

인터넷에서의 차별화된 서비스품질 제공 방안

김동철¹ · 장희선^{2*}

¹평택대학교 컴퓨터학과 / ²평택대학교 경상학부

Differentiated Quality of Service Model in the Internet

Dong-Chul Kim¹ · Hee-Seon Jang²

¹Department of Computer, Pyeongtaek University

²Division of Business Administration, Pyeongtaek University

The quality of service(QoS) model should be presented with the optimal network design to effectively provide the multimedia data services between users and converged services with mobile or TV in the next-generation Internet. In specific, the method to provide differentiated services for each user is needed in the given Internet node to offer the previously negotiated QoS with the user. In this paper, the performance of the QoS enabling technologies in the differentiated services(DiffServ) network domain is analyzed. The QoS offering model and QoS metrics are presented to analyze the performance of the major scheduling algorithms. Under the real network topology and virtual service scenarios in the university, the NS-2 network simulation based on the discrete-event is performed. The results show that the ratio-based scheduling method is more effective rather than the bandwidth-assignment method.

Keyword: differentiated services, quality of service, internet services

1. 서론

1970년대 중반, 초기 군사와 연구 목적으로 개발되기 시작한 인터넷은 1990년대 이후 상업화의 길로 접어들면서 일반인과 기업들에게도 보편적으로 이용할 수 있게 되었다. 이후 20여년 정도가 지난 지금, 대부분의 국가에서 인터넷이 유선과 이동전화와 더불어 없어서는 안 되는 상호 정보교환의 수단이 되었으며, 특히 최근에는 휴대 인터넷, IPTV, 디지털 인터넷 방송 등, 기존 기술들과의 융합을 통한 새로운 서비스 창출을 통하여 인터넷의 중요성이 다른 각도로 인식되고 있다. 더불어 인터넷의 생활화와 함께 사용자들의 다양한 욕구로 인한 멀티미디어 데이터의 효과적인 정보 교환, 차세대 인터넷 구조 및 서비스, 광대역 통합 네트워크 구축, 융복합 킬러 비즈니스 개발, 무한 경쟁의 통신 시장에서의 생존 등의 문제들의 해결 방안에 대한 논의가 필요한 시점이다(Gozdecki *et al.*(2003), Lee(2009),

Xiao and Ni(1999)). 그리고 사용자들의 다양한 욕구를 해결하기 위한 최우선 해결 과제로서 사용자별 차별화된 서비스품질(QoS : Quality of Service) 제공 기술의 구현과 이를 위한 QoS 제공 모델이 필요하다.

사용자에게 만족할만한 QoS 제공을 위하여 지금까지 주로 Best-Effort 위주의 단일 서비스 모델을 사용하여 왔으나 이는 사용자 증가에 따른 인터넷 트래픽 증가, 다양한 인터넷 어플리케이션 지원 및 다량의 멀티캐스팅 등의 최근 인터넷 사용 환경에서는 바람직하지 못한 QoS 모델로 인식되고 있다(Kim(2003)). 다양한 유형의 멀티미디어 서비스에 대한 QoS를 보장하기 위하여 다양한 QoS 제공 기술들과 함께 여러 QoS 구현 모델들이 제시되고 있다(Braden *et al.*(1994), Cho *et al.*(2002), Crawley(1998), Hardy(2001), Kim(2008)). 특히, 사용자별로 차별화된 QoS 제공을 위한 DiffServ(Differentiated Services) 모델은 기존 모델들과 달리 네트워크 구현과 유연성 측면에서 차세대 인터넷 구

이 논문은 2009학년도 평택대학교 학술연구비의 지원에 의하여 연구되었습니다.

*연락처 : 장희선 교수, 450-701 경기도 평택시 용이동 111 평택대학교 Tel : 031-659-8283, Fax : 031-659-8011, E-mail: hsjang@ptu.ac.kr
투고일(2010년 03월 02일), 심사일(1차 : 2010년 04월 04일), 게재확정일(2010년 04월 06일).

조에서 필수적으로 제공되어야 하는 구현 모델로 인식되고 있다(Blake(1998), Heinanen and Guerin(1999)).

본 논문에서는 DiffServ 네트워크에서 가상의 서비스 시나리오를 가정하여 서비스 등급별 차별화된 QoS 제공을 위한 구현 기술들의 성능을 비교, 분석한다. 이를 위하여 먼저, 지금까지 제시된 QoS 모델들을 상호 비교, 분석하고 사용자간 QoS 측정 기준을 제시한다. 그리고 DiffServ 도메인에서 사용자 종단 간 흐름제어를 위하여 RED(random early discard) 및 TCM(three-color marker) 기술을 분석하고 주요 스케줄링 알고리즘들의 성능을 비교한다. 네트워크 토폴로지를 모 대학 정보 시스템의 구조로 가정하고 강의실과 도서관 그리고 강의실-기숙사 사이의 소스-도착지간 서비스 시나리오를 가정하며, 범용 네트워크 시뮬레이션 도구인 NS-2(Fall and Varadhan(2009), SourceForge(2009))를 이용하여 discrete-event 기반의 시뮬레이션을 수행한다. 서비스 등급별 스케줄링 알고리즘으로서 RR(Round Robin), PRI(Priority), WRR(Weighted RR)의 성능을 비교, 분석하고 차세대 인터넷 구조에서 바람직한 QoS 구현 기술을 제시한다.

2. 서비스 품질

ETSI/ITU(ETSI(1994))에서는 서비스품질(Quality of Service)을 “The collective effect of service performance which determines the degree of satisfaction of an user of the service (서비스 사용자 만족의 정도를 결정하는 서비스 성능의 집합적인 효과)”로 정의하고 있다. 여러 문헌들과 인터넷 서비스 제공과 관련된 표준화 기관별로 서로 다른 각도로 QoS를 정의하고 있으며 이를 <그림 1>과 같이 나타낼 수 있다.

일반적으로 QoS는 서비스와 기술적인 면에서 구분하여 정의할 수 있으며, IETF에서는 주로 기술 측면에서 QoS를 정의한다. 반면, ETSI/ITU에서의 기술적인 측면에서는 네트워크 성능(NP : Network Performance)을 정의하고 서비스 측면에서 QoS를 네 가지 관점(고객 인지, 고객 요구사항, 프로바이더 제공, 프로바이더 구현)에서 정의한다. 이는 각각 Hardy(2001)의 본질적 QoS(Intrinsic, 기술 측면)와 인지적 QoS(Perceived, 서비스 측면)로 구분할 수 있으며, 아직까지 평가와 관련된 QoS(Assessed)에 대해서는 구체적인 개념 정의가 필요하고 이는 주로 서비스 가격, 고객 만족도 및 프로바이더의 정책적인 결정 등과 관련된다.

본 연구에서는 서비스 측면에서의 QoS 개념은 논외로 하고 주로 기술적인 측면에서의 QoS 개념을 사용하고자 한다. 기술적인 측면에 대해 주로 논의하고 있는 IETF RFC 2386(Crawley(1998))에서 QoS는 “A set of service requirements to be met by the network while transporting a flow (어떤 플로우를 전송하면서 네트워크에 의해 만족되어야 하는 서비스 요구사항들의 집합)”으로 이해하고 있으며, 이는 ETSI/ITU의 NP의 개념과 상응된다. 정의에서 사용하고 있는 flow(플로우)란 “사용자와 서버 혹은 사용자와 사용자 사이에 존재하는 어플리케이션의 세션”을 말하며, 예를 들어 FTP를 사용해서 파일을 전송하는 경우, 파일이 전송되는 동안 하나의 세션이 존재하게 되고 이러한 세션을 통해 전달되는 데이터를 트래픽 플로우라고 한다. Intrinsic QoS의 정량적인 값을 표현할 때 QoS의 특성은 여러 가지 파라미터를 사용하여 나타내며, 주요 QoS 파라미터(QoS Parameters 또는 QoS Metrics)를 정리하면 <표 1>과 같다.

위와 같은 서비스 품질을 향상시키기 위해 네트워크에서는 다양한 트래픽 엔지니어링(traffic engineering) 기술을 사용한다.

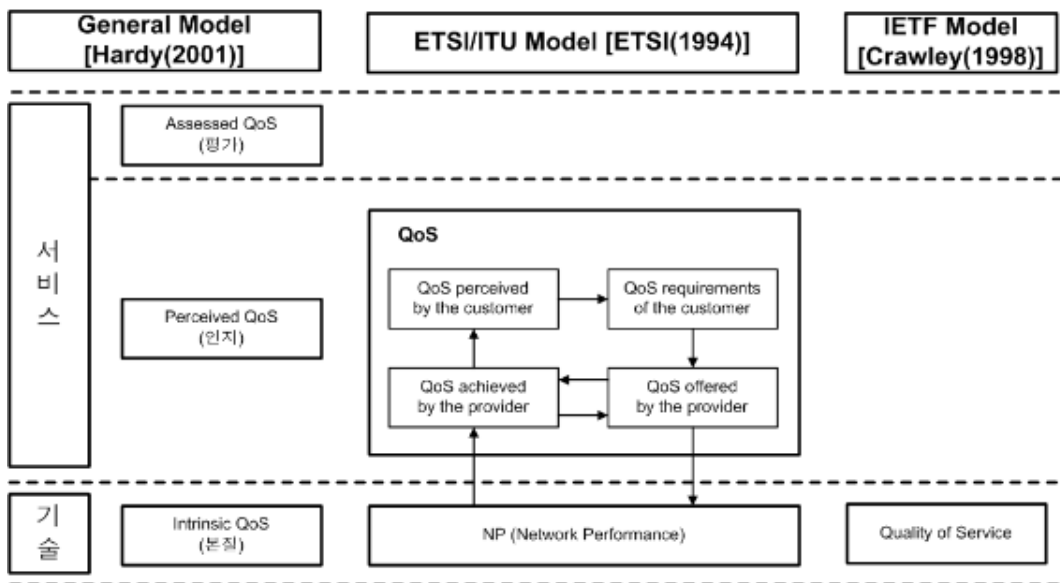


그림 1. 서비스 품질 모델

표 1. QoS 측정 기준

파라미터	주요 내용
대역폭(Bandwidth)	<ul style="list-style-type: none"> 특정 어플리케이션 플로우에 할당된 네트워크 자원의 양 서비스에 대해 가용한 사용자 데이터를 전달하는 속도 혹은 달성될 수 있는 목표 throughput Mbps, Gbps 등으로 측정 Sustained data rate, Peak data rate, min/max data rate 등의 파라미터로 사용
딜레이(Delay/Latency)	<ul style="list-style-type: none"> 네트워크를 통해 전달되면서 패킷이 경험하게 되는 지연 End-to-end 또는 특정 네트워크 구성요소에서 고려 End-to-end 딜레이는 패킷 처리와 관련하여 serialization(패킷 조립/분해), propagation(전달), switching(스위칭/라우팅) 지연의 합으로 구성
지터(Jitter)	<ul style="list-style-type: none"> IP 패킷 전달 딜레이의 변화 송신측에서 패킷 전송시 패킷들 사이의 시간에 대한 수신측에서 수신되는 패킷들 사이의 시간의 왜곡 정도 End-to-end 또는 특정 네트워크 구성 요소에서 고려
패킷 손실(Packet Loss)	<ul style="list-style-type: none"> 패킷 전달 과정에서 패킷 손실의 정도 보내어진 전체 패킷에 대한 전달되지 않은 패킷 수의 비율
기타	<ul style="list-style-type: none"> 패킷의 순서(packet order) 네트워크 가용성 및 신뢰성(network availability/reliability)

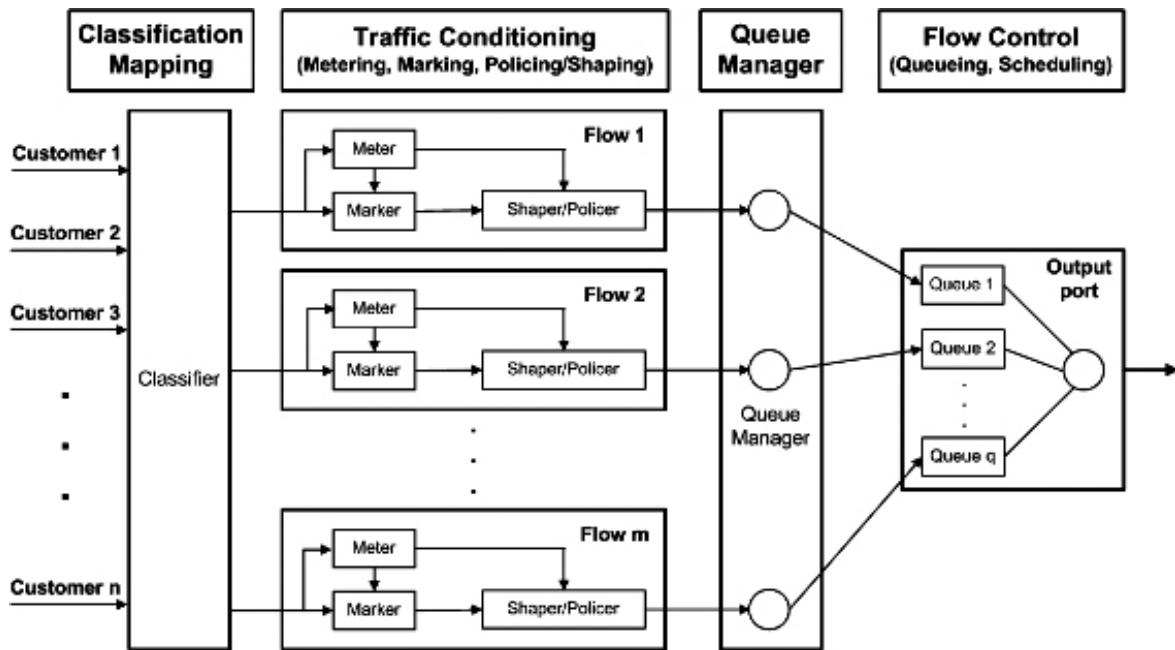


그림 2. 네트워크노드에서의 QoS 구현 기술

네트워크 노드에서 수행되는 대표적인 구현 기술을 <그림 2>와 같이 나타낼 수 있으며, 이를 간략히 설명하면 <표 2>와 같다.

IETF에서는 IP 네트워크에서 보다 효율적인 QoS를 제공하기 위해 여러 가지 표준 프로토콜과 네트워크 구조를 제안하였으며, 대표적으로 IntServ(Braden *et al.* (1994))와 DiffServ(Blake(1998))를 들 수 있다. IntServ에서는 RSVP(Resource Reservation Protocol) 시그널링 프로토콜을 기본적으로 사용하며, GS(Guaranteed Service)와 CL(Controlled Load Service)의 두 가지 서비스 클래스를 제공한다. 여기서 GS는 실시간 어플리케이션에 해당하고, CL

은 사전에 정의된 기준부하 하에서 기존 BE(Best Effort) 트래픽을 처리하는 것과 유사한 서비스 등급을 의미한다.

RSVP의 단점인 확장성(scalability) 문제를 해결하기 위해 제안된 DiffServ 구조에서는 6bits의 DSCP(Differentiated Services Code Point) 값에 의해 패킷을 서로 다른 등급(총 $2^6 = 64$ 개까지 가능)으로 나누어 차등화된 서비스를 제공한다. DiffServ 모델에서는 개별적인 플로우들이 사전에 미리 정의된 서비스들 중 하나로 처리되고 네트워크 상에서 동일한 등급의 플로우들은 같은 방식으로 처리된다. 동일한 등급으로 서비스 되는

표 2. QoS 구현 기술

기술	주요 내용
Classification	- 정의된 기준에 따라 입력 패킷을 여러 등급으로 분류 - Best Effort 등급을 포함하여 일반적으로 4개의 등급 분류 - TOS(type of service) 또는 DSCP 필드를 이용하여 구분
Mapping	- 서로 다른 계층(layer), 프로토콜, 서비스 도메인들 사이에서 QoS 특성을 서로 연결해 주는 기능을 수행 - IP 네트워크에서 주로 Layer 2와 Layer 3사이에서 수행 - 일관된 QoS 특성을 유지시켜 줌
Metering	- 입력되는 트래픽 플로우의 속도를 통해 대역폭을 측정 - 입력 플로우의 버스트(burst) 정도를 측정 - 사전에 정의된 트래픽 프로파일과 비교
Marking	- Metering 결과에 따라 마킹 - 트래픽 프로파일 만족(committed), 일정 범위내 초과(excess), 일정 범위 초과(violated)의 세 가지로 분류 - DiffServ의 경우 sr-TCM(single-rate three-color marker) 또는 tr-TCM(two-rate three-color marker) 이용
Policing/Shaping	- 트래픽 대역폭을 제어 - 초과하는 과도한 트래픽을 제한 - Policing의 경우 버퍼링을 사용하지 않고 초과 트래픽을 폐기하며, shaping의 경우 버퍼링 기능을 사용하여 어느 정도 버스트 트래픽을 수용 - Leaky bucket, Token-bucket 등의 기법을 사용
Queuing	- 한 번에 처리할 수 없는 패킷들을 버퍼에 저장 - FIFO, Priority, Round-Robin, Fair Queueing, Weighted Fair Queueing(WFQ) 등의 기법을 사용 - WFQ의 경우 각 큐에 weight를 할당하여 이 값에 비례하도록 서비스 하며, 보장 대역폭과 연계하여 동작
Scheduling	- 버퍼에 저장된 패킷들을 서비스하는 방식 - 일반적으로 Queuing 기법에 따라 스케줄링 방식이 결정됨
Congestion/Flow Control	- 네트워크 노드들이 계속해서 과부하 상태에 있는 것을 방지 - 입력되는 트래픽을 제어 - 대표적으로 TCP 트래픽을 제어하기 위한 RED(random early discard) 및 WRED(weighted RED) 기법을 이용 - RED 기법에서는 큐에 두 가지 임계값을 설정하여 구간별 서로 다른 패킷 drop 확률을 적용함 - WRED 기법에서는 트래픽 등급에 따라 서로 다른 RED 함수 적용

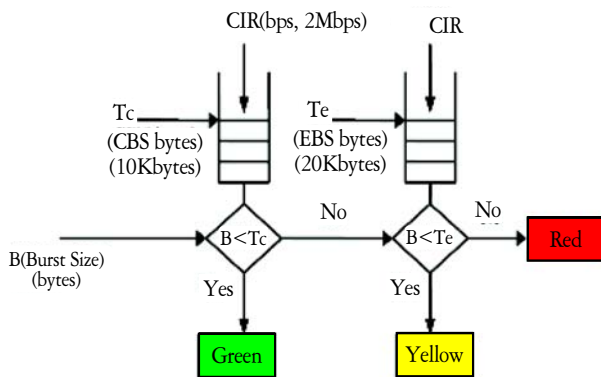


그림 3. srTCM의 동작 원리

패킷(플로우)들은 하나로 통합(aggregate)되며, 동일한 수준의 QoS를 경험한다. 현재 서비스 등급은 EF(Expedited Forwarding, 실시간 트래픽, GS), AF(Assured Forwarding, 기준부하 하에서의 적정대역폭 보장, CL), Default(Best Effort)로 정의되어 사용된다.

<표 2>의 구현 기술 중 DiffServ 서비스를 제공하기 위해 주

로 사용되는 QoS 구현 기술을 요약하면 다음과 같다.

1) **Marking:** RFC 2697 (Heinanen and Guerin(1999))에 기술되어 있는 srTCM(single-rate three-color marker)은 이중토큰 버킷 구조를 가지며, <그림 3>은 srTCM의 동작원리를 보여준다.

입력된 패킷의 크기 B가 토큰 카운터 Tc(CBS : committed bucket size)보다 작으면, Green 패킷으로 마크되고, Tc 보다 크고 Te(EBS: excess bucket size)보다 적은 경우 Yellow 패킷으로, Te보다 큰 경우에는 Red 패킷으로 마크된다. srTCM에서는 사용되는 토큰을 업데이트하기 위해 CIR(committed information rate)만의 single-rate를 사용하며, 반면 trTCM(two-rate three-color marker)에서는 이를 위하여 CIR과 PIR(peak information rate, Te를 업데이트하기 위해 사용)를 사용한다. 일반적으로 srTCM은 버스트 크기가 문제시 되는 플로우에 유용하고, trTCM은 최대 속도가 정확히 요구되는 플로우에 유용한 Marking 기술이다.

2) **Congestion/Flow Control :** 혼잡제어(congestion control)는 네트워크 장비가 처리할 수 있는 능력 이상으로 입력된 트래픽 혹은 특정한 포트에 집중된 트래픽으로 인해 발생하는 혼잡

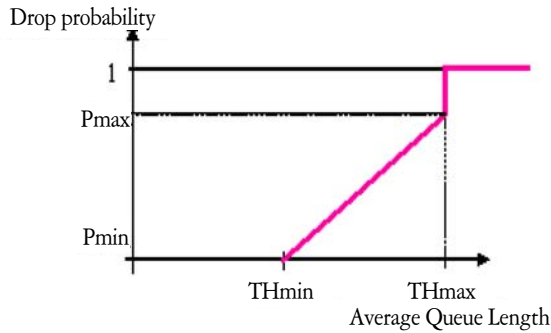


그림 4. RED 큐에서의 drop 확률 값

현상, 즉 장비가 계속해서 과부하 상태에 있는 것을 해결하는 기술을 의미한다. 혼잡제어를 위해 TCP의 동작 특성을 이용하는 방법 중 대표적인 것으로서 RED(random early detection/discard) 기법을 들 수 있다. 기본적인 동작은 큐에 두 개의 임계값(threshold : min, max)을 두고 세 구간에서 서로 다른 drop 확률값을 적용한다. 평균 큐 사이즈(AVG)에 따른 drop 확률 값을 나타내면 <그림 4>와 같다. 평균 큐 사이즈가 THmin 보다 작을 때까지 도착하는 패킷은 drop 되지 않고 큐에서 대기할 수 있으며, AVG=THmax에 도달 하는 경우 최대 drop 확률 Pmax로서 도착하는 패킷이 폐기된다. 이렇게 함으로서 혼잡의 정도가 심해질 수록 보다 많은 패킷을 버리게 되며, 입력되는 트래픽의 양을 줄일 수 있다. THmin과 THmax에 의해 구분되는 각 구간에서 패킷을 버리는 확률은 drop probability function이라는 확률 함수 혹은 확률 테이블로 작성되며, 가능한 큐 사이즈에 대해 drop 확률 값을 정리해 놓는다. 그리고 RED 기법을 사용하고자 할 때는 THmin, THmax, Pmin, Pmax 값을 적절하게 설정해 주어야 한다. 한편, WRED(Weighted RED)는 하나 혹은 여러 개의 서로 다른 클래스 트래픽에 서로 다른 특성을 갖는 RED 함수를 적용함으로써 보다 효과적인 혼잡 제어를 하는 기법을 말하며, 여기서 서로 다른 특성을 갖는 RED 함수란, THmin, THmax, Pmin, Pmax 값이 서로 다른 RED 패킷 drop 확률 함수를 말한다. 동일한 클래스 트래픽에 WRED를 적용하는 경우 같은 클래스에 속한 서로 다른 우선순위의 패킷들에 서로 다른 RED 함수를 적용한다. 예를 들어, 임의의 클래스에 3개의 서로 다른 drop 우선 순위가 존재하는 경우, 각각의 순위에 대해 서로 다른 RED 함수를 적용할 수 있다.

3) Queueing과 Scheduling: Queueing은 어떤 네트워크 장비가 처리할 수 있는 것 이상으로 패킷이 도착하거나 동시에 동일한 목적으로 향하는 패킷들이 존재할 때 발생한다. 즉, 한꺼번에 처리할 수 없는 패킷들을 잠시 동안 buffer에 저장했다가 나중에 서비스하는 것을 큐잉이라 한다. Scheduling은 버퍼에 저장된 패킷들을 서비스하는 방식을 총칭하는 용어이다. 대표적인 큐잉 및 스케줄링 방식으로 FIFO(First-In-First-Out), RR(Round-Robin), PRI(Priority), WRR(Weighted RR) 등이 있다. 주요 스케줄링 기법을 설명하면 다음과 같다.

- RR(Round Robin) : 라운드 로빈 스케줄링 원칙하에, 우선순

위 클래스별로 다시 분류되며, 클래스들간의 엄격한 우선 순위 서비스를 적용하기보다는 라운드 로빈 스케줄러는 클래스들간에 교대로 서비스를 받게 된다. Work-conserving 스케줄링 원칙은 전송하기 위해 대기하고 있는 패킷이 있는 경우에는 링크를 결코 놀리지 않는 것으로, 주어진 클래스에서 패킷을 찾고 없는 경우, 즉시 라운드 로빈 순서상의 다음 클래스를 조사하여 서비스한다.

- PRI(Priority) : Priority 스케줄링에서 출력 링크에 도착하는 패킷은 출력 큐에서 두 개 이상의 우선순위 클래스 중에 하나로 분류된다. 패킷 우선순위 클래스는 패킷 헤더에 전달되는 정보에 의해 결정된다. 각 우선순위 클래스는 전형적으로 자신의 대기 영역(queue)을 가지고 있다. 전송할 패킷을 선택할 때, priority 스케줄링 원칙은 큐가 비어 있지 않고 전송을 대기하는 패킷이 있는 가장 높은 우선 순위 클래스의 큐로부터 하나의 패킷을 전송한다. 그리고 동일한 우선순위 클래스의 패킷들 가운데 선택은 전형적인 FIFO 방식을 적용한다.
- WRR(Weighted Round Robin) : WRR은 상호 연결 요구들에 대해서 라운드 로빈 형태로 데이터를 처리한다. 하지만, 서비스되는 데이터의 양은 무한대로 적은 양 대신 패킷 단위로 처리하게 된다. 만약, 각 연결마다 다른 가중치를 가질 경우, 각 연결에서 서비스되는 패킷의 수는 상대적으로 가중치에 비례하게 된다.

3. 성능분석

3.1 모형

사용자별로 차별화된 서비스 품질을 제공하기 위한 모델의 성능분석을 위하여 <그림 5>와 같은 가상의 네트워크를 가정한다. <그림 5>의 구조는 모대학 내 네트워크에서 강의실, 도서관 그리고 기숙사에서 학생들이 상호 멀티미디어 데이터를 교환하는 환경을 가정하였다. 특히, 본 연구에서는 강의실(이공관)에서 도서관으로 데이터를 송신하고 강의실(인문관)에서 기숙사로 데이터를 보내는 정보 교환 구조를 가정하며, <표 3>과 같은 소스별 트래픽 발생과 파라미터 값을 가정한다.

EF, AF, BE의 세 가지 서비스 등급의 트래픽을 가정하며 EF는 srTCM, AF는 trTCM의 흐름제어를 사용하고 BE 트래픽에 대해서는 흐름제어 정책을 반영하지 않는다(Null). 소스에서의 패킷 발생은 CBR(constant bit rate)을 가정하고 하나의 패킷은 1,000bytes이며, 링크마다 5msec의 지연시간 후에 패킷을 전송한다. 각 소스에서는 DropTail 큐를 가정하여, 각 링크에서의 서비스 처리율(3Mbps) 이상의 트래픽이 발생하는 경우 폐기되며, DiffServ Edge 노드(노드 e0와 e1)에서는 RED 큐를 가정하고 관련 파라미터로서 THmin = 20, THmax = 40, Pmin = 0, Pmax = 0.02를 가정한다.

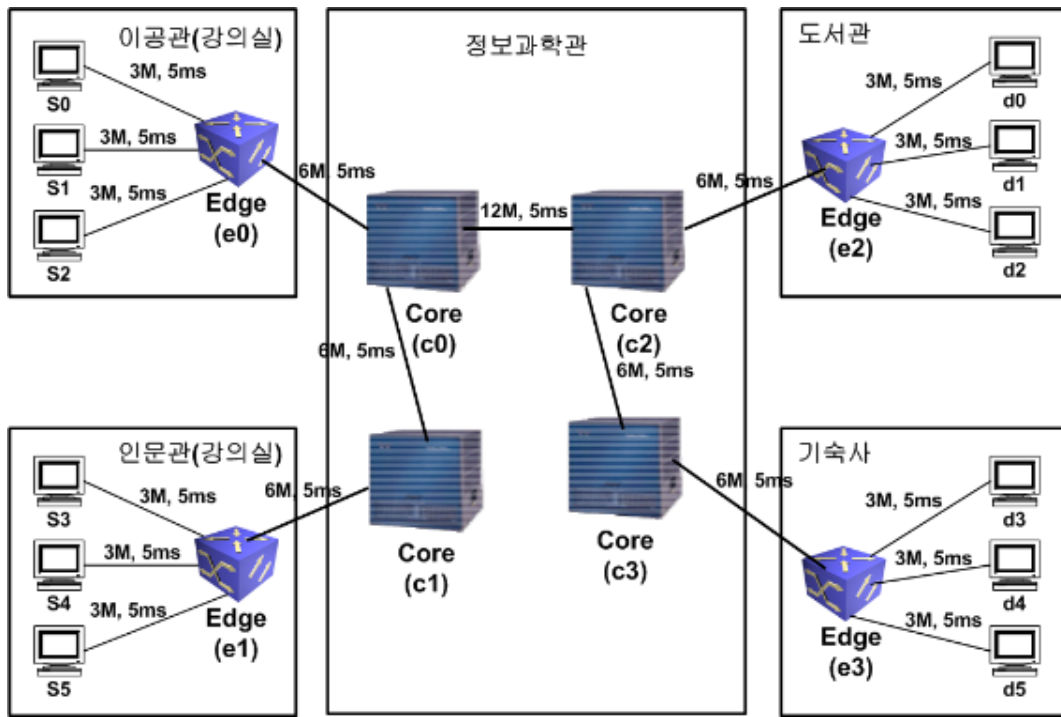


그림 5. 네트워크 구조

표 3. 트래픽 파라미터

서비스 등급(Class)	출발지(Source)	목적지(Destination)	흐름제어 정책(Policy)	파라미터
EF	S0 S3	d0 d3	srTCM	CIR = 1Mbps CBS = 3Kbytes EBS = 6Kbytes
AF	S1 S4	d1 d4	trTCM	CIR = 1Mbps CBS = 3Kbytes PIR = 1Mbps PBS = 6Kbytes
BE	S2 S5	d2 d5	Null	Null

3.2 시뮬레이션 결과

성능분석을 위하여 대표적인 범용 네트워크 시뮬레이터인 NS-2(Network Simulator version 2)를 사용한다. 콜롬비아 대학의 테스트베드를 기반으로 UC Berkeley에서 개발한 NS-2는 TCP, UDP, FTP, HTTP 등과 같은 TCP/IP 프로토콜 패밀리와 라우팅 프로토콜, 멀티캐스팅 프로토콜, 이동통신 망, WLAN, Ad Hoc 네트워크 등 무선 네트워크까지 지원할 수 있는 그 적용 범위가 넓은 시뮬레이터이다(Fall and Varadhan(2009), SourceForge(2009)).

서비스 등급별 성능을 평가하기 위하여 중단간 대역폭, 손실된 패킷의 수, 큐의 크기 및 지연시간을 구하였다. <그림 6>은 각각의 노드별로 단위 시간 동안 발생하는 패킷의 수를 증가시키는 경우 등급별 end-to-end 사이 처리 수율(대역폭, Mbps)을 나타낸다. 1초 동안 발생하는 패킷의 수가 250개 일 때 노드별로 총 2Mbps/sec(= 250 × 1,000 × 8/1,000,000)의 대역을 요구하

게 되며, 이 경우 노드 e0 → c0(또는 노드 e1 → c1)의 대역이 6Mbps로서 모든 등급의 패킷을 처리할 수 있다. 하지만, 단위 시간 동안 발생하는 패킷의 수가 증가할 때(333개 이상), 요구 대역폭은 증가하고 노드 e0 → c0에서 적절한 등급별 처리 알고리즘이 요구된다. 먼저, RR 스케줄링 방법에서는 등급별로 동일한 처리를 하게 되므로 등급에 따른 차이가 없으며 각각의 노드별로 총 2M의 데이터를 처리한다. 사전에 EF에게 2.7M, AF에게 2M의 대역을 할당해주는 PRI 방법에서는 발생하는 패킷의 수가 증가하더라도 EF 등급에게 2.7M, AF에게 2M의 대역을 할당하고 나머지 1.3M(6M-2.7M-2M)는 BE가 사용한다. 반면, EF : AF : BE = 3 : 2 : 1의 비율로서 등급별 서비스 대역을 할당해주는 WRR 스케줄링에서는 각각의 등급별로 최대한 사용할 수 있는 대역폭을 전체 가용 대역폭의 비율로서 사전에 지정해 줄 수 있다. 여기에서는 EF = 3M(= 6M × 3/6), AF = 2M(= 6M × 2/6), BE = 1M(= 6M × 1/6)의 대역을 할당하고 <그림 6>의

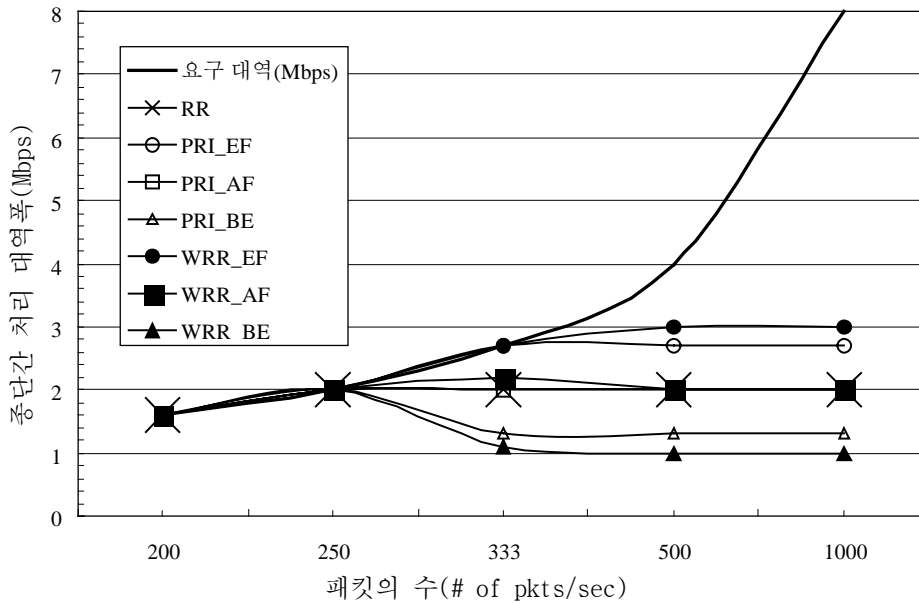


그림 6. 종단 간 처리 대역폭(Mbps)

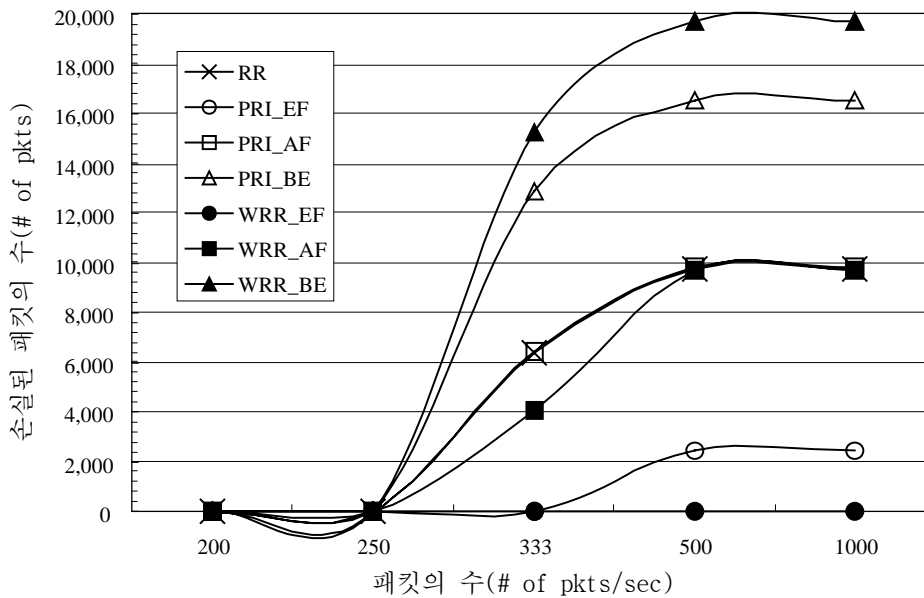


그림 7. 링크 Overflow에 따른 손실된 패킷의 수(# of pkts)

결과처럼 단위시간 동안 발생하는 패킷의 수가 500개 이상일 때 WRR 방법에서는 PRI에서 보다 EF에게 보다 많은 대역을 할당하고, 이로 인해 BE 등급의 대역폭이 다소 감소함을 알 수 있다.

<그림 7>은 노드 e0와 c0 사이의 링크에서 서비스 처리율보다 많은 트래픽 발생으로 인해 dropped된 패킷의 수를 나타낸다. PRI 스케줄링과 비교하여 WRR에서는 EF에게 보다 많은 대역을 할당해 주며 EF 트래픽은 평균 40% 정도 패킷 손실이 감소된다. 그러나 이로 인해 BE 트래픽은 11.3% 정도 패킷 손실이 증가하게 됨을 알 수 있다. <그림 7>에서 패킷 발생율이 500과 1000인 경우 AF 등급에 대한 패킷 손실율이 동일한 이유

는 PRI와 WRR에서 AF 등급에 대한 보장 대역폭이 2M로 동일하기 때문이다.

Edge 노드인 c0의 RED 큐에서의 평균 큐의 크기를 나타내면 <그림 8>과 같다. RR 스케줄링에서는 등급별로 동일한 데이터 처리를 해 줌으로서 큐의 크기가 일정하다. 그러나 PRI 방법에서는 발생하는 패킷의 수가 증가 할수록 WRR에 비해 큐에서 대기하고 있는 패킷의 수가 증가함을 알 수 있다. 이는 PRI 방법에서 EF 등급에게 2.7M만을 할당해 줌으로서 초기 EF 서비스 등급에 해당하는 패킷이 서비스되지 못하고 큐에서 대기하고 있기 때문이며, WRR에서는 EF 등급을 위하여 최대 3M의

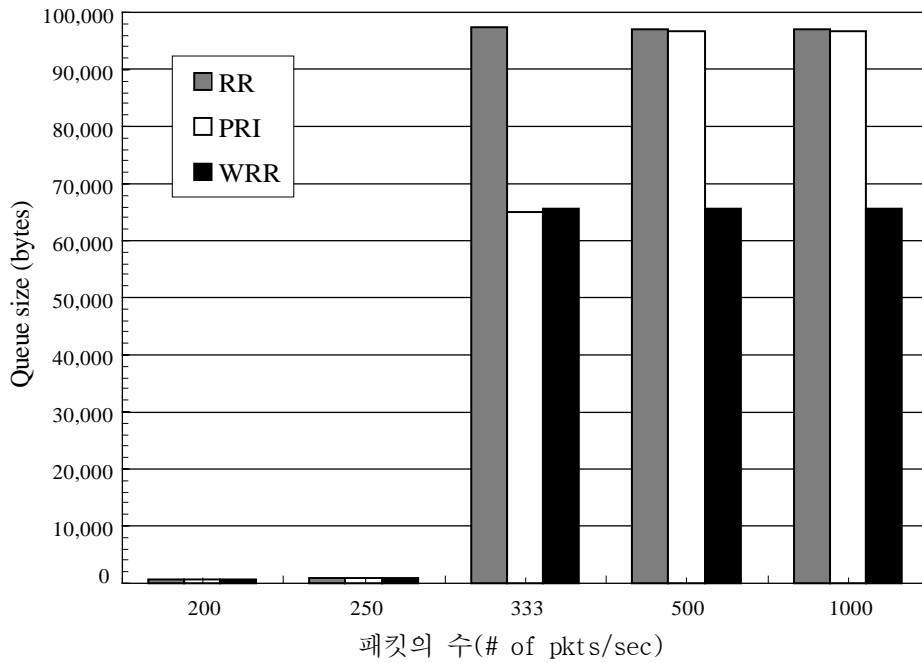


그림 8. 평균 큐의 크기(bytes)

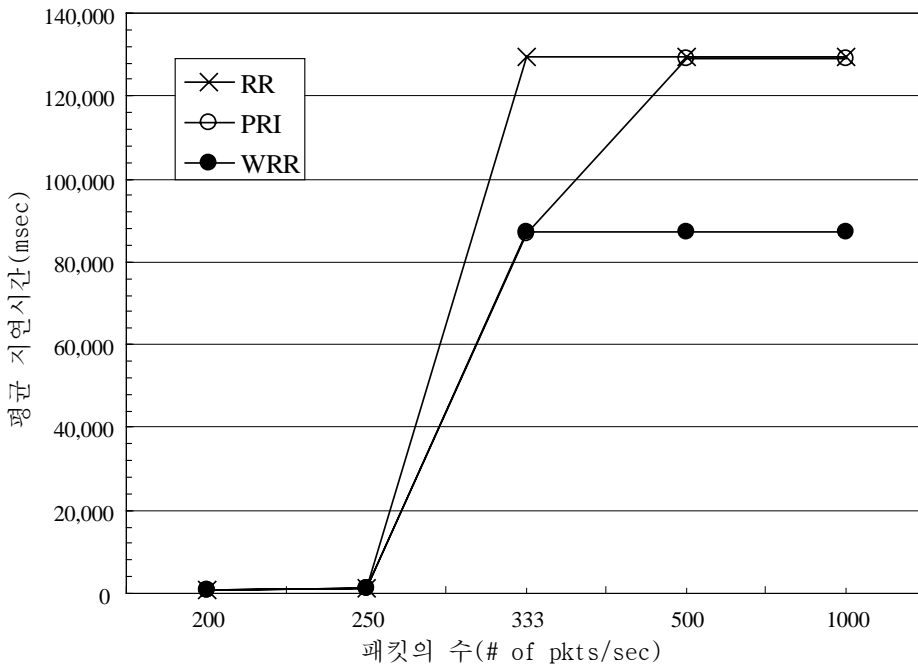


그림 9. 평균 지연시간(msec)

서비스 대역을 할당받게 되어 상대적으로 일정한 큐의 크기를 유지할 수 있다.

이러한 이유로 인해 <그림 9>에서와 같이 WRR에서는 PRI에서 보다 큐에서의 평균 대기시간이 상대적으로 적게 된다. 단위시간 동안 250개 이하의 패킷이 발생하는 경우, 노드당 2M의 트래픽이 발생하며, 이는 e0 → c0의 6M 대역의 용량으로 모든 데이터를 처리할 수 있다. 그러나 6M 이상의 트래픽이 발

생하면, e0 노드는 처리를 하지 못하는 패킷을 RED 큐에 저장하고 이를 사전에 정해진 스케줄링 기법에 따라 패킷을 서비스한다. 모든 등급에게 동일한 우선순위를 주는 RR 방법에서는 모든 패킷이 동일한 지연시간을 갖게 되고, 특정한 서비스 등급에게 우선순위를 주는 PRI와 WRR 기법에서는 도착하는 패킷마다 서로 다른 지연시간을 가진 후 서비스되거나 또는 폐기된다. DiffServ에서 정의하고 있는 PRI 기법에서는 EF와 AF

의 등급에게 우선순위를 줄 수 있는 방법으로서 각각의 등급의 트래픽이 사용할 수 있는 최대 사용 가능한 대역폭을 지정할 수 있다. 이 경우, EF 또는 AF에게 적절한 우선순위를 지정해 주지 못할 경우 <그림 8>과 <그림 9>에서와 같이 큐에서 대기하고 있는 패킷의 수가 증가하고 이에 따른 지연시간이 길어진다. 반면, WRR에서는 전체 서비스 대역폭에 대한 서비스 등급별 가용 대역의 비율을 유연하게 지정할 수 있으므로 PRI에서 발생할 수 있는 문제점을 해결할 수 있다. 아울러 라우터 설계시 특정 서비스에 대한 서비스 이용 대역을 지정하는 방법보다는 전체 가용 대역의 비율로서 정의하는 것이 트래픽 엔지니어링 기법에서 보다 효율적임을 알 수 있다.

4. 결론

1990년대 중반부터 일기 시작한 인터넷 붐은 일반 사용자에서 기업에 이르기까지 사회 전 분야에 주요한 구조 변화를 일으키고 있다. 이러한 구조 변화의 중심에는 크게 차세대 인터넷 구조, 인터넷 서비스, 휴대 인터넷/IPTV 등 인터넷 융합 기술, 멀티미디어 데이터 처리 등의 핵심 이슈와 함께 사용자들에게 만족할만한 서비스를 제공해 주기 위한 서비스품질(QoS: Quality of Service)에 대한 논의가 이루어지고 있다. 특히, 미래 융복합 서비스를 제공하기 위해서는 인터넷 인프라가 필수적으로 요구되고 사용자 간 멀티미디어 데이터의 처리를 위해 서비스 등급별 QoS 제공 기술이 반드시 필요하다. 지금까지 다양한 QoS 제공 기술들이 제시되고 있으며, 본 연구에서는 DiffServ(Differentiated Services)에서 제시된 서비스 등급별 차별화된 서비스 품질 제공 방안을 분석하였다. 이를 위하여 먼저, 기존 QoS 모델들을 비교, 분석하고 QoS 측정 기준을 제시하며, DiffServ에서 사용되고 있는 주요 QoS 구현 기술을 정리하였다. 주어진 네트워크에서 시뮬레이션 분석을 위하여 서비스 등급을 크게 EF(Expedited Forwarding), AF(Assured Forwarding), BE(Best Effort)의 세 가지 등급으로 나누고 종단 간(end-to-end) 흐름제어를 위하여 RED(Random Early Discard)와 TCM(Three-Color Marker) 기법을 사용하였다. 그리고 차별화된 서비스 처리를 위한 스케줄링 기법으로 RR(Round Robin), PRI(Priority), WRR(Weighted RR)의 성능을 비교, 분석하였다. 범용 네트워크 시뮬레이션 도구인 NS-2를 이용하며, 모 대학의 가상 네트워크 구조를 가정하여 소스-목적지 노드 사이 단위 시간 동안의 발생 패킷의 수를 증가시키면서 서비스 등급별로 차별화된 QoS의 성능을 비교하였다. 모든 등급에 대해 동일한 순위를 지원하는 RR에서는 각 노드별로 동일한 서비스 대역폭을 보장해 줄 수 있으나 특정 서비스에 대해 차별화된 서비스 처리(대역폭과 지연시간 관점에서 QoS 보장)가 불가능하다. 이를 해결하기 위해 PRI 방법에서는 사전에 EF와 AF에게 특정 범위의 대역을 보장해 주며, WRR에서는 전체 가용 서비스 대역의 일정 비율만큼을 할당한다. 시뮬레이션 분석 결과, PRI에 비해 WRR 기법에서 보다 유연하게 가용 대역

의 양을 지정해 줄 수 있으며, 이로 인해 RED 큐에서 대기하고 있는 패킷의 수가 감소하고 결국, 패킷 당 평균 지연시간이 감소함을 알 수 있다. 이외에도 라우터 설계시 PRI를 적용하고자 할 경우 각각의 서비스 등급에 대하여 특정 대역폭을 사전에 지정해 주어야 하는 번거로움이 있으나 WRR에서는 전체 가용 대역의 비율만을 지정함으로써 특정 서비스 등급에 대한 차별화된 QoS 제공이 가능함을 알 수 있다. 물론, 인터넷 서비스를 제공하는 인프라에서 소스-목적지 노드 사이 정확한 트래픽 정보가 있는 상황에서는 PRI 방법이 보다 효율적인 성능을 보일 수도 있으나, 인터넷 소스 트래픽의 버스트한 성질과 여러 유형의 멀티미디어 서비스를 처리해야 하는 미래 인터넷의 구조에서는 정확한 트래픽 예측이 불가능하다. 따라서 주어진 인터넷 구조에서 사용자간 멀티미디어 서비스를 제공하기 위해서는 전체 가용 서비스 처리 능력의 일정 비율로서 각각의 서비스 등급별 차별화된 서비스 품질을 제공하는 것이 보다 효과적임을 알 수 있다. 향후 실제적인 소스-목적지 노드간 인터넷 트래픽 상황에서 보다 현실적인 성능분석이 필요하며, WRR 기법 적용시 소스 트래픽의 다이나믹한 성질을 반영한 효과적인 알고리즘 개발이 이루어져야 하고, 아울러 소스 트래픽의 특징별로 개별적인 흐름제어와 스케줄링 기법의 개발이 필요하다.

참고문헌

Blake, S. (1998), An Architecture for Differentiated Services, IETF RFC 2476.
 Braden, R., Clark, D., and Shenker, S. (1994), Integrated Services in the Internet Architecture: an Overview, IETF RFC 1633.
 Cho, K. S., Jang, H. S., Lim, S. K., and Kim, Y. S. (2002), Quality of Service and Network Performance for the IMT-2000 Services, *IE Interfaces*, 15(3), 256-262.
 Crawley, E. (1998), A Framework for QoS-based Routing in the Internet, IETF RFC2386.
 ETSI (1994), Network Aspects(NA) : General Aspects of Quality of Service(QoS) and Network Performance(NP), Technical Report, ETR003.
 Fall, K. and Varadhan, K. (2009), The ns Manual(The VINT Project), <http://www.isi.edu/nsnam/ns/ns-documentation>.
 Gozdecki J., Jajszczyk, A., Stankiewicz, R. (2003), Quality of Service Terminology in IP Networks, *IEEE Communication Magazine*, 153-159.
 Hardy, W. C. (2001), QoS Measurement and Evaluation of Telecommunications Quality of Service, Wiley.
 Heinanen, J. and Guerin, R. (1999), A Single Rate Three Color Marker, IETF RFC 2697.
 Kim, H. K. (2003), Understanding of QoS Technologies, Network Manias White Paper, Netmanias-wp-sub-104, <http://www.netmanias.com>.
 Kim, T. J. (2008), Performance Evaluation of the RSVP-capable Router using Latency-Optimized Fair Queueing Scheduler, *Journal of Korea Multimedia Society*, 11(11), 1536-1546.
 Lee, Y. R. (2009), The Case Analysis of QoS Evaluation for Foreign Major Countries, KISDI Report, 21(14), 25-43.
 SourceForge (2009), The Network Simulator-ns-2, <http://sourceforge.net/projects/nsnam> (or <http://www.isi.edu/nsnam/ns/index.html>).
 Xiao, X. and Ni, L. M. (1999), Internet QoS : A Big Picture, *IEEE Network*, 13(2), 8-18.

**김 동 철**

University of Utah 학사

Oregon State University 석사

Brigham Young University 박사

현재 평택대 컴퓨터학과 교수

관심분야 : 컴퓨터네트워크, QoS 관리, 분산
시스템, 네트워크성능분석**장 희 선**

울산대학교 산업공학과

KAIST 산업공학과 석사

KAIST 산업공학과 박사

한국전자통신연구원 선임연구원

현재 평택대학교 경상학부 교수

관심분야 : 트래픽 엔지니어링