

시구간별 사용자 접속 패턴을 이용한 동적 자원 재분배

이진성* 최창열* 박기진** 김성수*

아주대학교 정보통신전문대학원* 안양대학교 소프트웨어학과**
{cslee77,clchoi,sskim}@ajou.ac.kr* kiejin@aycc.anyang.ac.kr**

Dynamic Resource Reallocation using User Connection Pattern per Timeslot

Jinsung Yi* Changyeol Choi* Kiejin Park** Sungsoo Kim*
Graduate School of Information and Communication, Ajou University*
Dept. of Software, Anyang University**

요 약

웹 서버 클러스터의 성능 개선을 위한 연구가 다양한 분야에서 이루어졌지만 로그 파일 분석과 같은 방식으로 접속 빈도를 통한 실시간 동적 자원 재분배에 관한 연구에만 대부분 초점을 맞추었다. 본 논문에서는 시구간별 접속 패턴 분석 결과를 기반으로 패턴을 예측하여 자원을 동적으로 재분배하는 메커니즘을 제안한다. 제안한 메커니즘은 불필요한 자원 낭비를 감소시켜 효율적인 자원 재분배를 통해 클러스터의 성능을 향상시킨다. 또한 시구간별 접속 패턴의 유사성을 증명한다.

1. 서 론

인터넷의 급속한 성장과 사용자의 급격한 증가로 웹 상에서의 트래픽은 폭발적으로 늘어났고, 동시에 접속 하는 많은 사용자로 인하여 대다수 웹 서버가 과부하를 처리하지 못하는 경우가 발생하였다. 이러한 순간적인 접속 따른 많은 양의 트래픽에 웹 서버 성능이 크게 저하되는 문제는 데이터를 포함한 자원의 적절하지 못한 구성이나 관리 때문이다. 이에 클러스터 시스템은 최소의 비용으로 효과를 극대화할 수 있는 효율적인 대안이다.

또한 사용자의 접근 패턴은 시간이나 서비스에 따라서 일정한 방식으로 나타나게 된다. 예를 들면, 인터넷의 사용자는 주기적으로 메일을 확인하고 아침에 신문을 읽으며, 저녁 여가 시간에 인터넷 쇼핑을 하며, 밤에는 주로 영화를 본다. 물론 이러한 패턴이 정확하지 않더라도 여러 서비스를 제공하는 업체에서는 일정한 접속 패턴을 경험하게 된다. 기존의 동적 자원 재분배의 연구는 대용량 로그파일 분석에 따른 과부하를 야기하고 지속적인 자원 이주와 복사로 인하여 불필요한 자원 낭비를 발생시키는 단점이 있다[1].

따라서 본 논문에서는 시구간별 접속 패턴 분석 결과를 기반으로 패턴을 예측하여 효율적인 자원 재분배를 수행하는 메커니즘을 제안한다.

본 논문의 2장에서는 관련 연구를 언급하고, 3장에서는 웹 서버 클러스터 모델을 제시한다. 4장에서는 시구간별 패턴 기반에 따른 자원 재분배 시뮬레이션을 통하여 본

논문에서 제안한 방법의 성능 평가를 수행하며 5장에서 결론을 내린다.

2. 관련 연구

자원을 적절하게 배치하기 위한 노력은 로그파일의 분석으로 시작하여 적절한 자원 분배 메커니즘의 성공 여부에 달려있다.

2.1 로그 분석 기법에 관한 연구

대부분의 웹 접근 로그 파일은 CERN과 NCSA의 규약에 따라 HTTP 프로토콜의 부분으로 규정되는 일반적인 로그 형식을 따른다. 로그 파일에서 데이터를 구문 해석(Parsing)하여 데이터베이스에 저장하거나 혹은 테이블을 통하여 보여주는 Web Analyzer[2,3,4]와 같은 많은 도구들이 존재한다.

2.2 자원 재배치에 관한 연구

정확하고 효율적인 자원 분배의 중요성은 이미 많은 연구를 통하여 증명이 되었으며 적절한 알고리즘을 사용함으로써 자원의 적절한 재배치가 일어났을 때 좋은 성능을 얻을 수 있다[5,6].

재배치를 위한 방법에는 정적 복제 방법과 동적 복제 방법이 있다. 정적 복제 방법은 기록 파일을 사용하여 주기적으로 데이터를 분산시키는 반면, 동적 복제는 사용자의 요청이나 선호도에 따라 실제 서비스 시간에 동적으로 데이터를 분산시킨다. 이 두 가지 방법은 기록 파일을 사용하여 정확한 예측을 할 수 없다는 단점이

이 논문은 2003년도 두뇌한국21사업에 의하여 지원되었음.

있다. 사용자의 패턴이 계속 변한다고 지속적으로 자원의 재배치가 일어난다면 클러스터 시스템 내에서의 트래픽의 증가로 실질적인 성능은 저하될 것이다.

3. 웹 서버 클러스터 시스템 모델

본 연구 모델은 기존의 웹서버 클러스터 시스템이며, 부하 분배기(Load Balancer)를 이용한 시스템의 구조와 자원 재분배기(Resource Reallocator)를 이용하고 구성도는 그림 1과 같다.

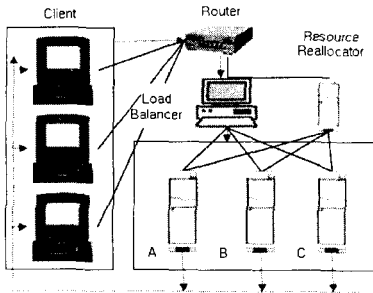


그림 1 웹 서버 클러스터 구성도

자원 재분배기는 부하 분배기와 실제 서버로부터 로그 파일을 가져와서 필요한 부분에 대한 구문을 해석한다. 해석된 정보는 데이터베이스에 저장되고 분석 후에 자원 재배치를 위한 정보로 사용된다.

4. 시공간별 패턴 기반 동적 자원 재분배

이 장에서는 실제 사이트의 시간대별 접속 빈도의 사례를 바탕으로 성능을 평가한다.

4.1 웹 사이트의 시공간별 접속 빈도 유사성

유사성은 표본으로부터 추출한 표본 상관 계수를 사용하며 가장 일반적으로 사용되는 것이 피어슨 상관 계수(Pearson Correlation Coefficient)이다. 일(day) 단위별 접속 빈도의 유사성 여부를 판단하기 위해 아래와 같은 식을 사용한다.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \quad (1)$$

데이터에 대한 빈도수는 다음과 같은 방법으로 구한다. 한 시간을 기준으로 i 번째 시간에 j 번째 자원이 접근되었다면 $R[i][j]$ 에 1을 증가시킨다.

$$\text{If (Access to } j^{\text{th}} \text{ data in time } i) \\ R[i][j] = R[i][j] + 1$$

i 와 j 는 관리자가 지정하는 시간주기로 디렉토리별 혹은 콘텐츠별 자원 단위로 정해질 수 있다. 그림 2는 노브레이크사(<http://www.cgiserver.net>)의 웹에 대한 시공간별 접속 빈도를 보여준다.

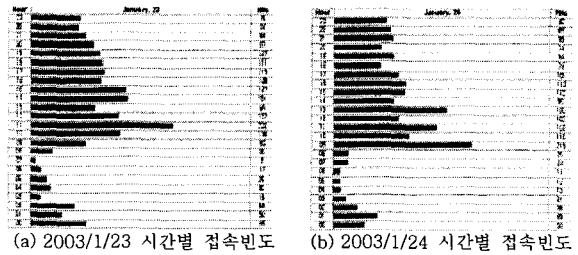


그림 2 <http://www.cgiserver.net> 의 홈페이지 접속 현황

그림 3은 그림 2의 자료를 기본으로 일(day)간 유사성 분석을 위한 그래프이다. 그림 3 (a)는 시공간별 접속 빈도를 그래프로 나타낸 것으로 각 시간마다 접속수도 비슷한 양상을 보여준다. 그림 3 (b)는 일정 시공간별 접속 빈도를 나타내며 하루를 8시간 단위로 시공간을 분리하였다.

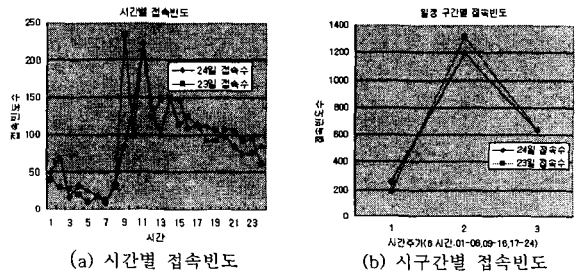


그림 3 시공간별 접속 빈도 그래프

또한 표 1은 그림 3 (b)를 정리한 자료로 유사성을 분석하기 위해 식 (1)에 대입해보면 r 값이 거의 1에 가까운 것을 볼 수 있다($r = 544163/544166 \approx 1.0$). 이는 일(day)간 접속 빈도의 차이가 거의 없다는 것을 의미한다. 따라서 일정 일자의 분석 결과를 가지고 다음날 접속 빈도에 대한 예측이 가능하다.

표 1 일(day)간 접속 빈도 예측을 위한 시공간별 접속빈도

시공간	평균	접속수	
		(2003/01/23)	(2003/01/24)
01-08시	218	252	184
09-16시	1259	1199	1319
17-00시	637	634	641

4.2 시스템 이용률 분석

일정한 접속 패턴을 보이는 환경에서 시스템의 성능은 이용률로 평가될 수 있다. 이용률은 총 가용 시간에 대한 실제 이용 시간의 비로 구할 수 있다. i, j 는 시간으로 $[i][j]$ 는 i 와 j 사이의 시간을 의미한다. 따라서 $Util[i][j]$ 는 i 와 j 시간 사이의 이용률을 나타내고 $TimeSlot[i][j]$ 는 i 와 j 시간 사이의 접속 수이다. $MaxToCon[i][j]$ 는 i 와 j 시간에서 가능한 최대 접속수이다. i 시간부터 j 시간까지 시간당 서버 이용률에 대한 계산 알고리즘은 다음과 같다.

```

Loop{
  IF(i < Accessed Time < j)
    Loop(k=m to n){
      TimeSlot[i][j] = TimeSlot[i][j] + R[i][k]
    }
}
Util[i][j] = TimeSlot[i][j]/MaxToCon[i][j]
    
```

예를 들어, 한 웹 서버에서 1시간 동안 처리 가능한 서비스의 수가 최대 2000명이라 가정하고 이용률을 살펴보면 아래와 같은 결과를 도출할 수 있다. 표 2는 8시간을 시구간으로 분할할 때 웹 서버에서의 이용률을 보여준다.

표 2 분할된 시구간의 서버 이용률

01 - 08시	218/2000 = 10.9%
09 - 16시	1259/2000 = 62.95%
17 - 00시	637/2000 = 31.85%

표 2에서 보듯이 실제 시스템에서의 이용률은 시간에 따라 편중되어 있는 양상과 함께 어제의 접속 빈도와 오늘의 접속 빈도가 비슷함에 따라 내일의 접속 빈도를 예측하기 위한 많은 정보를 제공한다. 하지만 이러한 평균을 기초로 다음을 예측하는 것은 시구간별 성향을 파악하기에는 유용하지만 많은 오차를 가진다. 따라서 그 시간대 가장 높은 접속을 보이는 값을 바탕으로 시구간을 곱하여 이용률을 구한다.

표 3 시구간 최대접속수에 따른 이용률

01 - 08시	544/2000 = 27.2%
09 - 16시	1872/2000 = 93.6%
17 - 00시	904/2000 = 45.2%

표 3은 23일과 24일중 시구간별로 가장 접속이 많았던 시간의 접속수에 시간차를 곱하여 이용률을 구한다. 그 시구간에서 가장 최대의 접속수를 기준으로 이에 오차율을 더하여 예측값을 설정한 경우이다.

4.3 자원 재분배에 따른 성능 분석

서비스를 제공하지 않는 서버 자원을 가용하게 만들어 성능을 향상시킬 수 있다. 이 절에서는 자원의 이용률과 사용되지 않는 자원들의 가용률을 바탕으로 성능을 분석한다. 가용률을 예측하기 위해서 시스템의 최대 이용률과 과거의 최대 이용률의 차를 가지고 이용 가능한 정도를 측정한다. MaxConPerTime은 특정 시구간에서 접속이 가장 많이 이루어졌을 때의 수로 다른 서비스를 위해 사용할 수 있는 가용률은 아래와 같다(가정 오차, ε: 0.2).

$$(1 - (j-i) * MaxConPerTime / MaxToCon[i][j]) + \epsilon$$

표 4는 표 3에서의 수치를 바탕으로 20%의 오차를 가정할 때 서버 이용률을 보여준다. 분석 결과를 바탕으로 자원 재구성을 하면 다른 서비스를 위해 평균적으로 하루당 평균 37.7%의 자원을 추가적으로

제공할 수 있다. 9시부터 16시까지는 서버 전체를 가동시켜야 하지만 해당 시구간 외에는 서버 자원의 45% 이상을 다른 서비스를 위해 할당할 수 있다. 20% 오차를 가정한 이용률이 다음날 재구성을 위한 설정값으로 사용될 수 있다. 오차를 가정하지 않은 경우 평균 44.67%의 서버가 서비스를 하지 않은 채 돌아 가고 있다.

표 4 오차를 가정한 이용률

01 - 08시	32.64%	67.36%
09 - 16시	100%	0%
17 - 00시	54.24%	45.76%

따라서 가장 서비스가 많이 요구되는 시구간에는 모든 서버 자원이 가용해야 하지만 다른 시구간에는 불필요하지 않은 서버를 다른 서비스 제공을 위해 사용하므로 전체 시스템 성능을 향상시킬 수 있다.

5. 결론 및 향후 방향

실시간 동적 복제를 하는 경우 복제로 인한 I/O비용, 네트워크 대역폭 소비, 시스템 자원 소모로 인해서 서비스의 증가에 따라 비용이 비례적으로 증가한다. 따라서 이를 해결하기 위해 본 연구에서는 시구간별 사용자 요청 패턴 정보를 바탕으로 부하가 가장 적은 시간대에 자원을 재분배 하여 이용률을 높이는 방안을 제안하였다. 또한 수요를 예측하고 부하가 가장 적은 시간대에 자원 복제, 이주가 일어나기 때문에 복제로 인한 과부하를 줄일 수 있다. 향후 과제로 예측을 불가능하게 하는 악의적 패턴 조작을 처리하기 위한 연구를 수행할 예정이다.

6. 참고 문헌

- [1] V. Kalogeraki, P.M Melliar-Smith, and L. Moser, "Dynamic Migration Algorithms for Distributed Object Systems," 21st International Conference on Distributed Computing Systems, pp. 119 -126, Apr. 2001.
- [2] J.H. Andrews, "Testing using Log File Analysis: Tools, Methods, and Issues," Proceedings of 13th IEEE International Conference on Automated Software Engineering, pp. 157-166, Oct. 1998.
- [3] F. Murtagh and F. Tao, "Towards Knowledge Discovery from WWW Log Data," International Conference on Information Technology: Coding and Computing, pp. 302-307, Mar. 2000.
- [4] M. Bamshad, D. Honghua, and N.T.L. Miki, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization," Data Mining and Knowledge Discovery, Vol. 6, No. 1, pp. 61-82, Jan. 2002.
- [5] Q. Li and B. Moon, "Distributed Cooperative Apache Web Server," Proceedings of the 10th World Wide Web Conference, pp. 555-564, May 2001.
- [6] J. Huang and R. Jhr, "RT-ARM: A Real-Time Adaptive Resource Management System for Distributed Mission-Critical Applications," IEEE Workshop on Middleware for Distributed Real-Time Systems and Services, Dec. 1997.