

N-정책하의 순환 서비스시스템의 평균대기시간 분석

홍정완* · 이창훈*

Mean Waiting Time Analysis of Cyclic Server System under N-Policy

Jeong Wan Hong* and Chang Hoon Lie*

Abstract

We consider a cyclic server system under N-policy. This system consists of multiple queues served in a cyclic order by a single server. In this paper, we consider the following control policy. Every time server polls one queue, the server inspects the state of the queue. If the total number of units is found to have reached or exceeded a pre-specified value, the server begins to serve the queue until it is empty. As soon as the queue becomes empty, the server polls next queue.

An approximate analysis of this system is presented. Server vacation model is used as an analytical tool. However, server vacation periods are considered to be dependent on the service times of respective queues. The results obtained from the approximate analysis are compared with simulation results.

1. 서론

순환 서비스 시스템(Cyclic Server System)은 단일 서버에 의하여 순차적으로 서비스를 받는 여러개의 대기(queue)가 존재하는 시스템이다. 이 시스템은 컴퓨터 통신 시스템에 많이 이용되고 있다. 특히, 토큰을 사용하는 근거리

통신망(Local Area Network : LAN)은 이러한 시스템의 대표적인 예이다. 최근에는 생산 시스템에도 많이 응용되고 있다.[10]

순환 서비스 시스템을 분석하는 일반적인 목적은 다른 대기 시스템과 마찬가지로 고객(customer)의 평균 대기 시간(mean waiting time)을 구하는 것이다. 다양한 순환 서비스 시스템 모형에 대한 지금까지의 분석 방법에

* 서울대학교 산업공학과

대해서는 Takagi[12], Levy와 Sidi[9]의 논문에 잘 요약되어 있다.

대기행렬 시스템을 경제적으로 운용하기 위해서는 대기행렬 시스템의 최적설계 및 제어정책(control policy)이 필요하다. 즉, 서비스를 할 것인가의 여부, 우선순위(priority)를 줄 것인가의 여부 등은 실제 현장에서 흔히 발생하는 문제이다. 단일 대기 시스템의 제어정책에 대해서는 Teghem[13]등의 논문에 잘 요약되어 있다. 최근 들어서는 순환 서비스 시스템에도 제어정책을 고려하고 있다. 특히, 어떠한 순서로 서비스를 할 것인가[3], 하나의 대기에서 얼마만큼 서비스를 할 것인가[2] 등의 문제에 대하여 많이 연구되고 있다.

본 연구에서는 서어버가 각 대기를 방문했을 때, 항상 서비스하는 것이 아니고, 정해진 수(N명) 이상의 고객이 존재해야만 서비스를 할 수 있는 순환 서비스 시스템을 분석하고자 한다. 이러한 제어 정책을 단일 대기 모형에서는 N-제어정책이라고 한다. 단일 대기 N-제어정책 모형에 대해서는 지금까지 많은 연구가 이루어져 왔다. Lee와 Srinivasan[8]은 휴가(vacation)가 있는 M/G/1 시스템에 대해서 기대 대기 시간을 구하였으며, 이효성[1]은 휴가가 있는 M^{*}/G/1 시스템에 대해서 안정 상태 확률(steady state probability)을 유도하였다. Hofri[4]는 두 개의 대기가 존재하는 시스템에 대하여 분석하였다. 그러나 여러 개의 대기를 가지는 순환 서비스 시스템에 대해서 아직 N-제어정책을 적용한 경우는 아직 찾을 수 없었다.

본 연구에서 분석하고자하는 순환 서비스 시스템은 다음과 같은 제어정책하에서 운용된다. 시스템내의 대기의 개수는 g개이며, 서비스를 받기위한 고객은 포아송 과정(Poisson pro-

cess)에 의해 용량이 무한한 각 대기에 도착한다. 서비스는 각 대기별로 선입선출방식(FIFO)에 의해 이루어지며, 서어버는 각 대기가 완전히 빌 때까지 서비스를 하고 난 후, 다음 대기로 넘어가는 고갈 서비스 정책(exhaustive service discipline)을 사용한다. 그리고 다음 대기에 도착한 서어버는 그 대기에 기다리고 있는 고객의 수를 관찰한다. 이 때 고객의 수가 일정한 수를 넘으면 서비스를 시작하고, 그렇지 않으면 서비스를 하지 않고 다음 대기로 넘어간다. 각 대기의 서비스 시간과 대기전이 시간(switchover time)은 일반분포(general distribution)를 따른다.

2. 기호 정의 및 접근방법

본 연구에서는 N-제어정책을 사용하는 비대칭적(asymmetric) 순환 서비스 시스템의 평균 대기 시간을 분석하려고한다. 제어정책을 사용하지 않는 비대칭적 순환 서비스 시스템의 경우, 지금까지의 연구에서는 고갈 서비스와 차단 서비스(gated service)의 경우에만 정밀해(exact solution)를 구하였으며, 그 외의 서비스 정책에 대해서는 근사해(approximate solution)를 구하는 데 그치고 있다.

본 연구에서 다루는 모형은 비록 서비스 정책이 고갈 서비스이지만, 고객이 있어도 서비스를 하지않는 제어정책의 특징 때문에, 지금까지의 분석 방법으로는 정밀해를 구할 수 없다. 따라서 본 연구에서도 각 대기를 독립적으로 계산하여 해를 구해내는 방법[5][7]을 사용하여 근사해를 구하는 알고리즘을 개발하였다. 이 방법은 M/G/1 휴가모형과 유사하지만,

휴가기간에 다른 대기의 특성이 반영되므로 연속되는 휴가 기간의 분포가 서로 다른 것이 특징이다.

본 연구에서 사용하는 중요한 기호들은 아래와 같다.

- N_i : i 번째 대기에서의 제어 파라미터
- λ_i : i 번째 대기에서의 도착률
- b_i : i 번째 대기에서의 서비스 시간 확률 변수
- ρ_i : i 번째 대기에서의 가동률($=\lambda_i b_i$)
- r_i : $(i-1)$ 대기에서 i 대기로의 전이(switchover)시간 확률변수
- B_i : i 번째 대기에서의 평균 가동 기간(busy period) 확률 변수
- $X_i^{(1)}$: 확률변수 X_i 의 1차 모멘트(first moment)
- $X_i^{(2)}$: 확률변수 X_i 의 2차 모멘트
- $X_i^{(r)}$: 확률변수 X_i 의 평균 잔여시간(mean residual life)
- $E[W_i]$: i 번째 대기에서의 평균 대기 시간
- $E[U_{ij}]$: 임의의 고객이 대기 i 에 도착했을 때 서어버가 대기 j 를 서비스하고 있는 경우, 대기 j 를 떠나고 난 후부터 대기 i 를 방문할 때까지 소요되는 시간
- $E[V_{ij}]$: 임의의 고객이 대기 i 에 도착했을 때 서어버가 대기 j 로 전이하고 있는 경우, 대기 j 를 방문한 후부터 대기 i 를 방문할 때까지 소요되는 시간
- PB_i : 임의의 고객이 시스템에 도착했을 때, 서어버가 대기 i 를 서비스하고

있을 확률

- PR_i : 임의의 고객이 시스템에 도착했을 때, 서어버가 대기 i 로 전이하고 있을 확률
- $P_i(l)$: 대기 i 에 l 명의 고객이 대기하고 있을 확률
- P_i : 어떠한 주기시간 동안 대기 i 를 서비스할 확률
- $h(l)$: 대기 i 에서 한 명의 고객을 서비스하는 동안 도착한 고객의 수가 l 명일 확률

3. 대기시간 분석

대기 시간을 고객이 서비스를 받기 전까지 대기에 머무르는 시간이라고 정의하자. 그러면 임의의 고객이 i 번째 대기에 도착했을 때의 대기 시간은 다음의 3가지 항목들로 구성된다.

- 도착했을 때부터 서비스 조건이 충족될 때까지의 시간
- 그 이후부터 서어버가 대기 i 를 방문할 때까지의 시간
- 자기 앞의 고객들을 서비스하는 시간

위의 항목들을 효율적으로 구하기 위해서는 i 번째 대기에 고객이 도착했을 때 i 번째 대기에 존재하는 고객의 수와 서어버의 위치를 다음의 3가지 경우로 나눌 수 있다.

- 1) i 번째 대기에 l 명의 고객이 기다리고 있고, 서어버가 i 번째 대기를 서비스하고 있는 경우
- 2) i 번째 대기에서 기다리고 있는 고객 수 l

- 이 $N_i - 1$ 보다 크거나 같고, 서어버가 i 번째 대기를 서비스하고 있지 않은 경우
- 3) i 번째 대기에서 기다리고 있는 고객 수 l 이 $N_i - 1$ 보다 작고, 서어버가 i 번째 대기를 서비스하고 있지 않은 경우

위의 각 경우를 조건으로 하여 대기시간을 각각 구하면 다음과 같다.

$$E[W_i | (1)의 경우] = ib_i^{(1)} + b_{0i}^{(1)} \quad (1a)$$

$$E[W_i | (2)의 경우] = ib_i^{(1)} + \frac{1}{1 - PB_i} \left\{ \sum_{j=i}^{\infty} PB_j (B_{0j}^{(1)} + E[U_{j\bar{i}}]) + \sum_{j=i}^{\infty} PR_j (r_{0j}^{(1)} + E[V_{j\bar{i}}]) \right\} \quad (1b)$$

$$E[W_i | (3)의 경우] = ib_i^{(1)} + \frac{1}{\lambda_i} (N_i - i - l) + \frac{1}{1 - PB_i} \left\{ \sum_{j=i}^{\infty} PB_j (B_{0j}^{(1)} + E[U_{j\bar{i}}]) + \sum_{j=i}^{\infty} PR_j (r_{0j}^{(1)} + E[V_{j\bar{i}}]) \right\} \quad (1c)$$

위의 조건들을 풀고 Little의 공식을 사용하여 $E[W_j]$ 를 구하면 다음과 같다.

$$E[W_i] = \frac{1}{1 - \rho_i} \left\{ PB_i \cdot b_{0i}^{(1)} + \frac{1}{\lambda_i} (1 - PB_i) \sum_{l=0}^{N_i-1} (N_i - l - 1) P_i(l) + \sum_{j=i}^{\infty} PB_j (B_{0j}^{(1)} + E[U_{j\bar{i}}]) + \sum_{j=i}^{\infty} PR_j (r_{0j}^{(1)} + E[V_{j\bar{i}}]) \right\} \quad (2)$$

여기에서 고객의 도착과정이 포아송이므로 다음과 같이 쓸 수 있다.

$$B_{0j}^{(1)} = \frac{B_j^{(2)}}{2B_j^{(1)}}, \quad r_{0j}^{(1)} = \frac{r_j^{(2)}}{2r_j^{(1)}}, \quad b_{0j}^{(1)} = \frac{b_j^{(2)}}{2b_j^{(1)}}$$

주기 시간(cycle time)을 한 대기를 떠난 후, 다음 그 대기를 떠날 때까지 소요되는 사이 시간이라고 정의하면, 평균 주기 시간을 아래와 표현할 수 있다.

$$C = \sum_{k=1}^{\infty} (r_k^{(1)} + P_k B_k^{(1)}) \quad (3)$$

그런데 시스템이 안정되어(stable) 있는 경우에, 한 주기 시간 동안 어떠한 대기에 도착하는 평균 고객수와 서비스를 받고 나가는 평균 고객 수는 같다. 따라서, 다음과 같은 관계식이 성립한다.

$$\lambda_j \sum_{k=1}^{\infty} (r_k^{(1)} + P_k B_k^{(1)}) = P_j \frac{B_j^{(1)}}{b_j^{(1)}} \quad (4)$$

이다. 위의 식 (4)를 풀면 다음과 같이 P_j 를 유도할 수 있다.

$$P_j = \frac{\rho_j}{B_j^{(1)}} \frac{r}{1 - \rho} \quad (5)$$

이다. 여기에서 $\rho = \sum_{k=1}^{\infty} \rho_k$, $r = \sum_{k=1}^{\infty} r_k^{(1)}$ 이다. 그리고 PASTA 정리에 의해서 PR_j 와 PB_j 는 아래와 같다.

$$PB_j = \frac{P_j B_j^{(1)}}{C}, \quad PR_j = \frac{r_j^{(1)}}{C}$$

$E[U_{j\bar{i}}]$ 와 $E[V_{j\bar{i}}]$ 를 유도하기 위하여 아래의 두 확률을 정의하자.

PU_{jk} : 임의의 고객이 시스템에 도착했을 때, 서어버가 대기 j 를 서비스하고 있으며, 그 후 대기 k 를 방문했을 때 대기 k 에서 서비스할 확률

PV_{jk} : 임의의 고객이 시스템에 도착했을 때, 서버가 대기 j 로 전이하고 있으며, 그 후 대기 k 를 방문했을 때 대기 k 에서 서비스할 확률

$$E[V_{jk}] = \sum_{k=j+1}^{\infty} r_k^n + \sum_{k=j}^{i-1} PV_{jk} B_k^i, \quad i > j \quad (6c)$$

$$E[V_{jk}] = \sum_{k=j+1}^{\infty} r_k^i + \sum_{k=i}^{j-1} r_k^i + \sum_{k=j+1}^{\infty} PV_{jk} B_k^i + \sum_{k=i}^{j-1} PV_{jk} B_k^i, \quad i < j \quad (6d)$$

따라서 $E[U_i]$ 와 $E[V_{jk}]$ 는 아래와 같이 표현된다.

$$E[U_{jk}] = \sum_{k=j+1}^i r_k^n + \sum_{k=j+1}^{i-1} PU_{jk} B_k^i, \quad i > j \quad (6a)$$

$$E[U_{jk}] = \sum_{k=j+1}^{\infty} r_k^i + \sum_{k=i}^{j-1} r_k^i + \sum_{k=j+1}^{\infty} PU_{jk} B_k^i + \sum_{k=i}^{j-1} PU_{jk} B_k^i, \quad i < j \quad (6b)$$

$E[U_{jk}]$ 와 $E[V_{jk}]$ 를 구하기 위해서는 PU_{jk} 와 PV_{jk} 를 알아야한다. C_k 를 임의의 고객이 시스템에 도착했을 때, 서버가 대기 j 를 서비스하고 있을 경우의 대기 k 의 평균 주기 시간이라고 하자. 그러면 고객이 도착하기 이전에 방문한 대기 l 에서 서비스할 확률은 P_l 이 되며, 도착한 이 후의 대기 l 에서 서비스할 확률은 PU_{jk} 이 된다. 따라서 C_k 는 다음과 같다.

$$C_k = r + 2B_{0j}^i + \sum_{l=k+1}^{\infty} P_l B_l^i + \sum_{l=i}^{j-1} P_l B_l^i + \sum_{l=j+1}^k PU_{jk} B_l^i, \quad k > j \quad (7a)$$

$$C_k = r + 2B_{0j}^i + \sum_{l=k+1}^{j-1} P_l B_l^i + \sum_{l=j+1}^{\infty} PU_{jk} B_l^i + \sum_{l=i}^k PU_{jk} B_l^i, \quad k < j \quad (7b)$$

따라서, 식 (5)의 경우와 같이 계산하면 PU_{jk} 은 다음과 같다.

$$PU_{jk} = \frac{\rho_k}{B_k^i} \frac{r + 2B_{0j}^i + \sum_{l=k+1}^{\infty} P_l B_l^i + \sum_{l=i}^{j-1} P_l B_l^i + \sum_{l=j+1}^k PU_{jk} B_l^i}{1 - \rho_k}, \quad k > j \quad (8a)$$

$$PU_{jk} = \frac{\rho_k}{B_k^i} \frac{r + 2B_{0j}^i + \sum_{l=k+1}^{j-1} P_l B_l^i + \sum_{l=j+1}^{\infty} PU_{jk} B_l^i + \sum_{l=i}^k PU_{jk} B_l^i}{1 - \rho_k}, \quad k < j \quad (8b)$$

유사한 방법으로 PV_{jk} 도 아래와 같이 유도할 수 있다.

$$PV_{jk} = \frac{\rho_k}{B_k^i} \frac{\sum_{l=j+1}^{\infty} r_l^i + 2r_{0j}^i + \sum_{l=k+1}^{\infty} P_l B_l^i + \sum_{l=i}^{j-1} P_l B_l^i + \sum_{l=j+1}^k PV_{jk} B_l^i}{1 - \rho_k}, \quad k > j \quad (8c)$$

$$PV_{jk} = \frac{\rho_k}{B_k^i} \frac{\sum_{l=j+1}^{\infty} r_l^i + 2r_{0j}^i + \sum_{l=k+1}^{j-1} P_l B_l^i + \sum_{l=j+1}^{\infty} PV_{jk} B_l^i + \sum_{l=i}^k PV_{jk} B_l^i}{1 - \rho_k}, \quad k < j \quad (8d)$$

위의 식 (8)을 k 가 $j+1$ 인 경우부터 풀면 모든 k 에 대해서 PU_{jk} 와 PV_{jk} 를 유도할 수 있다. 식 (8)을 식 (6)에 대입하면 $E[U_{ij}]$ 와 $E[V_{ji}]$ 를 구할 수 있다.

3.1 가동 시간 계산

가동 시간을 계산하기 위해서는 먼저 어떠한 대기에서 서비스가 시작될 때, 몇 명의 고객이 대기 중에 존재하는가를 알아야 한다. i 번째 대기에서 서비스가 시작될 때 i 번째 대기 중에 존재하는 고객의 수를 J_i 라고 하면, 가동 시간의 분포는 아래와 같이 표현된다.

$$P(B_i=x) = \sum_{k=N_i}^j P(B_i=x | J_i=k) P(J_i=k) \quad (9)$$

따라서, 이 가동 시간의 LST는 $B_i(s) = J_i(X_i(s))$ 이다. 여기에서 $J_i(\cdot)$ 는 J_i 의 z -변환을 의미하며, $X_i(s)$ 는 i 번째 대기의 1-가동시간(1-busy period)의 LST(Laplace Stieltjes Transform)를 의미한다. 따라서 가동시간의 1차 모멘트(moment)와 2차 모멘트는 각각 다음과 같이 표현된다.[6]

$$B_i^{(1)} = \frac{b_i}{1-\rho_i} J_i^{(1)}$$

$$\alpha_{ji} = \frac{\rho_i}{B_i^{(1)}} \frac{r + B_j^{(1)} + \sum_{k=i+1}^j P_k B_k^{(1)} + \sum_{k=i}^{j-1} P_k B_k^{(1)} + \sum_{k=j+1}^{i-1} \alpha_{jk} B_k^{(1)}}{j - \rho_i}, \quad i > j \quad (11a)$$

$$\alpha_{ji} = \frac{\rho_i}{B_i^{(1)}} \frac{r + B_j^{(1)} + \sum_{k=i+1}^{j-1} P_k B_k^{(1)} + \sum_{k=j+1}^i \alpha_{jk} B_k^{(1)} + \sum_{k=i}^{i-1} \alpha_{jk} B_k^{(1)}}{j - \rho_i}, \quad i < j \quad (11b)$$

$$\beta_{ji} = \frac{\rho_i}{B_i^{(1)}} \frac{r + \sum_{k=i+1}^j P_k B_k^{(1)} + \sum_{k=i}^{j-1} P_k B_k^{(1)} + \sum_{k=i}^{i-1} \beta_{jk} B_k^{(1)}}{1 - \rho_i}, \quad i > j \quad (11c)$$

$$\beta_{ji} = \frac{\rho_i}{B_i^{(1)}} \frac{r + \sum_{k=i+1}^{j-1} P_k B_k^{(1)} + \sum_{k=i}^j P_k B_k^{(1)} + \sum_{k=i}^{i-1} \beta_{jk} B_k^{(1)}}{1 - \rho_i}, \quad i < j \quad (11d)$$

$$B_i^{(2)} = \frac{b_i^2}{(1-\rho_i)^2} J_i^{(2)} + \frac{b_i^2}{(1-\rho_i)^2} J_i^{(2)}$$

위의 J_i 의 모멘트들을 구하기 위해 Kuehn[7]의 가정을 사용하였다. 대기 방문 사이 시간(intervisit time)은 대기간의 관계를 반영하여 두 가지 형태 즉, 그 대기를 서비스 했을 경우와 서비스 하지 않은 경우로만 분리하였다. 대기 j 를 서비스했을 경우의 대기 j 에 대한 방문 사이 시간을 V_{1j} 라 하고, 서비스하지 않은 경우의 방문 사이 시간을 V_{2j} 라고 하자. Kuehn의 결과를 응용하여 V_{1j} 와 V_{2j} 의 LST를 구하면 아래와 같다.

$$V_{1j}(s) = \prod_{i=1}^j r_i(s) \cdot \prod_{i=j}^i [z_{ij} B_i(s) + (1-z_{ij})] \quad (10a)$$

$$V_{2j}(s) = \prod_{i=1}^j r_i(s) \cdot \prod_{i=j}^i [\beta_{ij} B_i(s) + (1-\beta_{ij})] \quad (10b)$$

위의 두 식을 이용하면 V_{1j} 와 V_{2j} 의 모멘트들을 구할 수 있다. 여기에서 z_{ij} 와 β_{ij} 는 각각 대기 j 를 서비스한 경우와 하지 않은 경우를 조건으로 했을 때 대기 i 를 서비스한 확률을 나타낸다. 이 확률들을 식(5)를 유도한 방식과 마찬가지로 계산하면

그리고 V_1 와 V_2 시간 동안 대기 j 에 도착한 고객의 수가 i 일 확률을 각각 $a_{1j}(i)$ 와 $a_{2j}(i)$ 라고 하면, 이들의 확률 값과 z -변환은 다음과 같다.

$$a_{1j}(i) = \int_0^{\infty} \frac{e^{-\lambda_j t} (\lambda_j t)^i}{i!} dV_{1j}(t) \quad (12a)$$

$$a_{2j}(i) = \int_0^{\infty} \frac{e^{-\lambda_j t} (\lambda_j t)^i}{i!} dV_{2j}(t) \quad (12b)$$

$$a_{1j}(z) = V_{1j}(\lambda_j - \lambda_j z) \quad (12c)$$

$$a_{2j}(z) = V_{2j}(\lambda_j - \lambda_j z) \quad (12d)$$

위의 식(12a)와 (12b)를 풀기 위해서는 $V_{1j}(t)$ 와 $V_{2j}(t)$ 의 분포를 알아야한다. 그러나 위

의 휴가기간에 대한 모멘트들은 알 수 있지만 분포는 알 수가 없다. 따라서 본 연구에서는 실험적인 데이터를 가지고 분포를 유도하는 경우 가장 많이 사용되는 감마분포를 사용하였다. 즉, 위의 휴가기간을 두 개의 모멘트를 만족시키는 감마분포로 가정하였다.[11]

Kuehn의 가정에 의해서 서비스를 받고 난 후 첫번째 대기 방문 사이 시간만이 V_1 를 따르며, 두번째 대기 방문 시간부터 서비스 받기 전까지의 방문 사이 시간의 분포는 모두 V_2 를 따른다. I_j^m 을 m -제어정책을 갖는 단일대기 M/G/1 휴가모형에서 서비스를 시작할 때 대기에 머무르고 있는 고객의 수라고 하면 J_j 의 분포는 다음과 같다.

$$P\{J_j = k\} = \sum_{l=0}^{N_j-1} a_{1j}(l) P\{I_j^{N_j-l} = k-l\} + a_{2j}(k), \quad k \geq N_j \quad (13)$$

따라서 J_j 의 z -변환식은 아래와 같다.

$$J_j(z) = \sum_{k=0}^{N_j-1} a_{1j}(k) z^k I_j^{N_j-k}(z) - I_j^{N_j} + a_{2j}(z), \quad (14)$$

그리고 식 (14)를 미분하면 J_j 의 1,2차 팩토리얼(factorial) 모멘트들을 다음과 같이 구할 수 있다.

$$J_j^1 = \sum_{k=0}^{N_j-1} a_{1j}(k) (I_j^{N_j-k})^1 + a_{2j}^1 \quad (15a)$$

$$J_j^2 = \sum_{k=0}^{N_j-1} a_{1j}(k) [(I_j^{N_j-k})^2 + 2k(I_j^{N_j-k})^1] + a_{2j}^2 \quad (15a)$$

그런데 Lee와 Srinivasan[8]에 의하면 I_j^m 의 1,2차 팩토리얼(factorial) 모멘트들은 각각 아래와 같이 표현된다.

$$I_j^{m1} = \sum_{k=1}^{m-1} \frac{a_{2j}(k)}{1-a_{2j}(0)} I_j^{m-k-1} + \frac{a_{2j}^1}{1-a_{2j}(0)} \quad (16a)$$

$$I_j^{m2} = \sum_{k=1}^{m-1} \frac{a_{2j}(k)}{1-a_{2j}(0)} [2k I_j^{m-k-1} + I_j^{m-k-2}] + \frac{a_{2j}^2}{1-a_{2j}(0)} \quad (16b)$$

그리고 $(I_j^{(1)}) = \frac{a_3^{(1)}}{1-a_3(0)}$ 이고 $(I_j^{(2)}) = \frac{a_3^{(2)}}{1-a_3(0)}$

이므로, 이것을 식 (16)에 대입하여 순차적으로(recursively) 풀면 I_j^m 의 모멘트들을 구할 수 있다.

3.2 $P_i(I)$ 의 계산

3.1절에서 우리는 방문 사이 시간을 두 종류로 나누었다. 따라서 어떠한 대기에서 서비스가 끝난 시점을 재생점(regeneration point)이라고 할 수 있다. 즉, 어떠한 대기에서 서비스가 끝난 시점부터 다음 서비스가 끝날 때까지를 재생기간(regeneration cycle)이라고 할 수 있다.

본 연구에서는 i 번째 대기를 대상으로 하여 위의 재생 기간을 다음의 3가지 구간으로 분리하였다.

- 1 구간: N_i 명의 고객이 도착할 때까지의 기간
- 2 구간: 그 이후부터 서비스가 시작될 때까지의 기간
- 3 구간: 그 이후부터 서비스가 끝날 때까지의 기간

위의 3가지 구간을 모두 합하면 재생기간이 된다. 기대 재생 기간을 $E[R_i]$ 라고 하면,

$E[R_i]$ 는 다음과 같이 표현된다.

$$E[R_i] = \frac{N_i}{\lambda_i} + \sum_{j=1}^g PB_j(B_0^{(i)} + E[U_{ji}]) + \sum_{j=1}^g PR_j(r_0^{(i)} + E[V_{ji}]) + B_i^{(1)} \quad (17)$$

그리고 각 구간별로 상태 $I (< N_i)$ 에 머무르는 기대 시간을 계산하면, 1구간에서는 $1/\lambda_i$ 이고, 2구간에서는 0임을 쉽게 알 수 있다. 그리고 3구간에서의 시간은 TQ_{Ni} 이라고 하자. 그런데 재생과정(regeneration process)의 정리에 의하여, 재생 기간 동안 어떠한 상태에 머무르는 시간의 비율이 그 상태의 안정 상태 확률이므로 상태 I 의 안정상태확률은 다음과 같다.

$$P_i(I) = \frac{I/\lambda_i + TQ_{Ni}}{E[R_i]} \quad (18)$$

본 절에서는 TQ_{Ni} 을 유도하였다. TQ_{Ni} 은 i 번째 대기에서 N_i 명의 고객을 가지고 가동시간이 시작되었을 경우, 대기에 기다리고 있는 고객의 수가 I 명일 확률을 의미한다. 이 값은 이 효성[1]의 논문에서의 $T_{N_i, I, i}$ 과 동하다. 따라서 다음과 같이 쓸 수 있다.

$$TQ_{Ni} = T_{N_i, I, i} = \sum_{k=I}^{N_i} T_{ik} \quad (19)$$

그리고, T_{ik} 는 다음과 같다.[1]

$$T_{ik} = \frac{I}{h_i(0)} [t_i(k) + \sum_{l=1}^{k-I} T_{il} (I - \sum_{j=0}^{k-l} h_i(j))] \quad (20)$$

여기에서 $t_i(k)$ 는 한 명의 고객으로 가동시간이 시작되었을 경우, 이 한 명을 서비스하는 동안 시스템이 k 의 상태에 머무르는 기대 시간을 의미한다. 따라서 다음과 같이 표현된다.

$$t_i(k) = \frac{I}{\lambda_i} \sum_{j=k}^{\infty} h_i(j) \quad (21)$$

그리고 식(18)로 부터 $T_{ii} = t_i(1)/h_i(0)$ 이므로, 식(20)도 식(16)에서 처럼 순환적으로 계산할 수 있다.

3.3 예제

본 연구에서는 대칭(symmetric) 시스템인 경우와 비대칭 시스템인 경우로 나누어서 실험해보았다. 두 경우 모두 서비스 시간은 평균이 1인 일양분포(uniform distribution)를 따르며, 전이 시간은 평균이 0.1인 일양분포를 따른다. 그리고 대기의 갯수는 10개이다. <표 1>은 대칭시스템의 결과이며, <표 2>는 비대칭시스템의 결과이다. <표 2>의 경우 $\lambda_1 = \lambda_2 = 5\lambda$ 이며, $\lambda_3 = \lambda_4 = 2\lambda$ 이고, $\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = \lambda_9 = \lambda_{10} = \lambda$ 이다.

<표 1>과 <표 2>에서의 시뮬레이션 결과는 SIMAN을 사용하여 도출하였으며, 그 값들은

95% 신뢰구간(confidence interval)을 의미한다.

<표 1>에서 보는 바와 같이 대칭 시스템의 경우에는, 본 알고리즘이 매우 잘 맞는다는 것을 알 수 있다. 그러나 부하(load)가 커질수록 오차가 점점 커지는 경향이 있다. 이러한 경향은 대기간의 독립가정을 사용하는 일반적인 알고리즘에서는 흔히 발생하는 것이다. 비대칭 시스템의 경우에는 <표 2>에서 보는 바와 같이 대칭 시스템의 경우보다 정확도가 떨어진다.

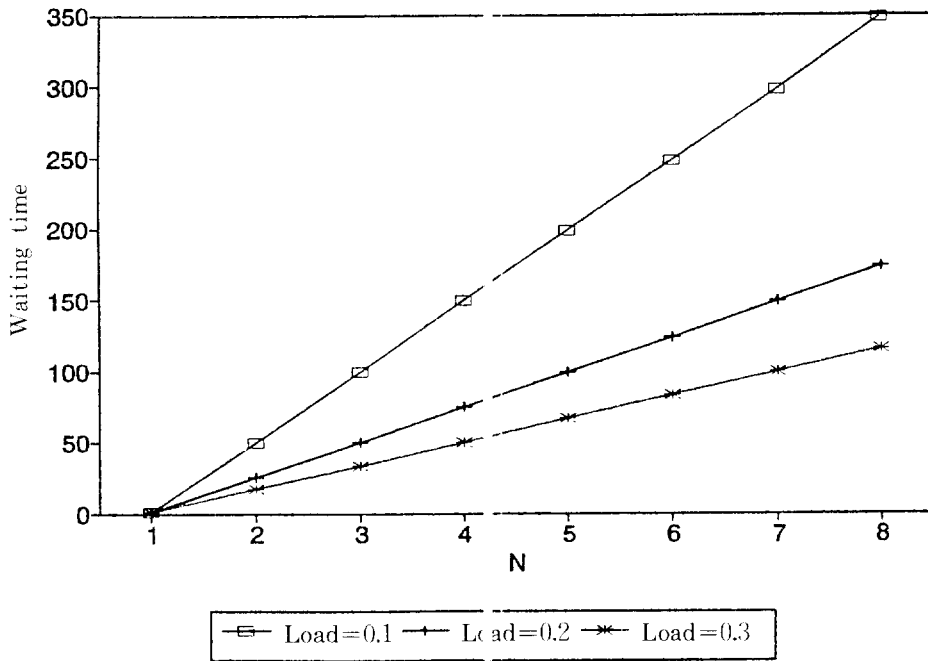
또한 대칭 시스템의 경우에 대하여, N 값을 변화시켜가면서 평균대기시간의 변화를 살펴보았다. [그림 1]은 트래픽이 작은 경우, [그림 2]는 중간 트래픽의 경우 그리고 [그림 3]은 트래픽이 큰 경우의 결과들이다. 서비스 시간과 전이 시간 그리고 대기의 개수는 <표 1>의 경우와 동일하다.

<표 1> 대칭 시스템인 경우의 평균대기시간(N=4)

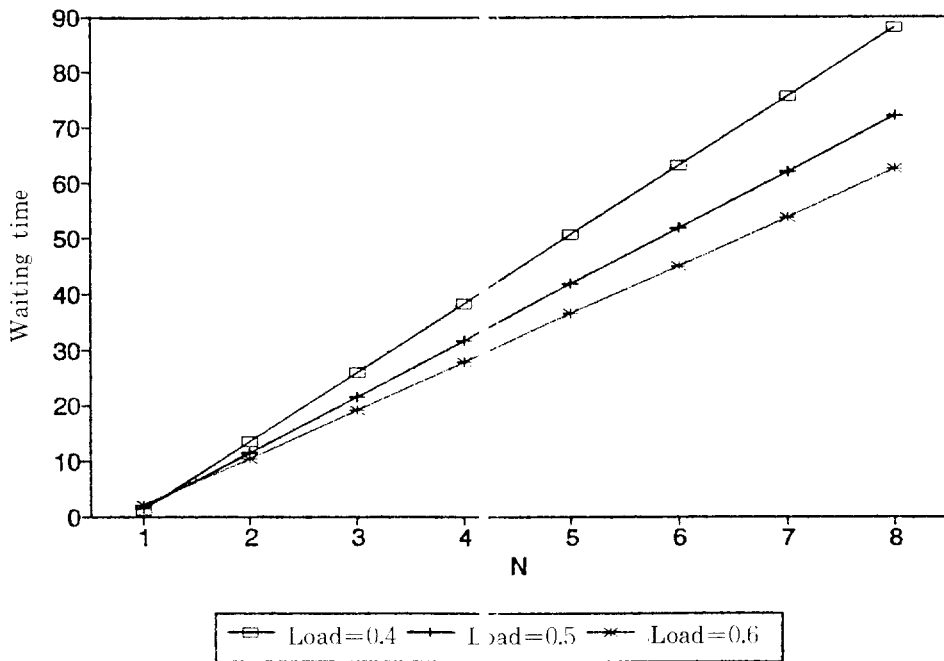
가동률	근사해	시뮬레이션	
0.1	149.21845	1.871734	± 11.48144
0.2	74.62999	5.11340	± 3.10853
0.3	50.15090	0.76228	± 1.45349
0.4	38.31274	8.70503	± 1.53307
0.5	31.67528	1.65205	± 0.74338
0.6	27.82971	7.82971	± 0.83597
0.7	25.86797	25.86797	± 0.48405
0.8	25.52535	25.52535	± 0.40687
0.9	27.92122	27.92122	± 0.41530

〈표 2〉 비대칭 시스템인 경우의 평균대기 시간(N=4)

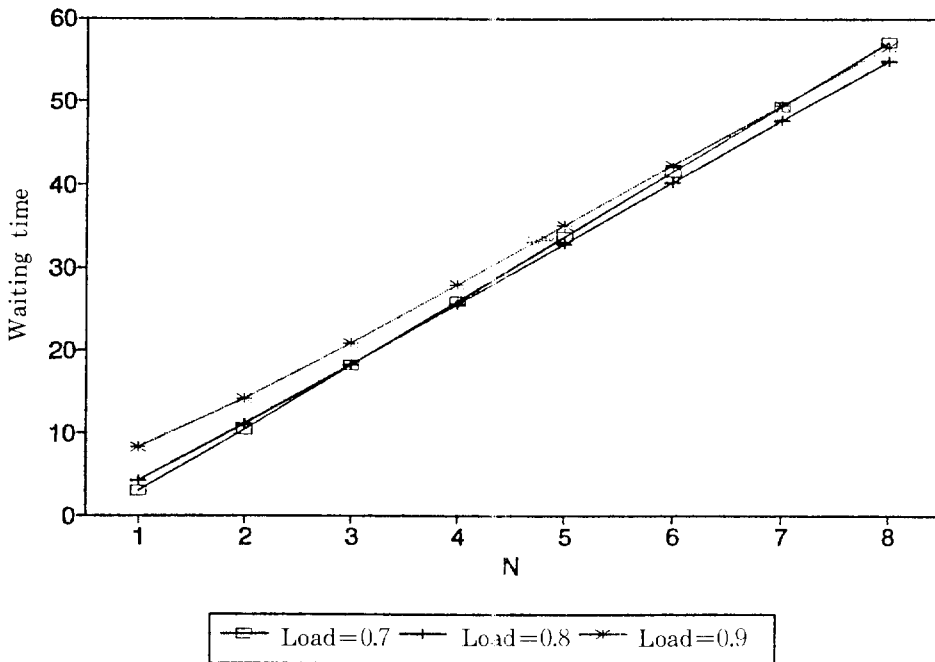
가동률	근사해	시뮬레이션		가동률	근사해	시뮬레이션	
대기 1과 6							
0.1	59.29126	60.41149	± 1.85061	0.6	14.30152	11.95600	± 0.11998
0.2	29.82170	30.81943	± 0.61481	0.7	15.21550	11.24353	± 0.09595
0.3	20.51040	20.85763	± 0.58633	0.8	17.64636	11.61565	± 0.12235
0.4	16.42161	16.09452	± 0.30822	0.9	23.31878	14.83161	± 0.30249
0.5	14.64567	13.48518	± 0.21428				
대기 2와 7							
0.1	299.19949	305.98125	± 19.71490	0.6	52.94018	51.98135	± 2.02193
0.2	149.60426	149.45868	± 10.18269	0.7	47.53305	46.28367	± 1.89142
0.3	100.12788	101.28556	± 5.06258	0.8	44.82681	42.13240	± 1.03197
0.4	75.81244	75.15953	± 3.94325	0.9	45.87399	42.1715	± 1.33729
0.5	61.72229	61.86269	± 1.96208				
대기 3과 8							
0.1	149.22888	152.99051	± 8.42615	0.6	28.59967	27.27745	± 0.57267
0.2	74.68401	74.90015	± 4.76850	0.7	27.09095	24.60062	± 0.53130
0.3	50.28331	51.11568	± 1.42481	0.8	27.53903	23.43398	± 0.28261
0.4	38.57950	38.52064	± 1.60084	0.9	31.45068	25.09499	± 0.45494
0.5	31.15030	31.57601	± 1.01745				
대기 4와 9							
0.1	299.21726	305.70750	± 27.96687	0.6	53.56995	52.40804	± 2.15613
0.2	149.66587	149.36815	± 5.61220	0.7	48.45377	46.13895	± 1.22116
0.3	100.26484	100.68998	± 5.52570	0.8	46.13997	42.47266	± 1.05662
0.4	76.06304	76.36244	± 3.46167	0.9	47.77471	42.20980	± 0.58569
0.5	62.13319	61.17712	± 2.06251				
대기 5와 10							
0.1	299.23150	302.96375	± 28.83650	0.6	54.01382	51.68768	± 2.39462
0.2	149.71223	151.71460	± 10.05968	0.7	49.10386	46.20046	± 1.49572
0.3	100.36474	100.72118	± 4.10870	0.8	47.07895	42.39187	± 1.00478
0.4	76.24248	76.07896	± 3.83524	0.9	49.18914	41.98754	± 1.37165
0.5	62.42409	62.53802	± 2.52197				



[그림 1] 트래픽이 작은 경우의 N에 따른 대기 시간의 변화



[그림 2] 중간 트래픽의 경우의 N에 따른 대기 시간의 변화



[그림 3] 트래픽이 큰 경우의 N에 따른 대기시간의 변화

4. 결 론

본 연구에서는 제어 운영 정책 하에 있는 순환 서비스 시스템의 평균 대기 시간을 구하는 효율적인 알고리즘을 개발하였다. 그리고 개발한 근사 알고리즘의 결과와 시뮬레이션의 결과를 비교하여 정확도를 검정하였다.

〈표 1〉에서 보는 바와 같이 대칭 시스템의 경우에 매우 유용한 알고리즘임을 알 수 있다. 비대칭 시스템의 경우에는 정확도는 떨어지지만, 대기시간의 독립가정을 사용하는 일반적인 알고리즘에서 반영할 수 없는 대기의 위치(location) 요소가 본 알고리즘에서는 반영된다는 장점을 가지고 있다.

본 연구에서 살펴 본 바와 같이, 가동율이 커짐에 따라 평균 대기 시간이 감소하다가 증가한다. 트래픽의 양이 커졌을 때, 평균 대기 시간이 증가하는 것은 대기하는 시간에 대해 제어 파라미터의 영향보다 트래픽의 영향이 더 커졌음을 나타낸다. 따라서 비용(cost)이나 준비 시간(set-up time)을 고려하여, 시스템에 맞는 제어 파라미터를 결정하는데 척도를 제시해 줄 수 있다.

참 고 문 헌

- [1] 이 효성, “제어운영 정책에 있는 집단으로 도착하는 서버 휴가모형의 안정상태 확률,” 「한국경영과학회지」, 제16권, 제2호 (1991), pp.36-48.
- [2] Boxma, O. J., H. Levy and U.Yechiali, “Cyclic Reservation Schemes for Efficient Operation of Multiple-Queue Single Server Systems,” *Annals Oper. Res.*, Vol.35(1995), pp.187-208.
- [3] Browne, S. and U. Yechiali, “Dynamic Priority Rules for Cyclic-Type Queues,” *Adv. Appl. Prob.*, Vol.21(1989) pp.432-450.
- [4] Hofri, M., “Queueing Systems with Procrastinating Server,” *Performance 86 and ACM Sigmetrics 1986, Proc. Joint Conf. Computer Performance Modeling, Measurement and Evaluation*, Raleigh, NC, (1986), pp.245-253.
- [5] Ibe, O. C. and X. Cheng, “Approximate Analysis of Asymmetric Single-Service Token-Passing Systems,” *IEEE Trans. Commun.*, Vol.37, No.6(1989), pp.572-577.
- [6] Kleinrock, L., *Queueing Systems* Volume 1: Theory, John Wiley & Sons, 1975.
- [7] Kuehn, P. J., “Multiqueue Systems with Nonexhaustive Cyclic Service,” *Bell Systems Technical J.*, Vol.58, No.3 (1979), pp.671-698.
- [8] Lee, H. S. and M. M. Srinivasan, “Control Policies for $M^x/G/1$ Queueing System,” *Management Science*, Vol.35, No.6(1989), pp.708-721.
- [9] Levy, H. and M. Sidi, “Polling Systems: Applications, Modeling, and Optimization,” *IEEE Trans. Commun.*, Vol.38, No.10(1990), pp.1750-1760.
- [10] Sarkar, D. and W. I. Zangwill, “Expected Waiting Time for Nonsymmetric Cyclic Queueing Systems-Exact Results and Applications,” *Management Science*, Vol.35, No.12(1989), pp.1463-1475.
- [11] Servi, L. D., “Average Delay Approximation of $M/G/1$ Cyclic Service Queues with Bernoulli Schedules,” *IEEE J. Select. Areas Commun.*, Vol. 4, No.6(1986), pp.813-822.
- [12] Takagi, H., *Analysis of Polling Systems*, Cambridge, MA: M.I.T. Press, 1986.
- [13] Teghem, J., “Control of Service Process in a Queueing Systems,” *European J. Oper. Res.*, Vol.23(1986), pp.141-158.