論文2003-40TC-12-15

NGN 서비스의 호 처리 차별화 방안 및 성능분석

(Differentiated message handling and performance evaluation for the NGN call control services)

鄭 文 照 *, 黃 燦 植 **

(Mun Jo Jung and Chan Sik Hwang)

요 약

본 논문에서는 NGN에서 차세대 멀티미디어 통신 서비스를 제공함에 있어서 다양한 유형의 서비스에 대한 연결제어를 담당하는 소프트스위치에 대하여 연결특성이 다른 두 가지 이상의 응용 서비스를 가정한 경우 연결형태별 호 처리를 위한 제어 메시지의 요구조건을 만족하는 서버를 설계하기 위한 호 처리 차별화 방안에 대한 몇 가지 대안들을 제안한다. 그리고 각 대안 별로 호 처리 모델의 성능분석을 통하여 제안된 호 처리 차별화 방안의 유효성을 입증하고자 한다.

Abstract

In this paper we propose service schemes for the control message of voice and data connections served by a Softswitch in NGN (next generation networks). After that we propose a method of evaluating the performance of a Softswitch that provides a limited delay to voice connections. Via numerical experiments, we verify the implication of the proposition in the design of a Softswitch, which simultaneously incorporates voice and data services in the NGN framework.

Keywords : NGN, Softswitch, Call Processing, Performance evaluation

I. 서 론

최근에 들어서 통신사업자는 새로운 수입원의 고갈, 기존 통신장비의 노후화 및 새로운 네트워크 기술의 발달 덕분에 네트워크 운용의 효율화를 기하기 위하여 NGN(Next Generation Networks)의 구축에 박차를 가하고 있다. NGN에서는 IP 망을 전달망으로 구성하여 기존 PSTN 및 데이터 서비스를 수용하며, 호/연결/세션제어를 수행하기 위하여 물리적 또는 논리적으로 분리된 별도의 제어망을 구축하고 활용할 것으로 예상된다. 그 중에서 제어망은 트렁크게이트웨이, 액세스게이트웨이, 레지덴셜게이트웨이 등과 연동하며, PSTN No. 7 시그널링 연동을 위하여 시그널링게이트웨이와도 연동한다. 또한 MEGACO, SIP 등 프로토콜을 이용하여 기존 PSTN 전화기와 새로운 형태의 IP 단말기를 수용하며 SIP 서버 등과 연동한다. 특히, IP 망에서 신규 멀티미디어 서비스를 수용하기 위하여 응용 서버 및 미디어 서버와 연동한다. 또한 과금, 망 운용 관리 등은 사업자 망 운용 방식에 따라 다양한 형태로 구축되며, 여러종류의 back-end 시스템들과 연동이 예상된다[1-3].

IP 기반의 차세대 개방형 망에서 다양한 멀티미디어 서비스를 제어하기 위한 NGN 핵심 망 요소로서 소프트

* 正會員, KT 技術硏究所

(KT Technology Laboratory)

** 正會員, 慶北大學校 電氣電子工學部

(Department of Electronic Engineering Kyungpook National University)

接受日字:2003年 10月10日, 수정완료일:2003年 12月17日

스위치(Softswitch, 이하 SSW로 부름)가 있다. 소프트스위치는 PSTN 망의 Class4/Class5 서비스와 IP 단말을 이용한 음성 및 데이터 서비스를 수용하여 VoIP 서비스의 제공뿐만 아니라 응용 서버 및 미디어 서버 등과의 연동하여 다양한 종류의 멀티미디어 서비스를 제공하는 것을 목표로 하고 있다. 이와 같은 추세에 부응하여 현재 전기통신망사업자들은 차세대 네트워크 서비스로서 음성 및 데이터 복합 서비스를 제공하기 위하여 소프트스위치의 도입을 계획하고 있고 장비 제조업체들도 소프트스위치의 개발에 박차를 가하고 있다.

그러나 소프트스위치의 도입을 앞두고 선행되어야 할 작업의 하나인 성능에 대한 정량적인 평가는 별로 이루어져 있지 않다. 따라서 이 문제를 해결하기 위하여 단일 클래스 서비스로서 음성 서비스를 제공하는 경우에 대하여 음성 서비스의 호 설정 지연을 보장하기 위하여 호 제어 메시지의 도착률에 대한 소프트스위치가 필요로 하는 서버의 용량을 설계하였다[4]. 그러나 이 경우에도 멀티미디어 서비스 환경이 아닌 단일 음성 서비스 환경이었으므로 NGN이 추구하는 실질적인 서비스 환경에 대한 성능분석을 제대로 할 수 없다는 한계가 존재하였다.

본 연구에서는 이전 연구의 한계를 보완하여 보다 현실적인 환경을 반영하기 위하여 NGN의 대표적 서비스인 음성 서비스와 인스턴트 메시징 서비스(Instant Messaging Service, 이는 데이터계 서비스로 분류됨)를 동시에 제공하는 서비스 환경을 가정한다. 그리고 성능평가모델로서 이전에 제안한 소프트스위치가 제공하는 서비스의 클래스를 음성 서비스에 대한 단일 호 제어 모델에서 음성과 데이터의 두 가지 클래스로 확장하고자 한다. 제어 메시지의 서비스 방법으로는 이전의 방식과 새로운 방식으로서 음성이 데이터보다 우선적으로 서비스를 받도록 하여 지연에 민감한 음성 서비스가 요구하는 호 설정 지연목표치를 보장하도록 하기 위한 소프트스위치의 서버용량을 정량적으로 설계하는 방안에 대하여 몇 가지 대안을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 제Ⅱ장에서는 NGN에서 음성 서비스 및 데이터 서비스의 제어를 위하여 필요한 소프트스위치의 호 처리 기능을 분석하고, 제Ⅲ장에서는 소프트스위치의 호 처리 차별화 모델을 수립하고 수립한 모델의 성능분석 방법을 제안한다. 제Ⅳ장에서는 수치실험을 통하여 제안된 방법의 물리적 의미와 유효성을 기술한다. 마지막으로 제Ⅴ장에서는 본 연구의 의미를 정리하고 향후 연구방향에 대하여 간략히 기술한다.

Ⅱ. 소프트스위치 호 처리 기능 분석

소프트스위치는 기본 호 처리 기능, 장치제어 기능, 응용 서버 인터페이스 기능, 주소번역 기능, 루팅 기능, 연결제어 기능, 사용자/서비스 프로필 관리 기능, 자원관리 기능, 운용관리 기능, 과금 기능 등을 수행한다[1-3]. 이 중에서 소프트스위치의 기본적이고도 가장 중요한 기능은 단대단 통신연결을 위한 제어를 수행하는 호 처리 기능이다. 이는 호의 연결, 진행, 감시 및 해제기능을 수행하는 것으로 정의한다. <그림 1>은 NGN에서 음성 서비스 및 IM 서비스를 위한 연결 상황을 개략적으로 나타낸 것이다.

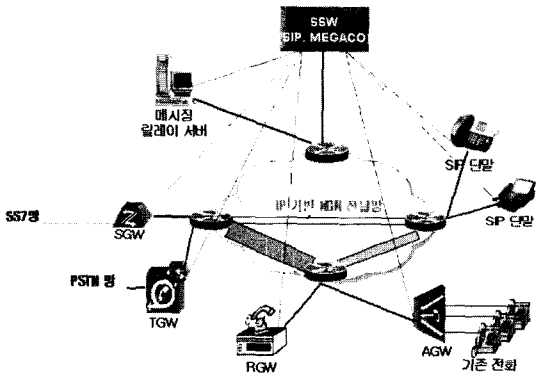

그림 1. NGN에서 음성 및 IM 서비스 구성 예
Fig. 1. An example of voice and IM service in NGN.

SIP를 이용한 음성호의 제어 메시지에 대한 메시지 전달 순서 및 그에 대한 지연요소를 분석하고 M/M/1 대기행렬모델을 이용하여 단일 소프트스위치에서의 평균지연을 분석하는 연구가 진행되었다[4]. 그러나 이 경우에 제안된 모델은 음성 서비스만을 제공하는 경우에 한정되었다. 따라서 여기서는 데이터 서비스를 추가한 경우로 확장하여 모델링하고, 음성/데이터 복합 서비스를 제공하는 경우의 성능평가를 위하여 실제 제어 메시지가 흐르는 과정을 살펴보고 지연요소에 대한 모델링을 하고자 한다.

소프트스위치의 성능을 평가하기 위하여 메시지의 송수신단간의 접속구조를 <그림 2>와 같이 나타낸다. <그림 2>에서 송신자는 TE1(Terminal equipment 1)

이고 수신자는 TE2로 가정하고, 메시지의 흐름을 송신측에서 수신측으로 향하는 메시지(Source-to-Destination Message, 이하 S2D메시지라 함)는 실선으로 나타내고, 수신측에서 송신측으로의 응답 메시지(Destination-to-Source Message, 이하 D2S 메시지라 함) 는 점선으로 나타내었다. <그림 2>의 메시지 흐름과 비교하면 INVITE, ACK, RE-INVITE 및 BYE 메시지는 S2D 메시지이고, 180 Ringing 및 200 OK 메시지는 D2S 메시지에 해당한다.

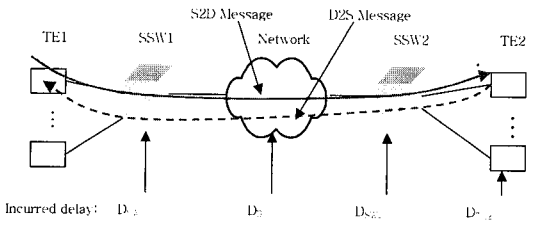

그림 2. 단대단 지연 성능평가 모델
Fig. 2. Performance analysis model for end to end delay.

<그림 2>에서 메시지가 TE1에서 출발하여 TE2를 거쳐 다시 TE1까지 돌아올 때까지의 왕복지연시간이 중요한 성능변수가 되는데, 이를 모델링하기 위하여 각 지점에서의 지연을 아래와 같이 정의한다. 송신측 TE1에서 출발한 패킷의 지연 발생요소는 소프트스위치1(SSW1), IP 네트워크, SSW2 및 TE2로 정의하고, 각 요소에서의 지연은 송수신방향에 대해서 균일하게 분포한다고 가정을 한다.

메시지가 SSW1, IP 네트워크, SSW2 및 TE2에서 겪는 지연을 각각 D_{SW1}, D_{IP}, D_{SW2}, D_{TE2}라고 가정을 한다. 이 때 D_{SW1}과 D_{SW2}는 소프트스위치에서 메시지가 겪는 지연이다. 이 지연치는 소프트스위치의 처리용량과 유입되는 패킷의 양에 직접적으로 의존하는 랜덤변수이며 본 연구에서 가장 초점이 맞추어져 있는 부분이다. 한편, D_{IP}는 양단의 소프트스위치간의 메시지가 IP 네트워크를 통과하는데 걸리는 시간을 나타낸다. 마지막으로 TE2에서 겪는 지연 D_{TE2} 는 수신단말에서 걸리는 지연으로서 단말의 메시지 처리성능에 의해서 결정되는 값이다.

위에서 한 메시지의 단방향 단대단 지연 D_{e2e}는 S2D 메시지가 D_{SW1}, D_{IP}, D_{SW2}, D_{TE2}의 합과 같다. 그러나 대부분의 제어 메시지는 [송신측 단말기->수신측 단말기->송신측 단말기]에 이르는 왕복지연시간(round trip time, RTT)을 측정해야 의미가 있기 때문에 다른 성능 파라미터로서 양방향 지연, D_{RTT}을 정의할 필요가 있다. D_{RTT}의 값은 위의 가정에 따라 D_{e2e}의 2배로 근사화 할 수 있다.

백본 네트워크에서의 지연 D_{IP}는 네트워크의 상황에 따라서 바뀌는 랜덤변수이나, 본 연구의 목적이 IP 백본의 전달 지연을 평가하는 것이 아니라 소프트스위치의 호 처리 지연성능이다. 따라서 백본 네트워크의 특성상 충분한 용량의 대역폭이 설계되어 있다고 가정을 하고 상수로 취급하기로 한다. 실제로 NGN에서 VoIP 서비스에 대한 데이터 및 호 설정을 위한 신호 데이터의 서비스 품질 클래스는 DiffServ의 EF(expedited forwarding) 클래스에 해당하고 EF 클래스의 패킷은 각 노드에서 다른 패킷의 과다유무에 영향을 받지 않고 최우선적으로 처리되기 때문에 거의 실시간으로 전송되도록 네트워크 자원을 송신 데이터률의 최대속도에 해당하는 대역폭을 할당하고 있다[5]. 그러나 복수 개의 패킷이 중첩되어 입력되는 노드에서 버퍼 내 대기 지연은 불가피한 상황이다. 따라서 호 설정 지연목표치는 ITU-T 및 제 외국 서비스 사업자가 권장하는 지연목표치를 근거로 상수값을 사용한다. 결국 D_{RTT}의 성능은 소프트스위치인 SSW1과 SSW2에서 일어나는 지연에 가장 크게 좌우된다.

메시지의 지연 목표치가 주어지면 상수항에 해당하는 네트워크 지연인 D_{IP}와 수신측 단말 지연인 D_{TE2}를 제외한 지연목표치를 구하고 그 값을 <그림 3>에서 나타낸 D_{SW1}과 D_{SW2}로 분산시킬 수 있음을 알 수 있다. 해석의 편의상 D_{SW1}=D_{SW2}라고 가정을 한다. SSW1과 SSW2는 IP 네트워크를 사이에 두고 TE1과 TE2간의 호 제어 메시지를 처리한다. 그리고 각 소프트스위치가 동시에 충분히 큰 규모의 복수의 연결을 독립적으로 처리하는 서버이기 때문에 메시지 시퀀스의 관계는 의미가 없게 되며, SSW1과 SSW2는 서로 독립된 서버라고 가정을 할 수가 있다. 따라서 TE1에서 생성된 메시지가 소프트스위치와 IP 네트워크를 통하여 TE2에 전달된 후 다시 자신에게 돌아오는데 걸리는 지연은 <그림 3>과 같은 등가지연 모델로 나타낼 수가 있다. <그림 3>에서 D_{SW1}, D_{SW2}, D_{IP}, D_{TE2}은 각 각 SSW1, SSW2, IP 네트워크 및 TE2에서의 지연을 나타낸다.

<그림 3>에서 D_{IP}와 D_{TE2}는 상수의 값을 취한다고

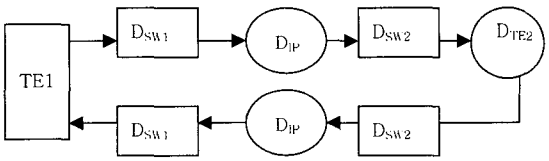

그림 3. 등가지연 모델
Fig. 3. Equivalent delay model.

앞에서 기술하였으므로 지연의 해석은 두 개의 소프트스위치에서 일어나는 지연인 D_{SW1} 및 D_{SW2}가 변수로 작용한다. 한편, 독립된 서버가 Tandem으로 연결된 네트워크의 성능에 대한 해석법은 대기행렬이론에서 Open Jackson Network의 개념으로 잘 정의되어 있으며, 결과를 요약하면 M/M/1 대기행렬이 K개 직렬 연결된 경우의 D_{e2e}은 단일 노드 지연의 K배이다[6]. 이 사실을 이용하면 결국 단대단 지연성능의 해석은 SSW1 혹은 SSW2에 대한 단일 노드해석으로 귀착된다. 송수신단간의 지연성능을 모델링함에 있어서 소스측 소프트스위치인 SSW1에 대해서 논의를 전개하기로 한다.

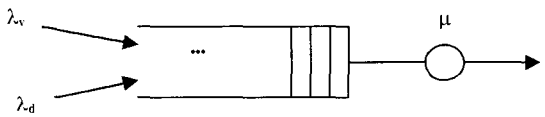

그림 4. 소프트스위치의 추상화 모델
Fig. 4. Simplified Softswitch model.

<그림 4>에서 각 메시지의 서비스 시간은 길이가 크게 차이는 나지 않으나 같지는 않다 (<표 2 ,3>참조). 따라서 이 메시지들의 길이는 평균이 1/μ인 지수분포를 따른다고 가정을 한다.

한편, 소프트스위치는 음성 서비스의 제공뿐만 아니라 응용 서버 및 미디어 서버 등과 연동을 통하여 데이터 및 비디오 통신 등 다양한 종류의 멀티미디어 서비스를 제공하여야 하므로 본 연구에서는 단일 음성 서비스 제공시의 모델링[4]을 확장한 모델로서 음성 서비스 이외에도 데이터 서비스까지도 제공하는 소프트스위치 모델을 제안하고자 한다. 소프트스위치에서 음성 서비스 이외에 데이터 서비스까지 제공하는 경우에 음성 서비스와 데이터 서비스에 대한 호 처리 특징과 소프트스위치에 대한 요구사항을 분석할 필요가 있다. 음성 서비스를 위한 제어 메시지를 처리하는 경우에는 기존의 PSTN에서와 마찬가지로 실시간 통화를 위하여 송수신단간에 일정한 시간이내에 연결설정이 완전히 이루어지도록 할 필요가 있다. 따라서 음성호의 연결에 필요한 메시지들의 지연이 정해진 지연목표치를 만족하도록 진행되어야 한다. 음성 서비스를 위한 제어 메시지의 절차도 및 메시지의 종류는 참고문헌[4]를 참고하기로 한다.

한편, 데이터 서비스를 위한 제어 메시지의 경우에는 <그림 5>의 첫 번째 블록에서 나타낸 바와 같이 서비스의 특성상 송신자가 세션설정이 이루어질 때까지 기다리는 최대시간(EXP=600)을 지정할 수 있다.

따라서 송수신단간의 연결설정이 임의의 유예시간을 가질 수 있고, 또 제어 메시지의 교환이 [가입자↔릴레이서버↔소프트스위치↔가입자]의 연결 관계상에서 직접 혹은 간접적으로 일어난다. 즉, 어떤 메시지는 [가입자↔릴레이서버] 간에 일어나고 어떤 메시지는 [가입자

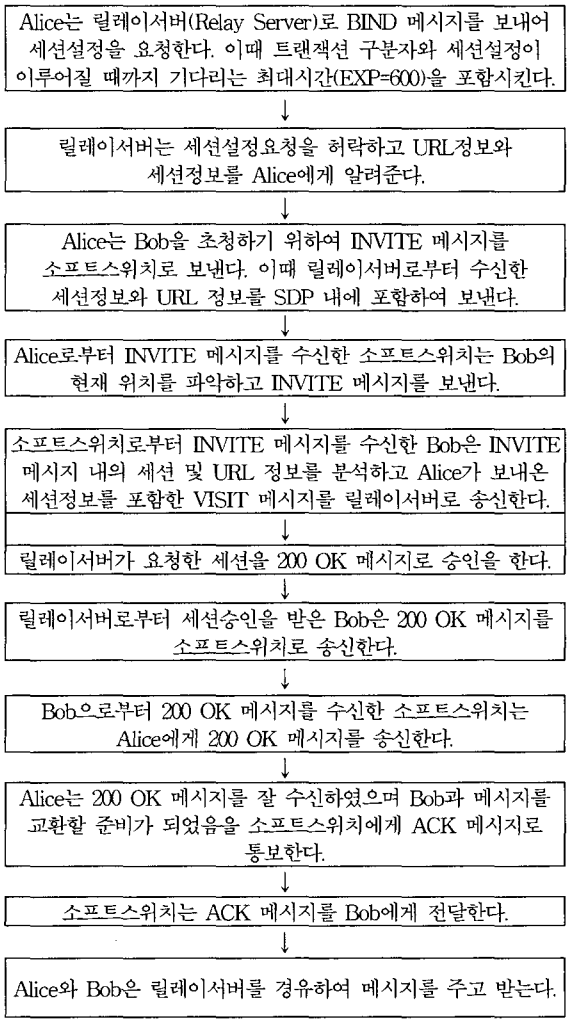

그림 5. IM 서비스 세션 설정 절차
Fig. 5. Session setup procedure for IM service.

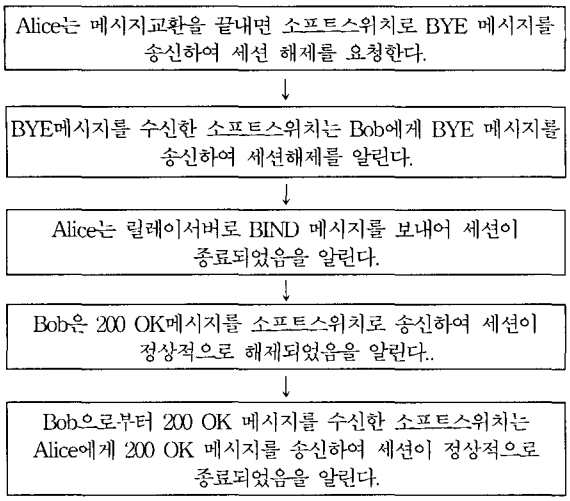

그림 6. IM 서비스 세션 해제 절차
Fig. 6. Session release procedure for IM service.

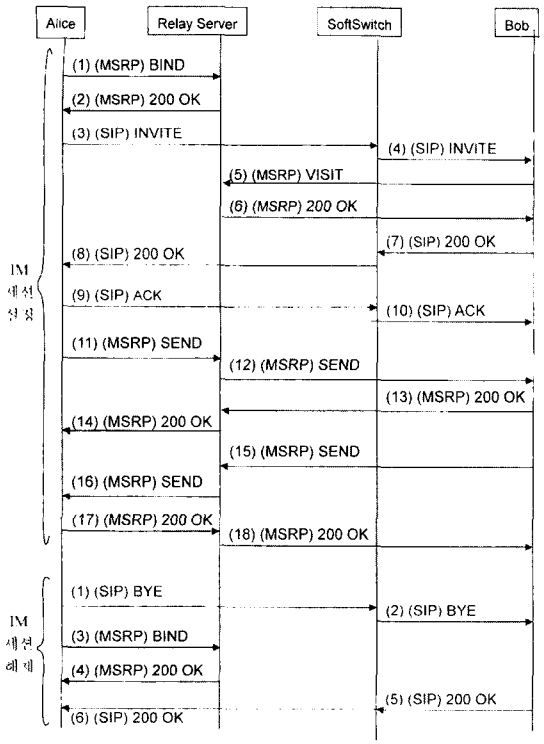

그림 7. IM 서비스 제어 메시지의 처리흐름도
Fig. 7. Control message flow for IM service.

↔소프트스위치] 간에 일어난다[7,8]. 따라서 음성 서비스를 위한 제어 메시지보다 연결설정에 필요한 메시지의 개수가 작다. <그림 4>에서 볼 수 있는 바와 같이 IM 서비스를 위한 세션설정을 위하여 세 개의 메시지(INVITE, 200 OK, ACK)가 소프트스위치를 경유하며 세션의 해제를 위해서 두 개의 메시지(BYE, 200 OK)가 소프트스위치를 경유한다. 결국 IM 서비스를 위한 세션의 설정 및 해제를 위해서 소프트스위치가 처리해야 할 메시지의 개수는 총 5개가 된다. 따라서 하나의 IM 서비스를 위한 세션을 설정하고 해제하는데 소요되는 부하는 이들 5개의 메시지(INVITE, 200 OK 2개, ACK, BYE)만 고려하면 됨을 알 수 있다. 단, 이 때 세션의 승인과 성공적 해제를 알리는 200 OK 메시지는 같은 메시지라고 가정을 한다. 이 개수는 참고문헌[4]에서 분석한 음성 서비스를 위한 제어 메시지의 수(INVITE, ACK 5개, BYE, TRYING, RINGING, 총 9개)에 비하면 작은 값을 가짐을 알 수 있다. 위에서 열거한 5개의 메시지를 제외한 나머지 메시지들은 가입자와 릴레이서버 간에 일어나는 메시지로서 이 들 메시지를 처리하는데 걸리는 시간은 소프트스위치의 입장에서 보면 메시지 도착 지연시간에 불과하다. 결국 소프트스위치에게는 메시지 도착간격이 커지는 효과로 가정할 수 있기 때문에 음성 서비스를 위한 제어 메시지에 비해 낮은 부하로 작용한다. <그림 5>는 IM 서비스를 위한 제어 메시지의 흐름을 나타낸 것이다. <그림 6>은 IM 서비스를 위한 세션 해제 절차를 나타낸 것이다. <그림 7>은 IM 서비스를 위한 제어 메시지의 전체 흐름도이다.

Ⅲ. 소프트스위치 호 처리 차별화 모델 및 지연해석

소프트스위치의 호 처리 성능을 모델화하기 위하여 몇 가지 파라미터를 정의한다. 가입자의 규모가 충분히 큰 네트워크를 가정하면 음성과 데이터 제어패킷의 도착분포는 Poisson 과정을 따른다. 음성 및 데이터 제어패킷의 평균 도착률은 각각 λ_v 및 λ_d이다. 패킷의 서비스 시간 분포는 음성과 데이터 간에 상호 독립이며 모두 지수분포를 따르며 평균 서비스 시간은 음성과 데이터 제어패킷이 같은 평균치를 가지며 $1/\mu$라고 가정한다. 서비스 시간의 분산을 σ^2라고 한다면 $\sigma^2=1/\mu^2$이다[9].

IM 서비스는 기존의 최선형 IP 망에서 메신저 서비스와 같은 서비스로서 NGN 환경에서도 현재와 마찬가지로 제공되어야 할 서비스이다. IM 서비스는 주로 사무실 PC 환경에서 서로 메시지를 주고 받는 형태로 통신하는 상대가 PC에서 작업을 하고 있다는 전제 하에서 전화를 하는 것 보다는 비실시간적이고 메일을 송수신

하는 것 보다는 실시간적인 서비스로 쉽게 상대와 통신할 수 있다는 장점을 가지고 시작된 서비스이다. IM 서비스가 비실시간성을 지닌 서비스인 관계로 현재로서는 IM 서비스를 위한 연결 설정시간에 대한 목표치가 없는 상태이다. 따라서 음성통신보다는 그 우선순위가 낮은 데이터 통신 서비스로 분류할 수 있다.

위에서 설명한 바와 같이 NGN 서비스가 다양화됨에 따라 실시간성을 요구하는 서비스와 그렇지 않은 서비스를 차별화함으로써 각 각의 요구성능을 만족시키는 문제가 중요한 과제로 대두되었다. 따라서 향후 소프트스위치에서도 가입자 혹은 호에 대하여 적절한 수의 클래스로 구분하는 우선순위를 지정하여 우선순위에 의거하여 서비스를 차별화하여 처리할 필요가 있다.

소프트스위치를 구성하는 방법은 여러 가지가 존재할 수 있다. 하나의 소프트스위치는 단일 프로세서 또는 복수 프로세서로 구성할 수 있으며, 시스템도 단일구조 또는 분산구조로 구성이 가능하며, 시스템의 용량 및 성능은 모듈단위로 확장이 가능하다. 따라서 최적의 네트워크 운용을 위해서는 차별화되는 서비스 간의 상호관계를 정량적으로 규명하여 최적의 자원을 설계하여 운용하여야 한다. 아래에서는 메시지의 서비스방식에 대한 몇 가지 대안을 소개하고 각 방식에 대하여 예상되는 지연성능을 예측하는 방안을 제안한다.

1. 단일 프로세서/서비스 크래스(Single Processor/Service Class : SPSC)구조

단일 프로세서/서비스 구조에서는 음성 서비스와 데이터 서비스를 위한 제어 메시지를 차별화하지 않고 같은 종류의 메시지로 인식하여 FIFO(First-In-First-Out)로 서비스 하는 구조이다. 음성과 데이터 서비스를 위한 제어 메시지의 도착률을 각각 λ_v, λ_d라 하고 서비스률은 같다고 하였으므로 μ라고 하면 SPSC 구조는 음성 및 데이터 메시지가 구별이 없이 단일 버퍼로 중첩되므로 중첩 도착률이 $\lambda=\lambda_v+\lambda_d$이고 서비스률이 μ인 M/M/1 대기행렬 시스템으로 모델링 할 수 있다. 그리고 시스템의 부하 ρ는 $\rho=\lambda/\mu$로 정의되며 만약 $\rho<1$이라면 이 시스템은 안정상태에 있고 참고문헌[4]로부터 시스템 내부에서의 응답시간 D^s_{sw}(지연시간+서비스시간)는 아래의 식과 같다.

$$D^s_{sw} = \frac{1}{\mu - \lambda} \tag{1}$$

2. 이중 프로세서/서비스(Dual Processor/Single service Class : DPSC) 구조

이중 프로세서/서비스 구조에서는 SPSC 구조와 같이 음성과 데이터 메시지를 하나의 버퍼에 수용한 후 두 개의 서버를 통하여 FIFO로 동시에 두 개의 메시지를 서비스하는 구조이다. DPSC 구조는 SPSC 구조에 비해서 서버가 두 개이므로 메시지 처리능력이 두 배로 늘어나는 효과를 얻을 수 있다. 따라서 음성 서비스용으로 설계된 소프트스위치에 데이터 메시지가 첨가됨에 따른 부하를 경감시키기 위해서 서버의 용량을 두 배로 늘이는 효과를 준다. 그러나 이 방식은 어느 한쪽의 서비스 요구가 순간적으로 늘어나면 다른 한쪽의 성능에 영향을 미치는 단점이 있다. 또 이러한 구조에서는 서버를 두 개를 가져야 하므로 설계 비용이 커지고 부하가 낮은 운용환경에서는 시스템의 효율이 떨어진다.

DPSC 방식에서는 음성과 데이터 메시지가 도착순서에 따라 FIFO로 버퍼에 저장된 후 head of line에 있는 메시지와 바로 다음에 있는 메시지가 두 개의 서버(dual processor)에 의해서 동시에 서비스 받는 구조이다. 3.1절에서와 같은 가정을 적용하고 해석을 하면 음성과 데이터의 도착률을 각각 λ_v 및 λ_d라고 하였을 때 총 도착률 λ는 $\lambda= \lambda_v+ \lambda_d$가 되고, 서버의 부하 ρ는 $\rho=\lambda/\mu$로 정의한다. 만약 $\rho<1$이라면 이 시스템은 안정상태에 있고 시스템 내부에서의 응답시간 D^d_{sw}(지연시간+서비스시간)은 아래의 식과 같다[10].

$$D^d_{sw} = \frac{\rho}{\lambda(1-\rho)(1+\rho)} \tag{2}$$

3. 단일 프로세서 우선 서비스(Single Processor/Priority Service : SPPS) 구조

단일 프로세서 우선 서비스 구조에서는 음성과 데이터 메시지를 별개의 버퍼에 각각 수용한 후 음성 메시지를 데이터 메시지에 대하여 상대적으로 우선 서비스 하는 구조이다. <그림 8>은 SPPS의 동작원리를 나타낸 것이다. <그림 8>에서 패킷 분류기는 도착하는 가입자의 요구 메시지를 분석한다. 음성과 데이터 메시지의 제어패킷은 각 각 음성 버퍼와 데이터 버퍼로 분배된다.

각 각의 버퍼로 분배된 메시지를 우선 처리하는 방법은 절대적 우선순위와 상대적 우선순위를 적용하는 방법으로 나눌 수 있다. 절대적 우선순위는 높은 우선순위의 메시지에 대해서 절대적으로 우선해서 서비스를 하

표 1. 각 방식의 비교
Table 1. Advantages and disadvantages on each service type.

서비스 방식	장 점	단 점
절대적 우선서비스 (strict priority)	가. 낮은 순위의 메시지의 양에 무관하게 높은 순위의 메시지의 서비스에 대해 요구되는 품질을 확실히 보장할 수 있다. 나. 해석적 방법에 의한 성능의 예측 및 구현이 용이하다.	높은 순위의 메시지의 양에 따라 낮은 순위의 서비스의 품질이 크게 영향을 받는다.
상대적 우선서비스 (relative priority)	가. 높은 순위의 메시지의 양에 관계없이 낮은 순위의 서비스의 품질을 일정한 레벨로 유지할 수 있다. 나. 메시지간에 공평성을 보장한다. 다. 메시지의 양에 따라 적응적으로 서비스률을 조절 할 수 있다.	해석적 방법에 의한 성능의 예측 및 구현이 어렵다.

며, 상대적 우선순위는 높은 우선순위의 메시지와 낮은 순위의 메시지를 임의의 서비스 비율로 서비스하는 방식을 말한다. 각 방식은 서로 장단점을 가지고 있으며, 이들을 요약하면 <표 1>과 같다.

본 연구에서는 소프트스위치의 성능설계를 염두에 둔 경우를 가정해서 해석의 용이성과 음성 서비스에 대한 호 설정 지연목표치에 대한 보장에 중점을 두고 절대적 우선순위방식을 채택한다.

음성 버퍼의 메시지는 도착순으로 버퍼에서 기다린 후 앞의 메시지가 서비스 된 후 자신의 차례가 되면 바로 서비스에 들어간다. 한편, 데이터 버퍼의 메시지는 음성 버퍼에 메시지가 없을 경우에 한해서 서비스를 받을 수 있다.

패킷 서비스 정책은 위에서 설명한 절대적 우선순위를 따른다. 즉, 서버는 음성 버퍼와 데이터 버퍼의 패킷을 서비스를 함에 있어서 엄격한 우선순위를 적용한다.

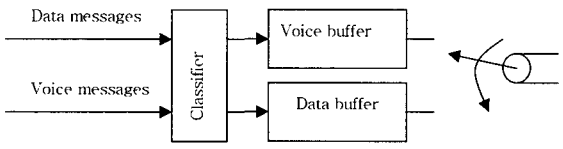

그림 8. SPPS구조.
Fig. 8. SPPS architecture.

먼저 서버는 음성 버퍼를 방문한다. 그리고 그 곳에 패킷이 존재하는 한 계속해서 그 곳의 패킷을 처리한다. 서버가 음성 버퍼의 패킷을 서비스 한 후 그 곳에 더 이상 패킷이 존재하지 않으면 즉시 데이터 버퍼로 이동하여 만약 패킷이 있다면 한 개의 패킷을 서비스 한다. 데이터 버퍼에서의 서비스가 종료되면 서버는 다시 음성 버퍼를 탐색한다. 그리고 위의 동작을 되풀이한다. 데이터 패킷의 서비스 도중에 음성 패킷이 음성 버퍼에 도착하더라도 일단 서비스를 시작한 데이터 패킷은 서비스가 끝날 때까지 계속 서비스를 받는다. 그리고 버퍼간 이동시간은 무한히 작은 것으로 가정한다.

W_v을 음성제어패킷의 버퍼에서의 대기시간이라고 가정한다. M/G/1 대기행렬에서 절대적 우선서비스 정책에 대한 지연성능의 해석은 높은 우선순위의 버퍼의 지연시간 W_v가 임의의 목표 임계치를 넘지 않도록 하는 것을 성능목표로 하여 해석하는 방법을 취한다. 대기행렬 해석 결과는 아래와 같다[9].

$$W_v = \frac{\lambda_v + \lambda_d}{\mu(\mu - \lambda_v)} \tag{3}$$

한편, 노드에 도착한 메시지는 자신보다 먼저 도착하여서 버퍼에 저장된 후 서비스를 받기를 기다리고 있는 메시지와 도착 당시 서버에서 서비스를 받고 있는 메시지의 서비스가 끝이 난 후에 서비스를 받을 수 있으므로, 노드에 도착한 후 서비스가 완료될 때까지 소요되는 시간(이를 D_{sw}^{sp}라고 둔다)은 다음과 같이 구할 수 있다. 음성 패킷이 시스템 내에서 체재하는 총 평균시간 D_{sw}^{sp}는 식 (3)에 메시지의 평균서비스 시간을 더함으로써 구할 수 있고 식 (4)는 그 결과이다.

$$D_{sw}^{sp} = W_v + \frac{1}{\mu} \tag{4}$$

<그림 3>의 등가지연모델에서 D_{RTT}는 식 (5)과 같이 정의할 수 있다. 식 (5)에서는 소프트스위치가 송수신단의 양측에 하나씩 있는 환경을 가정한 것이다. 실제로 네트워크를 구성할 때에는 네트워크의 형상에 따라서 두 개 또는 여러 개를 거쳐서 가는 경우도 있으나 편의상 두 개로 가정하였다.

$$D_{RTT} = 4 \times D_{sw} + 2 \times D_{IP} + D_{TE2} \tag{5}$$

단, 식 (5)에서 D_{sw}는 위에서 제시한 세 가지 서비스

방식에 무관하다고 가정을 할 수 있다 왜냐하면, 실제 수치실험에서는 D_{sw}가 소프트스위치가 허용할 수 있는 지연 목표치이기 때문에 서버의 방식에 관계없이 일정한 값을 대입하기 때문이다. 한편, 식 (3), (4) 및 (5)로부터 음성메시지의 지연목표치 D_{RTT}를 만족시키기 위한 서버의 용량 μ를 구할 수 있으며 식 (6)과 같이 나타낼 수 있다.

$$\mu = \frac{D_{sw}\lambda_v + 1 + \sqrt{(D_{sw}\lambda_v + 1)^2 + 4D_{sw}\lambda_d}}{2D_{sw}} \quad (6)$$

식 (6)에서 D_{sw}의 목표치는 식 (5)로부터 식 (7)과 같이 구한 값을 적용한다.

$$D_{sw} = \frac{D_{RTT} - 2D_{IP} - D_{TE2}}{4} \quad (7)$$

Ⅳ. 수치실험 및 결과분석

호 설정 절차에서 IETF(Internet Engineering Task Force)가 정의하고 있는 메시지의 크기는 메시지 헤드와 메시지 바디에 들어가는 내용에 따라 달라질 수 있다. 그러나 참고문헌[11]에 의하면 메시지 바디의 내용이 크게 차이가 나지 않음을 알 수 있는데, 음성 및 데이터의 각 메시지의 크기는 <표 2> 및 <표 3>과 같다. <표 2>에서 보는 바와 같이 각 메시지의 크기를 비교해 보면 BYE 및 200 OK 메시지가 조금 큰 값을 가지고 그 나머지의 패킷은 서로 크기가 비슷함을 알 수 있다. 위에서 가정한 M/M/1 대기행렬 모델을 이용하여 메시지의 평균대기시간을 구함에 있어서 각 메시지의 크기를 따로 고려해서 패킷 서비스 시간을 계산하는 것은 이론적으로는 불가능하며, 따라서 위에서 가정한 메시지의 평균 처리시간이 평균이 1/μ인 지수분포를 따른다는 사실을 이용해서 서비스 시간의 평균치를 구하는 접근법이 가장 현실적일 것이다. 각 패킷의 길이의산술적 평균은 424byte이며, 만약 서버의 링크의 속도가 C라고 한다면, 패킷의 평균 서비스 시간은 1/μ= 3816/C에 해당한다.

표 2. 음성 서비스 제어 메시지 유형별 크기
Table 2. Voice service control message and size.

패킷유형	메시지크기 (bytes)	패킷유형	메시지크기 (bytes)
INVITE	348	TRYING(100)	323
ACK	374	RINGING(180)	423
BYE	484	OK(200)	589

표 3. IM 서비스 제어 메시지 유형별 크기
Table 3. IM service control message and size.

패킷유형	메시지크기 (bytes)	패킷유형	메시지크기 (bytes)
INVITE	635	(MSRP)BIND	66
ACK	301	(MSRP)VISIT	69
BYE	301	(MSRP)OK(200)	78
OK(200)	359	(MSRP)SEND	52+text body

IM 서비스를 위한 메시지의 패킷 유형별 크기는 헤더 필드에 optional field가 있고 각 필드의 길이가 가변이다. 또 네트워크 내에서 거쳐야 할 hop 수가 늘어나면 via field 등이 추가되어 패킷의 길이가 그에 따라 늘어나므로 정확한 메시지의 크기는 경우에 따라 차이가 있다. 일반적으로 알려져 있는 결과로는 메시지가 평균 1-2 hop 정도를 거쳐서 전달된다고 알려져 있다. 따라서 기본적으로 사용되는 option field만으로 고려하여 정리하면 <표 3>과 같다. <표 3>에서 소프트스위치와 관계되는 메시지인 4가지 종류의 메시지(INVITE, 200 OK, ACK, BYE)의 길이를 관찰해 보면 음성 서비스를 위한 제어 메시지의 길이와 거의 같은 패턴을 따름을 알 수가 있다. 따라서 IM 서비스를 위한 제어 메시지의 경우에도 메시지의 평균 처리시간이 평균이 1/μ인 지수분포를 따른다는 가정을 사용할 수 있음을 알 수 있다.

한편, 세계 각국에서는 IP 네트워크의 SLA(Service Level Agreement)로서 백본 네트워크에서의 단대단 지연을 규정하고 있는데, 미국의 경우는 동서횡단 최대 허용 지연치로 약 80ms를 규정하고 있고 일본의 경우는 일본열도의 대륙 내에서 약 35ms를 규정하고 있다[12]. 우리나라는 일본보다도 국토의 종단간 거리가 더욱 짧으므로 IP 백본 내에서의 단대단 허용지연은 35ms를 넘지 않을 것으로 추정된다. 국내에서는 이에 대한 품질기준이 아직 수립되지 않은 상황이므로 수치실험을 위하여 D_{IP}를 본 연구에서는 35ms를 가정한다. D_{RTT}의 값은 ETSI의 요구치를 가정하기로 한다[13]. 지연품질을 Best class로 가정하는 경우 단방향 지연목표치가 E.164 Number translation to IP Address에 대해서 2초 이내

로 규정되어 있으므로 D_{RTT}의 값은 4초가 된다. 한편, D_{TP2} 값은 PC의 성능에 따라 크게 의존하나 1초 이내로 가정을 한다. 음성 서비스를 위한 제어 메시지와 IM 서비스를 위한 제어 메시지의 비율은 가변적이나 계산의 편이를 위하여 IM 서비스를 위한 제어 메시지는 백그라운드 트래픽으로 가정을 하여 도착률을 0.5로 고정시킨다. 반면, 음성 서비스를 위한 제어 메시지는 0.1부터 0.4까지 변하는 것을 가정하여 전체 도착률(input rate)이 0.6에서 0.9까지 변하는 시스템을 가정할 때 소프트스위치의 프로세서가 필요로 하는 소요용량 (output rate)을 계산한다. <그림 9>는 각 방식별 소요용량을 나타낸 것이다.

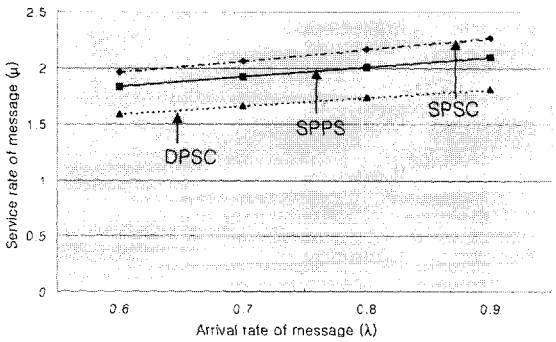

그림 9. 방식별 메시지 도착률과 서비스률의 관계
Fig. 9. Relationship between message arrival rate and message service rate

각 방식의 결과를 비교하면 서버의 소요용량은 SPSC, SPPS 및 DPSC의 순으로 큰 용량을 필요로 한다. 반면에 음성메시지의 지연보장에 대한 확실성은 그와는 역순이다. 이는 서버 시스템의 설계 및 네트워크 디멘저닝에서 나타나는 성능과 비용의 역관계에 대한 trade off 문제를 단적으로 보여주는 결과이다.

<그림 9>의 결과를 좀 더 자세히 분석하면 다음과 같은 결론을 도출할 수가 있다. 소프트스위치의 메시지 도착률 λ= 0.6일 때, D_{RTT}의 값을 4초 이내로 유지하기 위해서 소프트스위치의 서버에 필요한 소요용량은 SPSC, SPPS 및 DPSC의 각 방식에 대하여 각각 μ= 1.97, 1.84 및 1.6이며 이 때 소프트스위치의 최대 수용가능한 부하는 ρ=0.30, 0.33 및 0.38이 된다. 다시 말하면, D_{RTT}의 값을 4초 이내로 유지하기 위해서 소프트스위치가 수용할 수 있는 부하는 메시지의 서비스방식에 따라 0.30에서 0.38까지 다양한 범위의 값을 가짐을 알 수 있다. 이 경우에 SPSC는 허용가능 부하가 0.30으로 일반적인 데이터 네트워크에서 과도한 지연을 방지하기 위하여 요구되는 부하인 0.5 룰(부하가 0.5를 초과하면 지연이 급격히 커지는 현상) 보다 훨씬 더 작은 부하를 요구한다는 사실을 알 수 있다[14]. 부하가 이와 같이 낮아야 하는 이유는 SPSC가 음성과 데이터메시지 간에 서비스 구별이 없는 경우이기 때문에 음성 패킷에 대하여 위에서 가정한 바와 같이 낮은 지연을 보장하기 위해서는 서버의 효율을 낮출 수 밖에 없음을 나타낸다.

한편, DPSC의 경우에는 허용가능 부하가 0.38로 가장 높은 값을 유지하는데 이 경우는 서버가 두 개이므로 SPSC 보다 더 많은 양의 패킷을 수용할 수 있기 때문이다. 마지막으로 메시지에 우선순위를 도입하여 차별화 서비스를 제공하는 SPPS의 경우에는 <그림 9>에서 나타내었듯이 평균부하는 33%를 넘지 않아야 한다. 따라서 SPPS는 SPSC와 DPSC의 중간에 해당하는 성능을 가지고 있다.

위의 세 가지 방식을 비교하면 우리는 SPPS 방식이 가장 매력적인 해가 됨을 알 수 있는데 그 이유는 SPPS 방식이 SPSC의 문제점인 음성패킷에 대한 우선 서비스가 없다는 점을 해결하면서도 서버의 용량이 SPSC 보다 작아도 된다는 장점을 가짐을 알 수 있다. 또한 DPSC 방식은 서버를 두 개나 사용하면서도 역시 음성 패킷에 우선 서비스를 할 수 없었지만 SPPS는 DPSC 보다 적은 서버로 음성패킷에 대하여 만족할 만한 지연성능을 보장할 수 있으므로 결과적으로 SPPS 방식이 가장 적절함을 알 수 있다.

본 연구가 가지는 의미는 NGN의 소프트스위치가 호 설정 지연에 민감한 음성 서비스와 지연이 덜 민감한 데이터 서비스를 동시에 수용하는 환경에 대하여 음성 호 제어 메시지에 대하여 일정한 호 설정 지연목표치를 만족시킬 수 있도록 소프트스위치의 호 제어 메시지 도착률과 소프트스위치 서비스률 간의 관계를 정량적으로 나타내었다는 점에 주목할 필요가 있다. 따라서 향후 네트워크 사업자가 서비스를 도입할 때 수용 가능한 호의 규모를 예측 할 수 있을 경우에 호 설정 지연목표치와 호의 규모에 대한 통계치를 기반으로 본 연구에서 제안한 성능평가방법을 사용하면 소프트스위치의 적정 용량을 설계 할 수 있다. 반대로 소프트스위치의 용량과 지연목표치가 정해진다면 역으로 소프트스위치가 수용 가능한 음성 및 데이터 서비스 호의 규모를 예측 할 수 있을 것이다.

V. 결 론

본 연구에서는 NGN에서 다양한 서비스의 호/연결 제어를 담당하는 소프트스위치에서 음성 서비스와 데이터 서비스를 동시에 수용하는 경우를 가정하고, 이들 호/연결 제어와 관련된 호 설정 지연에 대하여 해석적인 방법을 이용하여 정량적으로 성능을 평가하는 방법을 제안하였다. 호/연결 제어에 대하여 호의 설정 및 해제에 이르는 각 단계에 대한 메시지의 교환절차 및 내역에 대한 분석을 수행하고 그 중 호 설정지연에 더욱 민감한 음성 서비스에 대하여 정해진 지연목표치 이내에 호를 처리하기 위한 서버의 메시지 처리방안에 대하여 몇 가지 대안을 가정하였다. 그리고 각 방안에 대한 해석을 수행하여 서버의 적절한 용량을 설계할 수 있도록 하였다. 또한 수치실험을 통하여 각 방안의 장단점을 비교분석하였다.

본 연구에서 제안한 방법을 이용하면 NGN에서 대표적인 실시간/비실시간 응용 서비스환경에서 호 레벨 서비스품질로 정의되는 호의 설정 지연목표치를 보장할 수 있도록 소프트스위치의 용량을 설계할 수 있을 것으로 기대된다. 그리고 소프트스위치의 서버에 다한 정량적 성능분석의 한 방법을 제시하였다는 점에 그 의의가 있으며, 앞으로 해결하여야 할 문제가 많이 남아있다. 그 중 대표적인 것으로는 NGN의 구조가 정해지면 IP 네트워크 내부에서의 지연요소를 좀 더 자세히 분석할 필요가 있고, 또 소프트스위치가 수용 가능한 가입자의 규모를 예측할 수 있으면 실제 서버 설계에 적용이 가능한 결과를 도출하는 방안을 찾아내는 것이 있다.

참 고 문 헌

[1] Frank Erfurt, "How to Make a SoftSwitch Part of the Distributed Service World of Converged Networks," Intelligent Network Workshop, 2001 IEEE, May 2001.

[2] Kyung Hyu Lee et al, "Architecture To be Deployed on Strategies of Next Generation Networks," Proc. of IEEE International Conference on Communications 2003, vol. 2, May 2003.

[3] Abdi R. Modarressi and Seshadri Mohan, "Control and Management in Next-Generation Networks: Challenges and Opportunities," IEEE Communications Magazine, vol. 38, issue 10, October 2000.

[4] 정문조, 황찬식, "NGN에서 음성서비스의 호 처리 성능해석," 대한전자공학회논문지, 2003년 11월

[5] 이훈, "NGN에서 품질보장형 음성서비스의 제공방안," KT 기술연구소, 2002년 9월

[6] P. Nain, Performance evaluation of Computer Systems, Lecture Notes of U.of Massachusetts, 1995.

[7] 이준원, "Instant Messaging Model," 2003 SIP 기반 VoIP 기술 하계세미나, 2003년 8월

[8] "Session Initiation Protocol (SIP) Extension for Instant Messaging," IETF RFC3428, December 2002.

[9] Hoon Lee, Lecture note in advanced queuing systems, Graduate school of IT at Changwon National University, 2003.

[10] 이호우, 대기행렬이론 -확률과정론적 분석-, 시그마프레스, 1998

[11] Uyless Black, Chapter 11 The Session Initiation Protocol(SIP), Internet Telephony-Call Procession Protocols, Prentice Hall PTR, 2001.

[12] Takizawa, "Focus on Internet QoS," Nikkei Communications, 1999.6.21.

[13] ETSI TR 101 329 V2.1.1, "Telecommunications and Internet Protocol Harmonization Over Networks(TIPHON);General aspects of Quality of Service(QoS)," June 1999.

[14] C. Filsfils and J. Evans, "Engineering a multiservice IP backbone to support tight SLAs," Computer Networks, vol. 40, 2002.

저 자 소 개

鄭 文 照(正會員)

1988년 : 경북대학교 전자공학과(공학석사). 1988년~현재 : KT 기술연구소. <주관심분야 : 차세대네트워크(NGN) 구조, 차세대네트워크 제어 및 운용관리>

黃 燦 植(正會員)

1979년 : 한국과학기술원 전자공학과(공학석사). 1996년 : 한국과학기술원 전자공학과(공학박사). 1979년~현재 : 경북대학교 전자전기컴퓨터학부 교수. <주관심분야 : 영상통신, 암호통신, 초고속망>