

고정 전송율 및 가변 전송율 서비스를 지원하는 다중 서비스 네트워크의 성능 분석 및 설계

정회원 이 태 진*

Performance Analysis and Design of Multiservice Networks for Constant- and Adaptive-rate Services

Tae-Jin Lee* *Regular Member*

요 약

본 논문에서는 고정 전송율 서비스와 가변 전송율 서비스를 동시에 지원하는 다중 서비스 네트워크의 성능을 분석하고 이를 설계에 이용하는 방법을 연구한다. 다중 서비스 네트워크에서 고정 전송율 서비스의 호 수는 시간에 따라 동적으로 가변하게 되는데, 이에 따라 전체 네트워크 링크의 용량중 가변 전송율 서비스가 사용가능한 대역폭(전송율)도 시간에 따라 가변하게 된다. 본 논문에서는 가변 전송율 서비스가 사용가능한 대역폭을 Erlang 고정점 방정식과 Little의 법칙을 이용해서 구하였다. 가변 전송율 서비스의 호들에 대한 대역폭 할당은 최대-최소 공정 대역폭 할당 방법에 따라 결정되는데, 본 논문에서는 먼저 가변 전송율 서비스가 얻을 수 있는 평균 최소 전송율의 상한값을 구하고 대용량 네트워크일 경우에는 이 상한값을 얻을 수 있음을 보였다. 네트워크의 예를 통해 이러한 성능 분석 방법을 검증하였으며, 최대-최소 공정 대역폭 할당 방법이 네트워크의 전체 수익을 최대화하지 못할 수도 있으므로, 네트워크 레벨에서의 우선 순위를 고려함으로써 네트워크의 효율을 근대화하는 방법을 제안하였다.

ABSTRACT

We consider a model describing large scale performance of multiservice networks supporting both constant- and adaptive-rate services. As the number of constant-rate connections changes over time in multiservice networks, the available bandwidth for adaptive-rate connections changes as well. We estimate the available bandwidth for adaptive-rate service in a dynamically changing traffic environment based on Erlang fixed point equation and Little's law. The bandwidth allocation for adaptive-rate service is determined by the max-min fairness criterion. We find an upper bound on the average minimum throughput for adaptive-rate service, and show that the bound is achieved in large-capacity networks. This is verified using network examples. Since max-min fair allocation may not maximize the total throughput or revenue in a network, we consider adjustments via network level priority schemes to increase efficiency.

1. 서 론

음성, 영상 등의 멀티 미디어 데이터를 지원하는 통합 서비스 네트워크에서는 다양한 QoS (Quality of Service)를 필요로 하는 여러 형태의 트래픽이 전송될 것이다. 이러한 다양한 트래픽을 전송하기

위해 고려되는 서비스에는 고정 전송율 서비스와 가변 전송율 서비스가 있다. 고정 전송율 서비스에서는 일정한 전송율로 데이터가 전송되도록 함으로써 멀티미디어 등의 데이터를 실시간으로 전송할 수 있다. 그리고, 가변 전송율 서비스는 고정 전송율 서비스의 호(connection)에 할당되고 남은 대역

* 성균관대학교 전기전자컴퓨터공학과 (tjlee@ece.skku.ac.kr)
논문번호 : 010174-0710, 접수일자 : 2001년 7월 10일

폭(전송율)을 이용함으로써, 한정된 네트워크 자원을 효율적으로 사용하면서, 적당한 수준의 QoS를 보장한다. 고정 전송율 서비스의 예로는 ATM 네트워크의 CBR(Constant Bit Rate) 서비스가 있고 가변 전송율 서비스의 예로는 ABR(Available Bit Rate) 서비스가 있다^{[1][2]}. 인터넷에서는 TCP를 이용하는데 이는 가변 전송율 서비스의 한 형태라고 할 수 있다^[3].

가변 전송율 서비스와 고정 전송율 서비스가 동시에 지원되는 네트워크 환경에서 가변 전송율 서비스에 할당되는 대역폭은 시간에 따라 가변하게 된다. 따라서 이러한 가변적인 대역폭이 가변 전송율 서비스를 이용하는 호들에 공정하게 할당되도록 하는 것이 중요하다. 기존에는 동적으로 고정 전송율 서비스의 호들이 변하는 경우에, 랜덤한 대역폭의 변화 및 공정한 전송율의 할당 방법 등이 서로 복잡하게 연관되어 있어서, 정량적으로 대역폭 및 각 호에 할당가능한 전송율에 대한 성능을 분석하는 것이 용이하지 않았다. 본 논문에서는 평균 대역폭 및 네트워크 서비스 사용자가 사용가능한 성능을 정량적으로 분석하는 방법을 제안하고 이를 네트워크 용량 설계에도 이용하고자 한다.

가변 전송율 서비스를 위해서는 현재 네트워크 링크에 얼마나 많은 대역폭이 고정 전송율 서비스에 할당되고 있는지를 알아내고, 이를 바탕으로 가변 전송율 서비스에 할당가능한 대역폭을 계산해 줄 필요가 있다. 그리고 가변 전송율 서비스를 이용하는 호들에 효율적으로 공정하게 대역폭을 분배해주어야 한다. 고정 전송율 서비스는 유효 대역폭(effective bandwidth)의 개념을 이용해 손실 네트워크 모델(loss network model)로, 즉 서킷-스위치 네트워크(circuit-switched network)로 모델링할 수 있다^[4]. 이러한 모델을 이용하면 가변 전송율 서비스에 할당 가능한 대역폭을 구해 낼 수 있다.

일반적으로 가변 전송율 서비스의 성능(호당 전송율 등)을 구하는 것은 간단하지 않다. 그 이유는 첫째, 고정 전송율 할당 서비스를 이용하는 호수가 시간에 따라 가변하고, 이러한 랜덤 환경하에서 가변 전송율 서비스에 할당될 대역폭을 구하는 것이 간단하지 않고, 둘째, 가변 전송율 서비스를 이용하는 호들에 대역폭을 공정하게 분배하기 위해 많이 사용되는 최대-최소 공정 전송율 할당 방법(max-min fair bandwidth allocation)이나 비례 공정 전송율 할당 방법(proportionally fair bandwidth allocation)과 같은 방법에서, 가변 전송율 서비스의 대역폭 성

능은 가변 전송율 서비스가 실제 사용가능한 대역폭의 복잡한 함수로 표현되기 때문이다.

본 논문에서는 고정 전송율 서비스의 호와 가변 전송율 서비스의 호들이 동적으로 동시에 존재하는 다중 서비스 네트워크에서 가변 전송율 서비스의 평균 성능, 즉 가변 전송율 서비스의 호들에 할당될 평균 최소 전송율을 구하고, 이를 이용해서 성능을 만족시키기 위한 네트워크 링크의 전체 대역폭(링크의 용량)을 설계하는 것을 목적으로 한다. 다중 서비스 네트워크에서처럼 고정 전송율 서비스의 호수에 따라 가변 전송율 서비스에 할당되는 대역폭이 변하는 경우에는 호들의 성능을 구하는 것이 쉽지 않다. 따라서, 최대-최소 공정 전송율 할당 방법하에서의 가변 전송율 서비스의 평균 최소 전송율을 구하기 위해, 본 논문에서는 먼저 평균 최소 전송율의 상한값을 구한다. 그리고, 링크의 용량과 고정 전송율 서비스 호의 도착율이 큰 대용량 네트워크에서는 이 상한값을 얻는 것이 가능하다는 것을 보인다.

이와 같이 동적인 고정 전송율 서비스 호들이 네트워크에 들어오고 나가는 환경에서, 고정 전송율 서비스에 할당되고 남은 대역폭을 효율적으로 이용하기 위해 가변 전송율 서비스를 이용할 수 있게 하면, 네트워크 서비스 제공자는 네트워크 자원의 이용 효율을 최대화할 수 있다. 또한, 가변 전송율 서비스 사용자가 적정한 수준의 성능(전송율)을 얻을 수 있도록 해주어야 하는데, 본 논문에서 제안한 분석 방법을 적용하면 동적인 다중 서비스 네트워크 환경에서 가변 전송율 서비스 사용자에게 보장할 수 있는 최소한의 전송율(각 가변 전송율 서비스 사용자당 전송율)을 예측, 분석하는 것이 가능하다. 그리고, 가변 전송율 서비스 호의 최소 필요 전송율이 주어져 있을 때, 이를 만족하기 위한 네트워크 링크의 용량 설계에도 활용할 수 있다.

본 논문의 구성은 다음과 같다. 먼저 II절에서 공정 전송율 할당 방법에 대해 살펴본 후, III절에서 고정 및 가변 전송율 서비스 및 그 성능 요인에 대해 알아본다. IV절에서는 최대-최소 공정 전송율 할당 방법으로 가변 전송율 서비스 호들에 공정하게 전송율을 할당할 때, 각 호에 할당되는 평균 최소 전송율의 상한값을 구한다. 그리고, 대용량 네트워크에서 이러한 상한값을 얻을 수 있음을 보인다. 가변 전송율 서비스 호들을 위해 가능한 전체 평균 전송율을 V절에서 구한다. 네트워크의 실제 예를 VI절에서 보이고, 전체 네트워크의 효율을 증가시키기 위한 방법을 VII절에서 제시한 후, 결론을 VIII

에 요약한다.

II. 공정 전송을 할당 방법

먼저 공정 전송을 할당 방법에 대해 알아보도록 한다. 호들의 집합 S 가 네트워크를 공유한다고 가정하자. 여기서 각각의 호 $s \in S$ 는 그 호가 지나가는 링크들인 L_s 를 갖게 된다. 한 개 이상의 호가 각각의 링크를 지날 수 있으므로, S_l 을 링크 l 이 지나가는 호들로 나타낸다. 그리고 $\vec{a} = (a_s, s \in S)$ 를 호들의 전송율로, b_l 을 링크 l 의 용량으로 나타내도록 한다. 호(사용자)들의 최소 효용(예: 호의 전송율)을 최대화하는 문제를 고려해 보자. 즉,

$$\max_{\vec{a}} \{ \min_{s \in S} u_s(a_s) \mid \sum_{s \in S_l} a_s \leq b_l, a_s \geq 0 \} \quad (1)$$

여기서 $u_s(a_s)$ 는 호 s 에 할당된 a_s 만큼의 전송율의 효용이다. 위의 최적화 문제를 풀면 최소 효용을 가지는 호들의 집합을 찾을 수 있다. 그러한 호들을 제거하고, 제거된 호들의 용량을 호들이 지나가는 링크의 대역폭에서 빼주면 원래의 네트워크보다 축소된 네트워크를 얻을 수 있다.

이러한 네트워크에 대해 앞 (1)에서와 같이 최적화 문제를 다시 적용하면 이번에는 두 번째로 작은 효용을 가진 호들의 집합을 얻을 수 있게 된다. 이러한 과정을 모든 호가 제거될 때까지 반복하면 계층적인 호들의 집합들(최소 효용을 가진 호들의 집합부터 최대 효용을 가진 호들의 집합까지)을 구할 수 있게 된다. 이 때, 각 호에 할당된 전송율이 \vec{a}^* 이고, $u_s(a_s) = a_s$ 일 때, 그 대역폭 할당 방법을 최대-최소 공정 전송을 할당 방법이라고 한다 [5][6]. 즉, 이 경우에는 호에 할당되는 최소 전송율이 최대가 되게 한다. 따라서, 최대-최소 공정 전송을 할당 방법은 대역폭이 한정되어 있을 때, 네트워크에서 계층적으로 최적화 문제를 푸는 것과 같다.

만일 네트워크의 전체 효용을 최대화하는 것이 목적이라면, 대역폭 할당 방법에서의 목표 함수는 호(사용자)의 효용 함수의 합으로 나타낼 수 있다. 즉,

$$\max_{\vec{r}} \{ \sum_{s \in S} u_s(r_s) \mid \sum_{s \in S_l} r_s \leq b_l, r_s \geq 0 \}, \quad (2)$$

여기서 $\vec{r} = (r_s, s \in S)$ 는 호 전송율의 벡터를 나타낸다. 이러한 최적화 문제를 풀면 호 전송율 \vec{r}^*

을 얻게 된다. 여기서는 최대-최소 공정 전송을 할당 방법과는 달리 각각의 호의 성능을 희생해서 전체 네트워크의 효용이 최대가 되도록 전송율을 할당한다. 최대-최소 공정 전송을 할당 방법은 전체 전송율이 최대가 되도록 할당하는 것이 아니므로 호에 우선 순위를 줌으로써 전체 네트워크의 성능을 향상시킬 수도 있다. 이 문제에 대해서는 VII절에서 살펴보고 성능 향상을 위한 설계 및 관리에 이를 이용하는 방법을 고려한다.

III. 가변 전송율 및 고정 전송율 서비스

용량이 $\vec{c} = (c_l, l \in L)$ 인 링크의 집합으로 이루어진 네트워크를 고려하자. 그리고, 고정 전송율 서비스의 호들이 지나가는 경로의 집합을 R 로 나타낸다. 여기서 각 경로 $r \in R$ 은 링크 L_r 들을 지난다. 하나 이상의 고정 전송율 서비스 경로가 각 링크를 지날 수 있으므로, R_l 을 링크 l 을 지나가는 고정 전송율 서비스 호의 경로들로 표시한다. 고정 전송율 서비스의 호들 $r \in R$ 은 도착율이 ν_r 인 Poisson 프로세스라고 가정한다. 또한, 호의 접속 시간은 단위 평균을 갖는다, 즉 $\mu_r^{-1} = 1$. 호의 접속 시간들은 서로 독립적이고, 이전 호의 도착 시간들과도 독립적이다. 편의상 고정 전송율 서비스의 호는 단위 전송율을 필요로 한다고 가정한다.

고정 전송율 서비스 트래픽이 필요한 만큼의 전송율(대역폭)을 확보한 후에 가변 전송율 서비스를 위해 할당 가능한 전송율을 다음과 같이 구할 수 있다. $\vec{B} = (B_l, l \in L)$ 을 각 링크가 가변 전송율 서비스를 위해 사용 가능한 전송율들로 나타내자. 여기서 \vec{B} 는 랜덤 벡터를 나타낸다. 그리고, 사용가능한 대역폭 \vec{B} 와 링크의 용량 \vec{c} 를 구분해서 나타낸다. 가변 전송율 서비스 호의 집합 S 가 네트워크를 사용한다고 생각하자. 여기서 각 호 $s \in S$ 는 링크 L_s 를 지난다. 집합 L_s 는 고정 전송율 서비스의 호 l 가 지나가는 링크들을 나타낸다. 한 개 이상의 가변 전송율 서비스 호가 하나의 링크를 지날 수 있으므로, 링크 l 을 지나가는 가변 전송율 서비스 호의 집합을 S_l 로 나타낸다. 여기서 가변 전송율 서비스 호는 고정되어 있는 것으로 가정한다. 즉, 한 네트워크에서 가변 전송율 서비스 호들의 수는 고정되어 있다 (예를 들어, 가변 전송율 서비스를 사용하는 LAN 호의 경우처럼), 그림 1 참조.

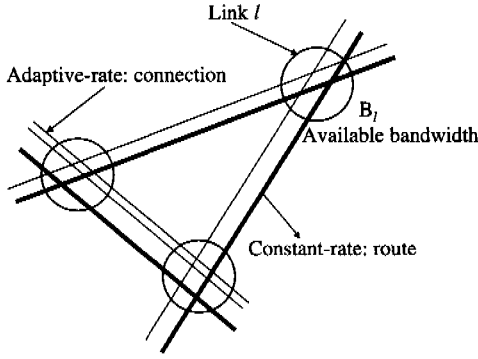


그림 1. 두개의 링크를 지나는 가변 전송률 서비스 경로와 고정 전송률 서비스 흐름.

여기서 링크들의 용량은 한정되어 있으므로, 링크에 도착하는 고정 전송률 서비스의 호는 차단될 수 있고, 이에 따라 가변 전송률 서비스 호의 전송률에 영향을 미칠 수 있다. 이러한 점에서 고정 전송률 서비스 호의 차단 확률 (Blocking Probability) 및 가변 전송률 서비스 호의 평균 전송률, 이 두가지를 각각의 성능 척도로 생각할 수 있다.

IV. 공정 전송률 할당 방법하에서 가변 전송률 서비스 호들의 평균 최소 전송률

가변 전송률 서비스가 사용가능한 전송률을 구하기 위해서는 먼저 고정 전송률 서비스의 호수에 대한 정보를 알아야 한다.

1. 고정 전송률 서비스 호수의 확률 분포

네트워크상에서 고정 전송률 서비스 호수의 확률 분포를 구하기 위해 [7]에서와 같은 손실 네트워크 모델을 이용한다. 고정 전송률 서비스 호수의 분포를 $\pi(\vec{n})$ 으로 나타낸다. 여기서 $\vec{n}=(n_r, r \in R)$ 이고, n_r 은 경로 r 을 사용하는 고정 전송률 서비스의 호수를 나타낸다.

$$\pi(\vec{n}) = G(\vec{c})^{-1} \prod_{r \in R} \frac{\nu_r^{n_r}}{n_r!}, \quad s \in S(\vec{c}) \quad (3)$$

여기서

$$S(\vec{c}) = \{ \vec{n} \in \mathbb{Z}_+^{|R|} \mid A\vec{n} \leq \vec{c} \},$$

$$G(\vec{c}) = \left(\sum_{\vec{n} \in S} \prod_{r \in R} \frac{\nu_r^{n_r}}{n_r!} \right).$$

여기서 $A=(A_{lr}, l \in L, r \in R)$ 은 0,1 행렬로 경로 r 이 링크 l 을 지나는지 않는지를 나타낸다. 만약

A_{lr} 이 1이면 경로 r 은 링크 l 상에서 하나의 호를 점유한다. \vec{N} 을 분포가 π 인 랜덤 벡터로 나타내자. 그러면 링크 l 에서 가변 전송률 서비스 트래픽이 사용가능한 전송률은 다음과 같이 주어진다.

$$B_l = c_l - \sum N_r, \quad l \in L. \quad (4)$$

여기서 N_r 이 랜덤 변수이므로 B_l 역시 랜덤 변수가 된다. 다음 절에서는 주어진 공정 전송률 할당 방법하에서, 가변 전송률 서비스 호들이 사용가능한 평균 전송률을 B_l 의 함수로 나타내는 방법을 알아본다.

2. 가변 전송률 서비스 호의 평균 최소 전송률

앞에서 기술한 것처럼 최대-최소 공정 전송률 할당 방법은 계층적인 최적화 문제가 된다. $\vec{B}=(B_l, l \in L)$ 을 링크 l 에서 가변 전송률 서비스 호들이 사용가능한 전송률 벡터라 하자. $\vec{B}=\vec{b}$ 일 때, 최소 전송률을 최대화하는 것은 다음 문제로 나타낼 수 있다.

$$W(\vec{b}) = \max_{\vec{a}} \{ \min_{s \in S} a_s \mid \sum_{s \in S_l} a_s \leq b_l, a_s \geq 0 \}, \quad (5)$$

여기서 $f(\vec{a}) = \min_{s \in S} a_s$ 는 오목(convex)하고 $g_l(\vec{a}) = \sum_{s \in S_l} a_s$ 는 볼록(concave)하므로 최소 전송률 $W(\vec{b})$ 도 \vec{b} 에 대해 오목하다 (강한 이중성 (strong duality)의 법칙 [8]). 따라서, 공정 전송률 할당 방법 W 하에서 가변 전송률 서비스의 호에 할당되는 평균 최소 전송률 $E[W(\vec{B})]$ 의 상한값은 다음의 정리로 주어진다.

정리 4.1 (평균 최소 전송률의 상한)

$$E[W(\vec{B})] \leq W(E[\vec{B}]). \quad (6)$$

증명: $W(\vec{b})$ 는 오목하므로 위의 정리는 Jensen의 부등식으로부터 구해질 수 있다. ■

여기서 우리는 이러한 결과가 다음 단계의 전송률(두 번째로 최소인 전송률)에도 동일하게 적용되는지 확인할 필요가 있다. 그러나, 두번째로 최소인 전송률은 원래의 용량인 \vec{b} 에 대해 오목하지 않을 수도 있다. 이러한 사실은 최대-최소 전송률 할당 방법이 최소 전송률 $W(\vec{b})$ 를 최대로 하기 위해 두 번째로 최소인 혹은 그 이상으로 최소인 전송률을

회생함으로써 얻어질 수도 있다는 것으로 설명할 수 있다. 이것은 최대-최소 공평 전송을 할당 방법이 전체 네트워크의 전송율이나 전체 네트워크의 효용을 최대화하지는 않는다는 사실과도 일치한다.

3. 대용량 네트워크에서 가변 전송을 서비스 호들의 평균 최소 전송율

가변 전송을 서비스의 평균 최소 전송율의 상한값(6) 참조)은 대용량 네트워크에서 얻어질 수 있다. 다음에서 이 사실을 증명한다. 순차적으로 커지는 연속되는 네트워크들을 고려하자. 즉 도착율 $\vec{\nu}^{(n)} = (\nu_r^{(n)}, r \in R)$ 과 링크의 용량 $\vec{c}^{(n)} = (c_l^{(n)}, l \in L)$ 이 동시에 증가하는 네트워크들을 고려하자. 그리고 $\vec{B}^{(n)} = (B_l^{(n)}, l \in L)$, $B_l^{(n)} = c_l^{(n)} - \sum_{r \in R_l} N_r^{(n)}$ 으로 나타내자. 여기서 (n) 은 n 번째 네트워크를 나타낸다.

$\vec{x}_r = \nu_r \prod_{l \in L_r} (1 - D_l)$ 을 [9]에 있는 것처럼 다음 최적화 문제의 해(확률 분포 (3) 하에서 가장 많이 존재할 가능성이 있는 상태 $\vec{x} = (x_r, r \in R)$)로 나타내자.

$$\begin{aligned} & \text{Maximize} \quad \sum_{r \in R} (x_r \log \nu_r - x_r \log x_r + x_r) \\ & \text{subject to} \quad A \vec{x} \leq \vec{c}, \quad \vec{x} \geq 0. \end{aligned} \quad (7)$$

마찬가지로 $\vec{x}_r^{(n)}$ 을 n 번째 네트워크에 해당되는 최적화 문제의 해로 나타내자. (7)의 라그랑지안은

$$\begin{aligned} L(\vec{x}, \vec{z}; \vec{y}) &= \sum_{r \in R} (x_r \log \nu_r - x_r \log x_r + x_r) \\ &+ \sum_{l \in L} y_l (c_l - \sum_{r \in R_l} A_{lr} x_r - z_l), \end{aligned} \quad (8)$$

여기서 \vec{z} 는 슬랙 변수의 벡터이고 \vec{y} 는 라그랑지안 승수의 벡터이다. 이 최적화 문제의 대칭 문제는 다음과 같이 표현할 수 있다.

$$\begin{aligned} & \text{Minimize} \quad \sum_{r \in R} (\nu_r \exp(-\sum_{l \in L_r} y_l A_{lr}) + \sum_{l \in L} y_l c_l) \\ & \text{subject to} \quad \vec{y} \geq 0, \end{aligned} \quad (9)$$

그리고, 그 해는 \vec{y}_l 이 된다. 여기서 $1 - D_l = \exp(-\vec{y}_l)$. 그리고, n 번째 네트워크에 대해서는

$$U_r^{(n)} = n^{-1/2} (N_r^{(n)} - \bar{x}_r^{(n)}). \quad (10)$$

위의 네트워크 확장이 $n \rightarrow \infty$ 에 따라 다음을 만

족한다고 가정하자.

$$\frac{\vec{\nu}^{(n)}}{n} \rightarrow \vec{\nu}, \quad \frac{\vec{c}^{(n)}}{n} \rightarrow \vec{c}. \quad (11)$$

우리는 여기서 두 가지 경우를 고려한다. 정부하 네트워크(용량과 도착 트래픽의 부하의 차가 $O(N^{1/2})$ 인 경우)와 저부하 네트워크(용량과 도착 트래픽의 부하의 차가 $O(N)$ 인 경우).

정부하 네트워크인 경우, Kelly가 [9]에서 보인 것처럼, $n \rightarrow \infty$ 에 따라

$$\frac{\bar{x}_r^{(n)}}{n} \rightarrow \bar{x}_r, \quad (12)$$

그리고,

$$U_r^{(n)} \rightarrow U_r \sim N(0, \bar{x}_r) \text{ in distribution.} \quad (13)$$

이러한 결과를 이용해서 우리는 다음의 정리를 구할 수 있는데, 이 정리로부터 평균 최소 전송율의 상한값을 극한적으로 구할 수 있다.

정리 4.2 (가변 전송을 서비스 호들의 평균 최소 전송율)

네트워크에 도착하는 호의 수와 네트워크 링크의 용량이 모두 증가해가는 연속 확장 네트워크들을 고려하자. 그 증가가 (11)에 따라 이루어질 때, 평균 정규화된 전송율은 확장이 계속됨에 따라 그 상한값에 수렴하게 된다. 즉,

$$\lim_{n \rightarrow \infty} E \left[\frac{1}{n} W(\vec{B}^{(n)}) \right] = W(\vec{\gamma}),$$

여기서 $\vec{\gamma}$ 는 부하 조건에 따라서 각각

1. 저부하인 경우:

$$\gamma_l = c_l - \sum_{r \in R_l} \nu_r$$

2. 정부하인 경우:

$$\gamma_l = c_l - \sum_{r \in R_l} \nu_r \prod_{l \in L_r} (1 - D_l).$$

증명: 먼저 정부하인 경우를 고려하면,

$$\lim_{n \rightarrow \infty} E \left[\frac{1}{n} W(\vec{B}^{(n)}) \right] = \lim_{n \rightarrow \infty} E \left[W \left(\frac{\vec{B}^{(n)}}{n} \right) \right].$$

그리고, (10)에 의해

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{B_l^{(n)}}{n} \\ &= \lim_{n \rightarrow \infty} \left(\frac{c_l^{(n)}}{n} - \frac{\sum_{r \in R_l} N_r^{(n)}}{n} \right) \\ &= \lim_{n \rightarrow \infty} \left(\frac{c_l^{(n)}}{n} - \frac{\sum_{r \in R_l} (\bar{x}_r^{(n)} + n^{1/2} U_r^{(n)})}{n} \right). \end{aligned}$$

따라서, (11)과 (12)로 부터

$$\lim_{n \rightarrow \infty} \frac{B_i^{(n)}}{n} = c_i - \sum_{r \in R_i} \bar{x}_r - \lim_{n \rightarrow \infty} \frac{U_r^{(n)}}{n^{1/2}}.$$

(13)을 이용하면

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{B_i^{(n)}}{n} &= c_i - \sum_{r \in R_i} \bar{x}_r \\ &= c_i - \sum_{r \in R_i} \prod_{l \in L_r} (1 - D_l). \end{aligned}$$

저부하인 경우에는, 차단율이 점점 작아지므로

$$\lim_{n \rightarrow \infty} E \left[\frac{1}{n} W(\bar{B}^{(n)}) \right] = W(\gamma).$$

앞에서의 확장해가는 네트워크에서의 정리는 다음과 같이 해석할 수 있다. 커다란 n 에 대해

$$E[W(\bar{B}^{(n)})] \cong W(E[\bar{B}^{(n)}]). \quad (14)$$

그러므로, 대용량 네트워크에서는 가변 전송율 서비스의 평균 최소 전송율의 상한값을 얻을 수 있음을 알 수 있다. (14)를 이용하면 가변 전송율 서비스의 호에 할당되어야 할 평균 최소 전송율을 구할 수 있고, 그 성능을 만족하기 위한 링크의 용량 c_l 을 구할 수 있다. 즉, 네트워크 링크의 대역폭을 설계하는 것이 가능하다.

V. 가변 전송율 서비스 트래픽을 위해 사용가능한 평균 전송율

다음으로 가변 전송율 서비스가 사용가능한 평균 전송율 $E[\bar{B}]$ 를 구하는 방법에 대해 알아본다. 차원이 $|R|$ 로 상태 공간이 작은 경우라도 정상 상태의 분포 (3)하에서, $G(\bar{c})$ 나 호의 수를 구하는 것은 간단하지 않은 문제이다. 이러한 경우에 취할 수 있는 방법은 실제 트래픽을 측정하거나, 근사적으로 평균 호의 수 $E[\bar{N}]$ 을 구하는 것이다. Erlang의 고정점 방정식^[7]으로부터 가변 전송율 서비스가 사용가능한 평균 대역폭을 계산하는 방법을 고려할 수 있다.

E_l 을 링크 $l \in L$ 의 차단 확률로 나타내자. 그리고, 링크 l 의 호들이 서로 독립적이고, 또한 다른 링크들과도 독립적이라고 가정하자. ρ_l 을 링크 l 에 도착하는 호들의 유효 도착율로 나타내면,

$$\begin{aligned} E_l &= E(\rho_l, c_l) \\ &= \left(\sum_{n=0}^{c_l} \frac{\rho_l^n}{n!} \right)^{-1} \frac{\rho_l^{c_l}}{c_l!}, \quad l \in L, \end{aligned} \quad (15)$$

여기서

$$\rho_l = \sum_{r \in R_l} \nu_r \prod_{m \in L_r \setminus \{l\}} (1 - E_m), \quad l \in L. \quad (16)$$

이것은 축소 과정("thinning" process)이라고 할 수 있다. 즉, 트래픽이 링크 l 에 도착하기 전에 각 링크 $m \in L_r \setminus \{l\}$ 에서 $1 - E_m$ 만큼 트래픽의 양이 축소되는 것으로 생각할 수 있다. 식 (15)와 (16)은 유일한 해를 가진 고정점 방정식들로 이루어져 있다^[7]. 경로 r 에서의 실제 전송율 ξ_r 은

$$\begin{aligned} \xi_r &= \nu_r (1 - L_r) \\ &\cong \nu_r \prod_{l \in L_r} (1 - E_l), \quad r \in R, \end{aligned} \quad (17)$$

여기서 L_r 은 경로 r 에서의 손실 확률을 나타낸다.

전송율 ξ_r 이 주어져 있을 때, Little의 법칙과 평균 호의 점유 시간이 단위 시간이라는 가정에 의해, 경로 r 에서의 평균 호수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} E[N_r] &= \xi_r = \nu_r (1 - L_r) \\ &\cong \nu_r \prod_{l \in L_r} (1 - E_l), \quad r \in R. \end{aligned} \quad (18)$$

Kelly는^[9]에서 링크의 용량과 호의 도착율이 동시에 증가할 경우에 Erlang의 고정점 방정식으로부터 정확한 추정치를 구할 수 있다는 것을 보였다.

다음으로 각 링크에서 가변 전송율 서비스 트래픽이 사용가능한 평균 대역폭을 구하도록 한다. 즉,

$$E[B_l] = c_l - \sum_{r \in R_l} E[N_r], \quad l \in L \quad (19)$$

여기서 $E[N_r]$ 은 (18)로부터 추정할 수 있다.

VI. 네트워크의 예

앞에서의 정리를 확인하기 위해 실제 네트워크의 예를 보인다. 예에 사용된 네트워크는 다음과 같다 (그림 2 참조). 네트워크의 링크(스위치)는 3개이고, 각 링크마다 고정 전송율 서비스 호를 위한 경로가 하나씩 지나고 있고, 전체 링크를 모두 지나는 경로가 하나 있다. 즉, 4개의 경로를 통해 고정 전송율 서비스의 호들이 지나게 된다. 이 경로들을 통해 고정 전송율 서비스 호들이 동적으로 지나게 되고, 고정 전송율 서비스 호의 각 경로와 병행하여 가변 전송율 서비스의 호가 2개씩 있다. 즉, 각 링크에는 4개의 호 (긴 호 2개, 짧은 호 2개)가 전체적으로는 8개의 호(긴 호 2개, 짧은 호 6개)의 가변 전송율

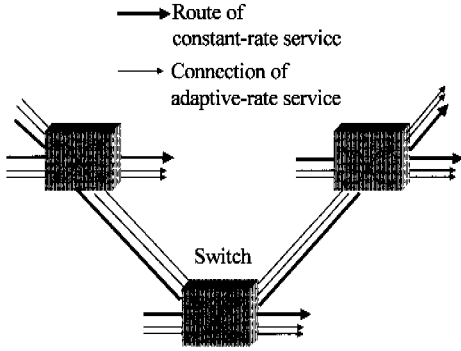


그림 2. 3개의 스위치와 4개의 고정 전송율 서비스 경로, 각 고정 전송율 서비스 경로마다 2개씩의 가변 전송율 서비스 호가 있는 네트워크의 예.

서비스 호가 네트워크 상에 존재한다. 고정 전송율 서비스의 호들에는 우선적으로 전송율이 할당되어 데이터를 전송할 수 있게 하고, 전체 대역폭 (각 링크의 용량) 중 고정 전송율 서비스에 할당되고 남은 대역폭을 최대-최소 공정 전송을 할당 방법에 따라 가변 전송율 서비스의 호들에 할당하게 된다.

본 예에서는 고정 전송율 서비스의 호당 필요 전송율(대역폭)이 1 Mbps인 경우를 고려한다. (즉, 1 Mbits/호). 그리고, 고정 전송율 서비스 호의 도착율 $\nu^{(n)} = 0.2n$ (호/초)이고, 각 링크의 용량 $c^{(n)} = 2n$ (호/초)로 가정한다. $\nu^{(n)}$ 과 $c^{(n)}$ 을 비트 단위로 생각하면 고정 전송율 서비스 호의 필요 전송율이 1 M (bits/호)이므로, 각각 $\nu^{(n)} = 0.2n$ (Mbps), $c^{(n)} = 2n$ (Mbps)가 된다. 예를 들어, n 이 10이면 $\nu^{(n)} = 2$ (Mbps), $c^{(n)} = 20$ (Mbps)가 된다. n 을 변화시키면서 (14)의 $E[W(\bar{B}^{(n)})]$ 과 $W(E[\bar{B}^{(n)}])$ 을 n 에 대해 정규화시켜서 나타내면, 그림 3과 같이 나타난다. 그림 3에 보인 것처럼 n 이 증가함에 따라 가변 전송율 서비스의 호에 할당될 평균 최소 전송율의 상한값인 400 kbps에 수렴해 가는 것을 볼 수 있다. 즉, 각 링크에서 1.6 Mbps(400 kbps x 4)는 가변 전송율 서비스에 할당되는 전송율로 생각할 수 있다.

이번에는 고정 전송율 서비스 호의 도착율이 링크 용량에 비해 상대적으로 큰 경우인 (앞의 예에 비해) $\nu^{(n)} = 0.5n$ (호/초)이고, 각 링크의 전체 용량 $c^{(n)} = 2n$ (호/초)를 고려한다. 고정 전송율 서비스 호당 필요 전송율은 앞의 예와 같다고 가정한다 (즉, 1 M (bits/호)). 이 경우에 각각 $E[W(\bar{B}^{(n)})]$ 와 $W(E[\bar{B}^{(n)}])$ 을 n 에 대해 정

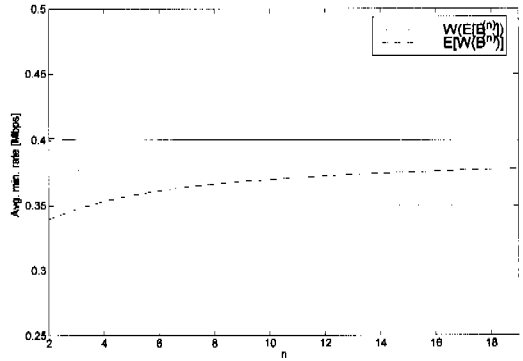


그림 3. 가변 전송율 서비스 호의 평균 최소 전송율 $\nu^{(n)} = 0.2n, c^{(n)} = 2n$.

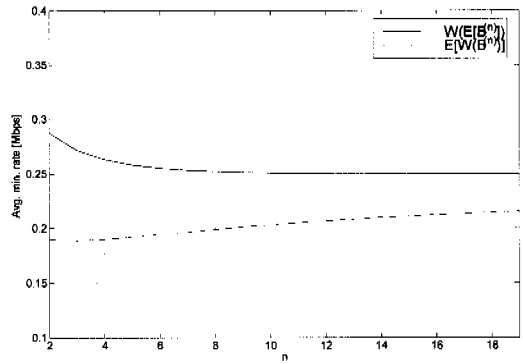


그림 4. 가변 전송율 서비스 호의 평균 최소 전송율 $\nu^{(n)} = 0.5n, c^{(n)} = 2n$.

규화시켜서 구한 값을 그림 4에 나타내었다. n 이 증가함에 따라 가변 전송율 서비스의 호에 할당되는 평균 최소 전송율은 그 상한값인 250 kbps에 수렴해 가는 것을 확인할 수 있다. 앞의 경우에 비해 고정 전송율 서비스호의 도착율이 2.5 ($\nu^{(n)} = 0.2n \rightarrow 0.5n$)배 증가했으므로, 고정 전송율 서비스의 호들에 보다 많은 대역폭이 할당되어야 하고, 결과적으로 가변 전송율 서비스에 할당될 대역폭은 더 적어지게 된다. 즉, 앞의 경우에서의 400 kbps보다 적은 250 kbps가 각각의 가변 전송율 서비스의 호들에 할당되게 된다 (각 링크당 1 Mbps (250kbps x 4)가 할당됨).

이와 같은 분석 방법을 활용하면 원하는 성능을 얻기 위한 링크의 용량도 설계할 수 있다. 고정 전송율 서비스의 도착율이 10 Mbps에서 50 Mbps까지 변할 때, 가변 전송율 서비스 호에 할당되어야 할 최소 전송율의 요구 조건을 만족하기 위해 필요한 링크의 용량을 그림 5에 나타내었다. 고정 전송율 서비스 트래픽의 도착율 ν 가 40 Mbps이상이

되면, 링크의 용량이 50 Mbps이하일 경우에 가변 전송을 서비스에 할당 가능한 최소 전송율이 1 Mbps를 넘을 수 없음을 확인할 수 있다. 따라서, 이 경우에 원하는 성능을 위해서는 링크의 용량을 50 Mbps이상으로 더 늘려야 함을 알 수 있다. ν 가 10 Mbps일 경우에는 링크의 용량이 20 Mbps만 되어도 1 Mbps의 전송율을 가변 전송을 서비스의 호에 보장해 줄 수 있다. ν 가 20 Mbps가 되면 같은 정도의 성능을 가변 전송을 서비스에 보장해 주기 위해서 35 Mbps의 링크 용량을 필요로 한다. ν 가 30 Mbps이면 같은 정도의 성능을 가변 전송을 서비스에 보장해 주기 위해서 45 Mbps의 링크 용량을 필요로 함을 알 수 있다. 이와 같은 방법으로 고정 전송율 서비스 호의 트래픽 정보를 알고 있고, 가변 전송율 서비스의 성능 요구 조건이 미리 주어져 있을 때, 네트워크 설계자의 입장에서 필요한 정도의 네트워크 용량을 설계할 수 있다.

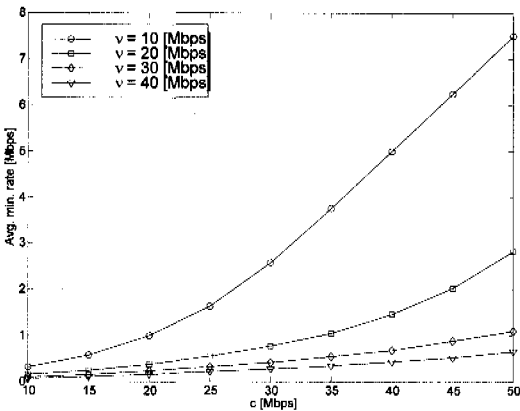


그림 5. 고정 전송율 서비스 트래픽의 도착율이 가변할 때, 링크의 용량과 가변 전송율 서비스 호에 할당될 평균 최소 전송율 (최대-최소 공정 전송율 할당 방법).

Ⅶ. 네트워크의 전체 효율을 증가시키기 위한 대역폭 할당 방법

최대-최소 공정 전송율을 할당 방법은 전체 전송율의 합을 최대화하지 않을 수도 있다. 따라서 가변 전송율 서비스의 호들에 우선 순위를 줌으로써 가중화된 최소 전송율이 최대가 되도록 하는 것도 가능한데, 여기서는 이러한 방법을 구체적으로 살펴보고자 한다. 가중 최대-최소 공정 전송율 할당 방법 (weighted max-min fair allocation of bandwidth)을 고려하자.

$$W_w(\vec{b}) = \max_a \{ \min_{s \in S} \frac{a_s}{w_s} \mid \sum_{s \in S} a_s \leq b_s, w_s \geq 1 \} \quad (20)$$

여기서 w_s 는 호 s 의 우선 순위로 호 s 에 가중치를 주기 위한 것이다. $f(\vec{a}) = \min_{s \in S} (a_s/w_s)$ 가 오목하고, $g_1(\vec{a}) = \sum_{s \in S} a_s$ 가 볼록하며, $W_w(\vec{b})$ 가 오목하다는 사실로부터, 가중 최대-최소 공정 전송율 할당 방법하에서 평균 최소 전송율의 상한 값을 구할 수 있다. 즉,

$$E[W_w(\vec{B})] \leq W_w(E[\vec{B}]) \quad (21)$$

4.3절에서와 마찬가지로 대용량 네트워크일 경우에는 이 상한값에 도달할 수 있음을 증명할 수 있다.

가중 최대-최소 공정 전송율 할당 방법을 적용해서 가변 전송율 서비스의 호에 대역폭을 할당할 경우의 예를 그림 6에 나타내었다. 사용된 네트워크 및 호들은 그림 2와 같다. 이 경우에 3개의 링크를 모두 지나는 가변 전송율 서비스 호(긴 호)에 주어진 우선 순위 $w_s=2$ 이고, 한 개의 링크만을 지나는 가변 전송율 서비스 호(짧은 호)에 주어진 우선 순위 $w_s=1$ 이라고 가정한다. 즉, 여러 개의 링크를 지나는 긴 호에 두 배의 우선 순위를 부여한다. 그림 6을 참조하면, ν 가 10 Mbps일 경우에는 링크의 용량이 20 Mbps만 되어도 1개의 링크를 지나는 가변 전송율 서비스 호 6개에는 각각 670 kbps를, 3개의 링크를 지나는 가변 전송율 서비스 호 2개에는 각각 1.34 Mbps를 보장해 줄 수 있다. (최대-최소 공정 전송율 방법에서는 8개 호에 각각 1 Mbps가 할당됨.) ν 가 20 Mbps가 되면 35 Mbps의 링크 용량으로 각각 700 kbps, 1.4 Mbps를, ν 가 30 Mbps이면 45 Mbps의 링크 용량으로 각각 590 kbps, 1.18 Mbps 전송율을 가변 전송율 서비스 호들에 보장해 줄 수 있다. 이처럼 모든 가변 전송율 호에 동일한 전송율을 할당하지 않고, 다른 대역폭 할당 방법을 적용함으로써, 가변 전송율 서비스 호의 우선 순위에 맞게 전송율의 할당을 유연하게 변경하는 것이 가능하다.

만약 네트워크 설계의 목적이 전체 효율이나 가치(수익)를 최대화하는 것이라면 대역폭 할당을 다음과 같은 방법으로 구할 수도 있다 (비례 공정 전송율 할당 방법¹⁰⁾):

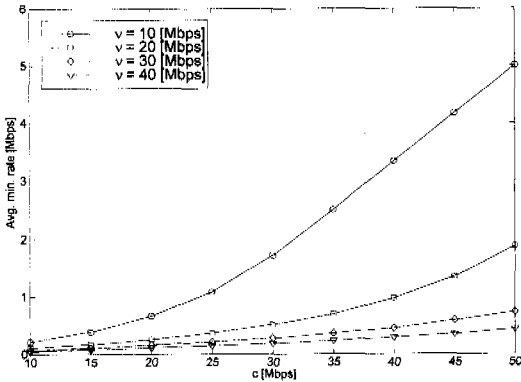


그림 6. 고정 전송율 서비스 트래픽의 도착율이 가변할 때, 링크의 용량과 가변 전송율 서비스 호에 할당될 평균 최소 전송율 (가중 최대-최소 공정 전송율 할당 방법).

$$W_m(\vec{b}) = \max_{\vec{a}} \left\{ \sum_{s \in S} u_s(a_s) \mid \sum_{s \in S} a_s \leq b_i, a_s \geq 0 \right\}, \quad (22)$$

여기서 $u_s(\cdot)$ 는 오목한 효용함수로 각 호 s 마다 주어진다. 그리고 $f(\vec{a}) = \sum_{s \in S} u_s(a_s)$ 는 오목하고, $g_i(\vec{a}) = \sum_{s \in S} a_s$ 는 볼록하며 $W_m(\vec{b})$ 는 오목하다는 것을 알 수 있다. 따라서, 전체 평균 효용은 다음과 같이 상한값을 갖는다.

$$E[W_m(\vec{B})] \leq W_m(E[\vec{B}]). \quad (23)$$

마찬가지로 대용량 네트워크일 경우에는 이 상한값에 도달할 수 있음을 보일 수 있다. (21)과 (23)을 이용하면 가변 전송율 서비스의 호에 할당되어야 할 평균 최소 전송율을 구할 수 있고, 그 성능을 만족하기 위한 링크의 용량 c_i 를 설계할 수 있다. 비례 공정 전송율 할당 방법으로 가변 전송율 서비스 호에 전송율을 할당하는 경우도 앞의 예에서처럼 성능을 분석하거나 설계하는데 활용할 수 있고, 최대-최소 공정 전송율 할당 방법 등과 달리 네트워크의 전체 효용이나 가치를 최대화하는데 이용할 수 있다.

VII. 결론

본 논문에서는 시간에 따라 가변하는 호수를 갖는 고정 전송율 서비스 트래픽과 고정된 호수를 갖는 가변 전송율 서비스 트래픽이 동시에 존재할 경우에 서비스 사용자의 성능 분석 및 네트워크의 용

량을 설계하는 방법에 대해 살펴 보았다. 최대-최소 공정 전송율 할당 방법에 따라 가변 전송율 서비스 호들에 공정하게 전송율이 할당될 때, 할당되는 평균 최소 전송율의 상한값은 먼저 가변 전송율 서비스 트래픽이 평균적으로 사용가능한 전송율을 구한 후, 최대-최소 공정 전송율 할당 방법에 의해 구해진 전송율이 됨을 보였다. 그리고, 대용량 네트워크일 경우에는 그 상한값에 도달할 수 있음을 보였다. 가변 전송율 서비스 호가 평균적으로 사용가능한 전송율은 손실 네트워크의 관점에서부터 구할 수 있었다. 이러한 결과를 이용하면, 원하는 성능을 얻기 위해 다중 서비스 네트워크 링크의 용량을 설계하는 것이 가능하다. 제안한 분석 방법을 네트워크의 예를 통해서 검증하였다.

각각의 사용자의 입장에서 최소 전송율이 최대가 되도록 하는 최대-최소 공정 전송율 할당 방법은 네트워크 운영자의 관점에서는 적합하지 않은 방법일 수 있다. 이 경우에 전체적인 네트워크의 효율을 향상시키는 방법으로, 가변 전송율 서비스 호의 효용 함수나 우선 순위를 고려 할 수 있는데, 이렇게 함으로써 전체 전송율의 합이나 전체 효용 가치가 최대가 되도록 할 수 있다 (예: 가중 최대-최소 공정 전송율 할당 방법, 비례 공정 전송율 할당 방법). 가변 전송율 서비스의 호들에 우선 순위를 부여하는 경우에, 가중 최대-최소 공정 전송율 할당 방법의 목적은 네트워크내에서 전체 성능이 최대가 되도록 하면서, 각 개개인(호)의 성능을 보장하도록 하는 것이다.

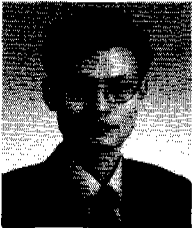
참고 문헌

- [1] F. Bonomi and K. W. Fendick, "The rate-based flow control framework for the Available Bit Rate ATM Service," *IEEE Network*, pp. 25-29, Mar/Apr., 1995.
- [2] P. Gevros, J. Crowcroft, P. Kirstein, and S. Bhatti, "Congestion control mechanisms and the best effort service model," *IEEE Network*, vol 15, no. 3, pp. 16-26, 2001.
- [3] G. de Veciana, T.-J. Lee and T. Konstantopoulos, "Stability and performance analysis of networks supporting elastic services," *IEEE/ACM Transactions on Networking*, vol 9, no. 1, pp. 2-14, Feb. 2001.
- [4] D. Mitra, J. A. Morrison, and K. G.

- Ramakrishnan, "ATM network design and optimization: A multirate loss network framework," *IEEE/ACM Trans. on Networking*, vol. 4, no. 4, pp. 531-543, Aug. 1996.
- [5] E. M. Gafny and D. Bertsekas, "Dynamic control of session input rates in communication networks," *IEEE Transactions on Automatic Control*, vol. 29, pp. 1009-1016, 1984.
- [6] T.-J. Lee and G. de Veciana, "A decentralized framework to achieve max-min fair bandwidth allocation for ATM networks," *Proc. IEEE GLOBECOM*, 1998.
- [7] F. P. Kelly, "Loss networks," *Ann. Appl. Prob.*, vol. 1, pp. 317-378, 1991.
- [8] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison Wesley, 1989.
- [9] F. P. Kelly, "Blocking probabilities in large circuit switched networks," *Ann. Appl. Prob.*, vol. 18, pp.473-505, 1986.
- [10] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33-37, 1997.

이 태 진(Tae-Jin Lee)

정회원



1989년 2월 : 연세대학교
전자공학과(공학사)

1991년 2월 : 연세대학교
전자공학과(공학석사)

1995년 12월:University of
Michigan, Ann Arbor
(M.S.E.)

1999년 5월 : University of Texas, Austin(Ph.D.).

1999년 8월~2001년 2월 : 삼성전자 중앙연구소
책임연구원

2001년 3월~현재 : 성균관 대학교 전기전자컴퓨터
공학부 전임강사

<주관심 분야> 통신 네트워크 성능 분석 및 설계,
무선 통신 시스템