
클라우드 환경에서 QoS 보장을 위한 SLA 관리시스템 구현

윤가람* · 이봉환**

Implementation of SLA Management System for QoS Guarantee in Cloud
Computing Environment

Ka-Ram Yoon* · Bong-Hwan Lee**

본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력양성사업으로 수행된 연구 결과임

요 약

인터넷이 보편화되고 확산됨에 따라 클라우드 서비스가 급속히 확산되고 있다. SLA란 서비스 제공자와 고객 간의 협약으로 서비스 제공자가 서비스 품질을 보증하기 위한 방법이다. 클라우드 컴퓨팅 환경에서의 SLA는 IT 서비스 제공자 입장에서 서비스 품질에 대한 고객만족도를 높이고 경쟁 사업자와의 차별화가 필요하므로 SLA를 통한 QoS 보장은 매우 중요한 요소이다. 그러나 클라우드 환경에서 SLA에 대한 연구는 아직 초기 단계에 머물고 있다. 본 논문에서는 클라우드 서비스를 위한 SLA 관리 지표를 정의하고, 이를 이용하여 QoS 보장을 위한 SLA 관리시스템을 구현하였다. 제안하는 SLA 관리시스템은 오픈소스를 이용한 클라우드 컴퓨팅 환경에서 모니터링 기반의 Migration 정책을 사용하여 Cluster Node들의 부하 분산이 할당된 가상머신들의 가용성, 응답시간, 처리율에 어떠한 영향을 미치는지 분석하고, 기존 클라우드 환경과 비교 분석하였다.

ABSTRACT

Internet is becoming more common with increasing cloud services have been spreading rapidly. A SLA is agreements between service providers and customers of service providers is how to ensure quality of service. SLA in cloud computing environments from the perspective of the IT service providers to customer satisfaction increase for service quality, it will need to differentiate between competing carriers QoS guarantees with SLA is a very important factor. However, the study of SLA in the cloud is staying in its infancy. In this paper, SLA indices for cloud services defining and use them SLA Management System for QoS guarantee has been implemented. The proposed system using monitoring-based migration policy an open source-based cloud computing platform which Cluster Nodes load distribution of virtual machines assigned to their availability, response time, throughput analysis of what affects and is compared with existing cloud.

키워드

클라우드 컴퓨팅, SLA, QoS, 클러스터, 가상머신

Key word

Cloud Computing, SLA, QoS, Cluster, Virtual Machine

* 준회원 : 대전대학교 정보통신공학과 석사 졸업

접수일자 : 2012. 09. 09

** 종신회원 : 대전대학교 정보통신공학과 교수(교신저자, blee@dju.kr)

심사완료일자 : 2012. 09. 27

Open Access <http://dx.doi.org/10.6109/jkiice.2013.17.2.302>

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

인터넷과 IT 서비스가 보편화되고 사용량이 폭발적으로 증가함에 따라 트래픽이 폭증하고 네트워크 부하가 증가하고 있다. 또한, 이와 같은 인터넷 서비스나 IT 서비스에 대한 수요를 공급 측면에서 적절히 대응해 주지 못함에 따라 서비스의 이용이 폭주하는 시간대에는 접속 성공률이 급격하게 떨어지거나 전송 성능이 급격히 저하되어 접속 중 끊김 현상이 발생하는 등의 문제가 대두되고 있다[1].

이와 같은 IT 서비스의 품질 관련 문제에 대해 사용자들은 점차 높은 수준의 서비스 품질을 요구하게 되었으며, 특히 기업용 서비스의 경우에는 다양한 어플리케이션이 시장에 출현하면서 IT 서비스에 대한 QoS(Quality of Service)의 중요성이 증대되고 있다. 최근 이런 경향을 반영하면서 새롭게 주목받기 시작한 것이 SLA(Service Level Agreement), 즉 서비스 수준 협약이다[2]. SLA는 서비스 이용자와 서비스 제공자 간의 일정한 서비스 제공 수준을 보장하기 위한 계약으로, 효과적이고 적절한 SLA 관리를 위해서는 제한된 자원을 효과적으로 활용해서 서비스 이용자에게 적절한 서비스 수준을 제공해 줄 수 있어야 한다. 이처럼 SLA를 미리 정의된 대로 관리, 지원할 수 있도록 제공되는 시스템이 바로 SLA 관리 시스템이다[1][3].

본 논문에서는 기존의 클라우드 환경에서 제안하는 SLA 관리시스템을 적용한다. 제안하는 시스템에서 사용되는 SLA 관리 지표는 가용성, 응답시간, 처리율로 정의하고, 이들 지표들에 대한 실시간 모니터링 기능과, 서버 환경의 장애 시 정보전달 서비스, QoS를 보장해 주기 위한 알고리즘으로 구성된다. 제안하는 시스템은 QoS를 보장해 주기 위해 모니터링 기반의 Migration 정책을 사용하여 각각의 Cluster Node들의 부하를 분산시킴으로써, 가용성, 응답속도, 처리율의 지표에 대한 QoS를 보장한다.

이에 따라 본 논문의 실험에서는 기존의 클라우드 환경과 제안하는 SLA 관리시스템이 적용된 클라우드 환경과의 분석을 통해 성능을 검증하였다.

II. 관련 연구

2.1. 클라우드 컴퓨팅 기술

클라우드 컴퓨팅 서비스는 급속한 IT 환경의 변화에 대처하고 조직의 IT 비용을 절감하기 위한 대표적인 수단으로 급부상하게 되었다. 클라우드 컴퓨팅을 한 문장으로 말하자면 서로 다른 물리적인 위치에 존재하는 IT 자원을 가상화 기술로 통합하여 사용자에게 인터넷을 통해 제공하는 서비스라 할 수 있겠다. 클라우드 서비스는 IT 기술의 집합체로 가상화, 빌려쓰기, 확장성, 표준화 및 자동화, 에너지 효율성 등의 특징을 내포하고 있다.

클라우드 서비스는 새로운 개념이라기보다 기존의 컴퓨팅 환경을 개선하여 효율적인 미래의 IT 서비스 환경을 제시하고 있다.

2.2. SLA 개요

SLA (Service Level Agreement)는 서비스 사업자가 제공하는 서비스를 대상으로 성능과 가용성 및 운영 체제 등 일정한 서비스 수준을 보장하기 위해 맺은 서비스 사업자와 사용자간의 계약을 말하는 것이다. 또한 서비스 사업자가 SLA의 명확한 이행을 위해 서비스 수준을 관리, 운영하는 제반 활동을 의미하는 SLM (Service Level Management)가 있다.

SLA 제도는 사용자에게는 서비스 품질 수준을 보장해 줌으로써 해당 서비스를 통한 비즈니스 가치를 높여 주고 서비스 공급업체는 제공하는 서비스 품질 수준에 적합한 요금을 부가할 수 있는 유효한 의사소통의 수단으로 부각되고 있다.

그러나 SLA 제도가 도입되어 있는 서비스를 중심으로 현재까지의 상황을 고려해 볼 때, SLA 제도 전반에 걸친 이해가 부족하거나 사용자 관점의 서비스와 서비스 사업자 관점의 서비스가 그 내용 및 수준에서 적절한 타협이 이루어지지 않고 있으며, IT 서비스나 통신서비스에 대해 비용을 지불하는 사용자의 불만이 서비스 사업자와 적절하게 타협이 이루어지지 않고 있다는 문제점을 안고 있다[4].

2.3. SLA 관리시스템의 개발 및 도입 필요성

다양한 기능을 보유한 SLA 관리시스템의 개발은 통신서비스나 IT 서비스에 대한 서비스 품질의 정확한 측정과 적절한 서비스 품질수준의 유지 및 가입자 정보의 효과적인 관리를 위한 체계적이며 안정적인 기반을 제공해 줄 수 있다. 이를 통해, 사업자 입장에서는 통신서비스나 IT 서비스와 같은 정보통신서비스를 제공하는 과정에서 필요한 네트워크와 인프라의 운용 및 고객서비스 제공 프로세스에 대한 내부 효율성 증대로 이어져 지속적인 품질 관련 운용비용 절감 및 고객만족도 제고에 기여할 수 있다. 또한, 서비스 제공 사업자나 서비스를 이용하는 고객 모두에게 SLA 제도의 도입 및 운영에 있어서 체계적인 관리를 지원해 줄 수 있는 SLA 관리시스템의 도입 및 운영은 SLA 제도를 통한 서비스 품질의 관리에 대한 실효성 및 효율성 전반에 걸쳐 중요한 결정변수로 작용할 수 있으며, 서비스 품질 전반에 걸친 체계적인 계획 수립과 함께 잘 정의된 SLA 관리시스템의 도입 및 적용은 효과적이고 적절한 SLA 도입 및 활용에 있어서 필수적으로 요구되는 요소라 할 수 있다.

III. 제안하는 SLA 관리시스템

클라우드 환경에서 QoS 보장을 위한 SLA 관리시스템의 구성과 프로세스 과정은 그림 1과 같다.

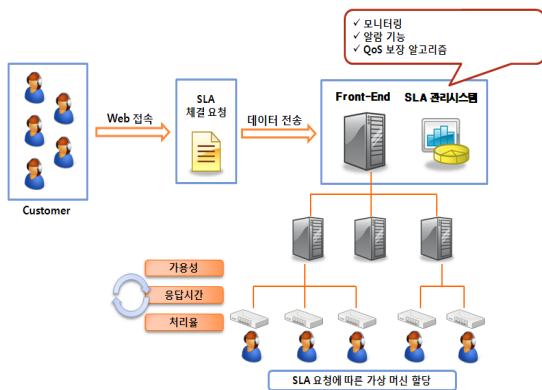


그림 1. 제안하는 시스템의 구성도
Fig. 1 Configuration of the Proposed System

서비스 사용자는 Web을 통해 SLA를 체결하고, 이 데이터는 Front-End 서버로 전송된다. Front-End에서는 이러한 사용자 요청에 맞게 각각의 Cluster Node에 가상머신을 할당하며, 본 논문에서 제안하는 SLA 관리시스템은 Front-End에 위치하고 다음과 같은 기능들을 제공한다.

- SLA 관리지표 실시간 모니터링

서비스 제공 성능과 자원 사용의 효율성을 예측할 수 있는 효과적인 모니터링을 통해 가용성 측면의 QoS를 보장한다

제안하는 SLA 관리시스템에서는 클라우드 서비스를 제공함에 있어 서비스 제공 성능과 자원 사용의 효율성을 예측할 수 있는 Monitoring 기능을 제공한다. SLA 관리시스템에서 Monitoring 할 수 있는 SLA 지표로는 가용성, 응답시간, 처리율이다. 앞에서 언급한 3가지 지표를 사용한 이유는 여러 가지 SLA 지표들 중에 클라우드 컴퓨팅 활용에 있어 중요한 요소로 작용하는 것이 바로 성능과 가용성이다. 따라서 본 논문에서 평가 항목으로 제시한 SLA 품질 지표로 가용성과 성능을 측정하기 위한 응답시간과 처리율을 사용한다.

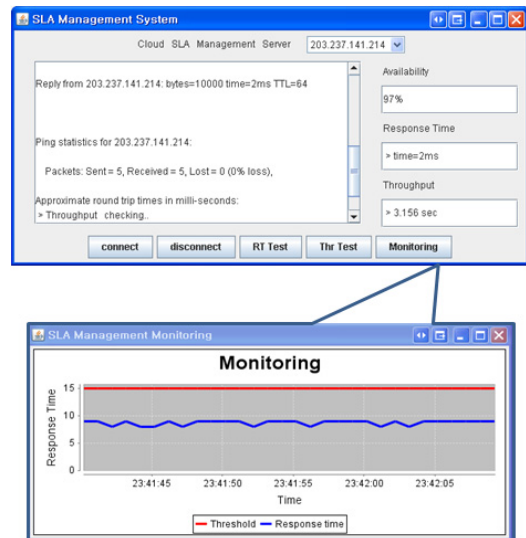


그림 2. SLA 관리시스템의 모니터링
Fig. 2 Monitoring of the SLA Management System

- 서버 환경의 장애 시 정보전달 서비스
 사용자가 할당받은 가상머신이 예기치 못한 장애가 발생하였을 경우, SLA 관리시스템에서 이를 감지하고 Front-End 서버로 메시지를 통한 정보 전달 서비스를 제공한다.

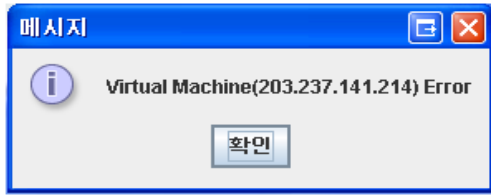


그림 3. 가상머신의 장애 정보 전달 메시지
 Fig. 3 Error Message Transfer from the Virtual Machine

- SLA 관리 지표를 통한 QoS 보장 알고리즘
 SLA 관리 지표는 가용성, 응답속도, 처리율로 정의하고 이러한 측정 항목을 기준으로 자원 사용의 효율성 및 서비스 품질의 QoS를 보장한다.

앞서 정의한 3가지 측정 항목들의 대한 서비스 품질을 보장해주기 위해 Cluster Node 들의 부하를 분산시켰고, 부하 분산을 위해 본 논문에서는 모니터링 기반의 Migration 방법을 제안하였다.

클라우드 컴퓨팅의 특성상 하나의 물리적 서버 자원을 다수의 가상머신이 공유하기 때문에 자원의 적절한 분배가 중요하다. 하나의 Cluster Node에 가상머신이 물리게 되면, 그 Cluster Node에 부하가 생기게 된다. 즉, Cluster Node에 부하가 생긴다는 것은 Node에 할당된 가상머신들의 운영에 영향을 준다는 것을 의미한다.

기존 연구들을 살펴보면 Cluster Node에 가상머신들이 물리게 되면서 Node의 부하가 심하게 발생했고, Node에 할당된 가상머신들의 성능이 현저하게 줄어들면서 가용성과 응답시간, 처리율에 대한 QoS를 사용자에게 보장해주지 못했다.

제안하는 SLA 관리시스템은 서비스 이용자에게 가용성, 응답시간, 처리율의 QoS를 보장해주기 위해 Migration 정책을 이용한다. Migration은 하나의 물리적인 Cluster Node에서 가상머신이 다른 Cluster Node로 이

동하는 것을 말한다. Migration을 사용하면 Cluster Node의 과도한 부하를 다른 유휴 자원이 있는 Node로 나누어 서비스 할 수 있어 가상머신의 동작 불능 상태를 예방할 수 있다.

기존 Migration 정책은 CPU의 점유율만을 고려하여 Migration을 수행한 반면, 제안하는 Migration 정책은 SLA 관리시스템을 통한 평가 지표들을 모니터링 하면서 지표들이 각각의 임계치에 도달하여 얼마 동안 머물러있는지에 대한 여부를 기준으로 Migration을 수행한다.

그림 4는 제안하는 SLA 관리시스템에서의 Migration 정책의 동작 과정에 대한 순서도를 나타낸다.

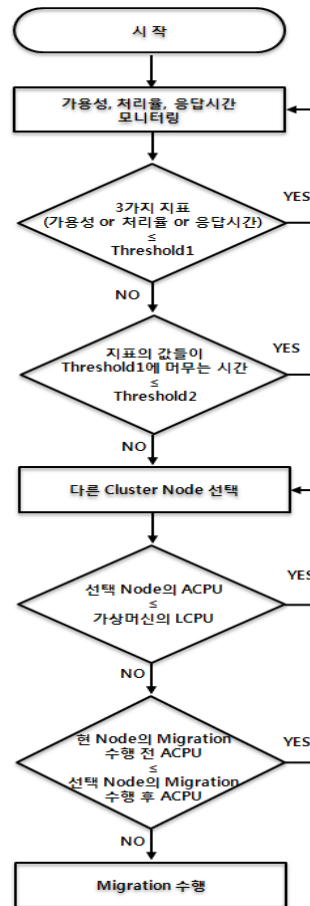


그림 4. Migration 동작과정
 Fig. 4 Migration Operation

그림 4에서 Threshold1은 3가지 지표들에 대한 각각의 평균값을 의미하고, Threshold2는 지표값들이 Threshold1에 도달하여 최대로 머물수 있는 시간을 의미한다. ACPU는 Cluster Node에 남아있는 CPU 자원을 말하며, LCPU는 사용자가 요구한 가상머신의 CPU자원을 의미한다.

IV. 제안하는 동적 전원관리 시스템의 성능 평가

본 절에서는 실험을 통해 기존의 클라우드 환경과 제안하는 SLA 관리시스템을 적용한 클라우드 환경에서의 부하S 분산 정도를 비교함으로써 부하 정도가 SLA 지표인 가용성, 응답시간, 처리율에 어떠한 영향을 미치는지 분석하고 제안하는 시스템의 타당성을 검증한다.

제안하는 SLA 관리시스템의 성능을 평가하기 위해 사용하는 SLA 지표들의 측정방법은 다음과 같고, 두 환경 모두 가상머신을 20개 생성하고, 가상머신의 사양은 동일하게 적용한다.

- 가용성 (Availability)

가용성은 총 운영시간에 대해 정상적인 기능을 수행한 시간이 어느 정도인지를 나타내는 지표로서 본 실험에서는 가상머신 생성 후 총 동작 시간에 대해 정상적인 'running' 상태가 어느 정도 인지를 %로 정의한다[4][5].

- 응답시간 (Response Time)

응답시간은 어떠한 시스템에 대하여 요청이나 명령을 하고 응답이 될 때까지의 걸린 시간을 나타내는 지표로서 본 실험에서는 가상머신에서 데이터 패킷을 Front-End로 보내고 다시 돌아오는데 걸리는 시간을 응답시간으로 정의한다[4][5].

- 처리율 (Throughput)

처리율은 단위 시간당 시스템이 처리할 수 있는 작업의 수를 나타내는 지표로서 본 실험에서는 윈도우 PI값을 계산하는 superPI Tool을 이용해 처리율을 측정한다[4][5].

그림 5은 실험을 통해 구축한 클라우드 환경에서 Cluster Node들의 CPU 사용률을 나타낸다.

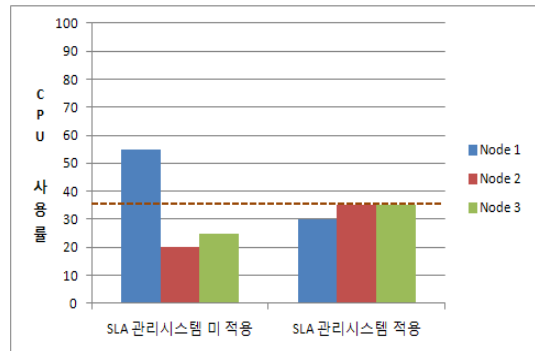


그림 5. Cluster Node의 CPU 사용률 비교
Fig. 5 Comparison of CPU Usage for Each Cluster Node

그림 5과 같이 SLA 관리시스템의 Migration을 수행하여 Cluster Node간 가상머신의 자원 분포를 고르게 하였다. 자원 분포가 고르다는 것은 하나의 Node에 가상머신이 몰리는 현상이 없다는 것이고, 이것은 각 Node간 부하를 분산 시켜준다는 의미가 된다.

위의 실험 결과를 바탕으로 SLA 관리시스템 적용 환경과 미적용 환경의 가용성, 응답시간, 처리율을 측정하여 비교 분석 하였다.

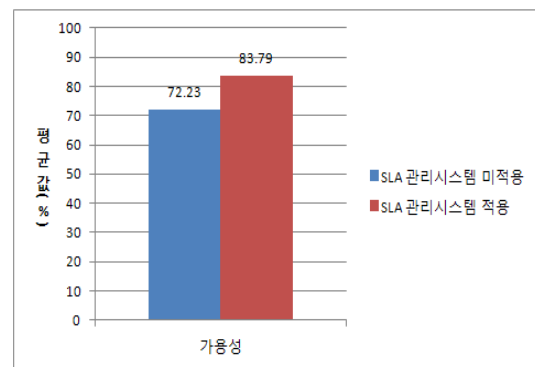


그림 6. 가용성 실험 결과
Fig. 6 Experimental Results for Availability

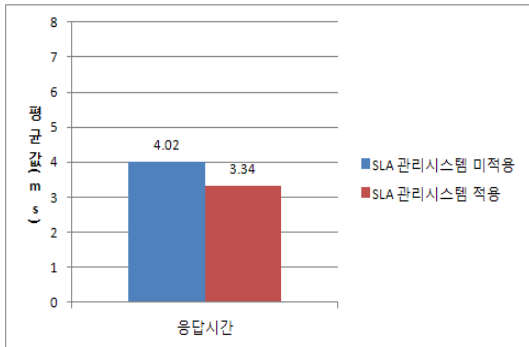


그림 7. 응답시간 실험 결과

Fig. 7 Experimental Results for Response Time

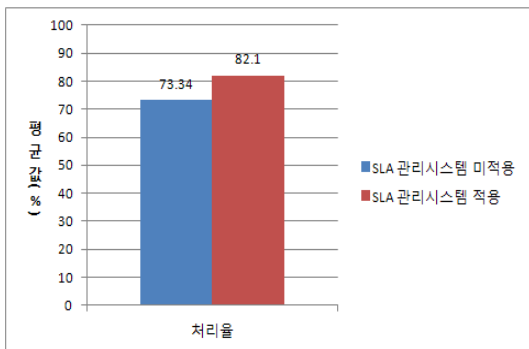


그림 8. 처리율 실험 결과

Fig. 8 Experimental Results for Throughput

본 논문에서 제안하는 클라우드 컴퓨팅 환경에서 SLA 관리시스템 도입은 앞의 실험 결과처럼 기존의 클라우드 컴퓨팅 환경보다 가용성, 응답시간, 처리율에서 우수함을 확인했다. 실험을 통해 가상화 기반의 클라우드 컴퓨팅 환경에서 가용성, 응답시간, 처리율은 각각의 Cluster Node의 CPU 사용률, 즉, Node의 부하에 많은 영향을 받는다는 결과를 얻었다.

실험 결과를 통해 제안하는 시스템은 기존 시스템보다 가용성은 평균 11.56%, 응답시간은 평균 20.35%, 처리율은 평균 8.76%의 향상된 성능을 보였다. 제안하는 시스템은 가상머신의 생성을 통해 발생하는 Cluster Node의 부하를 SLA 관리시스템에서 모니터링을 통한 Migration 정책을 적용함으로써 Node의 부하를 분산시키고, 부하가 분산됨에 따라 기존의 클라우드 환경에서 보다 더 좋은 QoS를 보장해줄수 있게 되었다.

V. 결론

본 논문에서는 클라우드 컴퓨팅 활성화를 위해 오픈 소스를 이용한 클라우드 컴퓨팅 환경에서 SLA 관리시스템을 제안하였다. 제안하는 SLA 관리시스템은 가용성, 응답시간, 처리율의 SLA 지표를 기준으로 서비스 제공자와 서비스 이용자가 동시에 측정 지표에 대한 모니터링이 가능하고, 실시간 모니터링 기능과 모니터링 기반의 Migration 정책을 사용하여 Cluster Node의 부하를 분산시킴으로써 가용성, 응답시간, 처리율에 대한 QoS를 사용자에게 보장해준다. 제안하는 시스템은 실험을 통해 동일한 조건에서 기존의 클라우드 환경보다 향상된 성능의 결과를 검증하였다.

참고문헌

- [1] 최재경, “SLA 관리시스템 도입시 고려요소”, ETRI 전자정보센터, pp 2-6 7-14, 2004. 8.
- [2] 방송통신위원회, “클라우드 SLA 가이드 도입 방안”, 연구보고서, pp.1-6, 2011.10.
- [3] 김대웅, 이길행, 김영선, “서비스수준협약(SLA) 기술 동향”, 전자통신동향분석 제 9권 제 6호 통권 90호, pp.55-65, 2004. 12.
- [4] 김종현, “클라우드 컴퓨팅 활성화를 위한 서비스 품질 관리에 관한 연구”, 성균관대학교 학위논문(석사) 2011.
- [5] 신선영, 권웅기, 김양우, “클라우드 컴퓨팅 SLA 수립을 위한 품질 요소”, 정보통신단체표준 2010, pp.1-10, 2010. 12.

저자소개

윤가람(Ka-Ram Yoon)



2010년 대전대학교 정보통신
공학과 졸업(학사)
2012년 대전대학교 정보통신
공학과 졸업(석사)

※관심분야: 클라우드 컴퓨팅, 네트워크 등



이봉환(Bong-Hwan Lee)

1985년 서강대학교 전자공학과
졸업(학사)

1987년 연세대학교 대학원
전자공학과 졸업(석사)

1993년 Texas A&M 대학교 대학원 전기 및 컴퓨터
공학과 졸업(박사)

현재 대전대학교 정보통신공학과 교수

※ 관심분야: 클라우드 컴퓨팅, 빅데이터, 네트워크
보안 등