JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 18, No. 4, pp. 747-751, July. 2017

네트워크 로그 및 SNMP 기반 네트워크 서버 관리 예측 시스템

문성주
원광대학교 컴퓨터공학과

Server Management Prediction System based on Network Log and SNMP

Sung-Joo Moon
Department of Computer Engineering, Wonkwang University, 460 Iksandae-ro, Iksan, Jeonbuk, 54538, Korea

[요 약]

네트워크 서버 접근시 발생되는 로그는 네트워크 관리에 필수적인 다양한 정보를 가지고 있다. 이러한 정보에서 네트워크 관리에 유용한 정보를 추출하여 사용자 접속량, 비정상적인 접근 등을 예측하여 네트워크 관리의 효율성을 높이고 비용을 줄일 수 있다. 네트워크 관리자는 SNMP를 활용하여 네트워크상 서버의 CPU, 메모리, 디스크 사용율과 같은 정보를 기반으로 서버의 상태를 실시간으로 파악할 수 있다. 본 논문에서는 네트워크 6가지 로그를 분석하여 사용자의 접속량을 예측에 필요한 정보를 추출한 후 시계열 분석 방법인 이동평균법과 지수평활법을 적용하여 실험하였다. 또한 SNMP 시뮬레이터를 활용하여 서버의 CPU, 메모리, 디스크 사용율에 관한 OID를 추출하여 서버의 상태와 장애 예측을 시계열 분석방법으로 실험한 후 엑셀과 R 프로그래밍언어를 통해 시각화된 예측 결과를 제시하였다.

[Abstract]

The log has variable informations that are important and necessary to manage a network when accessed to network servers. These informations are used to reduce a cost and efficient manage a network through the meaningful prediction information extraction from the amount of user access. And, the network manager can instantly monitor the status of CPU, memory, disk usage ratio on network using the SNMP. In this paper, firstly, we have accumulated and analysed the 6 network logs and extracted the informations that used to predict the amount of user access. And then, we experimented the prediction simulation with the time series analysis such as moving average method and exponential smoothing. Secondly, we have simulated the usage ration of CPU, memory, and disk using Xian SNMP simulator and extracted the OID for the time series prediction of CPU, memory, and disk usage ration. And then, we presented the visual result of the variable experiments through the Excel and R programming language.

색인어 : 네트워크 로그 분석, SNMP, 빅데이터 분석, 시계열 분석

Key word : network log analysis, SNMP, bigdata analysis, time-series analysis

http://dx.doi.org/10.9728/dcs.2017.18.4.747

Received 07 April 2017; **Revised** 16 April 2017
Accepted 28 July 2017

***Corresponding Author; Sung-Joo Moon**

Tel: +82-010-4176-8228
E-mail: mitmoon1@hanmail.net

Ⅰ. 서 론

IT 환경이 급속도로 발전되면서 네트워크 장비와 연관된 관리 기술의 고도화가 진행되고 있다. 특히, 고성능화된 보안 장비, 네트워크 서버 장비, 어플리케이션에서 생성되는 수많은 양의 데이터는 빅데이터는 활용 의도에 따라 많은 가치를 생성하고 있다. 네트워크에서 필연적으로 발생되는 로그는 접근자와 관리자에 유용할 수 있는 수많은 정보를 포함하고 있으며 정밀한 분석 기술을 통해 관리자의 의도에 따라 필요한 부분을 추출하여 네트워크 및 서버 관리에 효과적으로 활용할 수 있다. 네트워크 로그와 같이 실시간 또는 비실시간에 발생되는 수많은 로그 정보 또는 SNMP를 통해 취득할 수 있는 서버의 다양한 상태 정보 등은 빅데이터(Bigdata) 분석기술을 활용하여 기존에 발견되지 않았던 다양하고 신뢰할 만한 예측 결과물을 도출하여, 네트워크의 관리 및 네트워크상에 연결된 서버의 장애 예측 등에 효과적으로 응용할 수 있다[1].

SPLUNK[2]는 인간이 만들어 내는 휴먼 데이터와 기기 및 장비들이 만들어 내는 머신 데이터에 대해 수집 및 패턴분석을 수행하는 대표적인 도구이다. SPLUNK 솔루션은 웹서비스 기반이며 다양한 데이터 포맷 및 통신방식에서 데이터를 수집한 후 검색 가능한 머신 데이터 저장소에서 실시간으로 데이터를 획득하여 인덱싱한 후, 데이터 사이의 연관성을 분석한다. 분석된 정보는 그래프, 보고서, 경고, 대시보드 및 시각화된 결과를 생성한다. 본 논문에서는 네트워크상에 발생되는 수많은 로그에 대한 분석과 유용한 정보 추출을 위해 SPLUNK를 활용하였다.

SNMP(Simple Network Management Protocol)[3]는 TCP/IP로 구성된 네트워크에서 서버와 같은 노드들을 관리하기 위해 간단하고 운영 가능한 구조와 시스템을 제공하는 단순 전산망 관리 프로토콜로서 구현이 쉽고 간편한 장점을 가지고 있다. 본 논문에서는 SNMP를 통해 주변 서버의 CPU, 메모리, 디스크 사용률율 시뮬레이션을 통해 취득한 후 이를 시계열 분석을 통해 서버의 상태 정보에 대한 예측 실험을 진행하였다.

본 연구의 목적은 네트워크상에서 보안사고 및 위협에 선제적으로 대응하기 위해 네트워크 로그를 분석하고 접속량을 예측하여 예상 접속 시기에 집중적인 선제적 지원을 통해 접속자의 불편을 최소하고 관리자의 인프라 투입 비용을 절감하고자 한다. 또한 서버 상태에 대한 지속적인 데이터를 분석하여 임계점 이상의 사용률이 발생하는 경우에 대비하여 서버 장애 예측에 대응하고자 한다.

본 논문에서는 네트워크 로그와 SNMP를 통한 서버 상태를 분석하고 예측하여 네트워크 접속량을 예측, 서버 장애 예측 시기에 집중적 경계 및 지원을 통한 선제적 대응 모델을 설정하기 위해 R[5,6] 기반의 라이브러리를 이용하여 시계열 분석 기법 중 이동평균법과 지수평활법을 통해 실험하고 생성된 예측 결과를 제시하였다.

본 논문의 구성은 제2장에서 본 논문의 기반이 되는 예측 기법과 실험에 사용된 도구를 소개한 관련 연구 내용을 소개한다. 제3장에서는 네트워크 로그 분석 방법과 예측 시스템 결과를 제시하며 또한 SNMP를 통한 서버의 상태 정보 분석 및 예측 시스템을 소개한다. 제4장에서는 결론과 향후 연구 내용 및 방향을 설명한다.

Ⅱ. 연구배경 및 관련연구

2-1 예측 기법

통계학은 부분을 가지고 전체를 파악하는 기법으로서 표본에 담긴 정보를 사용하여 모집단의 성질을 추론하거나 검증한다. 다양한 통계 분석 방법중 시계열 분석은 주가, 환율과 같은 시계열 자료를 분석하는 방법으로서 시간의 흐름에 따라 관측된 데이터를 분석하여 예측 결과를 제시하는데 활용되고 있다. 시계열 분석 기법중 이동평균법은 비교적 간단한 시계열 자료 분석법 중의 하나로서 수집되는 데이터에 동일 가중치 부여하여 최근 몇 개 데이터의 단순 평균값을 다음 기간의 예측값으로 추정하는 방법이다. 지수 평활법은 과거의 모든 데이터를 사용하여 평균을 구하되 최근의 자료에 더 많은 가중치를 높게 부여하는 방법이다[4].

2-2 SPLUNK, SNMP 시뮬레이션

SPLUNK는 단일 소스 뿐만 아니라 수백 GB, TB에 달하는 데이터를 안정적으로 실시간 수집하고 인덱싱하여 사용자는 검색 및 분석을 통해서 보고서를 생성하는데 사용되고 있다. 특징적으로 단순화된 필드 추출기를 통하여 빠르게 데이터 추출할 수 있으며 키워드 검색, 추출된 필드의 즉시 검증이 용이하며 한 번의 동작을 통해 자동으로 탐지된 의미 있는 패턴을 볼 수 있다.

SNMP는 OSI 7계층인 응용계층에 존재하는 프로토콜로서 네트워크에 존재하는 다양한 장비들을 관리하기 위해 사용되는 간단한 프로토콜로서 네트워크를 관리하기 위한 표준 통신 규약이다. 하지만, 표준화 규약으로서 SNMP 자체에서 네트워크 내에 존재하는 장비를 관리하지는 않으며 현재 V.1~V.3까지 발표되었으며 본 논문에서는 V.3만 적용하였다. SNMP는 필요한 정보를 에이전트에 요청하는 메니저, 메니저가 요청한 정보를 수집하여 제공하는 에이전트, 정보를 저장하고 있는 MIB로 구성되어 있다. 그중, MIB가 저장하고 있는 대상 정보인 객체를 구분하기 위한 OID를 사용한다. 즉, 메니저가 객체의 인스턴스를 얻고 싶은 경우 메니저는 해당 객체의 인스턴스 ID가 필요하다.

본 연구에서는 Xian SNMP 시뮬레이터에서 CPU, 메모리, 디스크 사용률 정보를 획득에 필요한 OID를 분석한 후 그 값에 대한 시계열 분석을 통해 CPU, 메모리, 디스크 사용률에 관한 예측 실험을 진행하였다.

2-3 관련연구

[5]에서는 대용량 이기종의 보안 로그들을 통합적으로 수집 및 분석할 수 있는 MapReduce 기반의 보안 로그 분석 시스템을 제안한 후 방화벽, 침입탐지시스템, 웹 로그를 대상으로 정규화하고 공통 필드를 대상으로 병합하여 병합 비율과 수행시간 비교 연구를 진행하였다. [6]에서는 실시간 및 비실시간 예측 분석 엔진을 탑재하여 사이버 공격에 선제적으로 대응할 수 있는 프레임워크를 제안한 후 프로파일링 기법에서 회귀 분석이 적합함을 측정 및 분석을 통해서 제시하였다. [7]에서는 보안관련 인프라 로그를 분석하고 예측하여 예상 보안사고 시기에 집중적으로 경계를 통한 선제적 대응을 모색하기 R의 통계 기반 라이브러리를 활용한 예측 시스템 개발 방법을 제안하였지만 구체적인 구현 내용과 결과를 제시하지 못하였다.

Ⅲ. 네트워크 로그 분석 및 예측 시스템

3-1 네트워크 로그 예측 시스템

본 논문에서는 그림 1과 같이 다양한 네트워크 로그를 수집한 후 SPLUNK를 통해 네트워크 접속량 예측에 필요한 정보를 추출한 후 엑셀과 R 프로그래밍언어의 시계열 분석(이동 평균법, 지수 평활법) 라이브러리를 활용하여 향후 예상되는 접속량을 시각적으로 제시하였다.

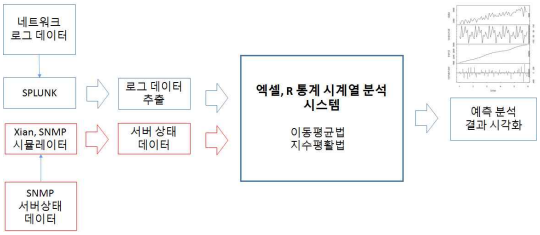

그림 1. 네트워크 로그 및 서버상태 기반 시계열 분석 예측 시스템
Fig. 1. Time Series Analysis Prediction System based on Network Logs and Server Status

네트워크 로그의 수집은 그림 2와 같이 6가지 서로 다른 로그에 대해 365일 동안 로그를 축적한 후 이를 시간대 별로 접속량을 예측하는 실험을 진행하였다. 실제 실험에서는 가상 로그의 생성과 분석에 신뢰성 높은 SPLUNK를 활용하였으며 시계열 분석 라이브러리의 적용은 엑셀과 R의 라이브러리를 활용하여 예측 결과의 유사도를 비교하여 유사한 예측 결과를 제시하였다.

수집된 로그는 SPLUNK를 통해 그림 3과 같이 다양한 분석 결과를 실시간의 시각화된 결과를 확인할 수 있으며 본 논문에서는 시계열 데이터에 해당되는 정보만 추출하여 데이터베이스에 저장하였다.

Index	로그이름	로그파일명	관련데몬	설명
1	시스템 로그	/var/log/messages	Syslogd	리눅스커널로그및주된로그
2	웹 로그	/var/log/httpd/access_log	Httpd	아파치로그저장
3	네임서버 로그	/var/log/named.log	Named	네임서버(DNS)로그
4	FTP로그	/var/log/xferlog	Ftpd	Ftp로그
5	보안로그	/var/log/secure	Xinetd	보안인증관련로그
6	메일로그	/var/log/maillog	Sendmail popper	메일로그

그림 2. 네트워크 로그 수집 종류 및 특성
Fig. 2. Catagory and Characteristic of Network Logs

그림 3. SPLUNK를 통한 로그의 시계열 정보 추출
Fig. 3. Time Serial Information Extraction using SPLUNK

3-2 네트워크 로그 예측 실험 및 결과

시스템 로그는 네트워크 시스템의 표준 에러 관련 메시자가 기록되는 로그 파일로서 날짜 및 시간, 메시지가 발생한 호스트 네임, 메시지를 발생한 내부 시스템이나 응용프로그램의 이름, 자체적으로 발생한 메시지 등의 정보를 가지고 있으며 이중 시계열 데이터를 추출하여 엑셀과 R을 통해 그림 4와 같은 네트워크 접속량 예측 정보를 생성하였다.

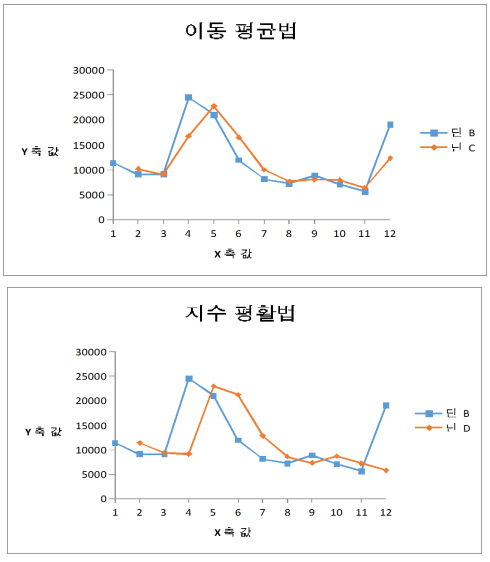

그림 4. 시스템 로그에 대한 접속량 예측 결과
Fig. 4. Prediction Results of System Log Traffic

또 다른 로그인 secure 로그에 대해서도 대용량을 수집하여 시계열 분석을 진행하기 위해 그림 5과 같이 수집하였다. secure 로그는 사용자 인증과 관련된 로그 및 커널, 데몬들에서 생성된 모든 로그를 가지고 있으며 rsh, rlogin, ftp, finger, telnet, pop3등에 대한 접속 기록 및 접속 실패 기록 등 시스템의 보안과 밀접한 관계에 있는 로그 파일이다.

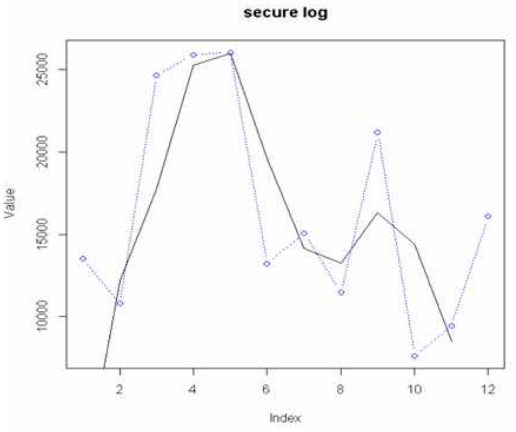

그림 5. secure 로그에 대한 접속량 예측 결과(R 프로그래밍 언어)
Fig. 5. Prediction Results of Secure Log(R Programming Language)

3-3 네트워크 로그 분석 및 예측 시스템

SNMP 시뮬레이션을 통한 서버의 CPU, 메모리, 이스크 사용량을 예측하기 위해 그림 6과 같은 Xian 시뮬레이터에서 시계열 데이터 수집에 필요한 OID를 결정하고 그 값을 0~23시까지의 1일 데이터 생성한 뒤 1년 데이터를 축적하였다.

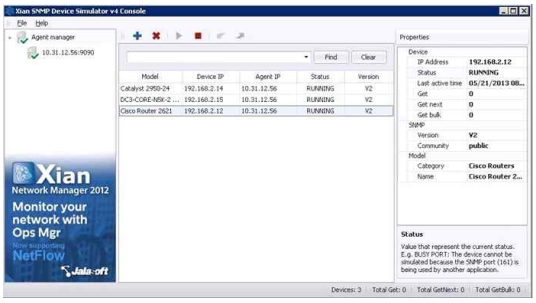

그림 6. Xian SNMP 시뮬레이터 설정 및 데이터 수집
Fig. 6. Configuration and Data Collection of Xian SNMP Simulator

시뮬레이션은 5개의 서버를 설정한 후 각 서버의 1년 동안 축적된 CPU, 메모리, 디스크 사용율 데이터중 시계열 분석이 가능한 OID만 그림 7과 같이 추출하여 이동평균법, 지수평활법을 엑셀과 R 프로그래밍언어를 이용하여 각각 생성하였다.

Disk 1	1.3.6.1.4.1.2021.9.1.1	dskIndex	Integer reference number (row number) for the disk mib.
Disk 2	1.3.6.1.4.1.2021.9.1.2	dskPath	Path where the disk is mounted.
Disk 3	1.3.6.1.4.1.2021.9.1.3	dskDevice	Path of the device for the partition
Disk 4	1.3.6.1.4.1.2021.9.1.4	dskMinimum	Minimum space required on the disk (in kBytes) before the errors are triggered. Either this or dskMinPercent is configured via the agent's snmpd.conf file.
Disk 5	1.3.6.1.4.1.2021.9.1.5	dskMinPercent	Percentage of minimum space required on the disk before the errors are triggered. Either this or dskMinimum is configured via the agent's snmpd.conf file.
Disk 6	1.3.6.1.4.1.2021.9.1.6	dskTotal	Total size of the disk/partion (kBytes). For large disks (>2Tb), this value will latch at INT32_MAX (2147483647).
Disk 7	1.3.6.1.4.1.2021.9.1.7	dskAvail	Available space on the disk. For large lightly-used disks (>2Tb), this value will latch at INT32_MAX (2147483647).
Disk 8	1.3.6.1.4.1.2021.9.1.8	dskUsed	Used space on the disk. For large heavily-used disks (>2Tb), this value will latch at INT32_MAX (2147483647).
Disk 9	1.3.6.1.4.1.2021.9.1.9	dskPercent	Percentage of space used on disk
Disk 10	1.3.6.1.4.1.2021.9.1.10	dskPercentNode	Percentage of inodes used on disk
Disk 11	1.3.6.1.4.1.2021.9.1.100	dskErrorFlag	Error flag signaling that the disk or partition is under the minimum required space configured for it.
Disk 12	1.3.6.1.4.1.2021.9.1.101	dskErrorMsg	A text description providing a warning and the space left on the disk.

그림 7. Xian SNMP 시뮬레이터을 이용한 디스크 OID 생성
Fig. 7. Di나 OID Generation using Xian SNMP Simulator

그림 7에서 생성된 디스크 OID를 활용하여 사용률이 예측된 결과는 그림 8과 같다.

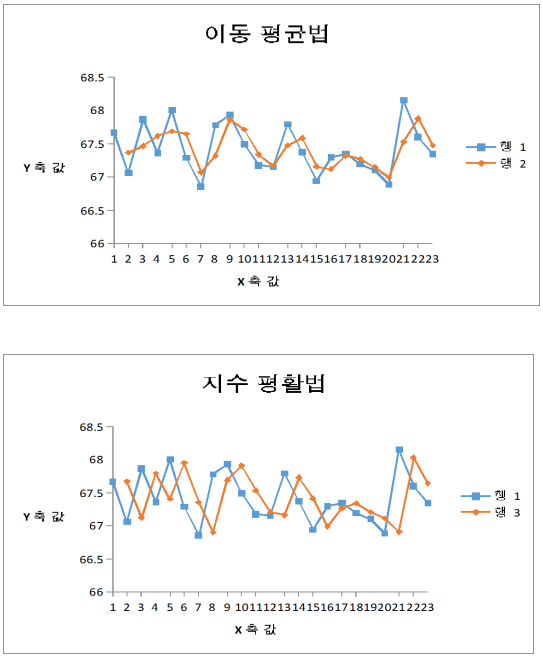

그림 8. Xian SNMP 시뮬레이터을 이용한 디스크 사용율 시계열 예측(엑셀 라이브러리 적용)
Fig. 8. Disk Using Ratio Time Serial Prediction using ian SNMP Simulator(MS Excel Library)

실제로 디스크 사용률에 대한 시계열 데이터를 수집하기 위해 OID 1.3.6.1.4.1.2021.9.9.번 핵심적으로 활용하였으며 디스크에 사용된 공간의 크기는 %단위이며 사용률에 대한 시계열 데이터는 1시간 마다 평균값을 산정하여 보다 신뢰성 높은 데이터를 확보하였다. 또한 예측된 결과값의 신뢰성 및 타당성을 검증하기 위해 동일 시계열 데이터에 대해 R 프로그래밍 언어의 시계열 분석 라이브러리를 적용하여 그림 9와 같은 동일한 결과를 도출하였다.

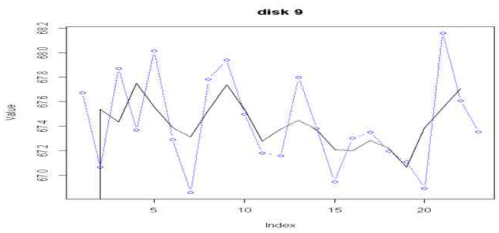

그림 9. Xian SNMP 시뮬레이터을 이용한 디스크 사용율 시계열 예측(R 라이브러리 적용)
Fig. 9. Disk Using Ratio Time Serial Prediction using ian SNMP Simulator(RI Library)

Ⅳ. 결론 및 향후 과제

본 논문에서는 6가지 네트워크 로그를 수집한 후 시계열 분석(이동평균법, 지수 평활법)에 필요한 정보를 SLPUNK를 통해 생성한 후 예측 모델 적용하여 각 로그에 대한 접속량을 예측하는 실험 결과를 제시하였다. 또한 네트워크상의 서버 상태 정보를 SNMP를 통해 확보한 후 서버 상태 정보중, CPU 사용률, 메인 메모리 사용율, 디스크 사용률 정보를 기반으로 서버 상태 시계열 예측 정보 생성하는 결과 Xian SNMP 시뮬레이터를 활용하여 실험결과를 제시하였다.

본 논문의 결과는 네트워크에서 발생되는 다양한 대용량의 정보를 분석하여 네트워크 관리 비용과 장애 등에 선제적으로 대응하는데 활용될 수 있다. 향후에 본 연구를 기반으로 보다 구체적인 정보를 획득과 분석을 통해 다양한 예측 기법을 적용하여 네트워크 효율적인 관리 및 비용 절감에 활용할 계획이다.

참고문헌

[1] Wei-Yu Chen, Jazz wang, “Building a Cloud Computing Analysis System for Intrusion Detection System,” CLOUD SLAM, April 2009.
[2] SPLUNK, http://www.splunk.com
[3] SNMP, http://en.wikipedia.org/wiki/Simple_Network_Management_Protocol
[4] Lee Woo-ri, Time-Series Analysis and Forecast : Understanding and Application, Tomming Publishing Co., Ltd, 2013.
[5] Choi dae soo, Moon Kil-jong, Kim Yong-min, Roh Bong-nam, " Mass Security Log Using MapReduce, " " Korea Information Processing Society, Article 9 of the Korean Information Processing Society, " Article 9, 2011.
[6] Real Time Forecasting System Design and Implementation of Real Time System Using Big Data Log, University of Korea Graduate School of Economics, 2015.
[7] Lee Sang-jun, " abnormal judgment system ", patent 10 - 1542534, UNET system, 2015.
[8] Jung Duk Won, " Real-time analysis and forecasting service framework of traffic giant, " Konkuk University Graduate School of Education, University of Korea,2014.
[9] Symbick-gil, " Real-time mass data analytics technology and application cases, " Ne xR, 2012.
[10]Ravi Kalakota, “USING BUSINESS INTELLIGENCE FOR REAL-TIME DATA MINING, CUSTOMER SEGMENTATION AND PREDICTIVE ANALYTICS”, DZone.com, 2013.

문성주(Sungjoo Moon)

1992년 2월 : 호원대학교 전자계산학
1997년 8월 : 원광대학교 컴퓨터공학 (공학석사)
2013년 2월 : 원광대학교 컴퓨터공학 (박사과정수료)

관심분야 : 소프트웨어 역공학, e-business, 임베디드시스템, 정보보안.