

## Estimation of Overflow Probabilities in Parallel Networks with Coupled Inputs<sup>1)</sup>

Jiyeon Lee<sup>2)</sup> and Min Hee Kweon<sup>3)</sup>

### Abstract

The simulation is used to estimate an overflow probability in a stable parallel network with coupled inputs. Since the general simulation needs extremely many trials to obtain such a small probability, the fast simulation is proposed to reduce trials instead. By using the Cramér's theorem, we first obtain an optimally changed measure under which the variance of the estimator is minimized. Then, we use it to derive an importance sampling estimator of the overflow probability which enables us to perform the fast simulation.

*Keywords* : Parallel networks, Importance sampling estimator, Overflow probability, Fast simulation

### 1. 소개

신뢰도가 아주 높은 시스템의 고장이나 안정된(stable) 대기 시스템에서의 과부하는 드물게 발생하는 희귀 사건(rare event)이다. 이러한 희귀 사건의 발생 확률을 일반적인 시뮬레이션 방법으로 추정하려고 하면 시뮬레이션의 시행 횟수를 아주 많이 해야 하는 문제가 생긴다 (Parekh and Walrand(1989), Lee(2000)). 이런 문제를 해결하기 위해 기존 시스템을 적당히 변화시켜 새로운 시스템을 얻고, 이 새로운 시스템에 대한 시뮬레이션 결과를 이용하면 희귀 사건의 발생 확률을 작은 시행 횟수로 빠르게 추정할 수 있다. 이때 주로 사용되는 방법이 중요 샘플링 방법(importance sampling method)(Glynn and Iglehart(1989))이다.

안정된 시스템  $S$ 에서 희귀 사건  $A$ 가 발생할 확률  $p := P(A)$ 에 관심이 있다고 하자. 기존의 시스템  $S$ 를 변형하여 새로운 시스템  $S'$ 과 사건  $A'$ 을 찾는다. 여기서 사건  $A'$ 은 시스템  $S'$ 에서 발생하는 사건으로서 시스템  $S$ 의 사건  $A$ 에 대응되는 것이다.  $S'$ 의 사건  $A'$ 이  $S$ 의 사건  $A$ 보다 훨씬 자주 발생하도록 시스템  $S'$ 을 찾는다면,  $S'$ 에서 사건  $A'$ 이 발생할 확률  $P'(A')$

---

1) This work was supported by Korea Research Foundation Grant (KRF-99-015-DP0045).

2) Assistant Professor, Department of Statistics, Yeungnam University, Kyeongsan 712-749, Korea.  
E-mail : leeje@yu.ac.kr

3) Graduate Student, Department of Statistics, Yeungnam University, Kyeongsan 712-749, Korea.

은 시뮬레이션을 통해 빠르게 추정할 수 있다. 이 값을 이용하면 기존의 시스템  $S$ 의 희귀 사건  $A$ 가 발생할 확률  $p$ 에 대한 추정도 가능하다.

일반적인 몬테카를로 시뮬레이션을 이용하면, 시스템  $S$ 에서  $n$ 번의 독립적인 시행(trial)으로부터  $p$ 의 추정량,

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \chi_A(\omega_i)$$

을 얻을 수 있다. 여기서  $\chi_A$ 는 지시함수(indicator function)로서

$$\chi_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

으로 정의되며  $\omega_i$ 는  $i$ 번째 시행의 결과(outcome)이다. 그리고 추정량  $\hat{p}_n$ 의 분산은

$$E[\hat{p}_n - p]^2 = \frac{1}{n}(p - p^2)$$

가 된다.

한편, 기존의 확률 측도  $P$ 가 변화된 시스템  $S'$ 의 확률 측도  $P'$ 에 대해 절대 연속(absolutely continuous)이면, 즉  $P(\omega) > 0$ 인 모든  $\omega$ 에 대해  $P'(\omega) > 0$ 이면

$$\begin{aligned} p &= \int \chi_A(\omega) dP(\omega) \\ &= \int \chi_A(\omega) \frac{dP(\omega)}{dP'(\omega)} dP'(\omega) \\ &= \int \chi_A(\omega) L(\omega) dP'(\omega) \end{aligned}$$

이 성립한다. 여기서  $L = dP/dP'$ 는 우도비(likelihood ratio) 혹은 Radon-Nikodym 도함수(Radon-Nikodym derivative)라고 부른다. 따라서 확률 측도  $P'$ 을 이용하면 확률  $p$ 에 대한 또 다른 추정량

$$\hat{p}_n' = \frac{1}{n} \sum_{i=1}^n \chi_A(\omega_i') L(\omega_i') \tag{1}$$

을 얻을 수 있다. 단,  $\omega_i'$ 는 확률 측도  $P'$ 을 갖는 시스템  $S'$ 에서 얻어지는  $i$ 번째 시행의 결과이다. 그리고 추정량  $\hat{p}_n'$ 의 분산은

$$E[\hat{p}_n' - p]^2 = \frac{1}{n} \left( \int \chi_A(\omega) L^2(\omega) dP'(\omega) - p^2 \right)$$

가 된다. 따라서 분산이 작은, 좋은 추정량  $\hat{p}_n'$ 을 찾기 위해서는  $\int \chi_A L^2 dP'$ 의 값이 가능한 작게 되도록 시스템  $S'$ 을 찾아야 한다. 이 때, 가장 작은 값을 갖는 시스템  $S^*$ 의 확률 측도  $P^*$ 를 최적의 변화된 확률 측도(optimally changed measure)라고 하고 확률 측도  $P^*$ 에 의해 얻어진 식 (1)의 추정량을 최적의 중요 샘플링 추정량(optimal importance sampling estimator)이라고 한다. 그러므로 중요 샘플링 방법에서는  $P^*$ 를 찾는 것이 중요한 문제가 된다. 본 논문에서 다루는 시스템  $S$ 는 마코비안 대기 네트워크(Markovian queueing network) 시스템으로서, 동일한 구조

(structure)를 갖는 대기 네트워크 시스템 중에서 최적의 변화된 시스템  $S^*$ 과 그 확률 측도  $P^*$ 를 찾는다. 마코비안 대기 네트워크의 확률 측도는 대기 네트워크를 구성하는 모수들, 즉 도착률과 서비스율, 그리고 경로확률(routing probability) 등에 의해 결정되기 때문에 이 경우에는 시스템  $S^*$ 를 구성하는 모수들만 찾으면 된다.

Sadowsky(1991)는 GI/GI/m의 대기 모형에서 과부하가 발생할 때까지의 평균 시간을 추정하기 위해 중요 샘플링 방법을 적용하였고, Parekh and Walrand(1989)와 Frater et al.(1991)은 잭슨 네트워크(Jackson network)에서 large deviation 이론(Shwartz and Weiss(1995), Varadhan(1984))의 Cramér 정리를 이용한 탐색적인 방법(heuristic method)으로 최적의 변화된 확률 측도를 찾았다. Frater와 Anderson(1994)은 이 방법을 적용하여 일반 서비스분포를 갖는 직렬 네트워크(tandem network)에서의 과부하 확률을 추정하였다. 이와 같이 지금까지의 연구는 단수 대기모형(single queueing model)이나 직렬 네트워크, 혹은 잭슨 네트워크와 같은 제한된 네트워크 모형에서만 가능하였다. 하지만 본 논문에서 다루게 되는 동시 입력이 있는 병렬 네트워크는 비잭슨 네트워크(non-Jackson network)로서, 그 적용 범위를 확장하는 것으로 볼 수 있다.

2장에서는 동시 입력이 있는 병렬 네트워크 모형을 소개하고, Parekh and Walrand(1989)의 탐색적 방법을 적용하여 최적의 변화된 확률 측도를 찾고 이것을 이용하여 과부하 확률의 중요 샘플링 추정량을 구한다. 3장에서는 시뮬레이션을 통해 일반 시뮬레이션과 빠른 시뮬레이션의 결과를 비교하고 마지막으로 4장에서는 그 결론을 정리한다.

## 2. 동시 입력이 있는 병렬 네트워크의 과부하 확률 추정

두 개의 노드가 병렬로 연결된 시스템이 있다. 각 노드에는 서비스율이 각각  $\alpha$ 와  $\beta$ 인 지수분포를 따르는 한 명의 서버가 있다. 그리고 손님은 도착률이 각각  $\lambda$ 와  $\mu$ 인 포아송 과정을 따르며 각 노드에 도착하고, 두 명의 손님이 도착률  $\nu$ 인 포아송 과정으로 같이 도착하여, 각 노드에서 따로 서비스를 받는다. 모든 손님은 각 노드에서 FIFO(First In First Out)의 규칙을 따른다. 이 모형은 Falto and Hahn(1984)에 의해 제기되었고 Wright(1992)에 의해 자세히 분석되어 Falto-Hahn-Wright 모형이라고도 부른다. 커풀이 함께 입장하여 각자 다른 서버에게 서비스를 받고 나가는 시스템에 적용할 수 있으며, 그러한 시스템의 신뢰도를 평가하는데 유용하게 사용될 수 있다. 본 논문에서는  $\lambda + \mu + \nu + \alpha + \beta = 1$ 로 가정하고 이 모형을 그림 1과 같이 모수  $\lambda, \mu, \nu, \alpha, \beta$ 들로 이루어진 시스템  $(\lambda, \mu, \nu, \alpha, \beta)$ 로 나타낸다.

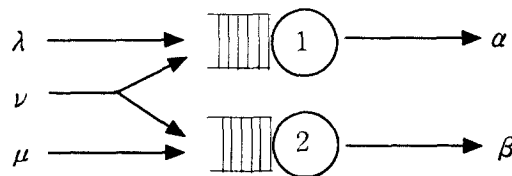


그림 1 시스템  $(\lambda, \mu, \nu, \alpha, \beta)$

시스템  $(\lambda, \mu, \nu, \alpha, \beta)$ 가 안정(stable)되기 위해서는 각 노드의 부하량(load)  $\frac{\lambda + \nu}{\alpha}$ 와  $\frac{\mu + \nu}{\beta}$ 에

대해서

$$\frac{\lambda + \nu}{\alpha} < 1 \text{ 그리고 } \frac{\mu + \nu}{\beta} < 1 \tag{2}$$

이 성립해야 함은 잘 알려져 있다. (Flatto and Hahn(1984), Flatto(1985) 그리고 Wright(1992)). Wright(1992)는 안정된 시스템  $(\lambda, \mu, \nu, \alpha, \beta)$ 에서 한 노드에 있는 손님의 수가 점점 많아질 때, 다른 노드에 있는 손님의 수의 상태를 분석하였다. 모수가  $\mu + \alpha\nu/(\lambda + \nu) > \beta$ 을 만족한다면, 노드 1에 손님이 많아질 때 노드 2도 불안정 상태가 되어 역시 손님이 점점 많아진다. 반대로  $\mu + \alpha\nu/(\lambda + \nu) < \beta$ 일 경우는, 노드 1에 손님이 많아지더라도 노드 2에 있는 손님 수는 안정 상태가 되어 일정하다. 두 노드는 대칭을 이루기 때문에 유사한 방법으로 노드 2에 손님이 점점 많아질 때 노드 1의 손님 수의 상태에 대한 설명도 가능하다.

본 논문에서는 시스템이 비어있는 상태에서 시작하여 다시 빈 상태가 되기 전에 시스템에 있는 총 손님의 수가 큰 수  $N$ 을 초과하는 사건  $A$ 가 발생하는 확률  $p$ 를 추정하고자 한다.  $N$ 을 시스템이 수용할 수 있는 최대 버퍼 수라고 하면, 확률  $p$ 는 시스템에서 버퍼 과부하(buffer overflow)가 발생하는 확률이 된다. 이 확률은 (2)의 조건을 만족하는 시스템에 대해서는 극히 작은 값을 가지기 때문에 일반 시뮬레이션으로는 추정이 힘들다. 따라서 중요 샘플링 추정량을 이용한 빠른 시뮬레이션을 구현하기 위해, 먼저 최적의 변화된 확률 측도를 구한다.

### 3.1 최적의 변화된 확률 측도

손님이 도착하거나 또는 서비스를 받고 시스템을 빠져나가는 것을 하나의 전이(transition)라고 하면, 전이가 일어나는 시점에서 각 노드에 있는 손님 수  $\{X(n) = (X_1(n), X_2(n)) \mid n = 0, 1, 2, \dots\} \in R^2$ 는 이산시간 마코프 체인(discrete-time Markov chain)으로 그림 2와 같은 전이 확률(transition probability)을 갖는다.

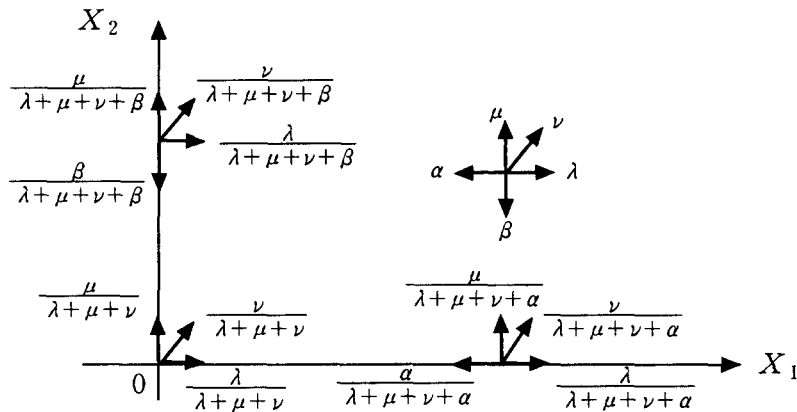


그림 2 시스템  $(\lambda, \mu, \nu, \alpha, \beta)$ 의 전이 확률

Parekh and Walrand(1989)의 탐색적 방법에 의하면 과부하 확률  $p$ 를 다음과 같이 추정할 수 있

다. (부록 참조)

$$p \approx \sum_T \sum_{\lambda'+\mu'+\nu'+\alpha'+\beta'=1} \exp \left\{ -T \left( \lambda' \cdot h_{\lambda} \left( \frac{1}{\lambda'} \right) + \mu' \cdot h_{\mu} \left( \frac{1}{\mu'} \right) + \nu' \cdot h_{\nu} \left( \frac{1}{\nu'} \right) + \alpha' \cdot h_{\alpha} \left( \frac{1}{\alpha'} \right) + \beta' \cdot h_{\beta} \left( \frac{1}{\beta'} \right) \right) \right\}, \quad (3)$$

여기서  $h_{\xi}(u)$ 는 평균이  $1/\xi$ 인 지수확률변수의 Cramér 변환(Cramér transform)으로서

$$h_{\xi}(u) = \begin{cases} \xi u - 1 - \log(\xi u), & u > 0 \\ 0, & \text{그외 } u \end{cases}$$

이고  $T$ 는 과부하가 발생할 때까지의 시간이다. 따라서 충분히 큰 값  $N$ 에 대해서는  $T \approx N \cdot R(\lambda', \nu', \mu', \alpha', \beta')$ 으로 나타낼 수 있다. 이때  $R(\lambda', \nu', \mu', \alpha', \beta')$ 은

$$R = \begin{cases} \frac{1}{\lambda' + \nu' - \alpha'}, & \lambda' + \nu' > \alpha' \text{ 그리고 } \mu' + \nu' < \beta' \text{ 일 때} \\ \frac{1}{\mu' + \nu' - \beta'}, & \lambda' + \nu' < \alpha' \text{ 그리고 } \mu' + \nu' > \beta' \text{ 일 때} \\ \frac{1}{\lambda' + 2\nu' + \mu' - \alpha' - \beta'}, & \lambda' + \nu' > \alpha' \text{ 그리고 } \mu' + \nu' > \beta' \text{ 일 때} \end{cases}$$

이다. 그러므로 식 (3)은 큰 값  $N$ 에 대해 근사적으로

$$p \approx \exp \{ -N \cdot \inf_{\lambda'+\mu'+\nu'+\alpha'+\beta'=1} H(\lambda', \mu', \nu', \alpha', \beta') \} \quad (4)$$

가 된다. 단,

$$H(\lambda', \mu', \nu', \alpha', \beta') := R(\lambda', \mu', \nu', \alpha', \beta') \left( \lambda' \cdot h_{\lambda} \left( \frac{1}{\lambda'} \right) + \mu' \cdot h_{\mu} \left( \frac{1}{\mu'} \right) + \nu' \cdot h_{\nu} \left( \frac{1}{\nu'} \right) + \alpha' \cdot h_{\alpha} \left( \frac{1}{\alpha'} \right) + \beta' \cdot h_{\beta} \left( \frac{1}{\beta'} \right) \right).$$

함수  $H$ 가 최소가 되는 값  $\lambda^*, \mu^*, \nu^*, \alpha^*, \beta^*$ 의 경로(path)들이 만약 과부하 확률  $p$ 의 대부분을 차지하는 우세한 경로들(dominated path)이라면, 이 때 얻어지는 값  $\lambda^*, \mu^*, \nu^*, \alpha^*, \beta^*$ 은 최적의 변화된 시스템을 나타내는 모수가 된다. 이것은 과부하가 발생하는 시점까지의 시스템의 평균 행태(average behaviour)가 최적의 변화된 시스템에서의 평균 행태와 일치하기 때문이다 (Parekh and Walrand(1989), Frater et al.(1991)).

함수  $R$ 의 각 범위에서 함수  $H$ 의 최소값을 찾자.

(i)  $\lambda' + \nu' > \alpha'$  그리고  $\mu' + \nu' < \beta'$ 일 때

함수  $H$ 는  $\lambda' = \frac{\alpha\lambda}{\lambda+\nu}$ ,  $\mu' = \mu$ ,  $\nu' = \frac{\alpha\nu}{\lambda+\nu}$ ,  $\alpha' = \lambda+\nu$ ,  $\beta' = \beta$ 에서 최소값  $\log\left(\frac{\alpha}{\lambda+\nu}\right)$ 을 가진다. 실제로 이 값들이 (i)의 범위를 만족하기 위해서는  $\mu + \frac{\alpha\nu}{\lambda+\nu} < \beta$ 이 성립해야 한다. 즉,  $\mu + \frac{\alpha\nu}{\lambda+\nu} < \beta$ 의 경우에는 노드 1의 도착률  $\lambda' + \nu'$ 이 서비스를  $\alpha'$ 보다 크기 때문에 손님의 수가 많아지고, 반대로 노드 2에서는 도착률  $\mu' + \nu'$ 이 서비스를  $\beta'$ 보다 작기 때문에 안정상태가 된다. 이것은 Wright(1992)가 얻은 결과와 일치한다.

(ii)  $\lambda' + \nu' < \alpha'$  그리고  $\mu' + \nu' > \beta'$ 일 때

함수  $H$ 가  $\lambda' = \lambda$ ,  $\mu' = \frac{\beta\mu}{\mu+\nu}$ ,  $\nu' = \frac{\beta\nu}{\mu+\nu}$ ,  $\alpha' = \alpha$ ,  $\beta' = \mu+\nu$ 에서 최소값  $\log\left(\frac{\beta}{\mu+\nu}\right)$ 을 갖기 위해서는  $\lambda + \frac{\beta\nu}{\mu+\nu} < \alpha$ 이 성립해야 한다. 그러므로 Wright(1992)의 결과와 같이,  $\lambda + \frac{\beta\nu}{\mu+\nu} < \alpha$ 를 만족하는 모수들에 대해서는 노드 2에서 만 손님의 수가 많아진다.

(iii)  $\lambda' + \nu' > \alpha'$  그리고  $\mu' + \nu' > \beta'$ 일 때

먼저 함수  $H$ 의 최소값을 갖게 하는  $\lambda'$ 를 찾기 위해,  $\lambda'$ 에 대하여 미분하고 조건  $\lambda' + \mu' + \nu' + \alpha' + \beta' = 1$ 을 이용하면

$$\nu\lambda'^3 + \lambda(\lambda + \mu)\lambda'^2 - \lambda^2\lambda' + (\alpha + \beta)\lambda^3 = 0 \tag{5}$$

을 얻는다. 이 식을  $\lambda'$ 에 대해 풀면 (iii)의 범위를 만족하는 해로서

$$\gamma := \lambda \cdot \frac{-(\lambda + \nu + \mu) + \sqrt{(\lambda + \nu + \mu)^2 + 4\nu(\alpha + \beta)}}{2\nu}$$

를 얻고  $\gamma > \lambda$ 인 것은 쉽게 보일 수 있다. 이 때, 나머지 모수들은 각각  $\mu' = \frac{\gamma}{\lambda}\mu$ ,  $\nu' = \left(\frac{\gamma}{\lambda}\right)^2\nu$ ,  $\alpha' = \frac{\lambda}{\gamma}\alpha$ ,  $\beta' = \frac{\lambda}{\gamma}\beta$ 가 된다. 이 값들이 실제로 범위 (iii)을 만족하기 위해서는

$$\lambda^2\gamma^2 + \nu\gamma^3 - \alpha\lambda^3 > 0 \tag{6}$$

$$\lambda\mu\gamma^2 + \nu\gamma^3 - \beta\lambda^3 > 0 \tag{7}$$

이 성립해야 한다.  $\gamma$ 는 식 (5)를 만족하기 때문에 식 (6)의 좌변은 조건  $\lambda + \mu + \nu + \alpha + \beta = 1$ 에 의해

$$\lambda^2\gamma^2 + \nu\gamma^3 - \alpha\lambda^3 = \lambda[\alpha\lambda(\gamma - \lambda) + \gamma\{\beta\lambda - \gamma(\mu + \nu)\}]$$

이 된다. 따라서, 만약  $\frac{\gamma}{\lambda} < \frac{\beta}{\mu + \nu}$  이면  $\lambda' + \nu' > \alpha'$ 가 성립함을 알 수 있다. 마찬가지로  $\frac{\gamma}{\lambda} < \frac{\alpha}{\lambda + \nu}$  이면 식 (7)의 조건도 만족되어,  $\mu' + \nu' > \beta'$ 가 성립함을 알 수 있다. 즉,  $\frac{\gamma}{\lambda} < \min(\frac{\alpha}{\lambda + \nu}, \frac{\beta}{\mu + \nu})$ 의 경우는, 두 노드 모두 손님이 많게 되고 이 때, 함수  $H$ 의 최소값은  $\log(\frac{\gamma}{\lambda})$ 이다.

$H$ 의 최소값을 찾기 위해서 위의 각 구간에서 구한 결과를 이용하면 다음과 같은 3개의 범위를 얻는다.

$$(a) \frac{\alpha}{\lambda + \nu} < \frac{\beta}{\mu + \frac{\alpha\nu}{\lambda + \nu}} < \frac{\beta}{\mu + \nu}$$

$$(b) \frac{\beta}{\mu + \nu} < \frac{\alpha}{\lambda + \frac{\beta\nu}{\mu + \nu}} < \frac{\alpha}{\lambda + \nu}$$

$$(c) \frac{\beta}{\mu + \frac{\alpha\nu}{\lambda + \nu}} < \frac{\alpha}{\lambda + \nu} < \frac{\beta}{\mu + \nu} \text{ 또는 } \frac{\alpha}{\lambda + \frac{\beta\nu}{\mu + \nu}} < \frac{\beta}{\mu + \nu} < \frac{\alpha}{\lambda + \nu}$$

함수  $H$ 는 (a)의 경우에  $\beta - \frac{\alpha}{\lambda + \nu}(\mu + \frac{\alpha\nu}{\lambda + \nu}) > 0$ 을 만족하고 안정성 조건  $\alpha/(\lambda + \nu) > 1$ 에 의해  $\beta - (\mu + \frac{\alpha\nu}{\lambda + \nu}) > 0$ 을 만족한다. 따라서 이 범위에서는 함수  $H$ 가 최소값  $\log(\frac{\alpha}{\lambda + \nu})$ 를 갖고 이때의 변화된 시스템의 모수는 (i)에서 구한 것과 같다. (b)의 경우에는,  $\alpha - \frac{\beta}{\mu + \nu}(\lambda + \frac{\beta\nu}{\mu + \nu}) > 0$ 이고 안정성 조건  $\beta/(\mu + \nu) > 1$ 에 의해  $\alpha - (\lambda + \frac{\beta\nu}{\mu + \nu}) > 0$ 이 성립하므로  $H$ 의 최소값은  $\log(\frac{\beta}{\mu + \nu})$ 가 되어 (ii)에서 구한 모수가 최적의 변화된 확률 측도를 나타낸다. 그리고 (c)의 경우에는  $\frac{\gamma}{\lambda} < \min(\frac{\alpha}{\lambda + \nu}, \frac{\beta}{\mu + \nu})$ 을 만족하여  $\log(\frac{\gamma}{\lambda})$ 가 함수  $H$ 의 최소값이 되고, 따라서 (iii)에서 구한 모수가 최적의 시스템을 나타낸다.

### 3.2 과부하 확률의 중요 샘플링 추정량

시스템  $(\lambda, \mu, \nu, \alpha, \beta)$ 에서 과부하가 발생하는 한 경로(sample path)  $\omega$ 에 대한 확률  $P(\omega)$ 를 구한다. 경로  $\omega$ 에 대해서 다음과 같은 변수를 정의하자.

$A_1$  : 노드 1에 도착률  $\lambda$ 로 도착한 손님 수

$A_2$  : 노드 2에 도착률  $\mu$ 로 도착한 손님 수

$A_3$  : 두 노드에 도착률  $\nu$ 로 동시에 도착하는 손님 쌍의 수

그리고  $N_1$ 을 과부하가 발생했을 때, 노드 1에 있는 손님의 수라고 하면, 노드 2에는  $N - N_1$ 의 손님이 있고, 노드 1에서 서비스를 받고 나간 손님의 수는  $A_1 + A_3 - N_1$ 이고 노드 2에서 나간 손님의 수는  $A_2 + A_3 - N + N_1$ 이 된다.

한편, 마코프 체인  $X(n)$ 은 그림 2에서 보는 바와 같이, 원점  $(0,0)$ 과 두 경계(boundary)  $B_i := \{(x_1, x_2) \mid x_{3-i} = 0, x_i = 1, 2, \dots, N-1\}$  ( $i = 1, 2$ )에서는 내부(interior)와 다른 전이 확률을 가진다.  $M_i$ 를 경로  $\omega$ 가 경계  $B_i$ 를 방문한 횟수라고 하고, 경계  $B_i$ 의 방문 후

- $I_i$  : 도착률  $\lambda$ 로 손님이 도착하는 전이가 발생하는 횟수
- $J_i$  : 도착률  $\mu$ 로 손님이 도착하는 전이가 발생하는 횟수
- $K_i$  : 도착률  $\nu$ 로 손님이 도착하는 전이가 발생하는 횟수

로 정의하면  $M_i - I_i - J_i - K_i$ 는 경계  $B_i$ 의 방문 후 손님이 서비스를 받고 나가는 전이가 발생하는 횟수가 된다.

시스템이 비어 있는 처음 상태  $(0,0)$ 에서 일어날 수 있는 전이는 그림 2에서처럼 손님이 시스템에 도착하는 것뿐인데 그 확률은 각각  $\frac{\lambda}{\lambda + \mu + \nu}$ ,  $\frac{\mu}{\lambda + \mu + \nu}$ ,  $\frac{\nu}{\lambda + \mu + \nu}$ 이다. 따라서  $P(\omega)$ 는

$$\begin{aligned} P(\omega) &= Q(\omega, \lambda, \mu, \nu, \alpha, \beta) \cdot \left( \frac{\lambda}{\lambda + \mu + \nu} \mu \nu + \frac{\mu}{\lambda + \mu + \nu} \lambda \nu + \frac{\nu}{\lambda + \mu + \nu} \lambda \mu \right) \\ &= Q(\omega, \lambda, \mu, \nu, \alpha, \beta) \frac{3\lambda\mu\nu}{\lambda + \mu + \nu} \end{aligned}$$

이다. 이때,  $Q(\omega, \lambda, \mu, \nu, \alpha, \beta)$ 는

$$\begin{aligned} Q(\omega, \lambda, \mu, \nu, \alpha, \beta) &= \left( \frac{\lambda}{\lambda + \mu + \nu + \alpha} \right)^{I_1} \left( \frac{\mu}{\lambda + \mu + \nu + \alpha} \right)^{J_1} \left( \frac{\nu}{\lambda + \mu + \nu + \alpha} \right)^{K_1} \\ &\quad \times \left( \frac{\alpha}{\lambda + \mu + \nu + \alpha} \right)^{M_1 - I_1 - J_1 - K_1} \left( \frac{\lambda}{\lambda + \mu + \nu + \beta} \right)^{I_2} \\ &\quad \times \left( \frac{\mu}{\lambda + \mu + \nu + \beta} \right)^{J_2} \left( \frac{\nu}{\lambda + \mu + \nu + \beta} \right)^{K_2} \left( \frac{\beta}{\lambda + \mu + \nu + \beta} \right)^{M_2 - I_2 - J_2 - K_2} \\ &\quad \times \lambda^{A_1 - I_1 - I_2 - 1} \mu^{A_2 - J_1 - J_2 - 1} \nu^{A_3 - K_1 - K_2 - 1} \\ &\quad \times \alpha^{A_1 + A_3 - N_1 - M_1 + I_1 + J_1 + K_1} \beta^{A_2 + A_3 - N + N_1 - M_2 + I_2 + J_2 + K_2} \end{aligned}$$

이다.



변화된 최적 시스템  $(\lambda^*, \nu^*, \mu^*, \alpha^*, \beta^*)$ 은 기존의 시스템과 같은 구조이며 다만 그 모수만 다르기 때문에, 시스템  $(\lambda^*, \nu^*, \mu^*, \alpha^*, \beta^*)$ 의 확률  $P^*(\omega)$ 도 같은 방법에 의해

$$P^*(\omega) = Q(\omega, \lambda^*, \nu^*, \mu^*, \alpha^*, \beta^*) \frac{3\lambda^* \nu^* \mu^*}{\lambda^* + \nu^* + \mu^*}$$

가 된다. 한편, 3.1절에서 얻은 모수  $\lambda^*, \nu^*, \mu^*, \alpha^*, \beta^*$ 에 대해서는  $P(\omega) > 0$ 인  $\omega$ 에 대해서  $P^*(\omega) > 0$ 을 만족하기 때문에, 우도비  $L(\omega) = \frac{dP(\omega)}{dP^*(\omega)}$ 은 모수들의 범위에 따라

$$L(\omega) = \begin{cases} \frac{\mu + \alpha}{\lambda + \mu + \nu} \left( \frac{\lambda + \nu}{\alpha} \right)^{N_1} \left( \frac{\mu + \alpha + \beta}{\lambda + \mu + \nu + \beta} \right)^{M_2}, & (a)\text{일 때} \\ \frac{\lambda + \beta}{\lambda + \mu + \nu} \left( \frac{\mu + \nu}{\beta} \right)^{N - N_1} \left( \frac{\lambda + \alpha + \beta}{\lambda + \mu + \nu + \alpha} \right)^{M_1}, & (b)\text{일 때} \\ \left( \frac{\lambda}{\gamma} \right)^N \left( \frac{\gamma + \left( \frac{\gamma}{\lambda} \right)^2 \nu + \frac{\gamma}{\lambda} \mu}{\lambda + \mu + \nu} \right) \left( \frac{\gamma + \left( \frac{\gamma}{\lambda} \right)^2 \nu + \frac{\gamma}{\lambda} \mu + \frac{\lambda}{\gamma} \alpha}{\lambda + \mu + \nu + \alpha} \right)^{M_1} \\ \quad \times \left( \frac{\gamma + \left( \frac{\gamma}{\lambda} \right)^2 \nu + \frac{\gamma}{\lambda} \mu + \frac{\lambda}{\gamma} \beta}{\lambda + \mu + \nu + \beta} \right)^{M_2}, & (c)\text{일 때} \end{cases}$$

로 얻어진다.

그러므로 시스템  $S^*$ 에 대한  $n$ 번의 시뮬레이션 시행에서

$$V_k = \begin{cases} 1, & k\text{번째 시행에서 과부하가 발생하면} \\ 0, & \text{그 외} \end{cases}$$

라 하고,  $L_k$ 를  $k$ 번째 시행의 우도비값이라고 하면 과부하 확률  $p$ 는

$$\hat{p}_n^* = \frac{1}{n} \sum_{k=1}^n L_k V_k$$

으로 추정할 수 있다.

### 3. 시뮬레이션

시스템  $(\lambda, \mu, \nu, \alpha, \beta)$ 의 몇 가지 모수들에 대해 일반 시뮬레이션과 최적의 확률 측도로 변화시켜 가속화한 시뮬레이션을 비교한다. 정확한 과부하 확률  $p$ 의 값은 Mathematica를 통해 얻은 것이다.

표 1에서 보는 바와 같이  $\lambda = 0.24, \mu = 0.01, \nu = 0.01, \alpha = 0.5, \beta = 0.24$ 는 2장의 (a)에 해당하는 경우로서  $N = 12$ 일 때, 정확한 과부하 확률은  $p = 3.269 \times 10^{-4}$ 이다. 일반 시뮬레이션의 결과를 보면,  $n = 5000$ 일 때  $\hat{p}_n = 3.100 \times 10^{-4}$ 이지만, 빠른 시뮬레이션의 경우  $n = 500$ 일 때

$\hat{p}_n^* = 3.216 \times 10^{-4}$ 로서 훨씬 더 작은 시행 횟수로 더 정확하게 추정할 수 있음을 볼 수 있다. 그리고  $\lambda = 0.01, \mu = 0.05, \nu = 0.1, \alpha = 0.59, \beta = 0.25$ 는 (b)의 범위에 해당하는 것으로, 이 경우도 빠른 시뮬레이션의 결과가 훨씬 뛰어남을 알 수 있다. 그러나 마지막 (c)의 경우에는는 다른 결과가 얻어졌다. 추정량  $\hat{p}_n^* = 5.9387 \times 10^{-11}$ 이  $p = 2.9665 \times 10^{-10}$ 와 많은 차이를 보이는 것이다. 그 이유는 두 노드의 부하량이 비슷하여 두 노드에서 모두 손님이 많아지는 경우는, 과부하 경로들 중에서 우세한 경로가 대부분의 과부하 확률을 나타낸다고 보기가 힘들기 때문이다.

표 2는 표 1의 시스템 I과 II에 대해 빠른 시뮬레이션을 20번 반복 실행하여 얻어진 중요 샘플링 추정량의 평균과 표준오차를 구한 결과이다.

표 1 동시 입력이 있는 병렬 네트워크의 시뮬레이션

방법	일반 시뮬레이션		빠른 시뮬레이션	
I. 시스템 (0.24, 0.01, 0.01, 0.5, 0.24) $N=12$				
$p = 3.269 \times 10^{-4}$				
$n$	2000	5000	100	500
$\hat{p}_n (\hat{p}_n^*)$	$3.000 \times 10^{-4}$	$3.100 \times 10^{-4}$	$3.192 \times 10^{-4}$	$3.216 \times 10^{-4}$
II. 시스템 (0.01, 0.05, 0.1, 0.59, 0.25) $N=20$				
$p = 6.608 \times 10^{-5}$				
$n$	10000	20000	300	500
$\hat{p}_n (\hat{p}_n^*)$	0.0	0.0	$6.6336 \times 10^{-5}$	$6.711 \times 10^{-5}$
III. 시스템 (0.2, 0.0452, 0.07, 0.48, 0.2048) $N=50$				
$p = 2.9665 \times 10^{-10}$				
$n$	10000	20000	1000	2000
$\hat{p}_n (\hat{p}_n^*)$	0.0	0.0	$5.9387 \times 10^{-11}$	$6.3004 \times 10^{-11}$

표 2 중요 샘플링 추정량의 평균과 표준오차 (20회의 시뮬레이션)

I. 시스템 (0.24, 0.01, 0.01, 0.5, 0.24) $N=12$		
$p = 3.269 \times 10^{-4}$		
$n$	100	500
평균	$3.196 \times 10^{-4}$	$3.2076 \times 10^{-4}$
표준오차	$2.325 \times 10^{-5}$	$1.619 \times 10^{-5}$
II. 시스템 (0.01, 0.05, 0.1, 0.59, 0.25) $N=20$		
$p = 6.608 \times 10^{-5}$		
$n$	300	500
평균	$6.6336 \times 10^{-5}$	$6.857 \times 10^{-5}$
표준오차	$8.8957 \times 10^{-6}$	$1.760 \times 10^{-6}$

#### 4. 결론

본 논문에서는 동시 입력이 있는 병렬 네트워크의 과부하 확률을 추정하기 위해 빠른 시뮬레이션 방법을 사용하였다. 일반적인 병렬 네트워크에서의 과부하 확률 추정을 위한 최적의 변화된 시스템은 노드의 부하량이 가장 큰 쪽에서 도착률과 서비스율이 서로 바뀐 시스템으로 얻어지는데 (Parekh and Walrand(1989)), 동시 입력이 있는 병렬 네트워크에서는 동시에 입장하는 손님의 도착 과정이 두 노드에 동시에 영향을 미치므로 일반적인 병렬 네트워크와 같이 두 노드의 부하량만으로는 두 노드의 상태를 설명할 수 없음을 알았다.

2장의 (a)와 (b)의 경우와 같이 두 노드의 부하량의 차이가 크게 되면 큰 부하량을 가진 노드에 손님의 수가 많아지고 그런 경로(path)가 과부하 현상을 잘 반영하고 또한 최적의 변화된 시스템도 설명해 준다. 하지만 동시 입력이 있게 되면 (c)의 경우처럼 두 노드가 동시에 과부하가 일어나는 경우가 있게 되는데, 이때는 Parekh and Walrand(1989)의 탐색적인 방법으로는 빠른 시뮬레이션을 위한 추정량을 찾을 수가 없음을 시뮬레이션의 결과로 확인하였다.

본 논문에서는 동시 입력이 있는 병렬 네트워크에 대해서 Parekh and Walrand(1989)의 방법으로 최적의 확률 측도를 찾을 수 있는 구간을 설명하였고, 그 구간 안에서는 시스템의 과부하 확률을 일반적인 시뮬레이션보다 빠른 시뮬레이션을 사용하여 더 빠르고, 더 효율적으로 추정할 수 있음을 보였다. 그러나, 두 노드의 부하량이 비슷하여 우세한 경로로 과부하 확률을 추정할 수 없을

때 즉, 시스템의 모수가  $\frac{\beta}{\mu + \frac{\alpha\nu}{\lambda + \nu}} < \frac{\alpha}{\lambda + \nu} < \frac{\beta}{\mu + \nu}$  또는  $\frac{\alpha}{\lambda + \frac{\beta\nu}{\mu + \nu}} < \frac{\beta}{\mu + \nu} < \frac{\alpha}{\lambda + \nu}$ 의 구

간 안에 있을 때의 과부하 확률을 추정하기 위한 방법은 아직 남아있는 과제로 앞으로의 연구 방향이 될 것이다.

#### References

- [1] Flatto, L. and Hahn, S. (1984), Two parallel queues created by arrivals with two demands I, *SIAM Journal of Applied Mathematics*, Vol. 44, 1041-1053.
- [2] Flatto, L. (1985), Two parallel queues created by arrivals with two demands II, *SIAM Journal of Applied Mathematics*, Vol 45, 861-878.
- [3] Frater, M. R., Lennon, T. M. and Anderson, B. D. O. (1991), Optimally Efficient Estimation of the Statistics of Rare Events in Queueing Networks, *IEEE Transactions on Automatic control*, Vol. 36, 1395-1405.
- [4] Frater, M. R. and Anderson, B. D. O. (1994), Fast simulation of buffer overflows in tandem networks of GI/GI/1 queues, *Annals of Operations Research*, Vol. 49, 207-220.
- [5] Glynn, P. W. and Iglehart, D. L. (1989), Importance Sampling for Stochastic Simulations, *Management Science*, Vol. 35, 1367-1392.
- [6] Lee, J. (2000), Fast Simulation for Excessive Backlogs in Tandem Networks, *The Korean Communications in Statistics*, Vol. 7, 499-512.
- [7] Parekh, S. and Walrand, J. (1989), A quick simulation method for excessive backlog in

networks of queues, *IEEE Transactions on Automatic Control*, Vol. 34, 54-66.

[8] Sadowsky, J. S. (1991), Large Deviations Theory and Efficient Simulation of Excessive Backlogs in a GI/GI/m Queue, *IEEE Transactions on Automatic Control*, Vol. 36, 1383-1394.

[9] Shwartz, A. and Weiss, A. (1995), *Large Deviations for Performance Analysis*, Chapman & Hall.

[10] Wright, P. E. (1992), Two parallel processors with coupled inputs, *Advances in Applied Probability*, Vol. 24, 986-1007.

[11] Varadhan, S. R. S. (1984), *Large deviations and applications*, SIAM, Philadelphia, PA.

### 부록

정리 (Cramér의 정리)(Varadhan(1984)) 실수  $R$ 에 정의되는 I.I.D. 확률 변수들  $X_1, X_2, \dots, X_n$ 이 분포함수  $F$ 와 평균  $m$ 을 가지며  $F$ 의 라플라스 변환(Laplace transform)

$$M(s) = \int e^{sz} dF(z)$$

가 0 근처에서 유한하다고 가정하자. 그러면  $a > m$ 에 대해서

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P\left\{ \frac{X_1 + X_2 + \dots + X_n}{n} > a \right\} = -h(a)$$

이다. 여기서 함수  $h$ 는

$$h(y) = \sup_{s \in R} (sy - \log M(s))$$

로 정의되며 Cramér 변환이라 부른다.

Parekh and Walrand(1989)는 위의 Cramér 정리를 이용한 탐색적인 방법으로 식 (3)을 유도하였다. 그 과정은 다음과 같다.  $\{Y_i^\xi, i=1, 2, \dots\}$ 를 평균이  $1/\xi$ 인 I.I.D. 지수확률변수들이라고 하자. 그러면 시스템  $(\lambda, \mu, \nu, \alpha, \beta)$ 에서  $Y_i^\lambda, Y_i^\mu, Y_i^\nu$ 는 각각의 도착 간격을 나타내는 변수들이고  $Y_i^\alpha$ 와  $Y_i^\beta$ 는 각 노드에서의 가상(virtual) 서비스 시간을 나타내는 변수들이다. 과부하 사건  $A$ 를 구성하는 경로 중에서 과부하가 발생할 때까지 일어난 전이의 총 횟수가  $T$ 인 경로들에 대해서는  $\lambda'T + \mu'T + \nu'T + \alpha'T + \beta'T = 1$ 이라고 하면, 총 전이 횟수  $T$  중에서  $\lambda'T$ 가 노드 1에 손님이 도착하는 전이가 일어난 횟수,  $\mu'T$ 는 노드 2에 손님이 도착하는 전이가 일어난 횟수,  $\nu'T$ 는 두 노드에 손님이 같이 도착하는 전이가 일어난 횟수,  $\alpha'T$ 는 노드 1에서 서비스를 끝내고 손님이 떠나는 전이가 일어난 횟수, 그리고  $\beta'T$ 는 노드 2에서 손님이 떠나는 전이가 일어난 횟수이다. 한편, 총 전이의 횟수  $T$ 는 균일화 기법(uniformization)에 의해 과부하가 발생할 때까지의 시간과 일치한다. 따라서 위의 Cramér 정리를 이용하면,  $T$ 시간 동안 노드 1에서  $\lambda'T$ 개의 입력 전이가 일어날 확률은 근사적으로

$$\begin{aligned}
 & P\{Y_1^\lambda + Y_2^\lambda + \dots + Y_{\lambda'T}^\lambda \approx T\} \\
 &= P\left\{\frac{Y_1^\lambda + Y_2^\lambda + \dots + Y_{\lambda'T}^\lambda}{\lambda'T} \approx \frac{1}{\lambda'}\right\} \\
 &\approx \exp\left(-\lambda'T h_\lambda\left(\frac{1}{\lambda'}\right)\right)
 \end{aligned}$$

이 된다. 마찬가지로 나머지 전이에 대해서도

$$\begin{aligned}
 & P\{Y_1^\mu + Y_2^\mu + \dots + Y_{\mu'T}^\mu \approx T\} \approx \exp\left(-\mu'T h_\mu\left(\frac{1}{\mu'}\right)\right) \\
 & P\{Y_1^\nu + Y_2^\nu + \dots + Y_{\nu'T}^\nu \approx T\} \approx \exp\left(-\nu'T h_\nu\left(\frac{1}{\nu'}\right)\right) \\
 & P\{Y_1^\alpha + Y_2^\alpha + \dots + Y_{\alpha'T}^\alpha \approx T\} \approx \exp\left(-\alpha'T h_\alpha\left(\frac{1}{\alpha'}\right)\right) \\
 & P\{Y_1^\beta + Y_2^\beta + \dots + Y_{\beta'T}^\beta \approx T\} \approx \exp\left(-\beta'T h_\beta\left(\frac{1}{\beta'}\right)\right)
 \end{aligned}$$

을 얻을 수 있다. 각 도착 과정과 서비스 분포는 서로 독립이므로 총 전이 횟수가  $T$ 인 경로들의 확률은 근사적으로

$$\begin{aligned}
 & \exp\left\{-T\left(\lambda' \cdot h_\lambda\left(\frac{1}{\lambda'}\right) + \mu' \cdot h_\mu\left(\frac{1}{\mu'}\right) + \nu' \cdot h_\nu\left(\frac{1}{\nu'}\right) \right. \right. \\
 & \quad \left. \left. + \alpha' \cdot h_\alpha\left(\frac{1}{\alpha'}\right) + \beta' \cdot h_\beta\left(\frac{1}{\beta'}\right)\right)\right\}
 \end{aligned}$$

이 된다. 따라서 과부하 사건  $A$ 의 확률  $p$ 는 근사적으로

$$\begin{aligned}
 p \approx \sum_T \sum_{\lambda'+\mu'+\nu'+\alpha'+\beta'=1} \exp\left\{-T\left(\lambda' \cdot h_\lambda\left(\frac{1}{\lambda'}\right) + \mu' \cdot h_\mu\left(\frac{1}{\mu'}\right) \right. \right. \\
 \left. \left. + \nu' \cdot h_\nu\left(\frac{1}{\nu'}\right) + \alpha' \cdot h_\alpha\left(\frac{1}{\alpha'}\right) + \beta' \cdot h_\beta\left(\frac{1}{\beta'}\right)\right)\right\}
 \end{aligned}$$

으로 추정할 수 있다.