

클라우드 환경의 교통정보 서비스를 위한 조건부 확률 추론을 이용한 가상 머신 프로비저닝 스케줄링

김재권¹ · 이종식^{1†}

Virtual Machine Provisioning Scheduling with Conditional Probability Inference for Transport Information Service in Cloud Environment

Jackwon Kim · Jong-Sik Lee

ABSTRACT

There is a growing tendency toward a vehicle demand and a utilization of traffic information systems. Due to various kinds of traffic information systems and increasing of communication data, the traffic information service requires a very high IT infrastructure. A cloud computing environment is an essential approach for reducing a IT infrastructure cost. And the traffic information service needs a provisioning scheduling method for managing a resource. So we propose a provisioning scheduling with conditional probability inference (PSCPI) for the traffic information service on cloud environment. PSCPI uses a naive bayse inference technique based on a status of a virtual machine. And PSCPI allocates a job to the virtual machines on the basis of an availability of each virtual machine. Naive bayse based PSCPI provides a high throughput and an high availability of virtual machines for real-time traffic information services.

Key words : Cloud Computing, Naive Bayse, Transport Information Services, Virtual Machine Provisioning, PSCPI

요약

전 세계적으로 자동차의 수요와 교통정보 서비스의 활용도가 높아지고 있다. 따라서 교통정보 서비스의 종류와 데이터의 양의 증가로 인해 많은 IT 자원 인프라가 필요하다. 인프라의 감소를 위해 클라우드 컴퓨팅이 주목을 받고 있으며, 자원관리를 위해 프로비저닝 스케줄링 기법이 필요하다. 본 논문에서는 클라우드 환경에서 교통정보 서비스를 위한 조건부 확률 추론을 활용한 프로비저닝 스케줄링(PSCPI: Provisioning Scheduling with Conditional Probability Inference)을 제안한다. PSCPI는 가상머신의 상태에 따라 나이브 베이즈 추론 기법을 사용하여 가상머신의 가용율에 따라 작업 할당을 할 수 있다. 나이브 베이즈 기반의 조건부 확률 추론 프로비저닝 스케줄링을 활용하여 교통정보 서비스에 높은 처리율과 활용율을 보인다.

주요어 : 클라우드 컴퓨팅, 나이브 베이즈, 교통정보 서비스, 가상머신 프로비저닝, PSCPI

1. 서론

세계적으로 자동차의 수요가 늘어나면서 교통정보 서비스의 활용도가 높아지고 있는 가운데 교통 환경 문제와 사고, 정체 등의 문제를 해결하기 위해 최근에는 IT시스템

을 접목한 첨단 교통체계(ITS: Intelligent Transport System)와 같은 지능형 교통 체계 구축을 추진하고 있다(김장욱 외 2008). 교통은 실생활에서의 삶과 매우 밀접한 관련이 있으며 정보 통신기술을 이용하여 사용자들은 다양한 실시간 교통 정보 서비스를 이용하고 있다(최훈 외, 2010).

서비스 제공자는 사용자수의 증가와 교통과 관련된 서비스가 많아짐에 따라 교통관련 서버에 많은 스토리지와 트래픽을 요구하게 되며 이를 해결하기 위해서는 많은 비용이 필요하다(장용구 외 2005). 따라서 점차 늘어나는 교통 서비스에 따라 유지보수의 비용이 증가하게 되어 자원 인프라의 중요성이 높아진다. 인프라 비용을 감소하며

*이 논문은 인하대학교의 지원에 의하여 연구되었음.
접수일(2011년 10월 26일), 심사일(1차 : 2011년 11월 22일),
게재 확정일(2011년 11월 30일)

¹⁾ 인하대학교 정보공학과

주 저 자 : 김재권

교신저자 : 이종식

E-mail: jslee@inha.ac.kr

QoS(Quality of Service)를 높이기 위한 기술로 교통 정보와 관련된 서비스에서 클라우드 컴퓨팅이 주목을 받고 있다(이승관 외, 2011).

클라우드 컴퓨팅은 가상의 환경에서 적절한 리소스를 제공하는 차세대 컴퓨팅으로서 대규모의 데이터 센터내의 자원 인프라의 효율성을 높일 수 있으며 QoS와 더불어 차세대 IT 서비스 제공이 가능하다(Daniel 외 2009).

클라우드 컴퓨팅 환경의 교통 정보 서비스에서 가용성 최적화를 위한 가상머신의 프로비저닝 스케줄링이 필요하며(장은영 외 2011), 교통정보 서비스의 원활한 제공을 위해서는 특정 조건에 따라 서비스를 제공할 수 있는 조건부 기반의 프로비저닝 스케줄링의 메커니즘의 새로운 방법과 정책을 정의하여 클라우드를 설계해야한다. 프로비저닝은 요청 서비스에 대해 빠르게 가상머신을 동적으로 할하는 방법으로서(Rodrigo 외 2011) CPU, RAM, Network 등의 가상머신의 자원 정보에 따라 작업을 처리할 가상머신을 선택하며, 이를 효율적으로 하기 위해서 조건부 확률 기법인 나이브 베이지안 기법이 필요하다(John, 1985). 나이브 베이지안은 불충분한 정보를 가진 환경을 표현하고 추론하는 대표적인 기법이며, 조건부 확률을 통한 의사결정에 많은 장점이 있다.

본 논문에서는 클라우드 환경의 교통정보 서비스를 위해 조건부 확률 추론을 이용한 가상머신 프로비저닝 스케줄링(PSCPI: Provisioning Scheduling with Conditional probability Inference)을 제안하며, 나이브 베이지안 추론 기법기반 프로비저닝 스케줄링의 새로운 방법과 구조에 대해 기술 한다. 제안하는 PSCPI는 클라우드 자원 리소스의 상태에 따라 나이브 베이지안(Naive Bayesian)추론 기법 기반 스케줄링(John, 1985)을 사용하여 가상머신의 가용율을 판단한 후 가용 율이 높은 가상머신에게 작업을 할당한다. 나이브 베이지안 추론 기법기반 프로비저닝 스케줄링은 클라우드 환경을 구성하는 각각의 가상머신의 작업 처리 개수, 성능, 대기시간 등의 데이터를 수집하여 가용율을 판단한다. 이에 가상머신 프로비저닝에서는 각각의 가상머신의 가용 율에 적합한 작업을 할당을 한다.

본 논문의 구성은 제 2장에서는 관련연구로서 클라우드 기반의 교통 정보 서비스와 자원관리 스케줄링에 대해 기술한다. 제 3장은 클라우드 환경의 교통정보 서비스를 위한 조건부 확률 추론을 이용한 가상머신 프로비저닝 스케줄링을 기술한다. 제 4장은 제안하는 기법의 실험 및 결과를 기술하며, 제 5장은 결론에 대해 기술한다.

2. 관련연구

2.1 클라우드 환경의 교통정보 서비스

클라우드 컴퓨팅은 서로다른 물리적 위치에 존재하는 자원들을 통합하는 가상화 기술과 자원을 배분하고 처리하는 분산 컴퓨팅 처리능력인 IaaS(Infrastructure as a Service)와 운영체제등의 응용 프로그램 집합체인 Paas(Platform as a Service)으로 구성된다. 또한 어플리케이션이나 인터넷을 통한 서비스 제공을 하는 SaaS(Software as a Service) 및 자원 사용량에 따른 과금 체계인 유틸리티 컴퓨팅 모델 등 최근의 기술동향과 많은 연관성이 있다(민영수 외, 2009).

교통정보 서비스는 유/무선망 및 개별 차량에서 수집된 데이터를 이용하여 기관, 기업, 차량 네비게이션 등에 교통 정보를 전달하여 미래의 U-City환경을 위한 중요한 비즈니스 모델로 연구가 진행 중에 있다(최훈 외 2010).

교통정보 서비스에서 클라우드 컴퓨팅의 접목은 서비스 제공자가 수행하는 테스트 비용을 줄일 수 있으며, 서버 자원 통합 구축을 통한 데이터 센터의 체계화, 유지비용의 절감 그리고 이를 통한 새로운 도로정보를 효율적으로 갱신이 가능하다.

이에 본 논문에서는 클라우드 기반의 교통 정보 서비스의 구조에 대해 설명을 하며 다수의 교통정보 서비스를 제공받는 사용자의 서비스 요청에 따른 클라우드 컴퓨팅 환경의 자원관리 스케줄링에 대해 논의한다.

2.2 자원관리 스케줄링

클라우드 기반 서비스를 배포하기 전에 서비스의 성능과 제공자의 수익을 예측하고 설계함으로써 자원 인프라의 접근 비용을 최적화하여 이익을 향상시킬 수 있어야 한다. 따라서 클라우드 서비스를 위해 이론적이나 막연한 측정이 아닌 시뮬레이션을 통해 자원 관리 상황을 예측하는 것이 필요하다. 클라우드 자원 관리를 위해서는 자원 할당이 필요하며, 자원 할당기법에 대한 연구가 진행 중이다. 첫 번째로, 라운드 로빈 토너먼트 스케줄링(Scheduling Round Robin Tournaments : SRRT) (Rasmus외, 2008)은 작업의 실행 순서 등을 결정하는 방법으로 CPU 스케줄링, 가상머신 스케줄링 등 여러 자원들의 선택이나 공유 문제에 많이 이용되고 있다. 두 번째로, 실행시간과 대기시간에 따른 스케줄링 기법인 위한 반응시간 최적화된 자원 관리 알고리즘(Response time Optimization-based Resource Management algorithm: RORM) (Assuncao 외 2009)은 큐잉이론을 바탕으로, 각 자원이 할당되는 시점과 이전의 자원 처리 시간을 비교하여 작업 처리시간이 가장

빠른 자원을 선택하는 기법으로, 짧은 대기시간과 활용율을 높였다. 본 논문에서는 클라우드 환경의 작업 스케줄링을 위해서 나이브 베이저안 추론 기법(Naive Bayesian Inference Method) 기반 스케줄링(John, 1985)을 가상머신 프로비저닝에 도입하기 위해 새로운 방법과 매커니즘을 제시한다.

3. 클라우드 환경의 교통정보 서비스를 위한 조건부 확률 추론을 이용한 가상머신 프로비저닝 스케줄링

3.1 클라우드 환경의 교통정보 서비스

클라우드 환경의 교통정보 서비스를 위한 조건부 확률 추론을 이용한 가상머신 프로비저닝 스케줄링(PSCPI: Provisioning Scheduling with Conditional probability Inference)을 설명하기에 앞서, 클라우드 환경의 교통정보 서비스를 정의하여 자원관리 필요성을 제시한다.

클라우드 환경에서 교통정보 서비스를 제공하는 방법은 그림 1과 같이 사용자 측면, 클라우드 시스템 측면, 응용 서비스 측면으로 나눌 수 있다.

사용자 측면에서는 교통 정보를 이용하는 사용자들이 서비스를 요청하고 제공받는 부분이다. 기존의 차량의 내비게이션이나 웹을 통해 서비스를 요청하고 제공받는 단일 교통 정보 서비스가 아닌 모바일이나 태블릿 PC를 통해서 서비스를 제공받을 수 있다.

클라우드 시스템 측면은 사용자로부터 요청되는 서비

스를 제공해 주기위한 측면으로서 가상머신, 분산처리 시스템, 분산 파일시스템, 분산 데이터베이스, 클러스터 등의 인프라 구성 요소와 가상 머신 안에 컴퓨팅 자원들이 포함되어있다(김창수 외, 2009). 이를 통해 사용자에게 원하는 서비스를 제공할 수 있게 되며 다수의 사용자의 접속을 허용한다.

응용 서비스 측면은 사용자에게 제공할 수 있는 여러 개의 프로그램들을 의미하며 기존의 GIS, GPS 등과 외부 서비스 제공업체에서 제공하는 서비스 들을 포함 시킬 수 있다. 이로서 복수개의 서비스 프로그램을 사용자가 이용할 수 있게 된다.

교통 정보 서비스의 특징은 불특정된 시간대에 사용자가 많은 서비스 요청을 하게 되며 클라우드 환경을 구성하는 각 자원의 리소스를 활용하기 위해서는 다수의 서비스 요청에 따르는 자원관리 방법이 필요하다. 교통 정보 서비스는 맵 다운로드 등 특정 서비스 이외에는 많은 컴퓨팅 처리를 요구하지 않지만 다수의 사용자의 접속으로 인해 사용자 접속과 자원 상황의 조건에 맞는 스케줄링 기법이 요구된다.

3.2 클라우드 환경의 프로비저닝 스케줄링

프로비저닝 스케줄링을 위해서는 클라우드 환경과 서비스에 따르는 정책과 매커니즘의 정의가 필요하다(Rodrigo 외 2011). 이에 본 논문에서는 자원 관리 시스템이 요구되며 사용자 접속수와 자원 상황의 조건에 맞는 자원관리 스케줄링을 위해 조건부 확률 추론 방법인 나이브 베이저

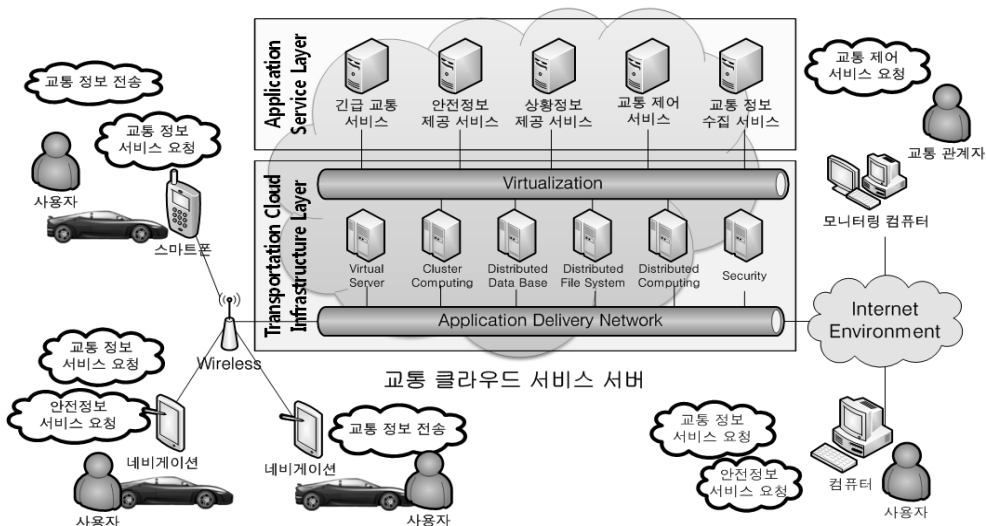


그림 1. 클라우드 환경에서 교통정보 서비스

안을 활용한다. 제안하는 나이브 베이지안 기반의 프로비저닝 스케줄링의 구조는 그림 2와 같다.

프로비저닝 스케줄링 구조는 SaaS, PaaS, IaaS로 구분된다. 첫 번째로, SaaS는 사용자가 서비스를 제공받기 위한 부분으로, 사용자는 작업을 요청하면 Admission Service로부터 인증을 받은 후 서비스의 이용이 가능하며, 인증이 완료되면 요청한 서비스에 대해 Generate Job에서 서비스를 제공하기 위한 작업을 생성한다. 두 번째로, PaaS는 SaaS로부터 작업요청을 받으면 해당 작업에 대한 배분을 결정한다. Workload Prediction은 작업을 요청받은 상태에서 요청 서비스의 종류에 대해 작업 계산한다. 계산이 완료되면 Application Provisioning은 서비스 종류에 따라서 해당 서비스의 내용이 저장되어 있는 IaaS내의 Data Center를 선택하는 부분이다. 마지막으로, IaaS는 요청받은 서비스에 대해 작업을 처리하는 부분으로서 해당 서비스의 Data Center에서 가상머신을 결정하기 위해 Virtual Machine Provisioning은 Naive Bayesian Method을 이용하여 가상머신을 할당한다. 또한 Update Current VM State로부터 작업 처리 상태에 따라 Application Provisioning과 VM

Provisioning의 추론에 활용된다. IaaS의 작업처리는 큐잉모델 기반이며, 작업 처리 구조는 그림 3과 같다.

큐잉모델 기반 작업처리 구조는 가상 머신 안에 작업을 처리하는 Physical Computing이 있으며, 이를 처리하기 전에 작업들은 큐에서 대기하며 FCFS방식이다. 첫 번째로 Application Provisioning은 서비스에 따르는 데이터센터를 정하게 되며, 패키지 도착시간, 서비스시간, 서버의 종류가 다르기 때문에 $m/m/\infty$ 이다. VM Provisioning은 1대의 서버를 사용하기 때문에 $m/m/1$ 로 구성이 된다. 또한 작업이 끝난 이후에는 Update Current VM Status로부터 현재의 가상머신의 상태에 대한 내용을 각 프로비저닝 모듈에 전송 한다.

3.3 나이브 베이지안 기반 프로비저닝 스케줄링

베이지안 기반 조건부 확률 추론은 불충분한 정보를 가진 환경을 표현하고 추론하는 대표적인 기법이며, 조건부 확률을 통한 의사결정에 많은 장점이 있다.

작업요청이 들어오면 가상머신 자원의 리소스 상태에 따라 가상머신을 선택하기 위해서 CPU 활용률, 주기억장치 활용률, 보조기억장치 활용률, 평균작업 처리 시간, 큐의 길이, 대기시간을 판단한다.

물리적 컴퓨터들의 정적 상태를 정의하여 가상머신을

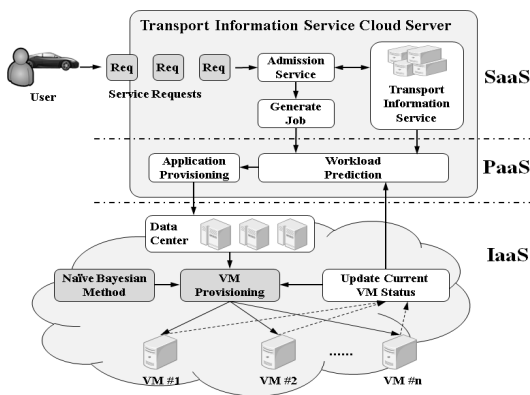


그림 2. 프로비저닝 스케줄링 구조

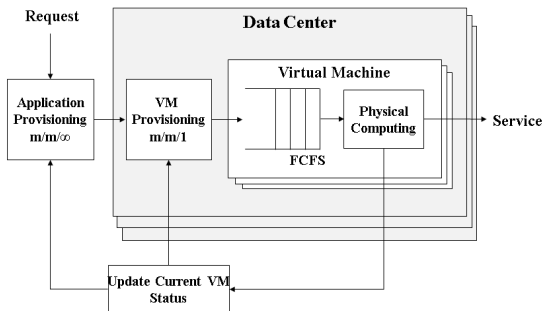


그림 3. 큐잉모델 기반 작업처리 구조

표 1. 물리적 자원의 정적 속성

Property Name	Description	Value Type	Unit
Physical ID	자원 식별자	Integer	1 ~ r
CPU	CPU 최대 클럭 속도	Float	Giga Hertz
RAM	주 기억장치 크기	Float	Mega Byte
Storage	보조기억장치 크기	Float	Giga Byte

표 2. 가상머신 사양 정적 속성

Property Name	Description	Value Type	Unit
VM ID	가상머신 식별자	Integer	1 ~ n
P_Num.	물리적 자원의 대수	Integer	1 ~ r
CPU Speed	가상머신 CPU 클럭속도	Float	Giga Hertz
RAM Size	가상머신 주 기억장치 크기	Float	Mega Byte
Storage Size	가상머신 보조기억 장치 크기	Float	Giga Byte
Rank (R)	가상머신의 순위	Integer	1 ~ n

표 3. 가상머신 동적 매개변수

Parameter	Description	Value Range	note
H_1	CPU Utilization	0 ~ 100	입력 변수1
H_2	RAM Utilization	0 ~ 100	입력 변수2
H_3	Storage Utilization	0 ~ 100	입력 변수3
H_4	Average Job Processing Time	0~Average Job Processing Time	입력 변수4
H_5	Queue Size	0~Queue Size	입력 변수5
H_6	Waiting Time	0~Estimated Time	입력 변수6
E	VM Utilization	0 ~ 1	출력 변수
Y	VM Utilization(70%), if($VM_n \leq 70\%$) True, else False	True, False	사후확률 변수

구성하기 위해 표 1에서는 물리적 컴퓨터 자원의 정적 속성을 정의하며, 표 2에서는 물리적 자원들로 구성된 가상 머신의 정적 속성을 정의한다.

위와 같이 물리적 자원의 정적 속성을 통한 가상머신을 구성하는 정적 속성을 구성 하였으며, n은 자원의 개수를 의미하며, P_Num.은 가상머신을 구성하는 물리적 자원의 개수(r)를 의미한다. Rank는 각 가상머신의 우선순위를 의미하며, 사후 확률을 계산한다. Rank는 각 가상머신의 성능별 오름차순으로 정의한다.

리소스 요청으로부터 가상머신의 선택을 추론을 하기 위한 동적 매개변수는 표 3과 같다.

$H_1 \sim H_6$ 는 입력 변수로서, 이를 통해 가상머신 활용율인 E를 추론하게 된다. Y는 군집을 위한 변수로서 사후확률을 계산하기 위해 E가 상위 30%일 경우 R을 True으로 표시하며, 그렇지 않을 경우 False로 표시한다. 이는 추론시간을 줄이기 위해 미리 활용성이 높은 가상머신을 결정한다.

각 가상머신의 정적속성과 추론을 위한 매개변수를 정의하였으며, 작업 요청부터 가상머신의 선택까지의 프로비저닝 스케줄링에 대한 순서도는 그림 4와 같다.

VM Provisioning으로부터 VM를 요청하게 되면 각 가상머신의 리소스의 현재 상태를 VM Status DB로 전송을 하게 되며, 매개 변수에 따라 활용 율이 높은 자원과 낮은 자원을 군집화 하게 된다. 가상머신들의 현재 상태와 최

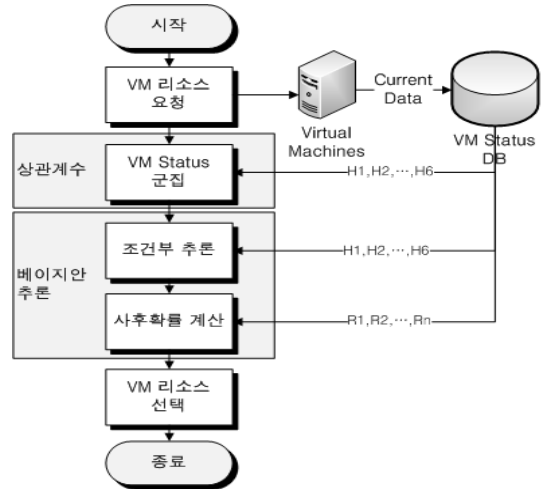


그림 4. 프로비저닝 스케줄링 순서도

적의 상태의 비교의 합을 피어슨 상관계수를 이용하여 식 1(강금식 외 2010)과 같이 각 매개변수의 합을 이용하여 가상머신들의 상위 70% 이상의 값을 = True로 반환한다.

$$VM_n = \frac{\sum_{x=1}^6 (H_{n,x} - \bar{H}_n)(O_{n,x} - \bar{O}_n)}{\sqrt{\sum_{x=1}^6 (H_{n,x} - \bar{H}_n)^2 \sum_{x=1}^6 (O_{n,x} - \bar{O}_n)^2}} \quad (1)$$

$$IF \left(\frac{VM_n}{\sum_{n=1}^{10} VM_n} * 100 \leq 70 \right) THEN Y(VM_n) = True$$

$$ELSE Y(VM_n) = False$$

VM_n 은 가상머신의 n번째 자원을 의미하며, $H_{n,x}$ 는 가상머신 n의 x번째 동적 매개변수를 의미한다. 여기서 H는 6개의 가상머신 동적 매개변수($H_1 \sim H_6$)를 의미한다. \bar{H}_n 은 동적 매개변수들의 평균값을 의미한다. $O_{n,x}$ 는 H의 매개변수에 대해 최적의 값을 의미하며 최적의 값은 해당 매개변수의 가장 가용 율이 높은 수를 의미한다(예를 들어 $H_1=0$). \bar{O}_n 은 최적의 매개변수에 대한 평균을 의미한다. 이를 통해 각각의 가상머신의 상관계수 합을 구하게 되며, $Y(VM_n)$ 값은 모든 가상머신의 상관계수의 합에서 VM_n 이 70%이상인 것은 True로 반환하고 그렇지 않은 것은 False로 반환한다.

가상머신들의 현재 상태의 군집화가 완료가 되면, True의 값으로 반환된 가상머신을 조건부 확률을 계산한다. 나이브 베이저의 조건부 확률 공식은 식 2(John, 1985)와 같으며, 속성집합 $H=\{H_1, H_2, \dots, H_6\}$ 는 앞서 정의한 동적

매개변수 6개의 속성으로 구성된다. 매개변수를 활용하여 가상머신의 활용 율 E 를 구하기 위한 공식은 식 2(John, 1985)와 같다.

$$P(H|E) = \prod_{i=1}^6 P(H_i|E) \quad (2)$$

$$= P(H_1|E) * P(H_2|E) * P(H_3|E) * P(H_4|E) * P(H_5|E) * P(H_6|E)$$

모든 H_n 는 연속적인 속성의 매개변수이며 확률 추론을 위해 가우시안 분포(Gaussian distribution)을 활용하며 식 3(Pang Ning 외, 2007)과 같다.

$$u_i = \frac{H_1 + H_2 + \dots + H_i}{i} \quad (3)$$

$$\sigma_i = \frac{(H_1 - u_i)^2 + (H_2 - u_i)^2 + \dots + (H_i - u_i)^2}{i(i-1)}$$

$$g(H_i, u_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[-\frac{(H_i - u_i)^2}{2\sigma_i^2}\right]$$

$$P(H_i|E) = g(H_i, u_i, \sigma_i)$$

H_i 는 동적 매개변수이며, μ 는 X_i 의 평균값(Mean)이며, σ 는 H_i 의 평방 편차(Standard deviation)이다.

$H_1 \sim H_6$ 는 군집 단계에서 True로 구성된 자원의 정보를 계산한 식이며, $P(H_i|E)$ 는 가우시안 분포(Gaussian distribution) $g(H_i, \mu, \sigma)$ 에 해당된다.

이를 통해 군집된 가상머신의 각각의 사전확률인 $P(H|E)$ 가 구해지면 마지막으로 사후확률을 구한다. 사후확률의 공식은 식 4(John, 1985)와 같다.

$$P(E)_i = \frac{P(H|E)_i * P(R|E)_i}{\sum_{i=1}^n P(H|E)_i * P(R|E)_i} \quad (4)$$

사후확률의 계산을 통한 자원활용율 $P(E)$ 는 사전확률 $P(H|E)$ 와 정적 매개변수인 R 을 이용한다. R 은 각 가상머신의 우선순위를 의미한다.

프로비저닝은 마지막으로 추론된 사후확률 중 가장 높은 확률을 가진 $P(E)$ 에게 자원을 할당 한다.

4. 실험 및 결과

본 논문에서 제안한 클라우드 환경의 교통정보 서비스를 위한 조건부 확률 추론을 이용한 가상 머신 프로비저닝

스케줄링이 효율적으로 서비스 요청에 대해 스케줄링 한다는 것을 증명하기 위해 가상머신으로 구성된 컴퓨팅 환경에서 작업을 분배하여 처리 할 때, 측정 가능한 성능 지표인 활용률(Utilization), 작업 처리율(Throughput), 평균 작업 처리 시간(Average of Turn Around Time)을 통해 비교 모델과 성능을 평가한다.

4.1 실험 테스트 환경 구성

본 논문에서 제안하는 PSCPI의 성능 측정을 하기 위해 DEVS 형식론(Zeigler 외, 2000)을 적용하여 그림 5와 같이 실험 테스트 환경을 모델링, 시뮬레이션 한다.

실험 시나리오는 교통 관련 서비스를 요청, 제공받는 사용자는 교통 클라우드 서버에 접속하여 서비스를 요청한다. 교통 클라우드 서버에서 서비스 요청이 오게 되면 VM Provisioning은 스케줄링 모델을 통해 작업을 수행할 가상머신을 결정한다. 총 10개의 가상머신으로 구성하였고, 총 200개의 서비스 요청을 한다. 서비스 요청시간은 시뮬레이션에서 1 Time 당 1개의 서비스가 요청 되며 총 200 Time 에 200개의 서비스가 요청이 된다. 각각의 가상머신은 물리적 자원을 가지고 있으며, 가상머신의 성능에 따라 Rank(R)을 지정을 한다. Processing Time은 가상머신의 동적변수인 $H_1 \sim H_6$ 의 성능을 요약하였다. 표 4는 실험을 위한 가상

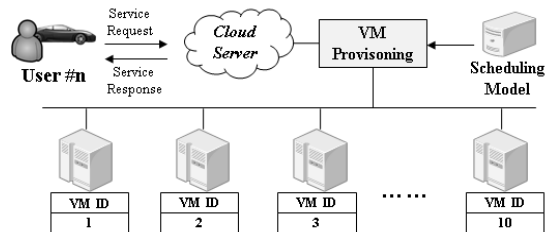


그림 5. 클라우드 테스트 환경

표 4. 가상머신 성능

VM ID	Processing Time (H_1, H_2, \dots, H_6)	Rank
1	32	10
2	30	9
3	28	8
4	26	7
5	24	6
6	20	5
7	18	4
8	16	3
9	14	2
10	10	1

머신의 성능이다.

실험을 위한 스케줄링 모델은 총 3가지로 작업의 실행 순서 등을 결정하는 방법인 라운드 로빈 토너먼트 스케줄링 (Scheduling Round Robin Tournaments : SRRT) (Rasmus 외, 2008)과 실행시간과 대기시간에 따른 스케줄링 기법인 위한 반응시간 최적화된 자원 관리 알고리즘 (Response time Optimization-based Resource Management algorithm: RORM) (Assuncao 외 2009)과 본 논문에서 제안하는 클라우드 환경의 교통정보 서비스를 위한 조건부 확률 추론을 이용한 가상 머신 프로비저닝 스케줄링(PSCPI: Provisioning Scheduling with Conditional probability Inference)을 스케줄링 모델로 비교 분석 한다.

4.2 실험 결과

SRRT, RORM, PSCPI의 성능을 측정하기 위해 본 논문에서는 활용률(Utilization), 작업 처리율(Throughput), 평균 작업 처리 시간(Average of Turn Around Time)을 통해 측정한다.

첫 번째로 활용률(Utilization)은 클라우드 환경에서 가상머신은 일정한 자원을 바탕으로 구성되어 있지만, 활용도는 스케줄링 방법에 따라 많은 차이가 있다. 본 논문에서의 활용율은 서비스 요청에 따라 가상머신이 요청 서비스 별로 얼마만큼의 서비스를 처리 했는지를 나타내는 척도로 정의한다. 활용률은 식 5와 같으며, 결과는 그림 6과 같다.

$$Utilization(\%) = \frac{\text{Number of Service Response}}{\text{Number of Service Request}} \quad (5)$$

서비스 요청 개수가 200이 되었을 때, 스케줄링 모델별 활용율은 SRRT는 0.825, RORM은 0.92, PSCPI는 0.935로 PSCPI가 다른 스케줄링 모델 보다 높은 활용율을 보인다. SRRT는 미리 정의되어있는 정적으로 가상머신을 할

당하기 때문에 활용률이 가장 떨어지며, RORM은 동적으로 할당을 하기 때문에 활용도가 높게 나타나지만 가상머신의 상태변수를 몇 가지 고려하지 않으며, 이보다 많은 변수를 고려하는 PSCPI보다 성능이 떨어진다. 즉, PSCPI는 클라우드 환경을 구성하는 가상머신의 높은 활용도를 보인다.

두 번째로, 작업 처리율(Throughput)은 클라우드 서버 내의 작업 처리의 성능을 나타내는 지표로서, 각각의 가상머신에서 작업이 완료된 시점을 기준으로 평가를 하게 된다. 작업 처리율을 구하기 위해 시뮬레이션 시간을 450 Time 까지 측정하였다. 작업처리율은 식 6과 같으며 결과는 그림 7과 같다.

$$Throughput(\%) = \frac{\text{Number of Service Response}}{\text{Service finish Time}} \quad (6)$$

시뮬레이션 시간이 450 Time이 되었을 때, 스케줄링 모델별 작업 처리율은 SRRT는 0.39468, RORM은 0.44124, PSCPI는 0.47072로 PSCPI가 다른 스케줄링 모델 보다 높은 작업 처리율 보인다. SRRT는 고정적인 작업 할당을 하기 때문에 작업 처리율이 낮으며 ROMA와 PSCIP는 동적으로 작업을 할당하기 때문에 상대적으로 높게 나타난다. 시뮬레이션에서의 ROMA와 PSCIP는 0.03정도의 약간의 차이를 보이지만 실험 환경에서 200개라는 서비스 요청을 했을 경우이며 만약 작업 요청 개수가 많아지면 상대적으로 큰 차이를 보이게 된다. 따라서 PSCIP는 다른 대조군보다 사용자의 서비스의 요청 수가 증가 할수록 서비스 제공이 원활하다.

마지막 실험인 평균 작업 처리 시간 (Average of Turn Around Time)은 각 가상머신별 서비스를 처리하는 평균 시간을 의미하며 서비스 처리시간이 낮을수록 빠른 서비스의 제공이 가능하다. 평균 작업 처리 시간의 공식은 식

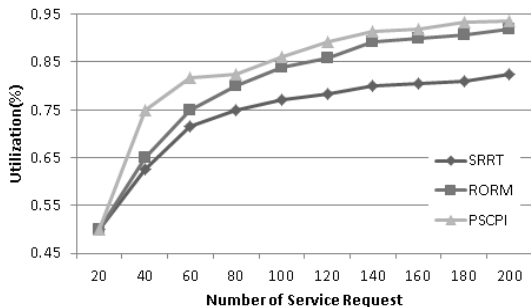


그림 6. 활용율(Utilization)

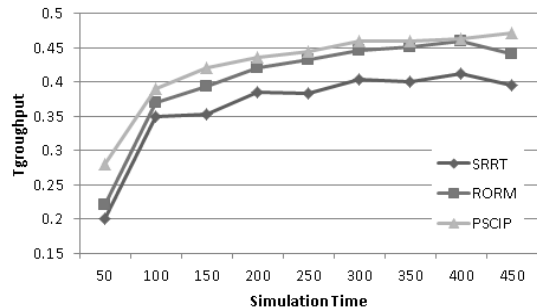


그림 7. 작업 처리율(Throughput)

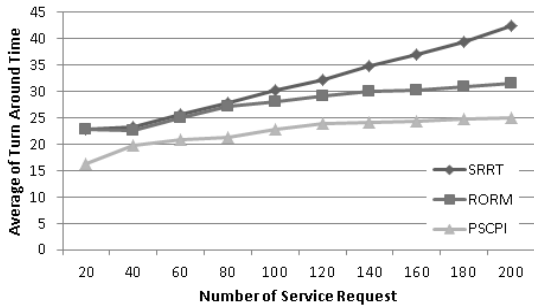


그림 8. 평균 작업 처리 시간(Average of Turn Around Time)

7과 같으며 결과는 그림 8과 같다.

$$ATA = \frac{\sum_{i=1}^n Service\ processing\ Time_i}{Number\ of\ Service\ Response(n)} \quad (7)$$

서비스 요청 개수가 200이 되었을 때, 스케줄링 모델별 평균 작업 처리 시간은 SRRT는 42.406, RORM은 31.478, PSCPI는 25.074로 나타났다. SRRT는 작업 요청의 개수가 늘어나는 만큼 평균 작업 처리 시간의 효율이 점차 떨어지게 되며, RORM은 동적으로 자원을 관리 하며 SRRT와 비슷한 성능을 나타내다가 서비스가 80개 이상부터는 처리 부하가 높아짐에 따라 SRRT보다 성능이 높아지고 있다. PSCPI는 RORM과 마찬가지로 동적으로 자원을 관리 하며 서비스의 개수가 많아질수록 높은 성능을 나타내지만 RORM보다 고려하는 가상머신의 변수가 많기 때문에 높은 성능을 나타낸다. 즉, PSCPI는 다른 스케줄링 모델에 비해 작업 처리속도가 빠르다는 것을 의미한다.

5. 결 론

교통정보 서비스는 점차 증가하고 있는 추세이며, 운전 중인 사용자의 서비스의 요청이 높아지고 있으며 이를 처리해야하는 비용을 줄이기 위해 클라우드 컴퓨팅이 주목을 받고 있다.

클라우드 컴퓨팅은 자원관리를 위한 가상머신의 프로비저닝 스케줄링의 도입이 필요하며 이에 본 논문은 클라우드 환경의 교통정보 서비스를 위한 조건부 확률 추론을 이용한 가상 머신 프로비저닝 스케줄링(PSCPI: Provisioning Scheduling with Conditional probability Inference)을 제안한다.

제안하는 PSCPI에 대한 구조적인 설명과 자원관리를

위한 나이브 베이지안 기반 클라우드 스케줄링 환경의 새로운 메커니즘에 대해 기술하였다. PSCPI의 성능을 검증하기 위해 클라우드 서비스의 모델링과 시뮬레이션을 했다. 실험결과, PSCPI는 다른 실험군보다 클라우드 환경에서 높은 활용율과 처리율을 보이며 낮은 작업 처리시간을 나타내어 다른 대조군들 보다 PSCPI의 우수한 성능을 나타낸다는 것을 확인했으며 클라우드 환경의 교통정보 서비스 제공에 적합하다고 할 수 있다.

향후 연구로는 스케줄링 기법에 로드 밸런싱을 적용하여 가상화 환경에서 자원들의 부하균형에 대한 연구를 통해 자원 활용도를 높이는 방법을 연구할 계획이다.

참 고 문 헌

1. 강금식 박정민(2010), *Excel 통계학 제3판*, 박영사, 서울, pp. 541-543.
2. 김장욱, 김태희, 이수범(2008), “교통정보 제공에 따른 이용자 만족도 모형 개발”, *대한토목학회 논문집 D*, 제28권, 제5D호, pp. 597-607.
3. 김창수, 김학영, 남궁환(2009), “클라우드 서비스를 위한 대규모 클러스터 관리 기술 개발”, *ETRI 전자통신동향 분석*, 제24권, 제4호, pp. 89-98.
4. 민영수, 김홍연, 김영균(2009), “클라우드 컴퓨팅을 위한 분산 파일 시스템 기술”, *한국정보과학회지*, 제27권, 제5호, pp. 86-94.
5. 이승관, 최진혁(2011), “클라우드 환경에서 이기종 네비게이션간의 새로운 도로정보 업데이트를 위한 XML 메타데이터 생성”, *한국콘텐츠학회논문지*, 제11권, 제4호, pp. 83-91.
6. 장용구, 구지희, 최현상, 이상훈(2005), “GPS / GIS 기술 기반의 실시간 교통정보시스템 개발”, *한국 GIS 학회 2005년 추계 학술대회*, pp. 45-56.
7. 장은영, 박춘식(2011), “클라우드 컴퓨팅 서비스의 가용성 최적화를 위한 모델링 및 시뮬레이션”, *한국시뮬레이션학회 논문지*, 제20권, 제1호, pp. 1-8.
8. 최훈, 유성열, 허갑수(2010), “U-시티환경에서 U-교통정보제어서비스를 위한 비즈니스모델”, *한국콘텐츠학회논문지*, 제10권, 제5호 pp. 351-359.
9. Assuncao, M. D. and Costanzo. (2009), “A.: Evaluating the Cost-Benefit of Using Cloud Computing to Extend the Capacity of Clusters”, *In: 18th ACM International Symposium on High Performance Distributed Computing*, New York, pp. 141-150.
10. Bernard P. Zeigler, Herbert Praehofer, Tag Gon Kim (2000), “Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems”, *Academic Press*, pp. 76-96.

11. Daniel Nurmi, Rich Wolski, Chris Grzegorzczak, Graziano Obertelli, Sunil Soman, Lamia Youseff and Dmitrii Zagorodnov (2009), "The Eucalyptus Open-source Cloud-computing System", *Proceedings of 9th IEEE International Symposium on Cluster Computing and the Grid*, pp. 124-131.
12. Jone A. Stankovic (1985), "An Application of Bayesian Decision Theory to Decentralized Control of Job Scheduling", *IEEE Transactions on computers*, vol. c-34, no. 2, pp. 117-130.
13. Pang Ning Tan., Michel Steinbach and Vinpin Kumar (2007), "Introducton to Data Mining", *Addison Wesley*, pp. 73-76.
14. Rasmussen Rasmus V. and Trick Michael A. (2008), "Round robin scheduling-a survey", *European Journal of Operational Research*, Vol. 188, Issue. 3, pp. 617-636.
15. Rodrigo Calheiros, Rajiv Ranjan and Rajkumar Buyya (2011), "Virtual Machine Provisioning Based on Analytical Performance and QoS in Cloud Computing Environments", *International Conference On Parallel Processing (ICPP)*, Taipei, Taiwan, pp. 295-304.



김 재 권 (jaekwonkorea@naver.com)

2011 가천의과대학교 정보처리과 학사
2011 ~ 현재 인하대학교 컴퓨터 정보과 석사과정

관심분야 : 클라우드 컴퓨팅, 인공지능, 모델링&시뮬레이션



이 종 식 (jslee@inha.ac.kr)

1993 인하대학교 전자공학과 학사
1995 인하대학교 전자공학과 석사
2001 미국 애리조나대 전기·컴퓨터공학과 박사
2001 ~ 2002 캘리포니아 주립대학교 전기·컴퓨터공학과 전임강사
2002 ~ 2003 클리블랜드 주립대학교 전기·컴퓨터공학과 조교수
2003 ~ 2006 인하대학교 컴퓨터정보공학부 조교수
2006 ~ 현재 인하대학교 컴퓨터정보공학부 부교수

관심분야 : 소프트웨어공학, 모델링&시뮬레이션