

# 패킷 교환망에서 가우스 분포 트래픽을 서비스하는 선형 시스템 접근법

## (A Linear System Approach to Serving Gaussian Traffic in Packet-Switching Networks)

정 송<sup>†</sup> 신민수<sup>\*\*</sup> 정현희<sup>\*\*\*</sup>  
(Song Chong) (Min-Su Shin) (Hyun-Hee Chong)

**요약** 이 논문에서는 자원을 공유하는 여러 개의 QoS(Quality of Service) 큐(queue)를 서비스하기 위한 새로운 서비스 규칙 - 선형 서비스 규칙을 제안하고, 그 특징을 분석하였다. 제안하는 선형 서버는 각각의 큐에 대한 출력 트래픽(traffic) 및 고객 수 과정을 입력 트래픽의 선형 함수로 만든다. 특히 입력 트래픽이 가우스 분포를 갖는 경우에는 큐 길이의 분포와 출력 트래픽 분포가 모두 가우스 분포를 갖게 하며, 그 분포의 평균과 분산이 입력 트래픽의 평균과 전력 스펙트럼(power spectrum)의 함수로 나타나게 한다. 중요한 QoS 척도인 버퍼 넘침 확률 및 지연 분포 역시 입력 트래픽의 평균과 전력 스펙트럼의 함수로 나타나게 된다. 이 연구는 네트워크의 각 노드를 하나의 선형 필터로 볼 수 있게 하므로, 선형 시스템 이론에 기초한 네트워크 전반에 걸친 트래픽 관리 기술의 새로운 방향을 제시하였다.

**키워드** : 서비스 품질, 서비스 방안, 가우스 분포 트래픽, 선형 시스템, 스펙트럼 분석

**Abstract** We present a novel service discipline, called linear service discipline, to serve multiple QoS queues sharing a resource and analyze its properties. The linear server makes the output traffic and the queueing dynamics of individual queues as a linear function of its input traffic. In particular, if input traffic is Gaussian, the distributions of queue length and output traffic are also Gaussian with their mean and variance being a function of input mean and input power spectrum (equivalently, autocorrelation function of input). Important QoS measures including buffer overflow probability and queueing delay distribution are also expressed as a function of input mean and input power spectrum. This study explores a new direction for network-wide traffic management based on linear system theories by letting us view the queueing process at each node as a linear filter.

**Key words** : Quality of Service, Service Discipline, Gaussian Traffic, Linear System, Spectral Analysis

### 1. 서론

이더넷(Ethernet) 트래픽의 자기닮음 특성의 발견은 트래픽 모델링 연구에서 중요한 이정표로 여겨지고 있다. 이후 인터넷 트래픽과 비디오 트래픽 역시 자기 닮음 특성을 가진 것으로 보고 되고 있으며, 이러한 발견은 효과적인 트래픽 관리를 위해 사용되고 있다[1]

[2][3].

그러나 빠르게 진행되고 있는 인터넷 트래픽 양의 증가는 링크 용량의 증가를 가져온다. 일반적으로 링크 용량이 증가할수록 하나의 링크가 포함하는 플로우(flow)의 수는 많아지며, 이러한 사실은 자연스럽게 중심극한 정리를 통해, 집합된 트래픽이 많아질수록 입력 트래픽의 분포가 가우스 분포와 더욱 유사하게 되는 것을 뜻한다. 현재의 트래픽이 가우스 분포와 차이가 있는 이유는 아직 집적되는 플로우의 수가 트래픽을 가우스 분포에 근접시킬 만큼 많지 않기 때문이며, 자기닮음 특성을 갖는 트래픽인 M/Pareto 모델로 생성된 트래픽의 경우 많이 집적될수록 가우스 분포 특성에 가까워지는 것을

† 정 회 원 : 한국과학기술원 전자전산학과 교수  
song@ee.kaist.ac.kr

\*\* 비 회 원 : 한국과학기술원 전자전산학과  
msshin@netsys.kaist.ac.kr

\*\*\* 비 회 원 : 삼성전자 네트워크사업부 연구원  
milou@samsung.com

논문접수 : 2002년 1월 12일

심사완료 : 2002년 6월 21일

볼 수 있다는 주장도 있다[4].

따라서 엄밀한 정확성 때문에 분석이 용이한 가우스 확률과정을 이용한 트래픽 모델링을 포기해야 하는가 하는 의문을 가질 수 있다. 만약 가우스 확률과정이 납득할 수 있는 정도의 정확도로 링크 내에 집합된 트래픽을 모델링할 수 있다면, 실용적인 인터넷 QoS의 방법으로 가우스 확률과정 모델링이 가능하다. 특히 현재 인터넷 QoS의 방향인 차등 서비스(differentiated service)의 경우 정확한 QoS 보장보다, QoS 차등화가 더 중요한 문제로 여겨지고 있으며, 가우스 확률과정 모델링이 QoS 차등화를 수행할 수 있는 정도의 정확성을 가지고 있다면 가우스 확률과정을 이용한 트래픽 모델링이 실제로 사용 가능할 것이다.

본 논문에서는 링크의 입력 트래픽이 시불변 가우스 과정이라고 가정한다. 이 논문에서 우리가 다룰 문제는 다음과 같다. i) 버퍼를 통과하는 트래픽의 가우스 분포 특성을 유지시키는 선형 서비스 규칙이 존재하는가? 그리고 ii) 이와 연관되는 대기행렬의 성능 - 큐 길이 분포, 버퍼 넘침 확률, 지연 현상 등이 선형 시스템 이론을 기반으로 쉽게 분석될 수 있는가?

일반적인 고정 서비스 용량을 가진 대기행렬 시스템은 바쁜 기간과 유향 기간을 반복하게 되므로 비선형성을 가지게 된다. 따라서 대기행렬의 분석이 매우 힘들어지며, 네트워크 전체의 대기행렬 문제를 풀기에는 적합하지 않다. QoS 보장이나 차등 서비스와 관련하여 많은 수의 서비스 규칙들이 제안되었으나, 어떤 서비스 규칙도 바쁜 기간과 유향 기간의 반복으로 인한 비선형성을 갖지 않는 것은 없다[5]-[10].

본 논문에서는 여러 개의 QoS 큐를 서비스하는 선형 서비스 규칙을 제안한다. 이 규칙은 입력 트래픽의 선형 함수 형태로 각 큐에 서비스 양을 할당한다. 선형 서버는 고객 수 과정보다 이탈 과정을 모두 입력 트래픽의 선형 함수 형태로 나타내며, 그 결과 큐의 길이, 지연, 출력 트래픽 분포 등의 통계량을 입력 트래픽의 통계량의 함수로 나타낸다. 특히 입력 트래픽이 가우스 분포를 갖는 경우에는 QoS 계산을 위해서 측정되어야 할 통계량은 단지 입력 트래픽의 평균값과 전력 스펙트럼(power spectrum)이 된다.

## 2. 네트워크 디자인

### 2.1 대기행렬 구조와 가우스 분포 입력

그림 1과 같이 각 출력 링크에는 논리적으로 분리된  $N$ 개의 QoS 큐가 있으며 하나의 최선형(best effort) 큐가 있다고 가정한다. 이러한 대기행렬 구조는 최근의

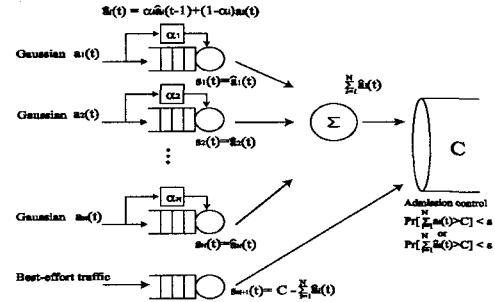


그림 1 네트워크 노드 모델

QoS를 고려한 스위치/라우터의 설계에 있어서 표준이며, 클래스별 큐를 지원하므로 입력 플로우의 수가 증가하여도 스케줄링의 복잡도가 증가하지 않아 확장이 용이하다. 또한 동일한 QoS 요구를 갖는 입력 플로우가 하나의 큐에 집합되므로, 집합된 입력 트래픽에 대해 통계량의 조절과 측정이 용이해진다. 예를 들어, 집합되는 플로우의 수가 많아질수록 그 트래픽은 중심극한정리에 의해 더욱 가우스 분포와 유사해지게 된다. 클래스별 대기행렬구조의 예로서는 인터넷 QoS를 위한 차등서비스 모델을 들 수 있다[11].

$i$  번째 QoS 큐에 시간 간격  $[t, t+1)$  에 들어오는 패킷의 수(또는 바이트 수)를  $a_i(t)$  라고 하며,  $a_i(t)$  는 중심극한정리에 의해 시불변 가우스 확률 과정이라 가정한다.  $a_i(t)$  의 시간과 무관한 통계량으로는 입력의 평균과 분산이, 시간에 관계있는 통계량으로는 자기상관(auto correlation)과 같은 고차 통계량이 있다.

실제로,  $a_i(t)$  의 분포는 관측 간격이 길어질수록 더욱 가우스 분포에 근접해지나, 이 경우 간격 내 짧은 시간 동안의 패킷 도착에 대한 정보를 잃게 된다.

### 2.2 서비스 규칙

$s_i(t)$  를  $i$  번째 큐에 대한 시간  $t$ 에서의 서비스 양이라고 하자. 즉,  $s_i(t)$  는 시간 간격  $[t, t+1)$  사이에 서비스 되어야 할 패킷의 수(또는 바이트 수)를 뜻한다. 우리가 제안하는 새로운 서비스 규칙은 다음과 같다.

우선 각 QoS 큐에 대해서  $s_i(t) = \hat{a}_i(t)$ ,  $\forall t, i=1, \dots, N$  이며,  $\hat{a}_i(t)$  는 아래와 같이 저대역 필터 식으로 정한다.

$$\hat{a}_i(t) = \alpha_i \hat{a}_i(t-1) + (1 - \alpha_i) a_i(t), \quad 0 < \alpha_i < 1 \quad (1)$$

이 때,  $\hat{a}_i(-1) = 0$ 으로 한다.

최선형 큐에 대해서  $s_{N+1}(t) = C - \sum_{i=1}^N \hat{a}_i(t)$ ,  $\forall t$  이며, 출력 링크의 용량  $C$  에 대해서 QoS 큐가 소비하고 남은 용량을 사용하는 것을 뜻한다.

제안하는 서비스 규칙의 특징은 각 큐의 서비스 양이 입력 트래픽의 양에 따라 적응적으로 변화한다는 것이다. 또한 각 QoS 큐가  $s_i(t)$ 만큼의 서비스량을 항상 제공 받기 위해서는  $\Pr[\sum_{i=1}^N \hat{a}_i(t) > C] < \epsilon$ 와 같은 입력 통계가 필요하며, 여기에서  $\epsilon$ 은 충분히 작은 양의 값이다.

제안된 서비스 규칙은 최선형 큐가 없다고 가정한다면, 비 일량 보존 특성(non work-conserving)을 갖게 된다. 즉, 출력링크의 전체 용량  $C$  중에, QoS 큐에서 사용한  $\sum_{i=1}^N \hat{a}_i(t)$ 만이 사용되게 되며, 나머지 양인  $C - \sum_{i=1}^N \hat{a}_i(t)$ 은 QoS 큐가 비어있는 것과 상관없이 낭비되게 된다. 일반적으로 일량 보존 서버는 같은 용량의 비 일량 보존 서버에 비해 더 많은 패킷을 처리할 수 있으나, 비 일량 보존 서버는 자격이 있는 패킷에 대해서만 서비스를 수행하므로 더 엄격한 제어가 가능하다는 장점이 있다. 제안된 선형 서비스 규칙은 트래픽의 가우스 분포 특성이 네트워크 노드를 이동하면서도 유지되도록 트래픽을 제어하는 특징이 있다.

$i$ 번째 QoS 큐의 시간  $t$ 에서의 길이를  $q_i(t)$ 라고 하면,  $q_i(t)$ 의 변화는 다음과 같은 식으로 나타낼 수 있다.

$$q_i(t+1) = [q_i(t) + a_i(t) - \hat{a}_i(t)]^+ \quad (2)$$

단,  $[x]^+ = \max[0, x]$ 이다.

식 (1)의 서비스량 계산에서 계수  $\alpha_i$ 는 저대역 필터의 정도를 결정하고, 저대역 필터의 컷오프 주파수  $f_c$ 는 식 (1)에 의해  $\cos(2\pi f_c \Delta) = 1 - \frac{(1 - \alpha_i)^2}{2\alpha_i}$ 를 만족하는 값이다.  $\Delta$ 는 우리의 모델에서 사용되는 이산 모델의 최소 시간 단위이다.  $\alpha_i$ 가 1에 가까워질수록  $f_c$ 는 0에 가까워지는 것을 알 수 있는데, 이는  $\alpha_i$ 의 값이 증가할수록 필터의 통과대역이 점점 좁아지는 것을 뜻한다.

$a_i(t)$ 와  $\hat{a}_i(t)$ 는 평균이 같은 가우스 분포를 가지며, 분산은  $\hat{a}_i(t)$ 가  $a_i(t)$ 에 비해 더 적거나 같기 때문에 입력 통계 규칙인  $\Pr[\sum_{i=1}^N \hat{a}_i(t) > C] < \epsilon$ 은  $\Pr[\sum_{i=1}^N a_i(t) > C] < \epsilon$ 로 강화될 수 있다. 그러나 이러한 입력 통계의 강화는, 플로우의 블로킹 확률(blocking probability)의 증가를 가져온다.

### 3. 네트워크 해석

이 절에서는 입력 트래픽이 가우스 분포를 가질 경우, 큐의 길이와 출력 트래픽 모두 가우스 분포를 갖는 것을 보이며, 더 나아가서 큐의 길이와 출력 트래픽의 평균 및 분산값이 입력 트래픽의 평균과 전력 스펙트럼의 명시적인 함수로 나타내어짐을 보인다. 또한 큐 내에서

의 지연 분포와 버퍼 넘침 확률 역시 입력 트래픽의 평균과 전력 스펙트럼의 함수로 나타낼 수 있음을 증명한다.

#### 3.1 고객 수 과정

식 (1), (2)를 이용해 시간  $t$ 에서의 고객 수  $q(t)$ 를 반복적으로 유도하면,  $q(t)$ 는 다음과 같이 나타낼 수 있다.

$$q(t) = [q(t-1) + a(t-1) - \hat{a}(t-1)]^+ = a^t a(0) + a^{t-1} a(1) + \dots + a a(t-1) \quad (3)$$

시간  $t$ 에서의 큐의 길이는 기하적으로 감소하는 가중치를 곱한 입력의 합으로 나타내진다. 이 식은  $q(t)$ 가 가우스 분포를 갖는다는 것과  $t > 0$ 일 때, 항상  $q(t) > 0$ 이라는 사실을 알려준다. 즉 큐가 비어있는 시간이 없게 되며, 이를 토대로 하여 식 (2)는 다음과 같이 바뀔 수 있다.

$$q_i(t+1) = q_i(t) + a_i(t) - \hat{a}_i(t), \quad \forall t > 0 \quad (4)$$

식 (4)는 이러한 대기행렬 시스템이 선형성을 갖는다는 것을 나타낸다. 고정용량을 갖는 보통의 대기행렬 시스템은 바쁜 기간과 유향 기간을 반복하므로 비선형성을 갖는다. 그 결과, 고객 수 과정과 이탈 과정의 해석은 매우 복잡하게 되고, 그 해석 결과가 네트워크 전반에 걸친 대기행렬 문제를 해결하는 것으로 확장되기는 매우 어렵다. 이에 반해 제안된 선형 서비스 규칙은 네트워크 노드를 하나의 선형 시스템으로 볼 수 있으며 입력 트래픽이 가우스 분포일 경우, 고객 수 과정과 이탈 과정 역시 가우스 분포를 따른다. 즉 풍부한 선형 시스템의 이론들이 네트워크 전반에 걸친 대기행렬 문제에 적용될 수 있음을 뜻한다.

다음으로, 가우스 분포를 갖는 고객 수과정의 평균과 분산을 구한다.  $\bar{q}$ ,  $\sigma_q^2$ ,  $\bar{a}$ 는 각각 큐의 길이의 평균과 분산, 그리고 입력 트래픽의 평균을 의미한다. 즉  $\bar{q} = \lim_{t \rightarrow \infty} E[q(t)]$ ,  $\sigma_q^2 = \lim_{t \rightarrow \infty} E[(q(t) - E[q(t)])^2]$ ,  $\bar{a} = \lim_{t \rightarrow \infty} E[a(t)]$ 이다.  $Q(z)$ 와  $A(z)$ 는 각각  $q(t)$ 와  $a(t)$ 의  $z$ 변환을 뜻한다. 식 (1), (4)와  $z$ 변환을 이용하여 다음과 같은 전달 함수를 얻을 수 있다.

$$H_q(z) = \frac{Q(z)}{A(z)} = \frac{az^{-1}}{1 - az^{-1}} \quad (5)$$

$z = e^{j\omega}$ 로 치환하면,

$$H_q(\omega) = \frac{ae^{-j\omega}}{1 - ae^{-j\omega}} \quad (6)$$

선형 시스템 이론을 이용하면, 우리는 다음과 같은  $\bar{q}$ 와  $\sigma_q^2$ 을 얻을 수 있다.

$$\bar{q} = H_q(0) \bar{a} = \frac{\bar{a}}{1 - \bar{a}} \quad (7)$$

$$\sigma_a^2 = \frac{1}{\pi} \int_0^\pi P_q(\Omega) d\Omega - \bar{q}^2 \quad (8)$$

$$= \frac{1}{\pi} \int_0^\pi |H_q(\Omega)|^2 P_a(\Omega) d\Omega - \bar{q}^2 \quad (9)$$

$$|H_q(\Omega)|^2 = \frac{\alpha^2}{1 + \alpha^2 - 2\alpha \cos \Omega}.$$

$P_q(\Omega)$ 와  $P_a(\Omega)$ 는 각각  $q(t)$ 와  $a(t)$ 의 전력 스펙트럼을 의미한다.

고객 수 과정의 해석 결과는 두 가지를 의미한다. 먼저 고객 수 과정의 가우스 분포는 주어진  $\alpha$ 에 대해 입력 트래픽의 평균과 전력 스펙트럼으로 완벽하게 나타낼 수 있으며, 두 번째로 주어진 입력 트래픽에 대해 적당한 계수  $\alpha$ 를 선택함으로써 고객 수 과정의 평균과 분산을 원하는 대로 제어할 수 있다는 것이다. 입력 트래픽의 평균과 전력 스펙트럼을 실시간으로 측정하는 것에는 이미 많은 방법들을 사용할 수 있으며, 한 예를 4절에서 보이고, 모의실험에 사용할 것이다.

### 3.2 버퍼 넘침 확률

고객 수 과정의 분포를 알았을 경우, 우리는 주어진 버퍼 크기  $B$ 에 대한 버퍼 넘침 확률을 예측할 수 있다. 이를 위해 우선 체르노프 상한(Chernoff's bound)을 이용한다[12].

$$\Pr[q > B] \leq e^{-\theta B} E[e^{\theta q}], \quad \theta > 0 \quad (10)$$

여기에서  $E[e^{\theta q}]$ 는  $q$ 의 모멘트 생성 함수이다.  $q$ 는 가우스 분포를 갖는 확률 변수이므로,  $\ln E[e^{\theta q}] = \bar{q}\theta + \frac{1}{2}\sigma_a^2\theta^2$ 이며, 상한은 다음과 같이 나타난다.

$$\ln \Pr[q > B] \leq \min_{\theta > 0} (-\theta B + \bar{q}\theta + \frac{1}{2}\sigma_a^2\theta^2) \quad (11)$$

식 (11)의 오른쪽 항은  $\theta = \frac{B - \bar{q}}{\sigma_a^2}$  일 때, 최소값을 가지므로 다음과 같은 버퍼 넘침 확률을 계산할 수 있다.

$$\ln \Pr[q > B] \sim -\frac{(B - \bar{q})^2}{2\sigma_a^2} \quad (12)$$

버퍼 크기  $B$ 에 대해 식 (12)를 통해, 목표로 하는 버퍼 넘침 확률을 갖기 위해 필요로 하는 버퍼 크기를 구할 수 있다.

$$B \sim \bar{q} + K\sigma_a \quad (13)$$

여기에서,  $K = \sqrt{-2 \ln \Pr[q > B]}$ 이다.

문헌 [13]의 개선된 큰 편차 근사(refined large deviation approximation)에 의하면,

$$\ln \Pr[q > B] \sim -\frac{(B - \bar{q})^2}{2\sigma_a^2} + \ln\left(\frac{\sigma_a}{\sqrt{2\pi(B - \bar{q})}}\right). \quad (14)$$

이며,  $K$ 는  $K^2 + 2 \ln(\Pr[q > B] \sqrt{2\pi} K) = 0$  을 만족하는 값이다.

또 다른 근사로서 문헌 [14]에 따르면,

$$\ln \Pr[q > B] \sim -\frac{(B - \bar{q})^2}{2\sigma_a^2} - \ln(\sqrt{2\pi}) \quad (15)$$

이 경우,  $K = \sqrt{-2 \ln \Pr[q > B] - \ln(\sqrt{2\pi})}$ 이다.

### 3.3 이탈 과정

큐가 유희기간을 가지지 않기 때문에, 시간  $t$ 에서 출력 트래픽은  $\hat{a}(t)$ 의 값으로 이탈하게 된다.  $\bar{a}$ 와  $\sigma_a^2$ 은 각각 정상상태에서  $\hat{a}(t)$ 의 평균과 분산을 뜻한다. 즉,  $\bar{a} = \lim_{t \rightarrow \infty} E[\hat{a}(t)]$ ,  $\sigma_a^2 = \lim_{t \rightarrow \infty} E[(\hat{a}(t) - E[\hat{a}(t)])^2]$ 이다.  $\hat{A}(z)$ 를  $\hat{a}(t)$ 의  $z$ -변환이라 하면, 식 (1)의  $z$ -변환을 이용하여 다음과 같은 전달 함수를 구할 수 있다.

$$H_a(z) = \frac{\hat{A}(z)}{A(z)} = \frac{1 - \alpha}{1 - \alpha z^{-1}} \quad (16)$$

$z = e^{j\Omega}$ 로 치환하면,

$$H_a(\Omega) = \frac{1 - \alpha}{1 - \alpha e^{-j\Omega}} \quad (17)$$

선형 시스템 이론을 이용하면, 우리는 다음과 같은  $\bar{a}$ 와  $\sigma_a^2$ 을 얻을 수 있다.

$$\bar{a} = H_a(0) \bar{a} = \bar{a} \quad (18)$$

$$\sigma_a^2 = \frac{1}{\pi} \int_0^\pi P_a(\Omega) d\Omega - \bar{a}^2 \quad (19)$$

$$= \frac{1}{\pi} \int_0^\pi |H_a(\Omega)|^2 P_a(\Omega) d\Omega - \bar{a}^2$$

$$|H_a(\Omega)|^2 = \frac{(1 - \alpha)^2}{1 + \alpha^2 - 2\alpha \cos \Omega}. \quad (20)$$

$P_a(\Omega)$ 는  $\hat{a}(t)$ 의 전력 스펙트럼을 의미한다.

### 3.4 지연 과정

$d(t)$ 를 시스템에 시간  $t$ 에 도착한 패킷이 겪게 되는 지연이라 하면, 그림 2에서 볼 수 있듯이, 주어진 상수  $\tau$ 에 대해 다음과 같은 관계식이 성립된다.

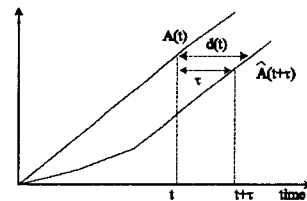


그림 2 시스템의 지연

$$\Pr[d(t) > \tau] = \Pr[A(t) - \hat{A}(t + \tau) > 0] \quad (21)$$

위 식에서  $A(t) = \sum_{k=0}^{t-1} a(k)$ ,  $\hat{A}(t) = \sum_{k=0}^{t-1} \hat{a}(k)$ 이다. 식 (21)은 시간  $t$ 에 도착한 패킷이 시간  $t + \tau$  이후에 시스템을 떠날 확률과 확률 변수  $A(t) - \hat{A}(t + \tau)$ 가 0보다

클 확률이 서로 같다는 것을 뜻한다.  $a(t)$ 와  $\hat{a}(t)$ 가 모두 가우스 분포를 갖는 확률 변수이므로,  $A(t) - \hat{A}(t+\tau)$  역시 가우스 분포를 갖는 확률 변수가 된다.  $A(t) - \hat{A}(t+\tau)$  는 다른 변수들을 이용하여 다음과 같이 나타낼 수 있다.

$$A(t) - \hat{A}(t+\tau) = \begin{cases} a(t) - \sum_{k=t}^{t+\tau-1} \hat{a}(k), & \tau > 0 \\ a(t), & \tau = 0. \end{cases} \quad (22)$$

$\bar{r}_\tau$ 와  $\sigma_{r_\tau}^2$ 은 각각  $r_\tau(t) \equiv A(t) - \hat{A}(t+\tau)$ 의 평균과 분산을 표시한다.  $R_\tau(z)$ 를  $r_\tau(t)$ 의  $z$ -변환이라 하면, 식 (1), (22)와  $z$ -변환을 사용하여,  $\tau > 0$ 일 때, 다음과 같은 전달 함수를 얻을 수 있다.

$$H_{r_\tau}(z) = \frac{R_\tau(z)}{A(z)} = \begin{cases} \frac{az^{-1} - (1-a)(1-z^\tau)/(1-z)}{1-az^{-1}}, & z \neq 1 \\ \frac{a}{1-a} - \tau, & z = 1 \end{cases} \quad (23)$$

$z = e^{j\Omega}$ 로 치환하면,

$$H_{r_\tau}(\Omega) = \begin{cases} \frac{ae^{-j\Omega} - (1-a)(1-e^{j\Omega\tau})/(1-e^{j\Omega})}{1-ae^{-j\Omega}}, & \Omega \neq 0 \\ \frac{a}{1-a} - \tau, & \Omega = 0 \end{cases} \quad (24)$$

선형 시스템 이론을 이용하면, 다음과 같은  $\bar{r}_\tau$ 와  $\sigma_{r_\tau}^2$ 을 얻을 수 있다.

$$\bar{r}_\tau = H_{r_\tau}(0) \bar{a} = \left( \frac{a}{1-a} - \tau \right) \bar{a} \quad (25)$$

$$\sigma_{r_\tau}^2 = \frac{1}{\pi} \int_0^\pi P_{r_\tau}(\Omega) d\Omega - \bar{r}_\tau^2 = \frac{1}{\pi} \int_0^\pi |H_{r_\tau}(\Omega)|^2 P_a(\Omega) d\Omega - \bar{r}_\tau^2 \quad (26)$$

$$|H_{r_\tau}(\Omega)|^2 = \frac{a^2 - 2a(1-a) \sum_{k=1}^{\tau-1} \cos k\Omega + (1-a)^2 \frac{1 - \cos \Omega\tau}{1 - \cos \Omega}}{1 + a^2 - 2a \cos \Omega} \quad (27)$$

식 (26)에서  $P_{r_\tau}(\Omega)$ 는  $r_\tau(t)$ 의 전력 스펙트럼이다.

체르노프 상한과 큰 편차 근사를 이용하여,  $\Pr\{d \geq \tau\}$ 에 대한 다음과 같은 근사값을 얻을 수 있다.

$$\ln \Pr\{d \geq \tau\} = \ln \Pr\{r_\tau > 0\} \sim -\frac{\tau^2}{2\sigma_{r_\tau}^2} \quad (28)$$

지연값  $d$ 는 이산 값을 갖기 때문에 식 (27)과 (28)을 사용하면 다음과 같은  $d$ 의 분포를 얻을 수 있다.

$$\Pr\{d = \tau\} = \frac{\Pr\{d \geq \tau - 1\} - \Pr\{d \geq \tau\}}{\Pr\{r_{\tau-1} > 0\} - \Pr\{r_\tau > 0\}} \sim e^{-\frac{\tau-1}{2\sigma_{r_{\tau-1}}^2}} - e^{-\frac{\tau}{2\sigma_{r_\tau}^2}} \quad (29)$$

리틀의 정리(Little's theorem)와 식 (7)을 이용하면,

우리는 다음과 같이 지연의 평균값은 입력 트래픽의 특성과 무관하게  $a$ 에 의해서만 결정됨을 알 수 있다.

$$\bar{d} = \frac{\bar{a}}{a} = \frac{a}{1-a} \quad (30)$$

### 3.5 입력 제어

각 QoS 류가 서비스 양  $\hat{a}_i(t)$ 를 항상 서비스할 수 있으려면,  $\Pr\{\sum_{i=1}^N \hat{a}_i(t) > C\} < \epsilon$  과 같은 입력 통제 규칙이 존재해야 한다. 여기에서  $\epsilon$ 은 충분히 작은 양의 값이다. 또한  $a_i(t)$ 와  $\hat{a}_i(t)$ 는 평균이 같은 가우스 분포를 가지며, 분산은  $\hat{a}_i(t)$ 가  $a_i(t)$ 에 비해 더 적거나 같기 때문에 플로우의 블로킹 확률을 증가시키면서  $\Pr\{\sum_{i=1}^N \hat{a}_i(t) > C\} < \epsilon$  은  $\Pr\{\sum_{i=1}^N a_i(t) > C\} < \epsilon$ 로 강화될 수 있다.

$\sum_{i=1}^N a_i = S$ ,  $\sum_{i=1}^N \hat{a}_i = \hat{S}$  라고 하면, 체르노프 상한과 큰 편차 근사를 이용하여, 입력 통제 규칙은 다음과 같이 근사화 될 수 있다.

$$\ln \Pr\{S > C\} \sim -\frac{(C - \bar{S})^2}{2\sigma_S^2} + \ln\left(\frac{\sigma_S}{\sqrt{2\pi(C - \bar{S})}}\right) < \ln \epsilon \quad (31)$$

$$\ln \Pr\{\hat{S} > C\} \sim -\frac{(C - \bar{\hat{S}})^2}{2\sigma_{\hat{S}}^2} + \ln\left(\frac{\sigma_{\hat{S}}}{\sqrt{2\pi(C - \bar{\hat{S}})}}\right) < \ln \epsilon \quad (32)$$

$\bar{S}$ 와  $\sigma_S^2$ 은 각각 입력 트래픽으로부터 측정할 수 있는 값이며,  $\bar{\hat{S}} = \bar{S}$ 이다. 또한 만약 모든  $i$ 에 대해서  $a_i = \alpha$ 라면,

$$\sigma_S^2 = \sigma_S^2 - \frac{1}{\pi} \int_0^\pi P_S(\Omega) \frac{2a(1 - \cos \Omega)}{1 + a^2 - 2a \cos \Omega} d\Omega \quad (33)$$

위 식에서  $P_S(\Omega)$ 는  $\sum_{i=1}^N a_i$ 의 전력 스펙트럼이며, 역시 입력 트래픽으로부터 측정 가능하다.

## 4. 입력 트래픽 측정

지금까지 우리는 여러 가지 QoS의 계산을 위해서 필요한 것들이 오직 입력 트래픽의 평균과 전력 스펙트럼이라는 것을 보였다. 이 절에서는 우리가 모의실험에 사용한 입력 트래픽의 측정 방법들을 제시한다.

### 4.1 평균

$M$ 개의 샘플들에 대해 우리는 입력 평균을 다음과 같이 측정하였다.

$$\bar{a} = \frac{1}{M} \sum_{n=0}^{M-1} a(n) \quad (34)$$

### 4.2 전력 스펙트럼

전력 스펙트럼의 측정에서는 좀 더 일관된 결과를 위해, 잘 알려진 피리어도그램 방법(periodogram) 대신

Welch의 방법(Welch's method)을 사용하였다[15]. M개의 샘플에 대해 전력 스펙트럼은 다음과 같이 측정된다.

$$P_a(\Omega) = \frac{1}{KLU} \left| \sum_{n=0}^{K-1} w(n) a(n+iD) e^{-i\Omega n} \right|^2 \quad (33)$$

위 식에서  $M=L+D(K-1)$ ,  $U = \frac{1}{M} \sum_{n=1}^{K-1} |w(n)|^2$  이며,  $w(n)$ 으로 해닝 윈도우(hanning window)를 사용하였다.

### 5. 모의실험 결과

본 논문의 해석적 결과를 검증하기 위하여, 선형 서버의 규칙을 따르는 하나의 큐 시스템에 대해 모의실험을 하였다. 버퍼의 크기는 무한대이며, 패킷의 크기는 일정하다는 가정 하에 세 가지 입력 트래픽이 사용되었다.

트래픽 A는 100개의 균일한 이산시간 ON-OFF 소스의 중첩이며, 각 소스의 활성 인자(activity factor)는 0.5, ON 상태에 머물 확률은 0.2이며, ON 상태에서 트래픽 발생량은 1 [패킷/Δ]이다. 따라서 트래픽 A는 이항 분포(binomial distribution)를 가지며, 평균과 분산은 각각 20 [패킷/Δ], 16 [패킷<sup>2</sup>/Δ<sup>2</sup>]이 되는 것을 알 수 있다[16]. 많은 수의 ON-OFF 소스가 중첩될수록, 그 이항 분포는 가우스 분포에 가까워짐은 잘 알려져 있다[12].

트래픽 B는 트래픽 A에 또 다른 400개의 균일한 ON-OFF소스의 중첩을 합한 것으로서, 이때 ON-OFF 소스는 활성 인자가 0.4, ON 상태에 머물 확률은 0.05, ON 상태에서 트래픽 발생량은 1 [패킷/Δ]를 가진다. 트래픽 B의 분포는 두 개의 서로 다른 이항 분포에 대한 합성곱(convolution) 형태를 가지며, 이 분포 역시 중첩되는 소스의 수가 많아질수록 중심극한정리에 의해 가우스 분포를 갖는다는 것이 알려져 있다.

트래픽 C는 NLNR[17]에 의해 수집된 10초 길이의 실제 인터넷 트래픽이다. 트래픽 C에서 우리는 Δ=1[ms]로 하고, 하나의 패킷이 40 바이트의 길이를 갖게끔 하였다. 인터넷 트래픽의 평균과 분산은 각각

51.6 [패킷/Δ], 533.0 [패킷<sup>2</sup>/Δ<sup>2</sup>]이며, 그림 3에서 인터넷 트래픽의 분포를 평균과 분산이 같은 가우스 분포와 비교하였다.

그림 4)는 모의실험에서 얻어진 큐 길이( $q$ ), 출력 트래픽( $\hat{a}$ ), 그리고 패킷 지연( $d$ )에 대한 평균과 분산 값을 해석적 결과와 비교하여 보여준다. 해석적 결과를 얻기 위하여, 트래픽 A와 트래픽 B에 대해서는 입력 트래픽 모델로부터 직접 구한 참값을 평균과 전력 스펙트럼에 사용하였고, 트래픽 C에서는 식 (36)과 식 (37)의 입력 트래픽 측정방법을 사용하여 측정된 평균과 전력 스펙트럼을 사용하였다. 전체적으로 모의실험 결과는 계수  $\alpha$ 와 상관없이 해석 결과와 거의 일치함을 보인다. 이는 사용된 트래픽이 가우스 분포를 갖는 경향을 보이며, 이에 따라 큐 길이와 출력 트래픽 역시 가우스 분포를 갖는 것을 의미한다. 그림 4의 (가)와 (나)는 각각  $\alpha$ 가 증가함에 따라 큐 길이의 평균과 분산이 증가하는 것을 보여준다. 이 결과는 식 (7) 과 식 (8)에서 증명한 바와 같이 서버의 통과대역이 좁아짐에 따라 나타나는 결과이다. 그림 4의 (다)와 (라)는 각각 출력 트래픽의 평균과 분산을 보이며, 출력 트래픽의 평균은 계수  $\alpha$ 에 상관없이 입력 트래픽의 평균과 같고 (트래픽 A : 20 [패킷/Δ], 트래픽 B : 40 [패킷/Δ], 트래픽 C : 51.6 [패킷/Δ]), 분산은  $\alpha$ 가 증가할수록 감소하는 것을 볼 수 있다. 이는 식 (18)과 식 (19)에서 증명한 결과와 일치한다. 마지막으로 그림 4의 (마)와 (바)는 패킷 지연의 평균과 분산을 보이며, 평균 지연 값은 식 (32)에서 증명한 바와 같이 트래픽 종류에 무관하며,  $\alpha$ 가 증가할수록 증가하는 것을 볼 수 있다. 반면에 패킷 지연의 분산은 각 트래픽 종류에 따라 다른 값을 보인다.

그림 5의 (가)와 (나)는 큐 길이 분포와 출력 트래픽의 분포를 보여준다. 도시한 분포는 모의실험의 결과 중, 트래픽 B에서  $\alpha=0.80$  일 때의 결과이다. 예상한 바와 같이 각 분포는 평균과 분산이 같은 가우스 분포에 매우 근접한다. 그림 5의 (다)는 시간에 따른 큐 길이의 변화를 도시하였다. 이 결과는 선형 서버에 의해 버퍼가 유희상태에 가지 않음을 보여준다.

모의실험을 통해 요구되는 버퍼의 크기  $B = \bar{q} + K\sigma_q$ 를 버퍼 넘침 확률의 목표 값의 함수로 구하였다. 여기에서  $\bar{q}$ 와  $\sigma_q$ 는 큐에 들어오는 입력 트래픽과 계수  $\alpha$ 에 의해 결정되는 값이다. 그림 5의 (라)는 모의실험에 의해 얻어진 버퍼의 크기가 식 (13), 식 (14), 식 (15)의 요구 버퍼 크기 상한에 대한 해석적 근사값과 잘 일치함을 보여준다.

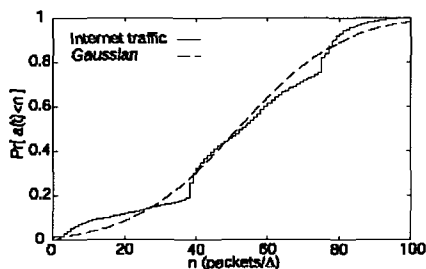


그림 3 인터넷 트래픽의 분포

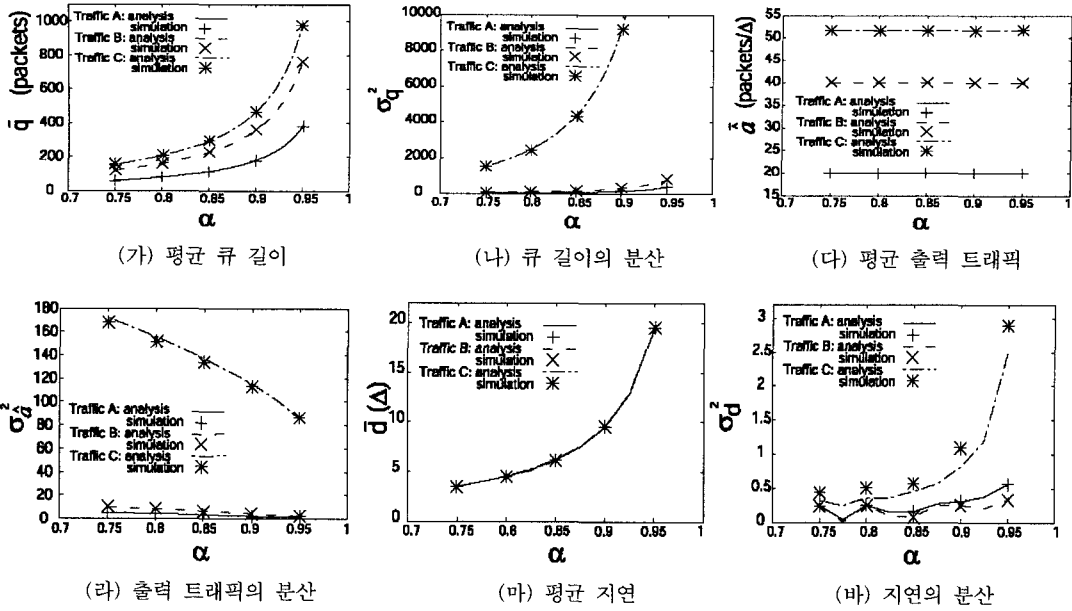


그림 4 모의실험 결과와 분석 결과의 비교

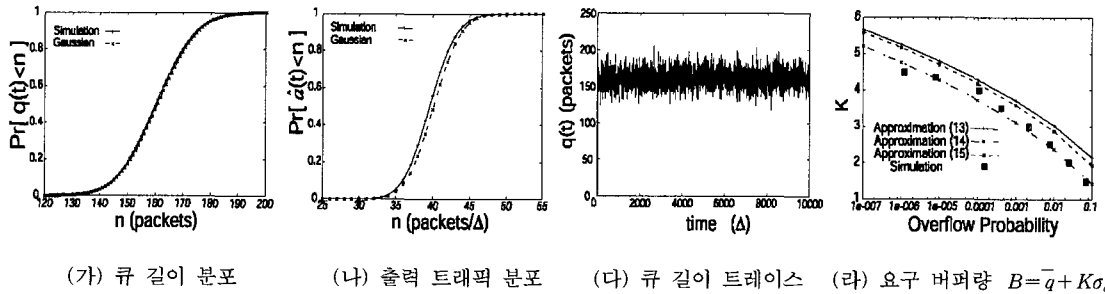


그림 5 모의실험과 분석 결과의 비교 (트래픽 B,  $\alpha=0.80$ )

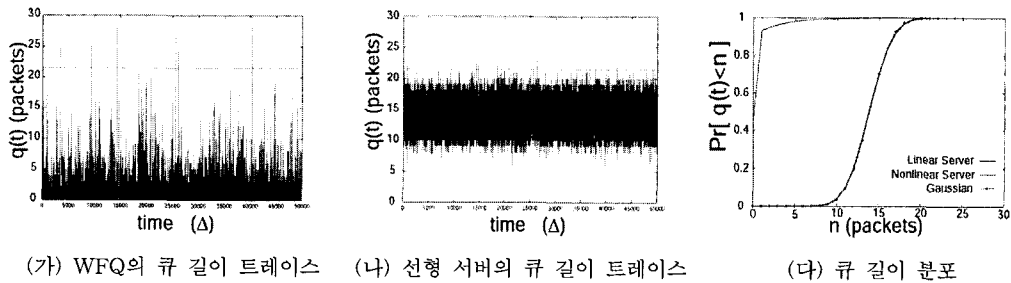


그림 6 선형 서비스 규칙과 WFQ의 비교 (트래픽 A,  $\alpha=0.40$ )

마지막으로 모의실험을 통해 선형 서버와 비선형성을 갖는 서버의 특징에 대해 비교하였다. 동일한 버퍼 크기와 버퍼 넘침 확률을 설정한 후, 가우스 분포를 갖는 트래픽을 입력 트래픽으로 하여 각 서비스 규칙을 적용한 후, 큐의 길이에 대하여 그림 6에 도시하였다. 그림 6의 (가)는 WFQ(Weighted Fair Queueing)를 사용한 큐의 길이를 나타내고, (나)는 선형 서비스 규칙을 사용한 큐의 길이를 나타낸다. 각 그림의 실선은 동일하게 설정한 버퍼 크기를 나타내며, 두 서버 모두 기준 버퍼 넘침 확률은  $10^{-4}$ 이다.

두 서버의 큐 길이 특성을 비교해보면, 선형 서버는 예측한 바와 같이 가우스 분포를 갖는 반면, 비선형 서버는 주어진 대기행렬 시스템의 해를 완전히 구하지 않고는 예측하기 힘든 분포를 가짐을 알 수 있다. 그림 6의 (나)에서 볼 수 있듯이 비선형 서버와는 달리 선형 서버는 큐의 유희기간이 존재하지 않으므로, 출력트래픽 또한 가우스 분포를 갖게 된다. 한편, 주어진 버퍼공간의 이용도 관점에서 선형 서버의 가우스 분포 특성은 버퍼 자원을 더 효과적으로 사용할 수 있게 한다.

## 6. 결론 및 향후계획

본 논문에서는 클래스별 대기행렬 구조를 서비스하기 위한 선형 서비스 규칙을 제안하고, 그 특성을 검증하였다. 선형 서비스 규칙은 큐에서의 서비스량을 입력 트래픽의 선형 함수로 나타내며, 그 결과 큐의 길이, 출력 트래픽의 분포, 지연, 버퍼 넘침 확률 등 여러 QoS 척도들을 입력 트래픽 통계량의 명시적 함수로 나타낼 수 있다. 또한 저대역 필터 계수  $\alpha$ 와 버퍼 크기  $B$ 를 이용하면, 지연 및 버퍼 넘침 확률 등을 원하는 양으로 조절할 수 있다는 장점이 있다. 특별히 입력 트래픽이 가우스 분포를 갖는 경우에는 큐 길이 분포 및 출력 트래픽의 분포가 가우스 분포를 유지하는 특징이 있으며 그 평균과 분산을 입력 트래픽의 평균과 전력 스펙트럼의 명시적 함수로 나타낼 수 있다.

이러한 선형 서비스 규칙은 차등 서비스에 사용될 수 있으며, 각 클래스 큐마다 서로 다른 필터 계수가 사용되므로, 클래스 큐의 특성에 맞는 서비스 제공이 가능하다.

본 논문에서는 단일노드에서의 선형 서비스 규칙에 대한 특성이 연구되었으며, 향후 멀티노드(multi-node) 환경으로 확장된 선형 서비스 규칙의 특성에 대한 연구가 필요하다.

## 참고 문헌

- [1] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Trans. on Networking*, vol. 2, no. 1, Feb. 1994, pp. 1- 15.
- [2] M. Crovella and A. Bestavros, "Self-Similarity in WorldWideWeb Traffic : Evidence and Possible Causes," *IEEE/ACM Trans. on Networking*, vol. 5, no. 6, Dec. 1997, pp. 835-846.
- [3] M. W. Garret and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic," *Proc. ACM SIGCOMM '94*, Sept. 1994, pp. 269-280.
- [4] R. Addie, M. Zukerman and T. Neame, "Broadband Traffic Modelling : Simple Solutions to Hard Problems," *IEEE Communications Magazine*, Vol. 36, No. 8, August 1998, pp88-95.
- [5] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks - the Single Node Case," *IEEE/ACM Trans. on Networking*, vol. 1, no. 3, June 1993, pp. 344-357.
- [6] A. Dmers, S. Keshav and S. Shenker, "Analysis and Simulation of a Fair queueing Algorithm," *Internetworking: Research and Experience*, vol. 1, no. 1, 1990, pp. 3-26.
- [7] L. Zhang, "VirtualClock: A New Traffic Control Algorithm for Packet Switching Networks," *ACM Trans. on Comp. Sys.*, vol. 9, May 1991, pp. 101-124.
- [8] S. Golestani, "A Self-Clocked Fair Queueing Scheme for Broadband Applications," *Proc. IEEE INFOCOM '94*, Apr. 1994, pp. 636-646.
- [9] D. Stiliadis and A. Varma, "Rate-Proportional Servers: A Design Methodology for Fair Queueing Algorithm," *IEEE/ACM Trans. on Networking*, vol. 6, no. 2, April 1998, pp. 164-174.
- [10] M. Shreedhar and G. Varghese, "Efficient Fair Queueing Using Deficit Round Robin," *Proc. ACM SIGCOMM '95*, Sept. 1995, pp. 231-242.
- [11] S. Blake, D. Black, M. Carlson, E. Davies, Z.Wang,W.Weiss, "An Architecture for Differentiated Services," *IETF-RFC2475*, December 1998.
- [12] A. Papoulis, Probability, Random Variables, and Stochastic Processes, 3rd Ed., McGraw-Hill, 1991.
- [13] A. Elwalid, D. Mitra and R. H. Wentworth, "A New Approach for Allocating Buffers and Bandwidth to Heterogeneous, Regulated Traffic in an ATM Node," *IEEE J. Select. Areas Commun.*, vol. 13, no. 6, Aug. 1995, pp. 1115-1127.



- [14] R. Guerin, H. Ahmadi and M. Nagshineh, "Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks," *IEEE J. Select. Areas Commun.*, vol. 9, 1991, pp.968-981.
- [15] M. Hayes, *Statistical Digital Signal Processing and Modeling*, 1st Ed., John Wiley & Sons, Inc., 1996.
- [16] M. Schwartz, *Broadband integrated Networks*, 1st Ed. Prentice Hall, Inc., 1996.
- [17] <http://moat.nlanr.net/Traces/>



정 송

1988년 2월 서울대학교 제어계측공학과 공학사. 1990년 2월 서울대학교 제어계측공학과 공학석사. 1995년 5월 텍사스 주립 대학교 오스틴 분교 공과대학 전기 및 컴퓨터 공학과 공학박사. 1994년 12월 ~ 1996년 8월 Bell Laboratories, Lucent Technologies 연구원. 1996년 12월 ~ 2000년 2월 서강대학교 전자공학과 교수. 2000년 3월 ~ 현재 한국과학기술원 전자전산학과 교수. 관심분야는 초고속 통신망, 고성능 스위치/라우터 시스템, 멀티미디어 네트워크



신 민 수

1998년 2월 한국과학기술원 전기 및 전자공학과(공학사). 2000년 2월 한국과학기술원 전기 및 전자공학과(공학석사). 2000년 3월 ~ 현재 한국과학기술원 전자전산학과 박사과정. 관심분야는 초고속 통신망, 인터넷 QoS



정 현 희

1998년 2월 서강대학교 전자공학과(공학사). 2000년 2월 서강대학교 대학원 전자공학과 졸업(공학석사). 2000년 3월 ~ 현재 삼성전자 네트워크사업부 초고속통신 사업팀