Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.30, NO.3, Jun. 2020 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.3.397

윈도우에서의 말랑말랑 톡카페 데이터베이스 암호화 프로세스 분석 및 삭제된 메시지 복구 연구*

윤 병 철,[1†] 김 소 람,[1] 김 종 성[2‡]
[1,2]국민대학교(대학원생, 교수)

Study on MalangMalang Talkafe Database Encryption Process and Recovering Its Deleted Messages on Windows*

Byungchul Youn,[1†] Soram Kim,[1] Jongsung Kim[2‡]
[1,2]Kookmin University (Graduate student, Professor)

요 약

실시간 대화, 멀티미디어, 파일 및 연락처 공유 서비스의 편리함으로 대다수 사용자는 인스턴트 메신저를 사용하며, 제공하는 기능이 다양해짐에 따라 메신저 이용 시간이 증가하고 있다. 이러한 이유로 메신저는 많은 양의 사용자의 행위 정보를 포함하고 있다. 따라서 디지털 포렌식 관점에서는 메신저 데이터를 사용자 행위 파악을 위한 유용한 증거물로 활용할 수 있다. 그러나 개인정보보호를 목적으로 중요 정보는 암호화하여 저장하기 때문에 증거로 사용하기 어려우며, 사용자가 고의로 메신저 데이터를 삭제한 경우에는 데이터 획득이 불가능하다. 따라서 메신저 데이터를 증거로 사용하기 위해서는 암호화된 메시지 복호화와 삭제된 메시지 복구 연구는 필수적으로 선행되어야 한다. 본 논문에서는 윈도우 환경에서 인스턴트 메신저인 말랑말랑 톡카페의 데이터베이스 암호화 프로세스를 분석하였으며, 복호화에 성공하였다. 또한, 삭제된 메시지를 식별하고 이를 휘발성 메모리 영역에서 복구하였다.

ABSTRACT

With the convenience of real-time conversation, multimedia file and contact sharing services, most people use instant messenger, and its usage time is increasing. Because the messengers contain a lot of user behavior information data, in the digital forensic investigation, they can be very useful evidence to identify user behavior. However, some of useful data can be difficult to acquire or recognize because they are encrypted or deleted. Thus, in order to use the messenger data as evidence, the study of message decryption process and message recovery is essential. In this paper, we analyze the database encryption process of the instant messenger, MalangMalang Talkafe, and propose the method to decrypt it. In addition, we propose the methods to identify the deleted messages and recover from the volatile memory area.

Keywords: Digital Forensic, Messenger, Decryption, Reverse Engineering, Memory Forensic

I. 서 론

인스턴트 메신저란 네트워크 통신을 이용하여 두 명 이상의 사용자가 즉각적 통신을 가능하게 하는 애플리케이션이다. 현재 현대인의 중요 통신수단으로 자리 잡았으며, 사용자의 계정 정보, 대화 내용 및

Received(03. 03. 2020), Modified(04. 29. 2020), Accepted(05. 04. 2020)
* 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2017-0-00520, SCR-Friendly 대칭키 암호 및 응용모드 개발)
† 주저자, qjcjf123@kookmin.ac.kr
‡ 교신저자, jskim@kookmin.ac.kr(Corresponding author)

일정과 같은 중요한 정보를 포함하고 있다. 이와 같은 정보는 디지털 포렌식 관점에서 사용자의 행위를 유추하기 위한 핵심 증거로 사용된다. 그러나 개인정보보호를 목적으로 보안 기능이 강화되면서 스마트폰을 통한 데이터 획득은 점차 어려워지고 있다.

스마트폰을 중심으로 제공되던 서비스는 데스크톱으로 확장되어 모바일과 동일한 기능을 이용할 수 있다. 또한, PC와 스마트폰의 데이터가 동기화되므로 스마트폰을 통해 데이터 획득이 어려운 경우에는 PC에서 스마트폰에 저장된 것과 동일한 데이터를 수집할 수 있다. 따라서 PC 메신저 데이터 획득 및 분석은 중요한 의미를 가진다. 그러나 메신저 프로그램 자체에서 데이터 암호화를 제공하고 있으므로, 실제 데이터를 획득하기 위해서는 메신저의 데이터 암호화 프로세스 분석이 동반되어야 한다.

본 논문의 구성은 다음과 같다. 2장에서는 인스턴트 메신저 관련 연구에 대해 언급하며, 3장에서는 암호화 대상 파일을 식별한다. 4장에서는 암호화된 데이터베이스의 복호화 방안을 제시하고 5장에서는 삭제된 메시지 복구에 관한 연구를 다룬다. 마지막 6장에서는 결론으로 마무리한다.

II. 관련 연구

기존 인스턴트 메신저에 관한 연구는 모바일 기기와 데스크톱 환경에서 사용자의 개인정보 및 대화 내역이 저장된 데이터베이스 복호화와 아티팩트 분석에 관한 연구가 진행되었다.

윈도우 환경에서 카카오톡 메신저의 백업 프로세스를 분석하여, 백업 데이터베이스 암호화 과정과 암호키 생성과정을 밝혀냈다[1]. 또한, 카카오톡, 네이트온 그리고 QQ 메신저의 데이터베이스 암·복호화 동작 과정 분석을 통해 사용자의 패스워드 없이 암호화된 데이터베이스 복호화에 성공하였다[2].

윈도우 환경 이외에도 모바일 앱에 대한 분석이 꾸준히 진행되고 있다. WhatsApp과 Viber 앱을 대상으로 다양한 기기와 안드로이드 버전에서 아티팩트 분석을 수행하였다[3]. 안드로이드 환경에서 말랑말랑 톡카페의 패스프레이즈 생성과정 및 암호화 프로세스 분석을 통해 암호화된 데이터베이스 복호화 방안을 제시하였다[4]. WeChat 메신저의 데이터베이스 파일을 분석하여 데이터의 파일 포맷, 경로 및 구조를 밝혔다[5,6]. 마지막으로 스마트폰의 휘발성, 비휘발성 메모리와 내부 저장소에 저장된 WhatsApp 메신저 데이터 추출에 관한 연구와 삭제된 데이터를 복구하는 방안에 관한 연구가 진행되었다[7].

III. 말랑말랑 톡카페 (PC 버전)

데스크톱 전용 말랑말랑 톡카페는 한글과 컴퓨터에서 개발한 인스턴트 메신저이며, 프로그램 설치 시 %LOCALAPPDATA%\Talkafe 위치에 사용자의 데이터를 저장 및 관리한다. 사진, 동영상과 같은 첨부 파일은 /cache 폴더에서 모두 관리한다. 사진의 경우 원본 파일과 썸네일을 동시에 저장하며, 동영상은 썸네일 파일만 저장된다. 메시지 내역과 같은 사용자의 개인정보는 데이터베이스에 저장되며, 개인정보보호를 위하여 데이터베이스가 암호화하여 관리한다. 암호화된 데이터베이스 파일명과 저장 경로는 다음과 같다(Table 1). generalSettings 데이터베이스(이하 Setting 데이터베이스)는 메신저 실행 후 최초 암호화되는 데이터베이스로써 사용자의 프로필 정보와 해당 애플리케이션에 관한 정보가 저장된다.

64바이트 길이의 헥사 값의 파일명을 갖는 데이터베이스 (이하 Chat 데이터베이스)에는 대화 내역과 같은 사용자의 개인정보가 저장된다. 해당 데이터베이스의 파일명 생성과정은 사용자의 고유 ID를 암호화한 값의 16진수 문자열이다. 암호화 방식은 이후 언급할 패스프레이즈 생성과정과 유사하다.

Table 1. List of Encrypted Databases

Name	Path
general Settings	%LOCALAPPDATA%\Talkafe\ settings\
64 Bytes Hex Strings	%LOCALAPPDATA%\Talkafe\

IV. 메신저 데이터 복호화

본 장에서는 메시지 데이터를 저장하고 있는 데이터베이스 복호화를 위해 메신저 프로그램 역공학을 수행한다. 이를 기반으로 암호화 프로세스를 식별하고 데이터 암호화에 사용한 패스프레이즈와 AES 키를 복구한다.

Table 2. Analysis Tool and Environment

Category		Name and Version
Virtual PC		VMware Workstation 12
OS		Windows 7 SP1 x86
Software		Talkafe 1.5.7
Analysis Tool	Dynamic & Static Analysis Tool	Ollydbg 1.1
		IDA Pro 6.8
	Memory Capture Tool	FTK Imager 3.2.0
	Hex Viewer	HxD 2.3.0
	SQLite Viewer	DB Browser 3.11.2

4.1 분석 환경

프로그램을 분석하기 위해 구축한 분석 환경 및 사용한 도구는 Table 2와 같다. 데이터베이스 암호화에 사용된 암호키 복구 및 데이터베이스 복호화를 진행하기 위해 Windows 7 가상환경을 구축하였으며, Ollydbg와 IDA Pro를 사용해 정적 및 동적 분석을 진행하였다. 또한, 메신저와 서버가 주고받는 네트워크 패킷을 분석하기 위해 WireShark를 사용하였으며, FTK Imager를 사용하여 메모리 포렌식적 관점에서 삭제된 메시지 복구 방안에 관한 연구를 진행하였다. 마지막으로 암호화된 데이터베이스의 구조 및 내부 데이터를 확인하기 위해 HxD 헥사 뷰어와 DB Browser를 사용했다.

4.2 AES 키 생성 및 데이터베이스 암호화 방식

데이터베이스 암호화 알고리즘은 AES-256-CBC이며, SQLCipher를 사용한다. 암호키는 패스프레이즈를 기반으로 생성하며, PBKDF2-HMAC-SHA1 알고리즘을 이용한다. PBKDF2-HMAC-SHA1 알고리즘 동작에 필요한 반복 횟수와 암호키 길이는 각각 64,000, 256이며, salt 값은 랜덤하게 생성된다. 암호화된 데이터베이스 내부에는 AES 키 생성 시 필요한 salt 값과 암호화 시 사용된 IV (Initial Vector) 값이 평문으로 저장된다. 따라서 암호화된 데이터베이스 파싱을 통해 salt 값과 IV 값을 획득한다면 AES 키 복구 및 암호화된 데이터베이스 복호화가 가능하다. (Fig. 1)은 데이터베이스 복호화 동작 과정을 나타낸다.

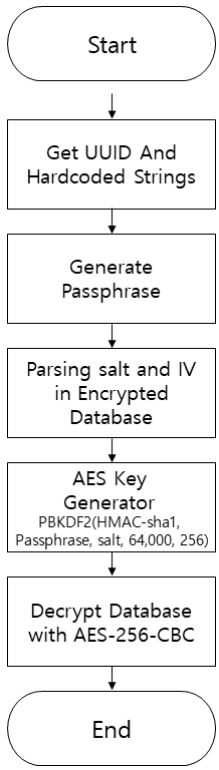

Fig. 1. Decrpytion Process

4.3 패스프레이즈 생성과정

Chat 데이터베이스와 Setting 데이터베이스 암호화에 사용되는 패스프레이즈 생성과정은 Fig. 2와 같다. 패스프레이즈는 사용자 PC의 UUID(Universally Unique Identifier)를 기반으로 생성된다. 따라서 메신저가 설치된 PC의 UUID를 획득하면 데이터베이스 복호화가 가능하다. UUID를 입력받으면 SHA256 해시 알고리즘으로 해싱된 값을 16진수

Input : $UUID(u)$
Output : *CHAT Database Passphrase*(p)
Variable : $Strings(S) = \{S_1, S_2, S_3, S_4\}$
1 : $i \leftarrow 0$
2 : $T \leftarrow Hex\,ToString(SHA256(u))$
3 : $T \leftarrow Hex\,ToString(SHA256(T))$
4 : *while*$(i < 3)$:
5 : if $(i == 2)$:
6 : $p \leftarrow Relocate\ T$ and $S_{i+1} + S_{i+2}$
7 : $T \leftarrow Hex\,ToString(SHA256(p))$
8 : *Break*
9 : *End* if
10 : $p \leftarrow String\ Relocation\ T$ and S_{i+1}
11 : $T \leftarrow Hex\,ToString(SHA256(p))$
12 : $i \leftarrow i+1$
13 : *End while*
14 : *Return* p

Input : $UUID(u)$
Output : *Setting Database Passphrase*(p)
Variable : $Strings(S) = \{S_0, S_1, S_2, S_3, S_4\}$
1 : $i \leftarrow 0$
2 : $T \leftarrow Hex\,ToString(SHA256(u))$
3 : $p \leftarrow Relocate\ T$ and S_i
4 : *while*$(i < 3)$:
5 : if $(i == 2)$:
6 : $p \leftarrow Relocate\ T$ and $S_{i+1} + S_{i+2}$
7 : $T \leftarrow Hex\,ToString(SHA256(p))$
8 : *Break*
9 : *End* if
10 : $p \leftarrow String\ Relocation\ T$ and S_{i+1}
11 : $T \leftarrow Hex\,ToString(SHA256(p))$
12 : $i \leftarrow i+1$
13 : *End while*
14 : *Return* p

Fig. 2. Passphrase Generation Process of Chat(Left) and Setting(Right) Databases

문자열로 변환하는 과정을 진행한다. 이후 16진수 문자열과 애플리케이션 내부에 하드코딩된 문자열 간의 재배열 과정과 해싱 과정을 총 3번 반복한다. 첫 두 번의 과정에서는 해싱된 64바이트 16진수 문자열과 32바이트 크기의 고정된 문자열을 재배열하며, 마지막 과정에서는 두 개의 32바이트 문자열을 합친 64바이트 문자열을 사용하여 재배치와 해싱 과정을 진행한다. 사용된 고정 문자열의 종류 및 적용된 데이터베이스 목록은 Table 3과 같다.

Setting 데이터베이스의 경우 초기 두 번의 해싱 과정을 갖는 Chat 데이터베이스 패스프레이즈 생성 과정과 달리 해싱 과정과 문자열 재배치 과정을 각각 한번 진행하여 생성된다. 이후 과정은 Chat 데이터베이스의 패스프레이즈 생성과정과 동일하다.

4.4 데이터베이스 복호화 결과

복호화된 Chat 데이터베이스와 Setting 데이터베이스의 내부 중요 데이터는 Fig. 3과 같다. Chat 데이터베이스에는 chats, lastindex, moimAlarms, rooms, sqlite_sequence와 verifies같이 6종류의 테이블이 존재하며 특히 chat 테이블과 rooms 테이블에 포렌식 수사관점에서의 중요 데이터가 저장된다.

chats 테이블에는 송·수신한 메시지 내역을 포함한 시간 정보, 송신자 고유 ID와 메시지 타입 등의 정보들이 저장되며, rooms 테이블에는 채팅방 고유 번호, 마지막으로 전송된 메시지와 관련된 정보들이 저장된다. Setting 데이터베이스의 경우 사용자의 이메일 정보와 마지막 접근 시간 그리고 소프트웨어 버전 등의 정보가 저장된다.

Table 3. List of HardCoded Strings

	Applicable Database	Hard Coding String
S_0	Setting Database	99786AD8C70F4E6F899143100B5B8374
S_1	Setting Database & Chat Database	4o8〈{X2(Z68}Z′!ws7eV48lU%Tj〈I!fl
S_2		O6(fz′!7*_:_Fh2J5〈E\\y!6Lsw3vtV\\U
S_3		z~x|{X2HR$Vx*|N`j68}Z′!w%D:,y!fl
S_4		oZlN0boZvxJ|**x:9FqzLN`Bs75V48lU

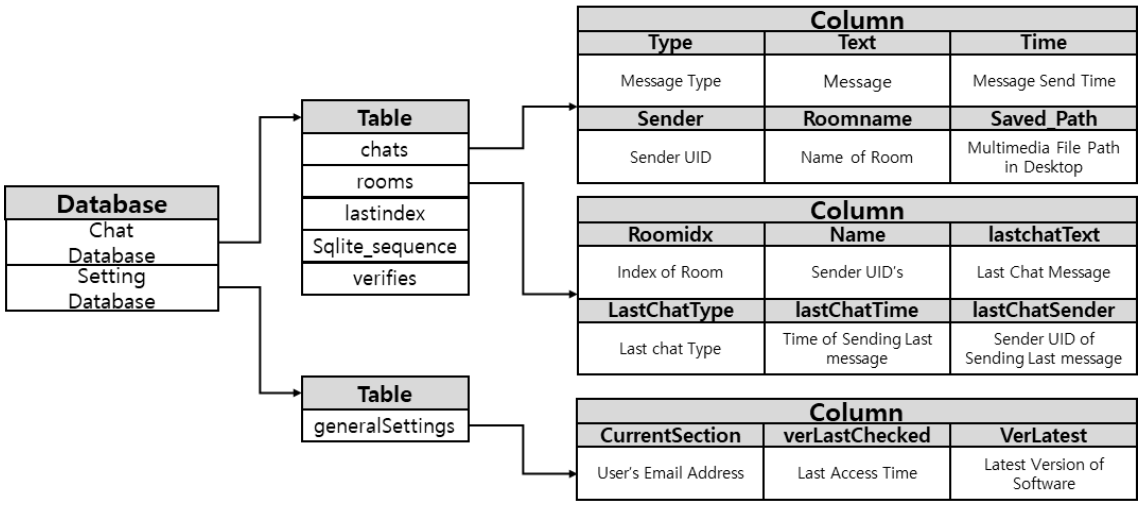

Fig. 3. List of Internal Data in Chat and Setting Database

V. 삭제된 메시지 복구

말랑말랑 톡카페는 사용자가 송신한 메시지를 취소할 수 있도록 삭제 기능을 제공한다. 대화 내역 삭제 기능은 메신저 사용자가 쉽게 사용할 수 있는 안티 포렌식 기법 중 하나로써 대화 내역을 삭제할 경우, 데이터베이스를 복호화하더라도 삭제된 메시지를 획득할 수 없다. 따라서 본 장에서는 메모리 데이터 획득 및 분석을 통해 삭제된 메시지 복구 방안에 대해 제시한다.

5.1 삭제된 메시지 식별

메시지 삭제 전후 암호화된 데이터베이스를 복호화한 결과는 다음과 같다(Fig. 4). 메시지 삭제 전 데이터베이스 내부에는 메시지 타입, 전송 시간, 메시지 내역 등의 정보가 저장된다. 메시지 삭제를 진행하여 삭제 전후 데이터베이스 내부 데이터를 비교 분석한 결과 메시지 타입 영역의 “text” 문자열이 “deleted” 문자열로 교체되며, 메시지 전송 시간은 삭제 이후 데이터베이스에서 제거되지 않는다. 또한, 삭제된 메시지 영역에 “%1님께서 톡을 취소하셨습니다.”라는 문자열이 utf-8로 변환되어 저장된다. 이를 통해 사용자가 고의로 삭제한 메시지를 식별할 수 있다.

5.2. 삭제된 메시지 복구 실험

메시지를 전송할 경우 대화 내역과 전송 시간 등의 데이터셋이 평문 형태로 서버로 전송되며, 이는

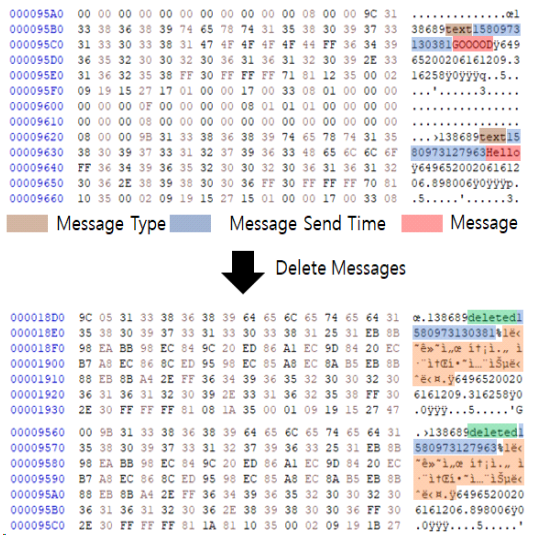

Fig. 4. Comparison Before and After Message Deletion

메모리에 적재된다(Fig. 5). 따라서 삭제된 메시지의 전송 시간을 키워드로 메모리 분석을 진행하면 삭제된 메시지를 복구할 수 있다. 메모리 영역에 남아있는 삭제된 메시지의 원본 데이터는 Fig 6.과 같다. 해당 영역의 데이터는 다른 프로세스에 의해 메모리가 덮어 씌워지지 않는 한 복구 가능하다.

대화 내역 전체 삭제 기능은 단일 메시지 삭제와 다르게 삭제된 메시지 식별이 불가능하다. 이 경우에는 메시지 타입을 나타내는 “text”를 키워드로 메모리 탐색을 수행한다면, 일부 메시지 복구가 가능하다.

Fig. 5. Message Data Set sent in Plain text to Server

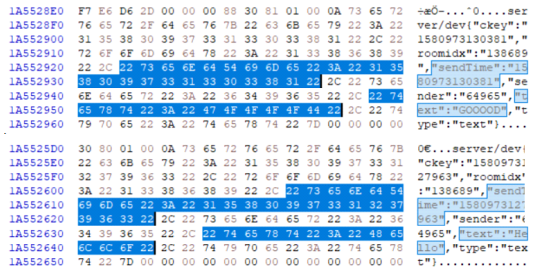

Fig. 6. Deleted Message Left in Volatile Memory

VI. 결 론

본 논문에서는 한글과 컴퓨터에서 개발한 데스크톱 전용 메신저인 말랑말랑 톡카페의 데이터베이스 암호화 과정 및 암호키 생성과정을 분석하여 복호화에 성공하였다. 데이터 암호화는 SQLCipher 모듈을 통해 수행하며, 데이터베이스 전체를 대상으로 한다. 암호화 시 사용하는 키는 사용자 PC의 UUID를 통해 생성된 패스프레이즈를 기반으로 생성된다. 따라서 UUID 획득을 통한 암호키 재현이 가능하므로 데이터베이스의 복호화가 가능하다. 또한, PC 메모리 영역에서 삭제된 메시지 식별 후 데이터 복구가 가능함을 보였다. 이와 같은 결과는 디지털 포렌식 수사 시 사용자 행위 파악에 유용하게 활용될 수 있을 것으로 기대한다.

References

[1] J. Choi, J. Park, and H. Kim, "Forensic analysis of the backup database file in KakaoTalk messenger," 2017 IEEE International Conference on Big Data and Smart Computing, pp. 156-161, Feb. 2017

[2] J, Choi, J. Yu, and H, Kim, "Digital forensic analysis of encrypted database files in instant messaging applications on Windows operating systems: Case study with KakaoTalk, NateOn and QQ messenger," Digital Investigation, vol. 28, pp. S50-S59, Apr. 2019

[3] A. Mahajan, MS. Dahiya, and HP. Sanghvi,"Forensic analysis of instant messenger applications on android devices," arXiv preprint arXiv:1304.4915, 2013.

[4] G. Kim, J. Lee, and S. Shin, "Study on The Decryption Method and Analysis of MalangMalang Talkcafe Application Datbase," Jonornal of The Korea Institute of information Security & Cryptology, 29(3), pp. 541-633, Jun. 2019

[5] G, Feng and Z. Ying," Analysis of WeChat on iPhone," 2nd international symposium on computer, communication, control and automation, Atlantis Press, 2013

[6] S. Wu, Y. Zhang, and X. Wang," Forensic analysis of WeChat on Android smartphones," Digital investigation, vol. 21, pp. 3-10, June. 2017

[7] THAKUR, N.S.," Forensic analysis of WhatsApp on Android smartphones," Master's Thesis, University Of New Orleans, 2013.

〈저자소개〉

윤 병 철 (Byungchul Youn) 학생회원
2019년 2월: 국민대학교 수학과 졸업
2019년 9월~현재: 국민대학교 금융정보보안학과 석사과정
〈관심분야〉 디지털 포렌식, 정보보호

김 소 람 (Soram Kim) 학생회원
2016년 2월: 국민대학교 수학과 졸업
2018년 2월: 국민대학교 금융정보보안학과 석사
2018년 3월~현재: 국민대학교 금융정보보안학과 박사과정
〈관심분야〉 디지털 포렌식, 정보보호

김 종 성 (Jongsung Kim) 종신회원
2000년 8월/2002년 8월: 고려대학교 수학 전공 학사/이학석사
2006년 11월: K.U.Leuven, ESAT/SCD-COSIC 정보보호 전공 공학박사
2007년 2월: 고려대학교 정보보호대학원 공학박사
2007년 3월~2009년 8월: 고려대학교 정보보호기술연구센터 연구교수
2009년 9월~2013년 2월: 경남대학교 e-비즈니스학과 조교수
2013년 3월~2017년 2월: 국민대학교 수학과 부교수
2014년 3월~현재: 국민대학교 일반대학원 금융정보보안학과 부교수
2017년 3월~현재: 국민대학교 정보보안암호수학과 부교수
〈관심분야〉 정보보호, 암호 알고리즘, 디지털 포렌식