http://dx.doi.org/10.14400/JDPM.2013.11.12.393

빅데이터 환경에 적합한 보안 인식 정보를 이용한 서비스 관리 기법

정윤수*, 한군희**
목원대학교 정보통신공학과*, 백석대학교 정보통신공학과**

Service Management Scheme using Security Identification Information adopt to Big Data Environment

Yoon-Su Jeong*, Kun-Hee Han**
Dept. of Information Communication & Engineeringe, Mokwon University*
Dept. of Information Communication & Engineeringe, Baeseok University**

요 약 최근 클라우드 환경에서 처리되고 있는 데이터의 양과 종류가 다양해지면서 서로 다른 네트워크 환경에 존재하는 이기종 장치에 저장된 빅 데이터에 손쉽게 접근하기 위한 방법이 요구되고 있다. 이 절에서는 클라우드 환경에서 빅 데이터를 사용하는 사용자의 프라이버시와 데이터를 보호하기 위해 사용자와 서버간 공유된 키를 부분키로 할당하여 빅 데이터와 용자의 속성정보를 연계하여 사용자가 다른 네트워크 환경에서 빅 데이터에 접근하는 것을 원활하게 하기 위한 보안 관리 기법을 제안한다. 제안 기법은 사용자가 생성한 임의의 비트 신호가 제3자에게 도청되거나 변조되더라도 높은 안전성을 가지며, 충분한 임의의 비트를 전달하여 사용자 보안 인식 정보를 공유하는데 사용한다. 또한, 보안 인식 정보를 생성하는 비트 수열이 제3자에게 불필요하게 노출되지 않도록 해쉬 체인한 값을 전달함으로써 사용자의 익명성을 보장받도록 하고 있다.

주제어 : 빅데이터, 접근제어, 보안, 인증

Abstract Recently, the quantity and type of data that is being processed in cloud environment are varied. A method for easy access in different network in a heterogeneous environment of big data stored in the device is required. This paper propose security management method for smoothly access to big data in other network environment conjunction with attribute information between big data and user. The proposed method has a high level of safety even if user-generated random bit signal is modulated. The proposed method is sufficient to deliver any number of bits the user to share information used to secure recognition. Also, the security awareness information bit sequence generated by a third party to avoid unnecessary exposure value by passing a hash chain of the user anonymity is to be guaranteed to receive.

Key Words : Big-data, Access control, Seucrity, Authentication

Received 25 October 2013, Revised 20 November 2013
Accepted 20 December 2013
Corresponding Author: Kun-Hee Han(Baeseok University)
Email: hankh@bu.ac.kr

ISSN: 1738-1916

1. 서론

최근 클라우드 환경에서 처리되고 있는 데이터가 이기종 장치에 저장되어 서로 다른 네트워크 환경에서 손쉽게 사용할 수 있는 빅데이터 서비스가 각광을 받고 있다[1,2]. 특히, 다양한 종류의 대규모 데이터에 대한 생성, 수집, 분석, 표현을 그 특징으로 하는 빅 데이터 기술의 발전은 다변화된 현대 사회를 더욱 정확하게 예측하여 효율적으로 작동케 하고 개인화된 현대 사회 구성원 마다 맞춤형 정보를 제공, 관리, 분석 가능케 하며 과거에는 불가능했던 기술을 실현시키기도 한다[3,4].

빅 데이터는 TB(테라바이트)단위의 데이터량으로 정의되거나 데이터 수집 및 분석에 장기적인 시간을 요하므로 데이터 양의 증가를 그 특징으로 하고 있다. 그러나 단순한 데이터 양의 증가를 넘어서서 빅 데이터는 크게 데이터 양(volume), 데이터 속도(velocity), 그리고 데이터 다양성(variety) 등 세 가지 요소의 복합적인 변화를 그 특징으로 한다[5]

빅 데이터는 정치, 사회, 경제, 문화, 과학 기술 등 전 영역에 걸쳐서 사회와 인류에게 가치있는 정보를 제공할 수 있는 가능성을 제시하며 그 중요성이 부각되고 있다. 그러나, 빅데이터의 문제점은 바로 사생활 침해와 보안 측면에 자리하고 있다[6]. 빅데이터는 수많은 개인들의 수많은 정보의 집합이다. 그렇기에 빅데이터를 수집,분석할 때에 개인들의 사적인 정보까지 수집하여 관리하는 빅브라더의 모습이 될 수도 있다. 그리고 수집된 데이터가 보안 문제로 유출된다면, 거의 모든 사람들의 정보가 유출되는 것이기 때문에 사회적으로 큰 문제가 야기될 수 있다.

본 논문에서는 타 네트워크로 이동하는 사용자가 이전 네트워크에서 제공받던 빅 데이터 서비스를 끊김없이 계속 서비스 받기 위한 서비스 관리 기법을 제안한다. 제안 기법은 사용자가 생성한 임의의 비트 수열을 해쉬체인하여 사용자 인덱스 값과 XOR 한 사용자 보안 인식 정보를 타 네트워크에 등록하여 사용자에게 제공되던 빅 데이터 서비스를 지속적으로 제공한다. 제안 기법은 사용자가 생성한 임의의 비트 신호가 제3자에게 도청되거나 변조되더라도 높은 안전성을 가진다. 특히, 제안 기법은 충분한 임의의 비트를 전달하여 사용자 보안 인식 정보를 공유하는데 사용한다. 또한, 보안 인식 정보를 생성하는 비트 수열이 제3자에게 불필요하게 노출되지 않도록 해쉬 체인한 값을 전달함으로써 사용자의 익명성을 보장받도록 하고 있다.

이 논문의 구성은 다음과 같다. 2장에서는 빅데이터의 정의 및 특징을에 대해서 알아본다. 3장에서는 보안 인식 정보를 이용한 이동 사용자의 빅 데이터 서비스 제공 기법을 제안하고, 4장에서는 제안 기법의 보안평가와 성능평가를 분석하고 마지막으로 5장에서 결론을 맺는다.

2. 관련연구

2.1 빅데이터

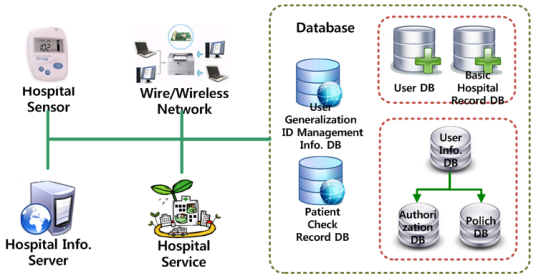
빅데이터란 과거 아날로그 환경에서 생성되던 데이터에 비해 그 규모가 방대하며 생성 주기가 짧고, 형태가 수치 데이터 뿐만 아니라 문자와 영상 데이터를 포함하는 대규모 데이터를 의미한다[1]. 최근 PC와 인터넷, 모바일 기기 등이 생활화 되면서 시간과 장소에 구애받지

〈Table 1〉 Character of big data environment

Division	Standard	big data environment
data	-Structured numerical data center	-Unstructured data of various -Characteristic data(SMS, Serarch keyword) -Location data
hardware	-Storage of high price -Database -Data-warehouse	-Possibility of cost-effective equipment usage such as cloud computing
software/analysis method	-RDMBS -Statistic package(SAS, SPSS) -Data mining -machine learning, knowledge discovery	-Free open-source software -Hadoop, NoSQL -Open source statistic solution(R) -text mining -Online buzz analysis(opinion mining) -sentiment analysis

않고 손쉽게 사이버 공간에서 사용 및 저장한 데이터가 기하급수적으로 증가하고 있다. 이 같은 현상은 사람과 기계, 기계와 기계가 서로 정보를 주고받는 사물지능통신(M2M, Machine to Machine)의 확산도 디지털 정보가 폭발적으로 증가하게 된 이유이다[7,9].

사용자가 직접 제작하는 UCC를 비롯한 동영상 컨텐츠, 휴대전화와 SNS(Social Network Service)에서 생성되는 문자 등은 데이터의 증가 속도뿐만 아니라, 형태와 질에서도 기존과 다른 양상을 보이고 있다. 특히, 블로그나 SNS에서 유통되는 텍스트 정보는 내용을 통해 글을 쓴 사람의 성향뿐만아니라 소통하는 상대방의 연결 관계까지도 분석이 가능하다. 또한 주요 도로와 공공건물은 물론 심지어 아파트 엘리베이터 안에까지 설치된 CCTV가 촬영하고 있는 영상 정보도 데이터로 저장되고 있다. 그리고, 민간 분야뿐만 아니라 공공 분야도 데이터를 양상 중인데 센서스(Census)를 비롯한 다양한 사회 조사, 국제자료, 의료보험, 연금 등의 분야에서 데이터가 생산되고 있다[10].



[Fig. 1] U-healthcare Concept with implantable device

2.2 빅데이터 특징

빅데이터는 일반적으로 3V,데이터의 양(Volume), 데이터 생성 속도(Velocity), 형태의 다양성(Variety) 등의 특징을 가진다. 빅데이터의 다양하고 방대한 규모의 데이터는 국가 경쟁력의 우위를 좌우하는 중요한 자원으로 활용되고 있지만 과거와 비교해 데이터의 양은 물론 질과 다양성 측면에서 패러다임의 전환이 필요하다.

빅데이터는 분산처리방식과 같은 기술을 활용해서 과거에 비해 대규모 고객정보를 빠른 시간 안에 분석하는 것이 가능해졌다. 트위터와 인터넷에서 생성되는 기업 관련 검색어와 댓글을 분석해 자사의 제품과 서비스에 대한 고객 반응을 실시간으로 파악해 즉각적인 대처를 수행할 수 도 있다.

빅데이터에서는 소프트웨어나 하드웨어도 오픈 소스 형태의 하둡(Hadoop)이나 분석용 패키지인 R 과 분석병렬처리기술, 클라우드 컴퓨팅 등을 활용하기 때문에 기존의 비싼 스토리지와 데이터베이스에 기반한 고비용의 데이터웨어하우스를 구축하지 않아도 효율적인 시스템 운용이 가능하다[4,8].

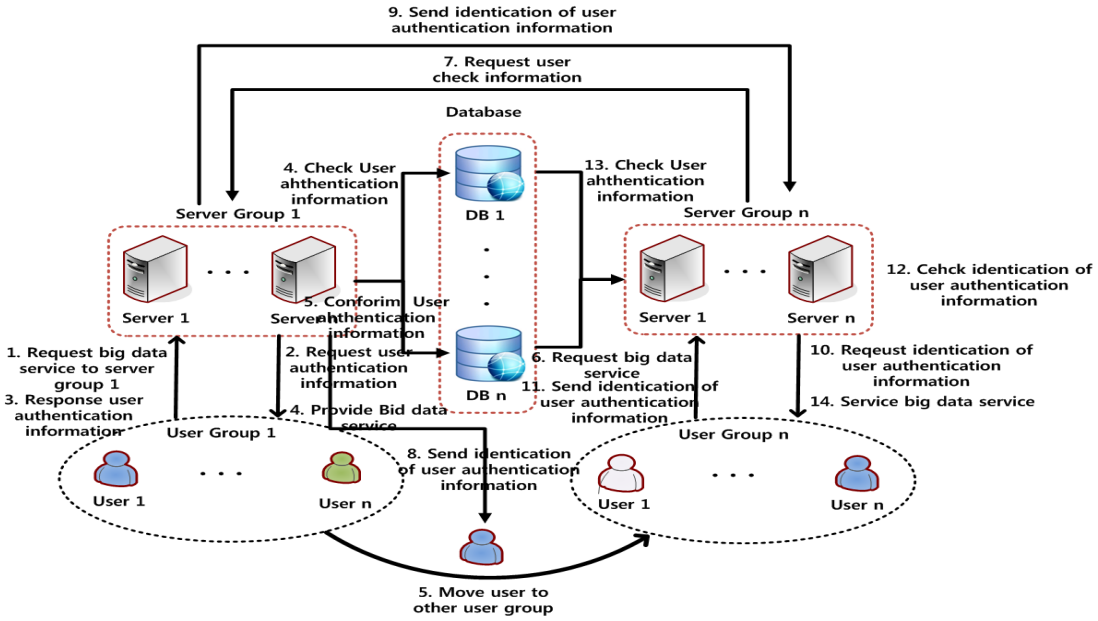
3. 보안인식정보를 이용한 이동 사용자의 빅 데이터 서비스 관리 기법

최근 클라우드 환경에서 처리되고 있는 데이터의 양과 종류가 다양해지면서 서로 다른 네트워크 환경에 존재하는 이기종 장치에 저장된 빅 데이터에 손쉽게 접근하기 위한 방법이 요구되고 있다. 이 절에서는 클라우드 환경에서 빅 데이터를 사용하는 사용자의 프라이버시와 데이터를 보호하기 위해 사용자와 서버간 공유된 키를 부분키로 할당하여 빅 데이터와 용자의 속성정보를 연계하여 사용자가 다른 네트워크 환경에서 빅 데이터에 접근하는 것을 원활하게 하기 위한 보안 관리 기법을 제안한다.

3.1 개요

클라우드 환경에서는 사용자가 언제, 어디서나 사용자가 원할 때 빅 데이터 서비스를 제공받아야 한다. 이 때, 빅 데이터를 제공하는 서버는 하나 이상이어야 하며, 서버의 위치 또한 동일하지 않아야 한다. 이 환경에서 사용자는 사용자가 원하는 빅 데이터 서비스를 하나의 특정 서버 지정하지 않고 사용자가 원하는 위치에서 원하는 빅 데이터를 서비스 받아야 한다.

그림 2은 제안 기법의 전체 동작을 나타내고 있으며, 각 서버는 서로 다른 네트워크에서 동기화가 이루어지면서 서비스를 제공한다. 사용자가 서로 다른 네트워크에서 빅 데이터 서비스를 제공받으려고 할 때 사용자는 이전 서버에서 인증 정보를 해쉬한 보안 인식자를 전달받아 사용자를 인증받아 빅데이터의 서비스를 제공받는다.



[Fig. 2] Overall process of proposal scheme

3.2 용어

<Table 1>]은 제안 모델에서 사용되는 용어를 정의하고 있다.

〈Table 1〉 Definition

Parameter	Definition
i	index number
s_i	i^{th} bit number
h_i	i^{th} hash chain
idx	User index value
SID	User Security Informaiton
SE_i, SE_j	Server

3.3 보안 인식자를 이용한 사용자 인증

서로 다른 네트워크 환경에서 빅데이터 서비스를 제공 받는 사용자를 서버가 안전하게 인식하기 위해서 제안 기법에서는 사용자의 보안 인식자를 최소 사용자 인증 과정에서 생성하여 사용자가 다른 네트워크에서 동일한 빅데이터 서비스를 제공받도록 한다. 다른 네트워크에 속한 서버가 이동한 사용자를 인식할 수 있도록 서버는 이전 서버에서 사용자 보안 인식자를 전달받도록 서버간 동기화를 수행하여 사용자를 인증한다.

3.3.1 초기화 과정

초기화 과정은 사용자가 빅데이터 서비스를 제공하는 서버에게 사용자 정보를 등록하는 단계로써, 사용자는 서버에게 사용자 인증과 관련된 정보를 전달한다. 초기화 과정은 2단계로 구성된다.

· 1단계 :

사용자는 빅데이터 서비스를 요청하기 위해서 SSL(Secure Socket Layer) 연결이 확립되면 사용자는 서버에게 사용자 정보(인식자, 패스워드 등)를 전달한다.

· 2단계 :

사용자는 환자는 임의의 비트 수열(0과 1로 구성된 수열)을 식 (1)처럼 N개 생성하여 서버에게 전달한다.

$$Generate \quad s \cong \{0,1\}^{*} \rightarrow \{0,1\}^{N} \qquad (1)$$

서버는 사용자로부터 전달된 임의의 비트 수열을 식

(2)처럼 해쉬체인 h_i로 생성한다. 이때, 서버는 해쉬체인 h_i로 생성된 값을 무작위로 선택한다.

$$\{h_i|s_i,\ i\in N\} \quad (2)$$

· 3단계 :

서버는 생성된 해쉬체인 h_i과 사용자 인덱스 값 idx과 함께 식 (3)처럼 XOR한 사용자 보안 정보를 데이터베이스에 저장한다.

$$SID = h_i \oplus idx \quad (3)$$

3.3.2 타 네트워크로 사용자 이동 과정

타 네트워크로 사용자 이동 과정에서는 사용자가 이전 네트워크에서 제공받고 있던 빅 데이터 서비스를 이동한 네트워크에서도 원활하게 빅 데이터 서비스를 제공받기 위한 과정이다. 이 과정에서는 이전 네트워크에서 빅 데이터 서비스를 제공받을 때 등록했던 보안 인식 정보 *SID*를 이전 서버 SE_i로부터 전달받은 후 새로운 서버 SE_j에게 전달한다.

· 1단계 :

서버는 사용자의 보안 인식정보를 사용자와 사용자가 이동하려는 타 네트워크내의 서버 SE_j에게 각각 전달한다.

· 2단계 :

사용자는 빅 데이터 서비스를 지속적으로 서비스받기 위해서 사용자의 보안 인식 정보 *SID*를 서버 SE_j 에게 전달한다.

· 3단계 :

서버 SE_j는 사용자로부터 전달받은 사용자의 보안 인식 정보 *SID* 가 정상적인 정보인지 확인하기 위해서 이전 서버 SE_i의 데이터베이스에 저장되어 있는 사용자의 보안 인식 정보 *SID*를 요청한다.

3.3.3 검증 과정

검증 단계는 사용자와 서버 SE_j사이에 사용자의 보안 인식 정보 *SID* 가 정상적인 정보인지 검증하여 사용자가 정상적인 사용자일 경우 빅 데이터 서비스를 지속적으로 서비스하는 단계이다. 사용자와 서버가 동일한 보안 인식 정보 *SID*를 사용하여 중간에 도청이 없었다면 빅 데이터 서비스를 그대로 제공하여도 무방하다.

· 1단계 :

서버는 이전 서버 SE_i로부터 사용자의 보안 인식 정보 *SID*를 전달받는다. 이 때, 서버 SE_j는 사용자의 보안 인식 정보 *SID*가 정상적인 정보인지를 검증하기 위해서 수집된 비트 수열 s 중에서 일부를 사용자에게 공개한다.

· 2단계 :

사용자는 서버가 공개한 비트 수열 정보 s를 확인한다.

$$Check \quad s \cong \{0,1\}^* \rightarrow \{0,1\}^N \quad (3)$$

· 3단계 :

사용자는 서버가 공개한 비트 수열 값 s이 정상이면 비트 수열 값 s와 보안 인식 정보 *SID*를 해쉬하여 서버에게 전달한다.

$$h(s \oplus SID) \quad (4)$$

· 4단계 :

서버는 사용자로부터 전달받은 $h(s \oplus SID)$를 확인하여 서버가 사용자에게 전달한 비트 수열 값과 사용자 보안 인식 정보를 비교하여 정상이면 빅 데이터 서비스를 사용에게 제공한다.

$$Check \quad s \quad (5)$$
$$Check \quad SID \quad (6)$$

4. 평가

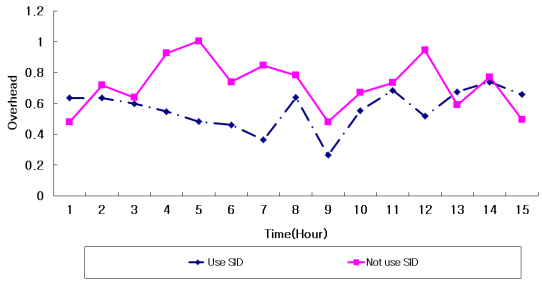
4.1. 보안평가

제안 모델은 사용자와 서버사이에 교환되는 사용자 보안 인식 정보 *SID*를 이용하여 이기종 네트워크간에 빅 데이터 서비스를 사용자가 끊김없이 서비스 받도록 무결

성을 보장받는다. 특히 제안 모델에서는 제3자가 보안 인식 정보 *SID*를 도청하더라도 사용자가 임의로 생성한 비트 수열 s를 알 수 없기 때문에 보안 인식 정보 *SID*를 서버에 전달하더라도 서버가 비트 수열 s과 보안 인식 정보 *SID*를 비교분석하기 때문에 빅 데이터 서비스의 무결성을 보장받는다. 빅 데이터를 제공받는 모든 사용자들은 보안 인식 정보 *SID*를 이용하여 사용자 인증이 수행되기 때문에 사용자와서버 사이에 공유된 키는 제3자에게 노출되지 않으면서 안전하게 사용되어 안전성과 신뢰성이 보장받는다.

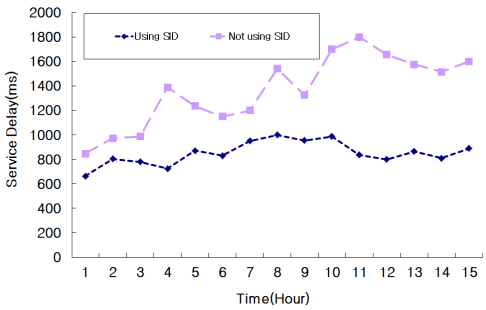
4.2. 성능평가

제안 기법에서는 이기종 네트워크에서 빅 데이터 서비스를 제공받는 사용자와 서버 사이의 통신 오버헤드와 지연시간을 평가하기 위해서 OPNET 시뮬레이터를 이용하였다. 그림 3처럼 제안 기법은 보안 인식 정보 *SID*의 사용 유무에 따른 통신 오버헤드를 평가한 결과, 보안 인식 정보 *SID*를 사용하지 않았을 때보다 보안 인식 정보 *SID*를 사용하였을 때가 통신 오버헤드가 5.6% 낮게 나타났다. 이같은 결과는 보안 인식 정보 *SID*를 사용하지 않았을 경우 사용자를 인증하기 위해 사용되는 정보가 추가로 사용되기 때문이다. 그러나 실험에 사용된 정보는 사용자를 인식하기 위해 최소 정보를 사용하였기 때문에 전체 처리량에 비해 큰 차이를 보이지 않았다.



[Fig. 3] Communication overhead between Using SID and Not using SID

그림 4는 사용자와 서버 사이에 빅데이터 서비스 지연에 대한 결과이다. 실험 결과 시간대별 서비스 지연은 980ms 차이를 보이고 있지만 평균 서비스 지연 시간은 평균 235ms로 큰 차이가 나지 않는 결과를 보였다.



[Fig. 4] Service Delay between User and Server

5. 결론

최근 빅데이터 서비스를 제공받는 사용자가 급속하게 증가하는 추세에서 서로 다른 네트워크에서 사용자의 빅 데이터 서비스 제공의 필요성이 요구되고 있다. 본 논문에서는 사용자의 보안 인식 정보를 이용하여 서로 다른 네트워크 환경에서 사용자가 서비스 받는 빅 데이터를 끊김없이 서비스 받기 위한 관리기법을 제안하였다. 제안 기법은 사용자와 서버 사이에서 임의의 비트 수열을 제3자가 도청하더라도 사용자와 서버간 쌍방 인증을 수행하기 때문에 안전성이 높게 나타났다. 향후 연구에서는 사용자와 서버 사이에 송수신되는 보안인식정보의 속성을 계층화하여 통합관리할 수 있는 모델을 설계 및 운영할 예정이다.

REFERENCES

[1] J. Manyika and M. Chui(2011), “Big data: The next frontier for innovation, competition, and productivity”, McKinsey Global Institute, pp. 1.

[2] Global Agenda Council on Emerging Technologies(2012), “The top 10 emerging technologies for 2012”, World Economic Forum.

[3] P. Russom(2011), “Big Data Analytics”, TDWI Research Fourth Quarter, pp. 6.

[4] O’Reilly Radar Team(2012). “Planning for Big Data”, O’Reilly.

[5] Y. C. Jung(2012). “Big Data revolution and media policy issues”, KISDI Premium Report, Vol. 12, No. 2, pp. 1-22.
[6] C. Tankard(2012), “Big Data Security”, Network Security, pp. 5-8.
[7] S. Y. Son(2013), “Big data, online marketing and privacy protection”, KISDI Premium Report, Vol. 13, No. 1, pp.1-26.
[8] CDWG(2013), “Proactive planning for big data”, CDWG - people who get it, pp. 1-8.
[9] A. Lane(2012), “Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments”, Securosis, pp. 1-16.
[10] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers(2011), “Big Data: The Next Frontier for Innovation, Competition and Productivity”, Mckinsey Global Institute, pp. 1-137.
[11] J. T. Kim, B. J. Oh, J. Y. Park(2013), “Standard Trends for the BigData Technologies”, 2013 Electronics and Telecommunications Trends, Vol. 28, No. 1, pp. 92-99.

정 윤 수(Jeong, Yoon Su)

・2000년 2월 : 충북대학교 대학원 전자계산학 이학석사
・2008년 2월 : 충북대학교 대학원 전자계산학 박사
・2009년 8월 ~ 2012년 2월 : 한남대학교 산업기술연구소 전임연구원
・2012년 3월 ~ 현재 : 목원대학교 정보통신공학과 조교수
・관심분야 : 센서 보안, 암호이론, 정보보호, Network Security, 이동통신보안
・E-Mail : bukmunro@gmail.com

한 군 희(Han, Kun Hee)

・2000년 2월 : 충북대학교 컴퓨터공학과(공학박사)
・2001년 3월 ~ 현재 : 백석대학교 정보통신학부 교수
・관심분야 : 멀티미디어, 정보보호
・E-Mail : hankh@bu.ac.kr