응용통계연구 (2010)
23(4), 747–757

멀티클래스 손실시스템의 마코프 모델링

나성룡[1]

[1]연세대학교 정보통계학과

(2010년 4월 접수, 2010년 7월 채택)

요 약

이 논문에서는 여러 종류의 고객을 서비스하는 멀티클래스 손실시스템의 상태를 마코프 확률과정으로 표현하고 분석하는 방법을 연구한다. 특히 손실시스템에 대한 유니트 개념을 설명하고 등확률 유니트 배정을 운용하는 경우를 중점적으로 다룬다. 상태방정식을 이용하여 극한확률을 구하는 방법을 연구하고 손실확률 등의 성능척도를 산출한다. 간단한 시스템에 대한 분석을 통하여 일반적인 시스템의 특성을 알아본다.

주요용어: 멀티클래스, 손실시스템, 마코프 확률과정, 유니트, 등확률 배정, 손실확률.

1. 서론

손실시스템은 컴퓨터 시스템, 통신 네크워크, 자원 배분 및 기계 수리 문제 등에서 많이 응용되고 있는데, 얼랑의 선구적인 업적이 발표된 후로 서비스 요구가 거절될 수 있는 상황에 대한 중요한 수학적인 모형으로 많은 관심을 받아 왔다. 전통적으로 마코프 성질 가정 하에서 손실시스템에 대한 해석적인 결과를 유도하게 되는데 기본적인 시스템의 결과는 Wolff (1989) 혹은 Medhi (2003) 등을 참고할 수 있다.

Kelly (1991)는 여러 종류의 고객이 서비스를 위하여 서로 다른 수의 서버를 점유하는 네트워크를 위한 멀티클래스 손실시스템을 이론적으로 연구하면서 평형 상태에서의 극한확률을 해석적으로 유도하고 이에 대한 근사식을 표현하였다. Kelly의 모형은 음성, 저속 및 고속 데이터 서비스를 동시에 제공하는 멀티미디어 통신시스템 등의 성능 분석을 위한 이론적인 모형으로 유용하게 사용된다. Koo 등 (2000)은 Kelly의 모형에 기초해서 멀티미디어 시스템의 용량 산출 방법을 연구했고 Koo 등 (2002)은 내용을 확장하여 채널 운용 방식에 따른 시스템 용량 산출을 연구하였다. 또 다른 연구로 Lee 등 (2002) 등을 살필 수 있다.

Ross (1995)는 멀티클래스 손실시스템을 확률적 배낭(stochastic knapsack) 문제로 형식화하고 점유확률에 대한 반복 계산식을 제안하고 있다. 그런데 Ross의 계산식은 본질적으로 계승(factorial)의 반복식과 유사하기 때문에 시스템의 크기가 크면 사용이 어렵다. 이를 해결한 효율적인 반복 계산식을 Na (2006)에서 제시하고 있으며 멀티클래스 손실시스템의 정확한 성능 분석에 이용하였다.

이 논문에서는 각 고객에게 동일 유니트에 속한 서버만을 배정하는 제약 조건의 멀티클래스 손실시스템을 다룬다. 서비스를 위하여 유휴 서버를 배정할 때 서로 다른 유니트의 서버는 동시에 1명의 고객에게 배정할 수 없다. 유니트는 인접성 등의 물리적인 구성을 반영할 수도 있고 시스템 운용 등을 위해서 필

[1](220-710) 강원도 원주시 흥업면 매지리 234, 연세대학교 정보통계학과, 부교수. E-mail: nasr@yonsei.ac.kr

요한 논리적 개념일 수도 있다. 전체 시스템은 여러 개의 유니트들로 구성되는 것으로 설정되는데 유니트는 일반적으로 동수인 여러 개의 서버로 구성된다. 신규 고객에 대한 유니트 배정 방식은 시스템 상태 전이를 변하게 할 수 있으므로 시스템 분석 결과에 영향을 준다. 이 논문에서는 유니트를 등확률로 배정하는 방식을 중점적으로 다룬다.

유니트 개념을 고려한 멀티클래스 손실시스템에 대한 연구는 기존 문헌에서는 찾아보기 어렵다. 초창기 통신시스템 개발에서 여러 개의 유니트를 동시에 제어하는 기술의 구현이 어려웠던 것 등을 이러한 방향의 연구가 이루어지지 않은 이유로 볼 수 있다. Koo 등 (2002)에서 통신네트워크의 서버 배정이 여러 개의 서브시스템 단위로 이루어지는 경우가 다루어지는데, 서비스를 요구하는 고객이 어느 하나의 서브시스템에 임의로 배정되고 서브시스템들이 서로 독립인 작은 용량의 멀티클래스 손실시스템처럼 작동한다는 점에서 이 논문의 유니트 개념과는 차이가 있다.

기존의 멀티클래스 손실시스템에 대한 이론을 2절에서 간단히 정리한다. 새로운 서버 배정 방식을 3절에서 소개하고 이에 따른 마코프 확률과정 표현과 분석을 연구한다. 4절에서 가장 간단한 시스템을 실제 예로 들어서 등확률 배정 방식에 따른 상태방정식과 집합 상태방정식을 구하고 이를 기초로 시스템 특성을 연구한다. 시스템 분석의 예를 통하여 유니트들이 서로 독립이 아니며 이 논문의 마코프 과정이 시간가역성을 만족하지 않는다는 새로운 사실을 얻게 된다. 이는 기존의 시간가역성에 기초한 해석적 해법이 더 이상 유효하지 않다는 사실과 새로운 분석 방법에 대한 연구가 이루어져야 함을 의미한다.

2. 멀티클래스 손실시스템 모형

기본적인 손실시스템에서는 모든 고객 혹은 서비스 요구에 대하여 동질성을 가정하며 각각의 서비스에 요구되는 서버 혹은 채널 자원의 양이 동일하게 된다. 반면에 멀티클래스 손실시스템에서는 여러 종류의 고객의 도착을 가정하는데, 이 논문에서는 모두 r 종류의 고객이 있음을 가정한다. $i = 1, \ldots, r$에 대하여 종류 i의 고객의 시스템 도착 시점은 도착률 λ_i의 포아송 확률과정을 이루며 r개의 포아송 확률과정은 서로 독립이다. 멀티클래스 시스템의 특징은 고객 종류에 따라 요구되는 서버의 수가 다르다는 점인데 종류 i의 고객을 위한 서비스에는 k_i개의 서버가 사용된다. 종류 i의 고객에 대한 서비스 시간은 평균이 μ_i^{-1}인 지수분포의 확률변수이고 모든 서비스 시간은 서로 독립임을 가정한다. 전체 시스템에는 c개의 서버가 있음을 가정하고 시스템에 새롭게 도착하는 고객은 필요한 수의 서버를 배정받지 못하면 즉시 탈락하고 시스템을 떠난다.

시간 t에서 시스템에서 서비스 중인 종류 i의 고객 수를 $N_i(t)$라고 $i = 1, \ldots, r$에 대하여 정의한다. 서비스를 위한 서버 배정에서 서버의 위치가 모든 종류의 고객에 대하여 문제가 되지 않는다면 서비스 탈락 여부는 유휴 서버의 수에 의하여 결정되게 되고 마코프 과정을 유도하기 위해서는 서비스 중인 고객의 수만 관찰하면 충분하다. 즉 r차원 확률과정 $(N_1(t), \ldots, N_r(t))$은 상태공간

$$S_1(c) = \left\{ (n_1, \ldots, n_r) \in \mathbf{Z}_+{}^r \,\middle|\, \sum_{i=1}^{r} k_i n_i \le c \right\} \tag{2.1}$$

에서 연속시간 마코프 확률과정이 되며, 여기에서 $\mathbf{Z}_+$는 0보다 크거나 같은 정수들의 집합이다.

시스템 상태의 변화는 새로운 고객이 서비스를 받기 위하여 시스템에 도착하거나 시스템에서 서비스를 받는 고객이 서비스를 마치고 시스템을 떠날 때 생길 수 있다. 종류 i의 고객이 시스템에 새롭게 도착하는 시점 t에서 시스템 유휴 서버의 수 $c - \sum_{i=1}^{r} k_i N_i(t)$의 값이 k_i 이상이면 서비스를 받을 수 있고 시스템 상태는 $(N_1(t), \ldots, N_i(t) + 1, \ldots, N_r(t))$로 i번째 값이 1 증가하도록 변한다. 반면에 서버의 수가 부족하면 도착하는 고객은 시스템에서 탈락하고 시스템 상태에는 변화가 없다. 시스템에서 서비스를 받

는 종류 i의 고객이 시점 t에서 시스템을 떠나면 시스템 상태는 $(N_1(t),\ldots,N_i(t)-1,\ldots,N_r(t))$로 i번째 값이 1 감소하도록 변한다. 상태간의 전이율을 살펴보면 $c-\sum_{i=1}^{r}k_iN_i(t)\geq k_i$ 조건을 만족하는 상태 $(N_1(t),\ldots,N_r(t))$에서 $(N_1(t),\ldots,N_i(t)+1,\ldots,N_r(t))$로의 전이율은 λ_i이고, $N_i(t)\geq 1$을 만족하는 상태 $(N_1(t),\ldots,N_r(t))$에서 $(N_1(t),\ldots,N_i(t)-1,\ldots,N_r(t))$로의 전이율은 $\mu_iN_i(t)$이다.

상태공간 $S_1(c)$ 안에서의 시스템 변화를 살펴보면 종류 i의 고객 수는 한 종류의 고객을 가정하는 $M/M/c_i/c_i$ 손실시스템의 고객 수와 유사한 확률적 특성으로 변하고 독립적인 고객 도착과 서비스 종료때문에 $N_i(t),\ i=1,\ldots,r$이 확률적으로 독립인 것처럼 보여진다. 여기에서 $c_i=[c/k_i]$이다. 그러나 사용중인 서버의 수가 전체적으로 c를 초과할 수 없으므로 독립은 성립하지 않는다. 결과적으로 $(N_1(t),\ldots,N_r(t))$는 r개의 서로 독립적인 $M/M/c_i/c_i$ 시스템의 고객 수로 정의되는 마코프 과정의 (2.1)로의 절단 형태로 이해할 수 있다.

$M/M/c/c$ 시스템이 시간가역성을 만족하는 마코프 과정임은 잘 알려진 사실이다. 따라서 독립적인 r개의 $M/M/c_i/c_i,\ i=1,\ldots,r$ 시스템의 고객 수로 정의된 r차원 확률과정 역시 시간가역적인 마코프 과정이고 이의 절단 과정 역시 시간가역적인 마코프 과정이 된다. 그리고 극한확률은 원래의 마코프 과정의 극한확률의 절단 분포와 같은데 Kelly (1979) 등을 참고할 수 있다. 이러한 사실을 바탕으로 멀티클래스 손실시스템 $(N_1(t),\ldots,N_r(t))$의 극한분포를 정리할 수 있다. 먼저 도착률이 λ, 서비스율이 μ인 $M/M/c/c$ 시스템의 극한확률을

$$p(n;c,\rho)=\frac{\rho^n/n!}{\sum\limits_{j=0}^{c}\rho^j/j!},\quad n=0,1,\ldots,c \tag{2.2}$$

식으로 정의하는데, $\rho=\lambda/\mu$는 시스템 부하량을 의미한다. $M/M/c/c$ 손실시스템이 평형 상태일 때 시스템에 n명의 고객이 서비스를 받을 극한확률은 식 (2.2)로 표현된다. 이때 멀티클래스 손실시스템 $(N_1(t),\ldots,N_r(t))$의 극한분포는

$$P[(N_1(t),\ldots,N_r(t))=(n_1,\ldots,n_r)]=\pi(S_1(c))^{-1}\prod_{i=1}^{r}p(n_i;c_i,\rho_i) \tag{2.3}$$

식으로 주어지며

$$\pi(S_1(c))=\sum_{(n_1,\ldots,n_r)\in S_1(c)}\prod_{i=1}^{r}p(n_i;c_i,\rho_i)$$

은 정규 상수이다. 극한확률 (2.3)은 Kelly (1991)와 Ross (1995)에서도 유도하고 있다.

위 식 (2.3)의 극한확률은 해석적으로 단순한 형태를 가지지만 시스템 용량이 크면 직접적인 계산이 쉽지 않은데 이는 식 (2.2)의 계산상 어려움에 기인한다. 따라서 구체적인 값을 계산하기 위한 얼랑-B식과 같은 반복 계산식의 필요한데, Na (2006)는 식 (2.2)에 대하여 다음의 효율적인 반복 계산식을 제공한다.

- 초기값으로 $p(0;0,\rho)=1$을 정의하고 $c=1,2,\ldots$에 대하여

$$p(c;c,\rho)=\frac{\rho p(c-1;c-1,\rho)}{c+\rho p(c-1;c-1,\rho)}. \tag{2.4}$$

- $n=c-1,c-2,\ldots,0$에 대하여

$$p(n;c,\rho)=p(n;c-1,\rho)[1-p(c;c,\rho)].$$

실제로 식 (2.4)는 잘 알려진 얼랑-B식에 해당하며 이를 먼저 계산하고 식 (2.2)의 다른 확률을 순서대로 구할 수 있다. 이 반복식을 이용하면 시스템 용량의 크기에 제한없이 극한확률에 대하여 정확하고 빠른 수치적 계산이 가능해진다.

손실시스템에 대한 성능척도는 시스템의 극한분포에 기초하여 정의된다. 가장 기본적인 성능척도는 서비스를 요구하는 고객이 서비스 받기를 개시하지 못하고 탈락하는 확률로 정의되는 손실확률인데 각각의 서비스 종류에 대하여 다르게 결정된다. 종류 i 고객의 손실확률 L_i는 $i = 1, \ldots, r$에 대하여

$$L_i = \sum_{(n_1,\ldots,n_r)\in B_i} \pi(S_1(c))^{-1} \prod_{i=1}^{r} p(n_i; c_i, \rho_i) \tag{2.5}$$

와 같은데, 여기에서 종류 i에 대한 손실집합 B_i는

$$B_i = \left\{ (n_1, \ldots, n_r) \in S_1(c) \,\middle|\, \sum_{i=1}^{r} k_i n_i > c - k_i \right\}$$

이다. PASTA 원리에 의하여 시스템에 도착하는 고객의 탈락 확률은 평형상태에서 시스템에 서버가 부족할 확률과 같으므로 식 (2.5)가 유도되며, L_i를 구하는 과정은 Na (2006)에 잘 설명되어 있다.

손실시스템의 대표적인 성능척도인 손실확률은 서비스 품질을 표현하는 가장 기본적인 값이다. 통신 네트워크, 컴퓨터 시스템 등의 설계와 분석에서 서비스 종류에 따른 손실확률의 계산은 필수적인 과정이며 이에 기초해서 시스템 자원 배분, 운용 등의 원칙이 결정되게 된다. 그리고 시스템 극한분포에 대한 수치적인 값은 손실확률 계산 외에도 서버 운용률 산출에 필요한 평균 사용 서버 수 등의 계산에 사용된다. 멀티클래스 손실시스템에 대한 다양한 성능분석의 예들은 Na (2006)를 참고할 수 있다.

3. 등확률 유니트 배정의 멀티클래스 손실시스템

특정 종류의 고객을 서비스하기 위해서는 일정한 집합의 서버가 동시에 사용되어야 하는 상황을 가정하고, 이러한 서버의 집합 단위를 유니트(unit)라 정의하자. 유니트는 인접한 서버들이나 논리적으로 미리 정해진 집합의 서버들로 구성된다. 서비스 종류별로 별도의 유니트를 구성하는 것을 고려할 수도 있지만 동일한 크기의 유니트들이 서비스의 종류와 관계없이 미리 정해져 있는 시스템을 고려하는 것이 일반적이다. 이 논문에서는 유니트들이 일정한 크기로 미리 정해져 있는 시스템을 고려하기로 한다.

전체 시스템이 M의 동일한 크기를 가지는 U개의 유니트로 구성됨을 가정하자. 이때 전체 서버의 수는 $c = UM$과 같다. 도착하는 고객에게 동일 유니트의 서버를 배정해야 하는 조건 때문에 시스템의 상태를 효과적으로 표현하기 위해서 모든 유니트에 대하여 종류별 고객의 수를 추적해야 한다. 앞 2절의 $N_1(t), \ldots, N_r(t)$만으로는 신규 고객의 탈락 여부를 결정할 수 없다. $u = 1, \ldots, U$에 대하여 $\mathbf{N}^u(t) = (N_1^u(t), \ldots, N_r^u(t))$를 u번째 유니트의 상태로 정의하는데 $N_i^u(t)$는 시간 t에 유니트 u에서 서비스 받는 종류 i의 고객 수를 의미한다. 이때 $N_i^u(t)$가 가질 수 있는 값은 k_i와 M의 값에 의하여 결정된다. 시간 t에서의 전체 시스템의 상태는 $(\mathbf{N}^1(t), \ldots, \mathbf{N}^U(t))$로 정의된다.

새롭게 도착하는 고객에 대한 유니트 배정 방식에 따라 시스템의 상태 전이가 결정된다. 신규 고객이 배정되는 유니트에 대해서는 고객의 수가 증가하는 방향으로 전이가 이루어지므로 배정 방식은 특히 고객의 도착과 관련한 상태 전이에 영향을 준다. 이 연구에서는 유니트에 대한 등확률 배정 방식을 다룬다. 즉 시스템에 도착하는 고객은 서비스를 위하여 필요한 수의 유휴 서버를 가지고 있는 유니트 중에서 랜덤하게 1개를 선택해서 서비스를 개시한다. 도착 시점에 필요한 수의 유휴 서버를 가진 유니트가 없다면 이 고객은 서비스를 받지 못하고 시스템에서 탈락하게 된다. 한편 이 방식은 필요한 수의 서버를 찾

을 때까지 전체 유니트를 랜덤하게 검색하는 방식과 실제적으로 동일한데 고객은 필요한 수의 유휴 서버가 있는 유니트를 검색하면서 서버가 부족한 유니트를 즉시 통과해서 다른 유니트를 조사하기 때문이다.

앞 2절에서와 같이 종류별로 고객의 도착 과정이 독립인 포아송 과정을 이루고 서비스 시간은 지수분포를 따름을 가정하자. 또한 $i = 1, \ldots, r$에 대하여 종류 i의 고객은 서비스를 위하여 k_i개의 서버를 점유함을 가정하는데 동일 유니트 안의 서버라는 제약이 추가된다. 앞에서 설명한 등확률 유니트 배정 방식을 가지는 멀티클래스 손실시스템은 시스템 상태를 $(\mathbf{N}^1(t), \ldots, \mathbf{N}^U(t))$의 확률과정으로 정의하면 마코프 모형이 성립한다. 시스템의 가능한 상태는 일반적인 r, U, M에 대하여

$$\left(\mathbf{n}^1, \mathbf{n}^2, \ldots, \mathbf{n}^U\right) = \left(n_1^1, n_2^1, \ldots, n_r^1; n_1^2, \ldots, n_r^2; \ldots; n_1^U, \ldots, n_r^U\right)$$

의 형태로 표현할 수 있는데, $n_i^u \in \mathbf{Z}_+$과 $\sum_{i=1}^r k_i n_i^u \leq M$이 모든 $u = 1, \ldots, U$와 $i = 1, \ldots, r$에 대하여 성립해야 한다. 따라서 상태공간은

$$S_2(U, M) = \left\{ (\mathbf{n}^1, \ldots, \mathbf{n}^U) \in (\mathbf{Z}_+{}^r)^U \,\middle|\, \sum_{i=1}^r k_i n_i^u \leq M,\ u = 1, \ldots, U \right\} \tag{3.1}$$

의 식으로 정리할 수 있다. 한편 식 (2.1)의 정의에서 $S_2(U, M) = (S_1(M))^U$의 관계가 성립함을 알 수 있고 이는 시스템의 복잡도가 기하급수적으로 증대함을 보인다.

시스템의 상태 전이는 새로운 고객이 도착하거나 시스템 내의 고객이 서비스를 마치고 시스템을 떠날 때 발생할 수 있다. 상태 전이와 전이율을 설명하기 위하여 필요한 몇 개의 기호를 정의하자. 먼저 $\underline{\mathbf{n}} = (\mathbf{n}^1, \ldots, \mathbf{n}^U) \in S_2(U, M)$에 대하여 $A_i(\underline{\mathbf{n}}) = \{u \in \{1, 2, \ldots, U\} | M - \sum_{i=1}^r k_i n_i^u \geq k_i\}$를 정의하는데, 이 집합은 현재 상태 $\underline{\mathbf{n}}$에서 종류 i의 고객 1명을 추가로 서비스할 수 있는 유니트의 집합을 의미한다. 또 집합 $E_i(\underline{\mathbf{n}}) = \{u \in \{1, 2, \ldots, U\} | n_i^u > 0\}$은 상태 $\underline{\mathbf{n}}$에서 종류 i의 고객을 적어도 1명 서비스하고 있는 유니트들의 모임을 나타낸다. $\underline{\mathbf{n}} \in S_2(U, M)$일 때 $u \in A_i(\underline{\mathbf{n}})$에 대하여 $I_i^u(\underline{\mathbf{n}})$를 $\underline{\mathbf{n}}$에서 u번째 유니트의 종류 i의 고객 수만 1 증가한 상태로 정의하고, $u \in E_i(\underline{\mathbf{n}})$에 대하여 $D_i^u(\underline{\mathbf{n}})$를 $\underline{\mathbf{n}}$에서 u번째 유니트의 종류 i의 고객 수만 1 감소한 상태로 정의한다. 집합 A의 크기를 원소의 수로 정의하고 $|A|$로 표시한다.

등확률 유니트 배정 방식을 고려하면 다음과 같은 상태 전이와 전이율이 가능하다.

(T1) $|A_i(\underline{\mathbf{n}})| > 0$이고 $u \in A_i(\underline{\mathbf{n}})$인 i와 u에 대하여 $\underline{\mathbf{n}}$에서 $I_i^u(\underline{\mathbf{n}})$로의 전이가 $\lambda_i / |A_i(\underline{\mathbf{n}})|$ 전이율로 발생한다.

(T2) $|E_i(\underline{\mathbf{n}})| > 0$이고 $u \in E_i(\underline{\mathbf{n}})$인 i와 u에 대하여 $\underline{\mathbf{n}}$에서 $D_i^u(\underline{\mathbf{n}})$로의 전이가 $n_i^u \mu_i$의 전이율로 발생한다.

위 (T1)의 전이는 종류 i의 고객이 시스템에 도착하면서 서비스가 가능한 유니트에서 1개를 임의로 선택해서 서비스를 받기 시작하는 것을 의미한다. 반면 (T2)의 전이는 시스템에서 서비스를 받던 종류 i의 고객 1명이 서비스를 마치고 시스템을 떠나는 사건을 나타낸다.

시스템 성능분석을 위해서는 식 (3.1)의 상태공간과 (T1), (T2)에서 설명한 가능한 전이를 기초로 평형 방정식을 수립하고 이의 해를 구하는 작업이 필요하다. 유한 상태공간이므로 마코프 과정의 극한분포는 존재하게 되며 상태 $\underline{\mathbf{n}} \in S_2(U, M)$의 극한확률을 $p(\underline{\mathbf{n}})$으로 표현하기로 하자. 이때 모든 상태 $\underline{\mathbf{n}} \in S_2(U, M)$에 대하여

$$p(\underline{\mathbf{n}}) \left[\sum_{i=1}^r \lambda_i I(|A_i(\underline{\mathbf{n}})| > 0) + \sum_{i=1}^r \sum_{u=1}^U \mu_i n_i^u \right] \tag{3.2}$$

$$= \sum_{i=1}^{r} \sum_{u \in A_i(\underline{\mathbf{n}})} p(I_i^u(\underline{\mathbf{n}}))\mu_i(n_i^u + 1) \tag{3.3}$$

$$+ \sum_{i=1}^{r} \sum_{u \in E_i(\underline{\mathbf{n}})} p(D_i^u(\underline{\mathbf{n}}))\lambda_i(|A_i(D_i^u(\underline{\mathbf{n}}))|)^{-1} \tag{3.4}$$

의 평형방정식을 얻는다. 여기에서 $I(|A_i(\underline{\mathbf{n}})| > 0)$는 $|A_i(\underline{\mathbf{n}})| > 0$이 성립하면 1, 아니면 0의 값을 가지는 지시함수이다. 상태 $\underline{\mathbf{n}}$을 벗어나는 전이율을 나타내는 식 (3.2)는 새로운 고객의 도착 비율과 서비스를 마친 고객이 떠나는 비율의 합으로 표현된다. 식 (3.3)에는 고객이 떠나면서 상태 $\underline{\mathbf{n}}$으로 전이하는 비율이, 식 (3.4)에는 고객이 새롭게 도착하면서 상태 $\underline{\mathbf{n}}$으로 전이하는 비율이 정리되어 있다.

평형방정식을 이용해서 극한확률을 구하면 이를 이용해서 손실확률을 얻게 된다. 등확률 유니트 배정을 하는 경우에 종류 i 고객의 손실확률은

$$L_i = \sum_{\underline{\mathbf{n}} \in \bar{B}_i} p(\underline{\mathbf{n}}) \tag{3.5}$$

식과 같으며, 여기에서 $p(\underline{\mathbf{n}})$는 식 (3.2)~(3.4)을 만족하는 극한확률이고 손실집합 $\bar{B}_i$는

$$\bar{B}_i = \{\underline{\mathbf{n}} \in S_2(U, M) |\ |A_i(\underline{\mathbf{n}})| = 0\}$$

이다. $\bar{B}_i$의 정의를 보면 k_i개 이상의 유휴 서버를 가지는 유니트가 없는 상태들로 이루어짐을 알 수 있는데 PASTA 성질에 의하여 식 (3.5)의 결과가 성립한다.

일반적인 r, U, M의 시스템의 경우에 식 (3.2)~(3.4)의 평형방정식을 직접 풀어서 해석적인 형태의 극한확률을 구하는 작업은 거의 불가능하다고 하겠다. 이 식들을 직접 풀기 위해서는 수치적인 방법을 사용해야 하는데 마코프 과정의 상태방정식은 자체로 자기 정의 형태이므로 반복 계산에 의한 근사법을 쉽게 적용할 수 있다. 이에 대하여 Rubinstein (1981) 등을 참고할 수 있다. 시스템이 특별한 성질을 만족하면 해석적인 형태의 극한확률도 가능한데 시간가역성, 유니트 독립성 등이 있겠다. 하지만 이러한 성질이 등확률 유니트 배정의 시스템에서는 일반적으로 성립하지 않음을 다음 절의 간단한 시스템을 직접 다루는 예에서 살펴볼 수 있다.

4. 두 종류 고객의 시스템 분석

두 종류의 서비스를 제공하는 시스템을 가정하자. 이러한 예는 음성과 데이터 통화를 제공하는 통신시스템을 예로 들 수 있다. 간단하게 $k_1 = 1$, $k_2 = 2$임을 가정하고 시스템의 유니트에 대하여 $U = 2$, $M = 2$을 가정하자. 각각 2개의 서버로 구성된 2개의 유니트를 가지는 시스템은 총 4개의 서버를 가지며 서버가 1개 필요한 종류 1의 고객은 시스템에 1개 이상의 유휴 서버가 있으면 서비스를 받을 수 있는데 종류 2의 고객은 필요한 2개의 서버가 동일 유니트 안에서 유휴 상태일 때만 서비스가 가능함을 알 수 있다.

서비스 중인 고객의 수를 유니트와 고객 종류별로 나타내는 시스템 상태는 $(n_1^1, n_2^1; n_1^2, n_2^2)$의 형태로 표현되며 각 $u = 1, 2$에 대하여 $n_1^u = 0, 1, 2$와 $n_2^u = 0, 1$의 값이 가능하고 $n_1^u + 2n_2^u \le 2$의 제약이 있다. 따라서 (n_1^u, n_2^u)은 $(0, 0)$, $(1, 0)$, $(2, 0)$, $(0, 1)$의 4가지 값이 가능한데, 각각 0, 1, 2, 3으로 간단하게 표현하기로 하자. 전체 시스템의 상태는 이들의 조합으로 16가지가 가능하므로 상태공간은

$$\begin{aligned} S_2(2, 2) &= \{ij | i = 0, 1, 2, 3,\ j = 0, 1, 2, 3\} \\ &= \{00, 01, 02, 03, 10, \ldots, 32, 33\} \end{aligned}$$

의 형태로 정리할 수 있다. 상태 $ij \in S_2(2,2)$에 대한 극한확률을 $p(ij)$라 하면 이 시스템에 대한 식 (3.2)~(3.4)의 평형방정식은 다음과 같다.

$$\begin{aligned}
&p(00)(\lambda_1+\lambda_2) = (p(10)+p(01))\mu_1 + (p(03)+p(30))\mu_2\\
&p(10)(\lambda_1+\lambda_2+\mu_1) = p(00)\frac{\lambda_1}{2} + p(11)\mu_1 + p(20)\cdot 2\mu_1 + p(13)\mu_2\\
&p(01)(\lambda_1+\lambda_2+\mu_1) = p(00)\frac{\lambda_1}{2} + p(11)\mu_1 + p(02)\cdot 2\mu_1 + p(31)\mu_2\\
&p(20)(\lambda_1+\lambda_2+2\mu_1) = p(10)\frac{\lambda_1}{2} + p(21)\mu_1 + p(23)\mu_2\\
&p(02)(\lambda_1+\lambda_2+2\mu_1) = p(01)\frac{\lambda_1}{2} + p(12)\mu_1 + p(32)\mu_2\\
&p(11)(\lambda_1+2\mu_1) = (p(10)+p(01))\frac{\lambda_1}{2} + (p(21)+p(12))\cdot 2\mu_1\\
&p(21)(\lambda_1+3\mu_1) = p(11)\frac{\lambda_1}{2} + p(20)\lambda_1 + p(22)\cdot 2\mu_1\\
&p(12)(\lambda_1+3\mu_1) = p(11)\frac{\lambda_1}{2} + p(02)\lambda_1 + p(22)\cdot 2\mu_1\\
&p(22)\cdot 4\mu_1 = (p(21)+p(12))\lambda_1\\
&p(23)(2\mu_1+\mu_2) = p(13)\lambda_1 + p(20)\lambda_2\\
&p(32)(2\mu_1+\mu_2) = p(31)\lambda_1 + p(02)\lambda_2\\
&p(13)(\lambda_1+\mu_1+\mu_2) = p(03)\lambda_1 + p(10)\lambda_2 + p(23)\cdot 2\mu_1\\
&p(31)(\lambda_1+\mu_1+\mu_2) = p(30)\lambda_1 + p(01)\lambda_2 + p(32)\cdot 2\mu_1\\
&p(03)(\lambda_1+\lambda_2+\mu_2) = p(00)\frac{\lambda_2}{2} + p(13)\mu_1 + p(33)\mu_2\\
&p(30)(\lambda_1+\lambda_2+\mu_2) = p(00)\frac{\lambda_2}{2} + p(31)\mu_1 + p(33)\mu_2\\
&p(33)\cdot 2\mu_2 = (p(03)+p(30))\lambda_2
\end{aligned}$$

상태의 수가 많은 관계로 평형방정식을 직접 푸는 작업은 실질적으로 불가능하며 해석적인 형태의 해는 얻기 어려운 것이 현실이다. 단 시스템이 시간가역성을 가지는 경우와 유니트별 고객 수가 서로 독립인 경우는 단순한 형태의 해를 유도할 수 있으므로 이에 대한 성립 여부를 체크해 보자.

역할이 대칭적인 관계로 평형 상태에서 각 유니트의 고객 수에 대한 주변분포는 동일하다. 이때 독립성을 가정하면 극한확률에 대하여

$$p(ij) = \bar{p}(i)\bar{p}(j), \quad ij \in S_2(2,2)$$

식이 성립하는데, 여기에서 $\bar{p}(\cdot)$는 동일한 주변분포이다. 이 식을 상태 00, 11, 22, 33에 대한 평형방정식에 대입하면

$$\begin{aligned}
&\bar{p}(0)(\lambda_1+\lambda_2) = \bar{p}(1)\cdot 2\mu_1 + \bar{p}(3)\cdot 2\mu_2\\
&\bar{p}(1)(\lambda_1+2\mu_1) = \bar{p}(0)\lambda_1 + \bar{p}(2)\cdot 4\mu_1\\
&\bar{p}(2) = \bar{p}(1)\frac{\lambda_1}{2\mu_1}\\
&\bar{p}(3) = \bar{p}(0)\frac{\lambda_2}{\mu_2}
\end{aligned}$$

식들이 유도되고 이를 정리하면

$$\bar{p}(1) = \bar{p}(0)\frac{\lambda_1 - \lambda_2}{2\mu_1} = \bar{p}(0)\frac{\lambda_1}{2\mu_1 - \lambda_1}$$

식을 얻게 되는데 일반적으로 성립할 수 없는 결과이다. 따라서 유니트별 고객 수는 서로 독립이 아님을 알 수 있다.

시간가역성이 성립하면 상세 평형방정식을 이용해서 극한확률을 쉽게 유도할 수 있게 된다. 시간가역성을 가지는 시스템은 콜모고로프 기준을 만족하는데 이는 임의의 상태에서 출발한 시스템이 다른 상태를 거쳐서 원래 위치로 돌아올 때 순환 방향에 관계없이 전이율의 곱이 동일하다는 성질이다. 그런데 $10 \to 11 \to 21 \to 20 \to 10$ 순환에 대한 전이율은

$$\frac{\lambda_1}{2} \cdot \frac{\lambda_1}{2} \cdot \mu_1 \cdot 2\mu_1 = \frac{\lambda_1^2\mu_1^2}{2}$$

이지만 반대방향 순환 $10 \to 20 \to 21 \to 11 \to 10$의 경우는

$$\frac{\lambda_1}{2} \cdot \lambda_1 \cdot 2\mu_1 \cdot \mu_1 = \lambda_1^2\mu_1^2$$

의 전이율을 가진다. 따라서 콜모고로프 기준을 만족하지 않음을 볼 수 있고 이 시스템은 시간가역성을 만족하지 않는다.

등확률 배정의 멀티클래스 손실시스템은 일반적으로 시간가역성을 만족하지 않고 유니트별 고객 수의 분포는 서로 독립이 아님을 확인할 수 있다. 결과적으로 단순한 형태의 극한분포를 유도하는 것은 어렵다는 사실을 알 수 있다. 이때 식의 수를 줄이는 방법으로 집합 평형방정식을 생각할 수 있겠는데 몇 개의 상태를 묶어서 정의된 상태집합에 대한 집합 평형방정식은 식의 개수가 적으므로 해석적인 혹은 수치적인 해의 산출이 용이할 수 있겠다. 또 상태집합의 극한확률을 통하여 손실확률 등의 성능척도를 쉽게 구할 수 있다.

이제 $U = 2$, $M = 2$의 시스템에 대하여 손실확률 산출이 가능하도록 하는 집합 평형방정식을 세우는 과정을 살펴본다. 상태집합은 동일한 위상을 가지는 상태들의 모임으로 정의할 수 있는데 상태집합의 각 상태는 시스템에 도착하는 고객에 대한 서비스 개시 혹은 탈락을 동일하게 결정한다. $U = 2$, $M = 2$인 시스템의 16개 상태는 $\{00\}$, $\{11\}$, $\{22\}$, $\{33\}$, $\{10, 01\}$, $\{20, 02\}$, $\{21, 12\}$, $\{23, 32\}$, $\{13, 31\}$, $\{03, 30\}$의 총 10개의 상태집합으로 분류할 수 있다. $i \neq j$일 때 $\overline{ij} = \{ij, ji\}$, $p(\overline{ij}) = p(ij) + p(ji)$의 표현을 가정하자. 앞의 평형방정식에서 다음의 집합 평형방정식을 얻을 수 있다.

$$p(00)(\lambda_1 + \lambda_2) = p(\overline{10})\mu_1 + p(\overline{03})\mu_2$$

$$p(\overline{10})(\lambda_1 + \lambda_2 + \mu_1) = p(00)\lambda_1 + p(11) \cdot 2\mu_1 + p(\overline{20}) \cdot 2\mu_1 + p(\overline{13})\mu_2$$

$$p(\overline{20})(\lambda_1 + \lambda_2 + 2\mu_1) = p(\overline{10})\frac{\lambda_1}{2} + p(\overline{21})\mu_1 + p(\overline{23})\mu_2$$

$$p(11)(\lambda_1 + 2\mu_1) = p(\overline{10})\frac{\lambda_1}{2} + p(\overline{21}) \cdot 2\mu_1$$

$$p(\overline{21})(\lambda_1 + 3\mu_1) = p(11)\lambda_1 + p(\overline{20})\lambda_1 + p(22) \cdot 4\mu_1$$

$$p(22) \cdot 4\mu_1 = p(\overline{21})\lambda_1$$

$$p(\overline{23})(2\mu_1 + \mu_2) = p(\overline{13})\lambda_1 + p(\overline{20})\lambda_2$$

$$p(\overline{13})(\lambda_1 + \mu_1 + \mu_2) = p(\overline{03})\lambda_1 + p(\overline{10})\lambda_2 + p(\overline{23}) \cdot 2\mu_1$$

$$p(\overline{03})(\lambda_1+\lambda_2+\mu_2)=p(00)\lambda_2+p(\overline{13})\mu_1+p(33)\cdot 2\mu_2$$
$$p(33)\cdot 2\mu_2=p(\overline{03})\lambda_2$$

그런데 $U = 2$, $M = 2$의 가장 단순한 시스템의 집합 평형방정식도 결국 미지수가 10개인 일차연립방정식으로 귀착되며 이 역시 간단한 형태의 해를 구하기는 쉽지 않다. 결과적으로 집합 평형방정식의 해 역시 해석적인 형태로는 얻기 어려우므로 수치적인 방법을 통해서 구해야 한다는 결론에 도달하게 된다. 수치적 해는 반복식을 통해서 계산할 수 있는데 평형방정식 혹은 집합 평형방정식의 우변에 과거값을 대입해서 좌변의 신규값을 구하는 절차를 반복하면 된다. 마코프 확률과정의 수렴 성질에 의하여 반복 계산의 결과는 빠른 속도로 극한분포로 수렴한다.

극한분포 혹은 집합 평형방정식의 해를 이용해서 성능척도를 산출할 수 있다. 종류 1의 고객은 시스템에 유휴 서버가 없을 때 탈락하므로 손실확률은

$$L_1=p(22)+p(33)+p(\overline{23})$$

식으로 주어진다. 종류 2의 고객은 유니트의 서버를 동시에 점유해야하므로 모든 유니트에 사용중인 서버가 1개 이상이면 탈락하게 된다. 따라서 종류 2의 고객의 손실확률은

$$L_2=L_1+p(11)+p(\overline{12})+p(\overline{13})$$

식과 같다.

5. 결론

이 논문에서는 등확률 유니트 배정 방식의 멀티클래스 손실시스템에 대하여 손실확률 등의 성능척도를 산출하기 위한 마코프 모델링을 연구했다. 마코프 확률과정을 정의하고 상태공간을 구하고 상태방정식을 수립하였다. 또한 종류별 손실확률에 대한 식의 형태를 살펴보았다. 기존의 멀티클래스 손실시스템에 비하여 마코프 모델링에 필요한 확률과정의 복잡성이 크게 증가함을 보았고 해석적인 극한확률의 계산이 거의 불가능하다는 사실을 알게 되었다. 기존 시스템에 대한 간단한 형태의 해석적 결과를 가능케 하는 시간가역성 등의 성질을 만족하지 않기 때문이다.

그럼에도 불구하고 현실의 다양하고 복잡한 통신 네트워크나 컴퓨터 시스템 등을 분석하기 위해서 이 논문에서 연구한 유니트 구조의 시스템이 중요한 역할을 하리라 기대한다. 또 이 연구의 결과는 수치 해를 구하기 위한 알고리즘 구성에 적절히 사용될 수 있다.

실제 시스템의 설계에서는 다른 형태의 배정 방식을 고려할 수 있고 이에 대한 분석은 또 다른 도전이 될 수 있다. 기존의 유니트 제약이 없는 시스템, 이 논문의 등확률 유니트 배정 방식, 또 다른 새로운 배정 방식의 상대적인 효율성을 비교하는 연구는 손실시스템 연구 분야에서 큰 의미가 있을 것이다. 이러한 목적을 위해서 뿐만 아니라 손실시스템의 실제 응용을 위해서 효율적인 수치적 방법의 연구가 병행되어야 하겠다.

참고문헌

Kelly, F. P. (1979). *Reversibility and Stochastic Networks*, Wiley.

Kelly, F. P. (1991). Loss networks, *Annals of Applied Probability*, **1**, 319–378.

Koo, I., Yang, J., Ahmad, A. and Kim, K. (2002). Erlang capacity analysis of hybrid FDMA/CDMA systems supporting multi-class services according to channel assignment methods, *International Journal of Communication Systems*, **15**, 867–880.

Koo, I., Yang, J. and Kim, K. (2000). Analysis of Erlang capacity for DS-CDMA systems supporting multi-class services with the limited number of channel elements, In *Proceedings of IEEE Conference on Wireless Communications and Networking 2000, USA*, 355–359.

Lee, J., Na, S., Choi, Y. and Lim, S. (2002). Performance analysis on channel distribution algorithm of selector in CDMA2000 1x system, In *Proceedings of Korea Information and Communications Society, summer 2002*.

Medhi, J. (2003). *Stochastic Models in Queueing Theory*, (2nd ed.), Academic Press.

Na, S. (2006). Exact performance evaluation of multi-class loss systems with applications to telecommunications networks, *Telecommunications Review*, **16**, 1084–1091.

Ross, K. W. (1995). *Multiclass Loss Models for Broadband Telecommunication Networks*, Springer.

Rubinstein, Y. R. (1981). *Simulation and the Monte Carlo Method*, Wiley.

Wolff, R. W. (1989). *Stochastic Modelling and the Theory of Queues*, Prentice Hall.

Markov Modeling of Multiclass Loss Systems

Seongryong Na[1]

[1]Department of Information and Statistics, Yonsei University

(Received April 2010; accepted July 2010)

Abstract

This paper studies the Markov modeling of multiclass loss systems supporting several kinds of customers. The concept of unit for loss systems is introduced and the method of equal probability allocation among units is especially considered. Equilibrium equations and limiting distribution of the loss systems are studied and loss probabilities are computed. We analyze an example of a simple system to gain an insight about general systems.

Keywords: Multiclass, Loss system, Markov process, Unit, Equal probability allocation, Loss probability.

[1]Associate Professor, Department of Information and Statistics, Yonsei University, 234 Maeji, Heungup, Wonju 220-710, Korea. E-mail: nasr@yonsei.ac.kr