한국컴퓨터정보학회 동계학술대회 논문집 제20권 제1호 (2012. 1)

클라우드 컴퓨팅 환경에서 가상머신의 효율적인 자원 감축 기법

김창현*, 이원주°, 전창호*
*한양대학교 컴퓨터공학과
°인하공업전문대학 컴퓨터정보과
e-mail: ctcquatre@hanyang.ac.kr, wonjoo2@inhatc.ac.kr, chj5193@hanyang.ac.kr

An Efficient Reduction Scheme of Virtual Machine Resource in Cloud Computing Environment

Chang Hyeon Kim*, Won Joo Lee°, Chang Ho Jeon*
*Dept. of Computer Science & Engineering, Hanyang University,
°Dept. of Computer Science, Inha Technical College

● 요 약 ●

본 논문에서는 가상머신에서 수행되는 어플리케이션의 서비스 품질을 보장하고, 클라우드 클러스터의 운영비용을 절감시킬 수 있는 자원 할당 감축기법을 제안한다. 이 기법은 가상머신의 자원 사용량 변화 추세를 분석하고 이를 토대로 확률적인 접근을 사용하여 새로운 자원 할당 감축량을 결정한다. 가상머신의 자원 사용량이 할당량을 초과하면 가상머신의 이주가 필요하다. 이때 발생하는 다운타임동안 가상머신의 어플리케이션은 서비스를 수행할 수 없기 때문에 클라우드 시스템의 성능이 저하된다. 따라서 성능평가에서는 가상머신의 자원 사용량이 할당량을 초과하는 횟수를 측정하여 Greedy 기법과 비교 평가함으로써 제안한 기법이 자원 할당 감축에 우수함을 검증한다.

키워드: 클라우드 컴퓨팅(cloud computing), 가상머신(virtual machine), 이주(migration)

I. 서론

클라우드 컴퓨팅의 기술적 기반에는 자원 가상화 기술이 핵심 요소이다. 가상화란 물리적 자원을 가상적인 논리 자원으로 만드는 것을 의미하며, 프로세서, 메모리와 같은 물리 자원들을 가상화하고, 가상화된 자원들을 조합하여 가상머신을 구성하는 것이다.

가상머신은 클라우드 컴퓨팅 서비스를 제공하는 핵심기능을 수행하는 요소이다. 또한 가상머신에 주어진 자원 용량은 어플리케이션 서비스의 성패와 클라우드 운영 효율에 영향을 미친다. 따라서 클라우드 컴퓨팅의 서비스 품질을 보장하고, 운영 비용을 절감하기 위해서는 가상머신에 적절한 자원을 할당하는 것이 필수적이다.

II. 관련 연구

가상화 기술을 통하여 단일 물리머신의 자원으로 다수의 가상머신을 구성할 수 있다. 이러한 가상화 기술은 집체적 물리자원의 활용률을 높여서 클라우드 운영비용을 절감하는데 크게 기여하고 있다. 클라우드 컴퓨팅에서는 가상머신의 자원을 효율적으로 관리하는 것이 중대한 과제 중의 하나이다[1]. 이러한 관점에서 국외에서는 가상머신의 자원 할당과 관리에 관한 연구가 활발하게 진행되고 있다.

T. Wood[2]은 최대 자원사용량에 근거하여 가상머신의 할당량을 산정하고 그 만큼의 자원을 제공할 수 있는 물리머신을 찾는 방법을 제시하였다. N. Bobroff [3]은 예측모델을 사용하여 가상머신의 할당량을 계산함으로써 최소의 물리머신으로 필요한 가상머신을 배치하였다.

III. 가상머신의 자원 감축 기법

본 논문에서는 가상머신에서 운영되는 어플리케이션의 서비스 품질을 보장하고 클라우드 클러스터의 운영비용을 절감시킬 수 있는 가상머신의 자원 할당 감축기법을 제안한다. 이 기법은 가상머신의 자원사용량의 변화 추세를 계산하여 추세의 유용성을 판단하고 확률적 접근에 기반하여 자원 감축량을 결정한다. 사용량 변화 추세는 가상머신으로부터 모니터링 된 자원 사용량을 바탕으로 선형회귀분석을 수행하여 계산할 수 있다. a^{ψ}와 b^{ψ}가 k개의 타임 슬롯동안에 대하여 계산된 자원 사용량 변화 추세라고 할 때 특정 τ시점의 자원 사용량은 다음과 같다.

$$e^{\psi} = a^{\psi}\tau + b^{\psi}$$

예측의 정확성을 향상시키기 위하여 가상머신에 할당될 수 있는 자원 최대량을 여러 개의 구간으로 나누고 각 구간의 a^{ψ}와

b^{ψ}을 a_i^{ψ}와 b_i^{ψ}을 표기한다면 i구간에 대한 자원 사용량과 예측 오차는 다음과 같이 구할 수 있다. u는 마지막에 측정된 자원 사용량이다.

$$e^{\psi} = a_i^{\psi}\tau + u + \epsilon$$
$$\epsilon \approx N(\mu_i, \sigma_i)$$

새로운 자원 할당량r이 주어질 때 가상 머신이 r을 초과하여 자원을 사용하는 경우는 다음 조건을 만족할 때 이다.

$$\epsilon > r - a_i^{\psi}\tau + u$$

따라서 r이 주어질 때 $P(r - a_j^{\psi}\tau + u < \epsilon \leq \infty)$는 자원 사용량을 초과할 확률이다. 이를 바탕으로 초과확률 허용치 δ가 주어지면 $P(r - a_j^{\psi}\tau + u < \epsilon \leq \infty) < \delta$을 만족하는 r을 찾음으로써 가상머신의 새로운 할당량을 결정할 수 있으며 이전 할당량과 새로운 할당량을 고려하여 감축량을 결정할 수 있다.

Ⅳ. 성능 평가

제안한 자원 할당 감축기법의 성능평가를 위하여 가상머신의 자원 사용량이 할당량을 초과(overflow)하는 횟수를 측정한다. 가상머신의 자원사용량이 할당량을 초과하면 가상머신의 이주가 필요하다. 이때 발생하는 다운타임동안 가상머신의 어플리케이션은 서비스를 수행할 수 없다. 따라서 할당량 초과 횟수는 클라우드의 서비스 품질을 측정하는 척도로 사용할 수 있다. 제안한 기법(δ=0.3, 0.7)의 성능은 기존의 Greedy 기법과 비교한다. Greedy기법은 할당시점의 자원사용량을 조사하여 운영에 필요한 최소의 자원량을 할당하는 기법이다. 성능평가는 100개의 가상머신과 100의 물리머신으로 수행한다.

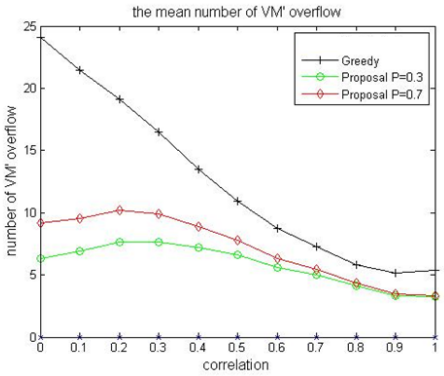

그림 1. 자원 할당량 초과 횟수

그림 1에서 correlation은 가상머신에서 운영되는 어플리케이션의 자원 사용 불규칙 정도이며 0에 가까울수록 자원 사용량 변화가 불규칙한 것을 의미한다. 불규칙 정도에 관계없이 Greedy 기법에 비해 제안한 기법이 우수함을 알 수 있다. 특히 correlation이 0.4이하에서는 제안한 기법과 Greedy 기법의 성능 차가 더욱 커지는 것을 볼 수 있다.

Ⅴ. 결론

본 논문에서는 가상머신에서 수행되는 어플리케이션의 서비스 품질을 보장하면서 클라우드 클러스터의 운영비용을 절감할 수 있는 자원 할당량 감축기법을 제안하였다. 제안한 기법은 자원 사용량을 예측하고 확률적 접근을 통하여 적절한 자원 할당량을 결정한다. 성능평가를 통하여 제안한 기법이 감축시점의 자원 사용량만을 고려하는 기법보다 감축량을 효과적으로 결정하였음을 확인하였다.

참고문헌

[1] Y.C. Lee, A.Y. Zomaya, “Energy efficient utilization of resources in cloud computing systems,” Journal of Supercomputing, Vol. 53, pp. 1-13, 2010

[2] T. Wood, P. Shenoy, A. Venkataramani, M. Yousif, “Sandpiper: Black-box and gray-box resource management for virtual mahcine,” Computer Networks, Vol. 53, pp. 2923-2938, 2009

[3] N. Bobroff, A. Kochut, and K. Beaty, “Dynamic placement of virtual machines for managing SLA violations. Integrated Network Management,” in proc. of 10th IFIP/IEEE International Symposium, pp. 119-128, May 2007