

클라우드 컴퓨팅 서비스의 가용성 최적화를 위한 모델링 및 시뮬레이션

장은영¹ · 박춘식[†]

A study of Modeling and Simulation for the Availability Optimization of Cloud Computing Service

Eun-young Jang · Choon-sik Park

ABSTRACT

Cloud computing emerges as a new paradigm for deploying, managing and offering IT resources as a service anytime, anywhere on any devices. Cloud computing data center stores many IT resources through resource integration. So cloud computing system has to be designed by technology and policy to make effective use of IT resources. In other words, cloud vendor has to provide high quality services to all user and mitigate the dissipation of IT resources. However, vendors need to predict the performance of cloud services and the use of IT resources before releasing cloud service. For solving the problem, this research presents cloud service modeling on network environment and evaluation index for availability optimization of cloud service. We also study how to optimize an amount of requested cloud service and performance of datacenter using CloudSim toolkit.

Key words : Cloud Computing, CloudSim, Queueing model, Availability optimization

요 약

클라우드 컴퓨팅은 장소나 장비에 제한 없이 네트워크를 통해 IT자원을 서비스 형태로 제공받을 수 있는 새로운 패러다임이다. 클라우드 컴퓨팅 환경은 데이터센터에 많은 IT자원이 집약된 형태로 효율적인 인프라구조를 위한 기술과 정책을 적용하여 시스템을 설계해야 한다. 즉, 클라우드 서비스를 효율적으로 제공하여 사용자의 요구를 만족시켜야 하며, 사업자는 불필요하게 낭비되는 자원으로 인한 손해가 없어야 한다. 그러나 최적 시스템을 구축하기 위해서는 서비스를 배포하기 전에 서비스 제공 성능과 자원 사용의 효율성을 예측할 수 있어야 한다. 본 논문에서는 이러한 클라우드 컴퓨팅 시스템 설계과정의 문제를 해결하기 위해 네트워크 환경에서의 클라우드 서비스 모델을 모델링하고 클라우드 서비스의 가용성 최적화를 위해 가용성 평가 지표를 산출하였다. 또한 클라우드 환경이 적용된 CloudSim 시뮬레이터를 이용해 클라우드 컴퓨팅 서비스 요구와 데이터센터 성능에 대한 가용성을 최적화하는 방법을 모색하였다.

주요어 : 클라우드 컴퓨팅, CloudSim, 큐잉모델, 가용성 최적화

1. 서 론

클라우드 컴퓨팅은 통합된 대량의 컴퓨팅 자원을 가상화 기술을 통해 인프라자원과 플랫폼, 소프트웨어 등을

다양한 서비스형태로 제공하는 새로운 기술 방식이다. 클라우드 컴퓨팅 환경에서의 컴퓨팅 자원은 수천 개 이상의 서버가 설치된 데이터센터에 의해 제공된다. 대량의 IT자원이 통합된 형태이기 때문에 효율적인 클라우드 서비스 제공을 위해 클라우드 인프라구조 관리를 위한 매커니즘과 새로운 방법 및 정책을 정의하여 클라우드 기술을 개발 및 설계를 해야한다. 그러므로 사업자는 클라우드 서비스를 배포하기 전에 실제 구축환경에서 제공하는 서비스의 성능과 사업자의 수익을 예측하고 설계함으로써 자원에 대한 접근 비용을 최적화하여 이익을 향상시킬 수 있어야 한다. 특히 클라우드 서비스 요청 및 제공에 따른

*이 연구는 2011학년도 서울여자대학교 교내학술연구비의 지원을 받았음

접수일(2010년 4월 5일), 심사일(1차 : 2010년 5월 6일, 2차 : 2011년 1월 15일), 게재 확정일(2011년 1월 21일)

¹⁾ 서울여자대학교 정보보호학과

주 저 자 : 장은영

교신저자 : 박춘식

E-mail: csp@swu.ac.kr

병목 현상을 조정 할 수 있어야 한다. 따라서 이론적이나 막연한 측정이 아닌 시뮬레이션을 통해 실제 서비스 제공 상황을 예측하는 것이 필요하다. CloudSim은 가상머신을 통해 데이터 센터 호스트의 IT자원을 공유 및 분리하여 서비스로 제공하는 클라우드 컴퓨팅 환경을 테스트할 수 있는 유일한 시뮬레이터이다.

본 논문에서는 데이터 센터의 자원 사용의 효율성 및 서비스 생성 및 실행 비용과 서비스의 품질간의 trade-off 관계를 적절히 조정하여 서비스 성능의 최적화 값을 찾는 방법을 연구하였다. 클라우드와 관련된 핵심 하드웨어 구조에서 서비스 요청을 처리하는 데이터 센터 구성요소를 실제 클라우드 서비스 환경으로 모델링하였으며, 네트워크 클라우드 서비스에 대한 가용성 평가를 위하여 CloudSim의 클래스를 수정하여 가용성 지표를 출력하였다. 또한, 시뮬레이션을 통해 요청된 서비스가 정상적으로 얼마나 빠른 시간 안에 사용자에게 제공되는 지를 테스트하고 그 결과를 기반으로 서비스 품질과 데이터센터의 성능, 수익성을 예측하여 구성요소 설정 값의 최적화 값을 추출하였다.

2. 관련연구

2.1 클라우드 컴퓨팅

클라우드 컴퓨팅은 “IT자원을 서비스로 제공하는 컴퓨팅”으로 인터넷 기술을 기반으로 데이터 센터의 자원을 다수의 사용자에게 다양한 방식의 서비스 형태로 제공하는 기술이다. 클라우드 컴퓨팅 서비스는 다양한 통신수단을 이용한 서비스로 언제, 어디서나 사용과 서비스 구매가 제한되지 않는다. 또한, 사용자의 요구만으로 인프라 및 플랫폼, 소프트웨어 등의 서비스의 확장이 용이하고 유연한 적용이 가능하여 비용절감에 큰 효과가 있다. 클라우드 컴퓨팅 자원의 공유 및 관리에 보안적인 특성으로 물리적, 가상적으로 할당된 자원의 정확한 위치를 사용자가 알 수 없으며, 사업자는 자원의 사용을 모니터링하고 기록하여 최적화된 서비스를 제공한다. 즉, 클라우드 컴퓨팅은 가상화, 자동화, 백업, 스냅샷(모니터 및 기록 등), 컴포넌트(서비스)가 모두 제공되어야 클라우드 컴퓨팅이라고 말할 수 있다(Hayes, 2008; Mell, Grance, 2009).

최근 클라우드 컴퓨팅은 가상화 환경에서의 접근제어 관리, 데이터 암호화, 분산관리, 가용성 및 무결성 등을 보안사항으로 요구하고 있다(CSA, 2009). 가상화는 물리적인 IT자원을 하나로 통합하거나 분리할 수 있는 기능을 말하며, 분산관리나 가용성은 통합된 데이터 자원을 관리

하기위해 필요한 기술이다. 또한, 접근 제어 관리, 데이터 암호화와 무결성은 공유되는 IT 자원을 보호하기위해 필요한 기술이다. 이와 같이 클라우드 컴퓨팅에서 발생할 수 있는 위험을 제거하기위한 연구는 계속 진행되고 있다.

2.2 CloudSim

CloudSim은 클라우드 컴퓨팅 환경에서의 서비스 제공을 시뮬레이션 할 수 있는 유일한 시뮬레이터이다. CloudSim은 JAVA언어로 구성된 텍스트 기반의 시뮬레이터이며, 이벤트 시뮬레이션인 SimJava, 그리드 모델링을 지원하는 GridSim를 상속받는 시뮬레이터로 GridSim과 다르게 가상 기술을 이용한 대량 배포와 툴을 제공한다. CloudSim은 클라우드 자원(Cloud Resource)과 사용자 인터페이스(User Interface Structures), 클라우드 서비스(Cloud Services)와 가상머신 서비스로(Virtual Machine Service) 구성된다. 클라우드 자원과 사용자 인터페이스는 시뮬레이션 기간 동안 Virtual Machine(가상머신), Host(호스트), Datacenter(데이터센터), 어플리케이션의 실행 관리 및 대규모 클라우드 인프라 구조를 관리한다. Cloudlet(클라우드 서비스)과 가상머신 서비스는 Virtual Machine을 기반으로 사용자가 요청한 Host의 권한 설정, 어플리케이션 실행관리, 동적인 모니터링을 수행한다. Host 할당정책 등의 효능을 연구하려면 User Code에서의 전략적 구현으로 가상머신의 설정 기능 등을 확장해야 한다. 즉, CloudSim은 클라우드 기반의 어플리케이션 서비스를 위한 호스팅 환경과 실행 역할을 하는 가상화 클라우드를 가진다.

CloudSim을 구성하는 클래스는 아래의 그림 1과 같으며 크게 Cloudlet, Virtual Machine, Datacenter, Datacenter-Broker(사용자 대신 가상머신을 관리), Host, SANStorage(LAN에 연결된 하드디스크들로 구성된 저장 영역 네트워크), Policy(정책), Provisioner(할당자), Scheduler(스케줄러)가 있다. 그 중 Datacenter, Host, Virtual Machine, Cloudlet은 개발자가 시뮬레이션을 위해 상세 성능 값을 설정 할 수 있다. 클라우드 컴퓨팅 환경의 구성요소로 Datacenter는 Host로 구성되어 있으며, Host는 VMAllocationPolicy(할당 정책)에 따라 Virtual Machine의 컴퓨팅 파워의 자원을 제어하고 Virtual Machine은 VM-Scheduler(가상머신 스케줄러 정책)에 따라 컴퓨팅 파워의 특정 양이 실행 엔진 내에 호스트되는 개별적인 작업 단위를 제어한다. 사용자는 하나 이상의 Virtual Machine을 수행할 수 있으며, Virtual Machine을 통해 Cloudlet을 제공받을 수 있다(Calheiros 등, 2009; Buyya 등, 2009).

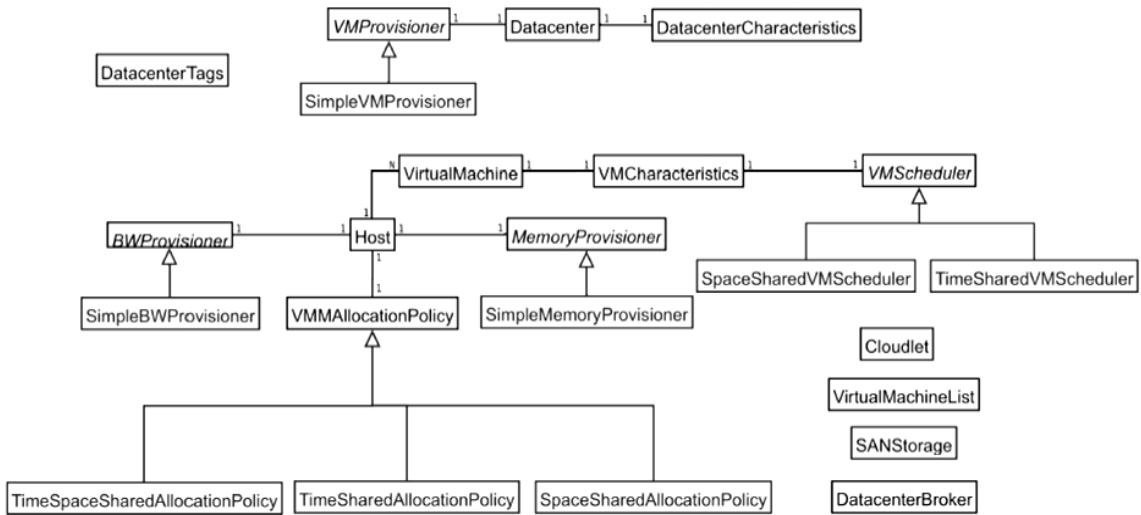


그림 1. CloudSim 클래스 다이어그램

본 논문에서는 클라우드 서비스를 제공할 때 발생하는 비용과 데이터 센터 IT자원의 사용률, 클라우드 서비스의 품질을 비교하기 위해 Datacenter, Host, DatacenterBroker, SpaceShared VMScheduler 클래스를 분석하고 수정하였으며, 위와 같은 Datacenter, Host, Virtual Machine, Cloudlet의 상세 값을 설정하였다.

2.3 큐잉 이론

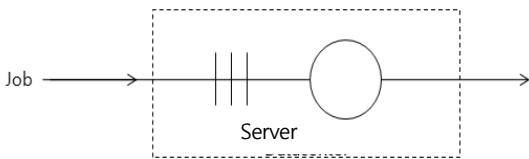


그림 2. 큐잉 모델

그림 2와 같은 큐잉모델은 시스템이나 프로세싱 순서를 결정하는 연산이나 커뮤니케이션 장치 프로세싱 등에 적용될 수 있다. 큐잉 모델은 M/M/1, M/M/c, M/G/1, G/M/1¹⁾등이 존재하며, 각 모델은 작업 요청 시간(패킷 도착 시간)/서비스 시간/ 서버 수로 모델링되어 있다. 이러한 큐잉 모델에서 고려해야 하는 지표는 다음과 같다

1) M/M/1 : 지수분포/지수분포/서버 1대
 M/M/c : 지수분포/지수분포/동일한 병렬의 다중 서버
 M/G/1 : 지수분포/동일분포/서버 1대
 G/M/1 : 동일분포/지수분포/서버 1대

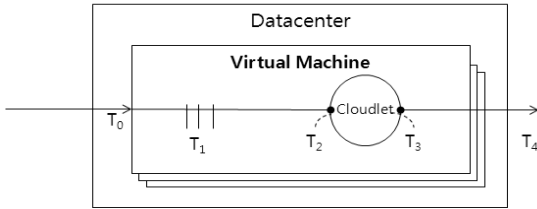
(Adan 등, 2002).

- 작업 요청 시간 : 실제 네트워크 환경에서 사용자는 포아송 분포와 지수 분포에 따라 작업 요청을 한다. 단위시간당 도착하는 사용자의 수가 포아송 분포를 따르며, 작업 요청 간의 경과시간이 지수 분포를 따른다.
- 대기 시간 : 사용자의 서비스 요청 후 서비스를 제공 받기 전의 대기 시간이다.
- 서비스 시간 : 요청된 작업을 완료하는 시간이다.
- 소요시간 : 사용자가 서비스를 요청하고 시스템에서 작업을 수행하는 시간의 합이다.
- 서비스 규율 : 시스템에서 적용한 정책에 의해 서비스를 요청하는 작업의 수행이 결정된다.
- 서비스 성능 : 시스템의 성능으로 서버의 수, 컴퓨팅 파워 등이다.
- 큐(대기열) 길이 : 버퍼 길이로 시스템에서 동시에 수용할 수 있는 사용자의 수 결정한다.

본 연구에서는 클라우드 컴퓨팅 환경에서 여러 대의 서버에 지수분포의 서비스 요청과 지수 분포로 서비스가 수행하는 M/M/1과 M/M/c 모델을 적용하여 모델링 하였다.

3. 클라우드 서비스의 가용성 최적화

3.1 클라우드 서비스 모델링



T0 : Cloudlet.SubmissionTime
 T1 : Cloudlet.WaitingTime
 T2 : Cloudlet.ExecutionStartTime
 T3 : Cloudlet.FinishTime
 T4 : Cloudlet.SuccessTime

그림 3. 클라우드 서비스 모델

클라우드 서비스의 가용성 테스트를 위해 CloudSim에서 제공하는 Datacenter, Host, Virtual Machine, Cloudlet을 이용하여 클라우드 컴퓨팅 네트워크 환경의 서비스 모델을 모델링하였다. Virtual Machine 내에 Cloudlet을 관리하는 큐가 존재하므로 큐잉이론 관점에서 시스템은 Virtual Machine이다. 그러므로 그림 3은 Datacenter 내 Host의 자원을 서비스 형태로 제공하는 Virtual Machine이 다중인 M/M/c 모델이다. Virtual Machine의 SpaceSharedVMScheduler(공간분할 가상머신 스케줄러) 큐잉이론을 기반으로 서비스를 제어한다. 그러므로 Virtual Machine이 SpaceSharedVMScheduler를 적용할 때, 서비스 큐를 갖으며 클라우드 서비스 모델이 큐잉 모델이 된다(Adan 등, 2002; Zeigler 등, 2000).

클라우드 서비스 모델은 그림 3과 같이 T0에 서비스를 요청하는 “작업 요청 시간”과 Data center의 자원이 충분하지 않다면 자원을 할당받기 전에 큐에서 대기 하는 시간인 T1인 “대기시간”이 존재한다. T0은 실제 네트워크에서 작업을 요청하는 분포인 지수 분포로 발생하도록 구현하였으며, 이러한 요청은 난수로 생성된다. 서비스 처리는 작업을 요청한 순서대로 진행되고 사용가능한 PE(Processing Element :처리요소)가 없다면 큐에 대기한다. 요청된 서비스가 자원을 할당받으면 T2에 작업이 실행되고 T3에 작업이 완료되어 종료된다. T4는 사용자가 서비스를 모두 제공받은 시간이 된다.

위와 같은 모델을 기반으로 CloudSim에서 제공하는 Datacenter 클래스의 predictFileTransferTime()함수와 DatacenterBroker 클래스의 submitCloudlet(), VM 클래스

스의 모든 함수, SpaceSharedVMScheduler 클래스의 updateVM Processing()과 CloudletSubmit()함수를 상세히 분석하고 수정하여 네트워크 서비스에 대한 가용성 지표 값을 출력하였다. 클라우드 서비스의 가용성을 평가하기 위한 지표는 다음과 같다.

- 평균 대기 시간 = $E(T_1)$
- 평균 소요 시간 = $E(T_4 - T_0)$
- 평균 큐의 길이
- 작업 처리율
- 서버 활용률 = $\lambda E(T_3 - T_2)$, 서버 활용률 < 1
- 서비스 처리 비용 = $\text{Datacenter.costPerBW} \times \text{Cloudlet.file_size} + \text{Datacenter.cost} \times \text{Cloudlet.ExecuteTime} + \text{Datacenter.costPerBw} \times \text{Cloudlet.ouput_size}$
- 데이터센터 비용 = $\text{VM_Num} \times (\text{Datacenter.costPerMem} \times \text{VM.memory} + \text{Datacenter.costPerStorage} \times \text{VM.size})$

평균대기 시간은 지수분포의 난수로 발생하는 사용자의 요청이 큐에 대기하는 평균 시간이며, 평균 소요 시간, $E(T_4 - T_0)$ 는 사용자가 서비스를 요청한 후(T0) 서비스를 제공받는 데까지(T4) 걸리는 평균 시간을 말한다. 평균 큐의 길이는 서비스 요청에 따라 Virtual Machine에 존재하는 가변하는 큐의 평균 길이이다. 작업 처리량은 일정시간 당 서버가 처리하는 작업량이고, 서버 활용률은 서버가 구동되는 동안 지수분포가 난수(λ)로 발생하는 T0의 사용자의 요청을 처리하는 총 서버 구동시간 비율이다. CloudSim에서 Datacenter 내의 IT자원을 사용할 때 발생하는 비용은 Datacenter_Debt (Datacenter에 Virtual Machine을 생성할 때 발생하는 비용)와 Processing Cost (Cloudlet을 실행하는 비용)로 발생한다. 전자는 Datacenter 내에 물리적인 자원을 제공받는 경우로 Virtual Machine가 Datacenter에서 사용하는 메모리(Datacenter.costPerMem \times VM.memory)와 저장 공간(Datacenter.costPerStorage \times VM.size)의 비용의 합이다. 후자는 Cloudlet을 수행하는 동안 발생하는 비용으로 클라우드 서비스의 입출력 크기 (Cloudlet.file_size, Cloudlet.output_size)에 따라 데이터 센터의 대역폭에서 발생하는 비용(Datacenter.costPerBW)과 클라우드 서비스의 실행시간(Cloudlet.ExecuteTime) 동안 발생하는 데이터 센터의 비용(Datacenter.cost)이다. 그러므로 클라우드 컴퓨팅 환경에서 서비스를 제공하는 데 드는 비용은 이 두 가지 비용의 합이 된다.

표 1. 시뮬레이션 입력 상수 값

상수 명		값
Datacenter	Time_zone	8.0(seoul)
	Cost	0.1
	Cost per memory	0.05
	Cost per storage	0.02
	Cost per bandwidth	0.1
	Bandwidth speed	80
	Architecture	x86
	OS	Linux
	Quantity	1
	Policy	FCFS(SpaceShared)
Host	Memory	1024x4(MB)
	Bandwidth	100(MB)
	The number of processing element	enterprise user × virtual machine
	Performance of PE	1000(MIPS)
	Storage	10(TB)
	Quantity	10
	Policy	SpaceSharedAllocationVM
VirtualMachine	Size	$\frac{\text{storage of Datacenter}}{\text{number of virtual machine}}$
	Bandwidth	$\frac{\text{bandwidth of Datacenter}}{\text{number of virtual machine}}$
	Memory	$\frac{\text{memory of Datacenter}}{\text{number of virtual machine}}$
	The number of core	1
	VMM	Xen
	Quantity	10
	Policy	SpaceSharedScheduler
Cloudlet	Length	2000000(100만 명령어)
	file size(Input size)	400(100만 명령어)
	Output size	400(100만 명령어)

3.2 시나리오

데이터센터를 소유하는 하나의 사업자는 하나의 Datacenter에 Host를 10개 보유하고 있으며 각각의 Host는 Host와 최적화된 성능을 갖는 Virtual Machine을 하나씩 갖는다. Datacenter의 IT자원은 FCFS(First-Come First-Served: 선착순처리)정책으로 관리하며 사업자는 IT자원을 사용자에게 서비스로 제공하기 위해 가상화 기술인 Virtual Machine을 사용하는데, Virtual Machine은 Host에 의해 SpaceSharedAllocationVM(공간분할가상머신할당)으로 할당된다. 그리고 Virtual Machine은 사용자에게 SpaceSharedScheduler(공간분할스케줄러)를 이용하여 Cloudlet을 제공한다. 각각의 시스템의 성능의 세부 값은 표 1과 같다. 사용자의 요청이 증가 할수록 Cloudlet

이 활성화되므로 Datacenter의 IT자원의 사용이 증가한다. 이때, 사업자는 시스템이 감당할 수 있는 사용자의 요청 수 정도와 활용 가능한 컴퓨팅 자원, 사용자 부담의 서비스 비용 및 서비스 대기 시간을 비교, 분석하여 사용자 증가에 따른 가용성 측면에서 최적화된 서비스를 제공해야 한다(Gabriel 등, 2008; Tidwell, 2001; Cohen, 1999).

3.3 시뮬레이션

3.3.1 평균 큐 길이

Datacenter의 성능은 제한되어 있으므로 한 시점에 사용할 수 있는 Cloudlet 요청 량(The number of service)도 한정되어 있다. 수용되지 않는 요청들은 이전 작업의 수행이 종료되기 전까지 큐에 쌓여있게 된다. 그러므로

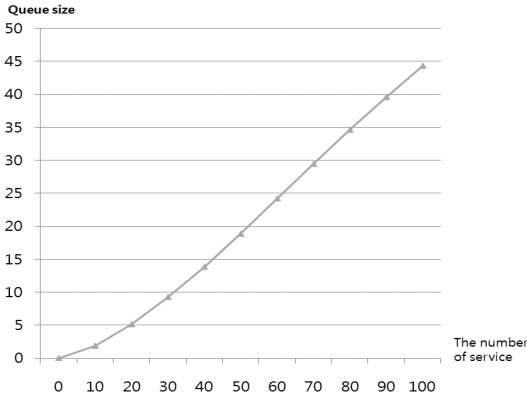


그림 4. 가상머신의 큐 길이

시스템에 큐의 길이(Queue size)가 사용자의 요청 수 보다 작다면, 사용자의 요청 수가 많아질수록 그림 4와 같이 큐의 길이는 점점 증가한다. 큐의 최대 길이는 Virtual Machine에서 설정할 수 있다. 이때, 큐의 길이가 크면 사용자들은 Cloudlet을 제공받을 수 있는 기회는 증가하지만, 너무 크면 사용자의 사용자가 인내하고 기다릴 수 있는 대기 시간을 넘는다. 그리고 큐의 길이가 너무 작으면 사용자들의 요청이 드롭 될 확률이 많으므로 많은 사용자를 수용하는 것이 불가능해 질 수 있다. 잦은 드롭과 서비스 지연은 Cloudlet에 대한 신뢰성과 안전성을 낮추는 결과를 초래할 수 있으므로 사용자의 요구량을 기반으로 적절한 평균 큐의 길이를 설정해야 한다.

3.3.2 대기시간과 소요시간

Cloudlet을 요청하는 사용자가 증가할수록 그림 4와 같이 Queue Size는 증가하며, 이와 비례하게 그림 5와 같이 대기시간(Waiting time)과 소요시간(Turnaround Time)이 증가한다. 즉, Cloudlet의 요청이 증가할수록 큐에서 대기하는 시간과 이로 인한 소요시간이 증가하는 것을 알 수 있다. 사용자가 대기 할 수 있는 시간 중 웹 다운로드 기준시간이 8.6초라고 할 때, Cloudlet의 수행시간은 8.6초 이상이라면 Virtual Machine은 한번에 8개의 Cloudlet을 처리하므로 사용자의 요청이 8일 때, 사용자의 Cloudlet 서비스 요청이 지속 될 것이다. 이러한 값을 기준으로 최적의 평균 대기시간과 평균 소요시간을 지정할 수 있다.

3.3.3 서버 활용률

Datacenter의 활용률(Utilization)이 높을수록 자원의 효율성도 높아진다. 그림 6은 사용자의 서비스 요청 량이

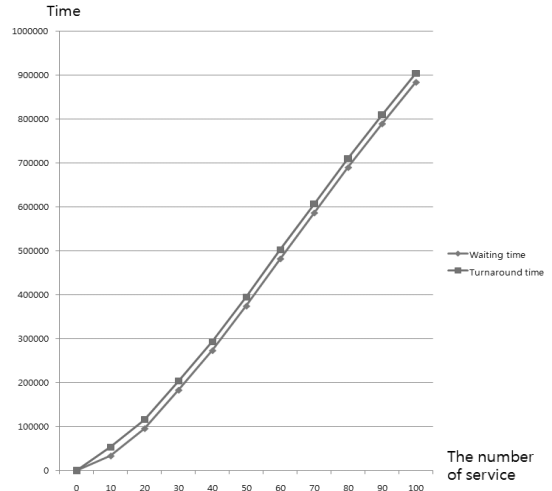


그림 5. 서비스 대기시간과 소요시간

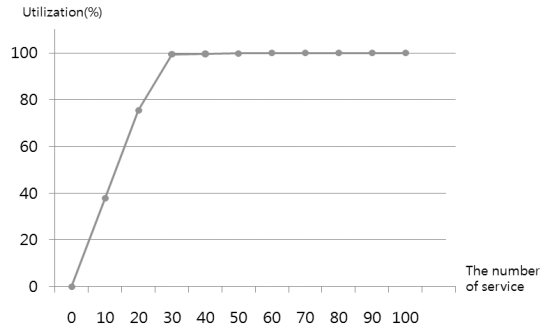


그림 6. 클라우드 자원의 활용률

30개가 될 때, 활용률이 100% 가까이에 도달하여 서버가 쉬지 않고 사용되고 있음을 알 수 있다. 활용률이 높을수록 Datacenter의 IT자원을 모두 활용하기 때문에 자원 낭비를 줄일 수 있지만 Cloudlet을 제공하는데 있어서 과부하가 발생할 수 있다. 그러므로 과부하가 될 때를 대비하여 서버의 활용률을 결정하고 서버의 구조를 적절하게 설계해야 한다. 그림 6은 시스템은 사용자의 요청이 28개 미만일 때, 정상적인 Cloudlet의 제공이 가능하다는 것을 알 수 있다.

3.3.4 작업 처리율

Datacenter의 작업 처리율(Throughput)은 서버 활용률과 같이 증가한다. 그러므로 그림 7도 그림 6과 같이 가상머신이 Cloudlet을 28개 이상 처리할 때 최대 처리량에 도달하는 것을 알 수 있다. 작업 처리량도 활용률과 같이 처리량이 높을수록 Datacenter의 IT자원이 적절하게 사

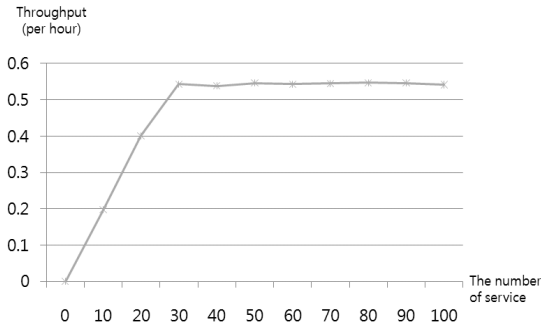


그림 7. 작업 처리율

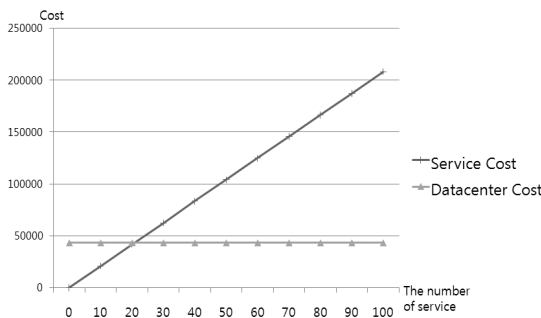


그림 8. Datacenter와 클라우드 서비스 비용

용된다. 그러나 서버의 성능이 고정되어 있기 때문에 작업 처리량은 한정되어 있다. 그러므로 사용자가 집중적으로 Cloudlet을 요청할 시점을 대비하여 작업 처리량의 적정선을 결정해야 한다. 즉, 사용자의 요구를 30 미만으로 제한해야 한다.

3.3.5 클라우드 서비스 비용과 데이터센터 비용

클라우드 서비스 실행 비용(Service Cost)은 Cloudlet을 요청한 사용자가 지불해야 하는 비용이고 Datacenter 비용(Datacenter Cost)은 Datacenter 내에서 Virtual Machine이 IT자원을 생성할 때 발생하는 비용이다. 현재 Datacenter가 1개이며 사업자가 1개이기 때문에 그림 8과 같이 Datacenter의 비용은 일정하게 부과된다. 그러나 전체 클라우드 서비스 비용은 Cloudlet의 요청 량과 비례하게 증가한다. 사업자가 Datacenter에서 Virtual Machine을 관리하는 비용으로 인해 적자가 나지 않기 위해서는 클라우드 서비스 비용이 Datacenter 비용보다 많아야 한다. 그러므로 적어도 20개의 서비스를 제공해야 한다.

위와 같은 입력 지표로 공간분할 정책을 적용한 Datacenter는 서비스의 품질 만을 고려하면 서비스의 수가 8개

일 때, 수월한 서비스 제공이 가능하다. 서버의 활용률을 고려하면 서비스의 수가 28개일 때, Datacenter의 자원을 효율적으로 사용하고 사용자에게도 적절한 서비스를 제공할 수 있다. 또한, 사업자의 수익 측면에서는 서비스의 수가 20개 이상일 때 수입이 증가한다. 그러므로 사업자는 표 1과 같은 성능을 가진 시스템이 20개 이상 28개 이하의 서비스를 수행 할 수 있다고 판단 할 수 있다. 사업자가 추구하는 지표에 따라 이 값은 움직일 수 있다.

4. 결 론

본 논문에서는 클라우드 컴퓨팅의 네트워크 환경에서 제공하는 클라우드 서비스에 대해 가용성을 최적화시키기 위한 모델링 및 시뮬레이션을 연구하였다. 큐잉 이론을 기반으로 클라우드 서비스를 모델링 하였고, 클라우드 컴퓨팅 환경의 시뮬레이션을 제공하는 CloudSim을 수정하여 네트워크 서비스의 가용성 지표를 산출하였다. 또한, 이를 토대로 CloudSim 시뮬레이터를 이용하여 클라우드 서비스의 가용성을 테스트하고, 가용성이 최적화 되는 지점을 찾았다. 그러므로 개발자들은 클라우드 컴퓨팅 환경을 구축하기 전에 이와 같은 방법을 이용하여 클라우드 서비스를 제공하기 위해 적절한 Datacenter의 서버의 수, 성능, 구성 등을 결정할 수 있을 것이다. 결과적으로 클라우드 서비스를 배포하기 전에 시스템과 서비스의 가용성을 최적화 할 수 있으므로 시스템을 재구축하는 비용과 수고를 감소시킬 수 있을 것이다.

향후에는 CloudSim에서 제공하는 다양한 정책들을 이용하고 다양한 모델링을 통해 시뮬레이션을 수행 할 것이며, 적합한 클라우드 시스템의 구성 및 정책들을 개발할 것이다. 또한, 개발자들이 CloudSim을 이용하기 편리하도록 텍스트 기반의 CloudSim을 그래픽 기반의 시뮬레이터로 확장하여 개발할 것이다.

참 고 문 헌

1. Cloud Security Alliance, "Security Guidance for Critical Areas of Focus in Cloud Computing," 2009.
2. B. Hayes, "Cloud computing," Communications of the ACM, vol. 51, Issue 7, pp. 9-11, 2008.
3. P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, Information Technology Laboratory, Version 15, 2009.

4. R.N.Calheiros, R.Ranjan, De Rose, and R.Buyya, "CloudSim: A novel framework for modeling and simulation of cloud computing infrastructures and services," Technical Report, Grid Computing and Distributed Systems (GRIDS) Laboratory, Department of Computer Science and Software Engineering, The University of Melbourne, 2009.
5. R. Buyya, R.Ranjan, and R.N. Calheiros, "Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: challenges and opportunities," Proceedings of the 7th High Performance Computing and Simulation Conference, Leipzig, Germany, 2009.
6. I. Adan, and J.Resing, "Queueing Theory," Department of Mathematics and Computing Science Eindhove University of Technology, P.O.Box 513, 5600 MB Eindhoven, 2002.
7. B.P.Zeigler, H.Praehofer, and T.G. Kim, "Theory of Modeling and Simulation," 2th Ed., Academic Press, Jan 2000.
8. M. Li, "An approach to reliably identifying signs of DDOS flood attacks based on LRD traffic pattern recognition," Computer & Security, vol. 23, pp. 549-558, 2004.
9. G.Macia'-Ferna'ndez, J.E.Dí'az-Verdejo, and P. Garcí'a-Teodoro, "Evaluation of a low-rate DoS attack against application server," COMPUTER & SECURITY, vol. 27, pp. 1013-1030, 2008.
10. T.Tidwell, "Modeling internet attacks," Proceedings of the 2001 IEEE Workshop on Information Assurance and Security United States Military Academy, pp. 54-59, 2001.
11. J.Mirkovic, J.Martin, and T. Reiher, "A taxonomy of DDoS attacks and DDoS defense mechanisms," ACM SIGCOMM Computer Communication Review, vol. 34, Issue 2, pp. 39-53, 2004.
12. F. Cohen, "Simulating cyber attacks, defences, and consequences," Computer & Security, vol. 18, pp. 479-518, 1999.



장은영 (elishajey@swu.ac.kr)

2008 서울여자대학교 정보보호공학과 공학사
2009~현재 서울여자대학교 컴퓨터학과 석사과정

관심분야 : 개인정보보호, 네트워크 보안, 취약점 분석 및 모델링, 클라우드 컴퓨팅 보안, 접근제어



박춘식 (csp@swu.ac.kr)

1995 일본동경공업대 공학박사
1982~1999 한국전자통신연구원 책임연구원
2000~2008 국가보안기술연구소 책임연구원
2009~현재 서울여자대학교 정보보호학과 교수
클라우드컴퓨팅연구센터

관심분야 : 개인정보보호기술, 클라우드컴퓨팅보안