

유한 재시도 $M/M/1/1$ 대기행렬의 근사적 분석

성기원 · 채경철

한국과학기술원 산업공학과

An Approximate Analysis of the $M/M/1/1$ Queue with Finite Number of Retrials

Ki-Won Sung · Kyung-Chul Chae

We consider the $M/M/1/1$ retrial queue where the maximum number of retrials is fixed by a constant. We present an efficient approximate procedure for mean performance measures and the loss probability. The approximate results are satisfactory when compared with simulation results.

1. 서론

전화를 걸었다가 통화중이면 얼마 후에 재시도(retrial)하는 현상은 Cohen(1957)이 처음 대기행렬로 모형화했다. 이에 따른 후속 연구는 특히 최근에 활발한데, 이는 재시도 모형이 통신시스템에 널리 활용되기 때문이다. 조사(survey) 논문으로는 Falin(1990)과 Yang and Templeton(1987)을 꼽을 수 있으며, 재시도 모형만을 다룬 책(Falin and Templeton, 1997)도 나와 있다.

기존의 재시도 모형에서는 대개 재시도 횟수를 제한하지 않는다. 이러한 “무한” 재시도 모형은 크게 두 가지로 나눌 수 있다. 첫째는 재시도가 성공할 때까지 계속 재시도를 반복하는 경우로서 수학적 난이도가 제일 낮은 경우이다. 둘째는 기하손실(geometric loss) 재시도 모형이라 불리는 경우인데, 첫 시도를 실패한 고객은 P_1 의 확률로 재시도를 포기하고(시스템을 이탈하며) $1-P_1$ 의 확률로는 얼마 후에 재시도를 한다. 이후, 재시도가 실패할 때마다 P_2 의 확률로 더 이상의 재시도를 포기하고 $1-P_2$ 의 확률로는 얼마 후에 다시 재시도한다. 물론 $P_1=P_2=0$ 인 특수한 경우가 위의 첫 번째 경우이다.

본 연구에서는 재시도 횟수를 최대한 K 회로 제한하는 모형을 다룬다. 이는, 예를 들어, 전송한 패킷(packet)이 충돌하면 얼마 후에 재전송을 하는데, 이 때 프로토콜로서 재전송 횟수를 K 회로 제한하는 경우에 해당된다. 이러한 “유한” 재시도 모형에 대한 연구는 거의 없다. 저자가 알기로는 김영철(1997, 제2장)의 연구가 유일한데, 이는 $K=1$ 이고 $c \leq 2$ 인 경우에 대해서 $M/M/c/c$ 대기행렬을 분석한 것이다.

[비고 1] Falin and Templeton(1997)의 책이 나오기 전에 김영철(1997)의 논문이 발표된 것으로 사료된다. 그 이유는 김영철이 다룬 모형이 Falin and Templeton의 기하손실 재시도 모형(218쪽과 250쪽 참조)에 $P_1=0$ 과 $P_2=1$ 을 대입한 것과 동일하기 때문이다.

본 연구에서는 $K \geq 2$ 인 경우에 대해서 $M/M/1/1$ 대기행렬을 근사적으로 분석한다. 근사적 분석의 주된 목표는 손실(loss) 확률인데, 이는 첫 시도와 K 번의 재시도가 모두 실패할 확률이다(비고: 최적의 K 값을 결정하기 위한 가장 중요한 정보는 손실 확률임). 또한, 평균고객수와 평균체류시간 등의 평균 성능척도들을 근사적으로 구한다. 그리고 이러한 근사치들을 모의실험(simulation) 결과와 비교한다.

2. 모형, 용어 및 표기

고객은 도착률이 λ 인 포아송 과정에 따라 도착한다. 고객들의 서비스 시간은 평균이 μ^{-1} 인 지수분포를 따르며 서로(독립이고 또한 도착과정과도) 독립이다. 서버(server)는 한 명인데, 서버가 유휴중(idle)일 때 도착하는 고객은 바로 서비스를 받기 시작하지만(시작한 서비스는 쉬지 않고 끝까지 처리함) 서버가 바쁠(busy) 때 도착하는 고객은 웨도(orbit)라고 불리는 가상적인 대기장소에 들어간다. 웨도에 있는 고객들은 각각 독립적으로(그리고 기타 모든 것과 독립적으로) 평균이 γ^{-1} 인 지수 시간 후에 재시도를 한다. 그러나 K 번째 재시도를 실패한 고객은 더 이상의 재시도를 포기하고 시스템을 떠난다.

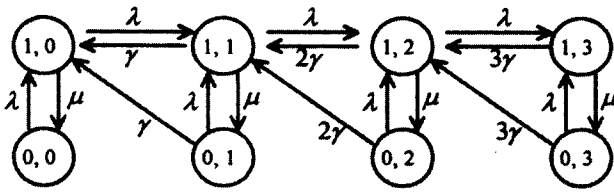


그림 1. 재시도가 1회로 제한된 M/M/1/1의 상태전이도.

안정상태(steady-state)에서 서비스를 받고 있는 고객의 수를 M 이라 하고, 큐에 있는 고객들 중에서 i 번째로 재시도할 고객의 수를 N_i 라 하자 ($i=1, 2, \dots, K$). 시스템의 상태를 벡터 $(M, N_1, N_2, \dots, N_K)$ 로 정의하면 상태의 전이는 연속시간 마야코프 연쇄를 따른다. 따라서, 평형방정식(balance equation)을 풀기만 하면 이로부터 각종 성능척도(의 확률분포 및 기대치)를 얻을 수 있다. 그런데 문제는 평형방정식을 풀기가 어렵다는 점이다.

[비고 2] 유한 용량(buffer) 시스템에서 안정상태 조건이 불필요하듯이, 유한 재시도 시스템에서도 안정상태 조건이 불필요하다. 다만, 비현실적인 상황을 피하기 위해서 $0 < \lambda, \mu, \gamma < \infty$ 라 가정한다.

3절에서 다룰 $K \geq 2$ 인 경우를 잘 이해하기 위해서, 먼저 $K=1$ 인 경우에 대해 알려진 결과([비고 1] 참조)를 요약하고 이를 의미있게 해석한다. 시스템의 상태를 (M, N_1) 으로 정의하면, 상태전이도는 <그림 1>과 같다.

그리고 $P_{m,n} = P(M=m, N_1=n)$ 이라 하면 평형방정식은 다음과 같다($n=0, 1, 2, \dots$).

$$(\lambda + n\gamma)P_{0,n} = \mu P_{1,n} \quad (1)$$

$$(\lambda + \mu + n\gamma)P_{1,n} = \lambda P_{0,n} + (n+1)\gamma(P_{0,n+1} + P_{1,n+1}) + \begin{cases} 0, & \text{if } n=0 \\ \lambda P_{1,n-1}, & \text{if } n \geq 1 \end{cases} \quad (2)$$

식 (1)과 (2)는 간단해 보이지만 풀이과정은 복잡하다. 그리고 대부분의 결과는 아래에 정의된 소위 Kummer confluent 함수 $\Phi(a, b; w)$ 에 $a = w = \lambda/\gamma$ 와 $b = (\lambda + \mu + \gamma)/\gamma$ 를 대입한 것으로 표현된다.

$$\begin{aligned} \Phi(a, b; w) &= \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{wt} t^{a-1} (1-t)^{b-a-1} dt \\ &= 1 + \frac{a}{b} w + \frac{a(a+1)}{b(b+1)} \frac{w^2}{2!} \\ &\quad + \frac{a(a+1)(a+2)}{b(b+1)(b+2)} \frac{w^3}{3!} + \dots \end{aligned} \quad (3)$$

예를 들어, 서버가 바쁠 확률을 P_B , 유휴중일 확률을 P_I 라 하면

$$P_B = \sum_{n=0}^{\infty} P_{1,n} = \Lambda / (1 + \Lambda) \quad (4)$$

$$P_I = \sum_{n=0}^{\infty} P_{0,n} = 1 / (1 + \Lambda) \quad (5)$$

인데, 이때 Λ 는 다음과 같다 ($\rho = \lambda/\mu$).

$$\Lambda = \rho \Phi(a+1, b; w) / \Phi(a, b; w) \quad (6)$$

[비고 3] $K=1$ 이더라도 $\gamma \rightarrow \infty$ 이면 $K=0$ 인 경우 (즉, 재시도가 허용되지 않는 M/M/1/1 경우)와 동일한 상황이 되기 때문에 $\Lambda \rightarrow \rho$ 가 된다.

큐에 있는 평균고객수를 L 이라 하고, 이를 편의상 $L = P_I L_I + P_B L_B$ 로 표현하자(비고: L_I 와 L_B 는 조건부 기대치인데, 각각 서버가 유휴중일 때와 바쁠 때의 평균고객수를 나타냄). 식 (1)을 모든 n 에 대해서 합치면 다음을 얻는다.

$$(\lambda + \gamma L_I) P_I = \mu P_B \quad (7)$$

식 (7)을 Little's 공식의 관점으로 해석하면 다음과 같다. 서비스를 받고 있는 평균고객수는 P_B 이고 고객의 평균체류시간은 μ^{-1} 이므로 단위시간당 서비스받기 시작하는 평균고객수는 $(\lambda + \gamma L_I) P_I$ 이다.

반면에, 큐에 있는 고객에 대해서 Little's 공식을 적용하면 다음을 얻는다.

$$L = \lambda P_B \cdot \gamma^{-1} \quad (8)$$

단위시간당 큐에 들어오는 평균고객수는 λP_B 명인데 이들이 큐에 체류하는 평균시간은 γ^{-1} 이다(비고: 식 (8)은 식 (1)과 (2)로부터 얻을 수도 있음). 식 (8)에서 유의할 점은 PASTA 속성(Wolff, 1982)이다. 즉, 포아송 과정으로 도착하는 고객들 중에서 서버가 바쁠 때 도착해서 큐로 들어가는 고객의 비율은 (시간평균 확률인) P_B 와 동일하다. 식 (8)은 또한 다음과 같이 표현할 수도 있다.

$$\lambda P_B = \gamma L = \gamma P_I L_I + \gamma P_B L_B \quad (9)$$

식 (9)에서 좌변은 큐에 대한 고객의 입력(input)률이며 우변은 출력(output)률인데, 단위시간당 큐를 이탈하는 평균고객수 γL 명 중에서 $\gamma P_I L_I$ 명은 재시도에 성공하지만 $\gamma P_B L_B$ 명은 재시도에 실패한다. 따라서 재시도가 실패할 확률 α 는 다음과 같다(비고: α 는 큐를 이탈하는 고객이 재시도에 실패할 "조건부" 확률임. 따라서, 손실확률을 P_{loss} 라 하면 " $P_{loss} = P_B \alpha$ "가 성립함).

$$\alpha = \frac{\gamma P_B L_B}{\lambda P_B} = \frac{\gamma}{\lambda} L_B \quad (10)$$

그런데 L_B 와 L 이 서로 다르므로, 재시도가 실패할 확률인

$\alpha = \gamma L_B / \lambda$ 는 첫 시도가 실패할 확률인 $P_B = \gamma L / \lambda$ 와 다르다.

[비고 4] α 에 대한 근사치로 P_B 를 사용하는 방법을 RTA (Retrials see time averages) 방법이라 한다 (Wolff, 1989, p. 369).

식 (3)의 적분이 용이한 $\lambda = \mu = \gamma$ 경우를 예로 들어보자([비고 2] 참조). $a = w = \lambda / \gamma = 1$ 과 $b = (\lambda + \mu + \gamma) / \gamma = 3$ 을 식 (3)에 대입하면 $\Phi(a, b; w) = e - 2$ 와 $\Phi(a + 1, b; w) = 1$ 을 얻는데, 이를 $\rho = \lambda / \mu = 1$ 과 함께 식 (6)에 대입하면 $A = (e - 2)^{-1}$ 을 얻는다. 그리고 A 를 식 (4)와 (5)에 대입하면 $P_B = (e - 1)^{-1}$ 과 $P_I = (e - 2) / (e - 1)$ 을 얻고 이들을 식 (7)~(10)에 대입하면 $L = (e - 1)^{-1}$, $L_I = (3 - e) / (e - 2)$, $L_B = e - 2$, $\alpha = e - 2$ 를 얻는다. 따라서, 재시도가 실패할 확률인 $\alpha = e - 2 \approx 0.7183$ 은 첫 시도가 실패할 확률인 $P_B = (e - 1)^{-1} \approx 0.5820$ 과 상당한 차이가 있음을 알 수 있다 ([비고 4] 참조). 또한 손실 확률로 $P_{loss} = P_B \alpha = (e - 2) / (e - 1)$ 을 얻는다.

3. $K \geq 2$ 경우에 대한 근사적 분석

$K \geq 2$ 인 경우에는 편의상 계도를 K 개로 나누어서 i 번째, $i = 1, 2, \dots, K$, 계도에 체류중인 고객수 N_i 의 기대치를 L_i 라 하고 이를 편의상 $L_i = P_I L_i^I + P_B L_i^B$ 로 표현한다(비고: L_i^I 와 L_i^B 는 조건부 기대치인데, 각각 서버가 유휴중일 때와 바쁠 때에 계도 i 에 체류중인 평균고객수를 나타냄). 그리고 $L = \sum_{i=1}^K L_i$, $L_I = \sum_{i=1}^K L_i^I$, $L_B = \sum_{i=1}^K L_i^B$ 라고 하자.

식 (7)은 $K \geq 2$ 인 경우에도 유효한데, 이는 서버가 여전히 한 명이기 때문이다. 반면에, 식 (9)와 (10)은 계도 1에서만 다음과 같이 적용된다.

$$\lambda P_B = \gamma L_1 = \gamma P_I L_1^I + \gamma P_B L_1^B \quad (11)$$

$$\alpha_1 = \frac{\gamma P_B L_1^B}{\lambda P_B} = \frac{\gamma}{\lambda} L_1^B \quad (12)$$

식 (11)의 우변은 단위시간당 계도 1을 이탈하는 평균고객수인데, 이 중에서 $\gamma P_I L_1^I$ 명은 재시도에 성공하고 $\gamma P_B L_1^B$ 명은 재시도에 실패해서 계도 2로 들어간다. 따라서 첫 번째 재시도가 실패할(조건부) 확률인 α_1 은 식 (12)로 주어진다.

계도 $i = 2, 3, \dots, K$ 에 대한 “입력률=출력률” 등식은 다음과 같다.

$$\gamma P_B L_{i-1}^B = \gamma L_i = \gamma P_I L_i^I + \gamma P_B L_i^B \quad (13)$$

식 (13)에서 $\gamma P_B L_{i-1}^B$ 는 단위시간당 $i-1$ 번째 재시도를 실패하는 평균고객수이고 γL_i 는 단위시간당 i 번째 재시도를 하는 평균고객수이다. 그리고 γL_i 명 중에서 $\gamma P_I L_i^I$ 명이 i 번째 재시도에 성공하고 $\gamma P_B L_i^B$ 명은 실패하므로 i 번째 재시도가 실패할 확률 α_i 는 다음과 같다(비고: α_i 는 계도 i 를 이탈하는 고객이 재시도에 실패할 “조건부” 확률임).

$$\alpha_i = \frac{\gamma P_B L_i^B}{\gamma P_B L_{i-1}^B} = \frac{L_i^B}{L_{i-1}^B}, \quad i = 2, 3, \dots, K \quad (14)$$

이제, 평균 성능척도들을 α_i 로 표현한다($i = 1, 2, \dots, K$). 식 (12)와 (14)로부터 $L_i^B = (\lambda / \gamma) \prod_{j=1}^i \alpha_j$ 를 얻고 이를 $L_B = \sum_{i=1}^K L_i^B$ 에 대입하면

$$L_B = (\lambda / \gamma) \sum_{i=1}^K \prod_{j=1}^i \alpha_j \quad (15)$$

를 얻는다. 또한, 식 (11)과 (13)을 (모든 i 에 대해서) 더하면 $\lambda P_B = \gamma P_I L_I + \gamma P_B L_K^B$ 를 얻는데, 이를 식 (7)과 연립으로 풀면

$$P_B = \lambda \left(\mu + \lambda \prod_{i=1}^K \alpha_i \right) \quad (16)$$

를 얻는다. 그리고 L_i 는 $P_I = 1 - P_B$ 와 식 (16)을 식 (7)에 대입해서 얻을 수 있으며, 이들을 모두 $L = P_I L_I + P_B L_B$ 에 대입하면 L 도 얻을 수 있다.

[비고 5] 식 (16)을 더욱 간단하게 얻는 방법은 다음과 같다. 손실확률인 $P_{loss} = P_B \prod_{i=1}^K \alpha_i$ 는 도착하는 고객이 서비스를 받지 못하고 시스템을 이탈할 확률이므로, 식 (7)의 좌변을 $\lambda(1 - P_{loss})$ 로 대체하여 P_B 에 대한 방정식을 얻은 다음 이를 풀면 식 (16)을 얻는다.

본 연구의 주된 목표는 $P_{loss} = P_B \prod_{i=1}^K \alpha_i$ 에 대한 근사치를 얻는 것이다. 재시도 횟수를 최대한 K 회로, $K \geq 2$, 제한하는 모형에 대해서는 본 연구가 처음이므로([비고 1] 참조), P_{loss} 의 근사치 역시 지금까지 구해진 적이 없다. 먼저, 무한 재시도 모형에 활용되는 RTA 방법([비고 4] 참조)으로 P_{loss} 의 근사치를 구한 다음, 이를 개선시킨 근사치를 제안한다.

RTA 방법에 의하면 $\alpha_i \approx P_B$ 이므로, $i = 1, 2, \dots, K$, α_i 및 P_B 의 근사치를 $\tilde{\alpha}$ 이라 하면

$$P_{loss} \approx \tilde{\alpha}^{K+1} \quad (17)$$

인데, $\tilde{\alpha}$ 에 대한 방정식은 식 (16)으로부터 다음과 같이 얻는다.

$$\lambda \tilde{\alpha}^{K+1} + \mu \tilde{\alpha} - \lambda = 0 \quad (18)$$

[비고 6] 식 (18)은 (0, 1) 구간에서 하나의 실근을 가진다. 이는 식 (18)의 좌변을 $f(\hat{\alpha})$ 라 할 때 $\lambda, \mu > 0$ 이므로 $f(0) < 0, f(1) > 0$ 이고 또한 (0, 1) 구간에서 $\frac{d}{d\hat{\alpha}} f(\hat{\alpha}) > 0$ 이기 때문이다.

RTA 방법은 한 마디로 $P_B \approx \alpha_1 \approx \dots \approx \alpha_K$ 인데, 본 연구에서 제안하는 개선책은 $P_B \neq \alpha_1 \approx \dots \approx \alpha_K$ 이다. 1 단계로, $\alpha_i \approx \alpha_1, i = 2, 3, \dots, K$ 를 식 (15)와 (16)에 대입하면 다음을 얻는다.

$$L_B \approx (\lambda/\gamma)(\alpha_1 + \alpha_1^2 + \dots + \alpha_1^K) \quad (19)$$

$$P_B \approx \lambda/(\mu + \lambda \alpha_1^K) \quad (20)$$

2단계에서는 α_1 의 근사치를 구한다. 편의상 서버가 바쁠 때 도착하는 임의의 고객을 시험(test) 고객이라 부르자. PASTA 속성에 의해서, 시험고객이 도착할 때 큐도에 체류중인 평균고객 수는 L_B 명인데 이를 N_B 의 기대치라 하자. 그리고 $N_B = n$ 이라는 조건하에 시험고객이 재시도에 실패할 확률을 β_n 이라 하면

$$\alpha_1 = \sum_{n=0}^{\infty} \beta_n \cdot P(N_B = n) \quad (21)$$

이 성립한다. 편의상 θ 와 ϕ 를 다음과 같이 정의한다.

$$\theta = \{\lambda + (1+n)\gamma + \mu\}^{-1}$$

$$\phi = \{\lambda + (1+n)\gamma\}^{-1}$$

그러면, β_n 에 대한 순환(recursive)식은 다음과 같다(<부록> 참조)(비고: $n=0$ 인 경우에는 $n\beta_{n-1}$ 을 0으로 간주함).

$$\beta_n = \theta\lambda \cdot \beta_{n+1} + \theta\gamma$$

$$+ \theta n\gamma \cdot (\beta_{n-1} \text{ 또는 } \beta_n) \quad (22)$$

$$+ \theta\mu(\phi\lambda \cdot \beta_n + \phi n\gamma \cdot \beta_{n-1})$$

근사적 방법의 2단계는 식 (22)에서 β_{n-1} 과 β_{n+1} 을 β_n 으로 대체하는 것인데, 이에 따라

$$\beta_n \approx \frac{\lambda + (1+n)\gamma}{\lambda + (1+n)\gamma + \mu} \quad (23)$$

을 얻는다. 그리고 식 (23)을 식 (21)에 대입하면 다음을 얻는다.

$$\alpha_1 \approx E\left[\frac{\lambda + (1+N_B)\gamma}{\lambda + (1+N_B)\gamma + \mu}\right] \quad (24)$$

α_1 에 대한 최종적인 근사치 $\hat{\alpha}$ 은 다음과 같다.

$$\alpha_1 \approx \hat{\alpha} = \frac{\lambda + (1+L_B)\gamma}{\lambda + (1+L_B)\gamma + \mu} \quad (25)$$

[비고 7] Jensen's 부등식에 의하면 식 (25)의 우변은 식 (24)의 우변보다 크거나 같다. 즉, 식 (23)의 우변을 $f(n)$ 이라 하면 $f(n)$ 은 오목(concave) 함수이므로 $f(E[N_B]) \geq E[f(N_B)]$ 가 성립한다.

식 (19)를 식 (25)에 대입하면 $\hat{\alpha}$ 에 대한 방정식으로

$$\lambda \hat{\alpha}^{K+1} + (\mu + \gamma)\hat{\alpha} - (\lambda + \gamma) = 0 \quad (26)$$

을 얻는데, 이는 (0, 1) 구간에서 하나의 실근을 가진다 ([비고 6] 참조).

평균 성능척도들은 모두 $\alpha_i, i = 1, 2, \dots, K$ 로 표현할 수 있으므로 α_i 에 식 (26)의 근을 대입하면 이들의 근사치를 얻을 수 있다. 또한, 식 (20)에 의해서 P_B 의 근사치 \hat{P}_B 와 P_{loss} 의 근사치 \hat{P}_{loss} 는 다음과 같다.

$$\hat{P}_B = \lambda/(\mu + \lambda \hat{\alpha}^K) \quad (27)$$

$$\hat{P}_{loss} = \hat{P}_B \hat{\alpha}^K = \lambda \hat{\alpha}^K / (\mu + \lambda \hat{\alpha}^K) \quad (28)$$

$K \geq 2$ 인 경우에 대한 \hat{P}_B 와 \hat{P}_{loss} 의 성능은 4절에서 모의실험을 통해서 보인다. 2절에서 예로 들었던 $K=1, \lambda = \mu = \gamma$ 인 경우에는 식 (26), (27), (28)로부터 $\hat{\alpha} = \sqrt{3} - 1 \approx 0.7321, \hat{P}_B = 1/\sqrt{3} \approx 0.5774, \hat{P}_{loss} = (\sqrt{3} - 1)/\sqrt{3} \approx 0.4226$ 을 얻는다. 이들을 참값과 비교하면 $\hat{\alpha}$ 은 $\alpha = e - 2 \approx 0.7183$ 보다 약간 크고 \hat{P}_B 는 $P_B = (e - 1)^{-1} \approx 0.5820$ 보다 약간 작으며 \hat{P}_{loss} 는 $P_{loss} = (e - 2)/(e - 1) \approx 0.4180$ 보다 약간 크다. 반면에, RTA 방법에 의한 근사치는 식 (18)과 (17)로부터 $\hat{\alpha} = (\sqrt{5} - 1)/2 \approx 0.6180$ 과 $P_{loss} \approx 0.3820$ 을 얻는데, 이들의 성능은 $\hat{\alpha}$ 과 \hat{P}_{loss} 에 비해서 떨어진다. $\hat{\alpha}$ 와 \hat{P}_B 에 관해서 주목할 점은 다음과 같다. $\hat{\alpha}$ 은 α 를 약간 과대평가하고 \hat{P}_B 은 P_B 를 약간 과소평가하므로 $\hat{P}_{loss} = \hat{P}_B \hat{\alpha}^K$ 에서는 과대평가 요인과 과소평가 요인이 어느 정도 상쇄된다.

4. 모의실험 결과 및 분석

본 연구에서는 재시도 횟수가 최대한 K 회로 제한된 M/M/1/1 대기행렬을 근사적으로 분석하였다. 그런데 $K \geq 2$ 경우에 대해서는 기존의 연구가 없으므로, 본 연구에서 제시한 근사법의 성능은 모의실험을 통해서 알아볼 수밖에 없다.

3절에서는 $K=1$ 이고, " $\lambda = \mu = \gamma$ "인 경우를 예로 들어서 P_B 와 P_{loss} 의 근사치인 \hat{P}_B 와 \hat{P}_{loss} 를 참값과 비교하였다. " $\lambda = \mu = \gamma$ " 경우는 (모든 K 에 대해서, $K \geq 1$) 다음과 같은 특징이 있다. 첫째로, 식 (16)에 $\lambda = \mu$ 를 대입하면 $P_B + P_{loss} = 1$ 이 된다. 둘째로, 식 (12)에 $\lambda = \gamma$ 를 대입하면 $L_1^B = \alpha_1$ 이므로, 이

표 1. $\lambda = \mu = \gamma$, $K=2\sim 6$ 경우에 대한 결과

K	모의실험 결과						근사치			
	α_1	α_2	α_3	α_4	α_5	α_6	P_{loss}	$\hat{\alpha}$	\hat{P}_{loss}	$\hat{\alpha}^{K+1}$
2	0.7525	0.7564	—	—	—	—	0.3627	0.7709	0.3728	0.3177
3	0.7764	0.7819	0.7847	—	—	—	0.3226	0.7976	0.3366	0.2755
4	0.7943	0.8004	0.8040	0.8070	—	—	0.2917	0.8175	0.3087	0.2451
5	0.8095	0.8150	0.8188	0.8231	0.8234	—	0.2682	0.8330	0.2862	0.2219
6	0.8211	0.8273	0.8312	0.8338	0.8364	0.8380	0.2480	0.8455	0.2676	0.2035

를 식 (14)와 묶으면 $L_i^B = \prod_{j=1}^i \alpha_j$ 를 얻는다, $i=1, 2, \dots, K$. 이러한 특징은 물론 대응하는 근사치들 간에도 성립한다. 즉, $\lambda = \mu$ 이면 $\hat{P}_B + \hat{P}_{loss} = 1$ 이고, $\lambda = \gamma$ 이면 $\hat{L}_i^B = \hat{\alpha}^i$ 이다, $i=1, 2, \dots, K$. <표 1>은 " $\lambda = \mu = \gamma$ " 경우에 대해서 K를 2에서 6까지 변화시킨 결과를 요약한 것이다(비고: $P_B = 1 - P_{loss}$, $\hat{P}_B = 1 - \hat{P}_{loss}$).

<표 1>로부터 유한 재시도 M/M/1/1에 대한 몇 가지 특징을 알아볼 수 있다. 첫째로, 주어진 K에 대해서 α_i 값은 ($i \leq K$) i가 커질수록 약간씩 증가하는 경향이 있다(비고: 이러한 경향은 $\lambda = \mu = \gamma$ 가 아닌 경우에 대한 모의실험 결과에서도 공통적으로 나타남). 그러나 그 증가율은 매우 작은 편이다. 따라서, 본 연구에서 제안한 근사법의 첫 단계인 $\alpha_1 \approx \alpha_2 \approx \dots \approx \alpha_K$ 라는 가정이 근사치의 성능에 큰 영향을 미치지 않는 것으로 사료된다.

둘째로, 주어진 i에 대해서 α_i 의 값은 K가 커질수록 ($K \geq i$) 약간씩 증가하는 경향을 보이는데, 이는 다음과 같이 설명할 수 있다. K가 커지면 그만큼 재시도 기회가 늘어나므로 큐도에 체류중인 고객수가 증가하는데, 이로 인해 전체적인 재시도율이 증가하여 P_B 를 증가시킨다. 그리고 P_B 가 증가하면 재시도율 실패할 확률 또한 증가한다.

셋째로, $\hat{\alpha}$ 은 α_i 를 전반적으로 과대평가하고 있다. $\hat{\alpha}$ 는 한마디로 $\alpha_1 \approx \alpha_2 \approx \dots \approx \alpha_K$ 라는 가정하에 α_1 의 근사치를 구한 것이다. 그런데 실제로는 i가 커질수록 α_i 가 증가하는 현상이 보이므로 $\hat{\alpha}$ 이 α_1 을 약간 과대평가하는 것이 오히려 바람직할 수도 있다. 그러나 <표 1>의 $\lambda = \mu = \gamma$ 경우에서는 $\hat{\alpha}$ 이 α_K 보다도 약간 큰데, 이는 근사법의 마지막 단계인 식 (25)에 의한 영향으로 사료된다([비고 7] 참조).

넷째로, \hat{P}_{loss} 는 P_{loss} 를 약간씩 과대평가하는 경향이 있다. 따라서, $\hat{P}_B = 1 - \hat{P}_{loss}$ 는 $P_B = 1 - P_{loss}$ 를 약간씩 과소평가한다. 그런데 \hat{P}_B 가 P_B 를 과소평가하는 정도에 비해서 $\hat{\alpha}^K$ 가 $\prod_{i=1}^K \alpha_i$ 를 과대평가하는 정도가 더욱 강하기 때문에, 결과적으로 $\hat{P}_{loss} = \hat{P}_B \hat{\alpha}^K$ 는 $P_{loss} = P_B \prod_{i=1}^K \alpha_i$ 를 과대평가하게 된 것이다. 그러나 \hat{P}_{loss} 가 P_{loss} 에 대한 상한치로 쓰일 수 있는지의 여부는 추후 연구를 통해 밝혀야 할 것이다.

다섯째로, RTA([비고 4] 참조)에 의한 P_{loss} 의 근사치인 $\hat{\alpha}^{K+1}$ 은 P_{loss} 를 과소평가하는데, 그 정도가 심한 편이다. 다만, 기존의 연구에서 더러 RTA에 의한 P_B 의 근사치인 $\hat{\alpha}$ 가 P_B 에 대한 상한치로 쓰이듯이(Greenberg and Wolff, 1987), 유한 재시도 M/M/1/1에서도 $\hat{\alpha}^{K+1}$ 이 P_{loss} 에 대한 하한치가 될 가능성이 있을 것으로 사료된다.

지금까지 " $\lambda = \mu = \gamma$ "로 고정하고, K를 2에서 6까지 변화시켰다. 지금부터는 " $K=3$ "으로 고정하고 $\rho = \lambda/\mu$ 와 $r = \gamma/\lambda$ 를 변화시킨다([비고 2] 참조)(비고: K값이 달라지더라도 ρ 와 r 의 변화에 따른 전반적인 경향은 달라지지 않음). 구체적으로, 첫 번째 실험에서는 ρ 를 0.05에서 1.2까지 0.05씩 변화시키되 $r=1$ 로 고정시키고, 두 번째 실험에서는 $\rho=1$ 로 고정시키고 r 을 0.05에서 3까지 0.05씩 변화시키면서 \hat{P}_B 와 \hat{P}_{loss} 값의 변화를 살펴보았다(비고: 편의상 모의실험의 결과를 P_B 와 P_{loss} 로 표기함).

<그림 2>는 첫 번째 실험에서의 P_B 를 나타내고 있다. RTA에 의한 근사치는 전반적으로 P_B 를 과대평가하는 경향이 있고, ρ 가 커질수록 오차가 커져서 상대오차가 7%를 넘는다. 본 연구의 근사법은 전반적으로 P_B 를 과소평가하는 경향이 있고, 역시 ρ 가 커질수록 오차가 커지는 경향이 있으나 상대오차가 2.5% 미만에서 유지된다. <그림 3>은 첫 번째 실험에서

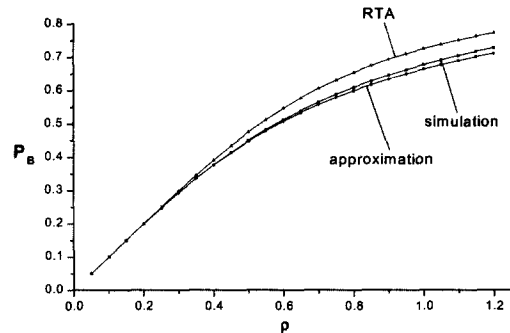


그림 2. ρ 의 변화에 따른 P_B .

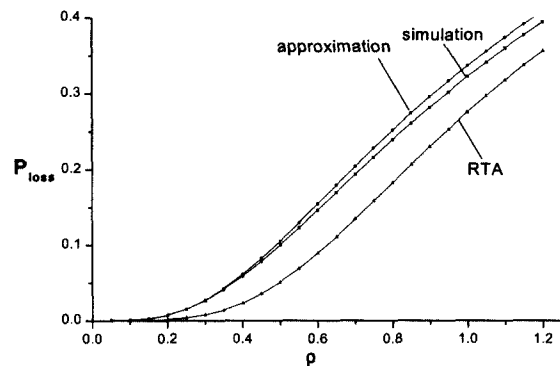


그림 3. ρ 의 변화에 따른 P_{loss} .

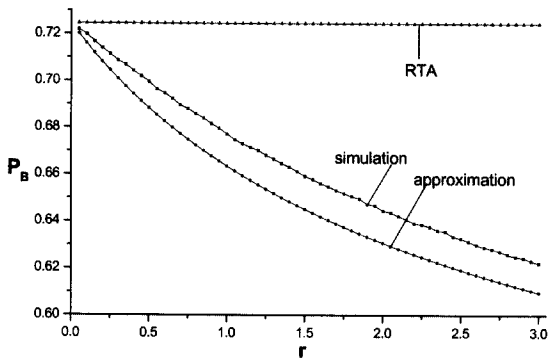


그림 4. r의 변화에 따른 P_B.

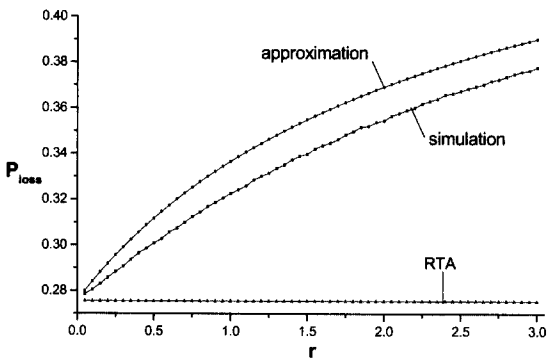


그림 5. r의 변화에 따른 P_{loss}.

의 P_{loss} 를 보여주고 있다. <그림 2>에서와는 반대로 RTA는 P_{loss} 를 과대평가하고, 본 연구의 근사법은 P_{loss} 를 과소평가한다. 근사법의 성능에서는 P_B 의 경우보다 차이가 두드러진다. RTA의 상대오차는 85%를 넘어가지만 본 근사법의 상대오차는 최대 5% 정도이다.

<그림 4>와 <그림 5>는 두 번째 실험에 대한 결과를 보여주고 있다. 식 (18)에서 알 수 있듯이 RTA 방법은 재시도율의 변화를 반영하지 못한다. 따라서 재시도율이 커질수록 오차가 증가하게 된다.

<그림 4>는 $r = \gamma/\lambda$ 를 변화시킬 때의 P_B 의 추세를 나타내고 있다. 본 연구의 근사법은 P_B 를 과소평가하는데, 상대오차가 2.5%를 넘지 않는다. 이에 비해 RTA는 P_B 를 과대평가하는데 r 이 증가함에 따라 오차가 점점 증가해서 $r=3$ 일 때는 상대오차가 16%를 넘는다. <그림 5>는 r 을 변화시킬 때의 P_{loss} 의 추세를 나타내고 있다. 첫 번째 실험과 마찬가지로 과대/과소평가 경향은 P_B 의 경우와는 반대가 된다. 본 연구의 근사법은 상대오차가 5% 미만에서 유지되는 것에 반해 RTA의 상대오차는 $r=3$ 일 때 27% 이상이다.

두 가지 실험 모두에서 본 연구가 제안한 근사법은 P_B 를 과소평가하고, P_{loss} 를 과대평가하고 있다. 본 연구의 근사치가 두 확률값에 대한 하한, 상한으로 쓰일 수 있는지의 여부는 추

후 연구를 통해 밝혀야 할 것이다. P_B 의 상대오차는 2.5% 미만이고, P_{loss} 의 상대오차는 최대 5% 정도이다. 이에 비해 RTA 방법의 상대오차는 매우 크게 나타났다.

참고문헌

김영철 (1997), 우선 순위 대기체계의 트래픽 제어에의 응용, 한국과학기술원 수학과 박사논문.
 Cohen, J. W. (1957), Basic problems of telephone traffic theory and the influence of repeated calls, *Philips Telecommunication Review*, 18, 49-100.
 Falin, G. I. (1990), A survey of retrial queues, *Queueing Systems*, 7, 127-167.
 Falin, G. I., and Templeton, J. G. C. (1997), *Retrial Queues*, Chapman & Hall.
 Greenberg, B. S., and Wolff, R. W. (1987), An upper bound on the performance of queues with returning customers, *Journal of Applied Probability*, 24, 466-475.
 Wolff, R. W. (1982), Poisson arrivals see time averages, *Operation Research*, 30, 223-231.
 Wolff, R. W. (1989), *Stochastic Modeling and the Theory of Queues*, Prentice Hall.
 Yang, T., and Templeton, J. G. C. (1987), A survey on retrial queues, *Queueing Systems*, 2, 201-233.

부록. 식 (22)의 유도

편의상 시스템의 상태를 벡터 (M, N_B) 로 정의한다. 상태 $(1, n)$ 을 보면서 도착하는 고객은 큐대로 들어가므로, 시험고객이 도착한 직후의 상태는 $(1, n+1)$ 이다. 이후 처음으로 발생하는 상태전이에 조건을 걸어서 시험고객의 (첫번) 재시도율이 실패할 확률을 구한다. 다음과 같이 네 가지의 상호배반인 사건을 정의하자.

- 사건 1: 새 고객의 도착이 먼저 발생
- 사건 2: 시험고객의 재시도율이 먼저 발생
- 사건 3: 다른 고객의 재시도율이 먼저 발생
- 사건 4: 진행중이던 서비스의 종료가 먼저 발생

첫째로, θ_1 의 확률로 사건 1이 발생하면 시험고객이 재시도에 실패할 확률은 β_{n+1} 인데, 이는 마치 시험고객이 도착시에 $n+1$ 명을 본 것과 같은 상황이기 때문이다. 둘째로, 서버가 바쁠 때 발생하는 재시도율은 실패이므로 θ_2 의 확률로 사건 2가 발생하면 시험고객이 재시도에 실패할 확률은 1이다. 반면에, θ_3 의 확률로 사건 3이 발생하는 경우에는 재시도에 실패하는 고객이 큐에 K 에 있던 고객인지 아닌지에 따라서 시험고객이 재시도에 실패할 확률은 β_{n-1} 또는 β_n 이 되는데, 이는 마치 시험고객이 도착시에 $n-1$ 명 또는 n 명을 본 것과 같은 상황이기 때문이다. 셋째로, 확률 θ_4 로 사건 4가 발생하면 상태는 $(0, n+1)$ 로 전이하는데, 이 경우에는 다시 다음 번 상태전이에 조건을 건다. 첫째로, ϕ_1 의 확률로 사건 1이 발생하면

시험고객이 재시도에 실패할 확률은 β_n 이다. 둘째로, 서버가 유휴중일 때 발생하는 재시도는 성공이므로 ϕ_γ 의 확률로 사건 2가 발생하면 시험고객이 재시도에 실패할 확률은 0이다.

반면에, $\phi_{n\gamma}$ 의 확률로 사건 3이 발생하면 시험고객이 재시도에 실패할 확률은 β_{n-1} 이 된다. 그리고 사건 4의 확률은 0이다.