

이질적 서버로 구성된 대기행렬 시스템의 도착시점상태 종속성

김준우*

요약

여러 가지 서비스 시스템의 평균 대기시간이나 평균 대기고객수와 같은 성능 변수를 산출하기 위해서는 흔히 대기이론이 사용된다. 이러한 목적으로 가장 많이 사용되는 것은 흔히 Erlang-C로 불리기도 하는 M/M/c 모형 및 그 변종들이다. 이러한 단순한 대기모형들은 시스템 내 서버들의 특성이 확률적으로 동일함을 가정한다. 그러나 현실적으로 서버들의 숙련도나 성향 등에 따라 서버들의 서비스율은 서로 상이할 수 있다. 본 논문에서는 확률적으로 동일한 고객들이 서로 다른 서비스율을 갖는 서버 2명에 의해 응대되는 마야코프 대기행렬 시스템에서 발생하는 도착시점상태 종속성의 특성에 대해 관찰한다. 이러한 시스템의 성능 변수를 계산하기 위한 간단한 방법을 소개하고자 한다. 뿐만 아니라 수치적인 분석 결과 높은 서비스율을 가진 서버가 있는 경우, 다른 서버 성능이 낮더라도 시스템의 성능변수들의 값이 상당히 개선됨을 알 수 있었다.

Arrival Point State Dependency of Queueing Systems Consisted of Inhomogeneous Servers

Jun-Woo Kim*

ABSTRACT

Queueing theory is widely applied to obtain the performance measures such as average waiting time and average queue length of various service systems. M/M/c model, so called Erlang-C, and its variants are widely used for this purpose, and such simple queueing models assume that the servers are stochastically identical, however, the service rates of the servers in practical service systems can be different from each other according to their skills or job preferences. This paper considers a Markov queueing system where stochastically identical customers are handled by inhomogeneous 2 servers, and analyzes the characteristics of the arrival-point-state dependency occurred in such systems. Next, this paper proposes a simple method for computing the performance measures of the inhomogeneous 2-server queueing system, and numerical tests reveal that a server with relatively high service rate can significantly enhance the performance measures of the system even if the other one has lower service rate.

Key Words : Queueing Theory, Inhomogeneous Servers, 2-Server System, Dependency, Operations Research

* 동아대학교 산업경영공학과 (✉kjunwoo@dau.ac.kr)

· 제1저자(First Author) : 김준우 · 교신저자(Correspondent Author) : 김준우

· 접수일(2012년 3월 15일), 수정일(1차 : 2012년 4월 13일), 게재확정일(2012년 4월 17일)

1. 서론

고객의 도착과정과 서버의 응대과정이 모두 확률적인 시스템의 성능을 분석하기 위해서는 시스템은 대기모형으로 모델링한 다음, 이 대기모형의 성능변수를 분석하는 것이 일반적이다[1].

특히 서비스업의 경우 고객(업무)의 도착 간격과 서버에 의한 1회 응대시간이 일정하지 않고 확률적이기 때문에 대기이론을 사용하여 평균 대기시간이나 평균 대기고객수와 같은 시스템의 성능변수를 분석하는 경우가 많다[2].

이러한 대기모형 중 가장 간단하면서도 널리 사용되는 것은 M/M/c 모형으로 이는 확률적으로 동일한 특성을 가진 고객들이 지수분포를 따르는 시간 간격을 두고 시스템을 방문하며, 도착한 고객들은 c명의 확률적으로 동일한 특성을 가진 서버들에 의하여 응대되는 것을 가정한다[3][4]. 여기서 서버들의 특성이 확률적으로 동일하다는 것은 서버들의 역량이 동일하여 임의의 고객 한 명을 응대하는 데 소요되는 평균 시간이 모두 같음을 의미하며, M/M/c 모형에서는 고객 한 명에 대한 응대 시간 역시 지수 분포를 따른다고 가정한다.

본 논문은 M/M/c 모형이 가지고 있는 다양한 가정 중, 서버들의 특성이 확률적으로 모두 동일하다는 점에 대해 논의하고자 한다. 전통적으로 제조업에 비해 서비스업의 산출물은 그 품질과 특성의 균질성이 떨어진다는 특성이 지적되어 왔다. 이에 따라 서비스 현장에서는 서버들에 대한 상시적 교육과 업무의 매뉴얼화 등을 통하여 모든 서버들이 균질한 수준의 응대를 하는 것을 추구하는 경향이 있다[5]. 그럼에도 불구하고 기계적인 작업보다는 사람의 판단과 역량에 의존하는 서비스 분야에서 서버들간의 개인 역량 차이는 불가피할 것으로 생각된다. 따라서 이중 서버들로 구성된 대기모형에 대한 분석이 필수적일 것이다.

본 논문은 서버 2명으로 구성된 대기모형 시스템의 특성을 자세히 고찰하고자 한다. 분석 대상 시스템은 M/M/2 모형과 나머지 특성이 동일하나 서버들의 역량에 차이가 있어 서버 1은 서비스율 μ_1 , 서버 2는 서비스율 μ_2 ($\mu_1 \neq \mu_2$)인 응대시간이 걸린다.

나아가 본 논문은 이러한 시스템에서 고객 1명에 대한 응대 시간이 그 고객의 도착시점에서 관찰되는 시스템 내부 고객수, 즉, 도착시점상태 종속적임을 보이고, 이러한 도착시점상태 종속성을 갖는 대기모형의 여러 가지 특성에 따라 시스템의 성능 변수를 분석하는 방법을 소개한다. 이를 통해 본 논문은 향후 보다 현실적인 서비스 시스템의 성능 분석에 기여할 것으로 기대된다.

본 논문은 2장에서 이중의 서버에 의한 응대가 일어나는 시스템에 대한 기존 문헌들을 간단히 소개하고, 3장에서 종래의 M/M/c 모형과 이중 서버 시스템의 차이를 비교 분석한 다음, 이중 서버 시스템의 성능변수를 분석하는 방법을 설명한다. 4장에서는 이 방법에 따라 실제 시스템의 성능변수를 수치적으로 계산한 결과를 살펴보고, 끝으로 5장에서 결론 및 추후 과제를 제시한다.

II. 문헌 연구

대기모형이 가장 활발하게 적용되고 있는 분야로 콜센터를 들 수 있다. 콜센터는 전통적으로 규모가 크고 노동집약적인 서비스 시스템으로 M/M/c 모형과 같은 단순한 모형부터 보다 복잡한 형태의 모형들이 다양하게 적용되어져 왔다[6][7].

전통적으로 콜센터 내의 서버들 역시 확률적으로 동일하다고 가정되어 왔지만 근래에는 서로 특성이 다른 서버들로 구성되는 콜센터들에 대한 연구들도 이루어져 왔다[7].

그러나 이러한 연구들에서 상이한 서버들은 서로 다른 집단을 구성하여 응대에 참여하는 것으로 취급되었고, 기존 문헌들은 이들 서버 집단 간의 인입호 배분이나 라우팅 방법에 초점을 맞추었다[8]. 구체적으로는 서로 다른 언어를 사용하는 서버 집단이나, 아웃소싱을 통해 외부 서버를 고객 응대에 참여시키는 경우가 이들에 해당하며, 전문 분야가 서로 다른 서버 집단 간의 라우팅 방법에 대한 연구[9][10] 및 고정된 인원수의 내부 서버를 채용하고 있는 콜센터에서 필요한 외부 서버의 최적 인원수를 산정하는 방법에 대한 연구[11][12][13][14] 등에 대한 기존 문헌들을 찾아볼 수 있다.

반면, 현실적으로는 동일 집단 내 서버들이 서비스율에 차이가 있는 경우가 많을 것으로 생각되고, 이러한 이종 서버들로 구성된 대기행렬 시스템의 특성 및 성능변수에 대한 분석이 필요할 것이다. 이러한 맥락에서 본 논문은 이종 서버가 동일 집단을 구성하여 고객 응대에 참여하는 서비스 시스템 중, 가장 단순한 형태인 2-서버 시스템을 분석 대상으로 정하고, 이 시스템의 특성 및 여러 가지 성능변수 산출과정에 대해 자세히 알아보고자 한다.

III. 이종 서버 대기모형

3.1 동종 서버로 구성된 M/M/2 모형

확률적으로 특성이 동일한 고객들이 도착률 λ 인 포아송 과정(Poisson process)에 따라 도착하고, 서버가 1명의 고객을 응대하는 데 걸리는 시간은 평균이 μ^{-1} 인 지수분포를 따르는 경우, 2명의 서버로 구성된 M/M/2 모형의 특성은 <그림 1>과 같은 상태전이 다이어그램으로 나타낼 수 있다. 단, <그림 1>에서 0, 1, 2, ... 가 기록된 원은 각각 시스템 내부에 존재하는 고객의 인원이 0, 1, 2, ... 인 상태를 의미한다.

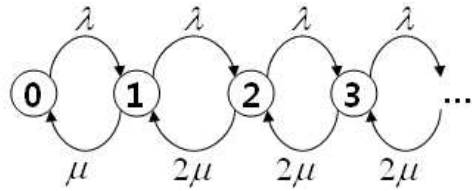


그림 1. M/M/2 모형 상태전이 다이어그램
Fig. 1. State-transition Diagram of M/M/2

이러한 M/M/2 모형의 특성은 잘 알려진 M/M/c 모형의 여러 가지 특성에 $c=2$ 를 대입하여 계산할 수 있다. M/M/c 모형의 경우, 안정 상태의 임의의 시점에 시스템을 관찰하였을 때, 내부에 고객 이 0, 1, 2, ... 명 있을 확률 P_0, P_1, P_2, \dots 가 다음과 같다.

$$P_0 = \left[\left(\sum_{k=0}^{c-1} \frac{a^k}{k!} \right) + \frac{a^c}{c!} \times \frac{1}{(1-\rho)} \right]^{-1} \quad (1)$$

$$P_n = \begin{cases} \frac{(c\rho)^n}{n!} P_0, & (1 \leq n \leq c-1) \\ \frac{(c\rho)^c}{c!} \rho^{n-c} P_0, & (n \geq c) \end{cases}$$

(제공로드 $a = \frac{\lambda}{\mu}$, 교통밀도 $\rho = \frac{\lambda}{c\mu}$)

또한, 이를 이용하면 평균 대기고객수 L_q 와 평균 대기시간 W_q 를 다음과 같이 얻는다.

$$L_q = \frac{\rho}{(1-\rho)^2} P_c \quad (2)$$

$$W_q = \frac{L_q}{\lambda} \quad (3)$$

따라서 확률적으로 특성이 동일한 2명의 서버로 구성된 2-서버 시스템의 경우, (1)~(3)을 이용하여 그 성능 변수들을 분석할 수 있다.

3.2 이중 서어버로 구성된 2-서어버 시스템

이제, 앞과 같은 2-서어버 시스템에서 두 명의 서어버의 특성이 상이함을 가정해보자. 이에 따라 서어버1의 서비스율은 μ_1 , 서어버2의 서비스율은 μ_2 이며, 편의상, $\mu_1 > \mu_2$ 라 하자. 즉, 서어버1은 서어버2에 비해 평균적으로 고객 응대에 걸리는 시간이 적으며, 이러한 상황은 두 서어버 간의 숙련도나 적성의 차이를 비롯하여 내부 전문 근로자와 파견 근로자의 동시 업무 수행 등 다양한 경우에 발생할 수 있다.

이러한 시스템에 도착하는 임의의 고객 1명의 평균 응대 시간은 당연히 그 고객이 어느 서어버에 연결되느냐에 따라 달라진다. 따라서 고객 1명의 평균 응대 시간을 파악하기 위해서는 각 고객이 각 서어버에 연결될 확률을 구해야 하고, 이 확률은 고객의 도착시점에서 관찰되는 시스템 내부 고객 인원 수, 즉, 도착시점 상태에 종속적이다.

예를 들어, 별도의 작업 배정 규칙이 없는 경우, 시스템 내부에 고객이 0명인 상태에서 도착한 고객은 무작위로 각 서어버에 연결된다. 따라서 (1/2)의 확률로 서어버1, 역시 (1/2)의 확률로 서어버2에 의해 응대된다. 만약, 시스템 내부에 고객이 2명 이상으로 서어버들이 모두 응대 중인 상태에서 새로운 고객이 도착하였다면, 이 고객은 대기상태로 들어가서 자신의 차례를 기다리게 된다. 그리고 이 고객이 서어버1, 서어버2에 연결될 확률 $P_{\rightarrow 1}, P_{\rightarrow 2}$ 는 각각 (4.1)~(4.2)와 같이 주어진다.

$$P_{\rightarrow 1} = \frac{\mu_1}{\mu_1 + \mu_2} \quad (4.1)$$

$$P_{\rightarrow 2} = \frac{\mu_2}{\mu_1 + \mu_2} \quad (4.2)$$

(4)와 같은 확률이 주어지는 이유는 다음과 같다. 대기상태의 고객이 대기행렬의 맨 앞까지 도달하였다면

이 고객은 더 이상 자신보다 앞의 대기자는 없는 상태에서 두 명의 서어버가 각기 다른 고객들을 응대하고 있는 상황을 보게 된다. 지수분포의 무기억 속성에 의하여 두 서어버의 잔여 응대 시간은 이때까지의 응대 시간에 무관하며, 따라서 앞으로 (4.1)의 확률로 서어버1이 먼저 응대를 마쳐 새로 도착한 고객이 서어버1에 연결되고, (4.2)의 확률로 서어버2가 먼저 응대를 마치게 되어 이 경우 서어버2와 연결된다.

추가적으로, 만약 시스템 내부에 고객이 1명 존재하는 상태에서 도착한 고객을 생각해보자. 만약 기존의 고객이 서어버1에 의해 응대 중이라면 이 고객은 서어버2에 연결되고, 서어버2가 응대 중이라면 서어버1에 연결된다. 따라서 기존 1명의 고객이 응대 중인 상황에서 이 고객이 각 서어버에 의해 응대 중일 조건부 확률을 고려하는 것이 필요하다.

두 서어버의 서비스율이 완전히 동일한 경우에는 $P(\text{서어버1 응대} | \text{고객 1명}) = P(\text{서어버2 응대} | \text{고객 1명}) = (1/2)$ 임이 자명하다. 이는 두 서어버의 성능이 동일하기 때문에 과업이 주어졌을 때 한 쪽 서어버가 상대방보다 일을 늦게 마칠 가능성 역시 동일하기 때문이다. 만약 서비스율이 서로 상이한 경우에는 한 쪽 서어버는 상대방이 먼저 작업을 마친 경우에 잔여 작업을 남겨두게 된다. 따라서 (5)와 같은 관계를 얻는다.

$$\begin{aligned} P(\text{서어버1 응대} | \text{고객 1명}) & \quad (5.1) \\ & = P(\text{서어버2 조기 응대}) = \frac{\mu_2}{\mu_1 + \mu_2} \end{aligned}$$

$$\begin{aligned} P(\text{서어버2 응대} | \text{고객 1명}) & \quad (5.2) \\ & = P(\text{서어버1 조기 응대}) = \frac{\mu_1}{\mu_1 + \mu_2} \end{aligned}$$

나아가, 시스템 내부 고객 수가 1명인 상태에서 도착하는 고객이 각 서어버에 연결될 확률이 다음과 같이 주어지고, 이는 (4)에서 주어진 시스템 내부 고객 수가 2명 이상일 때와 동일하다.

$$P_{\rightarrow 1} = P(\text{서버 2 응대} | \text{고객 1명}) \quad (6.1)$$

$$P_{\rightarrow 2} = P(\text{서버 1 응대} | \text{고객 1명}) \quad (6.2)$$

이상의 결과를 정리하면, 2명의 서버들의 서비스율이 서로 상이한 시스템에 도착하는 고객의 경우, 시스템 내부 고객 수가 0명일 상태에서 도착한 고객들은 각각 (1/2)의 확률로 서버 1, 2에 연결되나, 시스템 내부 고객 수가 1명 이상인 상태에서 도착한 고객들은 각각 (4)의 $P_{\rightarrow 1}$, $P_{\rightarrow 2}$ 의 확률로 서버 1, 2에 연결됨을 알 수 있다. 이를 이용하면 고객 1명에 대한 평균 서비스율 μ^* 가 다음과 같아진다.

$$\mu^* = P_0 \left(\frac{\mu_1 + \mu_2}{2} \right) + (1 - P_0)(P_{\rightarrow 1}\mu_1 + P_{\rightarrow 2}\mu_2) \quad (7)$$

(7)은 이중 서버로 구성된 대기행렬 시스템의 도착시점상태 종속성을 보여주고 있다. 일반적으로 대기행렬 시스템의 도착률이나 서비스율과 같은 파라미터들은 그 값이 미리 확정적으로 주어진다. 반면, 이중 서버들이 시스템을 구성할 경우, 시스템 파라미터 중 고객 1명에 대한 평균 서비스율은 시스템의 안정상태 확률 P_0 를 통해 구해진다. 따라서 기존의 동종 서버로 이루어진 M/M/2 모형과 같이 (1)~(3)을 직접 이용하여 그 성능 변수를 분석하는 것이 곤란하다.

3.3 이중 서버 시스템의 성능 분석

도착시점상태 종속성을 갖는 이중 서버 시스템의 경우, 먼저 P_0 를 결정하고, 이에 따라 평균 서비스율 μ^* 를 (7)에 의해 계산한 다음, 얻어진 μ^* 와 고객 도착률 λ 를 (1)~(3)에 대입하여 성능 변수를 분석해야 한다. 하지만 이러한 과정은 수치적으로 수행하기 곤란하기 때문에 시행착오법에 근거하여 이중 서버 시

스템의 성능 변수를 분석한다.

먼저, 분석 대상 이중 서버 시스템에 도착하는 고객들이 각 서버에 의해 연결될 확률이 항상 (1/2)이라고 가정하자. 그러면 이 시스템은 도착률 λ 와 평균 서비스율 $\tilde{\mu}$ 를 갖는 M/M/2 모형이 되므로 그 성능 변수 및 고객 수가 0명일 확률 \tilde{P}_0 를 (1)~(3)을 이용하여 쉽게 구할 수 있다. 여기서의 $\tilde{\mu}$, \tilde{P}_0 는 실제 이중 서버 시스템의 μ^* , P_0 와 일치하는 값은 아니나, 시행착오법을 시작하기 위한 이들의 초기 추정치로 사용한다. $\tilde{\mu}$, \tilde{P}_0 를 구한 다음에는 아래와 같은 과정을 통하여 실제 이중 서버 시스템의 P_0 의 정확한 값을 시행착오를 거쳐가며 얻는다.

(단계 1) $P_0 = \tilde{P}_0$ 로 두고, (7)에 의해 μ^* 를 계산한다.

(단계 2) 이중 서버 시스템을 파라미터 λ , μ^* 를 갖는 M/M/2 시스템으로 가정하고 (1)에 의해 P_0 의 새로운 추정치인 P_0' 를 계산한다.

(단계 3) P_0 와 P_0' 의 값을 서로 비교한다.

(단계 3-1) 만약 $P_0 = P_0'$ 이거나, 그 차이가 미세하다면 현재의 P_0 는 대상 이중 서버 시스템에 대한 정확한 P_0 값이므로 종료한다.

(단계 3-2) 만약 P_0 와 P_0' 의 차이가 큰 경우에는 현재의 P_0 에 적절한 값 Δ 를 더하여 $P_0 = P_0 + \Delta$ 로 두고, 변경된 P_0 를 이용하여 μ^* 를 다시 계산한 다음, (단계 2)부터 반복한다.

참고로 (단계 3-2)에서의 Δ 값의 경우, 적절히 작은 값을 사용하며, $P_0 < P_0'$ 인 경우에는 양수, $P_0 > P_0'$ 인 경우에는 음수를 사용한다. 또한, 위 단계를 거쳐 P_0 와 μ^* 의 정확한 값을 얻은 경우에는 대상 시스템이 평

균 서비스율 μ^* 를 갖는 M/M/2 모형임이 밝혀진 것이므로, 이들을 (1)~(3)에 대입하여 평균 대기고객수나 평균 대기시간과 같은 시스템의 성능 변수들을 계산할 수 있다.

IV. 시스템 성능 분석

본 장에서는 이제까지 설명한 이중 서버들로 이루어진 2-서버 시스템의 성능 분석을 실시한다. 구체적으로는 이중 서버에 대한 분석의 필요성을 보이기 위하여 종래의 M/M/2 시스템의 성능 변수를 계산한 다음, 두 명의 서버들의 서비스율 값들을 변동시켜가며 성능 변수들의 변화를 관찰하고자 한다.

예를 들어 도착률 $\lambda=3$ /분인 M/M/2 시스템에서 서버 한 명의 서비스율 $\mu=2$ /분인 경우를 생각해 보자. (1)~(3)에 의하여 이 시스템의 성능 변수는 <표 1>과 같이 구할 수 있다.

표 1. M/M/2 모형의 성능 변수
Table 1. Performance Measures of M/M/2 model

성능 변수	값
P_0	0.143
W_q	0.643 (분)
L_q	1.929 (명)

이제 이 두 명의 서버들의 서비스율의 평균은 앞의 M/M/2 모형에서처럼 2/분을 유지하면서 양쪽 서버의 서비스율이 차이가 나도록 조정해가며 시스템의 특성을 관찰한다.

먼저, 서버1의 서비스율 $\mu_1 = 2.5$ /분, 서버2의 서비스율 $\mu_2 = 1.5$ /분인 경우를 생각해 보자. 이 경우 (4)에 의하여 $P_{\rightarrow 1}=0.625$, $P_{\rightarrow 2}=0.375$ 로 주어지고, 서비스율의 산술 평균은 2/분으로 앞의 M/M/2 모형에

서 서버 한 명의 서비스율과 동일하다. 이러한 경우 흔히 M/M/2 모형을 사용하게 되기 쉬우나, 본 논문에서 제안한 분석 과정을 고려하면 <표 1>의 $P_0=0.143$ 은 실제 값에 대한 초기 추정치 \tilde{P}_0 이다. 따라서 이를 초기 P_0 로 두면 (7)에 의하여 $\mu^*=2.107$ /분이 된다. 이제 도착률 $\lambda=3$ /분, 서버 1명의 서비스율이 2.107/분인 M/M/2 모형의 P_0 를 (1)에 의하여 구하면 0.168이 되므로 초기 추정치 0.143과 차이가 있고, 아직 P_0 의 값을 정확히 추정하지 못하였음을 알 수 있다. 편의 상 $\Delta=0.005$ 로 두고 현재의 P_0 추정치의 값을 0.143, 0.145, 0.150, 0.155... 로 변동시키면 <표 2>와 같은 결과를 얻게 된다. 참고로 초기 추정치를 첫 번째 변동시킬 때는 편의 상 0.145로 0.002만큼 변경시켜 끝 자리를 5로 맞추었고, 그 다음부터는 Δ 만큼 값을 변경시켰다.

표 2. P_0 의 추정
Table 2. Estimation of P_0

P_0 추정치	μ^*	(1)로 얻은 P_0
0.143	2.1071/분	0.1683
0.145	2.1069/분	0.1683
0.150	2.1063/분	0.1681
0.155	2.1056/분	0.1680
0.160	2.1050/분	0.1678
0.165	2.1044/분	0.1677
0.170	2.1038/분	0.1675
0.175	2.1031/분	0.1674

<표 2>에서 시행착오를 통해 P_0 에 대한 추정치로 0.170을 사용하였을 때 추정치와 이를 (1)에 대입하여 계산한 값의 차이가 가장 적음을 알 수 있고, 이 때의 추정치와 계산값의 평균인 $(0.170+0.1675)/2 = 0.1688$ 을 실제 P_0 에 대한 추정치로 사용한다. 또한 이를 (7)

에 대입하여 이 시스템의 정확한 $\mu^* = 2.1039/\text{분}$ 을 얻을 수 있고, 이들을 (2), (3)에 대입하면 이 시스템의 성능 변수는 <표 3>과 같다.

표 3. 이종 서버 시스템의 성능 변수
Table 3. Performance Measures of inhomogeneous 2-server system

성능 변수	값
P_0	0.168
W_q	0.4914 (분)
L_q	1.4741 (명)

<표 3>에서 우리는 이종 서버로 구성된 시스템의 경우 평균 응대율이 동일한 기존의 M/M/2 시스템과의 성능이 차이가 남을 알 수 있다. 나아가, 평균 응대율을 고정시켜 놓고 각 서버들의 서비스율을 다양하게 변화시켜가며 구한 이종 서버 시스템의 성능 변수 값은 <표 4>에서 볼 수 있다.

표 4. 다양한 상황에서의 성능 변수
Table 4. Performance Measures under Various Conditions

μ_1	μ_2	P_0	W_q	L_q
2.2	1.8	0.147	0.6134	1.8402
2.5	1.5	0.168	0.4914	1.4741
3.0	1.0	0.229	0.2741	0.8223
3.5	0.5	0.300	0.1463	0.4388

<표 4>에 따르면 두 명의 서버들의 응대율 평균이 일정하더라도 서비스율의 차이가 크면 클수록 M/M/2 모형의 성능 변수에서 떨어지는 것을 알 수 있고, 한 쪽이라도 우수한 서버가 있는 경우 전체 시스템의 성능이 개선되는 것을 볼 수 있다. 따라서 이종 서버로 구성된 시스템의 성능 분석을 위해 본 논문에서 제안하는 방법이 필요함을 알 수 있다.

V. 결론 및 추후 연구 과제

본 논문에서는 확률적으로 동일한 서버들을 가정하던 기존 대기모형들과 달리 서비스율이 상이한 서버로 구성된 2-서버 시스템의 특성을 분석하였다. 이 경우, 확률적으로 동일한 고객들이 시스템을 방문하더라도 도착시점상태 종속성이 발생하여 시스템의 특성이 기존 M/M/2 모형과 달라짐을 알 수 있었다.

이러한 이종 서버 시스템의 성능 변수는 시행착오를 통하여 간단히 계산할 수 있으며, 특히 한 쪽이라도 우수한 상담원이 있는 경우, 전체 시스템의 성능이 크게 개선됨을 볼 수 있었다. 현실적으로 인간 서버들의 능력에 개인차가 있음을 고려할 때 이러한 분석 방법은 앞으로 보다 현실적인 시스템에 대한 분석에 크게 기여할 것으로 생각된다.

반면, 이번 논문에서의 분석 대상 시스템은 서버 두 명의 서비스율에 차이가 있다는 점 외에는 기존 대기모형과 거의 동일한 특성을 가졌으므로, 앞으로 보다 다양한 상황에서 이종 서버 시스템의 성능을 관찰하고, 이러한 시스템에 적용할 수 있는 작업 배정 규칙 등에 대한 연구를 계속할 계획이다.

참고문헌

- [1] Lee, H.W., "대기행렬이론: 확률과정론적 분석, 3rd edition," Sigma Press, 2006.
- [2] Metters, R.D., King-Metters, K.H., Pullman, M., and Walton, S., "Successful Service Operations Management, 2nd edition," South-Western Educational Publishing, 2002.
- [3] Koole, G., "Call Center Mathematics," Technical Report, <http://www.math.vu.nl/~koole/ccmath>
- [4] Brown, L.D., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L., "Statistical Analysis of a Telephone Call Center: A Queueing Science Perspective," *Journal of the American Statistical*

- Association*, Vol.100, No.469, pp.36-50.
- [5] Fitsimmons, J.A., and Fitsimmons, M.J., "Service Management: Operations, Strategy, Informatin Technology," McGraw-Hill, 2008.
- [6] Gans, N., Koole, G., and Mandelbaum, A., "Telephone Call Centers: A Tutorial and Literature Review," *Manufacturing and Service Operations Management*, Vol.5, No.2, pp.79-141, 2002.
- [7] Aksin, O.Z., Armony, M., and Mehrotra, V., "The Modern Call Centers: A Multi-disciplinary Perspective on Operations Management Research," *Products and Operations Management*, Vol.16, No.6, pp.665-688, 2007.
- [8] Garnett, O., and Mandelbaum, M., "An Introduction to Skill-based Routing and Its Operational Complexities," Teaching Notes, Technion, Israel, 2000.
- [9] Koole, G., Pot, A., and Talim, J., "Routing Heuristics for Multi-skill Call Centers," *Proceedings of the 2003 Simulation Conference*, Vol.2, pp.1813-1816, 2003.
- [10] Wallace, R.B., and Whitt, W., "A Staffing Algorithm for Call Centers with Skill-based Routing," *Manufacturing and Operations Management*, Vol.7, No.4, pp.276-294, 2005.
- [11] Gans, N., and Zhou, Y., "A Call Routing Problem with Service Level Constraints," *Operations Research*, Vol.51, No.2, pp.255-271, 2003.
- [12] Gans, N., and Zhou, Y., "Call Routing Schemes for Call Center Outsourcing," *Manufacturing and Service Operations Management*, Vol.9, No.1, pp.33-50, 2007.
- [13] Ren, Z., and Zhou, Y., "Call Center Outsourcing: Coordinating Staffing Level and Service Quality," *Management Science*, Vol.54, No.2, pp.369-383, 2008.
- [14] Kim, J.W., and Park, S.C., "Outsourcing Strategy in Two-stage Call Centers," *Computers and Operations Research*, Vol.37, No.4, pp.790-805, 2010.

감사의 글

이 논문은 동아대학교 교내연구비 지원에 의하여 연구되었음.

저자소개

김준우(Jun-Woo Kim)



2001년 한국과학기술원 산업공학과 졸업(공학사)

2003년 한국과학기술원 산업공학과 졸업(공학석사)

2009년 한국과학기술원 산업 및 시스템공학과 졸업(공학박사)

2009년~2010년 한국기술교육대학교 산업경영학부 대우교수

2011년~현재 동아대학교 산업경영공학과 조교수

※ 관심분야 : 데이터마이닝, 지능형 시스템, 서비스 관리, Operations Research, e-러닝, 융합 콘텐츠