

대용량 소프트웨어 스트리밍 서버를 위한 부하 분산 알고리즘 제안 및 구현

허성진^{1*}

A Proposal and Implementation of Load Balancing Algorithm for Large Scale Software Streaming Server Platform

Sung-Jin Hur^{1*}

요 약 본 논문은 소프트웨어 스트리밍 서비스의 특성이 일반 동영상 스트리밍 서비스와 달리 전송 대역폭이 가변적이며 초기에 집중적인 전송이 발생하는 사실에 근거하여 가장 효율적으로 서버 자원을 활용할 수 있는 부하 분산 알고리즘을 제안한다. 제안된 알고리즘은 지역적으로 산재한 대규모의 인터넷 사용자에게 소프트웨어 스트리밍 서비스를 지원하는 대용량 소프트웨어 스트리밍 서버 플랫폼에 실제 구현하여 동작함으로써 제안된 알고리즘이 적절하게 동작함을 보이고자 한다.

Abstract In this paper, I propose a load balancing algorithm based on the traffic characteristics of a software streaming service, and implement the proposed algorithm on the large scale software streaming server platform which provides a software streaming service to geographically dispersed internet users. By showing that the large scale software streaming server platform functions correctly, we know the proposed load balancing algorithm is operating properly and efficiently.

Key Words: 소프트웨어 스트리밍, QoS 특성, 부하 분산, 주문형 소프트웨어 서비스, 서버 자원 관리.

1. 서론

스트리밍 기술은 다운로드 방식과 대조적으로 사용자가 필요로 하는 데이터를 서버로부터 모두 가져와서 동작하는 것이 아니라 필요한 시점에 필요한 부분만을 가져와서 원하는 동작을 실행하는 기술로서 구체적인 적용 분야는 크게 동영상 스트리밍과 소프트웨어 스트리밍으로 구분할 수 있다. 동영상 스트리밍 서비스는 그 동작 특성상 서버로부터 전송되는 일부 화상 데이터가 손실되어도 사용자에게 지속적인 서비스가 가능한 반면에 소프트웨어 스트리밍 서비스는 실행 코드가 전송되는 특성상 전송 도중 한 비트의 오류가 발생하더라도 서비스가 지속될 수 없는 특성을 가진다. 이러한 차이점으로 인하여, 동영상 스트리밍 서비스와 응용 소프트웨어 스트리밍 서비스는 각각 UDP 및 TCP와 같은 서로 다른 성격의 하부

전송 프로토콜을 사용하여 화상 데이터 또는 실행 코드를 전송한다. 또한 동영상 스트리밍 서비스는 일반적으로 세션의 처음부터 종료될 때까지 요구하는 전송 대역폭이 일정한 반면에 소프트웨어 스트리밍 서비스는 요구하는 전송 대역폭이 시간에 따라 급격하게 변하는 특성을 가진다. 이로 인하여 동영상 스트리밍 서비스를 제공하는 서버와 소프트웨어 스트리밍 서비스를 제공하는 서버는 각각 서로 다른 정책에 기반하여 사용자에게 서버 자원을 할당한다. 그림 1은 소프트웨어 스트리밍 서비스의 개략적인 개념도를 나타낸다.

본 논문의 목적은 소프트웨어 스트리밍 서비스의 전송 대역폭 변화 특성을 분석하여 소프트웨어 스트리밍 서비스에 가장 적합한 부하 분산 알고리즘을 제안하고 이를 대용량 소프트웨어 스트리밍 서버 플랫폼에 구현함으로써 한정된 서버 자원을 보다 효율적이고 안정적으로 사용할 수 있는 기반을 마련하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 1장의 서론에 이어, 2장에서는 소프트웨어 스트리밍 서비스의 전송 대역

¹한국전자통신연구원 선임연구원

*교신저자: 허성진(sjheo@etri.re.kr)

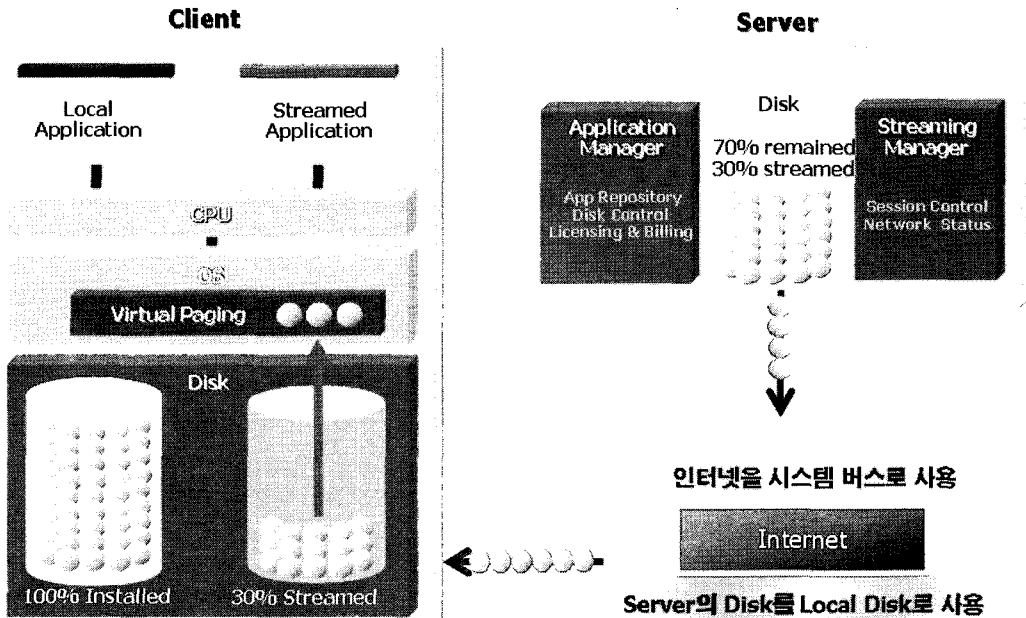


그림 1. 소프트웨어 스트리밍 서비스 개념도

폭 특성을 분석 및 설명하고 3장에서는 분석된 특성을 바탕으로 소프트웨어 스트리밍 서비스에 적합한 부하 분산 알고리즘을 제안하고 이를 대용량 소프트웨어 스트리밍 서버에 구현한 내용을 설명하며, 마지막으로 4장에서는 결론 및 향후 과제를 다룬다.

2. 소프트웨어 스트리밍 서비스의 전송 대역폭 특성

소프트웨어 스트리밍 서비스의 전송 대역폭 특성은 응용 소프트웨어가 실행되는 과정에서 요구되는 실행 코드의 양(페이지 수)과 밀접한 관련이 있다. 일반적으로 응용 소프트웨어는 사용자가 소프트웨어를 선정하여 초기 화면이 나타나는 시간까지의 실행 초기에는 많은 양의 실행 코드를 필요로 하는 반면에 그 이후에는 사용자가 요청한 기능을 지원하는 실행 코드만을 필요로 하는 특성을 가진다. 이는, 소프트웨어 스트리밍 서비스의 전송 대역폭 특성에 그대로 반영되는데, 사용자가 소프트웨어 실행을 시작하는 초기에는 대량의 전송 대역폭이 요구되는 반면에 그 이후에는 사용자의 요청이 있을 때만 간헐적으로 전송이 발생하므로 스파크 형태의 전송 대역폭 요구 특성을 관찰할 수 있다. 그림 2는 소프트웨어 스트리밍 서비스의 전형적인 전송 대역폭 특성을 보여주는데,

CAD를 소프트웨어 스트리밍 방식으로 실행했을 때 서버와 클라이언트 사이의 통신망에 나타나는 사용 대역폭 변화 추이를 실제적으로 보여주고 있다.

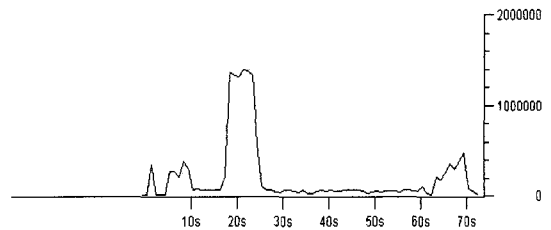


그림 2. 소프트웨어 스트리밍 서비스의 대역폭 특성

그림 2에서 나타난 바와 같이, CAD를 수행했을 때 초기 30초 이내에 대량이 전송 대역폭 사용이 일어나고 그 이후에는 간헐적으로 스파크 형태의 전송 대역폭 사용이 발생하고 있음을 볼 수 있다. 이러한 전송 대역폭 특성은 동영상 스트리밍 서비스의 전송 대역폭 특성과 매우 상이한 것으로서, 이는 곧 사용자의 요청에 대하여 서버 자원을 적절하게 할당하는 부하 분산기의 동작에 결정적인 영향을 미치게 된다. 이러한 특성에 기반하여 4장에서는 소프트웨어 스트리밍 서비스에 특화된 부하 분산 알고리즘을 제안한다.

3. 소프트웨어 스트리밍 서비스용 부하 분산 알고리즘

소프트웨어 스트리밍 서비스에 특화된 부하 분산 알고리즘을 제안하기 위하여, 본 논문에서는 먼저 소프트웨어 스트리밍 서비스의 세션을 크게 Download Phase와 Ondemand Phase의 두 단계로 나누어 관리하며 그림 3은 그 개략적인 구조를 보여주고 있다. 그림에서 나타낸 바와 같이, Download Phase는 사용자가 소프트웨어를 선정하고 해당 소프트웨어의 초기 화면이 나타나는 기간 동안의 집중적인 데이터 전송이 발생하는 단계인 반면에, Ondemand Phase는 초기 화면이 나타난 이후 사용자의 요청에 따라 데이터 전송이 간헐적으로 발생하는 단계로 정의한다.

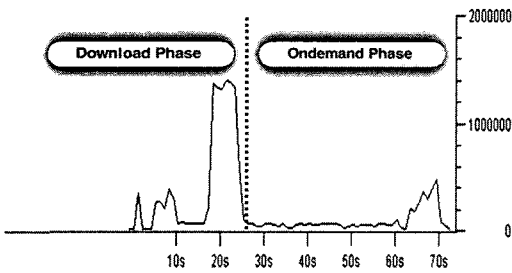


그림 3. Download Phase 와 Ondemand Phase

다음으로 본 논문에서는 Download Phase와 Ondemand Phase로 구분된 단계에 따라 서버 자원을 서로 관리하는데 그림 4는 제안된 부하 분산 알고리즘의 개

략적인 흐름도를 보여준다.

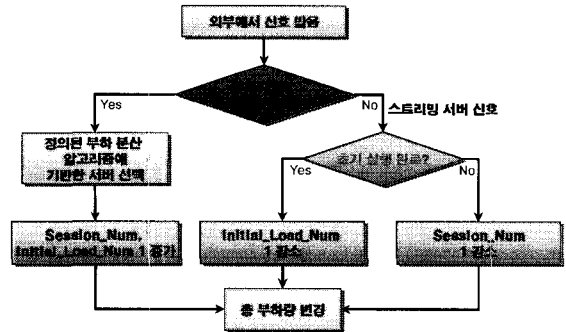


그림 4. 부하 분산 알고리즘 흐름도

- 외부에서 신호를 받은 서버는 해당 신호의 근원지를 파악한다.
- 클라이언트로부터 발생한 신호는 서비스 요청을 의미하고 이는 해당 서버에 연결된 세션 수(Session_Num) 및 Download Phase에 있는 세션의 수(Initial_Load_Num)를 하나 증가하게 된다.
- 스트리밍 서버로부터의 신호는 다시 초기 실행 완료 여부에 따라 구체적인 동작이 구분되는데, 초기 실행이 완료되는 경우 Initial_Load_Num을 1 감소하는 반면에, 초기 실행 완료가 아닌 경우에는 세션의 종료를 의미하게 되므로 Session_Num을 1 감소한다.
- 최종적으로 부하 분산기는 위의 알고리즘에 따라 각 스트리밍 서버가 유지하는 Session_Num 및 Initial_Load_Num 값에 의거하여, 사용자의 서비스 요청이 있을때 가장 적절한 서버를 선정한다.

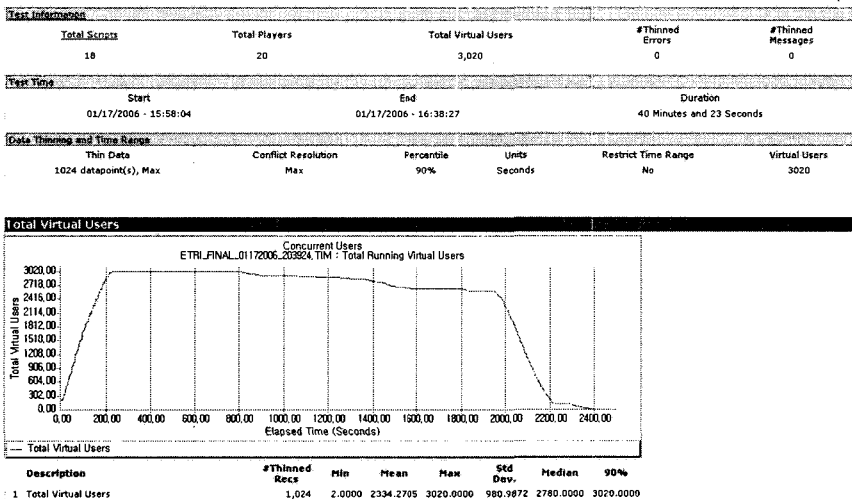


그림 5. 제안된 부하분산 알고리즘의 동작 결과

제안한 부하 분산 알고리즘은 대용량 소프트웨어 스트리밍 서버에 구현하였으며, 그림 5는 구현한 알고리즘에 따라 동작하는 대용량 소프트웨어 스트리밍 서버의 동작 결과를 보여준다.

4. 결론 및 향후 과제

본 논문에서는 소프트웨어 스트리밍 서비스의 전송 대역폭 특성에 기반한 부하 분산 알고리즘을 제안하고 제안한 알고리즘을 대용량 소프트웨어 스트리밍 서버에 구현하여 그 동작을 검증하였다. 소프트웨어 스트리밍 서비스는 일반 동영상 스트리밍 서비스와 달리 전송 대역폭이 가변적이며 세션 초기에 집중적인 데이터 전송이 발생한다는 사실에 기반하여, 세션을 두 단계로 나누어 관리함으로써 서버 자원을 보다 효율적인 활용한다. 향후 과제로는 제안된 부하 분산 알고리즘과 기존의 부하 분산 알고리즘에 대한 성능 비교 분석이 필요하며 이를 바탕으로 제안한 알고리즘의 효율성을 더욱 증대하는 과정이 필요하다.

참고문헌

[1] MPEG Official Homepage,
<http://www.chiariglione.org/mpeg/index.htm>
 [2] P. Kuacharoen, V. J. Mooney and V. K. Madisetti,
 "Efficient Execution of Large Applications on Portable and Wireless Clients," Proceedings of the Mobility Conference & Exhibition, Aug. 2004.

[3] T. Pavlidis, "Fundamentals of X programming: graphical user interfaces and beyond," New York, NY, Kluwer Academic, 1999.
 [4] D. Barron, "The World of Scripting Languages," Chichester, NY, Wiley, 2000.
 [5] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, The Macmillan Press Ltd., 1976
 [6] J. Kim, Jun Kim, S. Heo, K. Nam, W. Choi, "Measurement and Analysis of the On-Demand S/W Streaming Server Performance," Proceedings of the ICACT 2005, pp. 106-108, Feb. 2005
 [7] QALoad,
<http://www.compuware.com/products/qacenter/qaload.htm>
 [8] D. F. Stanat and D. F. McAllister, *Discrete mathematics in Computer Science*, Prentice Hall, 1977.

허성진(Sung-Jin Hur)

[정회원]



- 1990년 2월 : 경북대학교 전자공학과 (공학사)
- 1992년 2월 : 경북대학교 컴퓨터공학과 (공학석사)
- 1999년 2월 : 경북대학교 컴퓨터공학과(공학박사)
- 1999년 3월 ~ 2001년 2월 : 창신대학 인터넷정보과 전임교수
- 2001년 2월 ~ 2006년 10월 현재 : 한국전자통신연구원 온디맨드서비스연구팀 선임연구원

<관심분야>

멀티미디어통신, 소프트웨어스트리밍, 서버기술