論文2003-40TC-11-6

NGN에서 음성서비스의 호 처리 성능해석

(Performance Analysis for Call Processing in NGN Voice Services)

鄭 文 照 *, 黃 燦 植 **

(Mun Jo Jung and Chan Sik Hwang)

요 약

최근 차세대네트워크(Next Generation Network : 이하 NGN)란 이름으로 음성서비스 및 다양한 멀티서비스(multi-service)를 품질이 보장되는 IP 네트워크 상에서 제공하려는 움직임이 활발하다. 그리고 NGN에서 원활한 서비스 제공을 위하여 해결되어야 할 문제들은 전달망 구조 정립, IP 망의 QoS 보장, 서비스 생성 및 관리, 그리고 제어채널의 분리에 따른 제어망 설계분야로 나누어져 연구가 진행되고 있다. 특히, NGN에서는 호 및 연결 제어정보와 사용자 데이터를 전달하는 루트가 서로 분리된 형태로 구현될 예정이며, 호 처리 성능평가에 대한 방안은 조속히 정리되어야 할 주요한 문제 중의 하나이다. 그러나 아직 호 처리 성능에 대한 어떠한 정량적 평가방법도 제안되어 있지 않다. 본 연구에서는 NGN에서 음성서비스를 제공하기 위한 제어망 설계에 있어서 QoS 지표의 하나인 호 설정 지연 목표치를 만족시키기 위하여 소프트스위치가 갖추어야 할 서버의 성능을 정량적으로 분석하고 예측하는 방법을 제안한다.

Abstract

In this paper we propose a method of evaluating the performance of a Softswitch that provides call control to voice services in NGN (next generation network). First, we describe the architecture for voice services in NGN and anatomize the call control processes such as call initiation, call re-initiation and call release of a voice connection. After that we propose a method of estimating appropriate server capacity of the Softswitch using approximate queuing model. Via numerical experiments we illustrate the implication of the work

Keywords : Softswitch, Call Processing, Performance modeling and analysis

Ⅰ. 서 론

최근에 들어서 차세대네트워크(Next Generation Network : 이하 NGN)의 구현에 대한 움직임이 활기를 띠기 시작하였다. NGN은 각기 다른 기술로 출발한 기존의 음성전화와 인터넷 데이터 서비스, 그리고 향후 도입이 유망한 영상전화 등 별개의 네트워크에 의해서 제공되는 음성, 데이터 및 비디오의 다양한 서비스를 하나의 공통된 IP 네트워크 상에 구현하고자 하는 기술이다. 한편, NGN은 가입자간 또는 가입자와 응용 서버간의 베어러 신호는 QoS가 보장되는 IP 망을 통하여 전달되고, 가입자단말 간의 통신을 위한 제어, 즉, 호처리는 소프트스위치를 통하여 처리되는 망이다. 이는 기존의 교

* 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8月14日, 수정완료일:2003年11月1日

환기에서 제어신호와 베어러 신호의 처리를 단일 시스템 내에서 하던 방식에서 가입자를 접속하는 접속장치부와 망으로 접속링크를 제공하는 중계장치부, 그리고 시그널링을 처리를 하는 제어부를 각 각 분리된 시스템으로 독립시켰다는 부분이 크게 다르다.

IP 네트워크에서 음성패킷에 대해서 QoS가 보장되는 서비스를 제공하기 위한 성능해석과 QoS 보장 문제는 가입자 접속부 및 백본 네트워크부의 양 분야에 대해서 오래 전부터 많은 연구가 되어왔다[6-8]. 그러나 최근에 대두된 NGN에서 소프트스위치를 통하여 처리되는 가입자단말 간의 통신을 위한 제어에 대한 성능분석에 대한 연구는 참고문헌[3, 12]을 제외하고는 거의 이루어져 있지 않다. 참고문헌[3]에서는 IP 네트워크에서 음성호 처리 시스템의 성능평가를 위한 네트워크의 구성과 SIP(Session Initiation Protocol)을 중심으로 호 제어신호 프로토콜에 대해서 기술을 하고 호 처리 시스템이 가져야 할 기능과 예상되는 성능에 대한 평가항목을 정의하였다. 그러나 참고문헌[3]에서는 호 처리 시스템의 성능에 대한 어떠한 정량적 평가방법도 제안되어 있지 않다. 한편, 참고문헌[12]에서는 H.323 프로토콜에 대하여 IP 백본 내에서 발생하는 신호전달지연에 대하여 시뮬레이션을 통한 성능예측을 하였다. 그러나 참고문헌[12]에서는 소프트스위치의 성능분석에 대해서는 언급이 없을 뿐만 아니라 실제로 IP 백본은 대용량의 대역과 초고속의 라우터로 구성되어 있고 호제어와 같은 신호 트래픽은 ITU-T Y.1541 의 QoS 보장 구조 및 IETF의 DiffServ구조에서 다른 트래픽에 비해서 상대적으로 우선 서비스를 받도록 되어 있기 때문에 IP 백본 내에서의 지연은 거의 고정치에 가까운 값으로 안정적으로 동작할 것으로 예상된다[13].

본 연구에서는 이점에 착안하여 NGN에서 VoIP (Voice over IP) 서비스를 제공하기 위하여 핵심요소이면서 병목이 예상되는 소프트스위치의 호 처리 성능에 대한 정량적 성능평가방법을 고안하고 대기행렬이론에 근거한 성능평가모델을 제시하여 VoIP 호제어 메시지에 대한 지연을 보장하기 위한 소프트스위치 적정 부하를 예측하는 방법을 제시함으로써 향후 서비스가 도입될 때 사전에 적절한 호처리 시스템의 용량을 설계하는데 사용할 수 있는 이론적 기반을 구축하고자 한다.

본 고의 구성은 다음과 같다. 제II장에서는 NGN에서 음성서비스의 구조를 설명한다. 제III장에서는 NGN 제어평면에서의 호처리 절차를 기술한다. 제IV장에서는 제III장에서 기술한 호처리 절차에 대한 성능평가를 위한 모델을 제안한다. 제V장에서는 제IV장에서 제안된 모델을 이용한 수치실험결과를 기술한다. 마지막으로 제VI장에서는 본 연구의 의미를 살펴보며 향후의 연구방향에 대하여 기술한다.

II. NGN 음성 서비스 구조

NGN이란 가입자간 또는 가입자와 응용서버간의 베어러 신호는 QoS가 보장되는 IP 네트워크를 통하여 전달되고, 가입자단말 간의 통신을 위한 제어, 즉, 호처리는 소프트스위치를 통하여 처리되는 망이다. 이는 기존의 교환기에서 제어신호와 베어러 신호의 처리를 단일 시스템 내에서 하던 방식에서 가입자를 접속하는 접속장치부와 망으로 접속링크를 제공하는 중계장치부, 그리고 시그널링을 처리를 하는 제어부를 각 각 분리된 시스템으로 독립시켰다는 부분이 크게 다르다. 또한 제어신호 및 사용자 데이타를 IP 네트워크를 통하여 교환하도록 설계되었다는 점이 크게 차이가 나는 부분이다.

NGN을 통하여 전달될 수 있는 데이터의 형태는 음성, 영상, 텍스트 등이다. 그러나 NGN이 구축되는 과정에서 초기에 제공될 실시간성 서비스로서는 음성서비스가 될 것으로 사료된다. <그림 1>에 초기 NGN에서 음성서비스를 제공하기 위한 네트워크 구조 및 기능을 나타내었다[5].

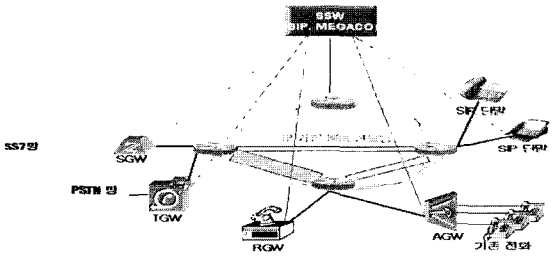


그림 1. 음성 서비스 제공을 위한 NGN 네트워크 구조
Fig. 1. NGN physical configuration for voice service.

<그림 1>에 나타낸 것과 같이 NGN 망요소로서 소프트스위치는 PSTN 전화 가입자와 SIP 프로토콜을 이용한 전화 가입자에 대한 호 제어를 제공하게 된다. AGW (Access Gateway)는 기존 가입자 전화를 집선하여 망으로 접속하는 가입자 접속 및 집선기능을 수행하는 장치이다. SGW(Signaling Gateway)는 No.7 신호방식과 PSTN 시그널링망과 연동하는 시그널링 연동장치이다.

TGW(Trunk Gateway)는 NGN내 전화가입자와 PSTN 전화 가입자간 음성을 전달하는 베어러 채널에 대한 연동 기능을 수행하는 장치이다. 소프트스위치와 TGW, AGW 간에는 MEGACO(MEdia GAteway COntrol) 프로토콜을 사용한다. 또한 NGN 망에서 전화 서비스를 위한 또 다른 장치로는 SIP 단말이 있다. SIP 단말기는 IP 기반의 단말장치로 호제어는 SIP 프로토콜을 사용한다. IP 단말기를 가진 가입자를 집선하는 가입자 접속장치로는 RGW(Residential Gateway)가 있으며 실제로 망에 도입될 액세스 노드의 실제 모습은 다양한 형태로 전개될 것으로 예상된다. 소프트스위치와 소프트스위치 간에는 SIP을 사용한다.

<그림 1>에서 보는 바와 같이 소프트스위치는 IP 망 요소인 라우터에 접속이 되고, 소프트스위치와 AGW, TGW, SGW, RGW 및 SIP 단말간에 논리적으로 각각 독립된 제어경로가 구성이 된다. 소프트스위치와 AGW 및 TGW 간에는 MEGACO 프로토콜이 사용이 되고, 소프트스위치와 SGW 간에는 SIGTRAN 프로토콜이 사용된다. 그리고 소프트스위치와 IP 단말, RGW간에는 SIP 프로토콜이 사용된다.

본 고에서는 NGN 망에서 SIP 단말을 이용한 음성전화 서비스를 전제로 SIP 단말 간에 호의 설정, 유지 및 해제에 필요한 제어메시지의 처리에 대한 지연성능을 모델링을 하고 이를 이용하여 수치 실험결과를 통해 제안하는 방법의 의미를 살펴보고자 한다.

Ⅲ. NGN 제어평면에서의 호 제어 절차

NGN에서 전화 서비스를 위한 호 제어 프로토콜은 MEGACO, SIP, SIGTRAN 등이 있다[5]. 본 고에서는 향후 NGN에서 가장 일반적인 프로토콜로서 널리 사용될 것으로 보이는 SIP에 대하여 그 동작을 기술한다.

SIP를 이용한 음성 호 처리 절차는 일련의 메시지들이 가입자단말(SIP전화기)-소프트스위치-가입자단말 간을 순서적으로 오고 감으로써 성립이 된다. 송수신단 간의 호 제어 메시지는 크게 호설정, 통화연장 및 호해제의 3단계로 구성이 되는데, 각 단계의 메시지 전달순서를 요약하면 다음과 같다.

호설정 메시지 전달 순서:

① SIP 전화기1은 송.수신 목적지주소(전화번호, E-mail address, IP address 등)를 INVITE 메시지에 포함하여 소프트스위치1로 송신한다.

② 소프트스위치1은 해당 목적지를 확인하고 소프트스위치2로 INVITE 메시지를 전달한다.

③ 소프트스위치2는 SIP전화기2로 INVITE 메시지를 전달한다.

④ SIP전화기2는 180 Ringing 메시지를 소프트스위치2로 송신한다.

⑤ 소프트스위치2는 소프트스위치1로 180 Ringing 메시지를 전달한다..

⑥ 소프트스위치1은 SIP전화기1로 180 Ringing 메시지를 전달한다.

⑦ SIP전화기1은 180 Ringing 신호를 울린다.

⑧ SIP전화기2가 호를 받아들일 수 있으면 200 OK 메시지를 소프트스위치2로 송신한다.

⑨ 소프트스위치2는 수신한 200 OK 메시지를 소프트스위치1로 전달한다.

⑩ 소프트스위치1은 200 OK 메시지를 SIP전화기1로 전달한다.

⑪ 200 OK를 수신한 SIP전화기1은 ACK 메시지를 소프트스위치1로 송신하고 소프트스위치1은 수신한 ACK 메시지를 소프트스위치2로 전달한다.

⑫ 소프트스위치2는 ACK 메시지를 SIP전화기2로 전달하고 호설정이 완료된다.

SIP에서는 호 설정 시에 세션에 대한 시간(통화시간)을 지정할 수 있고 필요에 따라서 지정된 시간이 완료되어 호가 해제되기 이전에 발신가입자가 세션에 대한 통화시간을 재조정할 수 있다. 세션에 대한 시간을 재조정하기 위하여 RE-INVITE 메시지를 사용한다.

통화연장 메시지 전달 순서:

① SIP전화기1은 통화의 연장을 요구하는 RE-INVITE 메시지를 소프트스위치1로 송신한다.

② 소프트스위치1은 소프트스위치2로 RE-INVITE 메시지를 전달한다.

③ 소프트스위치2는 SIP전화기2로 RE-INVITE 메시지를 전달한다.

④ SIP전화기2는 200 OK 메시지를 소프트스위치2로 송신한다.

⑤ 소프트스위치2는 수신한 200 OK 메시지를 소프트스위치1로 전달한다.

⑥ 소프트스위치1은 200 OK 메시지를 SIP전화기1로 전

달한다.

⑦ 200 OK를 수신한 SIP전화기1은 ACK 메시지를 소프트스위치1로 송신하고 소프트스위치1은 수신한 ACK 메시지를 소프트스위치2로 전달한다.

⑧ 소프트스위치2는 ACK 메시지를 SIP전화기2로 전달하고 통화연장이 완료된다.

만약 통화연장을 다시 하려면 위의 ①~⑧의 절차를 반복하여야 한다.

마지막으로 호의 해제를 위한 메시지의 순서는 아래와 같다.

통화해제 메시지 전달 순서:

① SIP전화기1은 호의 해제를 요구하는 BYE 메시지를 소프트스위치1로 송신한다.

② 소프트스위치1은 소프트스위치2로 BYE 메시지를 전달한다.

③ 소프트스위치2는 SIP전화기2로 BYE 메시지를 전달한다.

④ SIP전화기2는 200 OK 메시지를 소프트스위치2로 송신한다.

⑤ 소프트스위치2는 수신한 200 OK 메시지를 소프트스위치1로 전달한다.

⑥ 소프트스위치1은 200 OK 메시지를 SIP전화기1로 전달한다.

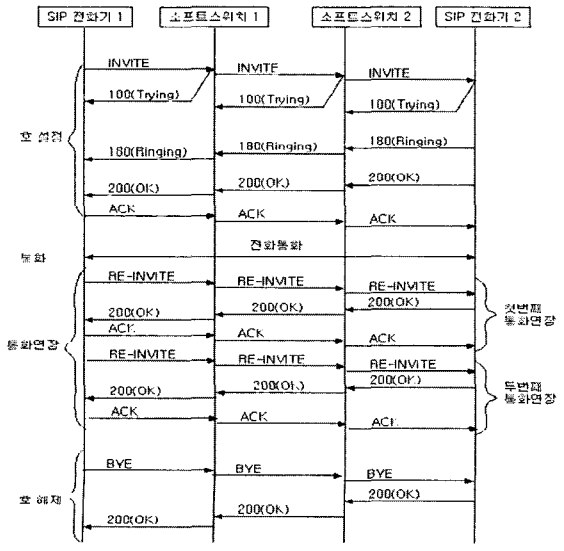


그림 2. SIP를 이용한 호 제어 절차
Fig. 2. SIP call control procedure.

SIP를 이용한 호처리 절차의 예를 <그림 2>에 나타내었다.

Ⅳ. 성능평가 모델

본 논문에서 제안하고자 하는 성능평가의 모델을 위하여 Ⅱ장과 Ⅲ장의 분석을 기반으로 실제 제어 메시지가 흐르는 과정을 살펴보고 지연요소에 대한 모델링을 하고자 한다. 이를 위하여 <그림 2>의 호 설정 절차를 대상으로 하여 소프트스위치의 성능을 평가하기 위하여 <그림 2>의 메시지의 송수신단간의 접속구조를 <그림 3>과 같이 나타낸다. <그림 3>에서 송신자는 TE1 (Terminal equipment 1)이고 수신자는 TE2로 가정하고, 메시지의 흐름을 송신측에서 수신측으로 향하는 메시지 (Source-to-Destination Message, 이하 S2D 메시지라 함) 는 실선으로 나타내고, 수신측에서 송신측으로의 응답 메시지 (Destination -to- Source Message, 이하D2S 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.

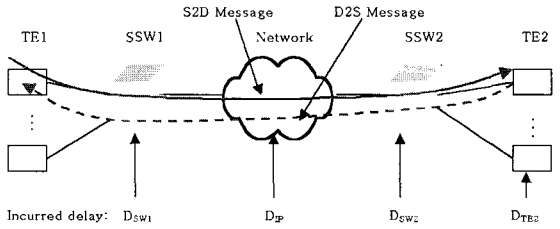


그림 3. 성능평가 모델
Fig. 3. Performance analysis model.

이들 메시지의 시퀀스는 크게 두 가지 경우로 나눌 수 있다. 한 가지는 호 설정 요구 시에 정한 시간 내에 통화를 마치는 경우이고 다른 한 가지는 <그림 2>의 통화연장에서 나타낸 것과 같이 호 설정 요구 시에 정한 시간 내에 통화를 마치지 않고 RE-INVITE 메시지를 이용하여 통화시간을 늘여서 통화를 하는 경우이다. 본 논문에서는 이 두 가지 경우가 확률적으로 동일하게 일어난다고 가정을 한다.

<그림 3>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 Network, SSW2 및 TE2로 정의하고, 각 요소에서의 지연은 송수신방향에 대해서 균일하게 분포한다고 가정을 한다.

메시지가 SSW1, IP Network, 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 임의변수(random variable)이며 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}은 S2D 메시지가 D_{SW1}, D_{IP}, D_{SW2}, D_{TE2}의 합과 같다. 그러나 대부분의 제어 메시지는 [송신단말 → 수신단말 → 송신단말]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 근사화할 수 있다.

백본 네트워크에서의 지연 D_{IP}는 네트워크의 상황에 따라서 바뀌는 임의변수이나, 본 연구의 목적이 IP 백본의 전달 지연을 평가하는 것이 아니므로 백본 네트워크의 특성상 충분한 용량의 대역폭이 설계되어 있다고 가정을 하고 상수로 취급하기로 한다. 실제로 NGN에서 음성서비스에 대한 데이터 및 호설정을 위한 신호데이터의 서비스품질 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[6]. 그러나 복수개의 패킷이 중첩되는 노드에서 버퍼 내 대기 지연은 불가피한 상황이므로 ITU-T 및 제 외국 서비스사업자가 권장하는 지연목표치를 근거로 상수 값이 사용된다. 결국 D_{RTT}의 성능은 소프트스위치인 SSW1과 SSW2에서 일어나는 지연에 가장 크게 좌우된다. 메시지의 지연 목표치가 주어지면 상수항에 해당하는 네트워크지연인 D_{IP}와 수신측 단말지연인 D_{TE2}를 제외한 지연 목표치를 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 TE2 간의 호제어 메시지를 처리하는데 각 소프트스위치가 동시에 충분히 큰 규모의 복수의 연결을 독립적으로 처리하는 서버이기 때문에 메시지 시퀀스의 관계는 의미가 없게 되며, 서로 독립된 서버라고 가정을 할 수가 있다. 따라서 TE1에서 생성된 메시지가 소프트스위치와 IP 네트워크를 통하여 TE2에 전달된 후 다시 자신에게 돌아오는데 걸리는 지연은 <그림 4>와 같은 등가지연 모델로 나타낼 수가 있다. <그림 4>에서 D_{SW1}, D_{SW2}, D_{IP}, D_{TE2}은 각각 SSW1, SSW2, IP 네트워크 및 TE2에서의 지연을 나타낸다.

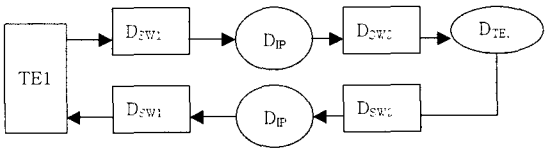


그림 4. 등가지연 모델
Fig. 4. Equivalent delay model.

<그림 4>에서 D_{IP}와 D_{TE2}는 상수의 값을 취한다고 앞에서 기술하였으므로 지연의 해석은 두개의 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 있다. 결과를 요약하면 M/M/1 대기행렬이 K개 직렬 연결된 경우의 D_{e2e}은 단일 노드의 지연의 K배이다[4]. 이 사실을 이용하면 결국 단대단 지연성능의 해석은 SSW1 혹은 SSW2에 대한 단일 노드해석으로 귀착된다. 송수신단간의 지연성능을 모델링함에 있어서 소스측에 유입되는 메시지가 그 원형을 가장 잘 보존하고 있으므로 SSW1에 대해서 논의를 전개하기로 한다.

앞 절에서 <그림 2>의 메시지를 세 가지로 구분하였는데 호의 설정을 위한 메시지로는 REGISTER 및 INVITE가 있다. REGISTER는 단말이 호 접속을 하기 전에 소프트스위치에게 자신의 위치를 등록하는 절차이며, 따라서 엄격한 의미에서 이 메시지는 호 설정 지연에 포함시키지 않는다. INVITE는 호 접속 요구 신호이므로 이 메시지는 지연성능에 포함되며, 이 메시지는 정

상적인 동작에서 호당 한 번씩 발생되는 메시지이다. INVITE 메시지의 발생에 대해서 180 Ringing 과 200 OK의 응답메시지가 소프트스위치로 부과된다. 그리고 단말에서 접속된 호의 연결상태를 연장하기 위한 메시지로서 RE-INVITE가 있는데, RE-INVITE 메시지의 발생간격은 호의 접속 시에 단말의 요구에 의해서 결정된다. 따라서 이 메시지는 2회 이상 발생할 수 있다. RE-INVITE 메시지에 대한 응답으로 200 OK 메시지가 소프트스위치에 부과된다. 마지막으로 단말에서 호 접속을 해제하는 메시지로서 BYE가 있다. 이 메시지는 발신측 단말이 통화를 완료한 후 호를 해제하겠다는 뜻을 소프트스위치에게 알리는 메시지이다. 만일 단말이 소프트스위치에게 호 해제 의사를 밝히지 않을 경우에는 소프트스위치가 일정한 시간임계치(세션타이머라고 함)만큼 기다린 후 호를 개방할 수 있다. 따라서 세션타이머가 임계치의 초과 여부를 검지하는 행위를 소프트스위치가 BYE 메시지를 처리하는 것으로 간주할 수 있다. 이 메시지도 REGISTER나 INVITE 메시지처럼 호당 한 번만 생성된다. 이 메시지 역시 응답메시지로서 200 OK 메시지가 소프트스위치에 부과된다.

위에서 분류한 세 가지 메시지에 대하여 다음과 같이 변수를 지정한다. REGISTER 메시지는 호의 설정 이전에 행해지므로 호 설정 지연에서 제외시킨다. INVITE 메시지의 도착률을 λ_I라 한다. INVITE 메시지에 대한 응답 메시지인 180 Ringing과 200 OK 메시지 도착률을 합쳐서 $\lambda_{I\ ACK}$로 정의한다. 그리고 200 OK 메시지에 대한 모든 응답을 λ_{ACK}로 정의한다. RE-INVITE 메시지 도착률은 λ_{RI}로 정의하자. λ_{RI}는 INVITE 메시지의 도착률에 의존하는데 호가 설정된 이후에 가입자의 통화지속시간의 장단의 여부에 따라 달라지므로 λ_{RI}는 λ_I의 값에 비례하는 임의변수로 볼 수 있다. 한 호당 RE-INVITE 메시지의 평균 발생횟수를 α로 가정하여 $\lambda_{RI} = \alpha\lambda_I$로 둔다. RE-INVITE 메시지의 응답인 200 OK 메시지의 평균발생률을 $\lambda_{RI\ ACK}$ 로 정의한다. 마지막으로 호의 해제를 위한 메시지인 BYE메시지의 도착률을 λ_B라고 하면 위의 정의에서 $\lambda_B = \lambda_I$임을 추측할 수 있다. BYE 메시지의 응답인 200 OK 메시지의 도착률을 $\lambda_{B\text{-}ACK}$라 둔다.

위에서 정의한 메시지의 전달 프로세스를 분석해 보면 하나의 음성호 접속과 해제를 위해서 TE1에서 TE2로 최소한 5개의 메시지가 전송되고, TE2에서 TE1으로는 최소한 4개의 메시지가 전송되어야 한다. 물론 RE-INVITE 메시지의 발신횟수가 늘면 그만큼 메시지 교환횟수는 늘어난다. 위에서 정의한 변수들은 소프트스위치가 수용하는 가입자의 수가 아주 큰 (수만 내지 수십만 가입자) 경우를 고려하면 하나의 가입자로부터 발생하는 일련의 메시지들이 여러 명의 가입자에 의해서 발생되는 다른 메시지들과 통계적 다중화 효과를 일으켜서 결국은 소프트스위치로 볼 때에는 메시지의 길이는 다르지만 메시지의 목적과는 상관없는 무수히 많은 메시지가 Poisson 분포를 따라서 계속해서 소프트스위치(이하 서버라고 부름)로 도착하는 프로세스로 모델링을 할 수 있다.

현재 VoIP에서 호의 설정, 접속확인 및 해제에 대하여 어느 특정한 메시지에 우선순위를 주지 않고 대등한 서비스를 제공하는 경우가 대부분이다[3]. 따라서 서버가 이 메시지들을 FIFO(First In First Out)로 서비스 한다고 가정을 하면 위의 논의는 <그림 5>와 같이 추상화할 수 있다. <그림 5>에서 버퍼의 크기는 패킷(이하, 패킷과 메시지는 동의어로 사용된다)의 손실이 일어나지 않도록 충분히 크다고 가정을 하며, 성능평가의 목적은 메시지의 처리시간이 일정한 임계치를 넘지 않도록 적절한 서버의 용량(<그림 5>에서 μ에 해당하는 값)을 계산하는 것이다. 이 때 메시지의 처리시간은 메시지가 버퍼에서 기다리는 시간과 서비스를 받고 처리가 완료되는 시간을 합한 값으로 정의한다.

그림 5. 소프트스위치의 추상화 모델
Fig. 5. Simplified Softswitch model.

<그림 5>에서 각 유형의 메시지는 각 각 Poisson 분포를 따라서 서버로 유입되며, 둘 이상의 Poisson 도착을 중첩을 하여도 역시 Poisson 도착을 따르므로 전체 도착률을 λ로 두고, 다음과 같이 나타낸다.

$$\lambda = \lambda_I + \lambda_{I\text{-}ACK} + 2\times\lambda_{ACK} + \lambda_{RI} + \lambda_{RI\text{-}ACK} + \lambda_B + \lambda_{B\text{-}ACK}. \quad (1)$$

한편, 각 메시지의 서비스 시간은 길이가 크게 차이는 나지 않으나 같지는 않다 (<표 1>참조)[1]. 따라서 이 메시지들의 길이는 평균이 $1/\mu$인 지수분포를 따른다고 가정을 한다.

단일노드에서의 M/M/1 대기행렬시스템의 평균대기시

간에 대한 해석은 네트워크의 해석에 있어서 잘 알려져 있으며, 평균메시지 도착률과 서비스률을 각각 λ 및 μ라고 하고, 평균부하를 ρ라고 하면, ρ=λ/μ이며, 단일노드(본 연구에서는 소프트스위치에 해당함)에서 노드에 새로운 메시지가 도착하였을 때 먼저 도착한 메시지가 서버로부터 서비스를 받기 위하여 기다리고 있는 평균 메시지의 수를 N_{SW}라고 두면, N_{SW}는 다음과 같다[4].

$$N_{sw} = \frac{\rho}{1-\rho} \quad (2)$$

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}라고 둔다)은 다음과 같이 구할 수 있다.

$$D_{sw} = N_{sw} \times \frac{1}{\mu} + \frac{1}{\mu} \quad (3)$$

위의 식을 간소화하면, 다음의 결과를 얻을 수 있다.

$$D_{sw} = \frac{1}{\mu - \lambda} \quad (4)$$

이 결과를 이용하면, <그림 4>의 등가지연모델에서 D_{RTT}는 식 (5)와 같이 정의할 수 있다.

$$D_{RTT} = 4 \times D_{SW} + 2 \times D_{IP} + D_{TE2} \quad (5)$$

<그림 2>에서 정의된 호설정 - 재설정 - 호해제의 진행절차를 분석하면 한 가입자가 호설정에서 해제에 이르기까지 각 단계에서 발생되는 메시지의 비율은 4:3:2임을 알 수 있다. 한편, 가입자의 규모가 충분히 큰 경우, 소프트스위치에 부과되는 9가지의 메시지가 랜덤한 분포를 따라서 발생한다고 근사할 수가 있으므로 호당 전체 도착률 λ는 다음과 같이 나타낼 수 있다. 이는 수식 (1)에서 $\lambda_I = \lambda_{RI} = \lambda_{RI-ACK} = \lambda_B = \lambda_{B-ACK} = \lambda_{ACK}$이고 $\lambda_{I-ACK} = 2*\lambda_I$인 경우이다.

$$\lambda = 9\lambda i \quad (6)$$

단 식 (6)에서 첨자 i는 각종 메시지를 대표하는 것으로 하여 편의상 individual의 첫 자를 나타낸다.

한편, 호의 설정, 재설정 및 호의 해제에 필요한 최대 지연목표치(D_{RTT})가 정해지면 하나의 소프트스위치의 서버용량은 다음의 식으로 정할 수 있다.

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

실제 호설정지연목표치는 D_{RTT}의 두 배가 되지만 여기서는 단위 메시지의 지연관점만 고려하였다.

V. 실험 및 고찰

<그림 2>의 SIP를 이용한 호 설정 절차에서 IETF (Internet Engineering Task Force)가 정의하고 있는 메시지의 크기는 메시지 헤드와 메시지 바디에 들어가는 내용에 따라 달라질 수 있다. 그러나 음성 서비스의 경우에는 메시지 바디의 내용이 크게 차이가 나지 않으므로 실제 큰 차이가 없을 것으로 사료되어 참고문헌[11]의 샘플 메시지 크기를 인용하였다. 각 메시지의 크기를 계산하면 <표 1>과 같다.

표 1. 패킷유형별 메시지의 크기
Table 1. Packet type and size

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

<표 1>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 것으로 여겨진다. 위의 각 패킷의 길이의 산술적 평균은 424byte이며, 만약 서버의 링크의 속도가 C라고 한다면, 패킷의 평균 서비스 시간은 1/μ=3816/C에 해당한다.

세계 각국에서는 IP 네트워크의 SLA(Service Level Agreement)로서 백본 네트워크에서의 단대단 지연을 규정하고 있는데, 미국의 경우는 동서횡단거리가 먼 관계로 약 80ms를 규정하고 있고[6], 일본의 경우는 일본열도의 대륙 내에서 약 35ms를 규정하고 있다. 우리나라는 일본보다도 국토의 종단간 거리가 더욱 짧으므로 IP 백본망 내에서의 단대단 허용지연은 35ms를 넘지 않을 것으로 추정된다[8]. 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[9]. 지연품질을 Best class로 가정하여 단방향 지연 목표치가 E.164 Number translation to IP Address에 대해서 2초 이내로 규정되어 있으므로 D_{RTT}의 값은 4초가 된다. D_{TE2} 값은 PC의 성능에 따라 크게 의존하나, 1초 이내로 가정을 한다.

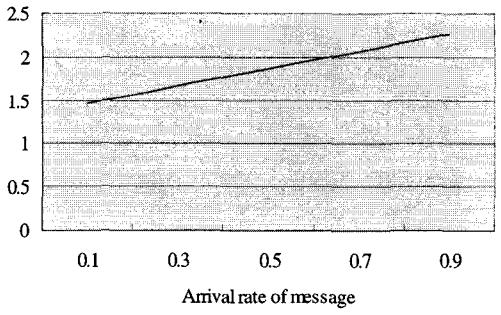


그림 6. 호 제어 메시지의 도착률과 소프트스위치의 서비스율간의 관계
Fig. 6. Relationship between control message arrival rate and Softswitch service rate.

위의 가정에 기반하여 <그림 6>에서 소프트스위치의 서버의 소요용량은 식 (7)에서 두개의 항으로 표시된 바와 같이 우선 입력메시지의 도착률에 비례하는 고정적인 값을 가지고 그와 함께 메시지가 통과하는 네트워크 요소의 지연치에 의해서 결정되는 λ=0.5일 때 D_{RTT}의 값을 4초 이내로 유지하기 위해서 필요한 소요대역은 μ=1.87이며 이 때 ρ=0.27이 된다. 다시 말하면, D_{RTT}의 값을 4초 이내로 유지하기 위해서 소프트스위치의 평균부하는 27%를 넘지 않아야 한다. 이 결과는 VoIP 네트워크의 데이터전달계에 있어서 라우터나 링크의 안정적인 운용을 위해서 필요한 최대부하수준으로 제시되는 25%룰의 결과와 아주 유사한 결과임을 주목할 필요가 있다[14]. 즉, NGN에서 데이터전달계 뿐만 아니라 제어계 요소인 소프트스위치에 대해서도 호 제어에 있어서 지연보장을 위한 안정적 운용수준으로서 권고될 만한 스위치의 부하는 약 27%정도로 유지할 필요가 있다는 것을 의미한다.

한편, <그림 6>의 결과는 본 연구에서 우선적으로 목표로 하는 NGN 음성서비스 네트워크의 소프트스위치에 대하여 일정한 호 제어 지연을 만족시키기 위하여 소프트스위치의 호 제어 메시지 도착률과 소프트스위치 서비스율 간의 관계를 정량적으로 나타내었다는 점에 주목할 필요가 있다. 따라서 향후 네트워크 사업자가 음성서비스를 시작할 시점에서 수용이 가능한 가입자의 규모를 예측할 수 있을 경우에 지연목표치와 가입자의 규모에 대한 통계치를 기반으로 본 연구에서 제안한 성능평가방법을 사용하면 소프트스위치의 적정 용량을 설계할 수 있으며, 반대로 소프트스위치의 용량과 지연목표치가 정해진다면 역으로 소프트스위치가 수용할 수 있는 음성서비스 가입자 규모를 예측 할 수 있을 것이다. 이 문제에 대해서는 향후 연구과제로 남겨두고자 한다.

VI. 결 론

본 연구에서는 NGN에서 음성서비스의 호 제어를 담당하는 제어계 평면에서 호 제어와 관련된 서비스 품질을 결정짓는데 있어서 가장 핵심적인 요소인 소프트스위치의 호 처리와 관련된 지연성능을 해석적인 방법을 이용하여 정량적으로 평가하는 방법을 제안하였다. 음성서비스의 호 제어는 호의 설정, 재설정 및 해제에 이르는 3단계의 제어절차로 이루어져 있으며 각 단계에 대한 메시지의 교환내역에 대한 분석을 통하여 그들을 정해진 지연목표치 이내에 처리하기 위한 서버의 용량을 설계하는 방법론을 제시하였다. 본 연구에서 제안한 방법을 이용하면 NGN에서 호 레벨 서비스 품질로 정의되는 호 설정지연의 보장이 가능하도록 소프트스위치의 용량을 설계할 수 있을 것으로 기대된다. 향후계획으로는 IP 네트워크 내부에서의 지연요소를 좀 더 자세히 분석하고 가입자의 수용규모를 고려하여 실제 설계에 적용이 가능한 결과를 도출하는 방안과 호 제어 단계 중 특정한 단계에 대한 메시지의 지연품질을 향상시키는 방안으로 호의 설정에 관한 메시지를 호의 해제와 관련된 메시지보다 더 우선적으로 처리하는 것 등을 대상으로 메시지의 우선처리에 대한 타당성 등을 검토할 예정이다.

참 고 문 헌

[1] RFC3261(SIP) : Session Initiation Protocol.

[2] RFC3263 Session Initiation Protocol(SIP) : Locating SIP Servers.

[3] N. Shiomi, T. Oka and E. Katayama, "A study of performance on Voice over IP system," Technical Report of IEICE NS2002-148, 2002. 10.

[4] P. Nain, Performance evaluation of Computer Systems, Lecture Notes of University of Massachusetts, 1995.

[5] Abdi R. Modarressi and Seshadri Mohan, "Control and Management in Next-Generation Networks : Challenges and Opportunities," IEEE Communications Magazine, 2000.

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

[7] Hoon Lee, "Dimensioning links for NGN VoIP networks," 한국통신학회 논문지, 제28권, 제8B호, 2003. 8.

[8] Hoon Lee, "Anatomy of VoIP Network and Quality of Service Provisioning," 대한전자공학회 추계학술대회, 2003. 11.

[9] ETSI TR 101 329 V2.1.1 "Telecommunications and Internet Protocol Harmonization Over Networks(TIPHON) ; General aspects of Quality of Service(QoS)," 1999. 6

[10] RFC 3264 An Offer/Answer Model with the Session Description Protocol(SDP).

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

[12] T. Eyers and H. Schulzrinne, "Predicting Internet telephony call setup delay", IPTel 2000 (First IP Telephony Workshop), Berlin, April 2000.

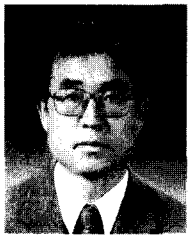
[13] 이훈, "NGN QoS 보장을 위한 요소기술 및 품질 기준 연구," KT 기술연구소 최종 연구보고서, 2003. 9.

[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>