

인터넷에서 멀티미디어 서비스를 위한 QoS 협상

고희창*, 이세훈**, 성미영***
(주)인프라밸리 기술연구소*
인하공업전문대학 전자계산기과**
인천대학교 전자계산학과***

Digital QoS Negotiation for Multimedia Service over Internet

Hee-Chang Koh*, Sei-Hoon Lee**, Mee-Young Sung***
R&D Center, Infracvalley, Inc.*
Dept. of Computer Engineering, Inha Technical College**
Dept. of Computer Science and Engineering, University of Incheon***

요약

인터넷의 발전에 따라 서비스에 대한 사용자 요구가 다양해지는 동시에 서비스 품질에 대한 요구가 높아지고 있다. 멀티미디어 서비스는 대량의 데이터로 구성되어 네트워크에 트래픽 부하를 증가시키며 이로 인한 지연, 동기, 패킷 손실과 같은 문제가 발생되나 서비스의 품질을 협상할 수 있다는 특성도 가지고 있다. 이 연구에서는 인터넷에서 제한된 대역폭의 사용 효율성을 높이면서 멀티미디어 서비스의 품질(QoS)을 보장하기 위한 기법으로 동기화된 멀티미디어 서비스의 QoS와 통계에 기반을 둔 지수적 무게 이동 평균 기법을 제안하였고, 협상을 위한 모듈을 설계하였다.

1. 서론

인터넷은 공개된 네트워크라는 특성에 따라 뛰어난 확장성을 가지고 있으며 또한 정보 공유가 용이하여 매우 빠르게 이용이 확산되고 발전되고 있다. 특히 하이퍼링크와 WWW의 개념을 도입한 멀티미디어 서비스가 제공된 이후로 사용의 편의성으로 인해 인터넷이 대중에게 급속도로 보급되었고, 새로운 대규모 시장이 형성됨에 다양한 응용이 개발되어 시도되고 있다. 이러한 시도 중에는 인터넷을 이용한 음성전화, 화상전화 및 화상회의 등 실시간 양방향성의 특성을 갖는 서비스들이 있다. 이러한 실시간 서비스들을 효과적으로 지원하기 위해 일정 규격의 서비스 품질(QoS: Quality of Service) 특성을 지닌 네트워크 제어 메커니즘이 개발되고, 최근에는 자원 사용의 효율성을 고려한 서비스 차별화를 위한 메커니즘과 신호 규격에 대한 연구가 활발히 진행되고 있다[1].

또한 콘텐츠의 제공에 있어서도 다양한 미디어를 혼용하여 정보의 전달력을 극대화하고자 하는 노력이

표준화로 이어져 SMIL(Synchronized Multimedia Integration Language)이 탄생하였다[2]. 이 연구에서는 최선형(best effort) 서비스의 특성을 기반으로 하는 인터넷에서 다양한 멀티미디어 응용의 서비스 품질을 제공하기 위해 품질에 영향을 미치는 요소들을 분석하고, 네트워크 자원을 효율적으로 사용하면서 일정 수준 이상의 서비스 품질을 제공하기 위한 통계적 지수 무게 이동 평균 기법을 제안하고, QoS 협상을 위한 모듈을 설계한다.

2. 멀티미디어 서비스의 품질

2.1 멀티미디어 서비스의 품질 특성

이 절에서는 멀티미디어 서비스 품질에 영향을 미치는 요소들에 대해 분석한다.

인터넷을 이용한 멀티미디어 서비스는 크게 두 부류로 분류할 수 있다[1,3]. 먼저, 원거리의 인간과 인간이 다양한 미디어를 이용하여 대화환경을 실현하는 것이다. 여기에는 실시간 처리가 요구되는 특성이 있

다. 두 번째는 특정 컴퓨터에 저장된 정보를 역시 다양한 미디어를 이용하여 제공하는 것이다. 여기에는 실시간 처리 요구는 상대적으로 적어지지만, 미디어 간의 동기를 맞추어야 하는 요구가 따른다. 인터넷 멀티미디어 서비스는 컴퓨터와 인간 혹은 인간과 인간 사이에 메시지를 효율적으로 전달하기 위하여 문자, 그래픽과 이미지, 오디오, 정지영상 및 동영상 등의 미디어를 통합 및 동기화 시켜 하나의 객체로 표현할 수 있어야 한다.

멀티미디어는 대량의 데이터로 구성되어 네트워크에 트래픽 부하를 증가시킨다. 반면, 멀티미디어 서비스는 서비스 품질을 절충할 수 있다는 특성도 가지고 있다. 예를 들면, 시스템의 부하에 따라 음질을 낮추거나, 화상의 크기, 해상도, 초당 프레임의 수를 어느 정도 조절할 수 있다는 것이다. 즉 시스템의 부하가 높으면 서비스 품질을 낮추고, 시스템의 부하가 적으면 반대로 서비스 품질을 의도적으로 높일 수 있다.

2.2 멀티미디어 서비스 품질에 관한 연구

화상 전송과 같은 서비스는 넓은 대역폭을 요구하고, 인터넷 방송과 같이 높은 대역폭과 실시간 전송을 요구하는 서비스에 대해, 많은 사용자가 요구하는 멀티미디어 서비스를 지원하기 위하여 망을 확장하는 동시에 기존 망을 효율적으로 최대 활용하는 방안에 대한 연구가 활발히 진행되고 있다[3,4,5].

기존의 인터넷은 최선형(best effort) 서비스만 지원하며, 품질에 대한 고려는 미약하며, IPv4에서는 TOS(Type of Service)란 8비트 필드가 서비스 클래스를 구분할 수 있지만 거의 사용되지 않고 있다. 서비스의 품질을 고려하지 않으면, 패킷을 빨리 처리할 수는 있겠지만, 전송 지연의 문제가 남는다[1,4].

제한된 대역폭 자원으로 보다 우수한 멀티미디어 서비스의 품질을 제공하기 위한 연구로는 혼잡 상황(congestion)을 미리 감지하여 패킷의 손실을 막는 RED(Random Early Detection)와 Generic Traffic Shaping 기법에 관한 연구, 큐잉 기법에 관한 연구로 Priority Queuing, Custom Queuing, Weighted Fair Queuing 등이 있다. 실시간 트래픽을 효과적으로 전송하기 위한 RTP(Real Time Protocol)과 RTCP(Real Time Control Protocol)를 보완하는 성격의 연구로서 RTP-HC(RTP Header Compression)와 LFI(Link Fragmentation and Interleaving)에 대한 연구가 있다. RSVP(ReSerVation Protocol)는 요구된 QoS에 해당하는 대역폭을 수신 측에서 계산하여 경로 전 구간에 대해 미리 확보하는 자원 예약 프로토콜이다.

DiffServ는 IP 헤더의 TOS 비트들에 IP 우선권(precedence), 폐기 우선권(drop precedence), 지연 표시(delay indication)를 나타내도록 하여 차별화된 서비스가 가능하도록 하고 있다. 이와 유사한 것으로 클래스 기반 큐잉(CBQ: Class Based Queuing)에 관한 연구도 활발히 진행되고 있다[1,7].

3. 동기화된 멀티미디어 서비스의 QoS

3.1 멀티미디어 통신 계층

이 절에서는 상호 협력적 분산 멀티미디어 응용을 위한 멀티미디어 통신 계층의 요구 사항을 정의하고, 실제 스트림이 전송되는 방식을 정의한다. 그리고 이를 제공하기 위한 기본 연산들을 정의한다. 정의되는 인터페이스들은 이후에 IDL 형태로 정의되어 RPC나 Java RMI, CORBA, DCOM 등 구현 방식에 독립적으로 사용될 수 있다[8,9,10].

멀티미디어 통신에서의 연속 미디어, 즉 오디오/비디오 스트림은 화상 회의 등의 멀티미디어 응용에서 반드시 고려해야 할 기본적인 미디어 형태이며, 분산 멀티미디어 응용에서는 이러한 스트림의 일대일, 일대다, 다대다 통신 환경을 지원하기 위한 멀티캐스트 방식이 요구된다[8,9,10].

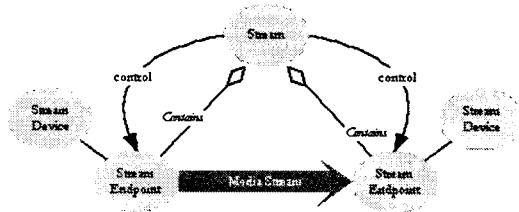
기본적으로 오디오/비디오 스트림은 일대일 전송을 기반으로 하고 있으며, 이를 일대다, 또는 다대다로 전송하기 위해서는 스트림을 복제하여, 일대일로 다수의 스트림을 전송하는 방식이 사용되는데, 이는 시스템의 효율에 막대한 영향을 줄 수 있다. 따라서 이 연구에서는 이러한 부작용을 방지하기 위해, 실제 미디어 스트림을 추상화하여, 추상화된 미디어 스트림에 일대다 통신을 위한 연산을 정의하여 사용한다.

3.2 스트림

이 연구에서는 오디오/비디오 스트림중 단일 미디어를 스트림(stream)으로 정의한다. 이러한 스트림은 근원(source)과 목적지(sink)를 가지는데 이 연구에서는 구분하지 않고 종단점(endpoint)으로 정의한다. 그리고 두 개의 종단점을 잇는 스트림에 방향을 지정할 수 있는 멤버를 정의하여 스트림의 전송 방향을 정의하는 방법을 사용한다. 이러한 방식은 실제 스트림의 근원과 목적지가 거의 동일한 연산을 수행하기 때문이다. 각 스트림의 종단점은 스트림 장치(stream device)와 연결되어, 해당 장치에서 발생하는 스트림을 전송, 제어한다.

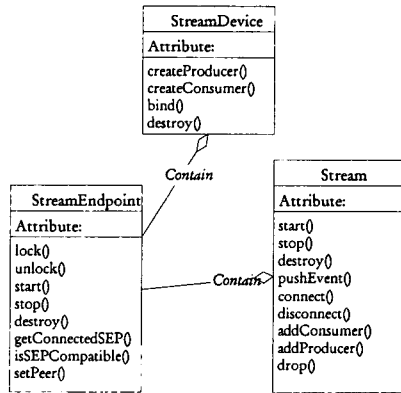
기본적으로 스트림은 일대일 통신을 고려하며, 다수의 스트림이 동시에 전송되는 경우는 다음절에서 언

급하게 될 스트림 컬렉션(stream collection)에서 통합하여 관리할 수 있도록 정의할 것이다. 단일 스트림을 전송하기 위해서는 해당되는 스트림의 종단점과 실제 장치를 추상화하는 장치, 그리고 이러한 것들을 제어하기 위한 스트림이 요구된다. 하나의 스트림은 스트림 종단을 포함하게 되며, 그것들을 제어하기 위한 기본적인 연산인 생성과 삭제, 재생에 관련된 연산들이 정의되어야 한다. [그림 1]은 이 연구에서 정의하고 있는 단일 스트림 통신의 형태를 나타내고 있다.



[그림 1] 단일 스트림 전송

스트림 관련 인터페이스에는 단일 스트림 전송을 위한 스트림 인터페이스와 스트림 종단점, 그리고 장치 추상화를 위한 스트림 장치가 포함된다. [그림 2]는 스트림 관련 클래스 다이어그램을 나타내고 있다.

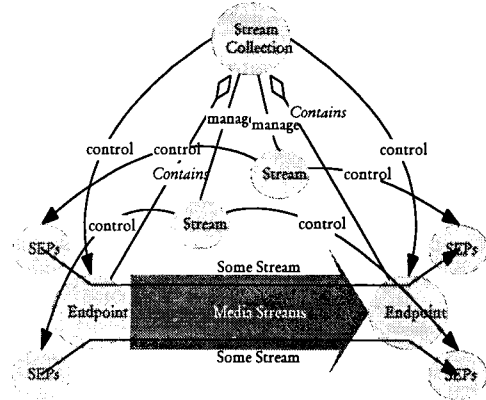


[그림 2] 스트림 관련 클래스 다이어그램

스트림 장치에는 종단점 생성을 위한 연산과 스트림 장치 접속을 위한 연산, 그리고 스트림 장치 삭제에 대한 연산이 정의된다.

3.3 스트림 컬렉션

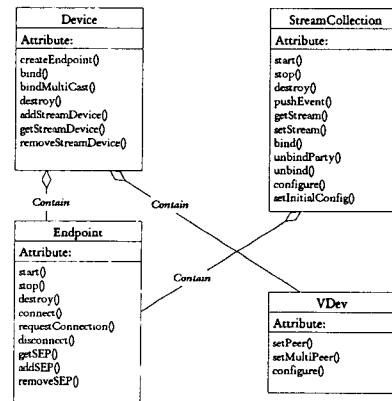
스트림 컬렉션은 하나 이상의 스트림을 포함할 수 있으며, 스트림 뿐만 아니라 미디어 파일도 포함할 수 있다. 스트림 컬렉션도 스트림과 마찬가지로 종단점을 가지며, 가상 장치와 연결되어 전체 스트림 컬렉션의 전송과 제어를 위해 스트림 종단점을 제어한다. [그림 3]은 스트림 컬렉션의 전송을 나타내고 있다.



[그림 3] 스트림 컬렉션의 전송

[그림 3]에서 스트림 컬렉션은 종단점을 포함하며, 각 스트림의 종단점은 스트림 컬렉션의 종단점을 통해 스트림을 전송한다. 물론 단일 미디어 스트림의 전송 시와 마찬가지로 각각의 스트림 종단점은 스트림에서 제어하게 되며, 이러한 스트림을 스트림 컬렉션에서 관리한다. 스트림 컬렉션은 다수의 종단점을 포함할 수 있는데, 스트림 컬렉션 내에서 복제 연산을 제공하게 되면 하나의 종단점에서 다수의 종단점으로 스트림들을 전송하는 브로드캐스트(broadcast) 방식을 구현할 수 있다.

[그림 4]는 스트림 컬렉션 관련 클래스 다이어그램을 나타내고 있다.



[그림 4] 스트림 컬렉션 관련 클래스 다이어그램

스트림 컬렉션에는 스트림에서와 마찬가지로 스트림의 재생과 중단, 삭제에 대한 연산들과 스트림을 관리하기 위한 연산, 장치의 접속과 해제, 그리고 장치들의 설정을 위한 연산들이 정의된다.

장치에는 종단점 생성을 위한 연산과 장치 접속을 위한 연산, 그리고 멀티캐스트 방식 지원을 위한 연산

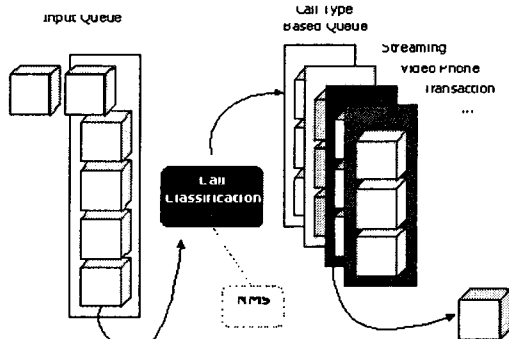
과 장치 삭제에 대한 연산이 정의된다. 또한 스트림 장치 관리에 대한 연산들이 정의된다. 가상 장치에는 가상 장치 설정에 관한 연산과 멀티캐스트 방식 지원을 위한 연산이 정의된다. 장치 인터페이스와 가상 장치를 분리하는 목적은 가상 장치 인터페이스에서는 실제 하드웨어 장치 제어에 대한 연산들을 추상화하며, 장치에서는 특정한 장치들을 추상화하여 사용하기 위해서이며, 장치는 전송되는 스트림의 종류에 따라 새로운 인스턴스를 생성하여 사용하게 된다.

3.4 클래스 기반 큐잉

기존의 FIFO(First In First Out) 큐잉은 네트워크 부하 시에 꼬리가 잘리는 현상(Drop Tail)이 발생할 수 있는 문제점이 있고, 우선순위(Priority) 큐잉은 높은 우선순위를 갖는 호 요청이 많을 경우, 상대적으로 낮은 우선순위를 갖는 호 요청은 계속 서비스를 받지 못하는(starvation) 현상을 가지게 되는 문제점 등이 있다. 클래스 기반 큐잉(Class Based Queuing)은 우선순위 큐잉 방식의 변형으로서, 하나의 출력 큐 대신에 여러 개의 출력 큐를 클래스별로 두어서 우선순위를 정하고 각 큐별로 서비스되는 트래픽의 양을 조절할 수 있는 방식이다. 이렇게 함으로써 어느 특정 클래스의 트래픽이 전 시스템 자원을 모두 독점하는 것을 방지한다.

즉, 클래스 기반 큐잉은 각 클래스별로 정해진 양의 대역폭을 보장할 수 있는 방식으로 알려져 있지만, 실제로는 완전 보장이라기보다는 우선순위 큐잉 방식보다는 약간 더 완화된, 클래스별로 자원이 완전 고갈되는 것을 막으면서도 각 클래스에 적절한 서비스를 제공할 수 있는 방식이다.

이 연구에서는 클래스 분류의 기준으로 트래픽의 종류와 서비스 우선순위 정보를 사용한다. [그림 5]는 이 연구에서 제안한 큐잉 알고리즘을 나타내고 있다.

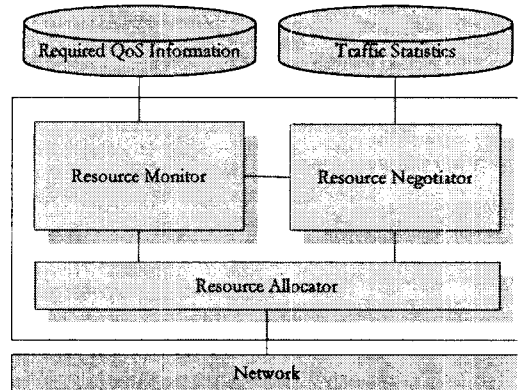


[그림 5] 호 종류 기반 큐잉

4. 멀티미디어 서비스를 위한 QoS 협상

4.1 QoS 협상 모듈의 설계

QoS 협상을 위하여 두 개의 데이터베이스가 사용된다. 클래스 기반 Queuing을 위한 요구 QoS 정보 데이터베이스, 그리고 예측기법 적용을 위한 장기 트래픽 통계 데이터베이스이다. 자원감시자(Resource Monitor)가 자원 할당 상태를 감시하고 새로운 요구가 있을 때 QoS 정보에 따라 클래스를 정한다. 자원협상자(Resource Negotiator)는 장기 통계 정보를 이용하여 할당 가능한 자원의 양을 파악하고 그에 따라 자원 감시자로부터의 클래스 정보에 따라 실제 할당할 자원의 양을 정해 자원 할당자(Resource Allocator)에게 전달한다. [그림 6]은 QoS 협상 모듈이다.



[그림 6] QoS 협상 모듈

4.2 대역폭 협상

이 절에서는 멀티미디어 서비스 요청에 대해 대역폭을 효율적으로 할당하기 위한 협상 메커니즘에 대해 기술한다. 이를 위하여 가용 자원(대역폭)의 양을 동적으로 산출하여 가급적 많은 대역폭을 할당할 수 있도록 한다. 가용 자원의 동적 산출을 위하여 과거($t-1$ 시점)의 변화량을 함께 고려하여 변화량의 추이를 유추하여 통계적 예측 값으로 활용할 수 있다. 변화량의 추이를 유추하는 방법으로는 호 수락 제어(call admission control), 트래픽 분석 및 예측(traffic analysis & prediction)에 사용되고 있는 통계 기법으로 지수적 무게 이동 평균(EWMA: Exponential Weighted Moving Average) 기법을 적용할 수 있다. EWMA는 $t-1$ 시점에서의 변화에 대한 가중치를 0.2 ~ 0.3으로, t 시점에서의 변화에 대한 가중치로 0.7 ~ 0.8을 부여하여 최근의 변화 추이를 모델링하고 이를 $t+1$ 시점에 대한 예측으로 활용하는 기법이다. 이 기법을 적용하여 변화량 추이를 다음과 같이 정의할 수

있다.

$$T_A = W_{ewma} \times \frac{\int_{t-2}^{t-1} Alloc - \int_{t-3}^{t-2} Alloc}{\int_{t-2}^{t-1} Alloc} + (1 - W_{ewma}) \times \frac{\int_{t-1}^{t-1} Alloc - \int_{t-2}^{t-2} Alloc}{\int_{t-1}^{t-1} Alloc} \quad (1)$$

where
 TA : Tendency of Allocation
 Alloc : Sum of allocated BandWidth at that time
 WEWMA : Weight of Exponential Weighted Moving Average

가용 대역폭은 다음과 같이 정의할 수 있다.

$$ABW = T - A \quad (2)$$

where
 ABW : Available BandWidth
 T : Total BandWidth (constant)
 A : Sum of current allocated BandWidth

요구된 자원을 할당할 수 있는 기준은 각각의 스트림에 대해 앞에서 기술한 가용 자원의 양과 할당 변화량 추이(TA)를 함께 고려하여 다음과 같은 경우로 나눌 수 있다.

- 가용 자원의 양이 요구 자원에 대해 충분하고, 전체 변화량 추이(TA)가 0보다 작은 경우
- 가용 자원의 양이 요구 자원에 대해 충분하고, 전체 변화량 추이(TA)가 0보다 큰 경우
- 가용 자원의 양이 요구 자원에 대해 부족하고, 전체 변화량 추이(TA)가 0보다 작은 경우
- 가용 자원의 양이 요구 자원에 대해 부족하고, 전체 변화량 추이(TA)가 0보다 큰 경우

이 네 가지 경우에 대해 각 서비스 타입별 특성과 가입자 서비스 프로파일을 고려하여 자원적용이 이루어지게 된다. 첫 번째의 경우는 요구 대역폭보다 많이 할당할 수 있는 경우이고, 두 번째는 요구 대역폭만큼을 할당하는 경우이다. 세 번째와 네 번째는 요구 대역폭만큼을 할당할 수 없는 경우이므로 대역폭을 변경하여 할당하는 오퍼레이션이 적용되어야 한다. 그리고 이러한 사항은 상향과 하향에 대한 요구 대역폭이 다른 비대칭 서비스가 있기 때문에 상향 대역폭과 하향 대역폭에 대해 함께 고려되어야 한다.

5. 결론

이 연구는 인터넷에서 많은 대역폭을 요구하는 멀티미디어 서비스에 대해 사용자가 요구하는 서비스 품질을 보장하면서 자원의 활용도를 높여 전체적인 서비스 품질을 향상시키도록 하였다. 이를 위하여 멀티미디어 서비스를 미디어 스트림의 종류에 따라 분류하였고, 클래스 기반 큐잉 기법을 적용하였다. 또한 트래픽 유형에 대해 장기적인 과거의 통계 정보를 이

용한 예측 기법을 적용함으로써 혼잡상황의 임계치가 가까울 때 발생할 수 있는 패킷 폐기를 최소화하였다.

참고문헌

- [1] 정일영, 강신각, "VoIP QoS를 위한 네트워크 기술", 정보통신 제18권3호, 2001.
- [2] W3C, Synchronized Multimedia Integration Language (SMIL 2.0) Specification, <http://www.w3.org/TR/smil20/>, 2001.3.
- [3] Cisco, "White Paper on IP QoS", Jan. 2000
- [4] J. Strassner, W. Weiss, D. Durham, A. Westerinen, "Information Model for defining the QoS Capabilities of Network Devices and Services", Internet Draft <draft-ietf-policy-qos-capabilities-schema-00.txt>
- [5] Y. Snir, Y. Ramberg, J. Strassner, R. Cohen, "QoS Policy Information model", Internet Draft <qos-policy-info-model-00.txt>
- [6] Errin W. Fulp and Douglas S. Reeves, "On-line Dynamic Bandwidth Allocation," Proceedings of the 1997 International Conference on Network Protocols (ICNP '97), 1997.
- [7] 고희창, 왕창중, "All-IP 네트워크에서 무선 자원 할당과 연계된 QoS 관리자의 설계", 정보처리논문지 제7권 제8호, 2000.8.
- [8] OMG, CORBA Telecoms Specification, <http://www.omg.org>, 2000
- [9] Iona Technologies, OrbixMX MediaStreams Programmer's Guide and Reference, 1998
- [10] 홍영래, 이승용, "CORBA A/V 스트리밍 서비스", 정보과학회지 제17권 제7호, 1999.