대기행렬모형에서 우선순위의 혜택과 비용 (Cost and Benefit of Priority in Queueing)*

남 익 현**

《目 次》

Ⅰ. 들어가며

우리는 다양한 기다림의 현상을 분석하고 필요한 의사결정을 하기 위해 대기행렬이론(Queueing Theory)을 이용하여 모형화를 한다. 그런데 고객의 서비스 순서에 관한 연구가 중요한 효과를 냄을 알 수 있다. 기본적인 서비스 순서인 FCFS(First Come First Serve), 즉 고객이 도착한 순서대로 서비스를 제공하는 것과 특정집단에 우선순위를 제공하는 경우를 비교하는 연구가 많이 이루어졌다. 특정 고객집단에 서비스순서에서의 우선순위를 제공한다는 것은 해당 집단에 속한 고객의 경우 다른 집단의 고객보다 서비스 순서에 있어 혜택을 본다는 것이다.

우선순위는 두 가지로 구분할 수 있는데, 탈취적 우선순위(preemptive priority)와 비탈취적 우선순위(non-preemptive priority)로 나눌 수 있다. 탈취적 우선순위는 우선순위가 높은 고객이 도착할 경우 현재 서비스를 받고 있는 고객을 바로 나오게 하고 자신이 먼저 서비스를 제공받는 것이다. 반면 비탈취적 우선순위는 현재 서비스를 받고 있는 고객은 해당 서비스를 마칠 때까지 기다려주고 이 고객이 서비스를 마치면 우선순위를 주장하는 경우이다. 본 논문에서는 분석의 편의상 비탈취적 우선순위를 적용하는 경우를 다루기로 한다.

여러 고객집단이 존재하는데, 이들의 평균 서비스시간에 차이가 있을 경우 우선순위를 적용하는 것을 검토해 보자. 다양한 연구에서 나온 결과는 서비스 시간이 짧은 고객부터 처리를 할 경우 고

* 본 연구는 서울대학교 경영정보연구소의 연구비 지원에 의해 이루어졌습니다.
** 서울대학교 경영대학 교수

객 1인당 총소요시간을 최소화할 수 있다는 것이다. 이를 SPT(Shortest Processing Time) 규칙이라고 한다.

우리가 본 논문에서 다루고자 하는 것은 SPT 규칙을 따르는 우선순위가 효과적이지만 이는 우선순위가 낮은 고객들이 서비스 순서를 양보해주어서 발생하는 것이다. 그리고 많은 경우 서비스 순서의 양보는 자발적으로 일어나지 않는 경우가 많다. 즉 우선순위가 낮은 고객들의 희생에 의해 전체 대기행렬시스템의 성과가 개선되는 것이다. 따라서 우선순위가 낮은 고객집단의 희생에 대한 부분을 명시적으로 다루는 것이 필요하다. 본 논문에서는 우선순위에 의해 일정 집단이 손해를 보게 되는 것을 비용함수로 명시하여 이를 의사결정에 반영해 보는 것을 다루고 있다.

Ⅱ. 부호와 우선순위 모형 재고찰

우리는 〔Gross & Harris, 1998〕에서 다룬 대기행렬의 우선순위 모형을 설명하고 이를 추가적인 논의에 활용하기로 하자. 우리가 다루려는 대기행렬 모형에서는 r개의 고객 집단이 있다. 임의의 고객집단 k는 도착률(arrival rate)이 λ_k인 포아송 분포(Poisson Distribution)를 따른다고 가정한다. 또한 서비스 시간도 서비스율(service rate)이 μ_k인 지수분포(Exponential Distribution)를 따른다고 가정한다. 즉 고객 집단 k에 속하는 고객의 경우, 평균 서비스 시간은 $1/\mu_k$이다.

우리는 대기행렬에서 우선순위(priority)를 다음과 같이 정의하기로 하자. 우선순위 k의 고객집단에 해당하는 고객이 도착하여 대기행렬에 들어가면, k보다 작은 고객집단(보다 우선순위가 높은 집단)보다 나중에 줄을 서고, k보다 큰 고객집단(보다 우선순위가 낮은 집단)보다는 앞서 줄을 서게 된다. 동일한 고객집단인 고객집단 k내에서는 도착한 순서대로 서비스를 받게 된다.

다음과 같이 부호를 정의하자.

$$\lambda = \sum_{k=1}^{r} \lambda_k,$$

$$\rho_k = \frac{\lambda_k}{\mu_k} (1 \le k \le r),$$

$$\sigma_k = \sum_{i=1}^{k} \rho_i (\sigma_0 = 0, \sigma_r = \rho).$$

우선순위가 i인 고객이 도착하였다고 하자. 이 고객이 서비스를 받기 시작할 때까지 기다리는 시

간을 T_i로 표시하자. 즉 T_i은 해당 고객의 대기행렬에서의 순수대기시간을 나타내는 것이다. 이 고객이 도착한 시점에 우선순위 i이하의 대기 중인 고객이 n_1명의 고객집단 1, n_2명의 고객집단 2, ..., n_i명의 고객집단 i이라고 하자. 이들의 서비스시간을 $S_1, S_2, ..., S_i$로 표시하자. 우선순위로 인해 추가로 고려하여야 할 사항이, 해당 고객이 기다리는 중에 보다 우선순위가 높은 고객이 오면 대기순번에서 양보를 해주어야 하고, 이것이 대기시간을 더욱 연장할 것이다. 해당 고객의 대기시간에 추가적으로 도착하는 높은 우선순위의 고객숫자를 $n'_1, n'_2, ..., n'_{i-1}$로 표시하고 이들의 서비스 시간을 $S'_1, S'_2, ..., S'_{i-1}$로 나타내기로 하자. 그리고 S_0를 해당 고객이 도착한 시점에 이미 서비스를 받고 있던 고객의 잔존 서비스 시간이라고 표시하자. 그러면 우리의 부호를 이용하여 다음과 같이 해당 고객의 대기시간을 표시할 수 있다.

$$T_i = \sum_{k=1}^{i-1} S'_k + \sum_{k=1}^{i} S_k + S_0.$$

위 식에 기댓값을 취하면

$$W_i = E[T_i] = \sum_{k=1}^{i-1} E[S'_k] + \sum_{k=1}^{i} E[S_k] + E[S_0].$$

S_0는 잔존서비스 시간이고, 대기행렬시스템이 유휴상태(idle state)이면, 즉 고객이 없는 상태에 도착을 하면 0이다. 따라서

$$E[S_0] = \Pr(busy) E[S_0 | busy] = \rho E[S_0 | busy].$$

이를 total probability를 이용하면 풀어서 표현하면 다음과 같다.

$$E[S_0 | busy] = \sum_{k=1}^{r} E[S_0 | busy - with - priority - k] \Pr(priority - k - customer).$$

이는 $\sum_{k=1}^{r} E[S_0 | busy - with - priority - k] \Pr(priority - k - customer) = \sum_{k=1}^{r} \frac{1}{\mu_k} \frac{\rho_k}{\rho}$.

여기서 $E[S_0 | busy - with - priority - k] = \frac{1}{\mu_k}$은 모든 고객의 서비스 시간이 지수분포를 따른다고 가정하였고, 따라서 지수분포의 memoryless property에 의해 기댓값이 계산된 것이다.

그러므로 $E[S_0]=\rho\sum_{k=1}^{r}\frac{1}{\mu_k}\frac{\rho_k}{\rho}=\sum_{k=1}^{r}\frac{\rho_k}{\mu_k}$로 유도된다.

그리고 개별 고객의 서비스 시간은 상호독립적이라는 가정하에, 즉 X_{kj}가 상호독립적이고 또한 n_k와도 독립적이므로 $E[S_k]=E[\sum_{j=1}^{n_k}X_{kj}]=E[n_k]E[X_{kj}]=\frac{E[n_k]}{\mu_k}$을 유도할 수 있다.

여기에 $L=\lambda W$로 표현되는 Little's formula를 적용하면

$$E[S_k]=\frac{\lambda_k W_k}{\mu_k}=\rho_k W_k.$$

동일한 방식으로 우리는 다음을 구할 수 있다: $E[S'_k]=\frac{E[n'_k]}{\mu_k}$.

여기서 n'_k는 해당 i고객이 대기하는 중에 추가로 도착하는 높은 우선순위 k의 고객숫자를 말하는 것으로, 포아송분포의 특징에 의해 $E[n'_k]=\lambda_k W_i$임을 구할 수 있다. 즉 해당 고객의 대기시간인 W_i 동안 도착하는 우선순위 k의 고객숫자의 기댓값은 해당 기간 W_i에 도착률 λ_k을 곱한 것이다.

따라서 $E[S'_k]=\frac{\lambda_k W_i}{\mu_k}=\rho_k W_i$.

이들을 모두 취합하면

$W_i=W_i\sum_{k=1}^{i-1}\rho_k+\sum_{k=1}^{i}\rho_k W_k+E[S_0]$이고, 우측 식의 첫 번째 항을 이항하면

$W_i=\frac{\sum_{k=1}^{i}\rho_k W_k+E[S_0]}{1-\sigma_{i-1}}$을 구할 수 있다. 이 식을 풀면 다음을 구할 수 있다.

$$W_i=\frac{E[S_0]}{(1-\sigma_{i-1})(1-\sigma_i)}=\frac{\sum_{k=1}^{i}\rho_k/\mu_k}{(1-\sigma_{i-1})(1-\sigma_i)}.$$

Little's formula를 이용하여 우선순위를 사용할 경우 평기대기고객숫자를 구하면

$$L=\sum_{i=1}^{r}\lambda_i W_i=\sum_{i=1}^{r}\frac{\lambda_i\sum_{k=1}^{i}\rho_k/\mu_k}{(1-\sigma_{i-1})(1-\sigma_i)}.$$

여기서 평균서비스시간이 짧은 순서대로 우선순위를 줄 경우, 즉 $\frac{1}{\mu_1} < \frac{1}{\mu_2} < \ldots < \frac{1}{\mu_r}$이 유지되도록 우선순위를 줄 경우 L, W가 최소화됨은 SPT(Shortest Processing Time) 법칙으로 널리 알려져 있다. 다음에는 우선순위를 적용할 경우 불가피하게 발생하는 순서역전에 따른 고객의 불만을 다루어 보기로 하자.

Ⅲ. 고객의 불만비용

대기행렬시스템에서 서비스시간이 짧은 것부터 우선순위를 제공할 경우, 우선순위기제에 따라 평균 대기시간과 대기인원을 최소화할 수 있다. 하지만 본 절에서는 이와 더불어 고객에게 발생하는 비용을 고려하고자 한다. 우리가 2절에서 유도한 식에서 보면 우선순위 i고객의 경우 본인이 기다리는 동안 도착한 높은 순위의 고객들로 인해 추가로 기다리게 되는 시간은 다음과 같이 유도된다.

$$W_i \sum_{k=1}^{i-1} \rho_k = \sigma_{i-1} W_i = \frac{\sigma_{i-1} \sum_{k=1}^{i} \rho_k / \mu_k}{(1-\sigma_{i-1})(1-\sigma_i)}.$$

여기서 특별한 경우로 i=1의 경우를 살펴보기로 하자. i=1의 경우 추가대기시간은 0이 되는데, 그 이유는 $\sigma_0 = 0$이어서 분자가 0이 되기 때문이다. 즉 i=1은 최고의 우선순위를 갖는 고객이므로 자신이 대기하는 동안 추가로 도착하는 고객에게 양보할 경우가 발생하지 않기 때문이다. 하지만 2 이상의 우선순위 고객의 경우 + 값의 추가대기시간이 발생함을 알 수 있다.

우리는 본 논문에서 우선순위의 기제를 이용할 경우, 종합적인 혜택이 발생하는 것에 반해 비용이 발생할 수 있음을 다루고자 한다. 즉 일반적으로 고객들이 공정하다고 받아들이는 대기행렬 순서는 FCFS(First Come First Serve)이다. 우선순위를 도입할 경우, FCFS가 고객의 우선순위에 의해 번복이 되게 된다. 따라서 이러한 순서의 번복에 대해 고객들은 부당함을 느끼고 이것이 해당 고객의 심리적 비용으로 나타날 수 있는 것이다. 이러한 서비스 순서 번복에 따른 비용을 불편비용이라고 부르기로 하자. 고객 i가 우선순위기제(priority mechanism)를 적용함으로써 추가로 대기하는 시간에 의해 인지하는 불만비용을 함수 $C_i(.)$로 표시하면, 우선순위에 따른 총평균 불만비용은 $\sum_{i=1}^{r} C_i\left(\frac{\sigma_{i-1} \sum_{k=1}^{i} \rho_k / \mu_k}{(1-\sigma_{i-1})(1-\sigma_i)}\right)$이 된다.

따라서 우리는 우선순위의 효과(고객의 평균대기 시간의 감소)와 함께 이로 인한 추가 대기시간에 따른 불만비용을 종합적으로 고려하여 의사결정을 하여야 한다. 우리가 다루는 상황에 따라 다양한 형태의 불만비용 함수를 고려할 수 있을 것이다. 앞서의 불만비용에서는 우선순위에 의해 자신의 순서가 추월되어 자신이 추가로 기다리는 시간에 대해 고객집단별로 인식되는 비용을 다룬 것이다. 다른 경우로는 모든 고객집단의 우선순위로 인해 순서에서 밀려 발생하는 추가대기시간의 총합에 대한 비용함수를 고려하는 것이 타당할 경우가 있을 것이다. 즉 고객집단별 추가대기시간의 총합인 $\sum_{i=1}^{r}\frac{\sigma_{i-1}\sum_{k=1}^{i}\rho_k/\mu_k}{(1-\sigma_{i-1})(1-\sigma_i)}$의 함수로 다음과 같이 총불만비용을 표현할 수 있는 것이다.

$$C(\sum_{i=1}^{r}\frac{\sigma_{i-1}\sum_{k=1}^{i}\rho_k/\mu_k}{(1-\sigma_{i-1})(1-\sigma_i)}).$$

또 다른 가능성의 하나로 다음 경우를 살펴보기로 하자. 고객의 불만이 자신이 대기하는 중에 추월당하는 것에 대한 불만이 있을 수도 있지만, 원래 FCFS에 따른 경우에 비해 평균대기 시간이 증가하는 것에 대한 불만이 발생할 수도 있다. 이때 우리가 유념하여야 하는 것은 FCFS에 비해 평균대기시간이 감소하는 경우에 대해서는 전혀 불만이 없을 것이라는 사실이다. 우선순위기제를 사용하지 않고 FCFS를 사용한 경우 고객집단 i의 평균대기시간을 No priority를 한다고 보고 N으로 표시하기로 하자. 따라서 우리는 고객집단 i의 평균대기시간을 W_i^N로 표시하기로 하자. 그러면 고객집단 i의 불만비용은 $C_i([W_i - W_i^N]^+)$로 표현할 수 있다. $[W_i - W_i^N]^+$은 고객집단 i가 FCFS에 따른 경우에 비해 우선순위기제를 따를 경우 평균대기 시간이 증가하는 것을 나타낸다. $[x]^+ = x$ if $x > 0$, 0 if $x \le 0$이므로, $W_i < W_i^N$인 경우 해당 함수 값은 0이 되므로, 즉 FCFS에 비해 평균대기시간이 감소하는 경우에 대해서는 전혀 불만이 없는 것이 됨을 알 수 있다. 이 경우 고객집단의 총불만비용은 $\sum_{i=1}^{r}C_i([W_i - W_i^N]^+)$ 혹은 $C(\sum_{i=1}^{r}[W_i - W_i^N]^+)$으로, 상황에 따라 달리 표현할 수 있다.

여기서 유념할 것이 있는데, 이를 살펴보기로 하자. r개의 고객집단에 대해 우선순위를 적용한다는 것은, 우선순위가 높은 집단은 순서의 혜택으로 인해 FCFS보다 평균대기시간이 감소하는 혜택을 얻는 반면, 우선순위가 낮은 집단은 FCFS에 비해 평균대기시간이 증가하는 것을 감수하게 된다. 평균서비스시간이 짧은 순서대로 우선순위를 적용하면 총체적으로 대기시간의 평균과 대기

인원의 평균을 최소화할 수 있지만, 이는 낮은 순위의 고객집단의 양보 내지는 희생을 감수함으로써 가능해 지는 것이다. 우선순위 1의 고객집단은 분명히 FCFS보다는 평균대기시간이 감소될 것이다. 또한 우선순위 r의 고객집단은 분명히 FCFS보다는 평균대기시간이 증가할 것이다. 우리는 임계치 k가 있어, k보다 우선순위가 높은 고객집단은 FCFS보다 평균대기시간이 감소하고, 즉 $W_i < W_i^N$이 성립하고 반면에 k 이상의 우선순위 고객집단의 경우 $W_i \geq W_i^N$이 성립하여 우선순위로 인해 개별 집단은 희생을 하게 되는 경우가 발생할 것이라고 추정할 수 있다.

방금 우리가 다루는 경우가, 앞서 다룬 추가대기시간의 발생에 따른 불만비용의 경우와는 분명한 차이가 있음을 유념하기 바란다. 추가대기시간의 경우에는 $i=1$인 경우를 제외하고, 즉 $i \geq 2$인 경우에는 반드시 + 값의 추가대기시간이 발생함을 알 수 있다. 따라서 $i \geq 2$인 경우 반드시 고객의 불만비용이 발생한다. 반면 FCFS와의 비교의 경우 우선순위가 가장 높은 순서를 고려할 경우 일정 수(대부분의 경우 1을 초과)의 고객집단에 대해서는 불만비용이 발생하지 않는다.

Ⅳ. 두 개 우선순위의 경우

여기서는 보다 구체적인 비교를 위해 2개의 고객집단이 있는 경우를 살펴보기로 하자. 먼저 부호를 정의하자. $\lambda = \lambda_1 + \lambda_2, \hat{\rho} = \lambda/\mu_1, \rho_1 = \lambda_1/\mu_1, \rho_2 = \lambda_2/\mu_2, \rho = \rho_1 + \rho_2$.

〔Gross & Harris, 1998〕의 내용에 따르면, 우선순위를 적용할 경우 고객집단별 평균대기인원숫자와 대기행렬시스템 전체 대기인원숫자의 평균은 다음과 같다.

$$L_1 = \rho_1 \hat{\rho} \frac{\lambda_1/\lambda + (\lambda_2/\lambda)(\mu_1/\mu_2)^2}{1-\rho_1},$$

$$L_2 = \frac{\lambda_2/\mu_1}{1-\rho_1} \hat{\rho} \frac{\lambda_1/\lambda + (\lambda_2/\lambda)(\mu_1/\mu_2)^2}{1-\rho_1-\rho_2},$$

$$L = L_1 + L_2 = \frac{1-\rho_1[\lambda_1/\lambda + \rho_2/\hat{\rho}]}{1-\rho_1} \frac{\rho_1\hat{\rho} + \rho_2\lambda/\mu_2}{1-\rho_1-\rho_2} = \frac{1-\rho_1[\lambda_1/\lambda + \rho_2/\hat{\rho}]}{1-\rho_1} \frac{\lambda(\rho_1/\mu_1 + \rho_2/\mu_2)}{1-\rho_1-\rho_2}.$$

비교를 위해 FCFS를 따른 경우를 살펴보면, 다음의 결과를 얻을 수 있다.

$$L_1^N = \rho_1 \hat{\rho} \frac{1-(1-\mu_1/\mu_2)\rho_2}{1-\rho_1-\rho_2},$$

$$L_2^N = \frac{\lambda_2}{\mu_1}\hat{\rho}\frac{(\mu_1/\mu_2)^2 + (1-\mu_1/\mu_2)(\lambda_1/\mu_2)}{1-\rho_1-\rho_2},$$

$$L^N = \frac{\rho^2}{1-\rho}\frac{\lambda_1/\lambda + (\lambda_2/\lambda)(\mu_1/\mu_2)^2}{[\lambda_1/\lambda + (\lambda_2/\lambda)(\mu_1/\mu_2)]^2}.$$

우리는 평균서비스 시간이 짧은 고객집단에 우선순위를 주는 경우, 즉 $\frac{1}{\mu_1} < \frac{1}{\mu_2} \Leftrightarrow \mu_1 > \mu_2$일 경우 $L < L^N$임을 알 수 있다. 또한 고객집단 1은 우선순위를 가져 대기시간이 짧아지는 등 혜택을 보고, 반면에 고객집단 2는 오히려 피해를 보게 된다. 즉 $L_1 < L_1^N, L_2 > L_2^N$. 하지만 총합의 경우를 보면 효과적임을 알 수 있을 것이다. 이는 고객집단 1의 혜택이 고객집단 2의 손해를 초과하기 때문에 발생하는 것이다.

구체적으로 $L < L^N$을 살펴보기로 하자. 위의 L을 다음의 식들을 이용하여 다시 정리해 볼 수 있다.

$$1-\rho_1[\lambda_1/\lambda + \rho_2/\hat{\rho}] = 1 - \frac{\lambda_1}{\lambda}(\rho_1+\rho_2),$$

$$[\lambda_1 + \lambda_2\frac{\mu_1}{\mu_2}]^2 = \mu_1^2(\rho_1+\rho_2)^2,$$

$$\mu_1^2(\frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}) = \mu_1\rho_1 + \frac{\mu_1^2}{\mu_2}\rho_2 = \lambda_1 + \lambda_2(\frac{\mu_1}{\mu_2})^2.$$

이들 식을 이용하면 우리는 다음과 같이 L을 새로운 형태로 표현할 수 있다.

$$L = \frac{\rho^2}{1-\rho}\frac{\lambda_1/\lambda + (\lambda_2/\lambda)(\mu_1/\mu_2)^2}{[\lambda_1/\lambda + (\lambda_2/\lambda)(\mu_1/\mu_2)]^2}\frac{1-\lambda_1\rho/\lambda}{1-\lambda_1\hat{\rho}/\lambda}.$$

우리는 이제 L, L^N을 비교할 수 있다.

$$\frac{L}{L^N} < 1 \Leftrightarrow 1-\rho\frac{\lambda_1}{\lambda} < 1-\hat{\rho}\frac{\lambda_1}{\lambda} \Leftrightarrow \rho\frac{\lambda_1}{\lambda} > \hat{\rho}\frac{\lambda_1}{\lambda} \Leftrightarrow \mu_2 < \mu_1.$$

이로써 평균서비스시간이 짧은 1에게 우선순위를 부여할 때 평균대기고객숫자가 감소한다는 사실을 증명한 것이다.

V. 현실 적용 및 결어

우리는 본 논문에서 대기행렬 모형에서 널리 알려진 우선순위에 대해 다루었다. 서비스 시간이 다른 다양한 고객군이 존재할 때 서비스 시간이 작은 순서대로 우선순위를 제공할 경우 전체 고객에 대한 평균 대기시간 및 대기고객의 숫자가 최소화된다는 사실은 알려져 있다. 우리는 본 논문에서 이와 같이 우선순위를 부여할 때 고객의 불만이 존재할 수 있음을 다루었다. 자신이 부여받은 우선순위보다 높은 우선순위의 고객이 도착할 경우 자신이 순서에서 밀려날 수 있고, 이에 대해 불만을 갖게 되는 것이 일반적이다. 대기행렬시스템 전체의 성과와 개별 고객집단의 성과 사이에 상충관계가 존재하게 되는 것이다. 물론 대기행렬시스템 전체를 고려할 때 적절한 우선순위 기제를 활용함으로써 최적의 성과를 낸다는 것은 개별 집단의 손해보다는 유익한 효과가 더욱 크기 때문에 총체적으로는 이익이 된다는 것을 의미한다. 하지만 개별 고객 집단이 어느 정도 순서에서 양보하여야 이러한 결과를 얻을 수 있는 것이다. 양보하는 것이 추가 시스템 전체적으로 얻게 되는 추가 효익의 양보다 작지만, 개별 고객집단의 입장에서는 손해를 보는 것이다.

대기행렬 모형에서 고객의 입장에서 가능한 대기시간을 최소화하는 것이 가장 바람직하다는 전제를 갖고 분석하는 경우가 많다. 하지만 동일한 시간을 기다리는 경우에도, 고객이 본인보다 나중에 도착한 고객이 본인보다 먼저 서비스를 받는 경우 매우 큰 불만을 느끼게 되는 경우가 많다. 본 논문에서는 이와 같은 고객의 불만을 다루고자 한 것이다. 이러한 개별 고객집단의 양보에 대한 불만을 비용화하여 우선순위로부터 얻는 효익과 비교를 하여야 한다는 인식을 다룬 것이다. 개별 고객집단이 순서에서 손해를 보는 것을 비용화하는 데에는 다양한 가능성이 있을 것이다. 또한 우리의 논문에서는 불만을 비용화하는 과정에 고객들이 합리적으로 이를 인식하고 계산할 수 있을 것이라는 가정이 전제되어 있다.

우리는 본 논문의 내용을 실제 상황에 적용하는데 있어, 다음을 고려하는 것이 필요하다. 불만비용은 추정하기가 어렵고 경우에 따라 볼록 급증(convex increasing)하는 함수의 형태를 띠는 경우가 있다. 우선순위 기제의 효과는 널리 알려져 있지만, 이의 바람직한 적용은 대기 고객이 우선순위를 적용하는 것을 관찰하지 못하도록 하거나, 관찰하지 못하는 경우에 한정하는 것이 보다 효과적일 수 있다. 가령 긴급출동 서비스의 처리순서를 정할 때 우선순위를 적용하는 것이 사회적으로 보다 바람직할 수 있다. 접수순서대로 출동을 하는 것이 아니라 긴급상황을 고려하여 우선순위를 부여할 수 있다. 또한 콜센터를 운영할 때 기업의 입장에서 우선순위를 부여하는 기제를 활용할 수 있다. 즉 고객집단별로 별도의 안내전화번호를 운영하는 것이다. 이와 같이 고객이 관찰하기 힘든 경우에는 우선순위를 통해 전체 효율을 최적화할 수 있다. 다만 고객이 대기행렬에서 기다리면

서 도착하는 고객을 관찰할 수 있는 경우에는, 본 논문에서 다룬 불만비용을 고려하는 것이 보다 타당한 최적화 모형이 될 것이다.

참 고 문 헌

1. Gross, Donald and Harris, Carl M., Fundamentals of Queueing Theory, John Wiley & Sons, Inc., 1998.