세개의 창구로 구성된 큐잉시스템의 최적순서에 관한 연구
-A Study on the Optimal Order of Queueing System with three Stations-

조 한 벽*
김 재 련**

Abstract

The one of the important problems in the design of queueing systems is the decision of the order of service stations. The object of this problem is the decision of the order that minimizes the expected sojourn time per customer in the given arrival process and service time distributions. In this paper, the tandem queueing system in series is studied with the emphasis on the optimal order of the tandem queueing system which has three stations with single servers.

In one system, customers arrive at the first station with Poisson process. This system is composed of service stations with a costant, a general distribution and a Exponential distribution is studied. To select the optimal order, after the orders of each pair of two stations is decided, it is compared the two orders of system. With this results, total expected delay of the systems which has three stations is compared. The result is the best that service station with constant time is on the first place, then the service station with general distribution and the service station with Exponential distribution is followed.

And the other system is consist of service stations with a costant and two probabilistic distributions. In this case, two probabilistic distributions has a non-overlapping feature. It is the optimal order that the service station with constant time is on the first place, then the service station with longer service time and the service station with shorter service time is followed.

1. 서 론

1.1 연구의 목적

최근 시스템이 복잡해짐에 따라 여러가지 설계 및 분석문제가 제기되고 있는데 시스템의 성능을 예측하고 향상시키는 기법으로 대기이론이 널리 이용되고 있다. 그 중에서도 탄뎀 큐잉시스템(tandem queueing system)은 통신 및 생산시스템을 모형화하는데 사용되기 때문에 중요하다. 그 중에서 대기이론에 관하여 알려진 결과의 대부분은 도착과정이 재생과정인 큐잉시스템에 관한 것이다[4]. 본 논문에서 다루는 서비스창구 순서의 설계문제에 대해서는 알려진 결과가 많지 않다. 이것은 두번째 서비스 창구에서 지연을 분석하기가 매우 어렵기 때문이며[1], 상대적으로 다수의 서비창구를 포함하는 간단한 큐잉시스템에서도 너무 복잡하여 정확하게 풀지 못하는 경우가 있다[3].

본 논문에서도 서비스 순서가 최종 품질에 영향을 미치지 않는다고 간주하고 주어진 외부 도착과정과 서비스 시간 분포에 대하여, 안정상태에서의 고객당 총 기대지연을 최소화하는 서비스창구의 순서를 결정하는 문제를 다룬다.

1.2 기존연구 고찰

큐잉시스템의 분석 분야중에서 서비스창구의 순서를 결정하는 문제는 출력과정의 어려움 때문에 많이 연구

*한양대학교 산업공학과 대학원
**한양대학교 산업공학과 교수
접수 : 1991. 8. 11.

되지는 않았다. 따라서 자연스럽게 근사법으로 해결하거나 한계치를 계산하는 해법이 개발되어 왔다[4].

두가지 특수한 경우에 대해서는 마지막 이탈과정이 순서에 무관하다는 것이 증명되었다. 임의의 도착과정과 모든 서비스창구에서 서비스 시간이 상수인 경우는 Friedman[6]이 증명하였고, 모든 서비스창구에서 독립적 지수분포인 경우는 Weber[6]와 Lehtonen[3]이 증명하였다. 결과적으로 이 상황에서 전체 시스템에서 체류시간의 분포는 서비스창구의 순서와 독립이다. 그러나 이 두 경우는 혼합되지 못한다. 즉 상수와 지수분포의 복합형태의 탄뎀 큐잉시스템에 대해서는 결과가 순서에 무관하다는 것이 성립하지 않는다.

Tembe와 Wolff[5]는 서비스창구가 두 개인 시스템에서 두 서비스창구중 하나는 상수이고 다른 하나는 확률적 서비스시간을 갖는 경우를 고려하였다. 각 고객의 지연은 결정적 서비스를 먼저 수행할 때 확률적으로 최소화 됨을 보였다.

또한 서비스시간분포가 비교차(nonoverlapping)되는 경우[$Pr(S_A \geq S_B)=1$]에는 먼저 느린 서비스창구에서 빠른 서비스창구의 순서로 배열한다면 고객의 지연은 확률적으로 최소가 된다고 증명하였다.

Niu[4]는 임의의 도착과정과 서비스 분포에 대해 첫째 서비스창구에서의 상수서비스시간은 둘째 서비스창구에서 기대지연을 감소시키고, 이 감소는 상수서비스 시간의 크기에 따라 단조적(monotone)이라고 추측하였다. Greenberg[1]는 경비트래픽 근사법과 과중트래픽 근사법을 이용하여 Niu의 추측을 뒷받침하는 증거를 제공하였다.

1.3 연구의 범위와 구성

본 연구에서는 다수의 서비스창구가 시리즈로 구성된 탄뎀 큐잉시스템을 다룬다. 우리의 목표는 시스템에서의 평형기대 체류시간이나 모든 서비스창구에서 서비스를 받기 위하여 대기하는 기대지연을 합한 총기대지연시간을 최소화하는 서비스창구의 배열순서를 찾는 것이다. 세개의 서비스창구로 구성된 시스템의 순서를 다룬다. 여기서는 지수분포를 포함하는 서비스창구수에 따라 시스템을 분류하여 각 경우에 대하여 살펴보았다.

제2장에서는 모형에 대한 설명을 제시한다.

제3장에서는 세개의 단일서버 서비스창구가 시리즈로 구성된 탄뎀 큐잉시스템 모델을 고찰한다.

지수서비스와 상수서비스와 일반서비스시간분포로 구성된 경우를 살펴본다.

다음으로 모든 서비스창구가 상수서비스시간을 갖는 경우와 서비스 시간이 교차되지 않는(Non-overlapping) 경우를 살펴본다. 제4장에서는 결론과 향후 연구되어야 할 과제에 대하여 기술하였다.

2. 모형설정

본 논문에서 다룰 모형은 세개의 서비스창구가 시리즈로 구성된 탄뎀 큐잉시스템에서 서비스창구의 순서를 결정하는 것인데 총 기대지연시간을 최소화하는 배열을 찾아 시스템을 이탈하는 시점이 가장 빠른 시스템의 배열순서를 찾는 것이다.

본 논문에 적용되는 가정은 다음과 같다.

- 모든 고객은 시스템의 각 서비스창구에서 동일한 순서로 서비스를 받는다.
- 시스템에 도착하는 입력과정은 도착률 λ의 포아손과정이다.
- 각 서비스창구에서 서비스는 상호 독립적이며 도착과정에도 독립적이다.
- 각 서비스창구의 대기창구는 무한이다.
- 서비스 정책은 선입선출이다.

본 논문에서 사용되는 용어의 정의는 다음과 같다.

X : 창구 X의 서비스시간의 확률 변수
Y : 창구 Y의 서비스시간의 확률 변수
Z : 창구 Z의 서비스시간의 확률 변수
X_e : 창구 X에서 남은 서비스시간
Y_e : 창구 Y에서 남은 서비스시간
Z_e : 창구 Z에서 남은 서비스시간

Y_x : 창구 X에서 서비스 중인 고객의 창구 Y에서 서비스시간
Z_x : 창구 X에서 서비스 중인 고객의 창구 Z에서 서비스시간
Z_y : 창구 Y에서 서비스 중인 고객의 창구 Z에서 서비스시간
G : 일반서비스 시간분포를 갖는 서비스창구
E : 지수서비스 시간분포를 갖는 서비스창구
D : 상수서비스 시간을 갖는 서비스창구
ρ_x : 창구 X의 이용율, $\lambda E[X]$
ρ_y : 창구 Y의 이용율, $\lambda E[Y]$
ρ_z : 창구 Z의 이용율, $\lambda E[Z]$
μ_x : 창구 X의 서비스율
μ_y : 창구 Y의 서비스율
μ_z : 창구 Z의 서비스율
λ : 시스템 도착률
d_{AT} : 시스템 A의 총기대 지연
d_{BT} : 시스템 B의 총기대 지연
S^i : i번째 창구의 서비스시간

3. 세개 창구의 순서 결정

3.1 개 요

본 논문에서 다룰 내용은 서비스 창구 세개가 연속적으로 연결된 탄뎀 큐잉시스템으로서 모든 큐는 무한 버퍼를 가지고 각 창구는 단일서버를 포함한다.

고려할 내용은 이 세개 서비스창구의 순서를 결정하는 것인데, 총 기대지연시간을 최소화하는 배열을 찾는데 목표를 둔다.

서비스창구 세개가 연속적으로 연결된 탄뎀 큐잉시스템의 구성은 그림 3-1 과 같이 표현할 수 있다.

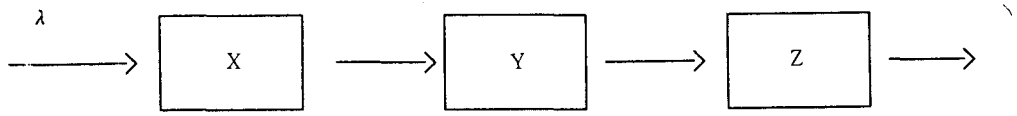

그림 3-1. 탄뎀 큐잉 시스템

본 장에서 고려할 시스템은 두가지로서 지수서비스와 상수서비스와 일반서비스의 경우와 지수서비스를 포함하지 않는 경우로 분류하였다.

3.2 일반분포와 지수분포와 상수로 구성된 시스템

이 절에서는 고려하는 시스템은 각 창구에서 서비스시간의 분포는 일반분포, 지수분포, 상수로 구성된다. 일반서비스 시간을 갖는 창구는 G, 지수서비스 시간을 갖는 창구는 E, 상수서비스는 D로 각각 표시하기로 한다. 고객은 포아슨 분포를 따르는 도착률을 가지고 첫째 창구에 도착한다고 가정한다. 각 창구에서 고객은 선입선출서비스 정책에 의하여 서비스를 받는다. 고려할 내용은 세개의 서비스 창구로 구성된 시스템에 있어 고객의 총 기대지연을 최소화하는 배열을 찾는데 목표를 둔다. 배열순서는 다음 절차에 의해서 결정된다. 먼저 D와 E의 순서와 D와 G의 순서를 결정하고[5], G와 E의 순서를 결정한다[2]. 다음 DGE와 GDE를 비교한다.

3.2.1 D와 E의 순서 및 D와 G의 순서결정

이 시스템에서 고려할 수 있는 가능한 배열은 모두 6가지이다. 먼저 지수분포와 상수서비스의 순서를 결정한다. 이 비교는 항상 상수서비스가 앞에 오는 것이 좋다. 순서가 문제가 되지 않을 경우는 상수서비스 시간이 지수서비스시간의 평균값보다 작은 경우이다.

다음에 상수서비스시간과 일반서비스시간을 갖는 창구의 순서를 결정한다. 이 비교에서도 항상 상수서비스가 앞에 오는 것이 좋다. 상수서비스는 출력과정을 정규화시킴으로써 다음 창구의 기대지연을 작게 하는 효과가 있다[4]. 따라서 확률분포 앞에 상수서비스를 놓는 것이 기대지연을 줄인다.

3.2.2 G와 E의 순서결정

이제 지수분포와 일반분포서비스 창구의 순서를 고려한다. 이 비교는 제2장에서 살펴본 바와 같이 일반분포의 라프라스 변환과 동일 평균을 갖는 지수분포의 라프라스변환을 비교하여 정할 수 있다. 일반분포를 앞에 놓는 것이 좋은 경우는 이 변환 값이 더 작은 경우이다. 마지막으로 일반서비스 분포를 갖는 창구를 지수서비스 앞에 놓는 것이 좋다는 가정하에서 최적 순서를 결정한다.

3.2.3 최적 순서 결정

이제 DGE. GDE와 GED를 비교하여 최적 순서를 결정하게 되는데 GED는 마지막 서비스 창구가 상수서비스가 있기 때문에 기대지연이 가장 크게 되어 비교대상에서 제외된다. 따라서 DGE가 GDE 보다 기대지연이 작다는 것을 보임으로써 확률적으로 고객의 총 기대지연을 최소화하는 서비스 창구의 순서를 결정하게 된다. 시스템 DGE를 A라 하고 시스템 GDE를 B라고 한다. 경비트래픽하에서 각 시스템의 기대지연은 다음과 같다.

$$d_{AT}=1/2\lambda d^2+\lambda dE[MAX(Y_s+Z_s-d-Y_a)^+,(Y_s-d)^+]+\lambda E[Y]E[MAX(Y_c+Z_s-d-Y_a)^+,(Y_c-d)^+]+\lambda E[Z]E[Z_c-d-Y_a]^++0(\rho^2) \quad (3.1)$$

$$d_{BT}=1/2\lambda E[Y^2]+\lambda E[Y]E[MAX(d+Z_s-Y_a-d)^+,(d-Y_a)^+]+\lambda dE[MAX(X_c+Z_s-Y_a-d)^+,(X_c-Y_a)^+]+\lambda E[Z]E[Z_c-Y_a-d]^++0(\rho^2) \quad (3.2)$$

Y는 평균 $2*1/\mu$인 얼랑분포를 갖는 확률변수이고, Z는 평균 $1/\mu$인 지수분포를 갖는 확률변수이다. 여기서 각 서비스창구의 기대 서비스시간은 얼랑서비스 평균시간이 상수서비스시간과 동일한 경우를 살펴보았다.

$$\begin{aligned}d_{AT}-d_{BT}=&1/2\lambda d^2+\lambda dE[MAX(Z-d)^+,(Y-d)^+]+\lambda E[Y]E[MAX(Y_c+Z-d-Y_a)^+,(Y_c-d)^+]+\lambda E[Z]E[Z_c-d-Y_a]^++0(\rho^2)\\&-1/2\lambda E[Y^2]+\lambda E[Y]E[MAX(d+Z_s-Y_a-d)^+,(d-Y_a)^+]-\lambda dE[MAX(X_c+Z_s-Y_a-d)^+,(X_c-Y_a)^+]-\lambda E[Z]E[Z_c-Y_a-d]^++0(\rho^2)\\&=1/2\lambda d^2+\lambda dE[(Y-d)^+]+\lambda E[Y]E[(Y_c-d)^+]\\&-1/2\lambda E[Y^2]+\lambda E[Y]E[(d-Y_a)^+]-\lambda dE[(X_c-Y_a)^+]\le 0\end{aligned} \quad (3.3)$$

윗 수식에 대한 상세한 계산은 부록에 첨부하였다.

따라서 G가 E 앞에 온다고 결정된다면 DGE가 최적순서임을 알 수 있다.

3.3 하나의 상수와 두개의 확률적 서비스로 구성된 시스템

이 절에서 다룰 내용도 서비스 창구 세개가 연속적으로 연결된 탄뎀시스템으로서 기본적으롱 동일한 가정하에서 내용이 전개된다. 이 절에서 고려하는 시스템은 각 창구에서 서비스 시간의 분포가, 일반 확률분포를 갖는 창구가 두개이고 상수인 창구가 하나로 구성되며, 이 시스템은 그림 3-2와 같이 표현할 수 있다. 여기서 G1의 서비스 시간은 G2의 서비스 시간보다 항상 확률적으로 큰 경우이다.

확률적 서비스 시간을 갖는 창구는 G, 상수서비스 시간을 갖는 창구는 D로 표시하기로 한다. 고려할 내용은

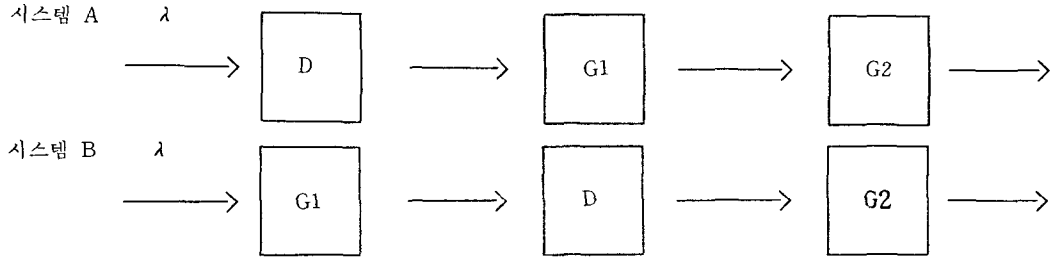

그림 3-2. 하나의 상수와 두 개의 확률적 분포로 구성된 시스템

이 세 창구의 순서를 결정하는 것인데 역시 기대지연시간을 최소화하는 것에 목표를 둔다.

이제 하나의 상수시간 서비스 창구와 두개의 확률적 서비스 창구로 구성된 탄뎀시스템을 살펴보자. 확률적 서비스 시간은 비교차($P(S_b > S_c)=1$)인 경우이다.

Tembe와 Wolff는 다음과 같이 이탈 시점을 표시하고 귀납법을 통하여 n번째 고객의 이탈시점을 구하였다. 순서에 대한 이탈시점을 비교하여 총대기지연시간이 작은 순서를 선택한다.

먼저 n번째 고객의 첫번째 창구에서의 이탈시점을 구한다.

$$T_n = \max[t_n,\ T_{n-1}] + S_n,\ (n=1,\ 2,\ \cdots\cdots)$$
$$= \max_{1 \le i \le n} [t_i + \sum_{j=1}^{i-n} S_j]\ (n=1,\ 2,\ \cdots\cdots) \tag{3.4}$$

n번째 고객의 첫번째 서비스창구의 이탈시점을 두번째 서비스 창구의 도착시점으로 하여 두번째 창구에 대한 이탈시점을 구한다.

$$T_{2n} = \max_{1 \le k \le n} [\ \max_{1 \le i \le k} (t_i + \sum_{k=1}^{k-1} S_k^a) + \sum_{k=1}^{k-n} S_k^b] \tag{3.5}$$

$$= \max_{1 \le i \le n} [t_i + \max_{i \le k \le n} (\sum_{k=1}^{k-1} S_k^a + \sum_{k=1}^{k-n} S_k^b)] \tag{3.6}$$

n번째 고객의 두번째 서비스창구의 이탈시점을 세번째 서비스 창구의 도착시점으로 하여 세번째 창구에 대한 이탈시점을 구한다.

$$T_{3n} = \max_{1 \le k \le n} [\ \max_{1 \le i \le k} \{ \max_{1 \le i \le k} (t_i + \sum_{l=1}^{l-1} S_k^a) + \sum_{l=1}^{l-k} S_l^b \} + \sum_{l=k}^{l-n} S_l^c]$$

$$= \max_{1 \le k \le n} [\ \max_{1 \le i \le k} \{t_i + \max_{i \le l \le k} (\sum_{l=1}^{l-1} S_l^a + \sum_{l=1}^{l-k} S_l^b)\} + \sum_{l=k}^{l-n} S_l^c]$$

$$= \max_{1 \le i \le k \le n} [t_i + \max_{i \le l \le k} \{ \sum_{l=1}^{l-1} S_l^a + \sum_{l=1}^{l-k} S_l^b\} + \sum_{l=k}^{l-n} S_l^c]$$

$$= \max_{1 \le i \le n} [t_i + \max_{i \le l \le k \le n} \{ \max_{i \le l \le k} (\sum_{l=1}^{l-1} S_l^a) + \sum_{l=1}^{l-k} S_l^b) + \sum_{l=k}^{l-n} S_l^c\}]$$

$$= \max_{1 \le i \le n} [t_i + \max_{i \le l \le k \le n} \{ \sum_{l=1}^{l-1} S_l^a) + \sum_{l=1}^{l-k} S_l^b + \sum_{l=k}^{l-n} S_l^c\}]$$

$$= \max_{1 \le i \le j \le k \le n} [t_i + \sum_{l=1}^{l-1} S_l^a + \sum_{l=1}^{l-k} S_l^b + \sum_{l=k}^{l-n} S_l^c] \tag{3.7}$$

이 식을 근거로 하여 각 경우에 대한 이탈시간을 구한다. S^a가 상수임을 고려하여 이에 대한 수식을 정리하고 일반분포의 비교차성질을 이용하여 계산을 수행한다. D G1 G2의 경우에 대한 이탈시간은

$$T_{3n}^{(1)} = \max_{1 \le i \le n} [t_i + \max_{i \le j \le k \le n} \{ \sum_{l=1}^{l-1} S_l^a + \sum_{l=1}^{l-k} S_l^b + \sum_{l=k}^{l-n} S_l^c\}]$$

$$= \max_{1 \le i \le n} [t_i + \max_{i \le j \le n} \{ \sum_{l=1}^{l-1} S_l^a + \sum_{l=1}^{l-n} S_l^b + S_n^c\}]$$

$$= \max_{1 \le i \le n} [t_i + \max_{i \le j \le n} \{(j-i+1)S^a + \sum_{l=1}^{l-n} S_l^b + S_n^c\}] \tag{3.8}$$

가 된다.

G1 D G2의 경우에 대한 이탈시간은

$$T_{3n}^{(2)} = \max_{1 \le i \le n} [t_i + \max_{i \le j \le k \le n} \{ \sum_{l=1}^{l-1} S_l^b + \sum_{l=1}^{l-k} S_l^a + \sum_{l=k}^{l-n} S_l^c\}] \tag{3.9}$$

$$= \max_{1 \le i \le n} [t_i + \max_{i \le j \le n} \{ \sum_{l=1}^{l-1} S_l^b + (n-j+1)S_l^a + S_n^c]$$

가 된다.

G1 G2 D의 경우에 대한 이탈시간은

$$T_{3n}^{(3)} = \max_{1 \le i \le n} [t_i + \max_{i \le j \le k \le n} \{ \sum_{l=1}^{l-1} S_l^b + \sum_{l=1}^{l-k} S_l^c + \sum_{l=k}^{l-n} S_l^a\}]$$

$$=\max_{1\le i\le n}\ [t_i+\max_{i\le k\le n}\ \{\sum_{l=i}^{l=k} S_l^b+S_k\}+(n-k+1)S^a] \qquad (3.10)$$

가 된다.

먼저 T_{3n}^{G2}와 T_{3n}^{D}를 비교한다. 이때에 G2와 D를 비교하여 G2가 D보다 큰 경우와 작거나 같은 경우로 나누어 생각한다.

G2가 D보다 큰 경우에는 S^c가 S^a보다 서비스시간이 긴 경우로서 순서가 문제가 되지 않는 경우이다. 이 경우 S^a가 상수이므로 지정된 k에 대하여 극대값은 k=n에서 발생한다.

G2가 D 보다 작거나 같은 경우에는 S^c가 S^a보다 서비스시간이 항상 짧은 경우로서 상수서비스 시간이 S^c와 비교차되는 경우로서 항상 상수 서비스시간을 앞에 놓는 것이 기대지연을 최소화한다. 따라서 이것은 D와 하나의 확률서비스의 순서를 결정하는 문제로 줄어든다. 그 결과 D G1 G2가 G1 D G2 보다 기대지연이 같거나 더 작게 된다.

다음은 T_{3n}^{G1}과 T_{3n}^{G2}를 비교한다. 여기서는 G1과 D를 비교하여 G1이 D보다 큰 경우와 작거나 같은 경우로 나누어 생각한다. G1이 D 보다 큰 경우에는 S^b가 S^a 보다 서비스시간이 긴 경우로서, 하나의 상수서비스와 하나의 확률적 서비스로 구성된 시스템의 순서를 결정하는 문제로 되어 순서가 기대지연에 영향을 미치지 않는 경우이다. 이 경우 S^a가 상수이므로 지정된 j에 대하여 극대값은 j=n에서 발생한다.

G1이 D 보다 작거나 같은 경우에는 S^b가 S^a 보다 서비스시간이 항상 짧아 상수서비스 시간이 S^b와 교차되지 않으므로 항상 상수 서비스시간을 앞에 놓는 것이 기대지연을 최소화한다.

따라서 이것은 D와 하나의 확률서비스의 순서를 결정하는 문제로 줄어든다. 그 결과 D G1 G2가 G1 D G2 보다 기대지연이 같거나 더 작게 된다.

그러므로 D G1 G2와 G1 D G2와 G1 G2 D에 대한 창구의 순서를 결정하는 문제는 D G1 G2 시스템이 구성되도록 D를 항상 가장 앞에 놓음으로써 기대지연을 가장 작게 하거나 최소한 같게 할 수 있다.

4. 결 론

본 논문에서 세개의 창구가 시리즈로 연결된 탄뎀 큐잉시스템에 대하여 고찰하였다. 이 시스템에서 총 기대지연을 최소로 하는 창구의 순서를 결정하는 문제를 다루었다. 지수서비스와 상수서비스 그리고 일반서비스분포를 갖는 창구로 구성된 시스템을 고찰하였다. 여기서는 일반서비스와 지수서비스의 순서를 결정하고 상수서비스를 앞에 놓음으로써 기대지연을 작게한다는 결론을 얻었다. 또한 하나의 상수서비스와 두개의 일반분포로 구성된 시스템을 고려하였다. 여기서는 일반분포가 비교차성질을 갖는 경우를 다루었다. 여기서 상수서비스를 가장 앞에 놓는 것이 기대지연을 작게한다는 결론을 도출하였다.

향후 연구될 과제는 다수의 일반적인 분포를 포함하는 시스템에 대한 순서를 결정하는 해법의 개발과 창구사이의 버퍼가 없는 경우와 유한한 경우에 최적 창구의 순서를 결정하는 문제를 고찰하는 것이다.

참 고 문 헌

[1] Greenbeng, B. S., Queueing Systems with Returning Customers and the Order of Tandem Queues, Ph. D. Thesis, University of California, Berkeley, 1986.

[2] Greenbreg B. S. and R. W. Wolff, "Optimal Order of Servers for Tandem Queues in Light Traffic," *Management Science, 34(4)*, April 1988, pp. 500-508.

[3] Lehtonen T., "On the Ordering of Tandem Queues with Exponential Servers," *J. of Applied Probability, 23(1)*, Jan. 1986, pp. 115-129.

[4] Niu Shun-Chen, "Bounds for the Expected Delays in Some Tandem Queues," *J. of Applied Probability, 17*, 1980, pp. 831-838.

[5] Tembe, S. V. and R. W. "Wolff, The Optimal Order of Service in Tandem Queues," *Operations Research, 24*, 1974, pp. 824-832.

[6] Weber R. R., "The Interchangeability of Tandem*/M/1 Queues in Series," *J. of Applied Probability, 16*, 1979, pp. 690-695.