

이질적 웹 서버 클러스터 환경에서 효율적인 우선순위 가중치 맵핑

(Effective Prioritized HRW Mapping in Heterogeneous Web Server Cluster)

김진영[†] 김성천^{**}
(Jinyoung Kim) (Sungchun Kim)

요약 인터넷 서비스에 대한 요청의 기하학적인 증가와 그 요구의 양과 질에 대한 다양성은 이질적인 능력을 지닌 서버들로 구성된 웹 서버 클러스터를 구축하게 되었다. 클러스터 환경에서 가장 결정적인 역할을 하는 요청 대 서버의 맵핑 기법이 최근 활발히 논의되고 있다. 기존의 맵핑 방식은 요청의 수를 기반으로 서버들이 동일한 양의 부하를 할당받는 것을 목적으로 한다. 하지만, 최근의 다양한 서비스의 증가로 인해 단순한 부하의 균등화만으로 적절한 지연 시간을 기대할 수 없게 되었다. 이에 요구되는 내용을 기반으로 맵핑을 수행하여 응답 시간을 단축하고 전체 서버들의 캐시 적중률도 높이는, 내용-기반 맵핑(content-based mapping)이 최근의 인터넷 환경에서 매우 긍정적으로 평가받고 있다. 이에 본 논문에서는 내용-기반 맵핑 방식을 이질적 환경에 적합하도록 개선시킨 우선 순위 최상 임의 가중치 맵핑(Prioritized Highest Random Weight Mapping, PHRW Mapping)을 제안하였다. 요구되는 내용을 기반으로 맵핑하며, 이질적 서버들에 대해 능력에 따라 우선 순위를 부여하여 높은 성능의 서버가 많은 요청들을 처리하는 방식이다. 이를 통해 이질적인 서버들로 구성된 클러스터에서 매우 효과적으로 적용하며, 특히 지연 시간에 제한을 갖고 있는 실시간 데이터 서비스의 지연 시간을 효과적으로 감소시킨다. 알고리즘과 시뮬레이션을 통하여, 제안된 기법을 사용하는 경우 짧은 지연 시간을 보장하여 전체 시스템의 성능이 향상됨을 증명하였다.

키워드 : 분산처리, 클러스터, 웹서버, 부하균등화

Abstract For many years clustered heterogeneous web server architecture has been formed on the internet because the explosive internet services and the various quality of requests. The critical point in cluster environment is the mapping schemes of request to server, and recently this is the main issue of internet architecture. The topic of previous mapping methods is to assign equal loads to servers in cluster using the number of requests. But recent growth of various services makes it hard to depend on simple load balancing to satisfy appropriate latency. So mapping based on requested content to decrease response time and to increase cache hit rates on entire servers - so called "content-based" mapping is highly valued on the internet recently. This paper proposes Prioritized Highest Random Weight mapping(PHRW mapping) that improves content-based mapping to properly fit in the heterogeneous environment. This mapping scheme that assigns requests to the servers with priority, is very effective on heterogeneous web server cluster, especially effective on decreasing latency of reactive data service which has limit on latency. This paper have proved through algorithm and simulation that proposed PHRW mapping show higher-performance by decrease in latency.

Key words : cluster, content-based, mapping, heterogeneous, priority

· 본 연구는 한국과학재단 목적기초연구 R01-2001-000-00356-0(2002) 지원으로 수행되었음

[†] 비회원 : LG전자 전자기술원
realzero@hotmail.com
^{**} 종신회원 : 서강대학교 컴퓨터학과 교수
ksc@arqlab1.sogang.ac.kr
논문접수 : 2001년 4월 14일
심사완료 : 2003년 8월 16일

1. 서론

인터넷은 폭발적으로 성장하여, 기존의 인터넷 서버들에 과중한 부담을 가져왔다. 이에 서버들로 클러스터를 구성하는 구조가, 확장성을 갖으며, 믿을 수 있고, 고성능을 수행하는 성공적인 대안으로 각광받고 있다[1]. 이와 같은 클러스터 기반 구조의 중요한 요구는, 입력된

요청들(requests)을 응답하기 가장 적절한 서버로 보내고, 라우팅해야 한다는 것이다.

기존의 요청-대-서버 맵핑은 요청의 수에 따라 부하 균등을 이루는 방식이었다. 이러한 방식은 상태 정보를 필요로 하지 않으면서 효과적으로 부하를 균등화하지만, 응답 시간의 최적화를 가져오지는 않는다. 따라서 응답 시간을 효율적으로 감소시키기 위해서, 서로 다른 대상들에 대한 요청들은 여러 개의 서버들로 나뉘어지는 반면, 같은 대상에 대한 요청들은 같은 서버로 보내지는 방식으로 요청들을 서버들로 맵핑시키는 '내용-기반' 맵핑(content-based mapping)이 제안되었다[2-4]. 특히 효과적으로 내용-기반 맵핑을 구현하기 위해 가중치 함수를 이용한 최상 임의의 가중치(Highest Random Weight, HRW) 맵핑이 제안되었다[4].

한편, 기존의 내용-기반 맵핑인 HRW 맵핑은 클러스터를 구성하는 모든 서버의 능력이 같다고 가정한 상태에서 가장 좋은 성능을 나타낸다. 이는 실질적인 최근의 네트워크 환경과 다르며, 점차로 요구가 증가되고 있는 이질적인 클러스터 환경의 이점을 사장시킨다[5,6]. 따라서 본 논문에서는 이질적인 클러스터에 적합하도록 HRW 맵핑에 우선 순위를 적용한 우선 순위 최상 임의의 가중치(Prioritized Highest Random Weight, PHRW) 맵핑을 제안한다. 그리고 기존의 대표적인 맵핑 방식인 라운드 로빈 맵핑, 기존의 내용-기반 맵핑 방식인 최상 임의의 가중치 맵핑과 제안 맵핑 방식을 비교하여 제안한 기법의 효율성을 증명하고 시뮬레이션을 통해 이를 검증한다.

2. 클러스터 환경의 기존 맵핑 기법과 문제점

기존의 맵핑 기법은 크게 상태 정보를 필요로 하는 기법들과 상태 정보를 필요로 하지 않는 기법들이 있다. 상태 정보를 필요로 하는 기법은 추가로 오버헤드가 증가하므로, 상태 정보를 무시하는 기법들이 일반적으로 쓰이고 있다. 요청들을 임의의 방법으로 선택된 서버로 보내는 임의 맵핑(Random Mapping)이 있으며, 가장 대표적으로 사용되는 라운드-로빈 맵핑(Round-Robin Mapping, RR Mapping)이 있다. 또한 RR 맵핑에 우선 순위를 부여하여 이질적 환경에 적용시킨 우선순위 라운드-로빈 맵핑(Priority Round-Robin Mapping, PRR Mapping)[7]이 있다.

연속적인 요청이 연속적인 서버들로 보내지는 RR 맵핑은 간단한 방식이다. 모든 서버들이 작동할 때, 진정한 라운드-로빈 방식은 클러스터로 보내진 n 번째 요청을 n 번째 서버(m 으로 나눈 값)로 맵핑시킨다. 따라서 요청 r 에 대한 맵핑 함수는 $F(r_n) = n \pmod{m}$ 이다. RR 맵핑은 보통 FIFO 큐로 실행된다. 첫 번째 요소는

큐의 앞부분으로부터 제거되고, 스케줄되고, 큐의 마지막 부분에 다시 삽입된다. 모든 요소는 똑같이 취급되고, 어떠한 우선 순위도 없다.

PRR 맵핑은 한 차원 발전된 RR 맵핑 방식이다[7]. RR 맵핑의 출력 민감성(sensitivity)과 기아-없는(starvation-free) 성능을 유지하면서, 우선 순위를 적용한다. PRR 맵핑의 수행을 위해서, RR 맵핑의 장점과 전체 정렬(full sorting)을 합친, 다중 단계 FIFO 큐에서 대강의 정렬을 한다. 단계들은 RR 순서로 처리되는 반면, 각 단계는 서로 다른 단계에 있는 요소가 선택되는 빈도에 영향을 미치는 우선 순위를 할당받는다. 기본적으로 더 높은 우선 순위를 갖는 요소들은 낮은 우선 순위를 갖는 요소보다 훨씬 자주 스케줄된다.

RR 맵핑은 부하 균등의 면에 있어 매우 좋은 성능을 보인다. 하지만 서버 부하의 균등이 맵핑을 선택하는 최우선의 기준은 아니다. 결국 더욱 중요한 것은 응답 시간의 최적화이다. 서버 부하 균등은 응답 시간의 최소화를 보장할 때 비로소 중요한 목적이 된다.

응답 시간 최적화는 기댓값 뿐만 아니라 응답 시간의 분산까지 감소해야 함을 의미한다. 도메인을 캐싱하는 RR 맵핑의 심각한 문제점은, 부하 균등에 의한 응답 시간의 감소가, 실제적으로는 캐시 적중 실패(cache miss)에 의한 탐색 지연으로 인한 지연 시간의 심각한 증가로 인해서 상쇄된다는 것이다. 각 캐시 적중 실패는 원격 제공자로부터의 탐색과 지역 캐시의 탐색보다 비싼 연산을 요구한다. 이는 서버 캐시의 데이터들의 중복을 가져온다. 완전한 중복(full replication)은 꽤 빨리 달성되며 클러스터의 유효한 캐시 크기는 빠른 속도로 감소한다.

이에 그림 1과 같이 요청들이 요구하는 대상의 이름으로 서버에 맵핑하는 내용-기반 맵핑 기법이 제안되었다. 내용-기반 맵핑 기법인 HRW 맵핑은 이전의 서버 상태에 대한 정보가 아닌, 대상 이름을 이용하여 서버로 요청을 맵핑한다. 대상 이름과 서버 주소 모두가 각 서버에 임의의 "가중치"를 할당하기 위해 사용된다. 그리고 서버들은 가중치에 의해 정렬되고, 요청은 가장 높은 가중치를 갖는 유효한 서버로 보내진다.

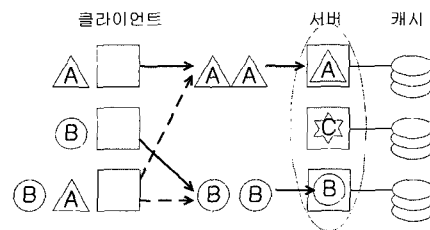


그림 1 내용-기반 맵핑

HRW 맵핑은 항상 주어진 대상 이름을 클러스터내의 같은 서버로 맵핑하기 때문에, 클라이언트들이 지역적 정보만을 바탕으로 즉각적인 결정을 할 수 있어서 낮은 부담을 가져오며, 부하 균등을 보장한다. 또한 각 서버들은 특정 대상만을 처리하면 되기 때문에, 데이터 중복의 부담이 없으므로 높은 적중률을 가져온다. 마지막으로 한 서버가 사용할 수 없게 되었을 때 그 서버에 맵핑되어 있던 대상들만을 다시 맵핑하면 되므로 최소 분열을 보장한다.

HRW 맵핑은 기본적으로 클러스터를 구성하는 서버들의 능력을 모두 같다고 가정했을 때 가장 좋은 성능을 나타내고 있다. 하지만 최근에 들어 실제 네트워크가 동질적인 서버들로만 구성되는 경우보다는, 이질적인 능력을 갖는 서버들로 구성되는 경우가 더욱 많으며, 이러한 경향은 앞으로 더욱 심화될 것이다. 따라서 만일 서버들이 서로 다른 능력을 갖는다면, 더 많은 능력을 지닌 서버들이 전체 부하에서 비효율적으로 더 많은 부분을 받도록 해야 한다. 이를 통해 개선된 서버들의 능력을 최대한 사용하면서, 기존의 서버들의 능력도 효율적으로 이용할 수 있게 된다.

3. 이질적 클러스터 환경에서 개선된 내용-기반 맵핑 기법

이질적 웹 서버 클러스터 환경에서 HRW 맵핑의 문제점을 살펴보았다. 따라서, HRW 맵핑의 특징인 낮은 부담, 부하 균등, 높은 적중률, 그리고 최소 분열을 유지하면서, 이질적인 환경에서 높은 효율을 가져올 수 있는 우선 순위 최상 임의 가중치 맵핑(Prioritized Highest Random Weight Mapping), PHRW 맵핑을 제안하고자 한다. PHRW 맵핑을 HRW 맵핑과 비교하였을 때, 가장 큰 차이점은 각 서버에 대해 그 능력에 따라 우선 순위를 부여한다는 것이다. 이를 통해 가중치를 스케일하여, 더 큰 능력을 지닌 서버가 상대적으로 더 많은 요청들을 받아 처리하도록 한다.

본 논문에서는 클러스터내의 서버들이 서로 다른 능력을 갖고 있는 이질적인 웹 서버 환경을 가정하고 있기 때문에 서버에 따라 다른 우선 순위를 갖는다. 즉, 각 서버는 자신의 IP 주소와 할당받은 우선 순위, 가지고 있는 대상의 이름을 가지고 가중치를 계산하고, 계산된 가중치에 따라 정렬되어 그 중 가장 높은 가중치를 갖고 있는 서버가 요청을 처리하게 된다.

PHRW 맵핑은 기존의 HRW 맵핑에서 가중치를 계산하는 의사입의 함수에 우선 순위를 적용한다. 따라서 최상의 성능을 가져오기 위해서는 효율적인 우선 순위를 결정하는 것이 중요하다. PHRW 맵핑은 서버의 주소, 대상의 이름, 그리고 우선 순위를 통해 가중치를 결

정하기 때문에, 웹 서버의 능력에 따라 각 서버에 우선 순위를 부여한다. 이때 평균 능력을 갖는 서버를 기준으로 우선 순위를 정의한다.

평균 능력을 갖는 서버의 우선 순위를 P_{medium} 으로 하고, 그보다 높은 능력을 갖는 서버에는 P_{high} 의 우선 순위를 할당한다. 또한 P_{high} 보다 좋은 성능을 보장할 수 있는 서버에는 $P_{highest}$ 의 우선 순위를 할당한다. P_{medium} 보다 낮은 능력을 갖는 서버에는 P_{normal} 의 우선 순위를 부여한다[10,11,12]. 실제로 이러한 조건에서 우선 순위는 다양한 값을 가질 수 있다. 우선 순위를 제한적인 실수 값으로 정의하는 것은 큰 의미를 갖지 않으므로, 우선 순위간의 관계만을 정립한다. 서버에 대한 우선 순위 값의 기본 단위를 $P_{priority}$ 라고 하면, $P_{normal} = P_{priority}$, $P_{medium} = 2P_{priority}$, $P_{high} = 3P_{priority}$, 그리고 $P_{highest} = 4P_{priority}$ 이다.

PHRW 맵핑의 성능에서 가장 결정적인 역할을 수행하는 부분은, 가중치를 결정하는 가중치 함수이다. 모든 클라이언트들에게 동일한 가중치 함수를 적용함으로써 부담을 낮출 수 있고, 효율적인 가중치 함수의 정의는 전체 성능의 효과적인 향상을 가져올 수 있다. W_{PHRW} 함수는 기 발표된 이름기반 맵핑 함수 W_{HRW} 에 우선 순위 개념을 할당하여 가중치 함수를 변경하였다[4]. 이는 각 서버의 능력에 따라 가중치를 돕으로써, 이질적인 웹서버 환경에서 다양한 데이터를 처리해야 하는 현실세계에서 보다 효과적으로 데이터를 처리할 수 있도록 한다. HRW 맵핑의 가중치 함수를 기반으로, PHRW 맵핑에 적합한 가중치 함수 W_{PHRW} 를 다음과 같이 제시한다.

$$W_{PHRW}(k, S_i, P_i) = (1103515245 \cdot P_i \cdot ((1103515245 \cdot 5 \cdot S_i + 12345) \text{ XOR } D(k)) + 12345)$$

위의 가중치 함수에서 $D(k)$ 는 대상 이름 k 를 31-bit로 나타낸 것이고, S_i 는 클러스터 내의 i 번째 서버의 주소이고, P_i 는 서버 S_i 에 할당된 우선 순위이다. 먼저 $D(k)$ 를 계산하고, 각 S_i 에 대해 $W_{PHRW}(k, S_i, P_i)$ 를 계산한다.

가중치 함수 W_{PHRW} 는, XOR 수행 이후의 임의화 단계에 우선 순위를 적용시킨다. 그 첫 번째 이유는 우선 순위가 상대적인 관계로 정의되어 있기 때문이다. 따라서 상황에 유동적으로 우선 순위를 결정해야 하므로, XOR 단계 이후의 임의화 단계에 우선 순위 값을 반영하는 것이 논리적이다. 둘째로, 가중치를 계산하는데 있어, 대상 이름 $D(k)$ 가 XOR에서 가장 결정적인 역할을 한다. 따라서 유동적인 우선 순위 값을 XOR 단계에 삽입하는 것은, 전체 가중치 값을 예측 불가능하게 만들며, 대상 이름이 정확하게 반영되지 않을 가능성을 포함하게 된다.

4. 시뮬레이션 검증

제안한 기법과 기존의 기법에 대하여 큐잉이론(queueing theory)에 따른 이산사건 모델(discrete event model)을 통해 시뮬레이션을 SUN ULTRA 60(Solaris 2.6) Ultrasparc-II360MHz CPU 시스템에서 SUN WorkShop Compilers C/C++ 4.2를 이용하여 수행하였다. 본 시뮬레이션에서는 요청이 입력되었을 때부터 맵핑 함수가 수행되어 서버를 선택하고, 그 서버에 할당되어 요청이 처리되기까지 걸리는 지연 시간을 측정한다. 측정된 지연 시간에 대한 비교를 통해 각 맵핑 기법들의 성능을 비교한다. 서버의 개수는 데이터 서비스를 위한 평균 지연 시간과 비례한다. 서버의 개수가 증가할수록 평균 지연 시간은 감소한다. 또한 서버의 개수가 많아질수록 각 서버가 가지는 데이터들의 중복성이 줄어든다. 서버의 개수를 1개에서 8개까지 제한한 것은 8개의 서버 이후에는 평균 지연시간의 감소가 많지 않았기 때문이다. 제안한 맵핑 기법은 기존 기법에 단순한 곱셈 단계가 추가되는 것이므로, 시간 복잡도의 증가를 가져오지 않으며 성능 평가에서 이는 무시될 수 있다. 또한 일반적으로 가장 많이 사용되는 RR 맵핑과 PRR 맵핑에 대한 제안 기법의 응답 시간 면에서의 성능 향상을 보이기 위해서, RR/PRR 맵핑 수행에 따른 서버 수에 비례하는 데이터 중복에 의한 오버 헤드를 성능평가에 반영하지 않았다. 가벼운 네트워크 부하 상황과 과중한 네트워크 부하 상황에 따라 서버의 수와 맵핑 방식에 따른 지연 시간의 관계를 알아본다[8,9]. SURGE를 사용하여 생성된 트래픽을 측정하여 얻은 결과로 UE (user equivalents)단위로 네트워크의 부하 상황을 판단할 수 있다. 가벼운 네트워크는 40UEs이고 과중한 네트워크는 520UEs로 가정하였다[13]. 이때 트래픽을 실시

간 데이터 서비스와 비 실시간 데이터 서비스로 구분하여 이들 각각의 평균 지연 시간을 측정하였다. 실시간 데이터는 지연에 제한적이지만 손실을 허용하는 반면 비 실시간 데이터는 지연에 대한 제한은 없지만 손실을 불허하는 특성을 갖는다. 과거 비 실시간 데이터들이 주로 서비스된데 반해, 갈수록 음성이나 비디오와 같은 실시간 데이터의 수요가 증가하고 있다. 여기서는 실시간 데이터와 비 실시간 데이터를 구분할 때 임의적으로 실시간 데이터는 500KB와 20KB로, 20KB와 1KB를 비 실시간 데이터로 구분하여 가정하였다.

실시간 데이터의 경우는 그림 2에서 나타나는 것처럼, RR 맵핑과 PRR 맵핑은 데이터 중복을 허용했음에도 불구하고 내용 기반 맵핑 기법인 PHRW 맵핑이나 HRW 맵핑보다 훨씬 성능이 떨어짐을 볼 수 있다 (PHRW 맵핑시 가벼운 네트워크 부하에서 평균 10.0%, 과중한 네트워크 부하에서 평균 7.4% 평균 지연 시간 단축). 이는 서버의 능력에 상관없이 요청들을 그 순서에 따라 서버들에 할당하는 맵핑 방식에 따른 당연한 결과이다. 또한 제안하는 PHRW 맵핑이 HRW 맵핑보다 개선된 성능을 나타냄을 알 수 있다. 이는 더 많은 능력을 가지고 있는 서버로 더 많은 요청이 할당될 수 있도록 우선 순위에 의해 스케일 했기 때문이다. 즉 더 높은 우선 순위를 갖고 있는 서버는 그렇지 않은 서버에 비해 같은 요청을 처리하는데 필요한 평균 지연 시간이 더 짧다. 그런데 우선 순위가 높은 서버에 상대적으로 많은 요청들이 할당되므로 PHRW 맵핑을 사용하는 경우에 평균 지연 시간은 단축된다.

반면 비 실시간 데이터의 경우 그림 3에서 보는 것과 같이, RR 맵핑과 PRR 맵핑이 PHRW 맵핑이나 HRW 맵핑보다 좋은 성능을 나타낸다. 이는 RR 맵핑과 PRR

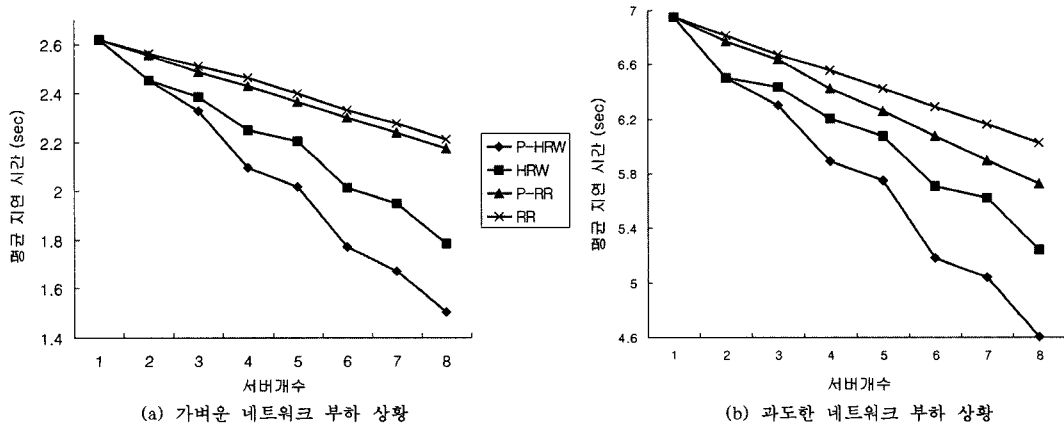


그림 2 실시간 데이터 서비스를 위한 평균 지연 시간

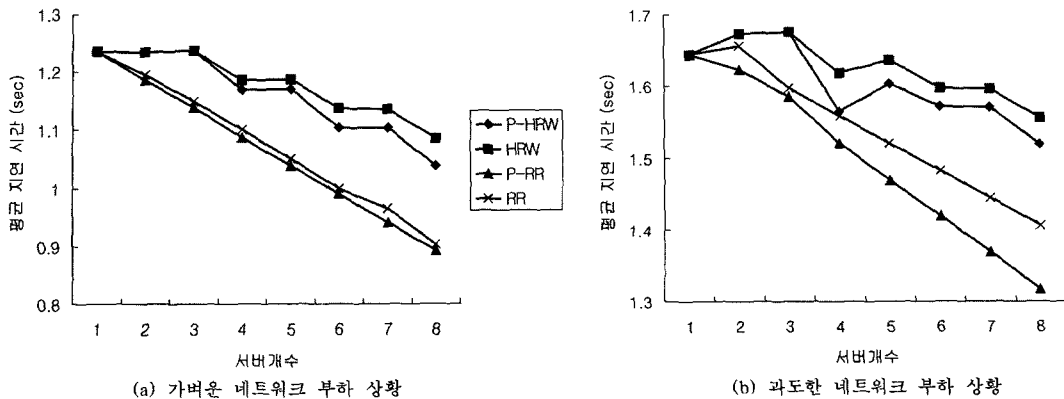


그림 3 비실시간 데이터 서비스를 위한 평균 지연 시간

맵핑은 서버의 능력을 고려하지 않고 요청들을 할당하기 때문에 새로 추가된 성능이 좋은 서버가 비 실시간 데이터를 처리할 경우가 $1/m$ (m = 서버의 개수)의 확률을 갖기 때문이다. PHRW 맵핑의 경우에는 새로 추가된 서버는 무조건 실시간 데이터를 처리하도록 할당되고, 이미 존재하는 상대적으로 성능이 낮은 서버들이 비 실시간 데이터를 처리하기 때문에 당연히 RR 맵핑보다 긴 지연 시간을 가져오게 된다. 하지만 실시간 데이터는 시간 지연을 허용하지 않지만, 비 실시간 데이터는 시간 지연을 허용하고 손실을 허용하지 않는다. 따라서 실시간 데이터의 지연 시간을 증가시키면서 비 실시간 데이터의 지연 시간을 감소시키는 것은 그다지 효율적이지 못함을 알 수 있다.

5. 결론

요청을 웹 서버에 맵핑하는데 사용된 기존의 방법은 주로 RR 맵핑으로, 이는 서버의 능력이나 요청의 내용, 서비스의 QoS에 대해서는 전혀 고려하지 않는 맵핑 방식이다. 따라서 이질적인 웹 서버 클러스터에 효율적인 새로운 맵핑 방식이 필요하다.

본 논문에서는 이질적인 웹 서버들로 구성된 클러스터 환경에 적합한 맵핑 기법으로 내용-기반 맵핑 기법을 기반으로 하여, 기존의 HRW(Highest Random Weight) 맵핑을 개선한 PHRW(Prioritized Highest Random Weight) 맵핑을 제안하고 성능 비교를 수행하였다. PHRW 맵핑은 요청되는 대상을 보고, 그 대상을 전문적으로 처리하는 서버로 요청을 할당하는 방식으로, 응답 시간을 감소시키고, 서버들간에 데이터의 중복이 없어서 캐시 적중률을 높이며, 새로운 서버의 삽입이나 기존 서버의 제거 시에도 최소의 부담으로 재 맵핑이 가능하도록 한다. 또한 HRW 맵핑과는 달리 이질적인

환경에서 적합하도록 각 서버에 우선 순위를 부여하여 상대적으로 높은 성능을 지닌 서버로 더 많은 요청들이 할당되도록 하여 전체적인 성능 향상을 꾀한다.

실제 시뮬레이션 결과 제안된 맵핑 기법이 전체적인 성능 향상뿐만 아니라, 일반적인 웹 트래픽의 성질에도 매우 적합함을 알 수 있다. 서버들에 분배하는 대상 자체에도 그 성질에 따라 우선 순위를 부여하였기 때문에, 지연 시간에 제한을 갖는 실시간 데이터 서비스에 있어서 최소 2.4%(과중한 네트워크 부하 시에는 2.1%), 최대 15.7%(과중한 네트워크 부하 시에는 12.1%)의 성능 향상을 가져왔다. 이는 PHRW 맵핑이 단순히 전체 트래픽에 대한 평균 지연 시간을 감소시키는 기법이 아닌, 실시간 데이터 서비스에 있어 매우 효율적인 지연 시간 감소를 가져오고 이를 통해 전체 평균 지연 시간을 감소시켰음을 보여준다.

참고 문헌

- [1] Mohammad Banikazemi, Vijay Moorthy, Dhavalleswar K. Panda, "Efficient Collective Communication on Heterogeneous Networks of Workstations," National Science Foundation, Jun. 2000.
- [2] Vivek S. Pai, Mohit Aron, Erich Nahum, "Locality-Aware Request Distribution in Cluster-based Network Servers," Proceedings of the 8th international conference on Architectural support for programming languages and operating systems, pp. 205-216, 1998.
- [3] Chu-Sing Yang, Mon-Yen Luo, "An Effective Mechanism for Supporting Content-based Routing in Scalable Web Server Clusters," International Conference on Parallel Processing, pp. 240-245, 1999.
- [4] David G. Thaler, Chinya V. Ravishankar, "Using Name-Based Mappings to Increase Hit Rates,"

- EEE/ACM Transactions on Networking, Vol. 6 No. 1, pp. 1-14, Feb. 1998.
- [5] Robert Elsasser, Burkhard Monien, Robert Preis, "Diffusive Load Balancing Schemes on Heterogeneous Networks," Twelfth annual ACM symposium on Parallel algorithms and architectures, pp. 30-38, Jul. 2000.
- [6] Pangfeng Liu, Tzu-Hao Sheng, "Broadcast Scheduling Optimization for Heterogeneous Cluster Systems," Twelfth annual ACM symposium on Parallel algorithms and architectures, pp. 129-136, Jul. 2000.
- [7] Chris Faisstnauer, dieter Schmalstieg, Werner Purgathofer, "Priority Round-Robin Scheduling for Very Large Virtual Environments," Proceedings of the IEEE Virtual Reality 2000 Conference, pp. 135-142, Mar. 2000.
- [8] Paul Barford, Mark Crovella, "Measuring Web Performance in the Wide Area," Special Issue on Network Traffic Measurement and Workload Characterization, Aug. 1999.
- [9] "SPECweb99 Release 1.02," <http://www.spec.org>, Jul. 2000.
- [10] Ron Cocchi, Deborah Estrin, Scott Shenker, Lixia Zhang, "A Study of Priority Pricing in Multiple Service Class Networks," Proceedings of the conference on Communications architecture & protocols, pp. 123-130, Sep. 1991.
- [11] "Layer 2 Traffic Prioritization," http://www.intel.com/network/white_papers/
- [12] Fred A. Maymir-Ducharme, "Dynamic Priorities, Priority Scheduling and Priority Inheritance," Proceedings of the fourth international workshop on Real-time Ada issues, pp. 39-45, Jul. 1990.
- [13] P. Barford and M. Crovella. "Generating representative workloads for network and server performance evaluation," Proceedings of ACM SIGMETRICS'98, pp. 151-160, June 1998.



김진영

1995년 3월~1999년 2월 서강대학교 전자계산학과. 1999년 3월~2001년 2월 서강대학교 컴퓨터학과 고성능시스템연구실. 2001년~2003년 삼성전자 통신연구소. 2003년~현재 LG전자 전자기술원

김성천

정보과학회논문지 : 시스템 및 이론
제 30 권 제 1 호 참조