

A Study on Times to the First Overflow in M/G/1/K/N Queueing Systems¹⁾

Kyu-Noh Lee²⁾, Hong-Gie Kim³⁾

Abstract

The main purpose of queueing theory is to find the optimal solution for maintaining systems such as service facilities. Analyzing the overflow process provides an important information for the solution in queueing systems with finite capacity.

In this thesis, we approximate the expected time until the first overflow in M/G/1/K/N queueing systems. Results will be applied to approximate the expected time until the first reduction of source population due to a bad service(long wait in the queue) in M/G/1/N queueing system. Simulation results show that our approximation is applicable to real situations.

1. 서 론

대기행렬에 관한 일반적인 이론은 Bunday(1986), Cohen(1982), Jaiswal(1968), Saaty(1961)등에 의해 연구되었으며, 유한 근원집단을 가진 M/G/1/N 대기행렬 이론은 Stecke(1982), Stecke and Aronson(1985), Takagi(1993)에 의해 연구되었다. M/G/1/N 대기행렬의 보다 일반적인 모델인 M/G/1/K/N 대기행렬은 최근 Takine, Takagi and Hasegawa(1993)에 의해 소개되었으며 앞으로 많은 연구가 예상되는 분야이다. M/G/1/K/N 대기행렬 시스템은 근원집단의 크기가 N이고, 서비스 시설의 용량은 K이며, 서비스하는 사람은 1명이고, 한 고객의 도착간격시간의 분포가 모수 λ 인 지수분포를 따르며, 서비스시간의 분포는 평균이 b인 일반적인 분포를 따르는 대기행렬 시스템이다.

시스템 용량초과 시간과정은 대기행렬 시스템에서 서비스 시설의 용량 부족으로 인해 고객의 서비스가 거부되는 시간과정으로서, M/G/1/K/N 모형에서 서비스 시설의 용량이 K이므로, 서비스 시설에 K명이 있을 때 도착하는 고객은 즉시 근원집단으로 되돌아 가게 된다. 따라서 시스템 용량초과로 인해 되돌아간 고객은 서비스 시설의 용량이 작음에 대해 불만족스러움을 느낄수 있으며, 이러한 고객이 많이 발생하게 되면 서비스 시설의 용량을 늘리거나, 서비스 시간을 단축시킬 수 있는 방법을 모색해야 하며, 이는 서비스 시설 운영비를 증가시킨다. 따라서 고객불만에 따른 손실비용과 시스템 운영비를 최적화 하기 위해 시스템 용량초과 시간과정에 대한 분석이 필요하게 된다.

Bhat(1975)는 M/G/1/K 모형에서 시스템 용량초과 시간과정을 분석하였으며, Halfin(1981)은

1) This paper is supported by Chungnam Research Foundation, 1998.

2) Senior Researcher, Agency for Defense Development, Taejon, 305-600, Korea

3) Associate Professor, Department of Statistics, Chungnam National University, Taejon, 305-764, Korea

M/G/1/1 모형에서 두 개의 연속적인 시스템 용량초과 시간의 길이인 τ_B 에 대한 분포를 유도하였으며, 그 결과에 의해 Takagi(1993)가 M/G/1/1/N 모형에서 두 개의 연속적인 시스템 용량 초과시간의 길이 τ_B 의 분포함수에 대한 LST(Laplace-Stieltjes transform)를 유도하였다.

이와 같이 근원집단의 크기가 무한인 경우에는 연구결과가 발표되어 있으며, 또한 근원집단의 크기가 유한인 M/G/1/K/N 모형에서도, K=1인 경우에는 비교적 분석이 간단하여 시스템 용량초과 시간과정에 대한 일반적인 확률분포를 유도하였으나, K가 2 이상인 경우에는 분석이 매우 복잡하여 일반적인 확률분포가 유도된바 없다. 따라서 본 논문에서는 M/G/1/K/N 모형에서 최초 시스템 용량초과 시간에 대한 평균의 근사값을 유도함으로써, K가 2 이상인 경우 시스템 용량초과 시간 분석을 통한 서비스 시설의 최적화에 기여하고자 한다.

2. M/G/1/K/N 대기행렬 시스템

Takine(1993)에 의하면 M/G/1/K/N 모형에서 서비스가 끝나는 시점의 서비스 시설에 남아 있는 고객수는 다음과 같은 확률과정으로 주어진다.

{ L_m ; $m=1,2,3,\dots$ }을 m번째 고객의 서비스가 끝난 직후 서비스 시설에 남아 있는 고객의 수라고 하면 { L_m ; $m=1,2,3,\dots$ }는 매장된 마코프 연쇄(imbedded Markov chain)가 되고, 이 마코프 연쇄의 전이 확률은 다음과 같다.

$$p_{jk} = p_{jk}^N \quad 0 \leq j \leq k+1; \quad 0 \leq k \leq K-2 \tag{2.1}$$

$$p_{jK-1} = \sum_{k=K-1}^{N-1} p_{jk}^N \quad 0 \leq j \leq K-1$$

이며, 여기서 p_{jk}^N 은 M/G/1//N 대기행렬 시스템의 전이확률로서 다음과 같이 주어진다.

$$p_{jk}^N = \begin{cases} \binom{N-1}{k} \int_0^\infty e^{-(N-k-1)\lambda x} (1 - e^{-\lambda x})^k dB(x), & j=0 \\ \binom{N-j}{k-j+1} \int_0^\infty e^{-(N-k-1)\lambda x} (1 - e^{-\lambda x})^{k-j+1} dB(x), & k \geq j-1 \geq 0 \\ 0, & j \geq 1, 0 \leq k < j-1 \end{cases} \tag{2.2}$$

여기서, $B(x)$ 는 M/G/1//N 모형에서 서비스 시간의 분포함수(distribution function)이다.

{ π_k ; $0 \leq k \leq K-1$ }을 서비스가 끝난 직후 서비스 시설에 있는 고객의 수에 대한 확률분포라고 하면, π_k 는 다음과 같다.

$$\pi_k = \frac{1}{\sum_{k=0}^{K-1} \pi_k^N} \pi_k^N \quad 0 \leq k \leq K-2$$

이며, π_k^N 은

$$\pi_k^N = \sum_{n=N-k-1}^{N-1} \binom{n}{N-k-1} p_n^N (-1)^{n-N+k+1} \quad 0 \leq k \leq N-1$$

이며, 여기서

$$p_n^N = \zeta_n \sum_{l=n}^{N-1} \binom{N-1}{l} \frac{\frac{1}{\zeta_l}}{\sum_{l=0}^{N-1} \binom{N-1}{l} \frac{1}{\zeta_l}}$$

이고

$$\zeta_0 = 1,$$

$$\zeta_n = \prod_{j=1}^n \frac{B^*(j\lambda)}{1 - B^*(j\lambda)}, \quad n = 1, 2, 3, \dots, N-1$$

이다. 여기서 $B^*(s)$ 는 $B(x)$ 의 LST로서 다음과 같다.

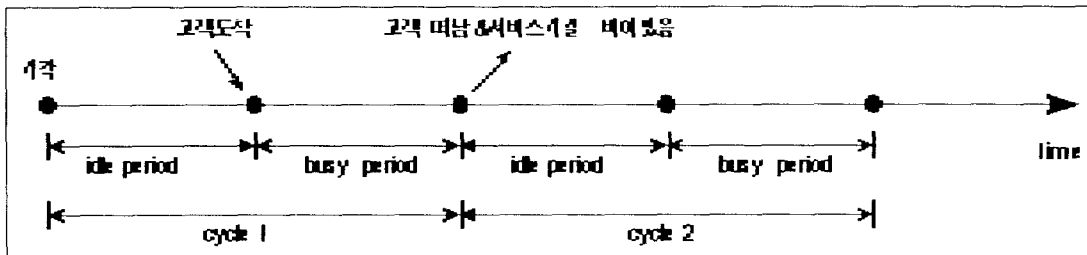
$$B^*(s) = \int_0^\infty e^{-sx} dB(x)$$

3. 최초 시스템 용량초과(overflow) 시간

3.1 최초 시스템 용량초과 시간의 평균

M/G/1/K/N 모형에서 최초로 시스템 용량초과가 발생할 때 까지의 시간을 T라고 하고 이에 대한 평균을 유도해 보기로 한다.

대기행렬 시스템은 (그림 3.1)과 같이 쉬는 시간(idle period)과 바쁜 시간(busy period)이 반복해서 나타나며 쉬는 시간과 바쁜 시간은 하나의 주기(cycle)를 이룬다.



(그림 3.1) 대기행렬 시스템의 주기(cycle)

여기서, X_1, X_2, \dots 는 순서대로 나타나는 쉬는 시간의 길이라고 하고, Y_1, Y_2, \dots 는 바쁜 시간의 길이라고 하면 이들은 서로 독립인 확률변수이며, X_1, X_2, \dots 는 각각 모수 λ 인 지수분포를 따르며, Y_1, Y_2, \dots 는 평균이 $\frac{b}{\pi_0}$ 인 확률변수이다. π_0 는 서비스가 끝난 직후 서비스 시설에 남아 있는 고객이 1명도 없을 확률이 되므로 $\frac{1}{\pi_0}$ 은 각각의 바쁜 시간에서 서비스를 받는 고객의 평균 수가 된다. 따라서 $E(Y_1) = \frac{b}{\pi_0}$ 임을 알 수 있다.

시스템 용량초과는 항상 바쁜 시간에서 발생하게 되므로, 임의의 바쁜 시간내에서 시스템 용량

초과가 일어날 확률을 P_K^* 라고 하고, M을 최초로 이러한 사건이 일어날 때 까지의 총 주기 수라고 하면, 확률변수 M은 모수 P_K^* 인 기하분포(geometric distribution)를 따른다.

따라서 유도하고자 하는 확률변수 T의 평균 $E(T)$ 는 다음과 같이 근사적으로 구할 수 있다.

$$\begin{aligned}
 E(T) &= E[E(T/M)] \\
 &= \sum_{m=1}^{\infty} E(T/M=m)P(M=m) \\
 &\approx \sum_{m=1}^{\infty} E[(X_1 + X_2 + \dots + X_m) + (Y_1 + Y_2 + \dots + Y_m)]P(M=m) \\
 &= \sum_{m=1}^{\infty} mE(X_1 + Y_1)P(M=m) \\
 &= E(M) E(X_1 + Y_1) \\
 &= \frac{E(X_1) + E(Y_1)}{P_K^*} \tag{3.1}
 \end{aligned}$$

여기서 $E(X_1)$, $E(Y_1)$ 은 각각

$$E(X_1) = \frac{1}{N\lambda}, \quad E(Y_1) = \frac{b}{\pi_0}$$

이다. 그러므로 P_K^* 의 값을 알면 $E(T)$ 를 근사적으로 구할 수 있다.

3.2 모수 유도

M/G/1/K/N 모형의 임의의 바쁜 시간 내에서 시스템 용량초과가 일어날 확률 P_K^* 는 M/G/1/N 모형을 이용하여 유도할 수 있다.

M/G/1/K/N 모형에서 시스템 용량초과는 바쁜 시간에서 서비스 시설의 고객의 수가 K명일 때 고객이 도착함으로 인해서 발생하게 되며, 이것은 M/G/1/N 모형에서 보면 바쁜시간에서 서비스 시설의 고객의 수가 K+1명 이상이 되는 확률과 같아진다. 또한 이 확률은 M/G/1/N 모형의 바쁜 시간내의 어느 마코프 시점에서의 고객의 수가 K명 이상이 되는 확률과 같다. 왜냐하면 M/G/1/K/N 모형에서의 마코프 시점은 고객이 서비스를 받고 떠나는 시점이므로, 이 마코프 시점에서 고객의 수가 K명 이라는 것은 이 시점 바로 직전의 고객수는 K+1명이 되기 때문이다.

따라서 $m < K$ 에 대해서 U_m 을 다음과 같이 정의하면,

$$\begin{aligned}
 U_m &= P(\text{M/G/1/N 모형의 마코프 시점에서 } m\text{명의 고객이 있다가 그 후 어느} \\
 &\quad \text{마코프 시점에서의 고객의 수가 0이 되지 않으면서 최소 } K\text{명이 됨})
 \end{aligned}$$

다음식이 성립한다.

$$P_K^* = U_0$$

또한, 이 경우 다음 식이 성립한다.

$$\begin{aligned}
 U_0 &= p_{01}^N U_1 + \dots + p_{0K-1}^N U_{K-1} + p_{0K}^{N*} \\
 U_1 &= p_{11}^N U_1 + \dots + p_{1K-1}^N U_{K-1} + p_{1K}^{N*} \tag{3.2}
 \end{aligned}$$

⋮

$$U_{K-1} = p_{K-1,1}^N U_1 + \dots + p_{K-1,K-1}^N U_{K-1} + p_{K-1,K}^{N*}$$

여기서, p_{ij}^N 은 M/G/1/N 모형에서 $\{L_m; m=1,2,3,\dots\}$ 의 전이확률로서 식 (2.2)에 주어져 있으며,

$$p_{iK}^{N*} = 1 - \sum_{j=0}^{K-1} p_{ij}^N \quad i=0,1, \dots, K-1$$

는 상태 i 에서 상태 K 이상으로의 전이가 일어날 확률이다.

식 (2.1)에서 항상 $p_{0j}^N = p_{1j}^N, j=1,2,\dots,N$ 기 때문에 식 (3.2)의 첫 번째 식과 두 번째 식은 같으므로 $U_0 = U_1$ 이며, 식 (3.2)를 행렬을 사용하여 표현하면 다음과 같다.

$$U = PU + P_K \tag{3.3}$$

여기서

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_{K-1} \end{pmatrix}, \quad P = \begin{pmatrix} p_{11}^N & \dots & p_{1,K-1}^N \\ \vdots & & \vdots \\ p_{K-1,1}^N & \dots & p_{K-1,K-1}^N \end{pmatrix}, \quad P_K = \begin{pmatrix} p_{1K}^{N*} \\ p_{2K}^{N*} \\ \vdots \\ p_{K-1,K}^{N*} \end{pmatrix}$$

이다.

따라서, 식 (3.3)으로부터

$$U = (I - P)^{-1} P_K \tag{3.4}$$

이 되며, 여기서 I 는 단위행렬(identity matrix)이다.

식 (3.4)에서 확률 $P_K^* = U_0 = U_1$ 을 얻을수 있으며, 따라서 식 (3.1)로부터 구하고자 하는 $E(T)$ 는

$$E(T) \approx \frac{\frac{1}{N\lambda} + \frac{b}{\pi_0}}{U_0} = \frac{\pi_0 + N\lambda b}{U_0 N\lambda \pi_0}$$

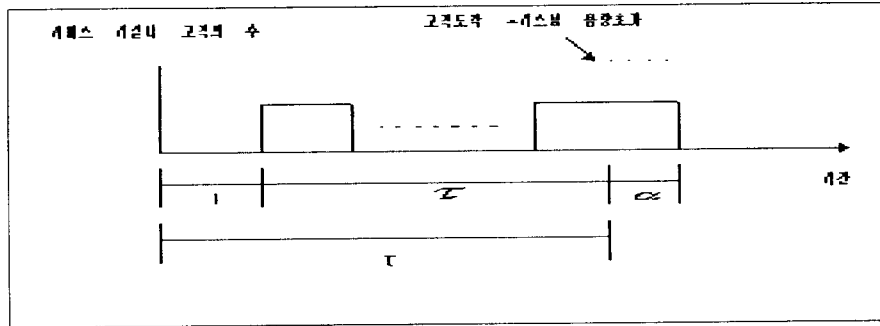
이다.

3.3 기존 연구결과와의 비교

Halfin(1981)은 M/G/1/1 모형에서 두 개의 연속적인 시스템 용량초과 시간의 길이 τ_B 와, 서비스 시작 시점에서 시스템 용량초과 시간까지의 길이 τ 에 대한 분포를 유도하였으며, 그 결과에 따르면 τ_B 와 τ 의 평균은

$$E(\tau_B) = E(\tau) = \frac{1}{\lambda} \left(1 + \frac{1}{\lambda b}\right)$$

이다.



(그림 3.2) M/G/1/1 모형에서 τ 와 T의 비교

여기서 τ 와 T의 관계는 (그림 3.2)에서와 같이 $T = I + \tau$ 로 주어진다. 이 사실을 서비스 시간의 분포가 평균이 b 인 지수분포를 따르는 M/M/1/1/N 모형에 적용하면 Takagi(1993)의 $E(\tau_B)$ 로부터 다음 사실을 얻을 수 있다.

$$E(\tau_B) = E(\tau) = \frac{1}{N\lambda} \left(1 + \frac{1}{N\lambda b} \right) \tag{3.5}$$

한편 본 논문에서는 $E(T)$ 를 근사적으로 다음과 같이 얻었다.

$$E(T) \approx \frac{b + \frac{1}{N\lambda}}{P_1^*} \tag{3.6}$$

여기서,

$$\begin{aligned} P_1^* &= P(\text{서비스 시간 동안 적어도 1명의 고객이 도착함}) \\ &= 1 - P(\text{서비스 시간 동안 1명의 고객도 도착하지 않음}) \\ &= 1 - \int_0^\infty P(\text{서비스 시간 동안 } N-1\text{명중 } 0\text{명이 도착함} / \text{서비스시간} = x) dB(x) \\ &= 1 - \int_0^\infty \binom{N-1}{0} [P(Y \leq x)]^0 [P(Y > x)]^{N-1} dB(x) \\ &= 1 - \int_0^\infty e^{-(N-1)\lambda x} dB(x) \\ &= \frac{(N-1)\lambda}{(N-1)\lambda + \frac{1}{b}} \end{aligned}$$

이며, 이 결과는 식 (3.4)에서도 유도될 수 있다.

따라서,

$$\begin{aligned} E(T) &\approx \frac{(N\lambda b + 1)[(N-1)\lambda b + 1]}{N\lambda(N-1)\lambda b} \\ &\approx \frac{(N\lambda b + 1)^2}{(N\lambda)^2 b}, \quad N \text{ 이 충분히 큰 경우} \\ &= \frac{1}{N\lambda} \left(N\lambda b + 2 + \frac{1}{N\lambda b} \right) \end{aligned}$$

$$\begin{aligned} &\approx \frac{1}{N\lambda} \left(2 + \frac{1}{N\lambda b} \right), \quad N\lambda b \text{ 가 충분히 작을 경우} \\ &= \frac{1}{N\lambda} + \frac{1}{N\lambda} \left(1 + \frac{1}{N\lambda b} \right) \end{aligned} \tag{3.7}$$

가 되어 N 이 충분히 크고, $N\lambda b$ 가 충분히 작을 경우(여기서 $N\lambda b = \rho$ 는 시스템의 부하량으로서 일반적으로 대기행렬 시스템에서는 1보다 상당히 작다.) Takagi 의 결과와 본 논문의 결과는 다음과 같이 근사적으로 같다.

$$\begin{aligned} E(T) &= E(I + \tau) \\ &\approx E(I + \tau + a) \end{aligned}$$

3.4 오차문제

여기서 유도한 $E(T)$ 의 근사값 $E(T) \approx \frac{\pi_0 + N\lambda b}{U_0 N\lambda \pi_0}$ 의 오차는 식(3.1)에서 보면, 바쁜시간의 평균 $E(Y_1)$ 이 시스템 용량초과가 일어나는 긴 바쁜시간으로 인해 과다추정됨을 알수 있다. 따라서 오차의 크기는 시스템 용량초과가 일어나는 바쁜시간의 길이가 길어질수록 커지며, P_K^* 가 커질수록 커진다. 시스템 용량초과가 일어나는 바쁜시기의 길이와 P_K^* 는 주어진 K 에 대해서 시스템의 부하량 $\rho = N\lambda b$ 에 비례해서 커지며, ρ 가 클 경우 식(3.7)에서도 알 수 있듯이 오차가 커진다. 하지만 ρ 가 큰 대기행렬 시스템에서는 시스템 용량초과가 빈번히 일어나게 되어 최초 시스템 용량초과 시간에 대한 분석의 의미가 적어진다. 이 오차의 크기에 대한 정량적인 분석은 어려우므로 ρ 가 작은 몇가지 경우에 대해서 시뮬레이션 결과와 비교하여 보았다.

4. 모의실험결과와의 비교

4.1 K=1 인 경우

$K=1$ 이고 서비스 시간의 분포가 평균 b 인 지수분포를 따르는 M/M/1/1/N 모형에 대해, 모의실험결과 $E(T)$ 는 시스템의 상태 0으로부터 시작하여 최초의 시스템 용량초과가 발생할 때까지의 시간을 1,000번의 모의실험을 통해 추정하였고, 근사적 $E(T)$ 는 식 (3.1)을 이용해서 구하였으며, Takagi(1993)의 결과와 비교하였다.

<표 4.1> M/M/1/1/N 모형에서 $E(T)$ 비교

$N\lambda, b$	N, λ	Takagi 결과	모의실험 결과 $E(T)$	근사적 $E(T)$	Takagi 결과 대비 상대오차
$N\lambda=1.0$ $b=0.5$	$N=50 \lambda=0.02$	4.0	3.94615	4.56122	14%
	$N=100 \lambda=0.01$	4.0	3.92143	4.53030	13%
$N\lambda=2.0$ $b=0.5$	$N=50 \lambda=0.04$	1.5	1.45462	2.02041	35%
	$N=100 \lambda=0.02$	1.5	1.44075	2.01010	34%

Takagi의 결과와 모의실험 결과 $E(T)$ 는 유사하게 나타났으나, 근사적 $E(T)$ 의 상대오차는 비

교적 크게 나타났다. 실제로 식(3.5)와 식(3.6)으로부터 M/M/1/1/N 모형의 경우 $\{ \text{근사적} E(T) - b \} \geq E(T)$ 이 성립한다.

4.2 K가 2 이상인 경우

K가 2이상인 경우에는 모의실험을 통하여 서비스 시간의 분포가 평균 b 인 확정분포 (deterministic distribution)를 따르는 M/D/1/K/N 모형과 지수분포를 따르는 M/M/1/K/N 모형에 대해서 모의실험 결과 $E(T)$ 와 근사적 $E(T)$ 를 비교하였으며, 그 결과는 <표 4.2>와 같다.

시스템의 부하량이 $\rho = N\lambda b = 0.5$ 인 경우 서비스 비율이 고객의 도착 비율의 2배이기 때문에 서비스 시설이 비교적 한산한 경우이며, 따라서 K가 증가함에 따라 시스템 용량초과가 일어날 확률은 매우 작아지고, 구하고자 하는 $E(T)$ 는 매우 빠른 속도로 증가하게 되며, K가 5이상인 경우 시스템 용량초과가 거의 일어나지 않으나, 시스템의 부하량이 증가함에 따라 시스템 용량초과가 빨리 일어난다.

<표 4.2> 근사적 $E(T)$ 와 모의실험결과 $E(T)$

N λ, b	K	N, λ	M/D/1/K/N 모형			M/M/1/K/N 모형		
			모의실험 $E(T)$	근사적 $E(T)$	상대오차	모의실험 $E(T)$	근사적 $E(T)$	상대오차
N λ=1.0 b=0.5	3	N= 50, λ=0.02	56.37235	56.92173	1%	28.17506	30.43252	8%
		N=100, λ=0.01	53.16357	53.69292	1%	26.77779	29.28483	9%
	4	N= 50, λ=0.02	221.55545	225.64779	2%	67.17133	69.05150	3%
		N=100, λ=0.01	199.74256	201.75685	1%	61.34755	64.29232	5%
	5	N= 50, λ=0.02	894.74713	922.53015	3%	149.12413	154.89105	4%
		N=100, λ=0.01	742.99775	766.25360	3%	135.39733	138.15470	2%

5. 결 론

M/G/1/K/N 모형에서 시스템 용량초과 시간과정에 대한 분석의 일환으로 최초 시스템 용량초과 시간 까지의 길이의 평균 $E(T)$ 에 대한 근사적 해를 유도하였다. K=1인 경우 Takagi(1993)의 결과와 이론적으로 비교하였고 모의실험을 통해 실제 계산결과를 제시하고, K가 2 이상인 경우 M/D/1/K/N 모형과 M/M/1/K/N 모형에 대해 모의실험을 통해 오차의 크기를 비교하였다.

추후 과제로서 과다추정되는 근사적 $E(T)$ 에 대한 오차수정 문제를 정량화 할 필요가 있으며, 최초 시스템 용량초과 시간에 대한 분석을 벗어나 연속적인 두 개의 시스템 용량초과 시간과정에 대한 일반적인 해를 구할 수 있는 방법이 고안되어야 할 것이다.

참 고 문 헌

[1] Bhat, U. N. (1975), 'Imbedded Markov Chains in Queueings M/G/1 and GI/M/1 with Limited Waiting Room', *Metrika*, Vol. 22, 153-160
 [2] Bunday, B. D., *Basic Queueing Theory*. Edward Arnold, London, 1986.

- [3] Cohen, J. W., The Single Server Queue. North Holland, Amsterdam, 1982.
- [4] Halfin, S. (1981), '*Distribution of the inter overflow time for the GI/G/1 loss system*', Mathematics of Operations Research, Vol. 6, No.4, 563-570
- [5] Jaiswal, N. K., Priority Queues. Academic Press, New York, 1968.
- [6] Saaty, T. L., Elements of Queueing Theory with Applications. McGraw-Hill Book Company, New York, 1961.
- [7] Stecke, K. E., Machine interference : the assignment of machines to operators. Handbook of Industrial Engineering, G. Salvendy, John Wiley and Sons, New York, 1982.
- [8] Steke, K. E. and Aronson, J. E. (1985), '*Review of operator/machine interference model*', International Journal of Production Research, Vol. 23, No. 1, 129-151
- [9] Takagi, H., Queueing Analysis : Volume 2, Finite Systems. North-Holland, Amsterdam, 1993.
- [10] Takine, T., Takagi, H. and Hasegawa, T. (1993), '*Analysis of an M/G/1/K/N Queue*', Journal of Applied Probability, Vol. 30, No. 2, 446-454