웹서버의 계층적 분류 및 QoS 보장 기술

이 상 문*, 박 종 규*, 김 학 배*
(*연세대 공대 전기전자공학부)

1. 서 론

오늘날 21세기 정보화 사회를 선도하고 있는 인터넷 산업분야는 초고속통신망의 구축으로 네트웍 환경이 크게 향상되고, 다양한 통신 서비스가 제공되고 있는 등 인터넷 이용의 보편화를 위한 많은 노력이 집중되고 있으며, 이에 따라 그 사용자도 급격히 증가하고 있다. 또한, 최근 인터넷을 이용해 크게 활성화되고 있는 E-business, E-commerce, Contents, IDC(Internet Data Center), 그리고 ISP(Internet Service Provider)등은 IT (Information Technology)분야의 핵심을 이루고 있다. 이와 함께, 통신기기, 가전·멀티미디어 기기, 자동화·제조기기등 여러 응용분야에서 인터넷을 통한 웹 응용 프로그램을 활용하기 위해 많은 노력을 하고 있다.

이와 같이, 인터넷을 기반으로 이루어지고 있는 정보통신 관련분야의 산업들이 날로 규모가 커지고 있으며,이를 이용하는 사용자도 꾸준히 늘어나고 있는 추세이다. 이로 인해 네트웍을 통해 이동하는 정보의 양이 급격히 증가하고 있으며, 결과적으로 정보를 저장하고 공급해주는 서버에 그 트래픽(traffic)이 집중되고 있어서 안정적인 동작을 보장해 줄 수 있는 서버의 중요성이 크게 부각되고 있다. 인터넷은 실질적으로 통신시스템에 의해 변화하고 있으며, 새로이 생겨나고 있는 무수한 서비스들을 처리하기 위한 매개체로 성장하고 있다. 사용자의 욕구를 충족시켜주기 위해 네트웍의 고속화와 더불어 보다 더 다양한 서비스가 제공되고 있으며, 이러한 인터넷 사용의 증가로 인해 트래픽도 급격히 증가하고 있다. 이와 함께, 사용자가 요구하는 인터넷 서비스에 대한 품질의 기대 수준도 점차 높아지고 있다. 그러나, 현재의 인터넷은 간단한 형태인 최선형(Best Effort)서비스만을 제공하고 있다. 즉, 네트웍 상에서 전송율, 에러율의 측정과 개선이 불가능하며, 트래픽이 증가할수록 그 성능을 유지하기가 힘들다. 이에 따라 QoS(Quality Of Service)개념의 적용으로 이러한 문제에 적절히 대응하려는 연구가 크게 주목을 받고 있다.

점차적으로 인터넷 서비스의 기반구조는 예측이 불가능하고 가변적인 네트웍 환경에서 서비스의 품질, 신뢰성, 그리고 보안성 보장에 대한 많은 요구사항들을 가지게 되는데, 그 변화는 웹서버를 중심으로 일어나고 있다. 웹은 표준화되고 유연한 호환성을 가진 응용서비스의 인터페이스를 제공해 주어, 많은 사용자들이 이러한 서비스를 이용할 수 있게 해주는 역할을 하고 있으며, 최근 새롭게 생겨나고 있는 performance-critical 또는 business critical한 응용서비스들이 요구하는 엄격한 QoS에 적합한 인터페이스를 제공해 줄 수 있다. 이러한 측면에서 다양한 사용자들 간의 성능 차별화(performance differe- ntiation), 독립된 서비스 간의 성능 분리화(performance isolation), 용량 계획(capacity planning)등에 대해 보장된 QoS를 제공해 주어야 한다[1].

영상, 음성 등과 같은 멀티미디어 통신에서 QoS를 보장해주기 위해서는 궁극적으로 단대단(end-to-end)에 대한 문제를 고려해주어야 한다. 예를 들어, 사용자가 순차적인 멀티미디어 데이터를 실행한다고 했을 때, 서버로부터 네트웍을 통해 멀티미디어 정보를 완벽히 전송받을 수 있는 확실한 QoS가 필요하다. 이를 위해서는 일반적으로 먼저 단대단(end-to-end) 수락테스팅(admission testing)과 자원예약(resource reservation), 말단 시스템에서의 디스크 정합(co-ordination)과 쓰레드(threads) 스케쥴링, 네트웍에서의 패킷/셀 스케쥴링과 흐름제어(flow control) 그리고 마지막으로 정해진 QoS의 실제적인 모니터링 및 유지가 요구된다[2].

이러한 의미에서 웹기반 인터넷 환경의 기반을 이루는 핵심 요소라 할 수 있는 웹서버에 대해 계층적으로 분류를 하고 그 응용분야들에 대해서 알아본다. 이와 함께, 새로운 인터넷 환경에서 향상된 품질을 보장해 줄 수 있는 QoS기술들에 대해 살펴본다.

2. 웹서버의 분류

서버(server)는 일반적으로 메인프레임의 기능을 하고 있는 시스템으로써, 특정 사용자에 의해 응용프로그램을 실행하기 위한 것이 아니라, 클라이언트 시스템들이 요구하는 여러 정보를 저장하고 있으며, 필요시에 이러한 정보를 원

하는 요구자에게 제공하는 역할을 한다. 웹서버, FTP서버 등이 여기에 해당한다. 이러한 여러 종류의 서버들을 성능, 가격, 사용목적 등 다양한 기준에 따라 분류될 수 있는데, 여기에서는 웹을 위한 PC서버에 대해 성능과 그 활용범위에 따라 low-end, mid-range, high- end 서버로 구분하였다. PC는 엔터테인먼트 및 인터넷, 기본적인 사무작업 등을 그 활용범위가 매우 다양하고 범용의 시스템이어서 PC의 성능에 따라서 서버의 역할을 수행하기도 하며, 최근 H/W 및 S/W기술의 향상으로 워스테이션의 수준의 성능을 가지기도 한다[3].

2.1 Low-end서버

Low-end서버는 일반적으로 단일 프로세서 플랫폼의 형태를 가지며 dual 프로세서로 업그레이드가 가능한 H/W구조를 가지고 있다. 주로 단일 사업자의 웹서비스에 이용될 수 있는데, 10대 이내의 낮은 성능을 가진 리얼서버(real server)들 즉, 저가 PC를 이용하여 최적화된 클러스터 환경의 웹 서버를 구축하여 운영할 수 있다. 또한 high-end급 PC의 소형 서버용 OS나 서버용 어플리케이션을 설치하여 사용할 수도 있다.

2.2 mid-range서버

Mid-rande서버는 low-end서버보다 성능이나 활용범위가 큰 중소규모의 서버를 말하며, 다양한 응용분야를 가지고 있다. 최근 가격이 저렴하면서도 성능이 우수한 H/W가 개발되고 있어, 이를 이용한 클러스터형 웹서버 개발이 활발히 이루어지고 있다. 여러 대의 PC를 하나의 클러스터로 구성하여 하나의 서버처럼 운영하는 것인데, 기존의 단일 서버방식에 비해 우수한 안정성과 확장성 또한 high-end급의 고성능의 제품이 개발되고 있다. 이렇게 볼 때, mid-range서버는 두 가지로 구분하여 생각해 볼 수 있는데, volume mid-range서버는 low-end급의 가격구조를 가지지만 H/W성능이나 확장성은 mid-range급을 형성하고 있다. 또 하나는 performance mid-range서버로서 가격은 mid-range급을 가지면서 성능은 high- end급을 보유할 수 있는 것으로 high-end서버의 volume화를 기대할 수 있게 할 것이다.

Mid-range급의 웹서버들은 100대 내외의 실서버들을 이용해 클러스터형 웹서버를 구축하여, 최근 크게 활성화되고 있는 IDC, VoD서비스, ISP등의 인터넷 사업과, 학교, 공공건물, 아파트 등의 중소규모 서버로 활용되고 있다.

2.3 high-end 서버

최근 PC에 사용되는 H/W부품들의 성능이 매우 향상되었기 때문에 high-end PC의 경우 워스테이션으로 사용되기도 한다. 또한 워스테이션에 PC와 똑같은 플랫폼이 사용되는 경우도 흔히 볼 수 있다. PC와 워스테이션의 경계가 사라져가고 있는 것이다. High-end PC서버는 기존의 서버 시장을 위한 제품과 함께 보다 기업환경에 적합한 더 높은 확장성을 추구하고 있다. 현재 PC서버에서 최다 프로세서인 4-way에서 8-way SMP플랫폼으로 개발되고 있으며 또한, 기존의 4-way 플랫폼을 클러스터링으로 구현하고 있다. 또한 mid-range급과 high-end급의 플랫폼은 랙(rack)에 장착하는 시스템 팩키지를 갖는 것 일반적인 구조이다. 이러한 high-end급 서버들은 대규모 사업장, ERP (Enterprise Resource Planning), OLTP(Online Transaction Processing), E-business, 데이타웨어하우스(data warehouse)등에 활용된다.

3. QoS 보장 기술

현재 계층적인 프레임워크에서 QoS의 상황은 다음과 같이 요약될 수 있다[2].

- **불완전성(incompleteness)** : 현재의 어플리케이션 프로그래밍 인터페이스들은 일반적으로 QoS환경을 가지고 있지 못하며, 멀티미디어 서비스를 위한 제어나 관리를 위해 필요한 설비의 일부분에만 QoS를 제공하고 있다.
- **QoS보장 지원기술 부족** : 정해진 레벨의 서비스가 예측 가능하고 확실성을 가지게 하도록 분산제어, 모니터링 그리고 QoS 유지에 대한 연구가 요구된다.
- **전반적인 프레임워크의 부족** : 서로 다른 시스템 레벨과 네트웍 구조에서 현재 제공하고 있는 QoS와 조화를 이루고 수행될 수 있는 전반적인 계층적 프레임워크의 개발이 시급하다.

이러한 한계를 극복하기 위하여 네트웍과 말단 시스템에서 QoS를 규정화하고 QoS제어 및 관리 기법을 통합화하기 위한 프레임 워크를 제공하여야 한다. 이를 위해서, 현재 인터넷 표준화 기구인 IETF(Internet Enginee- ring Task Force)는 인터넷에서 사용자의 요구사항에 따른 QoS를 제공하기 위해 여러 가지 서비스 모델과 메커니즘을 제시하고 있는데, 그 중에서 (i)Integrated services/ Resource Reservation Protocol(RSVP), (ii)the differentiated services (DS), (iii) multi-protocol label swi-tching(MPLS), (iv)traffic engineering, and (v)constraint based routing기술에 대해 알아본다[4-6].

3.1 Integrated services/ Resource Reservation Protocol (RSVP)

Integrated service모델은 Best Effort서비스와 더불어 두 개의 서비스 클래스를 제시하고 있다.

- **Guaranteed service** : 고정된 지연 경계값을 요구하는 어플리케이션을 위한 서비스
- **Controlled-load service** : 신뢰적이고 향상된 Best Effort서비스를 요구하는 어플리케이션을 위한 서비스

이러한 모델의 원리는 라우터가 특정한 사용자 패킷의 스트림(stream)이나 흐름(flow)을 위한 특별한 QoS를 제공해 주기 위하여 자원 예약을 할 수 있도록 피할 수 없는 요구를 하는 것이다. 이것은 라우터들에서 흐름특성 상태를

차례로 요구하게 된다. RSVP는 어플리케이션이 자원들을 예약하기 위하여 신호 프로토콜(signaling protocol)로 개발되었다. 하지만, 이와 같은 흐름별 상태와 흐름별 프로세싱에 관한 RSVP의 장점이 글로벌한 망에서는 확장성 문제를 야기시킴으로써 백본망에 적용하기에는 한계를 가진다. 최근 RSVP는 흐름의 집합화를 위해 자원 예약을 위한 몇가지 방식으로 수정과 확장을 하였다. 이는 QoS를 요구하는 ERs(Explicit Routes)를 설치하고, 다른 신호 작업들을 수행하기 위함이었다.

Integrated 서비스는 신호프로토콜(즉, RSVP), 수락제어 루틴(admission control routine), 분류자(classifier) 그리고 패킷 스케쥴러 등의 네 가지 요소에 의해 수행된다. 제어가 가능하거나 확실한 부하 서비스는 데이터를 전송하기 전에 미리 경로와 자원 예약을 하여야 한다. 라우터가 패킷을 받으면 분류자는 멀티필드(multifield)분류를 수행하고, 분류된 결과를 바탕으로 패킷을 특정한 큐(queue)에 넣는다. 그리고 나서 패킷 스케쥴러는 이것이 요구하는 QoS에 따라 패킷을 스케쥴하게 된다. Integrated service는 보다 폭넓은 서비스를 지원하기 위하여, 현재 비실시간 서비스와 그 필요성이 점차로 커지고 있는 실시간 서비스를 함께 지원하는 통합서비스 제공과 이를 지원하기 위하여 인터넷 구조와 프로토콜을 개선하기 위한 표준화 작업이 이루어지고 있다.

3.2 Differentiated Services

위의 integrated services와 RSVP은 그 수행이 매우 힘들기 때문에, differentiated services가 제안되었다. 이것은 integrated service와 매우 다른 방식이다. 우선, differentiated services필드에 의해 지정되는 서비스 클래스가 제한된 개수만을 가진다. 클래스의 작은 부분에 서비스들이 할당되기 때문에 상태정보의 양은 흐름의 수보다 클래스의 수에 비례하게 된다. 그러므로, different- iated services는 보다 더 확장성을 가진다. 두 번째로, 복잡한 분류, marking, policing, 그리고 shaping 작업은 단지 네트웍의 경계에서만 수행하며, 코어노드는 BA(Behavior Aggregate) 분류만 수행하여 확장성 문제를 해결한다. 일종의 우선권 개념이라 할 수 있는 differentiated services는 Best Effort 서비스 외에 프리미엄(premium) 서비스, Assured 서비스, 올림픽 서비스를 제공한다.

- **프리미엄 서비스** : 낮은 지연과 지터(jitter)서비스를 요구하는 어플리케이션을 위한 것이다.
- **Assured 서비스** : Best Effort서비스 보다 더 안정성을 요구하는 어플리케이션을 위한 것이다.
- **올림픽 서비스**는 품질에 따라 금, 은, 동의 서비스 등급을 주는 서비스를 말한다.

이 중에서 Assured 서비스는 망이 혼잡할 때 사용자에게 일정수준의 서비스 품질을 보장하는 서비스로서, 실제 트래픽을 SLA(Service Level Agreement) 준수여부에 따라 트래픽을 IN 또는 OUT 프로파일로 분류하여, Assured Queue에 넣고, RED(Random Early Detection) 또는 RED with In and Out(RIO)으로 스케쥴링한다.

3.3 Multiprotocol label switching(MPLS)

MPLS는 패킷 핸들링을 결정하기 위하여 고정된 길이의 라벨을 사용하려는 동기에서 개발되었다. OSI 7레이어 모델에서 2레이어(링크 레이어)와 3레이어(네트웍 레이어)사이에 위치하게 된다. MPLS는 라벨에 따라 패킷을 포워딩(forwarding)한다. MPLS는 라벨을 사용하여 도메인 내의 한 종단간에 다수의 경로를 설정할 수 있으며, 설정된 경로는 서로 다르고 가변적인 대역폭을 가져, 트래픽의 지연 및 손실 민감도에 따라 차별화된 서비스를 제공할 수 있다. MPLS에서 트래픽 전송은 다음과 같이 수행된다. 경계의 라우터는 헤더 정보를 분석하여 사용할 경로를 결정하며, 일치하는 경로가 있으면 이를 패킷에 붙여 전송한다. 코어 라우터는 라벨에 의해 지정된 경로를 따라 패킷을 단순히 전송한다. 이는 differentiated services와 유사한 구조를 갖지만, 그 차이는 입력 노드가 MPLS 헤더를 추가하고, 코어 노드에서 DSCP필드가 아니라 라벨을 참조하며, 출력 노드에서 MPLS 헤더를 제거한다는 점이다.

MPLS에서 QoS를 제공하기 위해서는 설정된 경로에 대역폭을 할당할 필요가 있으며, 이를 위해 ER-LSP (Explicitly Routed Label Switch Path)가 제안되고 있는데, 이것은 대역폭 정보와 정확한 경로를 표시하여 경로를 요청하는 것이다. ER-LSP를 구현하는 방법은 CR- LDP(Constraint-based Routed Label Distributed Protocol)를 사용하거나 RSVP를 확장하여 이용하는 방법이 IETF에서 논의 중에 있다. 기술적인 관점에서는 비슷한 신호 기능을 가지며, CR-LDP는 구현이 용이하고, 상호 호환성이 보장되며, 여러 업체가 참여하고 있는 장점을 가지고, RSVP를 사용하는 방법은 RSVP 메시지에 explicit-route 객체만 추가하는 약간의 수정으로 이미 구현·설치된 RSVP를 사용할 수 있는 장점이 있다. 특히, 빠른 패킷 분류, 포워딩, 그리고 효율적인 터널링(tunneling)구조를 제공하는 중요한 특성을 가지고 있다.

3.4 Ttraffic engineering and Constraint based routing

Integrated services/RSVP와 differentiated services와 같은 QoS기법은 근본적으로 트래픽 부하가 매우 클 경우 서로 다른 트래픽에 대해서 차별화되게 성능을 감쇠시킨다. 부하가 적을 경우에는 이들간에 별 차이가 없다. 이러한 트래픽을 발생을 미리 피하게 하는 것이 traffic engineering이다.

3.4.1 Traffic Engineering

네트웍 충돌은 네트웍 자원이나 트래픽의 불균일한 분배에 의해 발생한다. 첫 번째 경우는, 모든 라우터와 링크들은 과부하 상태가 되는 것인데, 그 해결책으로는 기본 네트웍 구조를 개선하여 보다 더 많은 네트웍 자원을 공급해주는 것이다. 두 번째 경우는, 네트웍의 몇몇 부분에 과부하가 발생하는 반면 부하가 적은 부분도 발생하는 경우이다. 이러한 불균일한 부하의 분포는 현재 사용하고 있는 RIP,

OSPF 그리고 IS-IS들과 같이 항상 최단 경로를 따라 패킷을 보내는 동적 라우팅 프로토콜을 사용하기 때문이다. 이 결과로 두 노드 사이의 최단 경로에 있는 라우터와 링크들은 부하가 많이 걸리며 네트웍상에 충돌이 일어나는 반면, 최적경로가 아닌 보다 긴 경로에 있는 라우터와 링크들은 idle한 상태가 된다. OSPF의 ECMP(Equal-Cost Multipath) 옵션과 최근의 개선된 IS-IS는 최단 경로가 다수일 때는 부하를 분배해 주는데 유용하지만, 최단 경로가 하나일 때는 효과가 없다. 따라서 여러 최단 경로가 존재하는 네트웍 구성이 필요하다.

이와 같이, traffic engineering은 불균일한 네트웍 이용에 의해 발생되는 충돌을 방지하기 위하여 트래픽의 흐름을 조정하는 것이다. 따라서 네트웍에서 QoS를 제공하기 위해 중요한 역할을 수행한다. 이에 비해 Constraint-based routing은 traffic engineering 과정을 자동으로 하게 하는 중요한 툴이다. 충돌을 피하는 것과 충돌시 성능을 서서히 감소시키는 것은 상호보완적이라 할 수 있을 것이다.

3.4.2 Constraint-based routing

Constraint-based routing은 policy와 같은 네트웍의 다른 제약을 고려해줌으로써 QoS라우팅을 확장시킨 것이다. 이것은 다음과 같은 목적을 가지고 있다.

- 어떠한 QoS요구도 만족시켜줄 수 있는 라우터 선택.
- 네트웍의 이용률 향상

라우터를 결정하는 동안, Constraint-based routing은 네트웍 topology뿐만 아니라 흐름의 요구, 링크의 자원가용도, 그리고 네트웍 관리자에 의한 다른 policy적용 여부 등을 고려해준다. 그러므로, 과부하가 걸린 최단 경로에 비해서 비록 길지만 적은 부하를 가진 최적의 경로를 찾을 수 있다. 이렇게 함으로써 네트웍의 트래픽을 고르게 분포시킬 수 있다.

4. 결 론

오늘날 급격히 증가하는 인터넷 산업으로 인해 네트웍을 통해 이동하는 데이터의 양이 급격히 증가하고 있으며, 결과적으로 이러한 정보들을 관리하는 서버에 그 트래픽이 집중되고 있다. 이에 따라, 인터넷의 기반을 이루는 요소 중의 하나인 서버의 중요성에 대한 인식을 새롭게 하고, 다양한 사용자의 요구와 인터넷 서비스의 품질향상을 위한 QoS에 대하여 알아보았다. 서버는 성능에 따라 low-end, mid-range, high-end서버로 분류하였다. QoS보장을 위한 기술로는 대표적으로 Integrated services/ Resource Reservation Protocol (RSVP), the differentiated services(DS), multiprotocol label switch- ing(MPLS), traffic engineering, and constraint based routing등에 대하여 살펴보았다.

참고문헌

[1] A. Tarek, N. Bhatti, " Web server QoS management by adaptive content delivery," *International Workshop on Quality Of Service*, London, UK, Jun. 1999.

[2] Aurrecoechea, A. Campbell, A.T. and L. Hauw, "A Survey of QoS Architectures", *ACM/Springer Verlag Multimedia Systems Journal*, Special Issue on QoS Architecture, Vol. 6 No. 3, pp. 138-151, May 1998

[3] http://www.gosamsung.co.kr/prodsub/semet/sn_first_ b.htm

[4] X. Xiao, L Ni, "Internet QoS: A big picture," *IEEE Network*, pp.8-18, Mar./Apr. 1999.

[5] A. Campbel, G. Coulson and D Hutchison, " A quality of service architecture," *ACM SIGCOMM Computer Communication Review*, vol. 24, no. 2, pp.6-27, Apr. 1994

[6] T. Abdelzaher, K. G. Shin, "End-host architecture for QoS-adaptive Communication," *IEEE Real-Time Technology and Applications Symposium*, Denver, Jun. 1998

저 자 소 개

이상문(李相文)

1971년 10월 12일생. 1997년 영남대 전기공학과 졸업. 1999년 연세대 대학원 전기공학과 석사. 1999년-현재 동 대학원 전기전자공학부 박사과정. 관심분야: 실시간 시스템, 고장포용 시스템.

박종규(朴種圭)

1972년 4월 12일생. 1998년 원광대학교 제어계측공학과 졸업. 2000년 원광대학 제어계측공학과 대학원 석사. 2000년 -현재 연세대학교 전기전자공학부 박사과정. 관심분야는 실시간 OS, QoS알고리즘

김학배(金鶴培)

1965년 10월 10일생. 1988년 서울대 전자공학과 공학사. 1990년 미국 미시간대 전기 및 컴퓨터공학과(EECS) 공학석사. 1994년 동대학 공학박사. 1994년 9월-1996년 8월 미국 National Research Council(NRC) Research Associate at NASA Langley Research Center. 1996년 9월-현재 연세대학교 전기컴퓨터 공학과 조교수, 2000년 9월-현재 전기전자공학부 부교수. 관심분야: 실시간 및 내장형 시스템, 인터넷 웹서버 기술, 디지털 시스템 고장포용 및 신뢰도 평가 분야.