

사용자 제한이 적용되는 2계층 대기행렬 네트워크 구조의 이탈과정에 관한 분석

이 영*

A Study of Departure Process on the Open and Nested Population Constrained Tandem Queueing Network with Constant Service Times

Young Rhee*

■ Abstract ■

In this paper, we consider the departure process from the open and nested tandem queueing network with population constraint and constant service times. It is known that the queueing network can be transformed into a simple queueing network which can be easy to analyze. Using this simple queueing network, upper and lower bounds on the inter-departure time are obtained. We prove that the variance of the interdeparture time is bounded within these two bounds. Validation against simulation data is shown that how it works the variance of the interdeparture time within two bounds. These bounds can be applied to obtain the better variance of the interdeparture time using a suitable method.

Keyword : Population Constraint, Sliding Window, Nested Queueing Network

1. 서 론

대기행렬 네트워크 구조에서 고객이나 데이터는

몇 개의 계층을 통과한 후에 네트워크를 이탈하는 구조가 일반적인 모형이라고 할 수 있다. 본 연구에서는 다중계층, 특별히 2계층구조를 갖는 대기행렬

논문접수일 : 2009년 01월 13일 논문게재확정일 : 2009년 11월 30일

논문수정일(1차 : 2009년 03월 09일, 2차 : 2009년 04월 21일, 3차 : 2009년 07월 13일, 4차 : 2009년 11월 12일)

* 계명대학교 공과대학 경영공학과

네트워크에서 이탈과정을 분석하고자 한다. 각 계층의 고객 제한은 세마포(semaphore) 대기행렬로 조절된다. 세마포 대기행렬은 계층의 제한인원을 반영하는 k 개의 토큰이 보관되는 토큰 풀(pool)과 고객이 대기행렬 네트워크에 진입하기 위한 외부대기(external queue)로 구성되어 있고 대기고객수(queue size)는 무한이다. 대기행렬 네트워크에 대한 고객의 도착은 외부대기를 통하여 이루어지며 토큰 풀에서 한 개의 토큰을 획득한 후에 대기행렬 네트워크 내부에 진입할 수 있다. 서비스를 전체를 마침 고객은 즉시 토큰 풀에 토큰을 반납한 후에 다음 계층에 진입하게 되며, 이후의 서비스를 계속하는 과정은 동일하게 진행된다. 만약 대기행렬 네트워크에 도착한 고객이 토큰 풀에 토큰을 없는 경우에는 자신에게 할당되는 토큰이 반납될 때까지 외부대기에서 대기한다. 따라서 세마포 대기행렬 구조가 적용되는 대기행렬 네트워크에서는 고객 초과에 따른 차단(blocking)으로 인하여 도착고객의 손실은 없다.

폐쇄형(closed) 대기행렬 네트워크의 종대종 원도 흐름제어(flow control)에 관한 연구는 고객의 손실이라는 가정에서 컴퓨터 통신 시스템에서 처음으로 적용되었다[6, 10]. 한편으로 세마포 대기행렬과 유사한 슬라이딩 윈도우(sliding window)기법을 적용한 상태 의존 서비스율을 기반으로 한 흐름등가 서버(flow equivalent server)에 관한 연구도 있다[5, 11]. 계층적(hierarchical) 분해(decomposition)와 결합(aggregation)방법을 적용하여 실제 데이터 전송 과정의 특성을 파악하였다[2]. 또한 도착과정이 일반 분포 하에 고객수를 제한하는 개방형 대기행렬 네트워크에 관한 연구는 서비스 분포가 상수라는 가정을 하여 진행되었다[7, 8].

다계층 구조의 통신 네트워크에 관한 분석은 통신 기능 자체가 각 계층의 제한적 기능과도 연관이 있는데, 각 계층의 상위 또는 하위 계층의 제한적 요소에 의존적이라는 것이다. 하위 계층은 고유의 원천적인 구조적 기능을 수행하는 반면에 상위 계층은 하위 계층의 흐름과 밀접한 관계가 있어서 기능 수행이 제한적이라는 것이다. 왜냐하면 하위 계

층의 이탈과정은 다음 상위 계층에 영향을 주기 때문이다. 다계층 모델을 폐쇄형 대기행렬 네트워크에 적용한 사례는 슬라이딩 윈도우 흐름제어를 모델화하는 것으로 시작하여[1], 등지(nested)형과 직렬형(tandem)이 결합된 슬라이딩 윈도우 흐름제어 네트워크를 분석하는 연구가 진행되었는데, 등지형 네트워크의 분석은 다계층 구조 네트워크 분석의 출발점이었다. 슬라이딩 윈도우 제어 상태의 각 계층을 흐름등가 방법을 적용하여 상태 의존적인 무한대개의 서버를 가정함으로써 진행되었다. 보다 현실적으로는 슬라이딩 윈도우 흐름 제어와 패킷 분할(packet fragmentation)을 수반하는 다계층의 싱글 홉(single hop) OSI 구조의 네트워크에 적용되었다. 이러한 대기행렬 네트워크는 다단계(multi-stage) 생산시스템에도 적용되었는데, 이러한 생산시스템을 분석하기 위하여 각 단계의 생산 활동을 기능적으로 결합시킴으로서 각 단계의 kanban 생산제어를 적용하기 위한 것 이었다[3]. 가장 최근에는 각 계층에 슬라이딩 윈도우가 적용되고, 일반적인 도착분포와 각 노드의 서비스 분포가 상수라는 가정과 함께 개방형 다계층 대기행렬 네트워크를 네트워크 차원(dimension)을 줄이는 모델을 제시하는 연구가 있었는데, 즉 다계층 구조의 아무리 복잡한 대기행렬 네트워크 일지라도 네트워크를 단순화하는 연구도 진행되었다[9]. 현대의 통신 네트워크의 프로토콜은 OSI 모델을 기본으로 계층적 구조라는 것을 의심하는 경우는 없다. 바로 슬라이딩 윈도우를 적용한 연구가 활발해진 이유이기도 하다. 통신 네트워크의 구조는 고전적인 데이터 통신, 인터넷 통신 등은 계층적 구조로 이루어져 있으며, 각 계층의 프로토콜에는 해당 계층에 대한 윈도우 사이즈가 있음도 알 수 있다.

본 연구는 [9]의 연구를 확장시키기 위한 것으로 제 2장에서는 분석하고자 하는 모델이 세부적으로 설명되며, 제 3장에서는 본 연구에서 성능 지표로 사용하고자 하는 출력 프로세스와 연관 있는 분산을 분석할 수 있는 상한과 하한을 제시하고, 이를 이용한 근사치 방법론이 제시되며, 제 4장에서는 실제 모의실험의 결과와 근사치 계산 결과를 비교분

석하며, 제 5장에서는 결론을 맺는다.

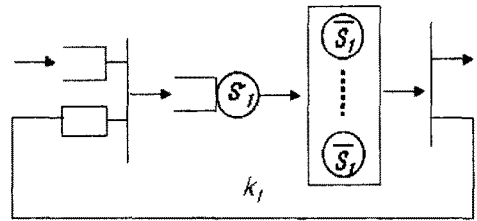
2. 연구모델

[그림 1]과 같은 등지형 대기행렬 네트워크를 고려해 보자. 각 노드의 서비스시간은 상수이며, 당연히 세마포가 적용되어 각 계층마다 제한된 고객의 수를 포함 가능하다고 가정한다.

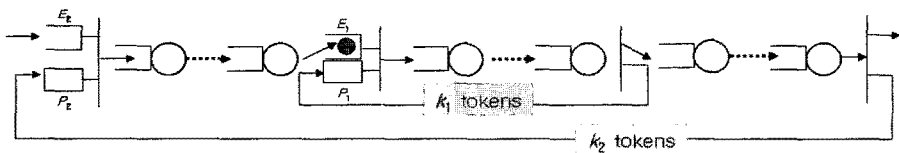
[그림 1]에서 일반적으로 네트워크의 내부, 즉 순수하게 등지화된 k_1 개의 토큰으로 구성된 일련의 네트워크를 하위계층이라고 부르며, 네트워크의 외부, 즉 k_2 개의 토큰으로 구성된 바깥 네트워크를 상위계층이라 한다. 또한 대기행렬 네트워크는 n 개의 노드로 구성된 하위계층과 m 개의 상위계층 노드를 가지고, 각 노드의 서버수는 1이며, 고객의 시스템에 대한 도착분포는 일반적이며 서로 독립적인 분포(general independent)를 따른다고 가정한다. 도착하는 고객은 토큰풀 P_2 에서 토큰과 함께 상위계층으로 진입하여 각 노드의 서비스를 완료하고, 하위계층 외부대기 E_1 에 진입한다. 이어서 P_1 에서의 토큰과 함께 하위계층 네트워크의 모든 노드에서 서비스를 완료한다. 하위계층의 서비스를 마친 후에는 토큰을 즉시 토큰풀 P_1 에 반납하며, 이후 상위계층의 나머지 노드에서 서비스과정을 종료한 후에는 토큰을 즉시 토큰풀 P_2 에 반납한다. 만약 도착하는 고객이 토큰풀 P_2 에 토큰이 없는 경우는 순

서에 의거 자신에게 토큰이 할당될 때까지 외부대기인 E_2 에서 대기한다. 하위계층의 입구인 E_1 에서도 토큰풀에 토큰이 없는 경우는 같은 상황이 되풀이된다. [그림 2]는 등지형 대기행렬 네트워크에서 고객과 토큰의 흐름을 묘사한 것으로서 점선은 상위계층과 하위계층의 토큰 흐름을 나타낸 것이며, 실선은 고객의 흐름을 나타낸 것이다.

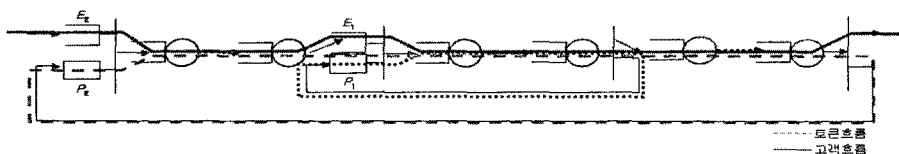
[그림 1]의 등지형 대기행렬 네트워크는 고객의 대기행렬 네트워크 내에서 머무르는 시간관점에서 일련의 몇 차례 변환과정을 거쳐 네트워크를 단순화시킬 수 있는데, 이를 네트워크의 단순화과정(reduction of network dimensionality)라고 한다. 진행되는 과정을 보면, 첫째로 등지화된 하위계층 대기행렬 네트워크는 두 개의 노드로 구성된 네트워크로 구성할 수 있음을 표본행로분석(sample path analysis)의 진화등식(evolutionary equation)을 적용하여 단순화된 2노드 대기행렬 네트워크로 변환할 수 있음을 [그림 3]과 같이 증명하였다[9].



[그림 3] 2노드 대기행렬 네트워크

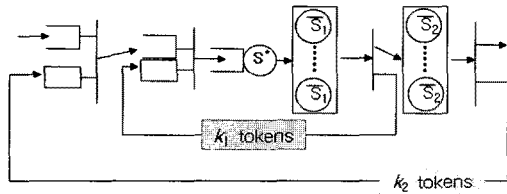


[그림 1] 등지형 대기행렬 네트워크



[그림 2] 고객과 토큰의 흐름도

[그림 3]에서 s_1^* 는 각 노드의 서비스 시간 중에서 가장 긴 시간을 의미하며, \bar{s}_1 는 총 서비스에서 s_1^* 를 제외한 값이다. 또한 2노드 대기행렬 네트워크에서 첫 번째 노드의 서버수는 1이며 두 번째 노드의 서버수는 무한수다. 등지화된 하위계층 대기행렬 네트워크의 단순화에 이어 전체 대기행렬 네트워크의 단순화 과정은 좀 더 복잡하지만 고객당 네트워크에 머무르는 시간 관점에서 6개의 노드로 구성된 대기행렬 네트워크로 변환되고, 다시 5개의 노드로 구성된 대기행렬 네트워크로 변환되어, 최종적으로 3노드 대기행렬 네트워크로 변환될 수 있다. 이에 대한 증명은 몇 차례의 진화등식을 이용하였다[9]. 3노드 대기행렬 네트워크 역시 마지막 노드의 서버수도 무한수다.



[그림 4] 3노드 대기행렬 네트워크

[그림 4]에서 s^* 는 상·하위계층 전체 노드의 서비스 시간 중에서 가장 긴 시간을 의미하며, \bar{s}_2 또한 상위계층 전체노드의 서비스 시간을 합한 값이다. [그림 4]의 모형은 계층의 토큰에 대한 제약성을 반영하기 위하여 $k_1 s^* \leq s^* + \bar{s}_1$ 와 $k_2 s^* \leq s^* + \bar{s}_1 + \bar{s}_2$ 를 가정한다.

3. 모델의 특성분석

본 절에서는 [그림 3]과 같이 2노드 대기행렬 네트워크의 출력 프로세스를 분석하기 위하여 세마포로 조정되고 있는 이탈사이시간(interdeparture time)의 분산을 분석하는 기본적인 특성을 소개하고자 한다. [그림 3]에서 첫 노드를 떠나는 고객들의 이탈사이시간에 대한 분산은 대기행렬 네트워크 전체, 즉 두 번째 노드를 떠나는 이탈사이시간에 대한 분

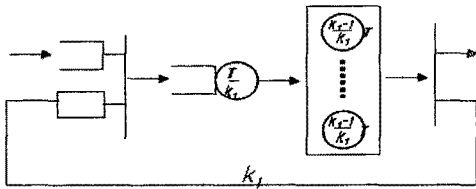
산과 동일하다. 이는 두 번째 노드의 서버수가 무한이기 때문이며, 모델을 좀 더 신중하게 살펴보면 대기행렬 네트워크에 허용되는 토큰수가 k_1 개이기 때문에 k_1 의 서버가 존재한다고 보면 된다. T 를 전체 서비스 시간의 합, 즉 $s^* + \bar{s}_1$ 이라 하면, $k_1 s_1^* \geq T$ 일 경우는 본 대기행렬 네트워크로부터 이탈사이시간 프로세스는 정확히 서비스 시간이 상수, s_1^* 인 G/D/1 대기행렬과 정확히 일치한다. $V[a]$ 를 도착분포의 분산이라 가정하고, $E[w]$ 를 고객의 시스템에서의 평균대기시간이라고 정의한다면, 서비스 시간이 s_1^* 인 G/D/1 대기행렬의 이탈사이시간 τ 의 분산은 익히 알려진 다음과 같은 결과로 정리할 수 있다 Marchal[4].

$$V[\tau] = V[a] - 2\left(\frac{1}{\lambda} - s_1^*\right)E[w] \tag{1}$$

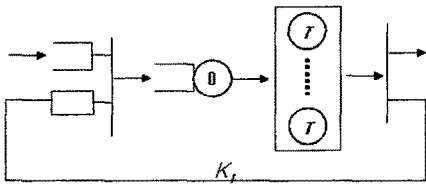
식 (1)에서 λ 는 고객의 대기행렬 네트워크에 대한 도착률을 의미하며, 식 (1)이 시사하는 점은 평균대기시간, $E[w]$ 에 의하여 이탈사이시간에 관한 분산 $V[\tau]$ 가 결정된다는 점이다. 대기행렬 네트워크에서 소요되는 평균대기시간에 대한 수리적인 연구는 특별한 경우를 제외하고는 명확한 방법이 제시된 바가 없다. 따라서 대기행렬 네트워크의 평균대기시간을 추정하는 방법론으로 대기행렬 네트워크를 대표할 수 있는 단순 대기행렬로 단순화 및 대표화하여 평균대기시간에 대한 상한과 하한을 연구하는 방법이 하나의 대안이 될 수 있다. 2노드 대기행렬 네트워크에 대한 평균대기시간을 구하는 방법은 이러한 상한과 하한에 기초한 연구 결과라고 할 수 있다[6]. 또한 평균대기시간의 상한이 식 (1)에서 보듯이 $V[\tau]$ 의 하한값을 결정하게 되고, 반대로 평균대기시간의 하한값은 $V[\tau]$ 의 상한값을 제공하기 때문이다. 평균대기시간의 상한은 다음과 같은 모형을 이용하였다.

본 연구에서는 [그림 3]의 2노드 대기행렬 네트워크에서 $k_1 s_1^* \leq T$ 를 가정하기 때문에 [그림 5]의 첫 노드에서의 고객 당 대기시간은 $\frac{T}{k_1} \geq s_1^*$ 가 성립

되어 [그림 3]의 모형에서 발생하는 고객 당 대기시간보다 증가하게 된다. 더욱 나아가 [그림 5]를 검토해보면 고객 당 대기시간은 두 번째 노드에는 최대 $(k_1 - 1)$ 명의 고객만이 서비스를 받을 수 있는 상황이 된다. 이는 두 번째 노드에서는 첫 노드에서의 서비스 시간이 $\frac{T}{k_1}$ 이기 때문에 고객들의 도착사이 시간이 $\frac{T}{k_1}$ 보다 크거나 동일하게 된다. 따라서 전체 유입 인원을 조정하는 세마포기능이 상실하게 되어 고객 당 대기시간은 세마포 기능이 없는 단단계의 노드로 구성된 대기행렬과 동일하게 된다. 따라서 [그림 5]의 고객의 대기시간 측면에서 상한 모형은 서비스 시간이 $\frac{T}{k_1}$ 인 GI/D/1 대기행렬과 동일하게 된다.



[그림 5] 평균대기시간, $E[w]$ 의 상한 모형

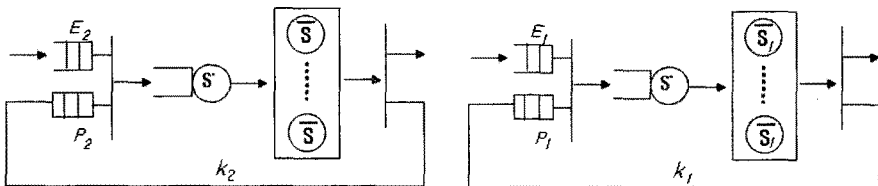


[그림 6] 평균대기시간, $E[w]$ 의 하한 모형

평균대기시간, $E[w]$ 의 하한 모형은 [그림 6]과 같

이 2노드 대기행렬 네트워크로 표현이 가능한데, 실제 첫 노드에서의 대기시간이 없으므로 도착하는 고객들은 외부대기에서만 기다린다고 생각하면 된다. [그림 6]은 다시 단순한 단단계의 노드로 구성된 대기행렬로 가능한데, 이는 상수 서비스 시간을 갖기 때문에 매번 k_1 번째 도착하는 고객은 같은 토큰을 이용할 수 있기 때문이다. 따라서 [그림 6]모형의 대기시간 측면에서 하한 모형은 서비스 시간이 상수라는 특성 때문에 $G/D/k_1$ 대기행렬과 동일하며, 달리는 k_1 곱슬(convolution) 도착분포, $G \otimes G \dots \otimes G/D/1$ 대기행렬로서 서비스 시간이 상수, T 와 동일하게 된다.

[그림 4]와 같은 등지형 대기행렬 네트워크에 관한 시스템 내에서의 평균대기시간을 추정하기 위한 상한과 하한을 제시하는 모형은 다소 복잡하나 다음과 같이 설명될 수 있다. [그림 4]와 같은 등지형 대기행렬 네트워크에서 상위계층과 하위계층의 세마포 제약조건 중에서 $k_1 \leq k_2$ 이라는 가정 하에서, 평균대기시간의 하한을 제시하는 모델은 k_1 개의 토큰으로 구성된 하위계층의 제약조건을 무력화하는 방법으로서 k_1 을 k_2 로 간주하면 하위계층의 외부대기, E_1 에서 대기시간이 발생하지 않으므로 평균 대기시간을 제시하는 하한모형은 [그림 3]과 같이 k_2 개의 토큰으로 구성되고, 첫 번째 노드의 서비스 시간은 s^* , 두 번째 노드의 서비스 시간은 \bar{s} 로서 $\bar{s}_1 + \bar{s}_2$ 이다. 하한모형의 다른 경우는 상위계층을 무력화하는 방법인데, 이는 k_2 를 무한수로 가정한다면 간단히 성립된다. 이 경우 k_1 개의 토큰으로 구성되고, 첫 번째 노드의 서비스 시간은 s^* , 두 번째 노드의 서비스 시간은 \bar{s}_1 로서 구성된 모형으로 평균대



[그림 7] 평균대기시간의 하한모형

기시간의 하한을 제시하는 모델은 [그림 7]과 같다.

이와 반대로 평균대기시간을 제시하는 상한모형은 [그림 4]에서 보듯이 상위계층의 토큰 수, k_2 를 k_1 으로 간주하면 하위계층의 외부대기, E_1 에서 대기시간이 발생하지 않게 된다. 따라서 평균대기시간의 상한 모형은 [그림 3]과 같이 k_1 개의 토큰으로 구성되고, 첫 번째 노드의 서비스시간은 s^* , 두 번째 노드의 서비스 시간은 \bar{s} 로 구성된 모형이라 할 수 있다. 또한 다른 경우의 평균대기시간에 대한 상한모형은 [8]에 제시되어 있다. 따라서 k_2 개의 토큰으로 구성되고, 첫 번째 노드의 서비스시간은 s^* 로 서 $\frac{s^* + \bar{s}_1}{k_1}$ 이며, 두 번째 노드의 서비스 시간은 \bar{s} 는 $(s^* + \bar{s}_1 + \bar{s}_2) - (\frac{s^* + \bar{s}_1}{k_1})$ 로 구성된 모형이라 할 수 있으며 [그림 8]과 같다.

4. 모의실험과 분석

본 절에서는 제 3장에서 제시한 2노드 대기행렬 네트워크에 대한 이탈사이시간 τ 의 분산에 대한 분석적인 결과와 실제 모의실험 도구인 SLAM을 이용하여 얻은 결과를 비교하여 이탈사이시간 τ 의 분산이 상한과 하한에 위치함을 결과로 제시하고, 나아가 본 연구에서 제시한 동지화된 대기행렬 네트워크의 이탈사이시간 τ 에 대한 분산을 같은 방법으로 제시할 것이다. 이탈사이시간 τ 에 대한 분산의 하한을 제시하는 $G/D/1$ 대기행렬의 이탈사이시간 τ 의 분산은 식 (1)을 이용하여 분석적으로 계산이 가능하나, $G/D/k_1$ 대기행렬 또한 앞서 밝힌 바와 같이 이탈사이시간 τ 에 대한 분산의 상한을 식 (2)를

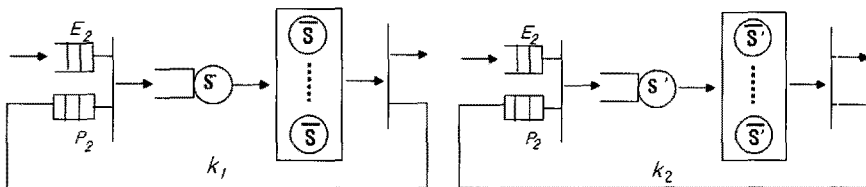
이용하여 계산이 가능하다. 그러나 $G/D/k_1$ 대기행렬의 이탈사이시간 τ 에 대한 분산은 정확도에 있어서 식 (1)보다 더 좋은 결과를 소개하고자 한다. [그림 6]과 같이 서비스 시간이 상수 T 를 갖는 $G/D/k_1$ 대기행렬의 이탈사이시간 τ 에 대한 분산분석은 Whitt [12]에 의하여 연구는 다음과 같다.

$$V[\tau] = (1 - \rho^2) V[a] + (\frac{T}{k_1})^2 (1 - k_1^{-0.5}) \quad (2)$$

이상에서 ρ 가 의미는 $\frac{\lambda T}{k_1}$ 를 말하며 식 (2)는 고객 대기행렬 네트워크에 도착하는 분포에 따라 정확도에서 약간의 보정이 필요한 것으로 밝혀졌다. 다시 말하면 식 (2)는 도착분포에 대한 분산계수(coefficient of variation)의 곱, c_a^2 이 1이 아닌 경우에 적용이 가능하나, $c_a^2 \geq 1$ 인 경우는 약간의 보정을 하면 더 나은 결과를 얻을 수 있다. 특별한 경우, 즉 c_a^2 이 1보다 월등히 큰 경우는 다음과 같은 보정이 필요하다고 할 수 있다[13].

$$V[\tau] = (1 - \rho^2) V[a] + \frac{1 + c_a^2}{2} (\frac{T}{k_1})^2 (1 - k_1^{-0.5}) \quad (3)$$

모의실험에서는 분석적인 방법, 즉 식 (1), 식 (2)와 식 (3)을 이용하여 상한과 하한을 계산하고, 실제 [그림 3]의 2노드 대기행렬 네트워크에 대하여 모의실험 결과가 제시될 것이다. 본 연구에서는 이탈사이시간 τ 에 대한 분산에 대한 상한과 하한, 그리고 모의실험 결과를 제시하고자 한다. 적용예제는 고객의 도착프로세스를 단계형태(phase type)



[그림 8] 평균대기시간의 상한모형

프로세스를 선정하였는데 다음과 같이 포아송 프로세스(λ, s^*, \bar{s}), Erlang 프로세스(단계수, λ, s^*, \bar{s}) 그리고 hyper 지수도착 프로세스($p_1, p_2, \lambda_1, \lambda_2, s^*, \bar{s}, c_1^2$)에 대하여 실험한 결과이다. 여기서 λ 는 도착률을 의미하며, p_1, p_2 는 분지확률을 의미한다.

<표 1> 포아송($\frac{1}{4}, 1.5, 16$)

토큰수(k_1)	5	6	7	8	9	10
분산값 상한	10.95	13.57	14.90	15.55	15.81	15.97
모의실험 분산값	9.38	11.79	12.92	13.46	13.67	13.84
분산값 하한	3.79	7.53	9.79	11.26	12.26	12.97

<표 2> Erlang(2, $\frac{1}{2}, 1.5, 16$)

토큰수(k_1)	5	6	7	8	9	10
분산값 상한	5.98	7.26	7.75	7.91	7.96	7.98
모의실험 분산값	5.51	6.69	7.12	7.26	7.30	7.33
분산값 하한	2.23	4.32	5.49	6.21	6.67	6.99

<표 3> hyper exponential($\frac{1}{3}, \frac{2}{3}, \frac{4}{70}, \frac{4}{7}, 2.5, 28, 3.25$)

토큰수(k_1)	5	6	7	8	9	10
분산값 상한	84.27	111.3	129.4	140.3	147.5	152.4
모의실험 분산값	67.77	94.07	115.4	132.7	138.2	143.7
분산값 하한	30.71	64.31	85.97	99.87	112.7	121.3

실험의 결과는 위 <표 1>~<표 3>에서 보듯이 2노드 대기행렬 네트워크에서 토큰 수가 적을 때는 분산이 상한에 접근되어 있다가 토큰 수가 증가함에 따라 분산이 하한으로 접근함을 알 수 있다. 토큰수를 증가시키는 2노드 대기행렬 네트워크에서 사용자수에 관한 제한이 약화되어 단순한 직렬(tandem) 대기행렬 네트워크와 같은 속성을 갖게 되어 분산값은 하한에 접근됨을 의미하기 때문이다.

동지형 대기행렬 네트워크에 관한 실험은 다소 제약사항이 있는데, 먼저 각 계층의 토큰수를 반영해야 하기 때문에 실험은 어느 한계층을 고정된 토큰수가 주어진 후에 다른 계층의 토큰수를 변화시

켜 가면서 진행되었다. 적용예제 역시 고객의 도착 프로세스를 단계형태(phase type) 프로세스를 선정하였는데 k_1 을 5로 고정한 후에 다음과 같이 포아송 프로세스($\lambda, s^*, \bar{s}_1, \bar{s}_2$), Erlang 프로세스(단계수, $\lambda, s^*, \bar{s}_1, \bar{s}_2$) 그리고 hyper 지수도착 프로세스($p_1, p_2, \lambda_1, \lambda_2, s^*, \bar{s}_1, \bar{s}_2, c_1^2$)에 대하여 실험한 결과이다. \bar{s}_1, \bar{s}_2 는 각각 [그림 4]의 각 계층의 값을 나타낸 것이다.

<표 4> 포아송($\frac{1}{7}, 2.5, 12.5, 14$)

토큰수(k_2)	5	6	7	8	9
분산값 상한	7.79	15.44	17.17	17.77	17.87
모의실험 분산값	7.51	15.33	17.17	17.77	17.87
분산값 하한	4.83	7.79	7.79	7.79	7.79

<표 5> Erlang(2, $\frac{1}{3.5}, 2.5, 12.5, 14$)

토큰수(k_2)	5	6	7	8	9
분산값 상한	5.53	8.71	9.29	9.29	9.29
모의실험 분산값	5.27	8.66	9.29	9.29	9.29
분산값 하한	3.21	5.53	5.53	5.53	5.53

<표 6> hyper exponential($\frac{1}{3}, \frac{2}{3}, \frac{1}{15}, \frac{1}{3}, 2.5, 12.5, 14, 2.31$)

토큰수(k_2)	5	6	7	8	9
분산값 상한	31.65	38.80	39.80	40.46	41.18
모의실험 분산값	14.18	33.63	38.32	40.11	40.84
분산값 하한	10.57	31.65	31.65	31.65	31.65

본 연구에서는 2노드 대기행렬 네트워크와 동지화된 대기행렬 네트워크의 고객이 서비스를 마친 후에 시스템을 벗어나는 이탈사이시간에 대한 분포를 연구하였다. <표 1>에서 <표 3>까지의 실험 결과를 살펴보면 토큰수가 적은 초기에는 분산의 상한에 접근되어 있다가, 차츰 토큰수가 증가함에 따라 분산의 하한으로 접근하는 형태를 유지하고 있다. 이러한 현상은 동지화된 대기행렬 네트워크에

서는 좀 다른 모습으로 결과를 얻었는데, 먼저 k_1 , k_2 의 크기에 매우 민감한 모습을 보인다. 2노드 대기행렬 네트워크와는 달리 등지화된 대기행렬 네트워크는 $k_1 \leq k_2$ 라는 조건이 필수적이므로, <표 6>에서 보듯이 초기에는 분산이 하한에 접근되어 있다가 k_2 가 증가하면서 급격히 상한값에 근접하는 형태를 보이고 있어 2노드 대기행렬 네트워크와는 다른 현상이 발생했다. <표 4>와 <표 5>에서는 처음부터 분산의 상한에 근접하여 결코 하한에는 접근되지 않은 형태로서, 이는 시스템의 분산이나 기타 성능의 척도가 k_2 에 의존적이지 못함을 알 수 있다. 반대로 이탈사이시간의 분산분석을 k_2 를 고정하면서 k_1 을 변화시키는 경우를 실험하였다. k_2 를 7로 고정한 경우의 실험결과는 다음과 같다.

<표 7> 포아손($\frac{1}{7}$, 2.5, 12.5, 14)

토큰수(k_1)	4	5	6	7
분산값 상한	13.24	17.17	19.04	19.64
모의실험 분산값	12.24	17.17	17.65	18.09
분산값 하한	2.31	7.79	12.26	17.73

<표 8> Erlang(2, $\frac{1}{3.5}$, 2.5, 12.5, 14) when $k_2 = 7$

토큰수(k_1)	4	5	6	7
분산값 상한	7.21	9.29	10.03	10.34
모의실험 분산값	6.21	9.29	9.36	9.38
분산값 하한	1.34	5.53	7.30	8.83

<표 9> hyper exponential($\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{15}$, $\frac{1}{3}$, 2.5, 12.5, 14, 2.31)

토큰수(k_1)	4	5	6	7
분산값 상한	25.42	39.80	43.58	44.03
모의실험 분산값	15.30	38.32	40.76	40.97
분산값 하한	5.30	31.65	37.84	39.66

<표 7>에서 <표 9>까지의 실험결과는 k_1 을 고정한 결과와는 정반대 현상으로서 초기에는 분산값 상한에 머물면서 k_1 이 증가함에 따라 분산값 하한

에 접근하는 결과를 얻었다. 이상을 종합하면, k_1 을 고정하는 실험에서는 하한모형에 머물던 분산값이 k_2 가 증가함에 따라 분산값의 상한모형으로 근접하는 결과를 보여주고, 반대로 k_2 를 고정하고 k_1 을 증가시키는 실험에서는 상한모형에 머물던 분산값이 k_1 이 증가함에 따라 분산값의 하한모형으로 근접하는 결과를 보여주고 있다. 등지형 대기행렬 네트워크는 상위계층의 토큰수, k_2 보다는, 하위계층의 토큰수, k_1 에 의존적임을 알 수 있다.

5 결 론

본 연구에서는 등지형 대기행렬 네트워크의 이탈사이시간에 대한 분산분석을 하였는데, 등지형 대기행렬 네트워크의 1계층을 먼저 분석하여 2노드 대기행렬 네트워크의 분석을 통하여 이탈사이시간에 대한 분산의 상한과 하한을 설정한 다음 분석적인 방법을 제시하였다. 이는 대기행렬 네트워크에 대한 분석이 특별한 경우를 제외하고는 정확한 분석이 가능하지 않으며, 따라서 시스템의 성능 척도를 근사적으로 계산하는 방법을 제공하기 위함이다. 특히 등지화된 대기행렬 네트워크는 $k_1 \leq k_2$ 라는 조건이 필수적이므로 k_1 을 고정시키고 k_2 를 증감시키는 방법에서는 2노드 대기행렬 네트워크와는 달리 상한에 급격히 접근되는 결과를 보였다. 이와 반대로 k_2 를 고정시키고 k_1 을 증감시키는 경우는 2노드 대기행렬 네트워크와 유사한 양상을 보이는 것으로 조사되었다. 이러한 결과를 종합하면 k_2 보다는 k_1 이 시스템에 더 큰 영향을 미치고 있음을 알 수 있다. 다중계층 구조의 대기행렬 네트워크의 분석은 정확성을 기하는데 많은 어려움이 있으나 상한 값과 하한 값을 구하여 이를 적절히 조정하여 보다 정확한 근사치를 구할 수 있다.

참 고 문 헌

[1] Fdida, S., H. Perros, and A. Wilk, "Semaphore Queue : modeling multilayered window

- flow control mechanism," *IEEE Trans. on Comm.*, Vol.83, No.3(1990), pp.309-317.
- [2] Gühr, O. and P. Kuehn, "Comparison of communication services with connection-oriented and connectionless data transmission," *Processing of the International Seminar on Computer Networking and Performance Evaluation*, Tokyo Japan, (1985) pp.431-444.
- [3] Liberopoulos, G. and Y. Dallery, "A unified framework for pull control mechanism in multi-stage manufacturing systems," *Annals of OR*, Vol.93(2000), pp.325-355.
- [4] Marchal, W.G., "Some simpler bounds on the mean queueing time," *Operations Research*, Vol.26, No.6(1978), pp.1083-1088.
- [5] Pennotti, M. and M. Schwartz, "Congestion control in store and forward tandem link," *IEEE Trans. on Comm.*, Vol.23(1975), pp.1434-1443.
- [6] Reiser, M., "A queueing network analysis of computer communication networks with window flow control," *IEEE Trans. on Comm.*, Vol.27(1979), pp.1199-1209.
- [7] Rhee, Y. and H. Perros, "Analysis of an Open Tandem Queueing Network with Population Constraint and Constant Service Times," *European J. Of Operations Research*, Vol.92(1996), pp.99-111.
- [8] Rhee, Y. and H. Perros, "On the Mean Waiting Time of a Population Constrained Open Tandem Queueing Network with Constant Service Times," *IIE Trans. on Design and Manufacturing*, Vol.30-10(1998), pp.973-979.
- [9] Rhee, Y., "Some notes on the reduction of the network dimensionality in the nested open queueing network," *European J. Of Operations Research*, Vol.174(2006), pp.124-131.
- [10] Schwartz, M., "Performance analysis of the SNA virtual routing pacing control," *IEEE Trans. on Comm.*, Vol.30(1982), pp.172-184.
- [11] Thomasian, A. and P. Bay, "Analysis of Queueing network models with population size constraint and delayed blocked customers," *ACM SIGMETRICS conference*, (1984), pp. 202-216.
- [12] Whitt, W., "The Queueing Network Analyzer," *The Bell System Tech. J.*, Vol.62-9(1983), pp. 2779-2815.
- [13] Rhee, Y., "Ph.D. Thesis," North Carolina State University, 1994.