

붕쇄를 갖는 순환 대기네트워크의 수율 분석*

Throughput Analysis for Cyclic Queueing Networks with Production Blocking*

김호균**, 이창훈***

H.G. Kim**, C.H. Lie***

Abstract

An approximation algorithm is presented for cyclic queueing networks with finite buffers. The algorithm decomposes the queueing network into individual queues with revised arrival and service process and revised queue capacity. Then, each queue is analyzed in isolation. The service process reflects the additional delay a unit might undergo due to blocking and the arrival process is described by a 2-phases Coxian (C_2) distribution. The individual queues are modelled as $C_2/C_2/1/B$ queues. The parameters of the individual queues are computed approximately using an iterative scheme. The population constraint of the closed network is taken into account by ensuring that the sum of the average queue lengths of the individual queues is equal to the number of customers of the network. Extensive numerical experiments show that this method provides a fairly good estimation of the throughput.

1. 서론

붕쇄 대기 네트워크 모형이 생산 시스템, 컴퓨터 시스템, 통신 시스템 및 FMS의 성능 평가를 위한 기법으로 널리 사용되고 있다. 붕쇄 대기 네트워크의 초기연구는 개방형 붕쇄네트워크(OQN-B ; Open Queueing Net-

work with Blocking)에서 시작되었다. OQN-B를 위해 개발된 대부분의 근사방법 [1,4,7,9,10]은 주어진 네트워크를 각각의 대기로 분해하고 그들을 분리하여 분석한다. 따라서 알고리즘들은 주변 대기길이 확률을 구하고 이들로부터 성능 척도를 계산하였다. 한편 폐쇄형 붕쇄네트워크(CQN-B ; Closed Queueing

* 본 논문은 1994년 동의대학교 자체학술연구 조성비의 일부 지원으로 연구되었음

** 동의대학교 산업공학과

*** 서울대학교 산업공학과

Network with Blocking)에 대한 관심은 최근에 고조되고 있다. 폐쇄형에서는 일정한 총 고객수가 네트워크 내를 순환하기 때문에 개방형 네트워크보다 더 복잡하다.

스테이션이 2개로 구성된 지수형 2-스테이션 CQN-B는 총 고객수를 수정하면 비봉쇄 네트워크와 동일한 전이율 행렬을 갖게 되어 승법형 대기길이 분포를 갖게 된다.[1,6,16] 다-스테이션 CQN-B에 대하여, 수율 또는 평균대기길이를 계산하는 알고리즘들이 개발되었다[2,5,14,16]. Suri & Diehl[16]과 Dallery & Frein[5]은 첫번째 대기가 무한 대기공간을 갖는 순환대기 네트워크를 분석하였다. Akyldiz[2]는 모든 대기가 유한한 봉쇄 대기 네트워크의 상태 수와 거의 일치하는 상태수를 갖는 비봉쇄 대기 네트워크를 찾아 그 수율을 봉쇄 대기 네트워크의 수율로 간주하였다. Onvural & Perros [14]는 모든 대기가 유한한 봉쇄 대기 네트워크의 몇몇 알고리즘 수율점을 통과하는 곡선을 적합시켜 총 고객수에 따른 수율 곡선을 제시하였다.

본 논문에서, 봉쇄를 갖는 순환 대기 네트워크의 성능을 분석하기 위한 근사 알고리즘을 제시한다. 이 알고리즘은 개방형 네트워크에 사용되는 분해 방법(decomposition method)에 기초한다. 분해방법은 일반성을 가지고 있으나 여기서는 첫번째 스테이션이 무한 대기능력(buffer capacity)을 갖는 지수형 봉쇄 순환대기 네트워크에 한정한다. 분해된 개별 대기는 $C_2/C_2/1/B$ 대기로 모형화되고 매개변수는 반복적으로 근사시킨다. 분석 과정은 개방 직렬형 네트워크의 분석과정(Jun & Perros[7])과 비슷하다.

2. 봉쇄 순환 대기네트워크

M개 스테이션과 총 고객수 N명을 갖는 순환대기네트워크(그림 1)에서 각 스테이션은 단일 대기과 단일 서버로써 구성된다. 각 스테이션에서의 서비스시간은 서비스율 μ_i 를 갖는 지수분포를 따른다. 첫번째 스테이션은 대기공간이 무한하며 나머지 스테이션들에서는 유한한 대기공간을 갖는다. 스테이션 i는 서버 앞의 공간을 포함하여 대기능력(buffer capacity) B_i 를 갖는다. 이는 스테이션 i에 동시에 있을 수 있는 최대 고객수이다. 관례로 $B_1 = \infty$ 로 둔다. 서비스는 선입선출(FIFO)의 규칙을 따른다. 스테이션 i($i=1, \dots, M-1$)에서 서비스를 받고 난 뒤 다음 스테이션 i+1로 나아간다. 스테이션 M에서 서비스가 끝난 뒤에는 스테이션 1로 되돌아 간다.

대기공간의 유한성 때문에 봉쇄가 발생할 수 있다. 여기서는 생산봉쇄(production blocking)체제[15]를 따른다. 즉 스테이션 i($i=1, \dots, M-1$)에서 서비스가 끝나 다음 스테이션 i+1로 나아갈려고 할 때 다음 스테이션 i+1이 가득차 있을 경우, 그 고객은 스테이션 i+1에 여유대기가 생길 때까지 스테이션 i에 머물러 있어야 한다. 이 시간동안 스테이션 i의 서버는 대기공간에 기다리고 있는 다른 고객에게 서비스를 제공하지 못하고 봉쇄당하고 있다.

총 고객수 및 대기능력에 대한 관계를 살펴본다. $N \leq B_{\min} = \min\{B_1, B_2, \dots, B_M\}$ 일 때는 봉쇄가 발생하지 않는다. $N = B_{\min} + 1$ 일 때는 변형된 승법형 해를 가지며(참조[1]), $N \geq \sum_{i=2}^M B_i + 1$ 일 경우 같은 수율을 가지므로(참조[13]) 총 고객수 N이 $B_{\min} + 2$ 에

서 $\sum_{i=2}^M B_i + 1$ 사이에 있을 때로 한정한다. 붕쇄를 갖는 순환 대기 네트워크를 분석하기 위한 알고리즘은 분해법에 기초한다.

분해방법은 일반적으로 대기 네트워크를 유한한 대기능력을 갖는 개별대기로 분해하여 각 대기의 도착과정, 서비스과정 및 대기능력을 특성화하여 대기 네트워크에서의 각 스테이션의 행태에 가깝게 근사시키는 방법이다. 개별대기에서의 서비스과정은 붕쇄순간 붕쇄당하는 지연시간을 반영하고 도착과정은 C_2 분포로 표현한다. 즉 개별대기를 $C_2/C_2/1/B$ 대기로 특성화 한다. 서비스 및 도착과정의 매개변수들은 Neuts[12]방법이나 Yao & Buzacott[18]방법을 이용하여 반복과정을 거쳐 계산된다.

SIMAN IV를 이용하여 시뮬레이션 모형을 개발하고 알고리즘의 정확도를 검증한다. 다양한 경우를 실험하여 시뮬레이션 결과와 비교한다.

이다.

총 고객수 N 과 같거나 큰 대기능력 B_i 를 갖는 스테이션은 전 스테이션을 붕쇄시키지 않는다. 이들 스테이션 집합을 E 즉, $E = \{i; B_i \geq N\}$ 라 한다. 스테이션 1은 항상 E 에 속한다. 각 대기에 있을 수 있는 수정된 대기능력 B_i^* 는 E 에 속하는 스테이션에 대응되는 대기에서는 N 이 되고 아닐 경우 붕쇄시키는 상태(state)를 표시하기 위하여 대기능력을 하나 증가시켜 $B_i + 1$ 이 된다. 실제로 붕쇄 고객은 붕쇄당한 스테이션에 있지만 다음 대기의 가상(fictitious)위치에 놓는다.

$$B_i^* = \begin{cases} N & ; i \in E \text{ 일 경우} \\ B_i + 1 & ; \text{다를 경우} \end{cases} \quad (1)$$

순환대기 네트워크에서 각 스테이션 수율은 시스템의 수율 λ 과 같다.

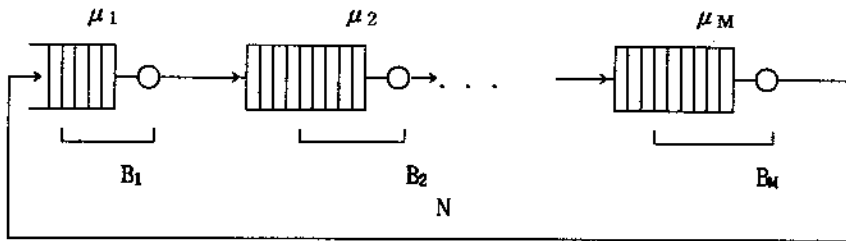


그림 1. 붕쇄 순환 대기 네트워크

3. 근사 알고리즘의 개발

분해방법은 일반적으로 대기 네트워크를 유한한 대기능력을 갖는 개별대기로 분해하여 각 대기의 도착과정, 서비스과정 및 대기능력을 특성화하여 대기 네트워크에서의 각 스테이션의 행태에 가깝게 근사시키는 방법

각 스테이션에서 실제(effective)도착률은 λ 이다. 각 대기에서 도착과정은 붕쇄 결과로 미지의 임의분포를 갖는다. 근사 알고리즘은 도착과정을 단계 1에서 포아송 과정으로 가정하고 단계 2 및 단계 3에서는 C_2 분포로 가정한다. 단계 3에서 분석순서는 단계 1에서와 같지만, 구현(implementation)은 도착과정

이 C_2 분포를 따르므로 같지 않다. 단계 2와 단계 3을 반복한 후 대기 길이 분포가 최종적으로 결정된다.

3.1 단계 1

이 단계는 Dallery & Frein[5]이 제안한 알고리즘에 근거한다. 분해된 각 대기의 도착 과정을 전반적(overall) 도착률 λ_i 를 갖는 포아송 분포로 가정하고, 서비스 과정을 매개변수 $br_i = (b_{i1}, b_{i2}, \beta_i)$ 를 갖는 2-phase Coxian 분포로 나타내 개별대기를 $M/C_2/1/B$ 대기로 근사시켜 스테이션 M부터 후진(backwards)으로 분석한다.

스테이션 i 에서 서비스를 마쳤을 때 스테이션 $i+1$ 이 가득찰 조건부 확률을 π_i 라고 한다. 스테이션 M에서는 봉쇄가 발생하지 못하므로 서비스 과정은 $br_M = (b_{M1}, b_{M2}, \beta_M) = (\mu_M, 0, 0)$ 를 갖는 C_2 분포가 된다. 따라서 Neuts[12]방법을 사용하여 스테이션 M을 $M/C_2/1/B_M^*$ 대기로 분석할 수 있다. 분석하기 위하여 λ_M 을 먼저 알아야 한다. λ_M 은 수율 λ 을 보장하기 위하여 다음의 고정점(fixed point)문제를 반복 계산하여 얻어진다.

$$\lambda_M = \lambda / [1 - P_M(B_M^*)] \quad (2)$$

π_{M-1} 은 Little의 관계식을 대기 M의 가상 위치 B_M^* 에 적용하여 스테이션 M의 대기 길이 분포에서 근사적으로 계산되어 진다.

$$\lambda \pi_{M-1} \cdot \frac{1}{\mu_M} = P_M(B_M^*) \quad (3)$$

여기서 $1/\mu_M$ 은 서버 M-1이 봉쇄당하는 순간 서버 M의 나머지(remaining)서비스 시간

이다.

스테이션 M-1에서 서비스를 마칠 때, 봉쇄가 발생하지 않아 확률 $1 - \pi_{M-1}$ 으로써 스테이션 M에 합류하거나 확률 π_{M-1} 으로써 스테이션 M-1이 봉쇄당할 수 있다. 봉쇄당하는 시간은 봉쇄당하는 순간 스테이션 M에서 서비스를 받고 있는 고객의 나머지(remaining)서비스 시간이다. 지수분포의 과거와 무관한(memoryless)성질 때문에 이 시간은 $1/\mu_M$ 이 된다. 실제(effective)서비스 시간은 매개변수 $br_{M-1} = (b_{M-1,1}, b_{M-1,2}, \beta_{M-1}) = (\mu_{M-1}, \mu_M, \pi_{M-1})$ 를 갖는 C_2 분포로 된다. 스테이션 M-1에서의 실제 서비스 시간의 Laplace 변환, $L_{M-1}^*(s)$ 는 다음 식으로 표현된다.

$$L_{M-1}^*(s) = (1 - \pi_{M-1})T_{M-1}^*(s) + \pi_{M-1} \cdot T_{M-1}^*(s) \cdot T_M^*(s) \quad (4)$$

여기서 $T_i^*(s) = \mu_i / (s + \mu_i)$, $i = M-1, M$ 이다.

따라서 스테이션 M-1을 $M/C_2/1/B_{M-1}^*$ 대기로 분석하여 λ_{M-1} 을 결정하고 대기길이 분포 $P_{M-1}(n)$, $n = 0, 1, \dots, B_{M-1}^*$ 를 얻게 된다. $w_{M-1,j}$, $j = 1, 2$ 를 스테이션 M-2가 봉쇄당하는 순간, 서버 M-1이 j 번째 phase에 있을 조건부 확률이라 한다. $i+1 \in E$ 일 때는 $w_{i1} = 1$, $w_{i2} = 0$ 이 된다. $w_{M-1,j}$ 및 π_{M-2} 는 $P_{M-1}(n)$ 에서 구할 수 있다.

$$w_{M-1,j} = P_{M-1}(B_{M-1,j}) / P_{M-1}(B_{M-1}) \quad (5)$$

여기서 $P_{M-1}(B_{M-1,j})$ 는 스테이션 M-1에 B_{M-1} 명의 고객이 있고 M-1번째 서버가 phase j 에

있을 확률이다. w_M 은 도착시점에서의 확률이지만 PASTA 이론[17]에 의해 안정상태의 확률로 근사시킨다. π_{M-2} 는 스테이션 M에서와 마찬가지로 Little의 관계식을 적용하여 다음 식에 의해 얻어진다.

$$\lambda \pi_{M-2} \cdot [w_{M-1}(\frac{1}{\mu_{M-1}} + \frac{\pi_{M-1}}{\mu_M}) + w_{M-1,2}] = P_{M-1}(B_{M-1}^*) \quad (6)$$

스테이션 M-2를 생각해 보자. 서비스를 마치고 확률 $1-\pi_{M-2}$ 로써 스테이션 M-1에 합류하거나 확률 π_{M-2} 로써 스테이션 M-2가 봉쇄당한다. 서버 M-2가 봉쇄당하는 순간, 서버 M-1이 확률 $w_{M-1,1}$ 로 phase 1의 서비스를 하고 있거나 확률 $w_{M-1,2}$ 로써 phase 2의 서비스를 하고있다. 따라서 스테이션 M-2에서의 실제 서비스 시간은 두 부분으로 구성되며[참조:그림2], 실제 서비스 시간의 Laplace 변환, $L_{M-2}^*(s)$ 은 다음 식으로 얻어진다.

$$L_{M-2}^*(s) = (1-\pi_{M-2})T_{M-2}^*(s) + \pi_{M-2} T_{M-2}^*(s) \cdot [w_{M-1,1}L_{M-1}^*(s) + w_{M-1,2} \frac{b_{M-1,2}}{s+b_{M-1,2}}] \quad (7)$$

실제 서비스 시간의 분포를 모멘트(moment)

method)방법[3,7,11]을 사용하여 C_0 분포로 표현할 수 있다. 구현을 쉽게하고 소프트웨어 개발을 가능하게 하기 위하여 2-phase Coxian 분포로 변형한다. 즉 $L_{M-2}^*(s)$ 를 다음으로 근사시킨다.

$$L_{M-2}^*(s) \cong (1-\beta_{M-2}) \frac{b_{M-2,1}}{s+b_{M-2,1}} + \beta_{M-2} \frac{b_{M-2,1}}{s+b_{M-2,1}} \cdot \frac{b_{M-2,2}}{s+b_{M-2,2}} \quad (8)$$

여기서 매개변수 $br_{M-2} = (b_{M-2,1}, b_{M-2,2}, \beta_{M-2})$ 는 모멘트 방법에 의해 결정된다. 스테이션 M-2가 $M/C_2/1/B_{M-2}^*$ 대기로 분석된다. 이와 같은 과정을 스테이션 1까지 후진(backwards)으로 수행한다.

단계 1의 과정이 다음으로 요약된다.

단계 1 :

1. $i = M$
2. $L_i^*(s)$ 및 br_i 를 얻는다.
 - $i+1 \in E$ 일 때, $L_i^*(s) = T_i^*(s)$
 - $i+1 \notin E$ 일 때, $L_i^*(s) = (1 - \pi_i) T_i^*(s) + \pi_i T_i^*(s)$.

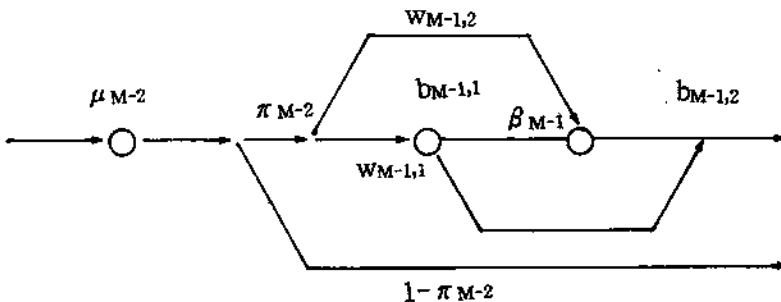


그림 2. 스테이션 M-2에서의 실제 서비스 시간

$$[w_{i+1,1}L_{i+1}^*(s) + w_{i+1,2} \frac{b_{i+1,2}}{s+b_{i+1,2}}]$$

특히 $i+2 \in E$ 일 때, $b_{r_1} = (b_{i1}, b_{i2}, \beta_i) = (\mu_i, \mu_{i+1}, \pi_i)$ 로 된다.

$i+1 \in E$ 일 때, 봉쇄당하지 않으므로 서비스 시간은 지수분포를 따른다.

$$\text{즉 } b_{r_1} = (\mu_i, \epsilon, 0)$$

3. 스테이션 i 를 $M/C_2/1/B_1^*$ 대기로 분석하여, $P_i^{(n)}$, $n=0,1,\dots,B_1^*$, w_{ij} , $j=1,2$ 및 π_{i-1} 을 계산한다.

$i+1 \in E$ 일 때, $w_{i1}=1$, $w_{i2}=0$

$i+1 \notin E$ 일 때, $w_{ij} = P_i(B_{ij}) / P_i(B_i)$

$i \in E$ 일 때, $\pi_{i-1}=0$

$i \notin E$ 일 때, $\lambda \pi_{i-1} \cdot [w_{i1}(\frac{1}{b_{i1}} + \frac{\beta_i}{b_{i2}}) + w_{i2} \frac{1}{b_{i2}}] = P_i(B_{i+1})$

4. $i>1$ 이면 $i=i-1$ 로 놓고 2로 간다.

3.2 단계 2

단계 2에서는 각 대기의 실제 서비스 시간은 단계 1에서와 같다고 가정하고 도착과정을 C_2 분포로 근사시켜 스테이션 1부터 분석한다. $d_i(t)(D_i^*(s))$, $e_i(t)(E_i^*(s))$ 및 $O_i(t)(O_i^*(s))$ 를 각각 스테이션 i 에서의 출발간(interdeparture)시간, 실제 도착간(effective interarrival)시간 및 전반적(overall) 도착간 시간의 pdf(그의 Laplace 변환)라 한다.

스테이션 1의 도착과정은 스테이션 M 의 출발과정에 의해 결정된다. 스테이션 M 에서의 출발간 시간의 pdf, $d_M(t)$ 를 알아야 한다. 연속하는(successive) 실제 서비스가 서로 독립이라 가정하면 그의 Laplace 변환 $D_M^*(s)$ 를 구할 수 있다. 한 명의 고객이 스테이션 M 을 떠날 때, 스테이션 M 에서 기다리고 있는 고

객이 있어 곧 서비스가 시작되거나 또는 기다리는 고객이 없을 수 있다. 앞의 경우에, 스테이션 M 에서 다음 출발까지의 시간은 실제 서비스 시간과 같은 분포를 한다. 고객이 없을 경우에는 다음 고객이 도착할 때까지의 시간간격과 그 고객의 서비스 시간간격과의 합이다. 따라서 $D_M^*(s)$ 가 다음 식으로 표현된다.

$$D_M^*(s) = (1-\delta)L_M^*(s) + \delta \frac{\lambda_M}{s+\lambda_M} L_M^*(s) \quad (9)$$

여기서 δ 는 스테이션 M 을 떠나는 고객이 스테이션 M 에 고객이 없음을 발견하는 확률이며, $P_M(n, j)$ 에서 얻어진다.

$$\delta = [(1-\beta_M)b_{M,1}P_M(1,1) + b_{M,2}P_M(1,2)] /$$

$$[(1-\beta_M)b_{M,1} \sum_{n=1}^{B_M} P_M(n,1) + b_{M,2} \sum_{n=1}^{B_M} P_M(n,2)] \text{이다.}$$

위의 출발간 과정의 Laplace 변환은 사실상 스테이션 1에서의 실제 도착간 시간의 Laplace 변환, $E_1^*(s)$ 과 같다. $E_1^*(s)$ 에서 $e_1(t)$ 의 1차, 2차 및 3차 모멘트 $m_{e_1,1}$, $m_{e_1,2}$, $m_{e_1,3}$ 를 얻는다.

스테이션 1을 개별 대기로 분석하려면 실제 도착과정 $e_1(t)$ 보다는 전반적(overall) 도착과정 $O_1(t)$ 가 필요하다. $e_1(t)$ 로부터 $O_1(t)$ 가 근사적으로 얻어질 수 있다. 재생 과정(renewal process)의 분해[8]를 사용하여, 다음 관계식을 이용한다.

$$E_1^*(s) = \frac{(1-f_1)O_1^*(s)}{1-f_1O_1^*(s)} \quad (10)$$

여기서 f_1 은 도착고객이 스테이션 1이 가득 차 있어 놓쳐 버린(lost) 확률이다. $q_1(n)$ 을 도

착고객이 대기 i 에 n 명을 발견할 확률이라 하면 $f_i = q_i(B_i^*)$ 이다. 이 확률은 시간 평균 확률 $P_i(n)$ 에서 얻어진다[7]. $O_i(t)$ 의 모멘트, $m_{o_i,1}, m_{o_i,2}, m_{o_i,3}$ 는 f_i 및 $e_i(t)$ 의 모멘트로써 표현할 수 있다.

$$\begin{aligned} m_{o_i,1} &= (1-f_i)m_{e_i,1} \\ m_{o_i,2} &= (1-f_i)(m_{e_i,2} - 2f_i m_{e_i,1}^2) \\ m_{o_i,3} &= (1-f_i)(m_{e_i,3} - 6f_i m_{e_i,2} \cdot m_{e_i,1} + 6f_i^2 m_{e_i,1}^3) \end{aligned} \quad (11)$$

모멘트 방법을 사용하여 $O_i(t)$ 를 C_2 분포로 표현하면, 스테이션 1을 매개변수 $ar_1 = (a_{21}, a_{22}, \alpha_1)$ 및 $br_1 = (b_{21}, b_{22}, \beta_1)$ 를 갖는 $C_2/C_2/1/B_1^*$ 대기로 분석할 수 있다. 여기서 $O_i(t)$ 의 모멘트는 미지인 f_i 의 함수이다. 단계 1에서 λ_i 를 결정할 때와 마찬가지로 대기 1의 수율이 λ 를 보장하는 고정점 문제, $1/m_{o_i,1} = \lambda_i/(1-f_i)$ 를 Yao & Buzacott[18]의 반복과정을 사용하여 수치적으로 풀어 f_i 를 결정한다. f_i 가 결정되면 $P_i(n)$ 이 구해진다.

스테이션 2를 분석하려면 $d_i(t)(D_i^*(s))$ 를 얻어야 한다. $D_i^*(s)$ 는 스테이션 1에서와 마찬가지로 연속하는 실제 서비스가 서로 독립이라 가정하면 다음 식을 갖는다.

$$\begin{aligned} D_i^*(s) &= \delta_1 O_i^*(s) L_i^*(s) + \delta_2 \frac{a_{12}}{s+a_{12}} L_i^*(s) \\ &+ (1-\delta_1 - \delta_2) L_i^*(s) \end{aligned} \quad (12)$$

여기서 $\delta_j, j=1, 2$ 는 서비스가 끝난 시점에서 스테이션 1은 비고(empty) 도착고객이 도착과정의 j 번째 phase에 있을 확률이며 안정상태 확률로부터 얻어질 수 있다. 대기 1

을 특성화할 때와 같은 과정을 수행하면 $O_2(t)$ 를 C_2 분포로 근사시킬 수 있다. 따라서 스테이션 2가 매개변수 $ar_2 = (a_{21}, a_{22}, \alpha_2)$ 및 $br_2 = (b_{21}, b_{22}, \beta_2)$ 를 갖는 $C_2/C_2/1/B_2^*$ 대기로 분석 가능하다. 이 과정을 마지막 스테이션 M 까지 수행하여 단계 2에서의 대기길이 확률 $P_i^{(2)}(n), n=0,1,\dots,B_i^*$ 를 얻게된다. 이 과정을 요약하면 다음과 같다.

단계 3에서 대기길이 분포가 수렴되지 않아 단계 2를 다시 반복할 때, 스테이션 M 의 도착과정이 C_2 분포로 주어졌으므로 $D_M^*(s)$ 가 다음 식으로 변화된다.

$$\begin{aligned} D_M^*(s) &= \delta_1 O_M^*(s) L_M^*(s) + \delta_2 \frac{a_{M2}}{s+a_{M2}} L_M^*(s) \\ &+ (1-\delta_1 - \delta_2) L_M^*(s) \end{aligned} \quad (13)$$

단계 2

1. $i=1$
2. $E_i^*(s), O_i^*(s)$ 및 ar_i 를 구축한다.

$$D_{i-1}^*(s) = E_i^*(s) \cong \frac{(1-f_i)O_i^*(s)}{1-f_i O_i^*(s)}$$

$$\begin{aligned} m_{o_i,1} &= (1-f_i)m_{e_i,1} \\ m_{o_i,2} &= (1-f_i)(m_{e_i,2} - 2f_i m_{e_i,1}^2) \\ m_{o_i,3} &= (1-f_i)(m_{e_i,3} - 6f_i m_{e_i,2} \cdot m_{e_i,1} + 6f_i^2 m_{e_i,1}^3) \Rightarrow ar_i = (a_{21}, a_{22}, \alpha_1) \end{aligned}$$

단, 처음 반복할 때

$$E_i^*(s) = (1-\delta) L_M^*(s) + \delta \frac{\lambda_M}{s+\lambda_M} L_M^*(s) \text{ 이다.}$$

3. 스테이션 i 를 $C_2/C_2/1/B_i^*$ 대기로 분석하여, $P_i^{(2)}(n)$ 및 $D_i^*(s)$ 를 얻는다.

f_i 를 결정하고 $P_i^{(2)}(n)$ 을 계산한다.

$$D_i^*(s) = \delta_1 O_i^*(s) L_i^*(s) + \delta_2 \frac{a_{i2}}{s+a_{i2}} L_i^*(s) + (1-\delta_1-\delta_2) L_i^*(s)$$

4. $i = M$ 이면 끝내고, 아닐 경우 $i=i+1$ 로 놓고 2로 간다.

3.3 단계 3

각 대기의 실제(effective)도착 과정, $e_i(t)$ 는 단계 2에서 구한 결과와 같다고 하고 실제 서비스시간을 다시 계산한다. 분석과정은 개념적으로 단계 1과 동일하지만 도착과정을 포아송 분포가 아니고 C_2 분포로 근사시키기 때문에 조정되어야 한다. $w_{i+1,j}$ 를 구하는 식 (5)가 다음으로 변경되어야 한다.

$$w_{i+1,j} = q_{i+1}(B_{i+1,j}) / q_{i+1}(B_{i+1}) \quad (14)$$

여기서 $q_{i+1}(B_{i+1,j})$ 는 도착하기 직전에 $i+1$ 번째 서버가 phase $j(j=1,2)$ 에 있고 스테이션 $i+1$ 에 B_{i+1} 명의 고객이 있을 확률이고 $q_{i+1}(B_{i+1})$ 은 도착하기 직전 스테이션 $i+1$ 에 B_{i+1} 명의 고객이 있을 확률이다. 이 확률들은 대기 $i+1$ 의 대기 길이 분포에서 얻어질 수 있다. 다음 스테이션 $i+1 \in E$ 일 때는 서비스를 마친 후에 봉쇄가 발생하지 않으므로 $L_i^*(s)$ 및 br_i 가 변화되지 않고 대기 길이 확률 $P_i^{(3)}(n)$ 은 단계 2에서 구한 $P_i^{(2)}(n)$ 과 같게 된다. $i+1 \notin E$ 일 때는 $P_i^{(3)}(n)$ 을 다시 구하여야 한다. 이를 위해 실제(effective)도착과정보다는 전반적(overall) 도착과정을 알아야 한다. 단계 2에서 구한 실제 도착과정, $e_i(t)$ 는 단계 3에서 변화되지 않는다고 가정한다. 전반적 도착과정 $O_i(t)$ 의 모멘트는 식(11)에 의해 f_i 의 함수로 표현된다. 단계 3에서 실제 서버

스 시간이 재계산되기 때문에 f_i 가 변화가능하다. 대기 i 의 수율이 λ 을 보장하도록 f_i 를 다시 결정하고 $P_i^{(3)}(n)$ 을 계산한다.

앞의 과정이 끝난 후 수렴도를 점검한다. $P_i^{(3)}(n)$ 이 $P_i^{(2)}(n)$ 에 충분히 접근할 때까지 단계 2 및 단계 3을 반복한다. 위의 과정을 요약하면 다음과 같다.

단계 3

1. $i=M$
2. $i+1 \in E$ 일 때, $P_i^{(3)}(n) = P_i^{(2)}(n)$ 이고 π_{i-1} 를 다시 계산한다.
 $i+1 \notin E$ 일 때, $L_i^*(s)$ 및 br_i 를 다시 계산한다.

$$L_i^*(s) = (1 - \pi_i) T_i^*(s) + \pi_i T_i^*(s) \cdot$$

$$[w_{i+1,1} L_{i+1}^*(s) + w_{i+1,2} \frac{b_{i+1,2}}{s+b_{i+1,2}}]$$

여기서 $w_{i+1,j} = q_{i+1}(B_{i+1,j}) / q_{i+1}(B_{i+1})$

3. 단계 2에서 구한 $E_i^*(s)$ 를 사용하여 스테이션 i 를 $C_2/C_2/1/B_i^*$ 대기로 분석하여 $1/m_{o_2,1} = \lambda_j(1-f_i)$ 를 만족시키는 f_i 를 결정하고 $P_i^{(3)}(n)$, $n=0,1,\dots,B_i^*$, w_{ij} , $j=1,2$ 및 π_{i-1} 을 다시 계산한다.

※ f_i 가 변화가능

4. $i>1$ 이면 $i=i-1$ 로 놓고 2로 간다.
5. 수렴도를 점검한다.

$$\max_n |P_i^{(3)}(n) - P_i^{(2)}(n)| < \epsilon, i=1,2,\dots,M$$

만족되면 끝내고, 아닐 경우 단계 2로 간다.

폐쇄형 대기 네트워크에서는 각 대기 에 있는 평균 고객 수, Q_i 의 합이 대기 네트워크 내에 있는 총 고객수 N 과 같아야 한다. E 에 속하지 않는 대기에서는, 봉쇄 상태를 나타내기 위하여 대기능력을 하나 증가 시켰으므로 스테이션 i 가 스테이션 $i-1$ 을 봉쇄시킬 때 i 대기는 B_i 명 있게 된다.

$$\sum_{i=1}^M Q_i = Q^* \quad (15)$$

$$\begin{aligned} \text{여기서 } Q_i &= \sum_{n=0}^N n P_i(n); i \in E \text{ 일 경우} \\ &\sum_{n=0}^{B_i} n P_i(n) + B_i P_i(B_i+1) \\ &\quad ; \text{다를 경우} \end{aligned}$$

$Q^*=N$ 이 되도록 수율 λ 이 다음과 같이 세 롭게 된다. $\sum_{i=1}^M Q_i > N$ 일 때 추정 산출률 $\hat{\lambda}$ 는 감소되어야 하고 다를 경우 증가 시킨다. 이 과정을 N 에 가깝게 될 때까지 반복한다. 근사 알고리즘의 전 과정은 다음으로 요약 된다.

계산 알고리즘

초기화 단계 :

수율 λ 에 대한 경계치를 정한다.

$$\lambda_{\min} = 0 \ \& \ \lambda_{\max} = \min\{\mu_1, \mu_2, \dots, \mu_M\}$$

반복단계 :

단계 0 : 수율의 추정치 $\hat{\lambda}$ 을 선택한다.

$$\hat{\lambda} = (\lambda_{\min} + \lambda_{\max}) / 2$$

단계 1 : $M/C_2/1/B$ 대기로 특성화

단계 2 : $C_2/C_2/1/B$ 대기로 특성화

단계 3 : $C_2/C_2/1/B$ 대기로 특성화

단계 4 : \cdot 평균대기길이 $Q_i, i=1, \dots, M$ 을

계산한다.

$$i \in E \text{ 일 때, } Q_i = \sum_{n=0}^N n P_i(n)$$

$$i \notin E \text{ 일 때, } Q_i = \sum_{n=0}^{B_i} n P_i(n) + B_i P_i(B_i+1)$$

$$\cdot Q = \sum_{i=1}^M Q_i$$

단계 5 : $\cdot Q$ 가 N 에 충분히 접근하면, 끝낸다.

\cdot 아닐 경우

$Q > N$ 이면, $\lambda_{\max} = \hat{\lambda}$ 로 놓고

$Q < N$ 이면, $\lambda_{\min} = \hat{\lambda}$ 로 놓아

단계 1로 간다.

위의 알고리즘은 바깥쪽(outer)루프와 3개의 안쪽(inner)루프(loop)로 구성된다. 바깥쪽루프는 각 대기 에 있는 전체 고객수가 총 고객수에 접근되도록 λ 를 찾고 안쪽루프(단계 1 ~ 단계 3)는 $\hat{\lambda}$ 을 보장시키는 λ_1 및 f_1 값을 찾는 것이다. 즉 Q^* 는 $\hat{\lambda}$ 의 비선형 함수, $Q^*=F(\hat{\lambda})$ 이고, $\hat{\lambda}$ 은 λ_1 및 f_1 의 비선형 함수 $\hat{\lambda}=G_1(\lambda_1)$ 및 $\hat{\lambda}=H_1(f_1)$ 이다. 따라서 바깥쪽루프는 $\hat{\lambda}$ 이 미지(unknown)일 때 $F(\hat{\lambda}) = N$ 의 해를 구하는 반면 안쪽루프는 λ_1 및 f_1 가 미지일 때 $G_1(\lambda_1)=\hat{\lambda}$ 및 $H_1(f_1) = \hat{\lambda}$ 의 해를 구한다. 알고리즘이 끝난 후, 실제확률 $P_i(B_i), i \notin E$ 는 $P_i(B_i)$ 및 $P_i(B_i^*)$ 의 합으로 얻어진다.

4. 실험 결과

제시한 근사 방법의 정확도를 검증하기 위

하여, 스테이션 수가 3개, 5개, 7개 일 때 다양한 경우에 대하여 수율을 시험하고 결과를 비교한다. 각 경우에 네트워크는 스테이션 수 M , 총 고객수 N , 대기공간 $B=(B_1, B_2, \dots, B_M)$ 및 서비스를 $\mu=(\mu_1, \mu_2, \dots, \mu_M)$ 로써 기술된다. 검증할 예제는 7개 군(표 1)으로 구성되고 46개 경우를 시험하였다. 3-스테이션 네트워크(예제 I, II, III)에서는 5가지 서비스를 고려하여 분석결과를 구하였고 분석결과는 정밀해와 비교된다. 예제 IV, V, VI은 5-스테이션 네트워크를 예제 VII은 7-스테이션 네트워크를 고려하였고 근사 분석결과는 정밀해가 가용하지 못하므로 시뮬레이션 결과와 비교된다.

표 2. 결과 비교의 요약

스테이션 수	검증 경우의 수	척도	분석해
3	15	평균 상대오차	-4.0%
		최대 상대오차	-7.5%
5	21	평균 상대오차	-4.7%
		최대 상대오차	-7.2%
7	10	평균 상대오차	-3.1%
		최대 상대오차	-5.9%

변화 기울기가 거의 완만해져 정확도가 낮아 지므로 제외 시켰다. 그림 3~5는 총 고객수 N 에 따른 수율의 변화를 나타낸다.

본 알고리즘은 수율을 과소 평가한다. 이

표 1. 검증 네트워크의 예제

스테이션수	총고객수	대기공간	서비스율	예제
3	4	$B=(\infty, 1, 3)$	$\mu=(0.1, 0.2, 0.2)$	I
		$B=(\infty, 3, 1)$	$\mu=(0.1, 0.2, 0.1)$	II
	5	$B=(\infty, 1, 3)$	$\mu=(0.1, 0.1, 0.2)$ $\mu=(0.1, 0.1, 0.1)$ $\mu=(0.1, 0.5, 0.1)$	III
5	5~11	$B=(\infty, 3, 3, 3, 3)$	$\mu=(1, 1, 1, 1, 1)$	IV
		$B=(\infty, 4, 2, 4, 2)$	$\mu=(1, 1, 1, 1, 1)$	V
		$B=(\infty, 2, 4, 2, 4)$	$\mu=(0.75, 1, 0.75, 1, 0.75)$	VI
7	5~14	$B=(\infty, 3, 3, 3, 3, 3, 3)$	$\mu=(1, 1, 1, 1, 1, 1, 1)$	VII

총 46개 경우의 실험결과 수율의 평균 상대 오차, 최대 상대오차가 각각 -3.1%, -5.9%이었다.(참조:표 2.) 스테이션 수와 고객의 수가 증가하여도 결과의 형태는 변화되지 않았다. 총 고객수 N 은 $B_{\min}+2$ 에서 $\sum_{i=2}^M B_i+1$ 까지 변화될 수 있으나, 포화상태 즉 $\sum_{i=2}^M B_i+1$ 근처에서는 총 고객수의 증가에 따른 수율의

는 스테이션 $i \in E$ 의 대기 공간을 $B_i^*=N$ 으로 수정한 결과 $n_i = N-1$ 에서 $n_i = N$ 으로 상태가 바뀔 때 실재는 봉쇄시키지 않으나 봉쇄시키는 것으로 해석되어, $P_i(B_i^*)$ 가 과대 평가되고 수율은 과소 평가되었다.

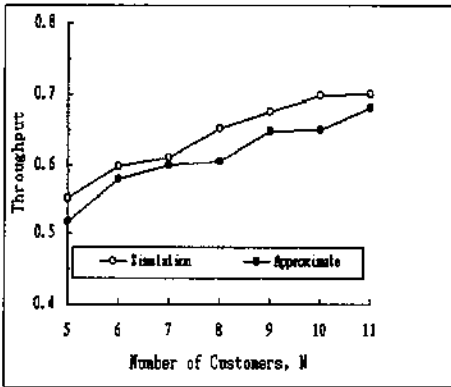


그림 3. 총 고객수에 따른 수율 (예제Ⅳ)

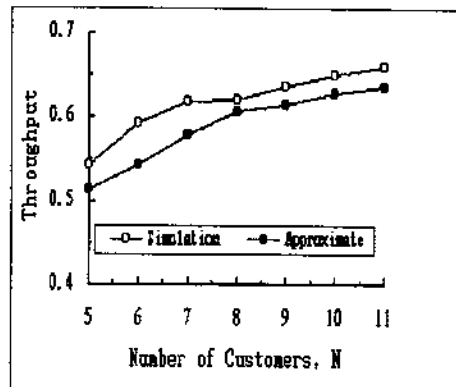


그림 4. 총 고객수에 따른 수율(예제Ⅴ)

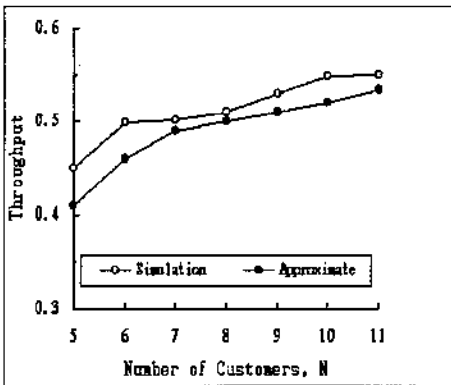


그림 5. 총 고객수에 따른 수율 (예제Ⅵ)

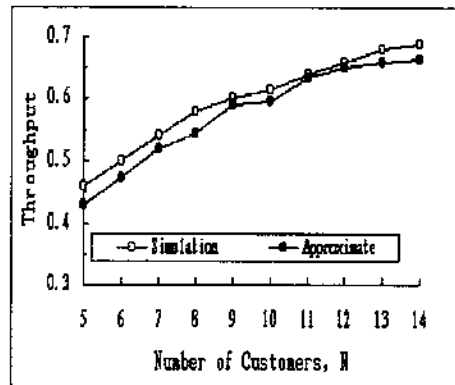


그림 6. 총 고객수에 따른 수율(예제Ⅶ)

5. 결론

대기 네트워크 모형은 생산 시스템, 컴퓨터 시스템 및 통신망을 모형화 하는데 사용되고 있다. 실제로 많은 대기 시스템이 유한한 대기공간을 갖는다. 유한한 대기공간 또는 붕쇄를 갖는 대기 네트워크를 분석하는 방법은 일반적으로 근사방법이 사용된다. 개방형 붕쇄 대기 네트워크를 위한 알고리즘은 많이 개발되었으나 폐쇄형 붕쇄 대기 네트워크에서는 적었다.

본 연구에서는 붕쇄를 갖는 지수형 순환

대기 네트워크의 분석을 위한 근사 알고리즘이 제시되었다. 알고리즘은 시스템을 개별 대기로 분해하고 시스템에서의 각 스테이션의 행태에 가깝도록 각 대기의 도착과정, 서비스과정 및 대기공간을 특성화한다. 각 대기에 있는 평균 고객수의 합이 대기 네트워크 내에 있는 총 고객수와 같게 함으로써 폐쇄형 네트워크의 일정 고객수 조건을 만족시킨다. 첫번째 스테이션이 무한 대기 공간을 갖는 순환 대기 네트워크에 알고리즘을 적용하였다. 수율에 대한 결과는 좋은 결과를 보여준다.

본 연구에서는 첫번째 대기공간이 무한하고 지수 서비스 분포를 갖는 순환 대기 네트워크를 분석하였다. 서비스 분포가 일반 분포를 따를 경우 본 알고리즘이 쉽게 연장될 수 있다. 분해 방법은 일반적인 접근법이고 큰(large)시스템에도 적용가능하다. 모든 상태이전이 유한한 대기공간을 가질 경우와 다른 봉쇄체제(서비스 봉쇄, 기각봉쇄) 및 일반적 위상을 갖는 경우가 추후 연구대상이다. 또한, 서버는 서비스를 완전하게 제공한다고 가정하였으나 현실적으로 고장이 발생되고 수리가 이루어지기도 한다. 서버의 신뢰성도 분석모형에 포함되어야 할 것이다.

참 고 문 헌

- [1] Akyildiz, I. F., "Exact Product Form Solution for Queueing Network with Blocking," IEEE Trans. Computers, C-36, pp.122-125, 1987.
- [2] Akyildiz, I. F., "On the Exact and Approximate Throughput Analysis of Closed Queueing Networks with Blocking," IEEE Trans. Software Engineering, SE-14, pp.62-70, 1988.
- [3] Altioik, T., "On the Phase-Type Approximations of General Distributions," IIE Trans. 17, pp.110-116, 1985.
- [4] Altioik, T. and H.G. Perros, "Approximate Analysis of Arbitrary Configurations of Open Queueing Networks with Blocking," Annals of Operation Research 9, pp. 481-509, 1987.
- [5] Dallery, Y. and Y. Frein, "A Decomposition Method for the Approximate Analysis of Closed Queueing Network with Blocking," In : H.G. Perros and T. Altioik, Ed., Proc. of the 1st Int. Workshop on Queueing Networks with Blocking, NCSU, Raleigh, NC, pp.193-215, 1989.
- [6] Gordon, W.J. and G.F. Newell, "Cyclic Queueing Systems with Restricted Length Queues," Operations Research 15, pp. 266-277, 1967.
- [7] Jun, K. P. and H. G. Perros, "An Approximate Analysis of Open Tandem Queueing Networks with Blocking and General Service Times," European Journal of Operational Research 46, pp. 123-135, 1990.
- [8] Kuehn, P. J., "Approximate Analysis of General Networks by Decomposition," IEEE Trans. Communications, COM-27, pp.113-126, 1979.
- [9] Lee, H.S., and S.M. Pollock, "Approximate Analysis for the Merge Configuration of an Open Queueing Networks Blocking," IIE Trans. 21, pp.122-129, 1989.
- [10] Lee, H.S., and S.M. Pollock, "Approximate Analysis of Open Acyclic Exponential Queueing Networks with Blocking," Operations Research 38, pp.1123-1134, 1990.
- [11] Marie, R., "An Approximate Analytical Method for General Queueing Networks," IEEE Trans. Software Engineering, SE-5, pp.530-538, 1979.
- [12] Neuts, M.F., MATRIX-GEOMETRIC

- SOLUTIONS IN STOCHASTIC MODELS,
The Johns-Hopkins Uni. Press, 1981.
- [13] Onvural, R.O. and H.G. Perros, "Equivalencies Between Open and Closed Queueing Networks with Finite Buffers," In: Y. Takahashi, Ed., Proc. Int. Seminar on Performance of Distributed and Parallel Systems, Kyoto, Japan, pp.213-228, 1988.
- [14] Onvural, R. O. and H. G. Perros, "Approximate Throughput Analysis of Cyclic Queueing Networks with Finite Buffers," IEEE Trans. Software Engineering, SE-15, pp.800-808, 1989.
- [15] Onvural, R. O., "Survey of Closed Queueing Networks with Blocking," ACM Computing Surveys 22, pp.83-121, 1990.
- [16] Suri, R. and G. W Diehl, "A Variable Buffer-Size Model and Its Use in Analyzing Closed Queueing Networks with Blocking," Management Science 32, pp. 206-224, 1986.
- [17] Wolff, R. W., "Poisson Arrivals See Time Averages," Operations Research 30, pp. 223-231, 1982.
- [18] Yao, D.D. and J.A. Buzacott, "Queueing Models for a Flexible Machining Station Part II: The Method of Coxian Phases," European Journal of Operational Research 19, pp.241-252, 1985.