대한산업공학회/한국경영과학회 2000 춘계공동학술대회 논문집. 2000년4월21-22일. 경남대학교
Session OB6.4

다중클래스 대기망의 안정성 향상을 위한 방법

Methods to stabilize multiclass queueing networks

윤 복 식
홍익대학교 기초과학과

Abstract

When there are several classes of customers demanding service times with different distributions at some stations of a queueing network, the stability problem becomes suddenly complicated compared with the single class case. Recently many researchers had tried to find some kind of stability conditions for multiclass queueing networks, but did not get significant results except in very limited 2-station cases. In this study, we try to develop some dynamic control techniques which can guarantee the stability under the nominal traffic condition. Our approach includes the randomization method and the leaky bucket control scheme. Also, we mention other possibilities such as the discrete-review approach and the generalized round-robin technique. Both theoretical and experimental results will be presented.

1. 서론

하나의 서비스 노드에 서비스 특성이 상이한 여러 종류의 고객들이 존재하는 다중클래스 대기망에서는 안정성을 보장해 주는 조건이 갑자기 복잡해진다. 비교적 단순한 구조의 재진입망의 경우에도 기존에 흔히 통용되던 통상적인 안정조건인 각 노드에서의 통합 부하율(workload)들이 모두 1보다 작은($\rho_i < 1$) 것으로 충분하지 않을 수 있다는 것이 밝혀진 후에 (Kumar and Seidman(1990), Lu and Kumar(1991), Seidman(1994), Bramson(1994a,b)) 안정성 조건을 찾기 위한 많은 연구가 수행되어 왔다 (Chen(1995), Dai(1995, 1996), Dai and Meyn(1995), Bertsimas et al(1996)).

본 연구는 다중클래스 망에서의 안정성 상실에 이르는 시스템의 동적 양태를 파악하고 유체흐름 모형을 이용한 접근 방법을 이용하여 안정성을 향상시킬 수 있는 방안을 연구하는 것을 주목적으로 한다. 물론 기존의 연구(Dai and Weiss(1996))에서 알 수 있듯이 매우 특수한 망에 대해서도 안정성의 일반적인 충분조건을 확립하기가 쉽지 않은데 다양한 서비스 규율에 따른 안정성 변화도 검토할 수 있도록 대기망 방정식을 만들고 유체모형으로 근사화 하여 접근하는 방법을 통해 어느 정도 접근이 가능하다. 기존의 재진입 망에서 Dai(1995)등에 시도된 방법과 유사하게 유체근사화(fluid approximation)와 Lyapunov 함수(Down and Meyn(1994))를 통한 선형계획 모형을 만들어 접근할 수도 있는데 본 연구에서는 상세한 이론적인 기술은 피하고 실용적으로 안정성을 파악하고 향상시키는 방법에 주안점을 둔다.

2. 다중클래스 대기망의 표현

2.1 다중클래스 대기망

각각 무한 대기공간을 갖는 단일 서버를 가진 J개의 서비스노드로 구성된 열린 대기망(open queueing network)을 고려하자. 각 노드를 방문하는 고객들은 모두 동일한 분포의 서비스 요구시간을 갖는 단일클래스 망과는 고객클래스 별로 다른 서비스 분포를 가질 수 있는데 총 $K(\geq J)$개의 고객클래스로 구성되어 있다고 가정하고 $\Omega(j)$를 노드 j에서 서비스 받는 고객클래스들의 집합이라고 하자. 각 노드는 여러 개의 고객클래스를 가질 수 있으나 각 고객클래스는 하나의 노드에서만 서비스를 받게 되는데 클래스 k고객은 $s(k)$에서 서비스를 받는다고 하자.

클래스 별로 외부로부터의 도착과정의 도착간격과 서비스 시간이 각각 평균이 $1/\alpha_k$, m_k이고 변동계수(coefficient of variation)가 c_{ek}, c_{sk}인 동일분포를 하고 다른 클래스와는 상호독립이라고 가정하자. 서비스를 끝낸 후에 클래스 k고객은 확률 P_{kl}로 클래스 l고객으로 바뀐다고 하자. 이때 행렬 $P=(P_{kl})_{K\times K}$을 경로이동행렬(routing matrix)이라고 부르는데 망 외부로 빠져나갈 수도 있으므로 P는 잠정적(transient)이 되어

$$Q=I+P^T+(P^T)^2+\ldots=(I-P^T)^{-1}$$

이 유한한 값으로 존재하게 된다.

2.2 통상적인 트래픽 조건

열린 대기망의 경우 클래스 k로의 명목상의 총 도착률(nominal total arrival rate) λ_k는 다음과 같은 트래픽방정식에 의해 구한다.

$$\lambda_k = \alpha_k + \sum_{l=1}^{K} \lambda_l P_{lk}, \qquad k=1,\ldots,K$$

위의 방정식은 $\lambda = \alpha + P^T\lambda$의 벡터형태로 나타낼 수 있는데 P가 잠정적이므로 유일한 해 $\lambda = Q\alpha$를 갖게 된다. 이것과 평균 서비스 시간 벡터 m을 이용하여 서버 j에서의 트래픽밀도(traffic intensity)를

$$\rho_j = \sum_{k \in \Omega(j)} m_k \lambda_k, \qquad j=1,\ldots,J$$

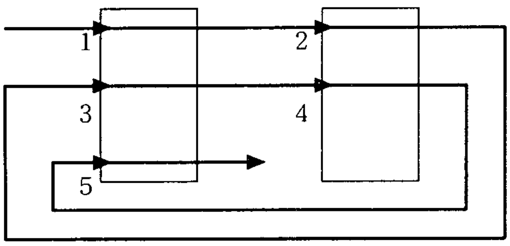
로 정의할 수 있는데 이것은 서버 j의 이용률(server utilization) 또는 작업중인 시간의 비율로 해석할 수 있다. 대기망이 안정적이기 위해서는 기본적으로

$$\rho_j < 1, \quad j=1,\ldots,J$$

의 조건이 필요한데 이를 통상적인 트래픽 조건(usual traffic condition)이라고 부르자.

3. 특수한 경우

3.1 간단한 2-노드 망의 예



[그림 1] 2노드 5 클래스 재진입망

그림 1에 주어진 2-노드 5-클래스 망의 경우 외부에서 클래스 1으로의 도착률을 1 이라고 하고 $m_1 = m_3 = m_4 = 0.1, \quad m_2 = m_5 = 0.6$이라고 하자. 이때 각 노드에서의 트래픽 밀도는 $\rho_1 = 0.8, \ \rho_2 = 0.7$이 되어 모두 1보다 작게 된다. 이제 한 노드에 클래스 a, b, c가 있을 때 서비스 규칙의 표기를 우선순위가 높은 클래스부터 집합기호 안에 나열하기로 하는데 이때 순위를 정할 필요가 없는 클래스는 쓰지 않기로 한다. 예를 들어 노드 1에서 우선순위가 5-3-1 순으로 설정되었다면 서비스 규칙은 {5, 3, 1}로 표기되고, 5와 1 만의 관계만 중요하면 {5,1}포 표기된다. 이제 규칙 {{5, 3, 1}, {2, 4}} 하에서 시뮬레이션을 해보면 (이때 서비스 시간은 모두 지수 분포) 대기자의 수가 폭증하여 망이 불안정해 지는 것을 관찰할 수 있다. 또한 클래스 2와 클래스 5의 고객은 동시에 서비스를 받지 않는다는 것을 관찰할 수 있다. Dai and VandeVate(1996b)는 preemptive의 가정 하에서 규칙 {{5,1},{2,4}}를 적용했을 때, 2와 5의 고객이 동시에 서비스되지 않음을 보이고, $m_2 + m_5 > 1$이 되면 시스템이 불안정해진다는 사실을 증명한 바 있으나 실제로는 preemptive의 가정이 필요 없을 것이라고 시뮬레이션 결과를 통해 짐작할 수 있다. 이 망의 안정성에 대해 다음과 같은 사실을 알 수 있다.

(1) 통상적인 트래픽 조건 외에 $m_2 + m_5 < 1$ (가상적인 노드조건)이 첨가되면 모든 비유휴(non-idling) 서비스 규율 하에서 안정적이 된다. (Dai and VandeVate(1996b, 1997a))

(2) 클래스 1에 최우선권이 주어지면 통상적인 트래픽 조건하에서 안정적이다. (Dai and VandeVate(1996b))

(3) 모든 노드에서 LBFS(last-buffer-first-service) 규칙을 사용하면 모든 재진입 망은 안정적이다.(Dai and Weiss(1996))

(4) {{3,5,1}, {4,2}}, {{5,1,3},{4,2}}도 통상적인 트래픽 조건하에서 안정적이다. (Chen and Zhang(1997))

(5) {{1,5}, *}, {*, {4,2}}는 통상적인 트래픽 조건하에서 안정적이다.

3.2 Kelly 형의 FIFO 유체모형

각 노드에서 모든 클래스의 평균 서비스시간이 동일할 때를 Kelly형의 망이라고 부른다. 이 경우 통상적 트래픽 조건이 만족되면 안정적이다. 이는 엔트로피 형태의 Lyapunov 함수를 사용하여 증명할 수 있다. 이를 개략적으로 설명하면, $h(x) = x \log x, \ x \geq 0$라고 하고

$$f(t) = \sum_k \int_t^{t+W_j(t)} \lambda_k h(\dot{D}_k(s)/\lambda_k)\, ds$$

로 정의하면 주어진 조건($\rho < 1$, FIFO)하에서

$W(t) = 0$ 일 때, $f(t) = 0$이고

$W(t) > 0$ 일 때 $f(t) > 0$

이 되는 것을 관찰할 수 있고 또한 정규점에서 $f'(t) \leq 0$이 되는 것을 보일 수 있다. 이 사실로부터 t가 충분히 커지면 $f(t)$=0이 됨을 기대할 수 있고 따라서 t가 충분히 커지면 $W(t)$=0이 되어 유체망은 안정적이 될 것이다.

4. 안정성 향상의 방법

4.1 제어노드의 설치

클래스 수가 K이고 노드가 J개인 다중클래스 대기망이 통상적인 트래픽 조건을 만족하고 있다고 하자. 이때 아직 망의 안정성을 보장할 수가 없는데 다음과 같이 각 클래스로의 서비스를 약간씩 지연시키면서 정규적으로 만드는 장치를 두어 안정성을 보장할 수가 있다. 모든 클래스 k에 대해 클래스 앞에 제어노드 $J+k$를 설치하여 원래 클래스 k로 들어오던 모든 고객을 클래스 $K+k$로 분류하여 서비스하도록 한다. 이때 평균 서비스 시

간을 $\left(\frac{1+\rho_{s(k)}}{2}\right)\frac{1}{\lambda_k}$로 설정하고 여기서 서비스를 받은 후에 항상 클래스 k로 가도록 한다. 클래스 k고객은 서비스를 받은 후에 원래 주어진 경로이동확률에 따라 움직이는데 이때 확률 P_{kl}로 (클래스 l 대신에) 클래스 $K+l$로 가도록 한다. 이렇게 만든 망은 $2K$의 클래스와 $J+K$의 노드를 갖게 되는데 이 경우 각 클래스의 명목상의 총 도착률 벡터 λ는 $\lambda_{K+k}=\lambda_k=\alpha_k$ 관계를 갖게 되고 모든 노드에서 통상적인 트래픽 조건을 만족하게 된다. 더욱이 이 망은 이 조건하에서 항상 안정적이 됨을 보일 수 있다.

실제로 이 방식은 각 클래스에 입력측 대기소 외에 출력 측 대기소를 의 구현할 수도 있다. 출력측 대기소에서 우선 다음에 갈 클래스 k를 경로이동 확률에 따라 확률적으로 결정한 후 k로 갈 고객들을 모아두고 $\left(\frac{1+\rho_{s(k)}}{2}\right)\frac{1}{\lambda_k}$만큼의 간격을 두고 하나씩 출발하도록 한다.

4.2 라운드 로빈(round-robin) 방식

각 서비스 노드에서 서비스 규율의 제어가 가능하면 안정성을 높이는 방법을 다양하게 고안할 수 있을 것이다. 라운드 로빈 방식도 그중 하나인데, 우선 각 클래스를 적당하게 순위를 정하여 그 순위에 따라 클래스 번호를 재배열 한 후에 양의 벡터 $\beta=(\beta_1,\ldots,\beta_K)$설정한다. 각 노드에서는 그 노드에 할당된 클래스들을 순위에 따라 서비스하는데 k 클래스라면 남아있는 고객이 없을 때까지 연이어 최대 β_k개의 k 클래스 고객들을 서비스해 주고 다음 순위 클래스로 넘어간다. 물론 해당 순위의 고객이 저장소에 없으면 그대로 통과한다. 이 방식에서 모든 클래스 k에 대해

$$\frac{\beta_k}{\sum_{l\in\Omega(j)}\beta_l m_l}>\lambda_k$$

의 조건을 만족하도록 β를 설정하면 대기망의 유체모형이 안정되는 것을 보일 수 있다.

4.3 랜덤화

3장에서 Kelly 형의 FIFO 망의 안정성을 알 수 있는데 이것의 특징은 각 노드에서 클래스들의 평균 서비스 시간이 모두 같다는 것이다. 만약 클래스 별로 평균 서비스 시간이 다르다면 각 서비스 시작 시점에서 대기 고객 중에서 확률적으로 선택하여 서비스를 해 주면 전체적으로 서비스 시간이 혼합(mixing)되는 효과가 있으므로 각 노드에서 서비스 받는 전체 고객을 동일한 서비스 시간 분포를 갖는 고객들로 취급할 수 있을 것이다. 만약 이때 노드 j에서 각 클래스의 서비스 선택 확률을 $\lambda_l/\sum_{l\in\Omega(j)}\lambda$로 설정하면 평균 서비스 시간은 $\sum_{l\in\Omega(j)}\lambda_l m_l/\sum_{l\in\Omega(j)}\lambda_l$이 되어 원래 망에 대한 통상적인 트래픽 조건 $\sum_{l\in\Omega(j)}\lambda_l m_l<1$ 하에서 Kelly 망으로 취급했을 경우의 트래픽 조건을 만족하게 되어 망의 안정성을 기대할 수 있게 된다.

4.4 기타 동적(Dynamic) 규율

망의 상태를 적당한 간격을 두고 검토하여 검토 시점에서의 대기열의 길이를 바탕으로 다음 검토 시점까지의 망의 운용계획을 수립하는 방식으로 망을 제어하는 Bigstep 방식(Harrison(1996))과 이것을 일반화한 이산적 검토 정책(discrete review policy, Maglaras(1998))을 사용하면 안정성을 보장할 수 있는 것으로 보고되고 있다.

5. 결론

다중클래스 대기망은 하나의 서비스 노드에서 다른 노드와의 트래픽의 연관성이 상이한 여러 종류의 고객을 서비스 해 주어야 하는데 따라 발생하는 고객 클래스간의 상호 블러킹 작용 때문에 예기치 못했던 대기량의 폭증 현상이 발생한다. 본 연구에서는 다중클래스 대기망이 안정적으로 작동할 수 있도록 보장해 주는 방법들에 주안점을 두었다. 기본적으로 정적인 서비스 규율 하에서는 안정성의 조건을 간편하게 찾아내기는 어렵다. 본 연구에서는 약간의 추가적인 지연을 허용하여 트래픽의 이동을 정규화 하는 방식(제어노드 첨가)이나 서비스 규율을 정교하게 하여 인위적으로 트래픽 처리의 균형을 맞추는 방식(라운드 로빈), 또는 랜덤하게 우선권을 배정하여 트래픽을 균질화 시키는 방식(랜덤화)이 제시되었고 Bigstep과 같은 동적인 제어 방법도 언급되었다. 유체모형과 대기망과의 관계는 생략하였고, 향후 보다 실험적인 연구가 뒤따를 것이다.

[참고 문헌]

Banks, J. and J.G. Dai, Simulation studies of multiclass queueing networks, *IIE Transactions*, 29, 213-219, 1997.

Bertsimas, D., D. Gamarnik and J.N. Tsitsiklis, Stability conditions for multiclass fluid queueing networks, *IEEE Transactions on Automatic Control*, 41, 1996.

Bramson, M., Instability of FIFO queueing networks, *Annals of Applied Probability*, 4, 414-431, 1994a.

Bramson, M., Instability of FIFO queueing networks with quick service times, *Annals of Applied Probability*, 4, 693-718, 1994b.

Chen, H., Fluid approximations and stability of multiclass queueing networks I: Work-conserving disciplines, *Annals of Applied Probability*, 5, 637-665, 1995.

Chen, H. and H. Zhang, Stability of multiclass queueing networks under FIFO service discipline, *Mathematics of Operations Research*, 22, 691-725, 1997.

Dai, J.G., On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models, *Annals of Applied Probability*, 5, 49-77, 1995.

Dai, J.G., A fluid-limit model criterion for instability of multiclass queueing networks, *Annals of Applied Probability*, 6, 751-757, 1996.

Dai, J.G. and S.P. Meyn, Stability and convergence of moments for multiclass queueing networks via fluid limit models, *IEEE Transactions on Automatic Controls*, 40 , 1889-1904, 1995.

Dai, J.G. and J.H. Vande Vate, Global stability of two-station queueing networks, in *Stochastic Networks: Stability and Rare Events*(K.S.P. Glasserman and D. Yao Eds.), 1-26, New York, Springer-Verlag, 1996a.

Dai, J.G. and J. Vande Vate, Virtual stations and the capacity of two-station fluid networks, Preprint, 1996b.

Dai, J.G. and J. Vande Vate, The stability of two-station fluid networks, Preprint, 1997a.

Dai, J.G. and J. Vande Vate, Stability of a three-station fluid network, Preprint, 1997b.

Dai, J.G. and G. Weiss, "Stability and Instability of Fluid Models for Re-Entrant Lines", *Mathematics of Operations Research*, 21, 115-134 (1996).

Down, D. and S. Meyn, Piecewise linear test functions for stability of queueing networks, *Proceedings of the 33rd Conference on Decision and Control*, 2069-74, 1994.

Harrison, J.M., The BIGSTEP approach to flow management in stochastic processing networks, *Stochastic Networks: Theory and Applications*(F. Kelly et al. Eds), 57-90, Oxford university press, 1996.

Kumar, P.R. and S.P. Meyn, Stability of queueing networks and scheduling policies, *IEEE Transactions on Automatic Control,* 40, 251-260, 1995.

Kumar, P.R. and S.P. Meyn, Duality and linear programs for stability and performance analysis of queueing networks and scheduling policies, *IEEE Transactions on Automatic Control,* 41, 4-17, 1996.

Kumar, S. and P.R. Kumar, Performance bounds for queueing networks and scheduling policies, *IEEE Transactions on Automatic Control,* 39, 1600-11, 1994.

Kumar, P.R. and T.I. Seidman, Dynamic Instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems, *IEEE Transactions on Automatic Control,* 35, 285-298, 1990.

Maglaras, C, Dynamic scheduling in multiclass queueing networks: Stability under discrete review policies, *Queueing Systems: Theory and Applications,* 1998.(submitted)

Lu, S.H. and P.R. Kumar, Distributed scheduling based on due dates and buffer priorities, *IEEE Transactions on Automatic Control,* 36, 1406-16, 1991.

Seidman, T.I., 'First come, first served' can be unstable!, *IEEE Transactions on Automatic Control,* 39, 2166-71, 1994.