

비고갈 서비스 하에서의 대칭적 폴링시스템의 순환시간에 대한 분석

이명용* · 강석호*

An Analysis on the Cycle Time of a Symmetric Polling System with Nonexhaustive Service

Myeong Yong Lee* and Suk Ho Kang*

Abstract

Queueing models with cyclic service can be applied in the area of performance studies of polling systems in data communication and switching systems. We provide a new approach for analysis on the cycle time of the multiqueue system consisting of statistically identical stations where at most one message is served for any station in a polling cycle.

We assume an independent Poisson arrival at each station. The message service time and switch-over time are assumed to be constant. Based on the dependence between the stations, cycle time distribution is derived. Then, this result is validated by a computer simulation.

1. 서론

폴링(polling)은 순환적 순서(cycle order)에 의하여 여러 사용자에게 서비스를 제공하는 방법으로서, 대기 이론의 관점에서 폴링 시스템은 multiqueue, cyclic service 시스템이다. 따라서, 폴링 시스템은 토큰패싱 프로토콜에 따라 운영되는 근거리 통신망(LAN)의 토큰링 및 토큰버스 네트워크의 수학적 모형화에 많이 이용된다.

본 연구에서는 하나의 서어버가 링 구조의

네트워크에 접속된 여러 스테이션들을 순환적인 순서로 방문할 때마다 스테이션에 대기 중인 메시지들 중에서 최대 하나의 메시지를 전송하는 nonexhaustive(또는 single) 서비스 원칙을 따르는 모형을 다룬다. 또한, 고려한 모형은 모든 스테이션에서의 메시지 도착율, 서비스율, switchover time 이 같은 대칭적시스템이다. Switchover time(또는 walking time, reply interval)이란 서어버가 한 스테이션에서 인접한 다음 스테이션으로 이동하는 데 걸리는 시간을 나타낸다.

폴링 시스템에 대한 기존의 연구들은 Leibowitz[9], Cooper와 Murray[2], Hashida와 Ohara[5], Eisenberg[3], Kuehn[8], Ferguson과 Aminetzah[4], Takagi[11], Pang과 Donaldson[10], Ibe와 Cheng[7] 등에 의하여 활발히 수행되어 왔다. 이러한 연구들은 시스템의 성능을 나타내는 평균 순환시간, 평균 메세지 대기시간 등의 수행도 평가지수(performance measure)들을 구하는 방법들을 언급하고 있다. 순환시간이란 임의의 스테이션이 폴링된 시점부터 다시 폴링될 때까지의 시간을 의미한다. Exhaustive 및 gated 서비스에 대해서는 비대칭적인 경우에도 정확한 해를 Ferguson과 Aminetzah[4]가 제시하였다. 그러나, nonexhaustive 서비스에 대해서는 그 동안의 많은 연구에도 불구하고 스테이션의 수가 2개인 경우에만 정확한 해를 제시하고 있다[1, 3].

본 연구에서는 nonexhaustive 폴링 시스템의 특수한 경우, 즉 모든 스테이션들이 동일한 대칭적 시스템에서 메세지 서비스 시간과 switchover time이 일정한 상수 값을 갖을 때 순환시간의 분포를 분석적으로 구하여 수행도 평가지수 중의 하나인 순환시간의 평균과 분산을 산출하는 새로운 접근 방법을 제시하고자 한다.

2. 모형설명

본 연구에서 다루는 모형은 다음과 같다.

- 대칭적 시스템
- nonexhaustive 서비스

- 이산 시간
- 무한 대기용량
- Poisson 도착

이러한 모형에서 사용되는 주요 변수들을 다음과 같이 정의한다.

g : 네트워크 상의 총 스테이션 수

λ_i : 스테이션 i 의 메세지 도착율($i=1$ to g)

h_i : 스테이션 i 의 평균 메세지 서비스 시간 ($i=1$ to g)

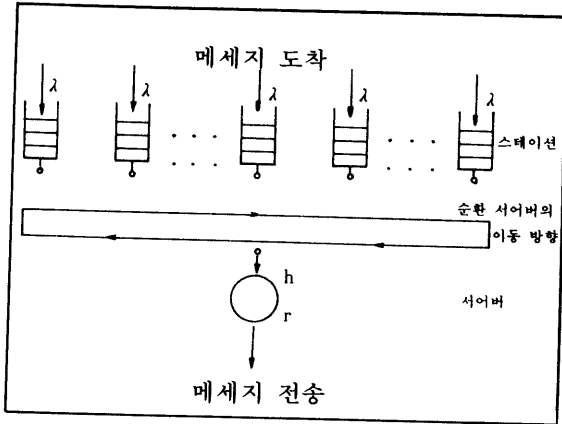
r_i : 스테이션 i 의 스테이션 ($i+1$)로의 switchover time($i=1$ to g)

(단, 스테이션 ($g+1$)은 스테이션 1이다)

기타 변수들에 대한 정의는 필요시에 하기로 한다. 또한, 접근 방법의 개발 과정에 필요한 가정들은 다음과 같다.

- (1) 외부로부터 도착하는, 메세지는 모든 스테이션이 평균 λ 의 동일한 도착율을 갖는 Poisson 과정을 따른다.
- (2) 모든 스테이션의 평균 메세지 서비스 시간은 상수값 h 이다. 즉 도착하는 모든 메세지들의 길이는 동일하다.
- (3) 다음 스테이션으로의 서어버가 이동하는 시간은 모두 상수값 r 로 동일하다.
- (4) 네트워크는 에러 없이 정상적으로 운영된다.

이상의 모형을 그림으로 나타내면 [그림 1]과 같다.



[그림 1] 순환 대기 모형

여기서, $c_0 = \sum_{i=1}^g r_i = gr$

$C_i^{(n)}$ 을 스테이션 i 가 n 번째 폴링된 시점부터 $(n+1)$ 번째 폴링될 때까지의 시간 간격을 나타내는 n 번째 순환시간이라고 하면, $C_i^{(n)}$ 이 $c_0 + jh$ 가 될 확률은 다음과 같이 표현된다.

$$\Pr\{C_i^{(n)} = c_0 + jh\} = \sum_{k=0}^g \Pr\{C_i^{(n)} = c_0 + jh \mid C_{i-1}^{(n)} = c_0 + kh\} \cdot \Pr\{C_{i-1}^{(n)} = c_0 + kh\}, \quad k=0 \text{ to } g$$

3. 접근방법

본 연구는 이산 시간의 모형을 다루고 있으므로 순환 시간 C 는 다음과 같이 구현된다.

$$C \in \{c_0, c_0+h, c_0+2h, \dots, c_0+gh\}$$

이때, $\Pr\{C_i^{(n)} = c_0 + jh \mid C_{i-1}^{(n)} = c_0 + kh\}$ 는 전이 확률로서 다음과 같은 추계행렬 P 를 구성한다.

	c_0	c_0+h	...	c_0+jh	...	c_0+gh
c_0	$P(0,0)$	$P(0,1)$...	$P(0,j)$...	$P(0,g)$
c_0+h	$P(0,0)$	$P(1,1)$...	$P(1,j)$...	$P(1,g)$
.
.
.
$P=c_0+kh$	$P(k,0)$	$P(k,1)$...	$P(k,j)$...	$P(k,g)$
.
.
.
c_0+gh	$P(g,0)$	$P(g,1)$...	$P(g,j)$...	$P(g,g)$

임의의 한 스테이션이 연속해서 두 번 폴링 되는 동안 g 개의 스테이션 중에서 k 개가 서비스를 받았다면, 이때의 순환시간은 $c_0 + kh$ 가 된다. 이러한 순환시간 동안 도착한 평균 메시지 수 α_k 는 다음과 같이 표현된다.

$$\alpha_k = \lambda(c_0 + kh) \quad , k=0 \text{ to } g$$

이 때, α_k 는 임의의 스테이션이 폴링되었을 때 하나의 메시지가 서비스 받을 확률과 같은데, 이는 nonexhaustive 서비스를 하는 대칭적 폴링시스템에서의 안정화 조건이 다음과 같이 주어지기 때문이다[6].

$$g\lambda(r+h) = \lambda(gr+gh) = \lambda(c_0+gh) < 1$$

추계행렬 P 를 구성하는 $P(k, j)$ 는 스테이션 $(i-1)$ 과 스테이션 i 간의 순환시간의 상호 연관성에 의해 표현되는데, $C_{i-1}^{(n)}$ 과 $C_i^{(n)}$ 의 관계는 [그림 2]에 나타난 바와 같다.

(1) $C_{i-1}^{(n)} = c_0$

$C_{i-1}^{(n)}$ 동안 모든 스테이션이 서비스를 받지

못한 경우를 나타내므로 $C_i^{(n)}$ 는 다음의 경우에만 구현된다.

$$C_i^{(n)} \in \{c_0, c_0+h\}$$

$C_{i-1}^{(n)} = c_0$ 는 $C_i^{(n)}$ 에 기여하는 스테이션들 중에서 [그림 2]의 스테이션 i 부터 스테이션 $(i-2)$ 까지는 서비스를 받지 못한 것을 의미하므로 $C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 에 메시지의 도착이 없으면 스테이션 $(i-1)$ 도 서비스를 받지 못하므로(이때의 확률이 $1 - \alpha_0$) $C_i^{(n)} = c_0$ 가 된다. 반면에, 스테이션 $(i-1)$ 이 서비스를 받으면(이때의 확률은 α_0) $C_i^{(n)} = c_0 + h$ 가 된다. 따라서

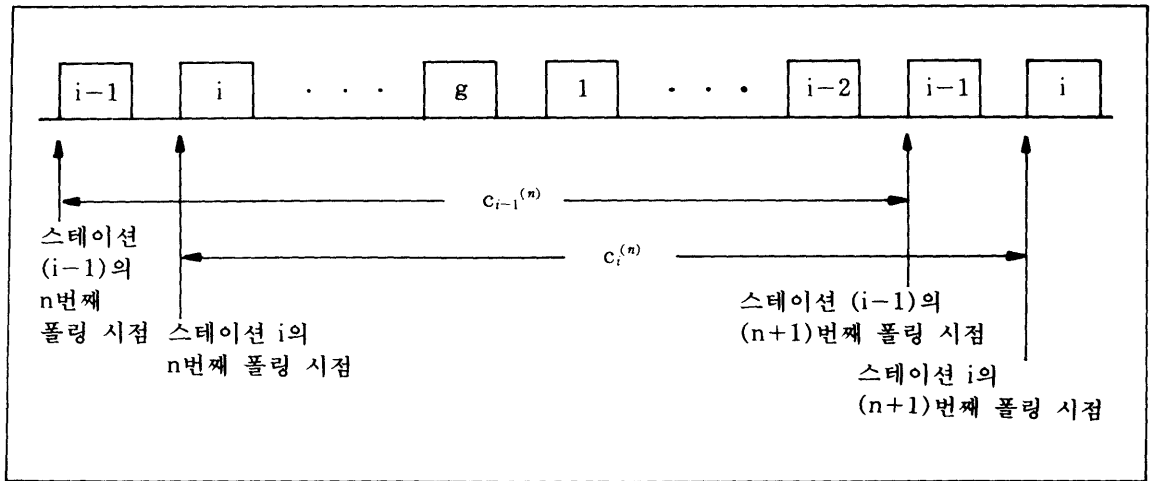
$$P(0,0) = 1 - \alpha_0, \quad P(0,1) = \alpha_0$$

(2) $C_{i-1}^{(n)} = c_0 + kh \quad (k=1, \dots, g-1)$

이 경우 $C_i^{(n)}$ 은 다음과 같이 구현된다.

$$C_i^{(n)} \in \{c_0 + (k-1)h, c_0 + kh, c_0 + (k+1)h\}$$

$C_{i-1}^{(n)} = c_0 + kh$ 는 $C_{i-1}^{(n)}$ 동안 k 개의 스테이션이 서비스를 받았으므로 스테이션 $(i-1)$ 이 서비스 받았을 확률은 k/g , 서비스를 받지 못하



[그림 2] 스테이션 간의 순환시간 관계

였을 확률은 $1-k/g$ 가 된다. $C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 이 서비스를 받았다면 $C_{i-1}^{(n)}$ 에 기여하는 스테이션들 중에서 [그림 2]의 스테이션 i 부터 스테이션 $(i-2)$ 까지 $(k-1)$ 개의 스테이션이 서비스를 받은 것을 의미하므로 $C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 에 메시지의 도착이 없어야만(이때의 확률은 $1-\alpha_k$) $C_i^{(n)}=c_o+(k-1)h$ 가 될 수 있다. 따라서,

$$P(k, k-1) = (1-\alpha_k) \cdot k/g$$

마찬가지로, $C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 이 서비스를 받았을 때 $C_i^{(n)}=c_o+kh$ 가 되려면

$C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 에 메시지의 도착이 없어야 한다(이때의 확률은 $1-\alpha_k$). 따라서,

$$P(k, k) = \alpha_k \cdot k/g + (1-\alpha_k)(1-k/g)$$

또한, $C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 이 서비스를 받았다면 $C_i^{(n)}=c_o+(k+1)h$ 가 될 수 없으므로, $C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 은 서비스를 받지 않은 상태에서 메시지의 도착이 없어야 한다.

따라서,

$$P(k, k+1) = \alpha_k(1-k/g)$$

$$(3) C_{i-1}^{(n)} = gh$$

$C_{i-1}^{(n)}$ 동안 모든 스테이션이 서비스를 받았으므로 $C_i^{(n)}$ 은 다음의 경우에만 구현된다.

$$C_i^{(n)} \in \{c_o+(g-1)h, c_o+gh\}$$

$C_{i-1}^{(n)}=gh$ 는 $C_i^{(n)}$ 에 기여하는 스테이션들 중에서 [그림 2]의 스테이션 i 부터 스테이션 $(i-2)$ 까지의 모든 스테이션이 서비스를 받은 것을 의미하므로 $C_{i-1}^{(n)}$ 동안 스테이션 $(i-1)$ 에 메시지의 도착이 없으면(이때의 확률은 $1-\alpha_g$) $C_i^{(n)}=c_o+(g-1)h$ 가 되고, 메시지의 도착이 있으면(이때의 확률이 α_g) $C_i^{(n)}=c_o+gh$ 가 된다. 따라서,

$$P(g, g-1) = 1-\alpha_g, P(g, g) = \alpha_g$$

이상의 과정을 거쳐 구한 추계행렬 P 는 다음과 같다.

	c_o	c_o+h	c_o+2h	...	c_o+kh	...	c_o+gh
c_o	$P(0,0)$	$P(0,1)$	0				
c_o+h	$P(1,0)$	$P(1,1)$	$P(1,2)$				ϕ
c_o+2h		$P(2,1)$	$P(1,2)$	$P(2,3)$			
.		.	.	.			
.			.	.	.		
.				.	.	.	
$P=c_o+kh$				$P(k,k-1)$	$P(k,k)$	$P(k,k+1)$	
.				.	.	.	
.		ϕ		.	.	.	
.				.	.	.	
c_o+gh				0	$P(g,g-1)$	$P(g,g)$	

스테이션 i 의 n 번째 순환시간이 $c_o + kh$ 가 될 확률을 $\Pi_k^{i,n}$ 이라 하고, 이러한 $\Pi_k^{i,n}$ 으로 구성된 벡터를 $\Pi^{i,n}$ 으로 나타내기로 한다. 즉

$$\Pi^{i,n} = [\Pi_0^{i,n}, \Pi_1^{i,n}, \dots, \Pi_g^{i,n}]$$

여기서,

$$\Pi_k^{i,n} = \Pr\{C_i^{(n)} = c_o + kh\}, \quad k=0 \text{ to } g$$

이때, $\Pi^{i,n}$ 은 스테이션 i 의 n 번째 순환 시간의 분포를 나타내므로 스테이션 $(i-1)$ 의 n 번째 순환 시간의 분포 $\Pi^{i-1,n}$ 과의 관계는 다

$$\Pi^{i,n} = \Pi^{i-1,n} \cdot P$$

이러한 관계식의 의하여 스테이션 i 의 안정상태에서의 확률 분포 벡터 Π^i 를 구할 수 있다. 즉

$$\begin{aligned} \Pi^{i,n} &= \Pi^{i-1,n} \cdot P \\ &= \Pi^{i-2,n} \cdot P \cdot P \\ &\vdots \\ &= \Pi^{i,n-1} \cdot P^g \end{aligned}$$

$$\lim_{n \rightarrow \infty} \Pi^{i,n} = \lim_{n \rightarrow \infty} \Pi^{i,n-1} \cdot P^g$$

$$\therefore \Pi^i = \Pi^i \cdot P^g \quad (1)$$

식(1)과 $\sum_{k=0}^g \Pi_k^i = 1$ 의 정규화 조건에 의하여 안정 확률 분포 Π^i 가 얻어지는데, 이 때의 분포는 모든 스테이션에서 같으므로 Π 로 나타낼 수 있다.

전이 행렬 P 에서의 모든 상태가 서로 도달 가능하고(irreducible), 상태의 수가 유한할 뿐만 아니라(recurrent nonnull), 모든 대각 원소가 0보다 크면(aperiodic)

$$\Pi = \Pi \cdot P$$

$$\Pi \cdot e = 1 \quad (e = [1, 1, \dots, 1]^T) \quad (2)$$

을 만족하는 Π 는 유일한 양의 해를 갖는다. 이때의 해를 Π^* 라고 하면, $n \in N = \{1, 2, \dots\}$

인 모든 n 에 대하여

$$\Pi^* = \Pi^* \cdot P = \Pi^* \cdot P^2 = \dots = \Pi^* P^n$$

이 성립한다. 따라서,

$$\Pi = \Pi \cdot P^g$$

$$\Pi \cdot e = 1$$

(3)

을 만족하는 해는 식(2)의 해와 같다.

따라서, 대칭적 폴링 시스템에 대한 안정 상태에서의 순환시간 C 의 확률분포가 $\Pi = [\Pi_0, \Pi_1, \dots, \Pi_g]$ 이면, 순환시간의 평균과 분산은 각각 다음과 같다.

$$E[C] = [\Pi_0, \Pi_1, \dots, \Pi_g] \begin{bmatrix} c_o \\ c_o + h \\ \vdots \\ c_o + gh \end{bmatrix} = \bar{C}$$

$$\text{Var}[C] = [\Pi_0, \Pi_1, \dots, \Pi_g] \begin{bmatrix} (c_o - \bar{C})^2 \\ (c_o + h - \bar{C})^2 \\ \vdots \\ (c_o + gh - \bar{C})^2 \end{bmatrix}$$

4. 계산결과

Nonexhaustive 서비스를 받는 GI/G/1 순환 서버 시스템에 대한 순환시간의 평균은 정확한 해가 알려져 있지만, 확률분포함수(pdf)의 정확한 해는 아직도 밝혀지고 있지 않다.

Hashida와 Ohara[5]는 순환시간의 분포함수 $F(t) = \Pr\{C \leq t\}$ 의 Laplace-Stieltjes 변환(LST)을 제시했는데, 이를 본 연구에서 다루는 모형에 맞게 바꾸면 다음과 같다.

$$\Phi_C(s) = [\Phi_R(s) \{\alpha \Phi_H(s) + (1 - \alpha)\}]^g \quad (4)$$

여기서,

R : switchover time의 r.v.

H : service time의 r.v.

$$\alpha = \lambda \cdot E[C]$$

식(4)에 의한 순환시간의 1차 모멘트와 2차 모멘트는 다음과 같다.

$$E[C] = \frac{d\Phi_C(s)}{ds} \Big|_{s=0} = g(r + \alpha \cdot h)$$

$$E[C^2] = \frac{d^2\Phi_C(s)}{ds^2} \Big|_{s=0} = g(g-1)(r + \alpha \cdot h)^2 + g(r^2 + 2\alpha \cdot r \cdot h + \alpha \cdot h^2)$$

$$\therefore \text{Var}[C] = E[C^2] - E[C]^2 = g \cdot \alpha(1-\alpha) \cdot h^2$$

이와 같이 구한 순환시간의 평균은 정확한 해 $c_0/(1-g \cdot \lambda \cdot h)$ 와 일치하지만, 분산은 시뮬레이션 결과보다 작다[8].

따라서, 본 연구에서 제시한 접근 방법의 정확도를 검증하기 위하여 여러 문제에 대하여 실험을 하여 산출된 순환시간의 평균과 분산을 Hashida와 Ohara의 결과 및 시뮬레이션 결과와 비교하였다. 시뮬레이션은 SIMAN을 이용하였는데, 일반적으로 폴링시스템의 시뮬레이션은 시스템이 안정상태에 도달할 때까지 상당한 시간이 걸린다. 따라서, 시뮬레이션으로 얻고자 하는 통계치는 시스템이 안정 국면에 들어선 후부터 수집하였다.

많은 문제에 대하여 실험을 한 결과 순환시

간의 평균은 정확한 해인 $c_0/(1-g\lambda h)$ 와 항상 일치함을 나타냈다. 따라서, 순환시간의 분산만을 Hashida와 Ohara의 결과치 및 시뮬레이션 결과치들과 비교하여 <표 1>, <표 2>와 같이 정리하였다.

일반적으로 외부로부터의 메시지 도착이 많을 수록 (λ 의 증가) Hashida와 Ohara의 결과는 시뮬레이션 결과보다 과소 추정(underestimation)하는 경향이 커진다[8]. 따라서, 본 연구에서는 다른 매개변수 값들을 동일한 상태에서 λ 값만을 변화시켜 순환시간의 분산이 어떻게 변화하는가에 대해 관심을 갖고 실험하였다. <표 1>, <표 2>는 이같은 실험들 중의 하나를 요약한 것이다.

실험 결과에 따르면, 본 연구에서 제시한 접근 방법에 의한 순환시간의 분산은 시뮬레이션 결과보다 과대 추정(overestimation)하는 경향이 있다. 그러나, 트래픽이 과중한(heavy load) 경우에 있어서 Hashida와 Ohara의 과소추정하는 비율보다는 본 연구 결과의 과대 추정하는 비율이 다소 적었다.

<표 1>

$g=8, h=0.1, c_0=1.0$

	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$	$\lambda=0.5$
시뮬레이션	0.0081	0.0156	0.0213	0.0240	0.0144
본 연구	0.0083	0.0170	0.0242	0.0269	0.0171
error (%)	3	8	14	12	19

<표 2>

$g=8, h=0.1, c_0=1.0$

	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$	$\lambda=0.5$
시뮬레이션	0.0081	0.0156	0.0213	0.0240	0.0144
Hashida와 Ohara	0.0078	0.0415	0.0191	0.0194	0.0111
error (%)	4	7	10	20	23

5. 결론

본 연구에서는 토큰링 및 토큰버스 등의 토큰패싱 프로토콜 하에서 운영되는 네트워크의 수학적 모형화에 많이 이용되는 폴링 시스템의 순환시간의 분포를 수치적으로 산출하는 새로운 접근 방법을 제시하였다. 이러한 접근 방법의 정확한 해와 항상 일치함을 보였으나, 분산은 시뮬레이션 결과치보다 과대 추정하는 경향을 보였다.

본 연구는 각 스테이션의 추계적 과정이 다른 스테이션들에 의하여 영향을 받는 관계를 고려하여 순환시간의 분포를 분석적으로 산출하는 방법을 새로이 제시한 데 그 의의가 있다.

참 고 문 헌

- [1] Boxma, O.J., "Two Symmetric Queues with Alternating Service and Switching Times," *Performance'84*(North-Holland, Amsterdam), pp. 409-431.
- [2] Cooper, R.B. and G. Murray, "Queues Served in Cyclic Oder," *Bell Syst. Tech. J.*, 48(1969), pp. 675-690.
- [3] Eisenberg, M., "Queues with Periodic Service and Changeover Time," *Oper. Res.*, Vol. 20(1972), pp. 440-451.
- [4] Fergerson, M.J. and Y. J. Aminetzah, "Exact Results for Nonsymmetric Token Ring Systems," *IEEE Trans. on Comm.*, Vol. 33, No. 3(1985), pp. 223-231.
- [5] Hashida, O. and K. Ohara, "Line Accomodation Capacity of a Communication Control Unit," *Rev. of Elec. Comm. Lab.*, NTT, Vol. 20(1972), pp. 231-239.
- [6] Ibe, O.C. and X. Cheng, "Stability Condition for Multiqueue Systems with Cyclic Service," *IEEE Trans. on Automatic Control*, Vol. 33, No. 1(1988), pp. 102-103.
- [7] Ibe, O.C. and X. Cheng, "Approximate Analysis of Asymmetric Single-Service Token Passing System," *IEEE Trans. on Comm.*, Vol. 37, No. 6(1989), pp. 572-577.
- [8] Kuehn, P.J., "Multiqueue Systems with Nonexhaustive Cyclic Service," *Bell Syst. Tech. J.*, 58(1979), pp. 671-698.
- [9] Leibowitz, M.A., "An Approximate Method for Treating a Class of Multiqueue Problems," *IBM J. of Res. and Dev.*, Vol. 5, No. 3(1961), pp. 204-209.
- [10] Pang, J.W.M. and R. W. Donaldson, "Approximate Delay Analysis and Results for Asymmetric Token-Passing and Polling Networks," *IEEE J. of Selected Areas in Comm.*, Vol. 4, No. 6(1986), pp. 783-793.
- [11] Takagi, H., "Mean Message Waiting Times in Symmetric Multi-Queue Systems with Cyclic Service," *Perform. Eval.*, Vol. 5(1985), pp. 271-277.
- [12] Takine, T., Y. Takahash, and T. Hasegawa, "Exact Analysis of Asymmetric Polling Systems with Single Buffers," *IEEE Trans. on Comm.*, Vol. 36, No. 10(1988), pp. 1119-1127.