ISSN 2234-4438
중소기업융합학회 논문지 제6권 제1호 pp. 17-23, 2016 DOI : http://dx.doi.org/10.22156/CS4SMB.2016.6.1.017

빅데이터를 이용한 APT 공격 시도에 대한 효과적인 대응 방안

문형진[1], 최승현[1], 황윤철[2*]
[1]백석대학교 정보통신학부, [2]한국교통대학교 정보공학과

Effective Countermeasure to APT Attacks using Big Data

Hyung-Jin Mun[1], Seung-Hyeon Choi[1], Yooncheol Hwang[2*]
[1]Division of Information & Communication, Baekseok University
[2]Department of computer science and information Engineering, Korea National University of Transportation

요 약 최근에 스마트 폰을 비롯한 다양한 단말기를 통한 인터넷 서비스가 가능해졌다. ICT 발달로 인해 기업과 공공기관에서 크고 작은 해킹사고가 발생하는데 그 공격의 대부분은 APT공격으로 밝혀졌다. APT공격은 공격의 목적을 달성하기 위해 지속적으로 정보를 수집하고, 장기간 동안 공격대상의 취약점을 분석하거나 악성코드를 다양한 방법으로 감염시키고, 잠복하고 있다가 적절한 시기에 자료를 유출하는 공격이다. 본 논문에서는 APT 공격자가 짧은 시간에 타겟 시스템에 침입하기 위해 빅데이터 기술을 이용하는 정보 수집 기법을 살펴보고 빅데이터를 이용한 공격기법을 보다 효율적으로 방어할 수 있는 기법을 제안하고 평가한다.

키워드 : 중소기업 정보시스템, APT 공격, 빅 데이터, 하둡

Abstract Recently, Internet services via various devices including smartphone have become available. Because of the development of ICT, numerous hacking incidents have occurred and most of those attacks turned out to be APT attacks. APT attack means an attack method by which a hacker continues to collect information to achieve his goal, and analyzes the weakness of the target and infects it with malicious code, and being hidden, leaks the data in time. In this paper, we examine the information collection method the APT attackers use to invade the target system in a short time using big data, and we suggest and evaluate the countermeasure to protect against the attack method using big data.

Key Words : SMB Information System, APT Attack, Big Data, Hadoop

1. 서론

최근에 많은 기업과 공공기관을 대상으로 해킹 사고가 빈번하게 발생하고 있다. 최근의 주요 해킹 공격에 사용되는 있는 기법이 바로 APT(Advanced Persistent Threat)공격이다[1]. APT공격은 새로운 공격기술이 아니라 기존에 존재하는 공격기법들을 복합적이고 지능적으로 사용하여 장기간 동안 특정 목적을 가지고 대상을 공격하기 위해 시도이다. 공격 목적은 정보시스템의 정보유출 또는 시스템의 서비스 중단시키는 것을 포함한다. 목적을 달성하기 위해서 공격자는 공격 대상의 다양한 정보를 수집하고 이를 활용하여 장시간 동안 해당 시스템의 취약점을 분석하여 복합적으로 공격을 시도하거나 신종 악성코드를 감염시켜 공격 시점을 기다리면서 잠복한다. 공격자는 공격 대상의 IT 인프라를 장악하고, 공격 대상의 재방문을 용이하게 만들기 위해 악성코드 등을 설치한다. 장기간 동안 특정 목적을 위해 공격이 진행되

Received 2016-03-03 Revised 2016-03-14 Accepted 2016-03-17 Published 2016-03-31
*Corresponding author : Yooncheol Hwang (dolpin98@nate.com)

지만 공격대상은 이를 인지하지 못하고 지나치는 경우가 많다. 뿐만 아니라 APT 공격은 특정 목적을 달성한 후에 흔적을 지우면서 공격하기 때문에 쉽게 공격여부를 확인하기 어렵다.

성공적인 APT 공격을 위해서는 정보시스템의 많은 정보를 수집하고, 분석하는 단계를 걸치기 때문에 많은 시간이 소요된다. 공격자 입장에서는 공격대상의 정보수집 및 분석 단계의 소요 시간을 최소화하는 것이 공격의 성패를 좌우하는 요소이다. 특히 공격자는 Google Search Engine을 대체할 목적으로 구글에서 만든 빅데이터 플랫폼인 Nutch를 이용하면 타겟시스템의 양질 정보를 수집 할 수 있고, 수집된 정보를 기반으로 공격을 시도하면 단 시간에 공격이 가능하다. 따라서 본 논문에서는 타겟 시스템의 정보를 수집하고 분석하는 시간을 단축하기 위해 빅데이터를 이용한 공격 시나리오를 제시하고, 이에 대한 효과적인 대응기법을 제안한다. 본 논문의 구성은 다음과 같다. 2장에서는 빅데이터와 하둡 플랫폼에 대해 설명하고, 3장에서는 빅데이터를 이용한 APT 공격시나리오를 소개하고, 4장에서는 공격에 대한 대응하는 방법을 논의한 후 5장에서 결론 및 향후 연구로 끝을 맺는다.

2. 관련연구

2.1 빅데이터

빅데이터는 디지털 환경에서 다양한 단말기를 통해 생성되는 많은 양의 데이터를 의미하며, 최근에는 데이터의 형태가 수치뿐만 아니라 문자와 영상처럼 데이터의 종류도 다양해지고 있다. 이런 데이터들은 저장되지 않았거나 저장되더라도 분석하지 못하고 버리게 되는 경우도 많았지만 최근 빅데이터 기술의 발달로 대량의 데이터를 통해 가치 있는 정보로 생성되고 있다. IT 기관에서 빅데이터를 Table 1과 같이 정의하고 있다[2].

빅데이터는 규모, 변화, 속도, 복잡성 등 다양한 특징을 가지고 있고 의미 없던 데이터가 처리 및 분석을 통해 가치를 창출하는 정보로 변환하는 기술이다. 시간의 흐름 속에서 IT는 하드웨어에서 소프트웨어로, 소프트웨어에서 데이터로 이동하기 시작하였고, 데이터의 단순 분석보다 데이터 속에 내포하고 있는 가치에 초점을 두고 발전하기 시작되었다[2].

빅데이터가 장점만 가지는 것은 아니다. Yun은 빅데이터가 가지는 위험요인에 대해 유형별로 분류를 제시하였다[3]. 정보화 사회에서 위험요인에 대한 인식하면서 이를 해결할 수 있는 기술 개발을 함께 도모해야 한다.

Table 1. Big Data Definition and Feature

Organization	Definition and Feature
Forrester	• Volume, Velocity, Variety ✓ Large-scale data with diversity ✓ Data to create economic value
SERI	• Sum of Large-scale data • Technology and tools related to Large-scale data
Gartner	• 3V : Volume, Variety, Complexity ✓ Volume : Large scale of data ✓ Variety : Diversification of the types of different data from the increase of the types of data such as log history, social and location information ✓ Complexity : The types of data vary being related to non-structured data, the difference of data storage mode, and duplication issue. Control and management method for complex data is necessary.
SAS	• 4V : Volume, Variety, Velocity, Value ✓ Volume, Variety similar to Gartner ✓ Velocity : Real time information increase regarding sensor, monitoring, IoT information, and streaming, Data generated in real time, Importance of the speed to process and analyze ✓ Value : A new value creation
Nomura LAB	• Three components of big data ✓ Human resource and organization to process big data ✓ Data process, collection, analysis techniques ✓ Data resource • Widespread application of three elements to real life with the characteristics of data and the development of computing power

2.2 하둡 생태계

하둡은 다중 컴퓨터 분산 프레임워크로 대표적인 공개 소프트웨어다. 하둡은 여러 개의 PC를 연결하여 마치 하나인 것처럼 묶어 대용량 데이터를 처리하는 기술로, 하둡은 수 천대의 분산된 x86 장비에 대용량 파일을 저장하는 기능을 제공하는 분산파일시스템(Distribute File System)과 저장된 파일 데이터를 분산된 서버의 CPU와 메모리 자원을 이용해 쉽고 빠르게 분석할 수 있는 컴퓨팅 플랫폼인 맵리듀스(MapReduce)로 구성되어 있다. 하둡은 하나의 마스터와 여러 개의 슬레이브로 구성된 아키텍처를 갖는다. 이런 하둡에 대한 여러 가지 측면에서 장점이 있기 때문에 하둡이 빅데이터 처리와 분석을 위한 플랫폼 시장에서 사용되기 시작하였다.

기본적으로 하둡 플랫폼은 MapReduce과 HDFS (Hadoop Distributed File System)를 지원하지만 분석과 저장의 기능만으로는 빅데이터를 제대로 활용하기 어려워서 이런 부분을 보완하기 위해 하둡 생태계(Hadoop ECO system)가 만들어졌다.

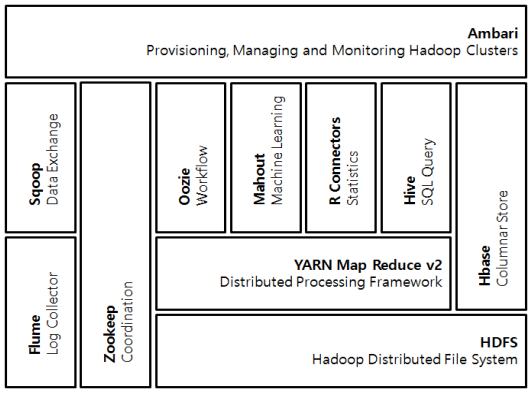

Fig. 1. Hadoop Ecosystem

Fig. 1은 하둡 생태계를 나타낸 그림이다[4]. 각각의 프로그램은 아래와 같은 기능을 수행한다[5,6].

- Flume는 방대한 양의 데이터 수집기능을 수행하여 하둡 파일시스템에 저장한다.
- Sqoop는 RDBMS에서 하둡으로 데이터 이전을 위해 설계된 프로그램으로 기존 레거시 시스템의 데이터를 하둡에 로딩하거나 처리결과를 RDBMS에 저장한다.
- Zookeeper는 분산 환경에서 서버 간에 상호조정 서비스를 제공한다. 로드분산처리, 처리결과의 동기화, 분산 환경을 구성하는 서버 환경설정을 통합 관리한다.
- Oozie는 하둡 작업을 관리하는 워크플로우 시스템이다.
- HBase는 HDFS 기반의 컬럼 기반의 DB로 BigTable 논문을 기반으로 개발되어 분산 클러스터 관리 및 복구기능 수행한다.
- Pig는 하이레벨 언어로 데이터 분석 플랫폼으로 대규모 병렬처리에 사용된다.
- Elastic Search는 아파치 Lucene 기반 분산형태인 오픈 소스 검색엔진으로 짧은 대기시간에 검색 및 인덱스 갱신이 가능하다.
- Mahout는 하둡 기반 데이터 마이닝의 중요한 알고리즘을 구현한 오픈소스이다.
- Hive는 하둡 기반의 dataware-housing을 위한 솔루션이다.

2.3 하둡 플랫폼

Fig. 2는 HDFS의 기본 개념도로서 파일을 64Mbyte 단위로 나누어 분산 저장하는 시스템으로 빠르고, 안정적이다. HDFS는 네임노드와 데이터노드로 분리되어 데이터노드에 장애발생시 새로운 데이터노드를 추가하여 시스템을 안전적으로 유지하도록 설계되었다[6,7].

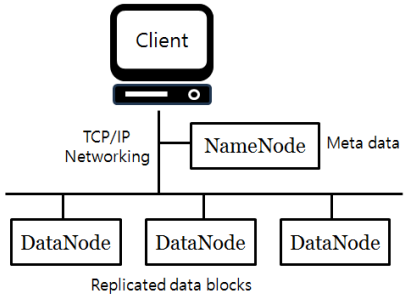

Fig. 2. Key Map of HDFS

Fig. 3은 MapReduce의 단순 개념도를 나타낸 것이다[6-8]. MapReduce는 하나의 큰 데이터를 여러 개의 조각으로 나누는 Map 함수와 처리된 결과를 하나로 취합하여 결과를 도출하는 Reduce 함수로 구성된다. 데이터를 입력받아 Map 함수를 통해 Key-Value 의 형태로 저장하고, Reduce 함수는 중간 데이터를 Key 중심으로 재분류하는 분석한다.

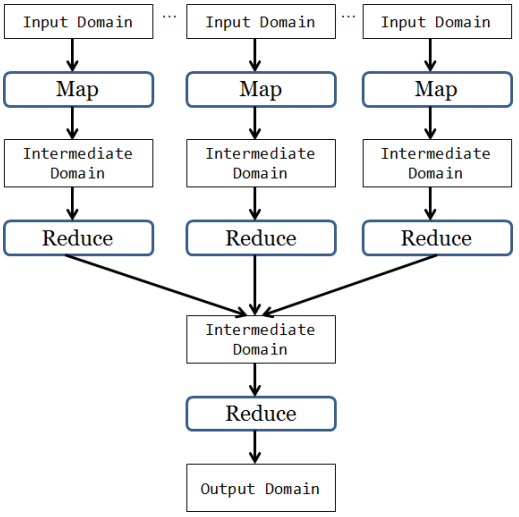

Fig. 3. Key Map of MapReduce

3. 빅데이터를 이용한 APT공격 시나리오

3.1 일반적인 APT공격 시나리오

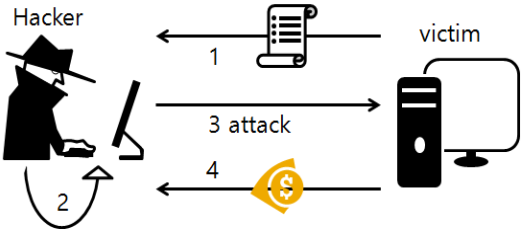

Fig. 4. APT Scenario

Fig. 4는 일반적으로 해커가 APT 공격하는 시나리오를 나타낸 것이다. 다음의 단계를 통해 공격을 수행한다.

① 해커가 구글 등의 검색엔진 등을 활용하여 직접 정보시스템의 정보를 수집한다.

② 해커가 수집된 모든 정보를 분석한다. 정보시스템의 열려진 포트가 무엇인지, 정보시스템의 운영체제 등 웹서버를 분석한다. 게시판 등에서 업로드 제한이 있는지, 취약점이 있는 시스템을 사용하는지 여부를 판단한다. 업로드 제한이 없다면 웹쉘(Web shell)의 업로드가 가능하다면 해킹 시나리오가 가능하다.

③ 해커 자신이 작성한 해킹 시나리오대로 APT 공격을 수행한다.

④ 해킹 성공 시 정보시스템을 장악하고, 정보시스템의 정보 유출 또는 정보 파괴 등 해커가 원하는 특정 목적을 수행한다.

3.2 하둡을 이용한 정보수집 및 분석 시나리오

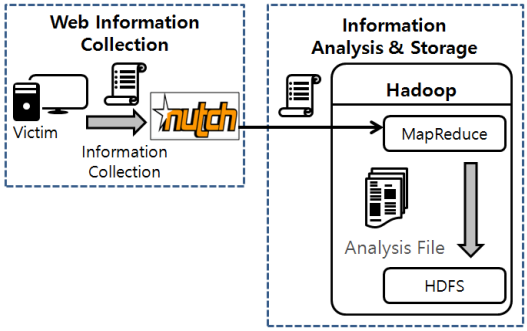

Fig. 5. Information collection and analysis using Hadoop

Fig. 5는 하둡을 이용하여 정보 수집 및 분석하는 시나리오를 나타낸 것이다. 다음과 같은 과정을 수행한다.

① 인터넷 상의 모든 데이터를 수집하기 위해 설계된 Nutch를 이용하여 해당 정보시스템의 모든 정보를 수집한다.

② 수집된 정보를 하둡에 존재하는 MapReduce에 정보를 전달한다. MapReduce는 사용자가 직접 원하는 데이터를 분석할 수 있도록 프로그래밍 할 수 있다.

③ 분석된 파일을 HDFS(Hadoop Distributed File System)에 안전하게 분산 저장한다.

3.3 빅 데이터를 이용한 APT공격 시나리오

APT 공격은 공격자가 공격대상인 시스템으로부터 공격을 위한 정보를 수집하고, 적절한 시기가 도래했을 때 공격을 시도한다. 하지만 이런 경우 APT 공격의 장기간의 준비시간이 소요된다.

하지만, 하둡을 이용하여 정보수집하고, 분석한 결과를 이용한 공격, 즉, 빅데이터를 이용한 발전된 APT 공격을 시도한다면 짧은 시간 내에 공격이 가능하게 된다.

Fig. 6은 빅데이터를 이용한 발전된 APT 공격 시나리오를 단계별로 나타낸 것이다.

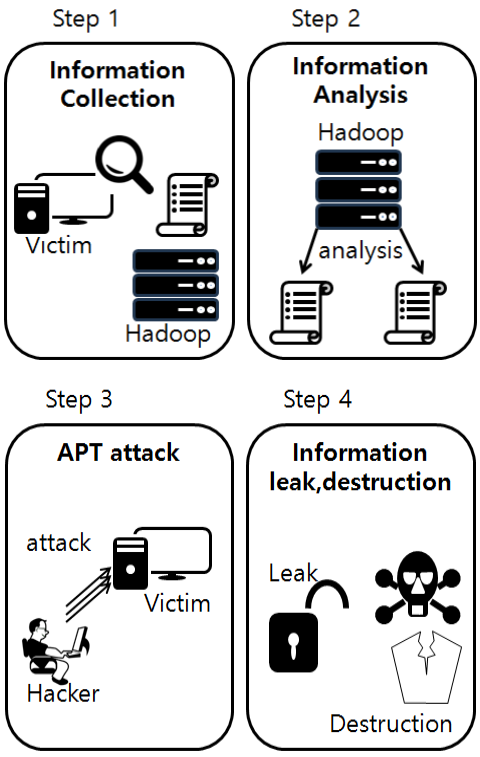

Fig. 6. APT Step Attack Scenario

① 공격자는 Hadoop을 이용하여 정보시스템에 대한 정보를 수집한다.

② Hadoop을 이용해 수집된 대규모의 정보를 분석한다. 분석된 정보를 통해 정보시스템에서 사용하는 시스템이나 프로그램 등을 알아내고, 그에 따른 취약점을 도출한다. Hadoop를 이용하여 정보시스템의 취약점을 짧은 시간 내에 도출한다. 도출된 취약점을 이용하여 해킹 시나리오를 작성한다.

③ 해커는 Hadoop을 이용해 만든 해킹 시나리오를 가지고 APT 공격을 수행한다.

④ 해킹하여 정보시스템을 장악하여 원하는 정보를 유출하거나 정보 파괴 등 해커가 의도한 특정 목적을 수행한다.

빅데이터를 이용한 APT 공격은 짧은 시간 내에 취약점을 찾아내고, 그에 따른 공격 시나리오를 작성하여 짧은 시간 내에 공격이 가능하다.

4. APT공격 차단 기법

APT공격을 효과적으로 막을 수 있는 방법은 기관내의 정보시스템의 접근 가능한 컨텐츠를 중요도에 따라 분류하고, 외부에서 접근 가능한 데이터와 접근이 가능하지 않는 데이터로 나누어 관리해야 한다. 침입공격이 성공했을 경우에도 중요한 데이터를 접근할 때 접근권한을 SSO 인증이 아닌 2 인증기법(2 factor)을 적용함으로써 접근을 차단해야 한다. 웹으로 접근 가능한 정보이외에도 DB에 저장된 정보에 대한 보호기법을 적용하는 것도 필요하다[9]. 공격자가 정보시스템의 취약점을 이용하여 DB관리자 권한을 취득한 경우에는 DB 의 모든 정보를 열람, 유출이 가능하기 때문이다[9-11]. Fig 7처럼 이를 해결하기 위해 민감하고, 중요한 데이터에 한하여 따로 암호화하여 저장하고, 암호화 키는 따로 저장함으로써 DB 관리자 권한으로 취득한 공격자가 정보 접근이나 유출을 막을 수 있다. 특히, 공격자의 공격기법이나 패턴, 유출피해정도 확인 등을 위해 DB의 접근이나 sql관련 명령어에 대한 로그 파일을 저장하고, 공격자가 로그파일을 접근할 수 없도록 분산저장하거나 백업파일을 물리적인 분리가 필요하다.

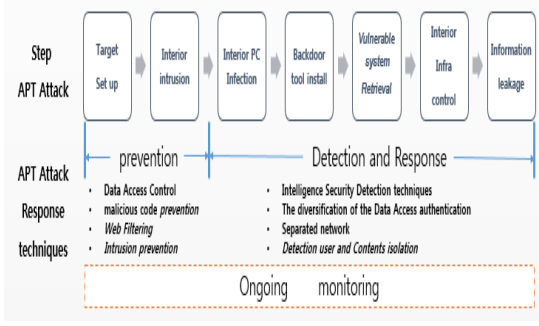

Fig. 7. APT Attack Step and Response

5. 결론

현재 빅데이터 기술은 불확실성, 리스크, 융합 등 미래사회에 대응하는 역할을 수행하면서 새로운 기회요인을 창출하고 있다[12,13]. 2011년 이머징 기술 하이프사이클(hype cycle)에서 Gartner는 빅 데이터를 기술발생 단계(technology trigger)로 분류하고 있다[14]. 다양한 단말기를 통해 생산되는 데이터를 이용하여 가치 있는

정보를 생성할 수 있는 빅 데이터는 마케팅, 스포츠, 의료, 금융 등 전 분야에서 사용되고 있다. 개인정보보호조치가 취해지지 않을 경우 빅 데이터에 개인의 위치, 진료, 금융 정보 등이 담겨있기 때문에 APT 공격에 노출될 수 있다.

빅 데이터 기술은 활용도가 급격하게 확장되어 공격자 역시 빅데이터 기술을 적절하게 이용할 경우 짧은 시간에 다양한 공격이 가능하다.

본 연구에서는 그 위험성을 알리고, 빅 데이터를 이용한 공격에 대한 대응방안 연구 필요성을 제기한다. 빅데이터를 통해, 공격뿐만 아니라 공격자의 패턴을 빅 데이터를 통해 찾아낼 경우 공격도구가 아닌 공격차단 도구로 활용이 가능하다.

향후 연구로는 빅 데이터를 이용한 공격자의 공격 패턴을 실시간으로 찾아내어 패턴기반 침입탐지시스템에 적용하는 것이 필요하다.

REFERENCES

[1] H. J. Mun, Y. C. Hwang, H. Y. Kim, "Countermeasure for Prevention and Detection against Attacks to SMB Information System - A Survey," *Journal of the Convergence Society for SMB*, Vol. 5, No. 2, pp. 1-6, Jun. 2015.

[2] Y. I. Cho, "Understanding and Major Issues of Big Data," *Journal of Korean Associastion for Regional Information Society*, Vol. 16, No. 3, pp. 43-65, 2013.

[3] S. O. Yun, "A Study on the Classification of Risks Caused by Big Data," *Journal of Korean Associastion for Regional Information Society*, Vol. 16, No. 2, pp. 93-122, 2013.

[4] The Big Data Blog, http://thebigdatablog.weebly.com/blog/the-hadoop-ecosystem-overview, 2014. 4.

[5] Hadoop ECO system, http://blrunner.com/18, 2012. 8.

[6] D. S. Park, *Big Data Computing Technology*, Hanbit, 2014.

[7] H. J. Lee, "Use of Big Data Hadoop Platform," *Journal of The Korean Institute of Communication Sciences*, Vol.29, No.11 pp. 43-47, 2012.

[8] K. -H. Lee, W. J. Park, K. S. Cho, W. Ryu, "The MapReduce framework for Large-scale Data Analysis: Overview and Research Trends", *Journal of Electronics and telecommunications trends(ETRI)*, Vol. 28, No. 6, pp. 156-166, Dec. 2013,

[9] H. J. Mun. "A Role based Personal Sensitive Information Protection with Subject Policy", *PhD Thesis of Computer Science Paper*. Chungbuk University, Korea, 2008.

[10] H. J. Mun, S. J. Oh, "Injecting Subject Policy into Access Control for Strengthening the Protection of Personal Information," *Wireless Personal Communications*, Vol. 89, Issue. 3. pp. 715-728, Aug. 2016.

[11] H. J. Mun, K. M. Lee, S. H. Lee, "Person -Wise Privacy Level Access Control for Personal Information Directory Services," *Embedded and Ubiquitous Computing, Volume 4096 of the series Lecture Notes in Computer Science*, pp. 89-98, 2006.

[12] NIA, *The World to evolve into a Big Data-Global Advanced Cases of Big Data, National Information Society Agency Big Data Strategy Research Center*, 2012. 5.

[13] NIA. *Big Data : Global teen Advanced Cases, National Information Society Agency Big Data Strategy Research Center*, 2012.

[14] Y. I. Cho, "The Big Data Technology and major of Issues in the smart era," *Institute of Control, Robotics and Systems*, Vol. 18, No. 4, pp. 23-33, 2012.

저 자 소 개

문 형 진(Hyung-Jin Mun) [정회원]

- 2008년 2월 : 충북대학교 전자계산학(이학박사)
- 2008년 3월 ~ 현재 : 백석대학교 강사

<관심분야> : 프라이버시 보호, 네트워크 및 웹 보안

최 승 현(Seung-Hyeon Choi) [학생회원]

- 2015년 3월 ~ 현재 : 백석대학교 정보통신학부 재학

<관심분야> : 정보통신, 네트워크

황 윤 철(Yooncheol Hwang) [정회원]

- 2008년 2월 : 충북대학교 전자계산학(이학박사)
- 2000년 3월 ~ 현재 : 한국교통대학교(경기도 의왕) 외래교수

<관심분야> : 네트워크 및 웹보안, 침입방지시스템