

Cache 기법을 도입한 복제 웹서버 시스템*

진상혁 정연진 이광모
한림대학교 정보전자공과대학 컴퓨터공학부

The Replica Web Server Systems Using Cache Mechanism

Sanghyouk Jin, Younjin Jeong, Kwangmo Lee
School of Computer Engineering, Hallym University

요 약

이 논문은 WWW과 데이터베이스의 시스템 통합에 있어서 독립 서버로 운영될 때 CGI 응용 프로그램의 과부하를 줄이는 복제 서버 시스템 모델[1]에서 데이터베이스 시스템에 매번 연결 및 종료 시 발생하는 접속 시간의 증가를 감소시킬 수 있는 cache 기법의 도입과 함께 WWW 서버의 시스템 큐(queue)의 길이로 시스템 작업량을 측정하여 가장 작업량이 적은 서버를 선택하는 조정자(coordinator)에 cache 정보를 기록하는 기능을 추가하여 보다 원활한 WWW 작업 분배를 기대할 수 있고 안정성이 향상된 모델을 제시한다.

1. 서론

오늘날 인터넷의 등장과 함께 발전된 WWW의 폭발적인 증가로 전자 상거래나 분산된 정보 검색 등의 서비스들이 가능하게 되었고 차바 및 멀티미디어, 데이터베이스까지 통합된 종합 시스템으로 발전하게 되었다. 이러한 WWW의 증가는 이질적으로 분산된 데이터를 Common Gateway Interface[2]를 통하여 검색할 있는 방법을 제공하게 이르렀다. WWW 시스템에서 CGI를 제공함으로써 기존의 시스템과 WWW 시스템을 통합하는 많은 노력이 있었다. 최근 국내에서도 대학, 회사 및 연구소 등에서 인트라넷을 구축을 하고 CGI를 이용한 게시판이나 상품 광고 및 관련 정보검색을 제공하고 있다.

이러한 환경에 있어서 제공되는 정보의 질과 마찬가지로 WWW 서비스 시스템에서 지속적이고 안정적인 정보제공이 무엇보다 중요한 요소 중의 하나이다.

이 연구의 목적은 단일 WWW 시스템에서 CGI 응용 프로그램과 데이터베이스 시스템의 통합 과정에 발생하는 성능 저하를 줄이고 지속적이고 안정적인 WWW 서비스를 제공하는 방법을 제안한 것이다.

이 논문의 구성은 다음과 같다. 2장에서는 Common Gateway Interface에 대하여 살펴보고 3장에서는 WWW 시스템과 데이터베이스 시스템에 대하여 기술하고 4장에서는 본 연구에서 제안한 캐시를 이용한 복제 서버 시스템을 제시하였으며, 5장에서는 결론과 향후과제를 기술하였다.

2. Common Gateway Interface

일반적으로 WWW을 이용한 정보제공에 있어서 정적인 문서들을 모아 하이퍼링크를 제공하는 방식과 사용자의 입력 요구에 따라 동적으로 문서를 제공하는 방식이 있는데 일반적으로 CGI를 통해 구현된다. 모든 WWW 서버에서는 CGI를 제공하며 다른 형태로 된 데이터를 쉽게 획득하고 외부 프로그램이나 소프트웨어를 실행할 수 있으며 기존에 운용중인 시스템과의 통합을 쉽게 할 수 있는 메커니즘을 제공하고 있다.

그림 1에서처럼 브라우저에서 사용자가 특정 요구사항을 Web 서버로 보내면 서버는 그 요구사항을 CGI 응용 프로그램을 실행시키고 처리된 결과를 받아서 사용자의 브라우저 화면에 출력하게 된다. 이 CGI는 서버에서 같은 방식을 제공하기 때문에 여러 플랫폼에서 실행할 수 있으며 URL을 직접 지정할 수 있

*본 연구는 '97년도 정보통신연구단의 대학기초연구지원사업과제
로 수행되었음

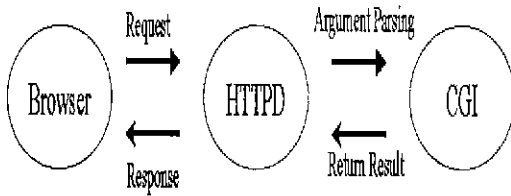


그림 1. CGI Mechanism

는 장점을 지니고 있으나 WWW 서버에서 CGI를 호출 할 때마다 하나의 독립적인 프로세서를 생성하기 때문에 사용자의 요구가 적을 시에는 문제가 되지 않으나 CGI 응용 프로그램이 빈번히 접속이 되는 경우에는 시스템 자원을 많이 차지한다는 단점을 지니고 있다[4].

3. WWW 시스템과 Database 시스템 통합

WWW 시스템과 Database 시스템간의 통합은 주로 Database 게이트웨이를 통하여 이루어지며 그 분류 방식[3]은 아래와 같다.

2.1. 서버 확장 방식

이 방법은 CGI를 이용한 게이트 구현 방법과 기존의 WWW 서버에 어플리케이션 서버를 결합한 방법으로 오라클에서 사용한 방식이다.

2.2. 클라이언트 확장 방식

이 방법에는 외부 프로그램을 사용하는 방법과, 데이터베이스 접속기능을 WWW 브라우저에 포함하는 브라우저 확장 방식이 있다.

이 연구에는 그림 2에서와 같이 서버 확장 방법 중 서버 사이트 확장 방식을 사용하였다.

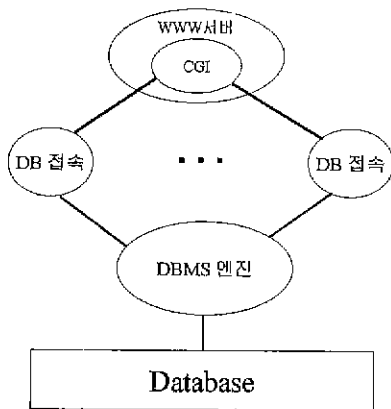


그림 2. 서버 사이트 확장 방식

이 방식은 기존 서버에 상관없이 단지 통로 (gateway)만을 개발함으로써 WWW 서버나 데이터베이스 시스템을 쉽게 통합할 수 있는 장점을 지니고 있으나 클라이언트가 증가할수록 CGI 응용프로그램이 데이터베이스에 접속하는 횟수가 증가함으로써 WWW 서버 성능이 현저하게 저하되는 단점을 지니고 있다.

4. Cache를 포함한 복제 서버 시스템

이 연구에서는 기존의 단일 서버에서 발생하는 CGI 응용 프로그램의 부하를 줄이기 위하여 제시한 복제 서버 시스템[1]에 추가적으로 데이터베이스 접속에서 발생하는 오버헤드를 줄이고자 cache 기법을 도입한 복제 서버 시스템 모델을 제시한다.

4.1 복제 서버 시스템에서 Cache 설계

제시한 복제 서버 시스템[1]에서 동일한 역할을 수행하는 각 복제 서버에서 CGI 응용 프로그램은 데이터베이스에 질의를 한 뒤 그 결과를 사용자의 화면에 보여주게 되는데 여기에 추가적으로 cache를 설정하여 자주 요구되는 질의를 HTML 형식으로 미리 저장하여 사용자 화면에 보여주게 되면 매 연결 때마다 데이터베이스 연결하는 시도 횟수가 현저하게 줄어들게 된다. 또한 복제 서버 시스템 작업량을 측정하여 가장 작업량이 적은 서버를 선택하는 조정자 (coordinator)에 부가적으로 cache의 정보를 기록하게 하였다. 표 1에서와 같이 조정자는 질의어에 대하여 각 서버에 해당되는 작업에 대한 큐의 길이를 기록하고 Cache 횟수를 기록하게 된다.

표1. 복제 서버에 대한 작업량과 cache 횟수

질의어	서버 A		서버 B		서버 C	
	작업량	캐쉬	작업량	캐쉬	작업량	캐쉬
heterophyd	02	17	05	20	1.2	5
malana	03	6	0.1	8	0.9	2

질의어가 특정 서버에서 cache counter 수가 높더라고 현재 작업량에 대한 큐의 길이가 다른 서버에 비해 상대적으로 높은 경우는 다른 서버를 선택하게 된다.

4.2 수행과정

Cache를 포함한 복제 서버의 수행과정은 이전에 제시한 복제 서버 시스템[1]과 유사하다.

- 1) 사용자는 WWW 서버가 제공하는 form 문장에 원하는 질의어를 입력하고 확인 버튼을 누른다.
- 2) WWW 서버에서 생성(fork)된 CGI 응용 프로그램

은 form parsing을 하고 cache 정보를 가지는 조정자를 호출한다.

3) 조정자에서는 CGI 응용프로그램으로부터 질의어를 받고 PVM(Parallel Virtual Machine)[5]으로 연결된 각 복제 서버에서 시스템 부하 정보(시스템 큐의 길이)를 수집하고 평가한다[6]. 조정자는 <hostname, queue_length> 함수 형태로 호스트 이름과 작업에 대한 큐의 길이를 받고 가장 작업 부하가 적은 복제 서버를 선택하고 캐쉬의 횡수를 증가시킨다. 이 때 캐쉬의 횡수가 늘더라도 큐의 길이가 상대적으로 짧게 되는 경우에는 다른 서버를 선택하게 된다.

4) 주 WWW 서버에서는 조정자로부터 호스트 이름을 받고 HTTP 프로토콜을 이용하여 선택된 복제 서버로 제어를 넘겨준다.

5) 선택된 복제 서버에서는 cache 내용을 살펴보고 그 내용이 cache에 있으면 데이터베이스에 접속하지 않고 직접 브라우저로 HTML 문서를 전송한다. 만일 Cache에 내용이 없다면 데이터베이스에 접속하여 그 결과를 저장하고 브라우저로 전송한다.

항은 다음과 같다.

- 주 WWW 서버와 복제 서버는 Apache 서버를 이용하였고 각 복제 서버는 Apache에서 제공되는 proxy 기능을 이용하였다.

- 각 복제 서버 cache의 내용은 WWW 서버의 로그를 분석하여 빈도수가 많은 것을 선택하도록 하였다.

5.결론 및 향후연구계획

이 논문에서는 단일 WWW 서버에서 발생하는 CGI 응용 프로그램의 기능을 축소화하고 프로그램 일부를 유틸 워크스테이션에 할당하여 기존의 CGI 응용 프로그램에서 발생하는 오버헤드를 줄이는 복제 서버 시스템[1]에 추가적으로 복제 서버 시스템에서 데이터베이스 시스템에 매번 요청에 따르는 접속 증가를 cache 기법을 도입함으로써 데이터베이스 연결 및 종료에서 발생하는 오버헤드를 줄일 수 있었다. 이를 통하여 사용자에게 정보 제공의 지속성 및 투명성과 서버에 대한 안전성 또한 증가하게 되었다.

현재 cache 기법을 도입한 복제 서버 시스템은 데이터베이스 시스템 내용이 부분 혹은 전체 내용이 변경되었을 때는 각 복제 서버에 있는 cache 내용이 전부 갱신되어야하는 단점을 지니고 있다. 이러한 단점을 보완하기 위해서는 독립적인 cache manager를 도입하여 데이터베이스 내용과 각 복제 서버에 있는 캐쉬 내용과 자료의 일치성을 유지시켜주는 정책이 필요하다. 또한 사전에 미리 설계된 정적 환경을 벗어나 동적 환경, 특히 이기종 환경이라면 PVM에서 임의의 시점에서 서버의 참여 및 해제에 대한 기능을 제공함으로써 부분 복제된 CGI 응용 프로그램 소스를 전송 받아 새로이 참여된 유틸 워크스테이션에서 수행할 수 있는 Run-time 환경에 대한 설계도 추가적으로 필요하다.

6.참고문헌

- [1] 진상혁,엄상용,최홍식,이광모, "서버 시스템 복제를 통한 WWW작업분배", 한국정보과학회 춘계 학술 발표회 논문집, 1998
- [2] NCSA, "The common gateway interface", tech report <http://hoochoo.ncs.uiuc.edu/doc-1.4/>, 1994.
- [3]김평철, "웹을 통한 데이터베이스 통로의 분류체계(1)", 제3회 WWW워크샵, 1996.
- [4] 배성민, "WWW Build, Refine and Management", 제 4 회 WWW Workshop강의 자료집, 1996.
- [5] Gest, A., Beguelin, A., Dongarra, J., Jiang W,Mancheck, R., and Sunderam, V., "PVM: Parallel Virtual Machine - A User Guide and Tutorial for Networked Parallel Computing", MIT press, 1994.
- [6] 엄상용, 지병준, 이광모, "이기종 분산 환경에서 부하 균형에 관한 연구," 한국정보과학회 춘계 학술발표회 논문집, 1997

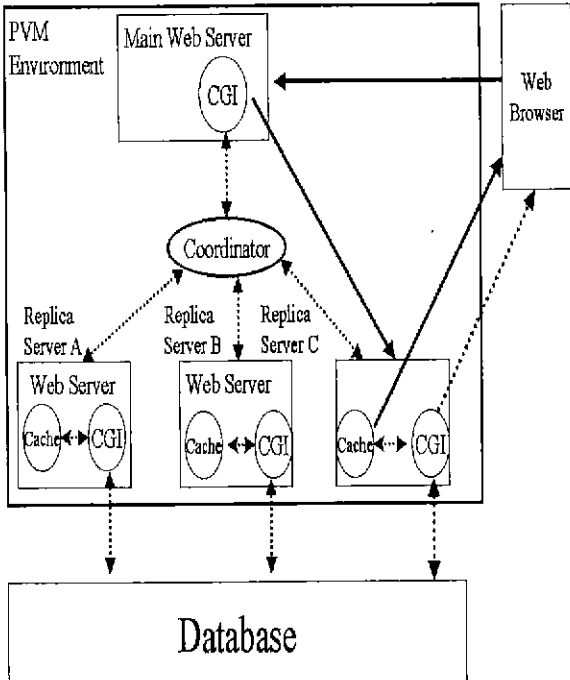


그림 3. Cache를 도입한 복제 서버 시스템

4.3 Cache를 포함한 복제 서버 시스템 환경

Cache를 포함한 복제 서버 시스템 환경은 이전의 제시한 서버 복제 시스템[1]과 동일하며 추가적인 사