Journal of the KSQC
Vol. 19, No. 2, Dec.1991

1991년 12월 韓國品質管理學會誌 제19권 제2호

시리즈로 구성된 큐잉망의 최적 순서에 관한 연구
A Study on the Optimal Order of Queueing Networks in Series

조 한벽*, 김 재련**
Cho, Han Byeog*, Kim, Jae Yearn**

In this paper, the queueing system in series is studied. The system is a tandem queueing system which has three stations. In system, one service station has a general distributions and two service stations have exponential distribution. Each station has a single server. The customer arrives with Poisson process and is serviced sequentially. It is assumed that the order of stations does not affect the quality of services. Using the light traffic approximations, the optimal order of the system which has the three stations is decided.

I. 서론

최근 시스템이 복잡해짐에 따라 여러가지 설계 및 분석문제가 제기되고 있는데 시스템의 성능을 예측하고 향상시키는 기법으로 대기이론이 널리 이용되고 있다. 큐잉시스템에 대한 중요한 설계문제중의 하나는 직렬로 구성된 다수의 서비스 창구의 최적 순서를 결정하는 것이다. 예를 들어 이 설계 문제는 생산라인을 구축하는 경우에 발생할 수 있는데, 어떤 시스템에서는 고객이 받는 서비스의 순서가 고정되어야 하지만, 다른 경우에는 서비스의 순서가

* 한양대학교 산업공학과 대학원
** 한양대학교 산업공학과 교수

고객이 시스템에서 전체적으로 받는 총(total) 서비스에 영향을 주지 않는다. 이 경우 총 작업(job)은 어떤 순서로도 수행될 수 있는 분리 과업으로 나누어질 수 있다. 그런 다음 목표는 모든 작업이 종료되기까지 지연이 최소화 되는 순서를 미리 결정하는 것이다.

이 설계문제에 대해선 알려진 것이 많지 않다. 일반적으로 탄뎀큐잉 시스템의 정확한 분석은 어려운데 그 이유는 첫째 서버에 도착과정이 포아슨일지라도 다음 서비스 창구에의 도착과정이 서비스시간 분포가 지수분포가 아닌한 포아슨 과정이 아니기 때문이다. 실제로 두번째(그리고 후속)서버에 도착과정은 일반적으로 재생과정조차 아니다. 이것은 두번째 서버에서 지연을 분석하기가 매우 어렵다는 것을 의미한다. [4] 또한 비교적 간단한 큐잉시스템에서도 서비스창구에 다수의 서버를 포함하는 경우에는 문제가 너무 복잡하여 정확하게 풀지 못하는 경우가 있다. [2]

본 논문에서는 탄뎀으로 구성된 두개 이상의 서비스 창구가 있는 시스템을 고려한다. 모든 고객은 동일한 창구순서에 따라 서비스를 받으며, 선입선출정책에 의하여 서비스를 받는다. 각 고객 (또는 작업)은 각 창구에서 한번 서비스를 받는다. 고객은 포아슨 분포를 따르는 도착률로 첫째 창구에 도착한다고 가정한다. 각 창구의 대기창구는 무한버퍼라고 가정한다. 창구들은 상호 독립적으로 서비스하고 도착과정에 독립이다.

본 논문에서도 서비스 순서가 최종 품질에 영향을 미치지 않는 경우에 주어진 외부도착과정과 서비스시간 분포에 대하여, 안정상태에서 고개당 시스템내 평균 기대체류(지연)시간을 최소화하는 창구의 순서를 결정하는 문제를 다룬다.

본 논문의 나머지는 다음과 같이 구성되었다. 제2장에서 세개의 단일서버 창구가 직렬로 구성된 탄뎀 큐잉시스템 모형을 고찰한다. 제3장에서 순서를 결정하는 해법을 제시하고, 제4장에서 결론을 기술하는 것으로 맺는다.

Ⅱ. 모형설명

본 논문에서 다룰 내용은 서비스 창구 세개가 직렬로 연결된 탄뎀시스템으로서 모든 서비스창구는 무한 버퍼를 가지고 각 창구는 단일서버를 포함한다. 각 창구에서 서비스 시간의 분포는 일반분포를 갖는 창구가 한개, 지수분포를 갖는 창구가 두개로 구성되는데 그림 1과 같이 표현할 수 있다.

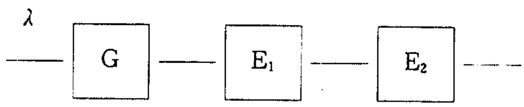

그림 1 하나의 일반분포와 두개의 지수분포를 갖는 세창구 탄뎀시스템

일반서비스 시간을 갖는 창구는 G, 지수서비스 시간을 갖는 창구는 E_1, E_2로 표시한다. 고려할 내용은 이 세 창구의 순서를 결정하는 것인데, 평균기대지연시간을 최소화하는 배열을 찾는데 목표를 둔다. 지수서비스 시간 분포의 확률변수 대소에 따라 두가지 경우를 다룬다.

본 논문에서 사용되는 용어의 정의는 다음과 같다.

X　: G의 서비스 시간의 확률 변수
Y　: E_1의 서비스 시간의 확률 변수
Z　: E_2의 서비스 시간의 확률 변수

Xe : G 에서 남은 서비스 시간
Ye : E_1에서 남은 서비스 시간
Ze : E_2에서 남은 서비스 시간
Yx : G 에서 서비스 중인 고객의 E_1에서 서비스 시간
Zx : G에서 서비스 중인 고객의 E_2에서 서비스 시간
Zy : E_1에서 서비스 중인 고객의 E_2에서 서비스 시간
λ : 도착률
1/K : 창구 X의 평균 서비스 시간
μ_1 : E_1의 서비스율
μ_2 : E_2의 서비스율
d_1 : 첫번째 창구에서의 기대지연
d_2 : 두번째 창구에서의 기대지연
d_3 : 세번째 창구에서의 기대지연

Ⅲ. 해 법

세개의 서비스 창구로 구성된 시스템에 있어 고객의 평균기대지연을 최소화하는 순서를 다음 절차에 의해서 결정한다.

1) G와 E의 순서를 결정한다.
2) GEE와 EGE를 비교하여 최적 순서를 결정한다.

이 시스템에서 고려할 수 있는 시스템은 모두 여섯가지이다. 먼저 지수분포와 일반분포의 순서를 결정한다. 이 비교는 두 지수분포에 대해 적용된다. 그런데 연이은 지수분포를 갖는 창구의 순서는 바뀌어도 출력분포에 영향을 미치지 않으므로 G E_1 E_2는 G E_1 E_2와 같은 결과를 산출한다. 마찬가지로 E_1 E_2 G와 E_2 E_1 G도 같은 결과가 된다. [3]

1. G와 E의 순서결정

포아손도착과 두 창구로 구성된 서비스 시스템에 대한 서버의 최적순서를 선택한다. 한 서버는 지수서비스시간분포, 다른 서버는 일반서비스시간 분포를 갖는다고하자. 모든 서비스시간은 독립적으로 선택한다.

정리 : 경미(light) 트래픽(traffic)에 대해 지수분포 앞에 일반 서비스 시간분포 G의 서버를 놓는 것으로 두 서버 탄뎀큐의 평균기대지연이 최소화되기 위한 필요충분조건은

$$L_X(\mu) \leq \frac{K}{K+\mu} \text{이다.}$$

여기서 $L_X(\mu)=E[\exp(-\mu X)]$는 일반분포의 Laplace변환이고 $\frac{K}{K+\mu}$는 일반 서비스분포와 같은 평균을 갖는 지수분포의 Laplace변환이다.

이 정리[1]가 의미하는 것은 경미 트래픽에 대해, 평균 1/K인 지수분포의 Laplace변환과 일반분포의 Laplace변환을 비교하여 적은 값을 갖는 분포를 먼저 서비스하므로써 탄뎀큐잉시스템에 대한 최적 순서를 결정할 수 있다는 것이다.

2. 최적 순서 결정

이제 GEE가 EGE보다 기대지연이 작다는 것을 보임으로써 확률적으로 고객의 평균 기대지연을 최소화하는 서비스 창구의 순서를 결정하게 된다. 시스템 GEE를 A라 하고 시스템 EGE를 B라고 하자. 각 시스템의 기대지연은 다음과 같다. [5]

$d_{A1} = 1/2\lambda\, E[X^2]$

$d_{B1} = \lambda/\mu_1^2$

$d_{A2} = \lambda\, E[X]\, E[Y_x - X_a] + \lambda/\mu_1\, E[Y_e - X_a]$

$d_{B2} = 1/2\lambda\, E[X^2]$

$d_{AT} = 1/2\lambda\, E[X^2] + \lambda\, E[X]\, E[\max(Y_x + Z_x - X_a - Y_a)^+, (Y_x - X_a)^+] + \lambda/\mu_1\, E[\max(Y_e + Z_x - X_a - Y_a)^+, (Y_e - X_a)^+] + \lambda/\mu_2\, E[Z_e - X_a - Y_a]^+ + O(\rho^2)$

$d_{BT} = 1/2\lambda\, E[X^2] + \lambda/\mu_1^2 + \lambda\, E[X]\, E[Z_x - X_a]^+ + \lambda/\mu_2\, E[Z_e - X_a]^+$

Y와 Z는 각각 평균 $1/\mu_1$과 $1/\mu_2$인 지수분포를 갖는 확률변수이다.

먼저 창구 Y에서 서비스 시간의 확률변수값이 창구 Z에서의 확률변수값과 비교해서 동일하거나 큰 경우를 살펴보자.

1) Y ≥ Z인 경우

$d_{AT} - d_{BT} = \lambda\, E[X]\, E[\max(Y_x + Z_x - X_a - Y_a)^+, (Y_a - X_a)^+] + \lambda/\mu_1\, E[MAX(Y_e + Z_x - X_a - Y_a)^+, (Y_e - X_a)^+] + \lambda/\mu_2\, E[Z_e - X_a - Y_a]^+ - \lambda/\mu^2 - \lambda\, E[X]\, E[Z_x - X_a]^+ - \lambda/\mu_1\, E[Z_e - X_a]^+$

$= \lambda\, E[X]\, E[Y_x - X_a]^+ + \lambda/\mu_1\, E[Y_e - X_a]^+ + \lambda/\mu_2\, E[Z_e - X_a - Y_a]^+ - \lambda/\mu_1^2 - \lambda\, E[X]\, E[Z_x - X_a]^+ - \lambda/\mu_1\, E[Z_e - X_a]^+$

$= \lambda/\mu_1\, E[Y_e - X_a]^+ + \lambda/\mu_2\, E[Z_e - X_a - Y_a]^+ - \lambda/\mu_1^2 - \lambda/\mu_1\, E[Y_e - X_a]^+ \le 0$

여기서 $\lambda\, E[X]\, E[Y_x - X_a]^+ + \lambda/\mu_1\, E[Y_e - X_a]^+ \le \lambda/\mu_1^2$ 이다.

따라서 Y ≥ Z인 경우에는 GEE가 EGE보다 기대지연시간이 더 작다.

2) Y < Z인 경우

$d_{AT} - d_{BT} = \lambda\, E[X]\, E[Y_x + Z_x - X_a - Y_a]^+ + \lambda/\mu_1\, E[Z_x - X_a]^+ + \lambda/\mu_2\, E[Z_e - X_a - Y_a]^+ - \lambda\, E[X]\, E[Z_x - X_a] - \frac{\lambda}{\mu_2} E[Z_e - X_a]^+ - \frac{\lambda}{\mu_1^2} = \lambda E[X] E[Z_a - X_a]^+ + \lambda/\mu_1\, E[Z_x - X_a]^+ + \lambda/\mu_2\, E[Z_e - X_a - Y_a]^+ - \lambda/\mu_1^2 - \lambda\, E[X]\, E[Z_x - X_a]^+ - \lambda/\mu_2\, E[Z_e - X_a]^+ \le 0$

Y<Z인 경우에도 GEE가 EGE보다 기대지연시간이 더 작다.

따라서 G가 E앞에 온다고 결정된다면 GEE가 최적순서임을 알 수 있다.

4. 결　론

본 논문에서는 기존에 연구된 두창구 시스템의 순서결정문제를 확장하여 직렬로 연결된 세 창구의 탄뎀시스템에 대하여 평균 기대지연을 최소화 시키는 창구의 순서를 결정하는 문제에 대하여 고찰하였다. 이 시스템은 하나의 일반 서비스 시간분포를 갖는 창구와 두개의 지수서비스 시간분포를 갖는 창구로 구성된다. 창구의 배열을 결정하기 위하여 일반 분포의 라프라스 변환과 이 분포와 같은 평균을 갖는 지수분포의 라프라스 변환을 비교하여 순서를 결정하고, 다른 지수 서비스시간 분포를 갖는 창구를 더한 각 시스템의 기대지연시간을 비교하여 최소값을 갖는 배열을 선택하였다. 나타난 결과는 지수분포 앞에 일반 서비스 분포를 갖는 창구를 놓는 것이 기대지연을 작게한다면 일반 서비스 분포를 갖는 창구를 가장 앞에 놓는 것이 좋다는 결론을 얻었다.

참고문헌

1. Greenbreg B. S. and R. W. Wolff, Optimal Order of Servers for Tandem Queues in Light Traffic, Management Science, Vol.34, No.4, April 1988, pp.500−508.
2. Letonen T., On the Ordering of Tandem Queues with Exponential Servers, J. of Applied Probability, Vol.23, No.1, Jan.1986, pp. 115−129.
3. Weber R. R., The Interchangeability of Tandem */M/1 Queues in Series, J. of Applied Probability, Vol.16, 1979, pp.690−695.
4. Whitt W., The best Order for Queues in Series, Management Science, Vol.31, No.4, 1985, pp.475−478.
5. Wolff R. W., Tandem Queues with Dependent Service in Light Traffic, Operations Research, Vol.30, 1987, pp.619−635.