韓國經營科學會誌
第22卷 第2號
1997年 6月

M/M/2 직렬-서어버 모형의 분석 및 응용

양원석* · 채경철**

A Markovian queue with two serial servers and its application to the double tollbooth system

Won S. Yang* · Kyung C. Chae**

Abstract

We consider an M/M/2 queue with two servers placed in series. System performance measures that we present in closed expressions are the first and the second moments for the system size, the queue waiting time and the sojourn time. We also present an algorithm for computing the queue waiting time distribution function based on the randomization method. As an application, we analyze the double tollbooth system and compare its performance with the conventional single tollbooth system's.

1. 서 론

단일 서어버 대기행렬 모형(single server queue)에서 과부하(heavy traffic)에 의해 시스템에 정체(congestion)가 발생하는 경우 서어버의 수를 늘려 정체에 따른 시간손실(time delay)을 줄일 수 있다. 이때 일반적으로 복수(multiple)의 서어버들은 병렬(parallel)로 배치됨을 가정한다. 그러나 공간의 제약에 따른 비용상의 문제로 서어버를 병렬로 배치하기 어려운 상황이 발생할 수 있으며 그런 경우에는, 〈그림 1〉과 같이, 서어버를 직렬(series)로 배치한 직렬-서어버 대기행렬 모형(queue with servers in series)을 고려할 수 있다.

〈그림 1〉 직렬-서어버 대기행렬 모형

* 한국과학기술원 산업공학과 박사과정
** 한국과학기술원 산업공학과 교수

직렬-서어버 대기행렬 모형에서는 시스템에 도착한 고객은 서어버 1과 2중 단 한곳에서만 서비스를 받고 시스템을 떠난다. 두 서어버가 모두 유휴상태(idle)일 때 도착한 고객은 서어버 2에서 서비스를 받는다. 서어버 2가 서비스중이고 서어버 1이 유휴상태 일 때 도착한 고객은 서어버 1에서 바로 서비스를 받는다. 그러나 서어버 1이 서비스중일 때 도착한 고객은 대기장소에서 기다린다. 서어버 1에서 서비스를 끝낸 고객은 서어버 2가 서비스중이면 시스템을 떠나지 못하고 그 자리에서 머물다가 서어버 2의 서비스가 끝나면 서어버 2에서 서비스를 마친 고객과 동시에 시스템을 떠난다.

고객들은 도착률이 λ인 포아송과정에 따라 도착한다. 서어버 j의 서비스 시간 S_j는 서비율이 μ_j인 지수분포를 따르며 서로 독립이라 가정한다, $j=1,2$. 서비스정책은 선입선출방식(FIFO)이고 대기장소의 크기는 무한대이다.

직렬-서어버 대기행렬 모형의 구체적인 예로는 복수창구 요금소 모형(double tollbooth system) [1,2]을 들 수 있다. 조면식 [1,2]은 고속도로 톨게이트의 심각한 정체 현상을 줄이기 위해 하나의 차선에 요금소를 직렬로 배치하는 복수창구 요금소 모형을 제안하고 실제 표본 자료를 바탕으로 시뮬레이션을 통해 분석하였다. 저자가 알기로 국내에서 아직까지 복수창구 요금소 모형에 대한 이론적인 연구가 없기 때문에 복수창구 요금소 모형을 이론적으로 분석해보고자 하는 것이 본 논문의 연구 동기이다.

직렬-서어버 대기행렬 모형의 주요한 연구로는 Clarke([4] pp. 232-245 참조)을 들 수 있다. Clarke은 알고리즘적인 접근방법(algorithmic approach)을 이용하여 서어버 1과 서어버 2 사이에 중간 대기장소가 있는 M/M/2 직렬-서어버 대기행렬 모형의 정상상태(定常,steady-state)에서의 고객수 및 대기시간(queue waiting time)을 분석하였다.

따라서 본 논문에서 다루는 직렬-서어버 모형은 중간 대기장소가 없는 Clarke의 모형과 동일하다. 그러나 본 본문에서는 대기시간과 체류시간(sojourn time)의 성능치(performance measure)들을 알고리즘(algorithm)이 아닌 간단한 형태(closed form expression)로 표현하고 이에 대한 해석을 제공한다. 그리고 대기시간의 분포함수를 계산하는 간단한 알고리즘을 제공한다. 아울러, 직렬-서어버 모형의 구체적인 예로 복수창구 요금소 모형을 다룬다.

본 논문의 구성은 다음과 같다. 2절에서는 행렬기하(matrix-geometric) 방법을 이용하여 정상상태에서 시스템내 고객수의 결합분포를 구한다. 그리고 시스템내 고객수 분포와 대기장소내 고객수 분포의 1차, 2차 모멘트를 간단한 형태로 나타낸다. 3절에서는 정상상태에서 대기시간의 Laplace-Stieltjes transform(LST)과 1차, 2차 모멘트를 간단한 형태로 나타내고 평균 대기시간을 해석한다. 아울러, 무작위화 방법(randomization method) [3]을 이용하여 대기시간 분포함수를 계산하는 알고리즘을 제공한다. 4절에서는 정상상태에서 체류시간의 LST와 1차, 2차 모멘트를 간단하게 나타내고 평균 체류시간을 해석한다. 5절에서는 간단한 수치예를 다룬다. 그리고 6절에서는 복수창구 요금소 모형을 다룬다.

본 논문에서 $\boldsymbol{e}$는 원소가 모두 1인 열벡터(column vector)이다. 그리고 첨자 t를 붙인 벡터들은 모두 열벡터들이다.

2. 시스템내 고객수

정상상태에서 시스템의 상태(state)를 $(n_q, \delta_1, \delta_2)$ 라 표기한다. 여기에서 n_q는 대기장소에서 기다리는 고객수이고, $n_q \geq 0$, δ_j는 서어버 j의 상태이며 다음과 같다.

$$\delta_1 = \begin{cases} 0 & \text{유휴상태} \\ 1 & \text{서비스중,} \\ b & \text{차단상태} \end{cases} \quad \delta_2 = \begin{cases} 0 & \text{유휴상태} \\ 1 & \text{서비스중} \end{cases}$$

그리고 다음과 같이 초상태(superstate)를 정의한다.

$$\begin{aligned} \boldsymbol{\beta} &= \{(0,0,0),\ (0,0,1)\} \\ \boldsymbol{i} &= \{(i,1,0),\ (i,1,1),\ (i,b,1)\},\quad i \geq 0 \end{aligned}$$

상태들을 $\{\boldsymbol{\beta}, \mathbf{0}, \mathbf{1}, \mathbf{2}, \cdots\}$의 순서로 늘어놓으면 무한소 생성자(infinitesimal generator) Q는 다음과 같이 주어진다.

$$Q = \begin{bmatrix} B_0 & C_0 & 0 & 0 & 0 & \cdots \\ B_1 & A_1 & A_0 & 0 & 0 & \cdots \\ B_2 & 0 & A_1 & A_0 & 0 & \cdots \\ 0 & A_3 & 0 & A_1 & A_0 & \cdots \\ 0 & 0 & A_3 & 0 & A_1 & \cdots \\ 0 & 0 & 0 & A_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{1}$$

여기에서 각 행렬들은 다음과 같다.

$$B_0 = \begin{bmatrix} -\lambda & \lambda \\ \mu_2 & -(\lambda+\mu_2) \end{bmatrix},\ B_1 = \begin{bmatrix} \mu_1 & 0 \\ 0 & 0 \\ \mu_2 & 0 \end{bmatrix},\ B_2 = \begin{bmatrix} 0 & \mu_1 \\ 0 & 0 \\ 0 & \mu_2 \end{bmatrix},\ C_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \end{bmatrix},\ A_0 = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

$$A_1 = \begin{bmatrix} -(\lambda+\mu_1) & 0 & 0 \\ \mu_2 & -(\lambda+\mu_1+\mu_2) & \mu_1 \\ 0 & 0 & -(\lambda+\mu_2) \end{bmatrix},\ A_3 = \begin{bmatrix} 0 & \mu_1 & 0 \\ 0 & 0 & 0 \\ 0 & \mu_2 & 0 \end{bmatrix}$$

(1)의 Q를 따르는 확률과정은 $\boldsymbol{\pi} A_0 \boldsymbol{e} < 2\boldsymbol{\pi} A_3 \boldsymbol{e}$이면 안정(stable)하고 [4], 이를 간단하게 나타내면 다음과 같다.

$$\lambda < \frac{2\mu_1\mu_2(\mu_1+\mu_2)}{\mu_1^2+\mu_1\mu_2+\mu_2^2} \tag{2}$$

여기에서 $\boldsymbol{\pi}$는 $A = A_0 + A_1 + A_3$의 정상상태 확률벡터이다. 정상상태에서 시스템 상태가 상태 (i, δ_1, δ_2)에 있을 확률을 $x_{(i,\delta_1,\delta_2)}$라 표기하고 다음과 같은 확률벡터를 정의한다.

$$\begin{aligned} \boldsymbol{x}_\beta &= (x_{(0,0,0)},\ x_{(0,0,1)}) \\ \boldsymbol{x}_i &= (x_{(i,1,0)}, x_{(i,1,1)}, x_{(i,b,1)}) \end{aligned}$$

그러면 확률벡터 $\boldsymbol{x}_i$는

$$\boldsymbol{x}_i = \boldsymbol{x}_0 R^i \quad, i = 0, 1, \cdots \tag{3}$$

를 만족시킨다. 행렬 R은 $R^3A_3 + RA_1 + A_0 = 0$의 비음최소해(minimal nonnegative solution)이고 다음과 같은 반복적인 절차를 통해 구할 수 있다.

$$\begin{aligned} R_0 &= 0 \\ R_{i+1} &= -A_0 A_1^{-1} - R_i^3 A_3 A_1^{-1}, \\ & \qquad i = 0, 1, \cdots \end{aligned} \tag{4}$$

또한 $\boldsymbol{x}_\beta$와 $\boldsymbol{x}_0$는

$$\begin{aligned} \boldsymbol{x}_\beta B_0 + \boldsymbol{x}_0 [B_1 + RB_2] &= \mathbf{0} \\ \boldsymbol{x}_\beta C_0 + \boldsymbol{x}_0 [A_1 + R^2 A_3] &= \mathbf{0} \end{aligned} \tag{5}$$

를 만족시키고

$$\boldsymbol{x}_\beta \boldsymbol{e} + \boldsymbol{x}_0 (I-R)^{-1} \boldsymbol{e} = 1 \tag{6}$$

에 의해 정규화(normalization)된다. (4)에서 구

한 R을 (5)와 (6)에 대입하면 $\boldsymbol{x}_\beta$ 와 $\boldsymbol{x}_0$ 를 쉽게 구할 수 있다. 그리고 $\boldsymbol{x}_0$ 를 (3)에 대입하면 시스템내 고객수의 결합분포를 모두 구할 수 있다.

시스템내 고객수를 N으로 대기장소내 고객수를 N_Q 로 표기하자. N과 N_Q 의 1차, 2차 모멘트는 다음과 같이 간단하게 표현된다.

$$\begin{aligned} E[N_Q] &= \boldsymbol{x}_0 R(I-R)^{-2}\,\boldsymbol{e} \\ E[N_Q{}^2] &= \boldsymbol{x}_0 R(I-R)^{-2} \\ &\quad [I+2R(I-R)^{-1}]\,\boldsymbol{e} \end{aligned} \tag{7}$$

$$\begin{aligned} E[M] &= 1+x_{(0,0,1)} + \boldsymbol{x}_0(I-R)^{-2}\,\boldsymbol{e} \\ &\quad - \boldsymbol{x}_0(I-R)^{-1}\,\boldsymbol{a}_1 \\ E[N^2] &= E[N_Q^2]+4E[N_Q]+4- \\ &\quad \boldsymbol{x}_0[2R(I-R)^{-2}+3(I-R)^{-1}]\,\boldsymbol{a}_1 \end{aligned} \tag{8}$$

여기에서 $\boldsymbol{a}_1=(1,0,0)^t$ 이다.

3. 대기시간

3.1 대기시간의 성능치

본 논문에서는 대기시간(queue waiting time)을 시험고객(test customer)의 도착 시점에서 그 고객의 서비스가 시작되기 직전까지의 시간으로 정의한다. 도착시점에서의 대기시간을 W_Q, 대기장소에서 기다리는 고객수를 N_Q 로 표기한다. PASTA [5]에 의해 N_Q 는 임의 시점의 대기고객수가 된다. 확률변수 $M=M_n=\max(S_1, S_2)$ 라 정의한다, $n=1,2,\cdots$. 그리고 대기시간 분포의 LST를 $W_Q^*(s)$, 서어버 j의 서비스 시간의 LST를 $S_j^*(s)$, 확률변수 M 과 M_n, $n=1,2,\cdots$의 LST를 $M^*(s)$라 표기한다.

시험고객이 도착한 상태에 조건을 걸면 W_Q 는 다음과 같은 확률합(random sum)으로 표현된다.

$$W_Q=\begin{cases} 0 & \text{if } \delta_1=0 \\ S_1+\sum_{n=0}^{[N_Q/2]} M_n & \text{if } \delta_1=1,\ \delta_2=0 \\ M+\sum_{n=0}^{[N_Q/2]} M_n & \text{if } \delta_1=1,\ \delta_2=1 \\ S_2+\sum_{n=0}^{[N_Q/2]} M_n & \text{if } \delta_1=b,\ \delta_2=1 \end{cases} \tag{9}$$

여기에서 $[a]$는 a를 넘지 않는 최대 정수이다. (9)의 양변에 LST를 취하면 $W_Q^*(s)$는 다음과 같이 표현된다.

$$\begin{aligned} W_Q^*(s) = &\ \boldsymbol{x}_\beta\,\boldsymbol{e} + \boldsymbol{x}_0(I+R) \\ & [I-M^*(s)R^2]^{-1}\,\boldsymbol{H}^*(s) \end{aligned} \tag{10}$$

이때 $\boldsymbol{H}^*(s)=(S_1^*(s), M^*(s), S_2^*(s))^t$ 이다. 그리고 (10)으로 부터 다음과 같은 대기시간 분포의 1차, 2차 모멘트를 얻는다.

$$\begin{aligned} E[W_Q] = &\ \boldsymbol{x}_0(I-R)^{-1}\,\boldsymbol{b}_1 + E(M) \\ & \boldsymbol{x}_0(I-R)^{-1}R^2(I-R^2)^{-1}\boldsymbol{e} \end{aligned} \tag{11}$$

$$\begin{aligned} E[W_Q{}^2] = &\ \boldsymbol{x}_0(I-R)^{-1}\,\boldsymbol{b}_2 + 2E(M)\ \boldsymbol{x}_0(I-R)^{-1}R^2(I-R^2)^{-1}\,\boldsymbol{b}_1 \\ & + 2E(M)^2\,\boldsymbol{x}_0(I-R)^{-1}R^2(I-R^2)^{-2}\boldsymbol{e} \\ & + E(M^2)\,\boldsymbol{x}_0(I-R)^{-1}R^2(I-R^2)^{-1}\boldsymbol{e} \end{aligned} \tag{12}$$

이때 $\boldsymbol{b}_1=(E[S_1], E[M], E[S_2])^t$, $\boldsymbol{b}_2=(E[S_1^2], E[M^2], E[S_2^2])^t$, $E[S_j]=1/\mu_j$, $E[S_j^2]=2/\mu_j^2$, $E[M]=\dfrac{\mu_1^2+\mu_1\mu_2+\mu_2^2}{\mu_1\mu_2(\mu_1+\mu_2)}$ 그리고 $E[M^2]=\dfrac{2}{\mu_1^2}+\dfrac{2}{\mu_2^2}-\dfrac{2}{(\mu_1+\mu_2)^2}$ 이다.

3.2 평균 대기시간의 해석

도착하는 고객의 관점에서 평균 대기시간은 다음과 같이 표현된다.

$$E[W_Q] = E[W_1] + E[W_2] \tag{13}$$

여기에서 W_1 은 시험고객의 도착시점부터 대기장소내 가장 앞에 있는 고객이 서비스를 받기 직전까지의 시간이고 W_2 는 대기장소내 가장 앞에 있는 고객이 서비스를 받는 시점부터 시험고객이 서비스를 받기 직전까지의 시간이다.

τ를 시험고객이 도착한 상태라 하자. 대기장소내 i 명의 고객이 있는 조건하에서

$$E[W_1|\tau] = \begin{cases} E[S_1] & \text{if } \tau=(i,1,0) \\ E[M] & \text{if } \tau=(i,1,1) \\ E[S_2] & \text{if } \tau=(i,b,1) \end{cases}$$

$$E[W_2|i\text{명 고객이 대기중}] = [i/2]E[M]$$

이다. 여기에서 [a]는 a를 넘지 않는 최대의 정수이다. 시험고객이 도착한 상태의 조건을 없애면

$$\begin{aligned} E[W_1] &= \textstyle\sum_{i=0}^{\infty} \boldsymbol{x}_i \ \boldsymbol{b}_1 \\ E[W_2] &= \textstyle\sum_{i=1}^{\infty} (\boldsymbol{x}_{2i} + \boldsymbol{x}_{2i+1})(iE[M]) \end{aligned} \tag{14}$$

이다. (3)을 (14)에 대입해 정리한 후 (14)를 (13)에 대입하면 (11)과 동일한 형태의 평균 대기시간을 얻는다.

3.3 대기시간의 분포함수

$W_Q(\cdot)$를 도착시점에서의 대기시간의 분포함수라 하자. 도착하는 고객의 관점에서 대기시간은 고객의 도착 이후 더 이상의 도착이 없다는 가정 하에서 시스템 상태가 도착 상태에서 초상태 $\boldsymbol{\beta}$로 흡수되기 직전까지의 시간이다. 이제, 초상태 $\boldsymbol{\beta}$를 흡수상태(absorbing state)로 갖는 흡수 연속 마코프사슬(absorbing continuous time Markov chain) $\{Y(t), t\geq 0\}$를 생각해보자. 여기에서 $Y(t)$는 시간 t에서 시스템의 상태이고 도착시점을 시간 0으로 정의한다. $\{Y(t)\}$의 무한소 생성자는 다음과 같이 주어진다.

$$\tilde{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ B_1 & D & 0 & 0 & 0 & \cdots \\ B_2 & 0 & D & 0 & 0 & \cdots \\ 0 & A_3 & 0 & D & 0 & \cdots \\ 0 & 0 & A_3 & 0 & D & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{15}$$

(15)의 $\tilde{Q}$는 (1)에서 $\lambda=0$ 그리고 B_0 에서 $\mu_1=\mu_2=0$으로 하여 얻은 것이다. 시간 t에서 시스템 상태가 상태 (i, δ_1, δ_2)에 있을 확률을 $y_{(i,\delta_1,\delta_2)}(t)$라 표기한다. 그리고 다음과 같은 확률 벡터를 정의한다:

$$\begin{aligned} \boldsymbol{y}_\beta(t) &= (y_{(0,0,0)}(t),\ y_{(0,0,1)}(t)) \\ \boldsymbol{y}_i(t) &= (y_{(i,1,0)}(t),\ y_{(i,1,1)}(t), \\ &\qquad y_{(i,b,1)}(t)) \quad, i\geq 0 \\ \boldsymbol{y}(t) &= (\ \boldsymbol{y}_\beta(t),\ \boldsymbol{y}_0(t),\ \boldsymbol{y}_1(t), \cdots) \end{aligned}$$

PASTA에 의해 초기조건은 $\boldsymbol{y}(0) = (\boldsymbol{x}_\beta, \boldsymbol{x}_0, \boldsymbol{x}_1, \cdots)$로 주어진다. $\boldsymbol{y}_\beta(t)$는 시간 t에 시스템 상태가 초상태 $\boldsymbol{\beta}$에 도달했을 확률이므로

$$W_Q(t) = P(W_Q < t) = \boldsymbol{y}_\beta(t)\,\boldsymbol{e} \tag{16}$$

이다. 이제 $\boldsymbol{y}_\beta(t)$를 구하면 대기시간의 분포함수를 구할 수 있다. $\boldsymbol{y}_\beta(t)$는 일반적으로 Kolmogorov 미분방정식 $\boldsymbol{y}'(t) = \boldsymbol{y}(t)\tilde{Q}$를 풀어서 구할 수 있다. 그러나 이 방법은 계산량이

너무 많고 수치적으로 불안정할 수 있기 때문에 본 논문에서는 연속적 마코프사슬(continuous time Markov chain)의 전이해(transient solution)를 구하는 무작위화 방법 [3]을 이용하여 $y_\beta(t)$를 구한다.

$P=(1/\Lambda)\tilde{Q}+I$라고 하자. 이때, $\Lambda=\mu_1+\mu_2$이고 I는 정방행렬(identity matrix)이다. 식 (15)의 $\tilde{Q}$를 대입하면 P는 다음과 같이 주어진다.

$$P = \begin{bmatrix} I & 0 & 0 & 0 & 0 & \cdots \\ P_1 & P_0 & 0 & 0 & 0 & \cdots \\ P_2 & 0 & P_0 & 0 & 0 & \cdots \\ 0 & P_3 & 0 & P_0 & 0 & \cdots \\ 0 & 0 & P_3 & 0 & P_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

여기에서 $P_0=(1/\Lambda)D+I$, $P_1=(1/\Lambda)B_1$, $P_2=(1/\Lambda)B_2$ 그리고 $P_3=(1/\Lambda)A_3$이다. [3]에 의하면 $\boldsymbol{y}(t)$는 다음과 같이 표현된다.

$$\begin{aligned} \boldsymbol{y}(t) &= \sum_{n=0}^{\infty} \boldsymbol{y}(0)P^n \frac{\exp(-\Lambda t)\,(\Lambda t)^n}{n!} \\ &= \sum_{n=0}^{\infty} \boldsymbol{\phi}(n) \frac{\exp(-\Lambda t)\,(\Lambda t)^n}{n!} \end{aligned} \tag{17}$$

여기에서 $\boldsymbol{\phi}(n)=\boldsymbol{y}(0)P^n=[\ \boldsymbol{\phi}_\beta(n),\ \boldsymbol{\phi}_0(n),\ \boldsymbol{\phi}_1(n),\cdots]$이다, $n=0,1,\cdots$. 행렬 P의 특수한 구조와 관계식 $\boldsymbol{\phi}(n)=\boldsymbol{\phi}(n-1)P$를 이용하면 $\boldsymbol{\phi}_\beta(n)$, $\boldsymbol{\phi}_0(n)$ 그리고 $\boldsymbol{\phi}_1(n)$는 다음과 같이 표현된다.

$$\begin{aligned} \boldsymbol{\phi}_\beta(n) &= \boldsymbol{\phi}_\beta(n-1) + \boldsymbol{\phi}_0(n-1)P_1 + \boldsymbol{\phi}_1(n-1)P_2 \\ \boldsymbol{\phi}_0(n) &= \boldsymbol{x}_0 F_n \\ \boldsymbol{\phi}_1(n) &= \boldsymbol{x}_1 F_n \end{aligned} \tag{18}$$

이때 $\boldsymbol{\phi}_\beta(0)=\boldsymbol{x}_\beta$ $\boldsymbol{\phi}_0(0)=\boldsymbol{x}_0$ 그리고 $\boldsymbol{\phi}_1(0)=\boldsymbol{x}_1$ 이고 F_n은 다음과 같다.

$$\begin{aligned} F_0 &= I \\ F_n &= F_{n-1}P_0+R^2F_{n-1}P_3, \quad n\geq 1 \end{aligned}$$

(16), (17) 그리고 (18)로부터 대기시간의 분포함수 $W_Q(t)$를 계산할 수 있다. 아울러, $W_Q(t)$를 미분하면 대기시간의 확률밀도함수 $f_{W_Q}(t)$를 얻는다.

$$W_Q(t) = \sum_{n=0}^{\infty}\left\{\frac{\exp(-\Lambda t)(\Lambda t)^n}{n!}\right\}\boldsymbol{\phi}_\beta(n)\ \boldsymbol{e} \tag{19}$$

$$f_{W_Q}(t) = \Lambda\exp(-\Lambda t)\left[-\ \boldsymbol{\phi}_\beta(0)\ \boldsymbol{e} + \sum_{n=1}^{\infty}\left(\frac{(\Lambda t)^{n-1}}{(n-1)!}-\frac{(\Lambda t)^n}{n!}\right)\boldsymbol{\phi}_\beta(n)\ \boldsymbol{e}\right] \tag{20}$$

4. 체류시간

체류시간은 시험고객이 도착한 시점부터 시스템을 떠나기 직전까지의 시간이다. 정상상태에서 도착 시점에서의 체류시간을 T, 그리고 T의 LST를 $T^*(s)$라 표기하자. 시험고객이 도착한 상태에 조건을 걸면 T는 다음과 같이 표현된다.

$$T=\begin{cases} Z_k & \text{if } (n_q,\delta_1,\delta_2)=(0,0,k) \\ Z_k+S_1+\sum_{j=0}^{m}M_j & \text{if } (n_q,\delta_1,\delta_2)=(2m+k,1,0) \\ Z_k+M+\sum_{j=0}^{m}M_j & \text{if } (n_q,\delta_1,\delta_2)=(2m+k,1,1) \\ Z_k+S_2+\sum_{j=0}^{m}M_j & \text{if } (n_q,\delta_1,\delta_2)=(2m+k,b,1) \end{cases} \tag{21}$$

여기에서 $k=0,1$이고 $Z=(S_2,M)^t$이다. (21)의 양변에 LST를 취하면

$$T^*(s) = \boldsymbol{x}_\beta\ Z^*(s)+\ \boldsymbol{x}_0\left[S_2^*(s)I+M^*(s)R\right]\left[I-M^*(s)R^2\right]^{-1} H^*(s) \tag{22}$$

이다. 여기에서 $Z^*(s)=(S_2^*(s), M^*(s))^t$ 이고 $H^*(s)=(S_1^*(s), M^*(s), S_2^*(s))^t$ 이다. 그리고 (22)로부터 체류시간의 1차, 2차 모멘트를 얻는다.

$$E[T] = \boldsymbol{x}_\beta\ c_1 + \boldsymbol{x}_0(E[S_2]I + E[M]R)(I-R^2)^{-1}\boldsymbol{e} + E[W_Q] \tag{23}$$

$$\begin{aligned} E[T^2] = {} & E[{W_Q}^2] + \boldsymbol{x}_\beta\ c_2 + \boldsymbol{x}_0(E[S_2^2]I + E[M^2])(I-R^2)^{-1}\boldsymbol{e} \\ & + 2E[M]\ \boldsymbol{x}_0(E[S_2]I + E[M]R)(I-R^2)^{-1}R^2(I-R^2)^{-1}\boldsymbol{e} \\ & + 2\ \boldsymbol{x}_0(E[S_2]I + E[M]R)(I-R^2)^{-1}\ \boldsymbol{b}_1 \end{aligned} \tag{24}$$

여기에서 $c_1=(E[S_2], E[M])^t$ 이고 $c_2=(E[S_2^2], E[M^2])^t$ 이다.

4.1 평균 체류시간의 해석

도착하는 고객의 관점에서 평균 체류시간은 다음과 같이 표현된다.

$$E[T] = E[W_Q] + E[S_T] \tag{25}$$

여기에서 S_T는 시험고객의 서비스가 시작된 시점에서 그 고객이 시스템을 떠나기 직전까지의 시간이다. τ를 시험고객이 도착한 상태라고 표기하면

$$E[S_T|\tau] = \begin{cases} E[S_2] & \text{if } \tau=(0,0,0) \text{ or } \tau\in \boldsymbol{2i} \\ E[M] & \text{if } \tau=(0,0,1) \text{ or } \tau\in \boldsymbol{2i+1} \end{cases}$$

이다. 시험고객이 도착한 상태의 조건을 풀면

$$\begin{aligned} E[S_T] = {} & \boldsymbol{x}_\beta\ c_1 + E[S_2]\Sigma_{i=0}^{\infty}\ \boldsymbol{x}_{2i}\,\boldsymbol{e} \\ & + E[M]\Sigma_{i=0}^{\infty}\ \boldsymbol{x}_{2i+1}\,\boldsymbol{e} \end{aligned} \tag{26}$$

이다. (3)을 (26)에 대입해서 정리하고 (26)을 (25)에 대입하면 식 (23)과 동일한 형태의 평균 체류시간을 얻는다.

5. 수치예

수치예로 $\lambda=\mu_1=\mu_2=1$인 경우를 다룬다. 먼저, 식 (4)에서 행렬 R을 계산하면 다음과 같다.

$$R = \begin{pmatrix} 0.548484 & 0.096968 & 0.048484 \\ 0.240597 & 0.481194 & 0.240597 \\ 0.048484 & 0.096968 & 0.548484 \end{pmatrix}$$

(5)와 (6)에 R을 대입한 후 선형 방정식을 풀면 $\boldsymbol{x}_\beta$와 $\boldsymbol{x}_0$를 다음과 같이 얻는다.

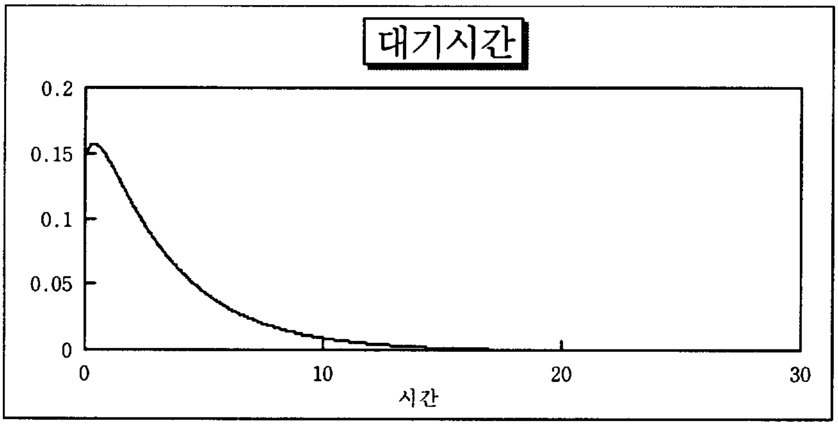
$$\begin{aligned} \boldsymbol{x}_\beta &= (0.213898\ ,\ 0.144409) \\ \boldsymbol{x}_0 &= (0.034744\ ,\ 0.069489,\ 0.034744) \end{aligned}$$

$\boldsymbol{x}_0$를 (3)에 대입하면 $\boldsymbol{x}_i$, $i\geq 1$들을 모두 계산할 수 있다. 그리고 (7), (8), (11), (12), (23), (24)로부터 〈표 1〉과 같은 시스템 성능치들을 얻는다.

〈표 1〉에서 리틀의 법칙(Little's law)이 성립함을 알 수 있다. 아울러, 식 (20)을 계산하면 〈그림 2〉와 같은 대기시간의 확률밀도함수를 얻는다.

〈표 1〉 $\lambda=\mu_1=\mu_2=1$일 때 성능치

$E[N]$	$E[T]$	$E[N^2]$	$E[T^2]$	$E[N_Q]$	$E[W_Q]$	$E[N_Q^2]$	$E[W_Q^2]$
3.358307	3.358307	26.369169	25.909345	2.144409	2.144409	16.005431	18.048323

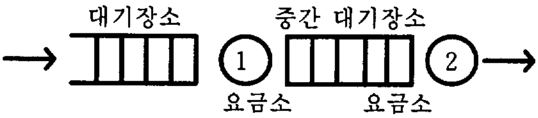


〈그림 2〉 $\lambda=\mu_1=\mu_2=1$일 때 대기시간의 확률밀도함수

대기시간의 확률밀도함수는 꼬리가 긴 형태이다.

6. 복수창구 요금소 모형

[2]에서 제안한 복수창구 요금소 모형에서는 〈그림 3〉과 같이 중간 대기장소를 둘 수 있다고 가정한다.



〈그림 3〉 복수창구 요금소 모형

중간 대기장소에 서비스를 받은 고객과 서비스를 받지 않은 고객이 혼재되어 있으면, 중간 대기장소에서 대기중인 고객이 요금소 2를 지나갈 때 이 고객이 요금소 1에서 서비스를 받았는지의 여부를 요금소 2에서 확인하기 어렵다. 따라서 본 논문에서는, 요금소 1에서 서비스를 마친 고객만 중간 대기장소에 대기시킨다. 그리고 요금소 2의 서비스가 끝나면 중간 대기장소에서 기다리던 고객들과 요금소 2에서 서비스를 마친 고객이 동시에 시스템을 떠난다. 대기장소의 크기는 무한대이고 중간 대기장소의 크기는 유한인 K이다. 시스템에 관한 다른 사항들은 1절의 직렬-서어버 모형과 동일하다.

(비고) 중간 대기장소에 서비스를 받지 않은 고객을 대기시키고 그 고객들을 요금소 2에서 서비스할 수도 있다. 그러나 그런 경우에는 대체적으로 시스템의 성능이 떨어지기 때문에 본 논문에서는 고려하지 않는다.

고속도로 톨게이트에서 교통수요가 작은 경우에는 단일창구 요금소만으로 고객들을 처리할

수 있다. 그러므로 과부하로 인해 상당수의 고객이 대기중인 정체 상태라야 비로소 복수창구 요금소 모형이 의미 있다. 정체 상태에서 어느 정도의 기다림을 고객들이 감수해야 하므로 일단은 시스템의 처리율(throughput)을 높이는 것이 급선무이다. 따라서 본 절에서는 먼저 상당수의 고객들이 대기중인 정체 상황에서의 평균 처리율에 초점을 맞추어 복수창구 요금소 모형을 분석하겠다. 그리고 그 결과 및 직렬-서어버 모형의 분석 결과를 바탕으로 복수창구 요금소 모형과 단일창구 요금소 모형의 안정성 및 성능을 비교한다.

본 논문에서는 처리율을 단위 시간당 시스템을 떠난 고객수로 정의한다. 상당수의 고객들이 대기중인 정체 상황에서의 복수창구 요금소 모형의 평균 처리율 r_K는 다음과 같이 표현된다.

$$r_K = \frac{(\mu_1+\mu_2)[2p-p^2-p^{K+2}]}{1-(1-p)p^{K+1}} \tag{27}$$

여기에서 $p=\mu_1/(\mu_1+\mu_2)$이다.

(증명)

서어버가 바쁠 때(busy)의 처리율이 도착률보다 커야 시스템이 안정하므로 [4]의 237쪽 식 (5.3.5)의 우변이 처리율임을 쉽게 알 수 있다. 아울러, 단위시간당 시스템을 떠나는 고객의 비율을 직접 구해도 식 (27)을 얻는다. (부록 1 참조)

복수창구 요금소 모형에서는 두 서어버의 서비스율이 다를 수 있다. 두 서어버의 서비스율이 다른 경우 서비스율이 큰 서어버와 작은 서어버의 배치 상황에 따른 처리율은 〈표 2〉와 같다.

중간 대기장소가 없는 경우에는 처리율에 차이가 없다. 그러나 중간 대기장소가 있는 경우에는 서비스율이 큰 서어버를 요금소 1에 배치할 때 처리율이 증가한다. 그 이유는 요금소 1과 2에서 동시에 서비스가 진행중인 경우, 요금소 1에서 서비스가 끝나면 서비스를 마친 고객이 중간대기장소로 들어가고 서어버 1이 다시 서비스를 할 수 있어 서어버 1의 이용률이 증가하기 때문이다. 그런데, 〈표 2〉의 (*) 부분처럼, 두 서

〈표 2〉 $\mu_1 \neq \mu_2$일 때 처리율

	$\mu_1 > \mu_2$			$\mu_1 < \mu_2$		
K	μ_1	μ_2	처리율(r_K)	μ_1	μ_2	처리율(r_K)
0	1.5	1.0	1.578947	1.0	1.5	1.578947
0	2.0	1.0	1.714286	1.0	2.0	1.714286
1	1.5	1.0	1.822430	1.0	1.5	1.592920
1	2.0	1.0	2.086957	1.0	2.0	1.680000
1(*)	2.5	1.0	2.269625	1.0	2.5	1.733746
2	1.5	1.0	1.943958	1.0	1.5	1.597338
2	2.0	1.0	2.301370	1.0	2.0	1.670886
2	2.5	1.0	2.570897	1.0	2.5	1.719610
2(*)	3.0	1.0	2.777293	1.0	3.0	1.754941
3	1.5	1.0	2.009787	1.0	1.5	1.598960
3	2.0	1.0	2.431718	1.0	2.0	1.668050
3	2.5	1.0	2.769493	1.0	2.5	1.715789
3	3.0	1.0	3.041357	1.0	3.0	1.751224
3(*)	3.5	1.0	3.262245	1.0	3.5	1.778713

〈표 3〉 $\mu_1=\mu_2=\mu$일 때 복수창구 요금소 모형의 처리율

K	0	1	2	...	∞
처리율(r_K)	$(4/3)\mu$	$(10/7)\mu$	$(22/15)\mu$	...	1.5μ

어버의 서비스율의 차이에 비해 중간 대기장소의 수가 상대적으로 적으면, 서비스율이 큰 서어버가 혼자 서비스를 하는 단일창구 요금소 모형보다 처리율이 오히려 작아지며 이 경우에는 복수창구 요금소 모형의 의미가 없게 된다. 그 이유는 요금소 2에서의 서비스가 늦어지면 요금소 1에서 서비스를 받은 고객들이 중간 대기장소에 점점 쌓이게 되고 그것이 정체를 유발시키기 때문이다. 실제로 고속도로 톨게이트에서 중간 대기장소를 여러개 두기가 곤란하므로 두 서어버의 서비스율의 차이가 작은 경우에야 비로소 복수창구 요금소가 의미 있다. 두 서어버의 서비스율이 $\mu_1=\mu_2=\mu$로 같을 때의 평균처리율은 〈표 3〉과 같다.

r_K는 $r_0=(4/3)\mu$ 인 증가함수이므로 서비스율이 μ인 서어버가 혼자 서비스를 하는 단일창구 요금소에 비해 언제나 효율적이다. 그러나 중간 대기장소를 1개씩 늘일 때마다 증가하는 처리율의 양은 0.1μ과 0.04μ로 미미한 편이다.

따라서 지금까지의 수치결과를 종합해 보면 실제 고속도로 톨게이트에서는 두 서어버의 서비스율이 같고 중간 대기장소가 없는 복수창구 요금소모형을 이용하는 게 적절하다.

중간 대기장소가 없고 두 서어버의 서비스율이 같은 경우 대표적으로 $\mu_1=\mu_2=1$일 때 복수창구 요금소 모형과 단일창구 요금소 모형의 평균 체류시간은 〈표 4〉와 같다.

〈표 4〉 $\mu_1=\mu_2=1$일 때 평균 체류시간

λ	$E[T]$	$E[T_1]$	$E[T]/E[T_1]$
0.50	1.458650	2.000000	0.729325
0.60	1.630824	2.500000	0.652330
0.70	1.857854	3.333333	0.557356
0.80	2.170344	5.000000	0.434069
0.90	2.627239	10.000000	0.262724
1.00	3.358307		
1.05	3.917321		
1.10	4.715853		
1.15	5.949853		
1.20	8.109196		
1.25	12.859463		
1.30	31.859712		

여기에서 T_1은 단일창구 요금소 모형의 평균 체류시간이다. 서비스 시간이 지수분포를 따르므로 단일창구 요금소 모형은 M/M/1 대기행렬 모형이며 $E[T_1]=1/(\mu-\lambda)$이다. 그리고 $E[T]$는 (23)에서 얻었다.

두 모형의 체류시간의 비율인 $E[T]/E[T_1]$를 살펴보면 교통 수요가 증가할수록 복수창구 요금소 모형의 평균 체류시간이 단일창구 모형에 비해 현저히 감소한다. 이론적으로 $\mu_1=\mu_2=1$인 경우 복수창구 모형은 도착률이 4/3 이하일 때 그리고 단일창구 요금소 모형은 도착률이 1.0 이하일 때 시스템이 안정하다. 그러므로 복수창구 요금소 모형은 단일창구 요금소 모형에 비해

시스템의 성능 및 안정성이 우수하다.

7. 결 론

본 논문에서는 서어버가 직렬로 배치된 M/M/2 직렬-서어버 모형을 분석하고 그 응용으로 복수창구 요금소 모형을 다루었다.

행렬기하 방법을 이용하여 M/M/2 직렬-서어버 모형의 대기장소내 고객수와 서어버의 상태에 대한 결합 확률을 구하였다. 이 결과를 이용하여 시스템내 고객수, 대기장소내 고객수, 대기시간, 그리고 체류시간의 1차, 2차 모멘트를 간단한 형태로 나타내고 평균 대기시간과 평균 체류시간의 해석을 제공하였다. 아울러, 무작위화 방법을 이용하여 대기시간의 분포함수를 계산하는 알고리즘을 제시하였다.

복수창구 요금소 모형에 대해서는 먼저 고객들이 상당수 대기중인 정체 상황에서의 처리율을 구하고 그 결과를 이용하여 서비스율의 다른 서어버의 배치 및 적절한 중간 대기장소의 수에 대해 토의하였다. 그리고 과부하인 경우 복수창구 요금소 모형과 단일창구 요금소 모형의 체류시간을 비교하였다. 수치분석 결과 복수창구 요금소 모형은 단일창구 요금소 모형에 비해 시스템의 성능 및 안정성이 우수한 것으로 판명되었다.

감사의 글

본 논문의 주제를 제시해주신 KAIST 산업공학과의 황학 교수님과 많은 조언을 주신 심사위원께 감사를 드립니다.

참 고 문 헌

[1] 조면식, 박윤선, "고속도로 매표방법 개선에 관한 연구", 「한국시뮬레이션 학회 논문지」, 제3권, 제1호(1994), pp. 99-105.

[2] 조면식, 전성하, "고속도로 요금소 설치 개선에 관한 연구", 「산업공학」, 제8권, 제4호(1995), pp. 109-120.

[3] Gross, D. and D. R. Miller, "The Randomization Technique as a Modeling Tool and Solution Procedure for Transient Markov Process", *Operations Research,* Vol. 32, No. 2(1984), pp. 343-361.

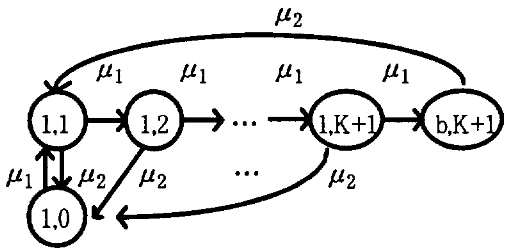
[4] Neuts, M. F., ***Matrix-Geometric Solutions in Stochastic Models : An Algorithmic Approach,*** The Johns Hopkins University Press, Baltimore, 1981.

[5] Wolff, R. W., "Poisson Arrivals See Time Averages", ***Operations Research,*** Vol. 30 (1982), pp. 223-231.

[6] Wolff, R. W., ***Stochastic modeling and the theory of queues,*** Prentice Hall, Englewood Cliffs, New Jersey, 1989.

부록 : 식 (27)의 증명

정체 상태에서 시스템 상태를 (n_1, n_2)로 정의한다. $n_1 = 1, b$로 1은 요금소 1이 서비스 상태이고 b는 차단상태를 뜻한다. n_2는 중간 대기장소와 요금소 2의 고객수이다. $n_2 = 0, 1, \cdots, K+1$이다. (n_1, n_2)로 정의된 연속적 마코프사슬 (continuous time Markov chain)의 상태전이는 〈그림 A1〉과 같다.



〈그림 A1〉 상태전이도

〈그림 A1〉에서 마코프사슬은 언제나 상태 (1,1)에서 시작해서 다시 상태 (1,1)로 돌아오는 재생싸이클(regeneration cycle)이 된다. 상태 (1,1)에서 출발해서 다시 상태 (1,1)로 돌아올 때까지의 시간을 T_{11}이라 정의한다. 그리고 T_{11} 동안 서비스를 받고 시스템을 떠난 고객수를 N_{11}이라 표기한다. 그러면

$$r_K = E[N_{11}] / E[T_{11}] \qquad \text{(A1)}$$

이다. 마코프사슬이 (1,1)을 출발해서 다시 (1,1)로 돌아올 때까지 어떤 경로를 따라 움직였는가에 조건을 걸면

$$E[N_{11}] = p^K(K+2) + \sum_{j=0}^{K-1}(j+2)(1-p)p^j \qquad \text{(A2)}$$

이다. 여기에서 $p = \mu_1 / (\mu_1 + \mu_2)$이다. π_{11}을 상태 (1,1)에 있을 정상상태 확률이라면 [6]의 식 (21)로 부터

$$1/E[T_{11}] = (\mu_1 + \mu_2)\pi_{11} \qquad \text{(A3)}$$

이다. 정상상태 확률을 계산하면

$\pi_{11} = \dfrac{p(1-p)}{1-(1-p)p^{K+1}}$이고 π_{11}을 (A3)에 대입하고 (A3)과 (A2)를 정리해서 (A1)에 대입하면 (27)을 얻는다.