과 같다. 본 논문에서는 최적 (p, T) 조합을 비용 측면에서만 분석하였는데 이를 이익 측면에서도 분석할 수 있을 것이다. 예를 들어 서버는 휴가기간 동안 비가동(switched-off) 상태가 아니라 가동(switched-on) 상태를 유지하고 다른 작업을 함으로써 수익을 낼 수 있는 경우이다. 그러면 기존의 비용함수에 새로운 수익항을 추가함으로써 더욱 정교하게 최적 (p, T) 설계를 할 수 있을 것이다(비고 : 상기 추후연구 과제는 본 논문의 심사위원이 제안하였음).

참고문헌

Daniel P. Heyman. (1977), The T-policy for the M/G/1 Queue, *Management Science*, **23**(7), 775-778.

Kim, D.J. and Moon, S.A. (2006), Randomized Control of T-policy for an M/G/1 system, *Computers and Industrial Engineering*, **51**(4), 684-692.

Lee, H. W. (2006), *Queueing Theory (3rd Edition)*, Sigma Press, Seoul, Korea.