

우선순위 방식 스케줄링에서의 가격선택 문제의 분석

(Analysis of the Price-Selection Problem in Priority-based Scheduling)

박 선 주 [†]

(Sunju Park)

요약 이 논문은 QoS (Quality of Service) 네트워크 서비스를 위한 우선순위 방식의 스케줄링에서, 각 서비스 레벨의 가격을 정하는 문제를 분석한다. 특히 본 논문에서는 균형 분석 (equilibrium analysis)에 근거한 가격정책의 문제점을 파악하는 것에 중점을 둔다. 균형분석은 다음과 같은 두 가정하에 이루어진다. 첫째, 각각의 사용자들이 시스템 전체에 미치는 영향은 극히 미약하여 무시할 수 있다. 둘째, 사용자들은 전체 시스템 상태에 대한 최신의 정보를 알고 있다. 그러나 이러한 가정들은 실제 상황을 정확하게 반영하지 못하는 경우가 종종 있고, 따라서 균형분석에 의한 가격의 결정은 문제점이 있다고 본다. 본 논문에서는 시스템 작동상황을 분석하기 위해 '동적모델'을 개발하고, 이를 이용하여 현존하는 균형분석 방법의 정확성을 평가하였다. 연구결과에 의하면 균형분석은 실제적인 환경에서의 시스템의 작동 현황을 정확하게 반영하지 못하는 경우가 많은 것으로 나타났다.

키워드 : 네트워크 서비스 가격정책, 우선순위방식의 서비스 네트워크, 균형분석

Abstract This paper analyzes the price-selection problem under priority-based scheduling for QoS (Quality of Service) network services, i.e., how to determine the price associated with each service level. In particular, we focus on the problems with the pricing mechanism based on equilibrium analysis. We claim that the assumptions needed to produce equilibrium may not hold in some important environments. Specifically, (a) the individual user's impact on the system is not infinitesimal and (b) users do not always have up-to-date global system-status knowledge crucial for optimal user decisions required for equilibrium. These may make the equilibrium models inaccurate in realistic environments. We examine the accuracy of some existing equilibrium methods by using a dynamic model that we have developed for system behavior analysis. The analysis indicates that equilibrium methods fail to model accurately the system behavior in some realistic environments.

Key words : network service pricing, priority service networks, equilibrium analysis.

1. 서론

네트워크 서비스의 가격을 어떻게 책정할지는 인터넷의 상업화 초기부터 계속하여 논란의 대상이 되어왔다. 역사적으로 보면, 사용자가 사용량에 무관하게 일정금액을 내는 일률 요금 정책(flat-fee scheme)이 인터넷 액세스를 위한 가격정책으로 쓰여져왔다(Odlyzko 2001). 그러나 최근들어 3G 이동전화 서비스나 차세대 인터넷 등의 영역에서는 일률요금정책과는 다른 방식의 가격정책을 쓰려는 움직임이 일고 있다. 예를 들어 대역폭 자

원을 낭비할 경우가 많은 일률요금정책을 대역폭이 부족하거나 비싼 이동전화 서비스에 그대로 적용하는 것은 무리가 있다. 또한 차세대 인터넷에서는 서비스 차별화를 필요로 하는 QoS(Quality of Service) 응용분야가 등장하고 있다. 이러한 속도에 민감한 QoS 서비스들과 이메일이나 화일 전송등의 전형적인 최선형(best-effort) 서비스들의 공존을 위해서는, 네트워크가 다른 질의 서비스에 해당하는 트래픽을 다르게 처리하는 능력뿐만 아니라 각 서비스들의 가격 또한 다르게 책정할 수 있어야 한다.

이러한 필요성 때문에, 사용자가 가격에 민감하다는 가정하에 서비스 제공자가 가격을 통해 사용자의 서비스 레벨 선택 결정에 영향을 미쳐 혼잡통제(congestion control)를 하는 여러 다른 종류의 가격정책이¹⁾ 제안되

· 본 연구는 2005학년도 연세대학교 학술연구비와 연세대학교 경영연구소의 부분적인 지원에 의해 이루어진 것임

[†] 통신회원 : 연세대학교 경영학과 교수
boxenju@yonsei.ac.kr

논문접수 : 2005년 4월 7일
심사완료 : 2006년 1월 9일

어왔다. 이 논문은 QoS를 지원하는 IETF 차별화 서비스에 가장 적합한 우선순위 방식의 스케줄링과 가격정책을 연구대상으로 선택하여, 우선순위 가격정책에서 가격선택의 방법으로 사용하는 균형분석의 문제점을 알아보는 것을 목표로 한다(Blake 1998).

우선순위 가격정책하에서 서비스 제공자는 사용자에 게 {서비스 레벨, 가격}의 쌍들을 제공하고, 자신의 트래픽의 값어치를 가장 잘 알고 있는 사용자들이 그 트래픽을 수행하기 위해 필요한 서비스 레벨과 가격을 선택한다. 따라서 서비스 제공자가 각 서비스 레벨의 가격을 잘 정하는 것은 중요한 문제이며, 이를 위해 시스템의 동작(즉 주어진 가격하에서 사용자가 어떤 레벨을 선택하고 그들의 선택이 시스템의 상황에 어떤 영향을 미치는지 등등)을 모델링하여야 한다(Cocchi 1991, Gupta 1997).

사용자들의 서비스 레벨 선택은 (1) 트래픽에 대한 유틸리티, (2) 각 서비스 레벨의 가격, (3) 각 서비스 레벨의 혼잡도에 영향을 받는다. 이때 혼잡도는 시스템을 사용하는 모든 사용자들의 선택의 결과이다. 예를 들어, 낮은 레벨의 서비스가 혼잡해지면 몇몇의 사용자들은 높은 가격을 내서라도 높은 레벨의 서비스를 받고자 할 수 있다. 이러한 선택의 결과는 높은 서비스 레벨의 혼잡도를 가중시켜 그 서비스 질을 떨어뜨리며, 어떤 사용자들은 가격에 비한 효용도가 떨어진다고 생각하여 낮은 레벨의 서비스를 선택할 수도 있다. 이러한 시스템의 다이내믹스는 종종 균형분석(equilibrium analysis)을 통해 모형화되어왔다(Hassin and Haviv 2003).

그러나 본 논문은 균형분석의 모델에 의거한 가격선택은 문제점이 있다고 본다. 왜냐하면 정적 균형점(static equilibrium)을 생성하기 위해 균형분석이 사용하는 가정들은 실제 시스템의 상황을 제대로 반영하지 못하기 때문이다. 특히 본 논문은 균형분석의 두가지 가정의 문제점에 초점을 맞추고 있다(Mendelson and Whang 1990, Mendelson 1985, Mandjes 2003). 첫째, 현존하는 많은 네트워크에서(특히 라스트 마일 인터넷 링크에서) 각 사용자가 시스템에 미치는 영향은 무시할 수 없다. 이는 균형분석에서 사용하는 '사용자가 무한히 많아서 각 사용자가 시스템에 미치는 영향은 무시할 수 있다'는 가정에 위배된다. 둘째, 현존하는 인터넷 프로토콜은 사용자들에게 각 서비스 레벨의 현재 큐의 길이나 트래픽의 도착률등을 효과적으로 알려주는 방법이 없다.

따라서 '사용자들이 시스템의 상태에 대한 가장 최신의 정보를 가지고 있다'는 균형분석의 가정은 현실에 위배된다. 결론적으로, 현실에서는 '유한한' 수의 사용자들이 시스템 상태에 대한 '정확하지 않은' 정보를 가지고 sub-optimal한 결정을 내리기 때문에 균형분석의 가정이 맞지 않게 되며, 따라서 안정된 균형점이 존재하지 않을 수 있다.

이러한 주장을 뒷받침하기 위해, 본 논문은 네트워크 시스템의 작동상황을 분석하기 위한 '동적모델'을 개발하고 현존하는 균형분석 방법의 문제점을 실제적인 네트워크 환경하에서 연구분석하였다. 균형분석과 달리, 동적모델은 큐를 명시적으로 모델화하고 사용자들이 각 큐에 대한 정보를 약간의 지연이 있는 후 알 수 있다고 가정한다(Balachandran and Srinidhi 1987, Hassin 1986). 실험 결과에 의하면 균형분석은 시스템의 행동양식을 제대로 반영하지 못 할 경우가 있다. 오래된 정보가 네트워크 서비스 가격정책에 미치는 영향은 이미 연구된 적이 있으나, 이들은 서비스 제공자가 실시간에 가격을 바꿀 수 있는 경우를, 본 논문은 한 번 정해진 가격들이 바뀌지 않는 경우를 연구하고 있다(Low and Lapsley, 1999). TCP 혼잡 정보의 지연으로 인한 효과에 관한 연구도 본 논문의 연구와 연계성이 있다고 볼 수 있다(Kelly 2001).

이 논문은 다음과 같이 구성되어있다. 제2절은 우선순위 방식의 가격정책 문제를 분석하기 위한 프레임워크(framework)를 설명한다. 제3절은 왜 균형분석이 실제 시스템을 제대로 모델링하지 못하는지를 살펴본다. 제4절은 동적모델을 소개한다. 제5절은 현존하는 3개의 균형분석방법을 동적 모델을 통해 분석한다. 마지막 절에서 결론을 내린다.

2. 우선순위 방식 서비스에 대한 가격정책 프레임워크

우선순위 방식 서비스를 제공하는 네트워크 시스템의 프레임워크를 정의한다. 사용자들에게 I 개의 서비스 레벨을 제공하는 우선순위 방식의 네트워크를 생각해 보자. 여기서 가장 좋은 (빠른) 서비스 레벨을 레벨 1, 두 번째로 좋은 서비스 레벨을 레벨 2, ..., 등등으로 정의 하자. 각 서비스 레벨 i 를 일정 데이터 단위만큼(예를 들어, 1 KB) 사용하는 금액은 P_i 로 정의한다.

서비스 제공자는 이득을 최대화한다든지 또는 전체 시스템의 유용성을 향상시킨다든지 하는, 주어진 목적을 만족시키는 I 개의 가격 벡터, $\{P_i\}$,를 결정하여야 한다. 이를 가격선택 문제라고 부른다. 본 연구의 프레임워크에서는 한 번 정해진 서비스 레벨의 가격들은 변하지 않는다. (즉, 이 논문은 서비스 레벨의 가격을 실시간에

1) 예를 들어, priority pricing(Cocchi 1991, Cocchi 1993), smart-market pricing(Mackie-Mason 1995), Paris-metro pricing(Odlyzko 1997), responsive pricing(Mackie-Mason 1997), edge pricing(Shenker 1996), proportional fairness pricing(Kelly 1997, Kelly 1998) 등등의 가격정책이 있다.

변화시키는 문제는 다루지 않는다.)

여러 사용자들은 데이터의 크기나 필요로 하는 최소 성능등이 다른 데이터를 만들어 낸다. 이러한 데이터들을 '트래픽'이라고 정의하자. 예를 들어 이메일 메시지를 보내는 것이나 웹페이지를 액세스하는 것등이 하나의 트래픽이라고 볼 수 있다. 본 논문에서는 j -타입 트래픽의 크기는 평균값 C_j 를 갖는 지수 분포(exponential distribution)를 따른다고 가정한다. 각 트래픽의 유용성은 유틸리티 함수에 의해 표현된다. 일반적으로 사용자들은 트래픽을 빨리 보내고 받기를 원하므로 대부분의 경우 전송의 지연정도가 성능의 척도가 된다. 따라서 이 논문은 트래픽 종료까지의 지연(traffic completion delay)을 성능 척도로 사용한다. j -타입 트래픽의 종료 지연시간이 T 라 할 때 그 트래픽의 유틸리티 함수를 $U_j(T)$ 로 정의한다. 이때 $U_j(T)$ 는 T 에 따른 단조감소 함수이다.

네트워크 서비스는 Non-Preemptive Strict Priority Scheduling(NP-SPS) 방식을 사용한다. SPS란 높은 레벨에서 서비스를 기다리는 트래픽들이 없는 경우에만 낮은 레벨의 트래픽들을 처리하는 것을 말한다. (즉 낮은 레벨의 트래픽은 높은 레벨의 트래픽을 다 처리한 후에야 처리된다.) NP는 새롭게 도착하는 높은 레벨의 트래픽이 현재 처리되고 있는 낮은 레벨의 트래픽을 중단(interrupt)시키지 않는 것을 의미한다. 같은 레벨의 트래픽은 FIFO 방식으로 처리된다.

사용자는 각 트래픽에 적합한 서비스 레벨을 선택하고 그에 해당하는 가격을 지불한다. 물론 개인의 결정은 현 시스템 상황에 영향을 받으므로, 같은 종류의 트래픽들도 그때그때의 상황에 따라 다른 서비스 레벨을 선택할 수 있다. 이때 사용자는 '이기적인' 결정을 한다. 즉, 서비스 레벨을 선택할 때 사용자는 시스템의 목적은 전혀 고려치 않고 오직 자신의 이익만을 고려한다. 이렇게 정해진 서비스 레벨은 일단 시스템에 들어온 후에는 바꿀 수 없다.

각 서비스 레벨에 도착하는 트래픽은 포아송 분포를 따른다고 가정한다. 서비스 레벨 i 에서의 트래픽 도착률을 λ_i 라 하면 $\{\lambda_i\}$ 는 트래픽도착률 벡터를 이룬다. 특수 경우로 λ_0 는 시스템에 도착하지 않은 트래픽들의 도착률을 나타낸다고 정의하자. (즉 λ_0 는 사용자가 전송을 포기한 트래픽들을 표현한다.) 이때 총 트래픽도착률 λ 는 각 서비스레벨의 도착률을 합한 것과 같다(즉, $\sum_{i=0}^I \lambda_i = \lambda$). 한편 네트워크의 트래픽처리율은 μ 로 표현한다. 대부분의 경우 μ 는 서비스 제공자의 상위 네트워크와의 연결링크(backhaul link)의 용량에 의해 결정된다. 시스템의 원활한 작동을 위해 실질적인 트래픽도착률($\lambda - \lambda_0$)은 μ 보다 작다고 가정한다.

3. 네트워크 시스템의 수렴(convergence)과 진동(oscillation)

균형분석은 우선순위 방식 네트워크의 가격을 결정하는데 많이 사용되어왔다. 균형분석은 각 사용자들이 자신의 이익을 최대화하는 이기적인 결정을 내려도 각 트래픽의 도착률이 어떤 일정한 값으로 수렴하여 균형상태(equilibrium state)에 도달한다는 가정을 바탕으로 하고 있다. 이러한 안정된 균형상태는 여러개가 존재할 수 있으며 그중 어느 균형상태로 수렴하는지는 사용되고 있는 가격 벡터에 의해 결정된다. 따라서 여러 균형상태중 정해진 시스템의 목표(예를 들어 이익의 최대화)를 극대화하는 균형상태를 식별해내고, 그러한 균형상태를 초래할 수 있는 가격벡터를 구하는 것이 그 목표이다.

그러나 균형분석은 미리 정해진 조건들이 만족되는 경우에만 적합하고, 이러한 가정들은 현실적인 환경에서는 불가능한 경우가 종종 있다. 즉, 우리는 시스템이 안정된 균형상태에 도달하는 것이 아니라 여러 상태를 진동할 수도 있다는 점을 지적한다.

균형점이 존재하기 위해 필요한 기본 가정중의 하나는, 각 트래픽이 전체 시스템에 미치는 영향을 무시할 수 있어야 한다는 것이다. 이는 현재 시스템내에 있는 트래픽의 갯수가 무한히 많아야함을 의미한다. 그러나 이러한 가정은 라스트 마일 인터넷 액세스를 사용자에게 제공하는 에지 네트워크(edge network)²⁾에 적용되지 않는다. 전형적으로 볼때, 에지 네트워크에서 활동하는 사용자의 수는 상대적으로 적다. 어떤 네트워크는 동시에 존재할 수 있는 사용자의 수에 물리적인 제한이 있기도 하다. 예를 들어 무선 데이터 네트워크의 가장 발전된 형태인 3G CDMA 네트워크는 각 셀에서 (보다 정확히는 각 캐리어에서) 동시에 존재할 수 있는 세션의 갯수가 64개로 제한되어 있다(Harte, Kitka, and Levine 2002).

따라서 시스템내에 존재하는 트래픽의 갯수가 적다면 각 트래픽이 시스템의 상태에 미치는 영향은 무시할 수 없게 된다. 즉 각 트래픽의 도착은 다른 트래픽들의 서비스 레벨 선택 영향을 주게 되며, 결과적으로 시스템이 안정된 균형상태에 이르지 못하게 만들 수도 있다. 다음에서 이러한 상황이 발생하게 되는 이유를 좀 더 자세히 살펴보자.

먼저 각 사용자가 어떻게 서비스 레벨을 선택하며, 어떤 방식으로 개인의 결정이 전체 시스템의 트래픽도착

2) 에지 네트워크는 병목현상을 일으키는 경우가 많고 또한 사용자가 처음으로 네트워크를 접하는 부분이기 때문에 가격정책에 관한 연구의 초점이 되어왔다. 이에 반해 코어 네트워크(core network)의 가격정책은 대부분 서비스 제공자들 사이의 서비스 레벨 계약(Service Level Agreement, SLA)에 의해 이루어진다.

를 결정하는지를 알아보도록 하자. 개인의 결정은 자신의 이득만을 고려하므로 j -타입의 트래픽을 가진 사용자는 자신의 유틸리티값에서 원가를 뺀 값을 최대화하는 서비스레벨 i 를 선택할 것이다. 즉 j -타입 트래픽을 가진 사용자의 의사결정은 다음의 식으로 표현할 수 있다:

$$\text{Argmax}_i [U_j(T_{ij}) - P_i \times C_j]. \quad (1)$$

이때 T_{ij} 는 j -타입의 트래픽이 i -레벨에서 서비스가 종료될 때까지 걸리는 기대 시간을, P_i 는 i -레벨의 가격을, C_j 는 j -타입 트래픽의 평균 크기를 말한다. 만약 유틸리티값에서 원가를 뺀 값이 음수가 되면, 그 트래픽은 시스템에 제출(submit)되지 않는다. (1)의 식에서 한 가지 유의할 사항은 SPS 스케줄링 방식에서는 종료시간을 보장할 수 없으므로 "기대" 종료 시간을 쓴다는 것이다. T_{ij} 의 값은 트래픽 도착 프로세스, 트래픽 처리 프로세스, 트래픽의 크기 등등의 여러가지 요인에 따라 달리 결정된다. 예를 들어 포아송 도착과 지수분포의 트래픽 크기를 가정한다면, j -타입의 트래픽이 i -레벨에서 종료까지 걸리는 기대 시간 T_{ij} 는 (2)의 식과 같이 $M/M/1$ 우선순위 대기행렬 모델³⁾로 구할 수 있다 (Bertsekas 1992).

$$T_{ij} = \frac{C_j}{\mu} + \frac{C^2/\mu^2 \cdot \sum_{k=1}^i \lambda_k}{(1 - C/\mu \cdot \sum_{k=1}^{i-1} \lambda_k)(1 - C/\mu \cdot \sum_{k=1}^i \lambda_k)}. \quad (2)$$

사용자의 결정은 식 (1)의 T_{ij} 에 의해 좌우되며, T_{ij} 는 식 (2)에서 보는 바와 같이 상위 레벨의 트래픽도착률에 의해 좌우되고, 트래픽도착률은 다른 사용자들의 결정에 의해 좌우된다. 이러한 상호의존성은 시스템이 진동할 수 있는 가능성을 불러 일으킨다.

예를 들어 두개의 우선순위 레벨이 있는 시스템을 생각해 보자. 만일 초기에 레벨 1 트래픽이 거의 없고 P_1 이 적당하다면, 많은 트래픽들이 레벨 1을 선택할 것이다. 만일 각 트래픽들이 시스템에 미치는 영향이 무시할 수 없는 정도라면, 이는 레벨 1의 트래픽도착률 (λ_1)을 높여서 결국 T_{1j} 를 높인다. 따라서 어떤 트래픽들은 레벨 1보다 레벨 2를 선택할 것이고, 이는 λ_2 를 증가시키고 λ_1 을 감소시키게 된다. 이렇게 되면 다시 레벨 1이 더 좋은 선택이 될 수도 있다. 그림 1은 이러한 사이클을 표현하고 있다. 결국 극단적인 경우에는 모든 들어오는 트래픽들이 한 레벨에서 다른 레벨로 왔다갔다하는 진동을 보일 수도 있다.

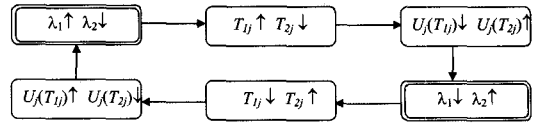


그림 1 두 서비스레벨이 있는 시스템에서의 진동

이제까지 각 트래픽들이 전체 시스템의 다이내믹스에 미치는 영향을 무시할 수 없는 경우에는 시스템이 진동할 수 있다는 것을 알아보았다. 그러나 만일 각 트래픽들의 영향이 무시할 정도로 미미하다고 가정해도, 또 다른 이유로 시스템이 수렴하지 않을 수 있다. 균형분석에서는 모든 사용자들이 최적의 의사결정을 내려야 하며, 이러한 최적의 의사결정을 위해서는 모든 레벨의 현재 트래픽도착률을 정확하게 알고 있어야 한다고 가정한다. (식 (2)에서 $\{\lambda_i\}$ 를 필요로 하는 것을 보면 알 수 있다.) 하지만 모든 레벨의 현재 트래픽도착률에 대한 최신 정보를 안다는 것은 현실적으로 불가능하거나 터무니없이 비싸게 된다. 특히 사용자들이 많아서 균형분석의 개개인의 영향은 무시할 수 있다는 첫번째 가정을 만족시킬 때에 더욱 그러하다. $\{\lambda_i\}$ 의 정확한 정보를 가지는 방법 중의 하나는 각 사용자가 다른 트래픽들을 계속 지켜보는 것이다. 그러나 이러한 스톱핑(snipping)은 보안상의 이유나 트래픽 라우팅의 기본적인 제한때문에 (예를 들어 모든 트래픽이 모든 사용자에게 보이는 것은 아니다) 불가능하다. 다른 방법으로, 사용자가 트래픽을 보내기 전에 현재의 $\{\lambda_i\}$ 값을 물어볼 수 있다. 하지만 이 방법은 (1) 엑스트라 트래픽을 만들어내어 효율성면에서 떨어지며, (2) 사용자가 사용하는 $\{\lambda_i\}$ 값이 정확하기 위해서는 대답을 받은 사용자가 트래픽을 보낼때까지 네트워크가 다른 트래픽들을 받지 않아야 하는 (따라서 트래픽 처리량을 엄청나게 떨어트리는) 바람직하지 않는 결과를 낳는다.

따라서 사용자가 현재 시스템의 상태 (즉, $\{\lambda_i\}$)에 대한 최신의 정확한 정보를 가지고 있지 않는다고 가정하는 것이 더 현실적이다. 예를 들어 네트워크가 정기적으로 (또는 $\{\lambda_i\}$ 가 주목할 정도로 크게 변할때) 가장 최근에 측정된 $\{\lambda_i\}$ 를 사용자에게 알려주는 것이 더 합리적이다. 이 경우 $\{\lambda_i\}$ 는 정기적으로 브로드캐스트되며, 사용자들은 (알려지지 않은 사실상의 $\{\lambda_i\}$ 대신) 가장 최근에 알려진 $\{\lambda_i\}$ 를 이용하여 서비스레벨을 선택하는 결정을 한다. 이때 이미 바뀌어졌을지도 모르는 정보를 사용하기 때문에 사용자의 결정은 sub-optimal할 수 있다. 균형분석에서 시스템의 수렴은 사용자가 최선의 정보로 최적의 결정을 내리는 것을 필요로 하므로, 정기적인 브로드캐스트에 의한 오래된 정보는 진동을 야기할 수도 있다.

3) 기대종료시간은 각 서비스레벨의 큐가 안정상태에 있을 때를 가정하여 구해진다. 이러한 가정은 업데이트 간격이 클 때에는 잘 맞는 가정이다. 업데이트 간격에 대해서는 이장의 마지막에 언급하고 있는데, 새로운 안정 상태에 도달하는 시간보다 긴 업데이트 간격은 sensible하다.

정기적 브로드캐스트 방법을 사용한다면, 네트워크 시스템은 $\{\lambda_i\}$ 에 대한 정보 업데이트의 빈도를 결정해야 한다. 업데이트의 간격이 크면 클수록 실제의 $\{\lambda_i\}$ 와 사용자가 사용하는 $\{\lambda_i\}$ 의 차이가 더 커지고 따라서 사용자의 결정의 sub-optimality도 커지게 된다. 브로드캐스트는 비싼 오퍼레이션이므로 IP 프로토콜의 브로드캐스트는 보통 초의 단위로 이루어진다. 예를 들어 RIP나 OSPF와 같은 IP 라우팅 프로토콜의 업데이트 간격은 1초에 한 번 정도이다⁴⁾(Huitema 1995). 다른 방법으로는 트래픽도착률을 이용하여 업데이트 간격을 정할 수도 있다. 즉, 트래픽도착률이 높으면 $\{\lambda_i\}$ 가 더 빨리 바뀔 것이므로 짧은 업데이트 간격을 선택할 수도 있다.

4. 동적 모델

앞에서 균형분석의 가정들이 잘 들어맞지 않는 네트워크 서비스의 실제 상황들을 알아보았다. 이러한 현실 상황을 정확히 모델링하기 위해서는 균형상태로의 수렴을 가정하지 않는 모델이 필요하며, 이를 위해 동적 모델을 개발하였다.

동적 모델은 가격 벡터 $\{P_i\}$ 에 의거하여 상태전이 (state transition) 그래프를 만든다. 상태전이 그래프상의 각 상태는 트래픽도착률 벡터 $\{\lambda_i\}$ 에 따라 정해지며, 상태의 전이는 각 사용자들의 결정에 의해 바뀌게 되는 트래픽도착률의 변화를 표현한다. 따라서 상태전이는 그림 2처럼 $\{\lambda_i\}$ 의 새 값이 브로드캐스트될 때 이루어진다. 이때 S^t 는 t 에서의 시스템 상태를 나타내며 $\{\lambda_i\}^t$ 는 S^t 에서의 트래픽도착률 벡터를 말한다.

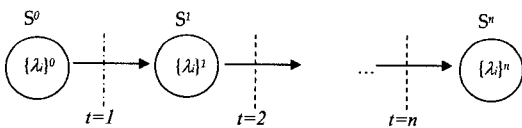


그림 2 시간에 따른 $\{\lambda_i\}$ 의 전이

앞서 말한 바와 같이 $\{P_i\}$ 에 따라 시스템은 안정된 상태로 수렴할 수도 있고 아닐 수도 있다. 동적 모델은 시스템의 수렴이 명백한(trivial) 경우와 시스템이 진동할 수도 있는 명백하지 않은(non-trivial) 경우를 나누어, 시스템이 수렴할 때에는 수렴하는 트래픽도착률 벡터 $\{\lambda_i\}$ 를 구하고 시스템이 진동할 경우에는 $\{\lambda_i\}$ 의 집합을 구해낸다. $\{\lambda_i\}$ 의 집합의 각 λ_i 는 진동하는 시스템의 순환(recurring) 상태를 나타내며, 얼마나 오랫동안 시스템이 각 상태에 머무르는지를 표현하는 확률을 가진다.

일단 두 레벨을 가진 시스템을 예를 들어 $\{\lambda_i\}$ 를 (또는 λ_i 의 집합을) 구하는 방법을 알아보고, 이를 다수의 레벨이 존재하는 시스템에 일반화하여 보자. 두 레벨을 가진 시스템에서는 사용자의 의사결정에 따라 네가지의 시나리오가 가능하다:

[시나리오-0] 모든 트래픽이 레벨 0을 선택한다(즉, 모든 트래픽이 전송을 포기한다.)

[시나리오-1] 모든 트래픽이 레벨 1을 선택한다.

[시나리오-2] 모든 트래픽이 레벨 2을 선택한다.

[시나리오-3] 어떤 트래픽들은 레벨 1을, 다른 트래픽들은 레벨 2을 선택한다.

결론적으로 말하면 시나리오-0, 1, 2의 경우에 $\{\lambda_i\}$ 를 구하는 것은 명백하나 시나리오-3의 경우는 명백하지 않다. 먼저 명백한(trivial) 경우를 살펴보자.

시나리오-0은 모든 사용자가 자신의 트래픽들을 레벨 1이나 레벨 2로 보낼 인센티브가 없을 때 발생한다. 이는 아래의 두 조건이 모두 만족하는 경우에 해당한다.

$$U_j(T_{1j}) - P_1 \times C_j < 0, \text{ for } \forall j \in J, \quad (A0)$$

$$U_j(T_{2j}) - P_2 \times C_j < 0, \text{ for } \forall j \in J. \quad (A1)$$

시스템의 상태를 λ_k 가 k 번째 서비스 레벨의 트래픽도착률을 나타내는 $(\lambda_0, \dots, \lambda_k, \dots, \lambda_I)$ 로 정의하자. 시나리오-0에서는 시스템이 $(\lambda, 0, 0)$ 의 상태로 수렴한다.

시나리오-1은 모든 사용자가 자신의 트래픽들을 레벨 1로 보낼 인센티브가 존재하고 (A2 조건), 모두가 레벨 1을 레벨 2보다 선호할 때 (A3 조건) 발생한다.

$$U_j(T_{1j}) - P_1 \times C_j \geq 0, \text{ for } \forall j \in J, \quad (A2)$$

$$U_j(T_{1j}) - P_1 \times C_j \geq U_j(T_{2j}) - P_2 \times C_j, \text{ for } \forall j \in J. \quad (A3)$$

시나리오-1에서는 시스템이 $(0, \lambda, 0)$ 의 상태로 수렴한다.

시나리오-2의 조건들도 비슷한 방법을 통해 구할 수 있다. 시나리오-2는 모든 사용자가 자신의 트래픽들을 레벨 2로 보낼 인센티브가 존재하고 (A4 조건), 모두가 레벨 2를 레벨 1보다 선호할 때 (A5 조건) 발생한다.

$$U_j(T_{2j}) - P_2 \times C_j \geq 0, \text{ for } \forall j \in J, \quad (A4)$$

$$U_j(T_{2j}) - P_2 \times C_j \geq U_j(T_{1j}) - P_1 \times C_j, \text{ for } \forall j \in J. \quad (A5)$$

시나리오-2에서는 $(0, 0, \lambda)$ 의 상태로 수렴한다.

위의 경우들을 일반화시키면, I 개의 서비스 레벨이 존재하는 시스템에서 모든 트래픽들이 k -번째 레벨을 ($k \neq 0$) 선택하는 조건은 다음과 같이 표현할 수 있다.

$$U_j(T_{kj}) - P_k \times C_j \geq 0, \text{ for } \forall j \in J,$$

$$U_j(T_{kj}) - P_k \times C_j \geq U_j(T_{ij}) - P_i \times C_j, \text{ for } \forall j \in J, \forall i \in I, i \neq k.$$

위의 조건이 만족하면 모든 트래픽들은 레벨 k 를 선택하고 시스템은 λ 가 $(k+1)$ 의 위치에 있는 $(0, \dots, \lambda, \dots, 0)$ 의 상태로 수렴한다.

시나리오-3은 안정 상태(steady state)로의 수렴을 보

4) 본 논문의 경우 브로드캐스트가 에지 네트워크의 마지막 홉(hop)으로 제한되므로, IP 라우팅 때보다는 더 잦은 업데이트도 수용할 수 있다.

장하지 못하므로, 시스템이 여러 상태를 진동할 수도 있다. 동적 모델은 진동 루프(loop)에 해당하는 $\{\lambda_i\}$ 들을 찾아내고 $\{\lambda_i\}$ 의 전이를 명시적으로 모델링한다. 이를 위해서는 새단계의 계산이 필요하다.

첫째, 현재의 시스템 상태에서 각 타입의 트래픽이 가장 선호하는 서비스 레벨을 예측한다. 즉 모든 (i, j) 쌍에 대해, 주어진 가격벡터가 $\{P_i\}$ 이고 시스템이 S^i 일 때 j -타입의 트래픽이 가장 선호하는 서비스 레벨이 i 이면 1을 취하는 $X'_{i,j}$ 를 구한다.

$$X'_{i,j} = \begin{cases} 1, & \text{if } U_j(T_j^i) - P_i \cdot C_j \geq 0, \text{ and } U_j(T_j^i) - P_i \cdot C_j \geq U_j(T_j^k) - P_k \cdot C_j, \text{ for } \forall k \in I, k \neq i \\ 0, & \text{otherwise} \end{cases}$$

여기에서 T_j^i 는 시스템이 S^i 상태일 때 j -타입의 트래픽을 i -레벨로 보내는데 필요한 기대 종료 시간이다. 주의할 것은 t 와 $t+1$ 사이의 레벨 선택 결정은 $\{\lambda_i\}^t$ 에 의해 이루어지므로 t 와 $t+1$ 사이에 도착한 트래픽들은 계산에 포함되지 않는다.

둘째, $X'_{i,j}$ 를 이용하여 S^i 에서 레벨 i 가 선택될 확률, $Prob(i, S^i)$,을 구한다:

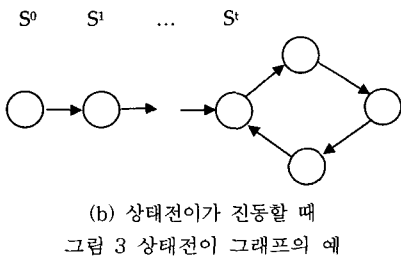
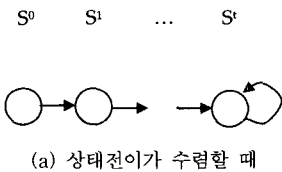
$$Prob(i, S^i) = \sum_j Pr(j) \cdot X'_{i,j} \quad (3)$$

여기에서 $Pr(j)$ 는 어떤 트래픽이 j -타입일 확률을 말한다.

마지막으로, $Prob(i, S^i)$ 를 이용하여 S^{t+1} 에서의 트래픽도착률, $\{\lambda_i\}^{t+1}$,을 구한다:

$$\lambda_i^{t+1} = \lambda \times Prob(i, S^i) \text{ for } \forall i \in I. \quad (4)$$

위와 같은 방법을 통해 동적 모델은 상태전이 그래프를 만들 수 있다. 시나리오-3의 경우 트래픽도착률 $\{\lambda_i\}$ 의 전이는 그림 3(a)처럼 결국 하나의 상태로 수렴할 수도 있고, 그림 3(b)처럼 여러 상태를 진동할 수도 있다. 그림 3(b)의 경우 진동 루프내의 각 상태 S 의 고정(stationary) 확률, $W(S)$,은 방문횟수를 이용하여 계산할 수 있다.



5. 균형분석법에 대한 평가

균형분석은 균형상태가 존재한다는 가정아래, 시스템의 목적을 극대화하면서 균형상태에 수렴하도록 만드는 최적의 가격벡터를 계산해낸다. 한편, 동적모델은 실제 시스템이 진동할 수도 있다는 가정에 어떤 주어진 가격벡터에 따른 시스템의 전이를 명시적으로 모델링한다.

이상에서는 현실적인 네트워크에서 (즉 각 사용자들이 시스템에 끼치는 영향을 무시할 수 없고 시스템 상태의 변화를 정기적으로 브로드캐스트하는 경우) 과연 시스템이 균형상태에서 계산되어진 안정된 균형상태에 도달하는지를 동적 모델을 이용하여 알아본다. 구체적으로 말해, 균형분석에서 얻어진 가격벡터를 동적 모델에 적용하여 두 모델에서 얻어지는 트래픽도착률 벡터를 비교해 보는 방법을 취한다. 비교대상으로는 본 논문과 비슷한 프레임워크를 쓰는 세가지의 내쉬 균형모델(Nash equilibrium model)을 선택하였다:

- Case I: 하나의 서비스 레벨을 제공하며 최대 이득을 노리는 시스템,
- Case II: 하나의 서비스 레벨을 제공하며 시스템의 가치(value)를 최대화하려는 시스템,
- Case III: 두개의 서비스 레벨을 제공하며 시스템의 가치를 최대화하려는 시스템.

위의 세가지 내쉬 균형모델들에서 j -타입의 트래픽의 가치는 식 (5)의 유틸리티 함수로 표현된다:

$$U_j(T) = V_j(\lambda_j^*) - v_j \times T. \quad (5)$$

이때 $V_j(\lambda_j^*)$ 는 한계가치함수(marginal value function)이고, v_j 는 지연비용(delay cost)이며, T 는 트래픽 종료까지의 지연시간을 표현한다. 가치함수(value function), $V_j(\lambda_j^*)$,는 실질적인 도착률이 λ_j^* 인 j -타입 트래픽들이 단위시간안에 얻는 총 기대값을 나타낸다.⁵⁾ $V_j(\lambda_j^*)$ 가 단조증가(monotone increase)하며, 오목(strictly concave)하고, 두번 미분이 가능(twice-differentiable)하다고 가정하면, 하나의 j -타입 트래픽의 가치는 한계 가치함수 $V_j(\lambda_j^*)$ 로 표현할 수 있다. 트래픽의 가치는 종료시간에 반비례하므로, 지연 가격 v_j 는 상수이다. 위의 세 모델에서 사용하는 유틸리티 함수에 대한 자세한 설명은 참고문헌에서 찾아볼 수 있다(Mendelson 1985, Mendelson and Whang 1990).

다음에서 각 경우에 해당하는 균형모델이 어떻게 균형상태를 구해내는지를 간단히 설명하고, 실제 시스템이 균형모델이 예측하는 균형상태로 수렴하는지 알아본다.

5.1 Case I

이득 중심의 가격구조(profit-center pricing struc-

5) λ_j^* 은 j -타입 트래픽도착률을 표현하고, s_i 은 i -레벨의 트래픽도착률을 표현하는 것에 주의한다.

ture) 모델은 크기가 일정한($C = 1$) 한 종류의 트래픽 ($J = 1$)이 하나의 서비스 레벨을 가진 ($I = 1$) 시스템에 있을 때를 살펴본다(Mendelson 1985). 따라서 모든 트래픽의 유틸리티 함수는 같고, 사용자들은 각 트래픽을 전송할지 안할지의 선택만을 가진다. 하나의 서비스 레벨만이 존재하므로 실질적인 트래픽 도착률 λ^* 은 레벨 1의 트래픽 도착률 λ_1 과 같다.

시간단위당 기대 이득을 나타내는 시스템의 목적 함수로 $\lambda^* \times P_I - \mu \times P_B$ 를 사용하며, 이때 P_B 는 시스템이 백홀 링크(backhaul link)에 내는 가격을 말한다. 또한 등탄력성 한계가치함수(iso-elastic marginal value function) $V'(\lambda^*) = K/\sqrt{\lambda^*}$ 를 가정한다. 이 균형모델은 내쉬 균형점(Nash equilibrium)의 조건을 만족하는 최적 가격 (P_I)과 최적 네트워크 용량(μ)을 동시에 구한다. 내쉬 균형점은 트래픽을 전송하거나 안하거나 상관없이 없는 (indifferent) 경우에 생긴다. 다시 말해, 균형상태에서는 식 (5)의 유틸리티 값이 서비스 가격과 같을때에 발생한다. (즉 $V'(\lambda^*) - v_I \times T_{II} = P_I$ 의 조건이 만족된다.) 따라서 최적의 균형상태에서는 시스템의 목적함수가 $\lambda^* \times [V'(\lambda^*) - v_I \times T_{II}] - \mu \times P_B$ 이 된다. 목적함수를 최대화하는 조건은 다음과 같다:

$$V'(\lambda^*) + V''(\lambda^*) = (v_I / \mu) \times f'(\lambda^*/\mu), \quad (6)$$

$$P_B = v_I \times (\lambda^*/\mu^2) \times f'(\lambda^*/\mu),$$

$$\text{where } f(\rho) = \rho/(1 - \rho), \quad \rho = \lambda^*/\mu. \quad (7)$$

위의 식들에 가정한 $V'(\lambda\mu^*) = K/\sqrt{\lambda^*}$ 를 적용하면, (6)의 식에서는 $K^2 = 4 v_I \times P_B \times f(\rho)$ 를, (7)의 식에서는 $\mu = K^2 \cdot \rho/(4 P_B^2)$ 를 구해낼 수 있다. 따라서 앞의 식에서 최적 시스템 활용률 ρ 를, 뒤의 식에서 최적의 시스템 용량 μ 를 구할 수 있다. (7)의 식에서 보다시피 $\rho = \lambda^*/\mu$ 이므로, 최적의 트래픽 도착률 λ^* 은 ρ 와 μ 로부터 구한다. 최적 가격 P_I 은 (2)의 식에서 구해지는 T_{II} 값을 균형조건에 적용하여 구한다. 보다 자세한 계

산은 Mendelson의 논문을 참조하도록 한다(Mendelson 1985).

이제 위와 같은 균형모델에서 구해진 트래픽 도착률을 동적 모델에서 구해진 도착률과 비교해 보자. 비교시에 사용되는 두 시스템으로 (1) $\{\lambda_i\}$ 의 업데이트가 빈번한(일초당 100번, 즉 $F = 100$) 큰 용량의 네트워크(큰 μ)와 (2) 업데이트가 빈번하지않은 (일초당 한두번, 즉 $F = 1$ 이거나 $F = 2$) 작은 용량의 네트워크(작은 μ)를 선택하였다.

표 1과 표 2는 비교결과를 정리하여 보여준다. 균형분석 모델은 P_B, K, v_I 를 입력으로 받아 균형상태에서의 최적의 가격 (P_I), 최적 용량 (μ), 최적의 트래픽도착률 (λ_i)을 구해낸다. 동적 모델은 (균형모델에서 사용한) P_B, K, v_I 와 (균형모델에서 얻어진) P_I 과 μ 를 입력으로 받는다. 동적 모델에서 총 트래픽도착률(λ)은 $0.99^* \mu$ 로 정한다. 여기에서 가장 중요한 비교대상은 실질적인 트래픽도착률 (λ_i)이다. 균형분석에서는 언제나 하나의 균형상태로 수렴하므로 λ_i 은 하나의 값으로 구해진다. 동적모델에서는 진동 루프내에서 여러개의 λ_i 값이 있는 경우도 있다. 예를 들어 $\lambda_i = \{73.8, 74.7\}$ 는 시스템이 $\lambda_i = 73.8$ 과 $\lambda_i = 74.7$ 인 두개의 상태를 진동하는 것을 표현한다.

표 1에서 보여주는 바와 같이, 예상한 대로 트래픽도착률에 관한 정보를 자주 업데이트하는 큰 용량의 네트워크에서는 균형분석의 결과와 동적모델의 결과가 거의 일치한다. 이는 시스템이 하나의 안정 상태로 수렴하지 않는다 해도 진동의 정도가 무시할 수 있을 정도로 작고 진동하는 상태들이 균형상태에 매우 근접하기 때문이다. 이러한 비교결과는 3장에서 설명된 두가지의 조건이 만족하는 경우에 균형분석이 실제 시스템의 작동상황을 잘 모델링하는 것을 보여준다.

이와는 대조적으로, 표 2를 보면 트래픽도착률의 브로드캐스트가 빈번하지 않은 작은 용량의 네트워크에서는

표 1 업데이트가 빈번한 큰 용량의 네트워크(case I)

(P_B, K, v_I)	균형모델		동적모델 ($F = 100$)
	(0.5, 10, 1)	$P_I = 1.08, \mu = 85.9$	$\lambda_i = 73.7$
(0.5, 10, 3)	$P_I = 1.16, \mu = 75.5$	$\lambda_i = 57.0$	$\lambda_i = \{56.6, 57.4\}$
(1.0, 20, 1)	$P_I = 2.11, \mu = 90.0$	$\lambda_i = 81.0$	$\lambda_i = \{81.0, 81.8\}$
(1.0, 20, 3)	$P_I = 2.21, \mu = 82.7$	$\lambda_i = 68.4$	$\lambda_i = \{67.8, 68.6\}$

표 2 업데이트가 빈번하지 않은 작은 용량의 네트워크(case I)

(P_B, K, v_I)	균형모델		동적모델 ($F = 2$)	동적모델 ($F = 1$)
	(1.5, 10, 1)	$P_I = 3.49, \mu = 8.4$	$\lambda_i = 6.3$	$\lambda_i = \{8.3, 4.2\}$
(1.5, 10, 3)	$P_I = 4.11, \mu = 6.4$	$\lambda_i = 3.7$	$\lambda_i = \{6.3, 3.2\}$	$\lambda_i = \{6.3, 0.0\}$
(2.0, 20, 1)	$P_I = 4.33, \mu = 21.5$	$\lambda_i = 18.4$	$\lambda_i = \{21.3, 10.6\}$	$\lambda_i = \{21.3, 0.0\}$
(2.0, 20, 3)	$P_I = 4.65, \mu = 18.9$	$\lambda_i = 14.3$	$\lambda_i = \{18.7, 9.4\}$	$\lambda_i = \{18.7, 0.0\}$

표 3 업데이트가 빈번한 큰 용량의 네트워크(case II)

(P_B, K, v_i)	균형모델		동적 모델 ($F = 100$)
(0.5, 10, 1)	$P_I = 0.5, \mu = 371.7$	$\lambda_I = 345.4$	$\lambda_I = \{345.3\}$
(0.5, 10, 3)	$P_I = 0.5, \mu = 351.0$	$\lambda_I = 308.0$	$\lambda_I = \{305.1, 308.6\}$
(1.0, 20, 1)	$P_I = 1.0, \mu = 380.0$	$\lambda_I = 361.0$	$\lambda_I = \{360.6\}^*$
(1.0, 20, 3)	$P_I = 1.0, \mu = 365.4$	$\lambda_I = 333.7$	$\lambda_I = \{332.2\}$

표 4 업데이트가 빈번하지 않은 작은 용량의 네트워크(case II)

(P_B, K, v_i)	균형모델		동적모델 ($F = 2$)	동적모델 ($F = 1$)
(1.5, 10, 1)	$P_I = 1.5, \mu = 39.0$	$\lambda_I = 34.2$	$\lambda_I = \{38.6, 19.3\}$	$\lambda_I = \{38.6, 0.0\}$
(1.5, 10, 3)	$P_I = 1.5, \mu = 35.0$	$\lambda_I = 27.6$	$\lambda_I = \{34.7, 17.3\}$	$\lambda_I = \{34.7, 0.0\}$
(2.0, 20, 1)	$P_I = 2.0, \mu = 92.9$	$\lambda_I = 86.4$	$\lambda_I = \{92.0, 46.0\}$	$\lambda_I = \{92.0, 0.0\}$
(2.0, 20, 3)	$P_I = 2.0, \mu = 87.8$	$\lambda_I = 77.0$	$\lambda_I = \{86.9, 43.5\}$	$\lambda_I = \{86.9, 0.0\}$

균형분석의 결과가 동적모델의 결과와 매우 다르게 나타나는 것을 알 수 있다. 빈번하지 않은 $\{\lambda_i\}$ 의 업데이트는 심각한 진동을 불러일으킨다. 업데이트 간격이 커질수록 진동의 간격도 커지며, 극한 상황에서는 한 업데이트 간격에서 모든 트래픽이 들어오고 다음 간격에서는 하나도 들어오지 않는 상황이 생길 수도 있다($F = 1$ 인 경우를 보면 알 수 있다). 이러한 결과는 균형모델이 가정하는 시스템의 수렴이 맞지 않는 것을 보여주며, 균형모델이 실제 시스템의 작동상태를 제대로 모델링하지 못한다는 것을 증명한다.

5.2 Case II

시스템의 총 가치를 최대화하는 (net-value maximization) 모델은 목적함수를 제외하고는 case I과 같다 (Mendelson 1985). 즉, $I = 1, J = 1, C = 1, V'(\lambda^*) = K/\sqrt{\lambda^*}$ 를 사용한다. 시스템의 목적함수는 $V(\lambda^*) - v_i \times L - \mu \times P_B$ 로 표현할 수 있다. 여기에서 L 은 안정상태에 있는 시스템에서의 평균 트래픽의 갯수를 나타내며, $M/M/1$ 모델에서 $L = f(\rho) = \lambda^*/(\mu - \lambda^*)$ 이 성립한다. 내쉬 균형점을 찾는 방법은 case I과 동일하다. 목적함수를 최대화하는 조건은 다음과 같다:

$$V'(\lambda^*) = v_i \times (\partial L / \partial \lambda), \tag{8}$$

$$P_B = -v_i \times (\partial L / \partial \mu). \tag{9}$$

가정한 $V'(\lambda^*)$ 를 적용하면 (8)의 식에서 $K^2 = v_i \times P_B \times f(\rho)$ 를, (9)의 식에서 $\mu = K^2 \cdot \rho / P_B^2$ 를 구해낼 수 있다. case I과 유사한 방법으로 두 식을 통해 μ 와 λ_i 를 구해내고, 균형 조건에 μ 와 λ_i 를 적용하여 P_I 을 구한다.

표 3과 표 4는 균형모델과 동적 모델의 결과를 비교하여 보여준다. 이러한 결과는 case I과 비슷한 경향을 나타낸다. 빈번한 업데이트가 있는 큰 용량의 네트워크의 경우에는 시스템이 균형상태에 이르는 경우가 많이 있다(표 3). 이와는 반대로 표 4는 균형분석이 업데이

트가 빈번하지 않은 작은 용량의 네트워크를 정확하게 모델링하지 못하는 것을 보여준다.

Case I과 case II를 통해 균형상태로의 수렴은 시스템의 목적과 무관하다는 것을 알 수 있다. 이는 사용자의 서비스 레벨 선택의 결정은 시스템의 목적과는 상관 없이 자신의 이득만을 고려하여 내리저기 때문이다. 또한 균형상태로의 수렴을 위한 조건이 만족되지 않았을 때는 시스템의 목적이 무엇이든간에 시스템이 진동하는 것을 알 수 있다.

5.3 Case III

Case III는 case II에 사용된 균형모델을 확장하여 다수의 레벨과 여러개의 트래픽 타입을 고려하는 균형모델을 다룬다 (Mendelson 1990). 새 모델은 주어진 시스템 용량 μ 하에서 시스템 전체의 가치 (value)를 최대화하는 경우를 다룬다. 시스템의 목적함수는 다음과 같이 표현된다.

$$\sum_{j \in J} V_j(\lambda_j^*) - \sum_{j \in J} v_j \times L_j(\{\lambda_j^*\}),$$

여기에서 λ_j^* 는 j -타입 트래픽의 도착률을, $V_j(\lambda_j^*)$ 는 j -타입 트래픽의 가치(value) 함수를, v_j 는 j -타입 트래픽의 지연비용을, $\{\lambda_j^*\}$ 는 j -타입 트래픽도착률 벡터를, $L_j(\{\lambda_j^*\})$ 는 안정상태의 시스템에 존재하는 j -타입 트래픽들의 평균 갯수를 나타낸다.

예를 들어, 두 개의 서비스 레벨과 ($I = 2$) 두 타입의 트래픽이 ($J = 2$) 존재하는 시스템을 생각해 보자. 시스템의 목적을 최대화하는 조건은 다음과 같다:

$$V_1'(\lambda_1^*) = v_1 \times (\partial L_1 / \partial \lambda_1^*) + v_2 \times (\partial L_2 / \partial \lambda_1^*), \tag{10}$$

$$V_2'(\lambda_2^*) = v_1 \times (\partial L_1 / \partial \lambda_2^*) + v_2 \times (\partial L_2 / \partial \lambda_2^*). \tag{11}$$

6) 시스템이 수렴하는 경우에도 동적 모델의 결과는 균형모델과 정확하게 일치하지는 않는다. 이는 단지 두 모델이 가지는 다른 가정 (무한한 트래픽과 시스템상태에 대한 정확한 정보)뿐만 아니라 반올림 (round-off)에 의한 오차때문이기도 하다. 예를 들어 브로드캐스트되는 $\{\lambda_i\}$ 의 값은 소숫점 아래 5자리로 잘려진다.

위의 식을 이용하면 주어진 가치 함수, $V_1(\lambda_1^*)$ 과 $V_2(\lambda_2^*)$ 에서 타입 1과 타입 2의 트래픽의 최적도착률을 구할 수 있다. 최적의 $\{\lambda_j^*\}$ 이 구해지면 균형조건, $V_1'(\lambda_1^*) = P_1 + v_1 \times W_1$ 와 $V_2'(\lambda_2^*) = P_2 + v_2 \times W_2$ 에 의해 최적가격 P_1 과 P_2 를 계산할 수 있다. 이때 W_j 는 j -타입 트래픽의 기대 종료 지연시간을 말한다. 이때 v_1 이 v_2 보다 크고 $C_1 = C_2 = 1$ 라고 가정하면 incentive-compatibility에 의해 타입 1의 트래픽들은 항상 레벨 1을, 타입 2의 트래픽은 레벨 2를 선호하게 된다(Mendelson 1990). 즉, $\lambda_1^* = \lambda_1$ 이고 $\lambda_2^* = \lambda_2$ 가 된다. 결과적으로 균형상태에서 W_1 은 T_{11} 과 같고 W_2 는 T_{22} 와 같게 된다.

위와 같은 균형모델을 분석하기 위해 Mendelson 논문에는 예를 이용한다(Mendelson 1990). 이 예는 $\lambda_1^* > 0.45$ 일 때 $V_1'(\lambda_1^*) = 9 - 20\lambda_1^*$ 를, $\lambda_2^* > 0.4$ 일 때 $V_2'(\lambda_2^*) = 12 - 30\lambda_2^*$ 를 사용한다. μ 의 값은 1.0으로 고정되어있고, v_1 은 2로 v_2 는 1로 되어있다. 케이스 II와 마찬가지로, 균형모델을 이용하여 최적의 가격벡터를 구한 후, 이를 동적 모델에 적용하여 시스템의 상태(즉, $\{\lambda_j\}$ 의 전이)를 알아본다. 표 5에 나타난 바와 같이, 케이스 III도 케이스 I과 케이스 II와 비슷한 경향을 보인다. $\{\lambda_j\}$ 의 업데이트 간격이 작을때는 시스템이 균형모델의 결과에 매우 근접한다. 다시 말해 F 가 100일 때, 트래픽들은 incentive-compatibility 원리를 따르고, 동적 모델에 의해 구해진 $\{\lambda_j\}$ 는 (하나의 값으로 수렴하지는 않으나) 균형모델에 의해 구해진 최적의 $\{\lambda_j^*\}$ 와 거의 일치한다. 그러나 $\{\lambda_j\}$ 의 업데이트 간격이 커지면 incentive-compatibility 원리가 더이상 맞지않게 되고 시스템은 진동하게 된다.

표 5 Case III의 비교결과

	$\{P_1, P_2\} = \{2.21, 1.61\}$ 일 때의 $\{\lambda_1, \lambda_2\}$
균형모델	$\{0.183, 0.278\}$ 로 수렴.
동적모델 ($F = 100$)	$\{0.183, 0.282\}$, $\{0.183, 0.277\}$, $\{0.188, 0.277\}$, $\{0.178, 0.277\}$, $\{0.183, 0.272\}$ 를 진동.
동적모델 ($F = 2$)	$\{0.0, 0.495\}$, $\{0.0, 0.742\}$, $\{0.0, 0.247\}$, $\{0.247, 0.247\}$, $\{0.247, 0.495\}$, $\{0.247, 0.742\}$, $\{0.0, 0.0\}$ 를 진동.
동적모델 ($F = 1$)	$\{0.0, 0.0\}$, $\{0.495, 0.0\}$, $\{0.0, 0.99\}$ 를 진동.

6. 결론

이 논문은 우선순위 방식의 네트워크 서비스에서의 가격선택 문제를 다루고 있다. 특히 이 문제의 해결책중 현재 가장 널리 사용되고 있는 균형분석모델에 초점을 맞추고 있다. 본 논문은 균형모델의 가정이 맞지 않는 경우에 시스템이 하나의 균형상태가 아닌 여러개의 상태를 진동하는 것을 보여줌으로써 균형분석의 문제점

을 지적하고, 정적 균형점이 존재하지 않을 수 있는 현실적인 QoS 네트워크를 모델할 수 있는 동적 모델을 개발하였다. 동적 모델을 이용한 우선순위 네트워크 서비스 가격문제의 해결책은 다른 논문에서 다루고 있다 (Park et al. 2006).

본 연구는 QoS 서비스를 위한 우선순위 가격정책을 모델링하는 새로운 기법을 제공할 뿐만 아니라 현존하는 균형분석의 문제점을 파헤침으로써 큰 의미를 가진다. 이제까지 가장 보편적인 모델링으로 쓰여왔던 균형분석이 에지 네트워크에 적용되기 어렵다는 것을 보임으로써, 라스트 마일 인터넷의 가격책정에 새로운 방법론을 제공하는 것은 기존의 균형분석 방법의 한계를 극복하는 패러다임의 변화를 가져오리라고 기대된다.

참고 문헌

- [1] Balachandran, K., and B. Srinidhi.(1987). A rationale for fixed charge application, *Journal of Accounting, Auditing and Finance*, Vol 15, pages 151-169.
- [2] Bertsekas, D., and R. Gallager. (1992). *Data Networks*. Prentice-Hall.
- [3] Blake, F., D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. (1998). An architecture for differentiated services, *IETF RFC 2475*.
- [4] Cocchi, R., D. Estrin, S. Shenker, and L. Zhang. (1991). A Study of Priority Pricing in Multiple Service Class Networks. *Proc. ACM SIGCOMM*, pages 123-130.
- [5] Cocchi, R., S. Shenker, D. Estrin, and L. Zhang. (1993). Pricing in Computer Networks: Motivation, Formulation, and Example. *IEEE/ACM Transactions on Networking*, Vol. 1, pages 614-627.
- [6] Gupta, A., D. Stahl, and A. Whinston. (1997). Priority Pricing of Integrated Service Networks. *Internet Economics*, MIT Press, pages 323-352.
- [7] Harte, L., R. Kitka, and R. Levine. (2002). *3G Wireless Demystified*. McGraw-Hill.
- [8] Hassin, R. (1986). Consumer information in markets with random products quality: The case of queues and balking, *Econometrica*, Vol. 54, pages 1185-1195.
- [9] Hassin, R., and M. Haviv. (2003). *To Queue Or Not To Queue: Equilibrium Behavior In Queueing Systems*. Kluwer Academic Publishers.
- [10] Huitema, C. (1995). *Routing in the Internet*. Prentice-Hall.
- [11] ITU. (2002). ICT Free Statistics on Information Technology. http://www.itu.int/ITU-D/ict/statistics/at_glance/Internet02.pdf
- [12] Kelly, F. (1997). Charging and Rate Control for Elastic Traffic, *European Transactions on Telecommunications*, Vol. 8, pages 33-37.

- [13] Kelly, F. (2001). *Mathematical Modeling of the Internet. Mathematics Unlimited - 2001 and Beyond*, Springer-Verlag, pages 685-702.
- [14] Kelly, F., A. Maullo, and D. Tan. (1998). Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability, *Journal of the Operation Research Society*, Vol. 49, pages 237-252.
- [15] Low, S., and D. Lapsley. (1999). Optimization Flow Control - I: Basic Algorithm and Covergence. *IEEE/ACM Transactions on Networking*, Vol. 7, pages 861-874.
- [16] Mackie-Mason, J., and H. Varian. (1995). Pricing Congestible Network Resources. *IEEE JSAC*, Vol. 13, No 7, pages 1141-1149.
- [17] Mackie-Mason, J., L. Murphy, and J. Murphy. (1997). The Role of Responsive Pricing in the Internet. *Internet Economics*, MIT Press, pages 279-303.
- [18] Mandjes, M. (2003). Pricing Strategies under Heterogeneous Service Requirements. *IEEE INFOCOM*.
- [19] Mendelson, H. (1985). Pricing Computer Services: Queuing Effects. *Communications of the ACM*, Vol. 28, No 3, pages 312-321.
- [20] Mendelson, H., and S. Whang. (1990). Optimal Incentive-Compatible Priority Pricing for the M/M/1 Queue. *Operations Research*, Vol. 38, No 5, pages 870-883.
- [21] Molina, M., P. Castelli, and G. Foddis. (2000). Web Traffic Modeling Exploiting TCP Connection's Temporal Clustering through HTML-REDUCE, *IEEE Network*, Vol. 14, No 3, pages 46-55.
- [22] Odlyzko, A. (1997). A Modest Proposal for Preventing Internet Congestion. <http://www.research.att.com/~amo>.
- [23] Odlyzko, A. (2001). Internet Pricing and the History of Communications. <http://www.research.att.com/~amo>.
- [24] Park, S., S. Han, and M. H. Rothkopf. (2006). A Dynamic Model Approach for Network Service Pricing, In preparation.
- [25] Russel, S., and P. Norvig. (1995). *Artificial Intelligence: A Modern Approach*. Prentice-Hall.
- [26] Shenker, S., D. Clark, D. Estrin, and S. Herzog. (1996). Pricing in Computer Networks: Reshaping the Research Agenda, *ACM Computer Communication Review*, Vol. 26, pages 19-43.

service level.

λ	Total potential job arrival rate ($\sum_{i=0}^I \lambda_i = \lambda$).
λ_j^*	Job arrival rate of type- j jobs.
C_j	The size of type- j job.
$Pr(j)$	The probability that a job is of type j .
$U_j(t)$	Utility function of a type- j job.
P_i	Price of level- i service.
$\{P_i\}$	Price vector for the system with I levels.
P_B	Price of the network backhaul capacity.
T_{ij}	Expected completion time of a type- j job at level i .

박 선 주

정보과학회논문지 : 정보통신
제 33 권 제 1 호 참조

Appendix: Notations

μ	Job service rate.
λ_i	Job arrival rate at level i .
λ_0	Rate of jobs that are not submitted to the network.
$\{\lambda_k\}$	Job arrival rate vector, where λ_k at the k -th position represents the arrival rate at k -th