

가중치 윈도우 기반의 미디어 캐쉬 교체 정책

오재학^o 차호정^h

^o 광운대학교 컴퓨터과학과

^h 연세대학교 컴퓨터과학과

ojh@cs.kwangwoon.ac.kr hjcha@cs.yonsei.ac.kr

A Media Cache Replacement Policy based on Weighted Window

Jaehak Oh^o Hojung Cha^h

^oDept. of Computer Science, Kwangwoon University

^hDept. of Computer Science, Yonsei University

요 약

본 논문에서는 스트리밍 미디어 캐싱 서버의 효율적인 캐싱 구조를 위하여 참조 횟수, 참조량, 참조 시간 등의 정량적인 인자들과 사용자 요구 주기를 적용하여 최근 참조 경향에 높은 가중치를 부여함으로써 변화하는 콘텐츠 선호 경향에 빠르게 적응하는 가중치 기반의 캐쉬 교체 정책을 제안한다. 성능 분석은 시뮬레이션 환경 구축을 통해 실험하였으며 LRU, LFU와 SEG 캐쉬 정책과 비교분석하여 향상된 결과를 보였다.

1 서론

최근 동영상 스트리밍 서비스는 네트워크 트래픽 증가에 따른 사용자 QoS 보장을 위해 선호하는 콘텐츠를 대상으로 콘텐츠 일부만 저장하고 관리하는 동영상 미디어 캐싱의 필요성이 대두되고 있다[1]. 동영상 미디어 캐싱 시스템은 사용자와 가까운 지역 네트워크에 위치하여 사용자가 선호하는 콘텐츠를 빠르게 전송하고 트래픽을 지역 네트워크로 한정하는 효과를 갖는다. 이러한 캐싱 시스템의 성능은 제한된 저장 공간에서 콘텐츠 혹은 콘텐츠 일부에 대한 가감을 관리하는 효율적인 캐쉬 교체 정책에 따라 많은 영향을 받는다[2].

동영상 미디어 캐싱은 기존 캐싱 환경의 캐쉬 오브젝트와 구별되는 대용량, 재생 연속성과 실시간의 특징이 있으며, 대용량 콘텐츠의 완전(full) 캐싱으로 인한 저장 공간 낭비를 막기 위해 사용자 참조 경향에 따라 캐쉬 오브젝트 가감을 조절할 수 있는 교체 정책을 필요로 한다. 즉, 캐쉬 교체 정책은 참조 빈도와 평균 참조량 등의 정량적인 인자를 통해 오브젝트 크기를 결정하고, 그 크기는 사용자 서비스를 위한 교체 정책에 역으로 반영되어 캐쉬 오브젝트의 중요도로 응용될 수 있다. 또한, 캐쉬 교체 정책은 참조 횟수, 참조량, 참조 시간과 오브젝트 크기 등의 정량적인 인자와 사용자 콘텐츠 요구 주기성을 기반으로 과거 참조 경향으로부터 미래 참조에 대한 예측을 고려하여야 한다. 동영상 콘텐츠의 사용자 참조는 시간 제한적 서비스이며 수 분에서 수 시간 동안 연속 참조되고, 동영상 콘텐츠에 대한 참조 경향과 시간대 요구의 편중성이 높다. 이는 정량적인 인자들을 현재 시점에서 과거 참조 발생 시점까지 차등적인 가중치를 적용하여 미래 참조에 대한 캐쉬 교체 정책의 정확도를 높일 수 있다.

본 논문에서는 참조 횟수, 참조량, 참조 시간과 오브젝트 크

기 등의 정량적인 캐쉬 인자들과 사용자 요구 주기를 적용하여 기존 캐쉬 교체 정책과 차등화된 개념의 가중치 윈도우에 기반한 캐쉬 교체 정책을 제시한다. 가중치 윈도우 기반의 캐쉬 교체 정책은 설정된 사용자 요구 주기에 따라 정량적인 인자들의 가중치를 차등 적용하고 최근 참조 경향에 높은 가중치를 부여하여 변화하는 콘텐츠 선호 경향에 빠르게 적응하는 구조이다. 논문의 구성은 다음과 같다. 2장에서 캐쉬 오브젝트와 가중치 윈도우와 교체 정책의 기본 개념을 살펴보고, 3장에서 시뮬레이션을 통해 교체 정책의 성능 평가와 경향을 분석한다. 4장에서 결론을 맺는다.

2 가중치 윈도우 기반의 캐쉬 교체 정책

다음은 스트리밍 미디어에 대한 캐쉬 교체 정책의 기본 개념과 가중치 윈도우 기반의 교체 정책을 제시한다.

2.1 캐쉬 오브젝트와 가중치 윈도우

캐쉬 오브젝트는 스트리밍 서버에서 전송받아 캐쉬 내에 저장되고 대용량, 연속성의 특징을 갖는 동영상 미디어이며 사용자 참조방식과 캐쉬 교체 정책에 따라 원본 미디어의 일부분이 캐쉬에 저장된다. 캐쉬 오브젝트는 사용자 참조에 따라 서비스 중인 액티브(active) 오브젝트와 사용자 참조가 없는 인엑티브(inactive) 오브젝트로 구분된다. 먼저, 사용자 콘텐츠 참조의 전제 조건은 콘텐츠 시작부터 끝까지 순차적으로 발생하고 임의 위치로 옮겨갈 수 없으며 임의 위치에서 중단할 수 있음을 가정한다. 액티브 오브젝트는 다수 사용자가 동시에 스트리밍 서비스를 받을 수 있으며, 선행 참조된 스트림에 의해서 콘텐츠 반입과 캐싱이 수행된다. 액티브 오브젝트는 순차 참조 경향에 따라 오브젝트의 임의 블럭을 삭제할 수 없기 때문에 캐쉬 블럭 요구의 교체 대상으로 적용되지 않는다. 따라서, 캐쉬는 액티브 오브젝트의 원본 미디어를 캐싱할 공간이 필요하며,

• 본 연구는 정보통신부에서 지원하는 정보통신기초기술연구지원사업으로 수행 하였습 (과제번호 : 2001-076-3)

이들의 총합이 캐쉬 저장 공간을 초과할 수 없다. 인덱스 오브젝트는 현재 사용자 참조가 없는 콘텐츠로 캐쉬 저장 공간 부족으로 발생하는 블럭 교체 요구의 대상이 되며, 이들은 캐쉬 교체 정책에 따라 가중치를 계산하여 우선 순위를 정해 교체 대상으로 선정된다.

캐쉬 오브젝트는 M_{size} , M_{mean} , M_{prefix} , $M_{replace}$, M_{suffix} 인자들로 구성된다. M_{size} 는 원본 미디어 크기이고, M_{mean} 은 참조 횟수로 전송량을 나눈 평균 전송량이다. M_{prefix} 는 캐쉬 내에 저장하고 있는 캐쉬 오브젝트의 크기이며, $M_{replace}$ 는 M_{prefix} 와 M_{mean} 의 차이량이다. M_{suffix} 는 전체 미디어를 재생하는 사용자들을 위해서 스트리밍 서버로부터 전송받을 양을 의미한다. 캐쉬 오브젝트에 대한 이러한 분류는 제한된 저장 공간을 효율적으로 사용하기 위해서 교체 가능한 블럭량과 캐쉬 오브젝트의 위급도를 나타내기 위한 목적이다. 우선적으로 교체 가능한 블럭인 $M_{replace}$ 는 클수록 위급도가 낮고, 작을 수록 위급도가 높아진다. 이와 같은 맥락에서 캐쉬 오브젝트의 M_{mean} 은 캐쉬 교체 정책을 반영하는 관리 단위로서 의미를 가지며 웹 오브젝트와 같은 독립적인 캐싱 단위로 의미를 부여할 수 있다.

가중치 윈도우는 현재 시점에서 일정 거리의 과거 시간간격을 시간 윈도우로 설정하여 그 시간 동안에 발생한 사용자 요구들 중에 현재에 가까운 요구일수록 높은 가중치를 부여하여 최근의 사용자 참조 경향을 잘 반영할 수 있는 개념이다. 가중치 윈도우는 패턴 주기 내에서 요구 편중 경향을 잘 반영하기 위해 최근 참조량이 많은 콘텐츠를 빠르게 높은 우선 순위로 도달시키며, 참조 빈도가 낮아지거나 없는 경우에 시간이 지남에 따라 이전 참조 경향에 대한 가중치를 낮추고 점차적으로 우선 순위를 하향시켜 재참조에 대한 위험성을 고려하였다. 또한 가중치 윈도우는 윈도우 범위를 일정 시간 간격으로 나누어 가중치 적용 단위를 설정하고 그 시간이 지남에 따라 윈도우를 이동시키는 슬라이딩 개념을 적용하였다.

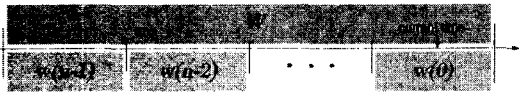


그림 2 가중치 윈도우

그림 2는 가중치 윈도우의 구조를 보여주고 있다. W 는 사용자 생활 패턴으로 전체 가중치 윈도우를 의미하는 주(major) 윈도우이다. $w(i)$ 는 W 를 일정 시간 간격으로 설정된 가중치의 적용 단위인 부(minor) 윈도우이다. 각 $w(i)$ 에 대한 가중치는 $f(i) = 1-i/n$, $i \in \{0, 1, 2, \dots, n-1\}$ 에 따라 적용되며, n 은 주 윈도우 W 내의 부 윈도우 수이다. W 에 대한 $w(i)$ 의 시간 간격 설정에 있어 다음을 고려해야 한다. $w(i)$ 가 짧을 수록 사용 경향을 정확하게 기록하고 경향을 잘 반영할 수 있지만 미디어 재생 시간 이하로 짧아지더라도 여러 부 윈도우에 경유하게 되 효과적이라 할 수 없으며, 오히려 시스템 처리 부하를 가중시킬 수 있다. 반면에 $w(i)$ 가 길수록 콘텐츠 참조 경향은 정확성이 낮아지게 된다. 그러므로, $w(i)$ 크기는 대상 미디어의 재생 시간에 준해서 설정해야 한다.

2.2 캐쉬 교체 정책

가중치 윈도우 기반의 캐쉬 교체 정책(이하 'WIN'이라고 함)은 제한된 시간 범위를 갖는 가중치 윈도우와 동적 크기의 캐쉬 오브젝트 특성을 반영한 블럭 단위의 관리와 캐쉬 오브젝

트의 후미 블럭 제거 방식을 제공하는 제한된 캐쉬 저장 공간에 대한 효율적인 활용 정책을 제시한다. W 의 가중치는 각각 $v(i)=a(i) \cdot (M_{mean}(i)/M_{prefix}(i))$, where $i = 0,1,2, \dots, n-1$ 이고. $w(i)$ 의 가중치는 $V=\tau \cdot \sum(1-i/n) \cdot v(i)$ 으로 계산된다. 여기서 τ 는 M_{mean}/M_{prefix} 이다. 부 윈도우의 가중치는 발생한 참조수 $a(i)$ 와 전송 평균값($M_{mean}(i)$) 분에 캐싱량($M_{prefix}(i)$)을 곱해서 계산된다. 주 윈도우의 가중치는 $v(i)$ 총합에 $w(i)$ 가중치를 적용하고 캐쉬 교체 정책 경향을 나타내는 τ 를 곱하여 계산된 것으로 W 의 중요도를 나타내고 높은 수치일 수록 콘텐츠 간에 높은 우선 순위를 나타낸다. τ 는 현재 캐쉬 오브젝트 크기에 대한 평균 미디어 전송량 비이며, 평균 미디어 전송량은 참조 적중수와 적중에 대한 전송량 합에서 계산된 평균 값이다. 따라서 τ 는 요구된 콘텐츠에 대한 위급도를 나타낸다. τ 값의 범위에 따라 캐쉬 교체 정책에 미치는 상관 관계는 다음과 같다. 첫째, τ 를 제한하지 않는 경우는 M_{mean}/M_{prefix} 에 따라 $0 \leq \tau \leq 1$ 의 범위를 갖는다. 캐쉬 오브젝트는 빈번한 참조와 참조량의 증가로 M_{mean} 이 커지거나 블럭 교체 대상이 되어 M_{prefix} 가 M_{mean} 보다 상대적으로 작아질수록 높은 우선 순위를 갖는다. 이는 M_{prefix} 가 M_{mean} 이하인 캐쉬 오브젝트에 사용자 참조가 발생했을 때, 캐싱량 부족으로 인한 사용자 QoS 저하와 스트리밍 서버로부터의 반입량 증가를 막기 위해서이다. 캐쉬 오브젝트가 낮은 참조 빈도를 갖더라도 캐쉬 저장 공간에서 제거되기보다는 M_{mean} 이하로 남아 있을 확률이 높아진다. 둘째, τ 를 제한하는 경우는 최대 값을 1 이하로 제한하여 $0 \leq \tau \leq 1$ 범위를 갖는다. 이는 첫째 경우에 비해 캐쉬 오브젝트 수는 감소하지만, M_{prefix} 가 M_{mean} 보다 작을 경우에 τ 를 1로 제한함으로써 참조 빈도가 낮은 오브젝트가 블럭 교체를 통해 작아진 M_{prefix} 로 인해 급격히 우선 순위 상승하는 것을 제한할 수 있다. 셋째, τ 를 상수로 고정하는 경우로 V 는 $v(i)$ 의 총합에 부 윈도우의 가중치를 곱한 미디어 가중치이다. 따라서, 최근에 발생한 참조 빈도와 전송량은 캐쉬 교체 정책 경향을 나타내는 중요한 인자이며 시간대와 콘텐츠 별로 편중하거나 콘텐츠 선호도 변화에 따라 캐쉬 오브젝트 교체에 효율적이다.

이와 같이 τ 범위에 따라 결정된 가중치 값 V 가 같은 경우, $M_{replace}$, M_{mean} , 참조 시간을 비교하여 우선 순위를 결정한다. 먼저, $M_{replace}$ 는 M_{mean} 에 대한 블럭 여유량으로 값이 작을 수록, 사용자 QoS에 대한 위급도 증가와 스트리밍 서버로부터 캐쉬 오브젝트의 반입량 증가로 비교 우위에 배정한다. $M_{replace}$ 값이 같은 경우, M_{mean} 을 기준으로 평균 참조 전송량이 높은 콘텐츠를 높은 우선 순위로 배정하고, M_{mean} 도 같다면 가장 최근에 참조된 콘텐츠에 높은 우선 순위를 배정한다.

3 성능 분석

가중치 윈도우 기반의 WIN 캐쉬 교체 정책은 참조시간에 대한 LRU 정책, 참조량에 대한 LFU 정책, 세스먼트 기반의 SEG 정책과 비교 분석한다. 성능분석은 각 정책에 대해 동일한 시뮬레이션 환경 하에서 측정하였으며 7일간의 시간 범위와 1,000개의 콘텐츠를 대상으로 하였고, τ , 콘텐츠 크기 종류, Zipf 요구 분포와 콘텐츠 요구 분포 이동을 적용하여 실험하였다. 캐쉬 용량은 최대 콘텐츠 크기와 최대 허용 스트림 수를 통해 최소 캐쉬 저장 공간을 계산하고 10 KBlocks 단위로 증가시켜 구성하였다. 요구 편중 시간은 요구 주기인 하루를 기준으로 22시로 설정하였다. 이와 같은 실험 환경을 기반으로 각 실험은 참조 적중률, 참조량 적중률, 지연된 참조률과 캐쉬 반입량을 측정하였다.

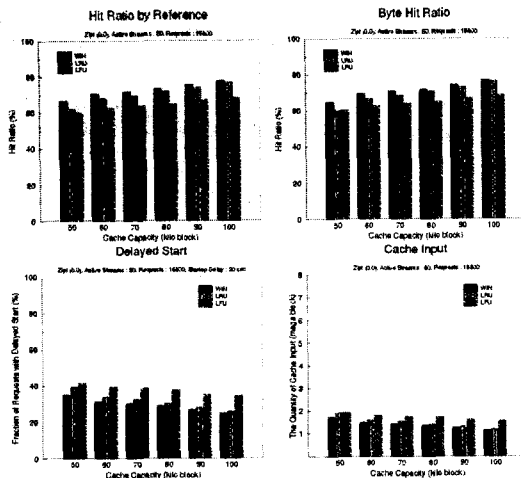


그림 3: WIN과 LRU, LFU의 성능 비교

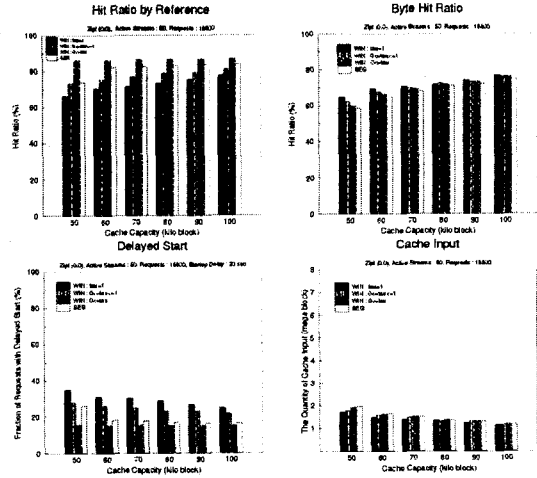


그림 4: WIN과 SEG의 성능비교

그림 3은 $\tau=1$, Zipf(0,0)과 콘텐츠 요구 분포일 때, LRU, LFU와 WIN의 성능 비교이다. 캐쉬 교체 정책은 신규 콘텐츠에 대한 집중적인 요구가 발생하면 캐쉬 오브젝트 등록과 함께 빠른 시간 안에 높은 우선 순위에도달되어야하고 최대한 많은 양의 콘텐츠를 캐쉬에 유지하는 것이 유리하다. 반면, 일정 기간 동안 선호 경향이 높았다가 점차적으로 선호 경향이 낮아지면 콘텐츠의 우선 순위도 점차적으로 낮아져야 하고, 최종적으로 캐쉬에서 제거되어야만 캐쉬 저장 공간을 효율적으로 사용할 수 있다. 이러한 사용자 요구 경향이 운영체제 캐쉬에서 지역 편중 참조 경향에 대한 참조율 누적 방식의 LFU 캐쉬 교체 정책에 응용하게 되면 요구 편중에 따라 최고 우선 순위가 될 때 좋은 성능을 나타낼 수 있지만, 참조율이 떨어지게 되는 경우 누적된 참조 횟수의 변화가 없어 우선 순위 이동이 느리거나 적용되지 않을 수 있다. 또한, 참조 시간에 따른 LRU 캐쉬 교체 정책은 선호도 경향이 쉽게 반영될 수 있지만, 그러한 변화가 일순간에 바뀔 수 있는 점에서 선호도가 점차적으로 줄어드는 경향에 대해서 효율적으로 적용되지 않는다. WIN 캐쉬 교체 정책의 성능 향상 요인은 콘텐츠 요구 분포의 이동으로 캐쉬 오브젝트가 교체되는 과정에서 LRU와 LFU의 참조 적중률과 참조량 적중률 등이 상대적으로 낮아지기 때문이다.

그림 4의 실험은 클라이언트의 초기 재생 지연 방지를 위해 사용자 QoS를 고려한 SEG 캐쉬 정책[4]과 τ 범위에 따른 WIN 정책을 비교하였다. SEG 캐쉬 정책은 캐쉬 블럭을 세그먼트 단위로 관리하며, i 번째 세그먼트에 2^{i-1} 블럭 개수를 할당한다. 세그먼트의 가중치는 현재시간(T)과 마지막 참조된 시간(T')의 차이값($T-T'$)에 세그먼트의 거리(i)를 곱한 식의 역수 ($1/((T-T') \times i)$)를 계산하여 작은 값을 갖는 세그먼트를 희생 세그먼트로 선택한다. 즉, 최근에 참조되지 않은 세그먼트 중에 가장 높은 순위의 세그먼트(가장 많은 데이터를 캐싱한 오브젝트)를 선택하여 캐싱 공간을 만든다. $0 \leq \tau \leq 1$ 일 때, WIN 정책이 SEG에 비해 참조 적중률과 지연된 참조물에서 높은 성능을 나타내는 이유는 WIN 정책에서 참조 빈도수가 낮은 콘텐츠에 대한 캐싱량이 작아질수록 τ 값이 역비례로 증가하여 전송 지연 방지량 이상의 캐싱량을 유지함으로써 평균 전송량 이하의 인덱트 오브젝트 수가 증가하고, SEG 정책은 전송 지연 방지량에 근사한 캐쉬 오브젝트 크기와 LRU 정책에 영향받는 다양한 크기의 콘텐츠들로 혼재되어 있기 때문이다. SEG 정책은

WIN 정책과 비교하여 $0 \leq \tau$ 와 $0 \leq \tau \leq 1$ 의 중간 성능을 보였으며, 참조량 적중률과 캐쉬 반입량에서 유사한 성능 경향을 보였다.

4 결론

본 논문에서는 미디어 캐싱 시스템의 효율적인 캐싱 구조를 위하여 가중치 윈도우에 기반한 캐쉬 교체 정책을 제시하였다. 제안된 캐쉬 교체 정책은 참조 횟수, 참조량, 참조 시간 등의 정량적인 인자들과 사용자 요구 주기를 적용하여 기존 캐쉬 교체 정책과 차별화된 개념을 도입하였으며, 최근 참조 경향에 높은 가중치를 부여함으로써 변화하는 콘텐츠 선호 경향에 빠르게 적용하는 구조를 제시하였다. 교체 정책 성능 분석은 시뮬레이션 환경 구축을 통해 실험하였으며 기존 캐쉬 교체 정책인 LRU, LFU와 SEG보다 참조 적중률, 참조량 적중률, 시작 지연률과 반입량에서 향상된 결과를 보였다.

본 연구의 향후 과제는 캐싱 시스템 구축에 적용하여 실제 사용자 요구 경향에 대한 실험을 통해 효율성을 검증하는 것이다.

참고문헌

- [1] S. Acharya, 'Techniques for Improving Multimedia Communication over Wide Area Networks,' *Ph.D. Thesis*, Cornell University, 1999.
- [2] M. Hofmann, T. S. Eugene Ng, K. Guo, S. Paul, and H. Zhang, 'Caching Techniques for Streaming Multimedia over the Internet,' *Technical Report*, Bell Laboratories, April. 1999.
- [3] G. K. Zipf, *Human Behavior and the Principle of Least Effort*, Addison Wesley, Reading, MA, 1948.
- [4] K. Wu, P. S. Yu, and J. L. Wolf, 'Segment-based Proxy Caching of Multimedia Streams,' *Proceedings of the 10th International WWW Conference*, May 2001, pp.36-44.