논문 2025-1-2 http://dx.doi.org/10.29056/jsav.2025.03.02

OTT 플레이어 에이전트 환경의 콘텐츠 이용 정보 민감도 분류에 따른 비식별화 프로토콜 설계

박병찬*, 장세영*, 유인재*, 김석윤*, 김영모*†

A Content Usage Information De-identification Protocol Design Based on Sensitivity Classification in the OTT Player Agent Environment

Byeong-Chan Park*, Se-Young Jang*, In-Jae Yoo*, Seok-Yoon Kim*, Young-Mo Kim*†

요 약

오늘날 OTT(Over-The-Top) 환경에서는 이용자 맞춤형 콘텐츠 제공을 위해 이용자의 데이터를 수집하고 분석하는 과정이 필수적이다. 그러나 이러한 과정에서 개인 정보 보호 및 데이터 보안이 중요한 이슈로 부각되고 있다. 본 논문에서는 OTT 플레이어 에이전트 환경에서 이용자 데이터의 프라이버시를 보호하기 위해 영지식 증명(Zero-Knowledge Proof, ZKP) 기반의 데이터 비식별화 프로토콜을 제안한다. 먼저, 기존의 비식별화 기술과 영지식 증명의 개념을 분석하고, OTT 플레이어 에이전트 환경의 콘텐츠 이용 정보 수집 과정에서 이를 적용할 수 있는 방안을 모색한다. 또한, 이용자의 민감도 분류에 따른 가명처리 및 암호화 기법을 활용하여, 데이터의 무결성을 유지하면서도 개인 정보를 보호하는 방안을 제시한다. 마지막으로, 제안된 방법론의 효과성을 평가하기 위해 다양한 시나리오에서의 적용 가능성을 검토한다. 본 연구는 OTT 플레이어 에이전트 환경에서 개인정보 보호 문제를 해결하는 데 기여할 수 있으며, 향후 데이터 보호 기술 발전에 중요한 기초 자료로 활용될 수 있을 것이다.

Abstract

In today's Over-The-Top (OTT) player agent environment, the process of collecting and analyzing user data is essential for providing personalized content. However, privacy protection and data security have emerged as critical issues in this process. This paper proposes a Zero-Knowledge Proof (ZKP)-based data anonymization protocol to safeguard user privacy in the OTT player agent environment. First, we analyze existing anonymization techniques and the concept of Zero-Knowledge Proofs to explore their applicability in the process of collecting content usage information in the OTT player agent environment. Additionally, we present an approach that utilizes pseudonymization and encryption techniques based on sensitivity classification to maintain data integrity while protecting personal information. Finally, we evaluate the effectiveness of the proposed methodology by examining its applicability in various scenarios. This study contributes to addressing privacy issues in the OTT player agent environment and serves as a valuable foundation for future advancements in data protection technologies.

한글키워드 : 영지식 증명 (ZKP), OTT 플랫폼, 데이터 비식별화, 시청률 조사, 개인정보 보호

keywords : Zero-Knowledge Proof (ZKP), OTT Platforms, Data Anonymization, Audience Measurement, Privacy Protection

* 숭실대학교 컴퓨터학과
† 교신저자: 김영모(email: ymkim828@ssu.ac.kr)

접수일자: 2024.12.31. 심사완료: 2025.03.13.
게재확정: 2025.03.20.

www.kci.go.kr

1. 서 론

OTT(Over-The-Top) 플랫폼의 보급이 확산됨에 따라, 시청률 조사를 위한 이용자의 데이터 수집과 분석이 필수적으로 자리 잡았다. 특히, OTT 플레이어 에이전트 환경에서는 콘텐츠 권리자의 권리 관리를 위해 정확한 시청률 조사가 필요하다[1]. 시청 데이터는 콘텐츠 제공자와 저작권자의 정당한 보상을 결정하는 중요한 요소이며, 공정한 수익 분배와 저작권 보호를 위해 신뢰할 수 있는 시청률 분석이 필수적이다[2]. 그러나 이 과정에서 개인정보 보호 및 데이터 보안 문제가 주요한 이슈로 대두되고 있다[3]. 이용자 데이터의 무분별한 수집과 활용은 개인 정보 유출 및 프라이버시 침해 가능성을 증가시킬 수 있다[4]. 이러한 데이터에는 개인을 직접 식별할 수 있는 이름, 나이, 성별, 이메일 주소와 같은 민감 정보뿐만 아니라, 콘텐츠 시청 기록, 플랫폼 사용 패턴 등 간접적으로 개인을 식별할 수 있는 정보도 포함하고 있다[5]. 이러한 데이터를 기반으로 한 시청률 분석, 개인화된 콘텐츠 추천, 플랫폼 최적화는 OTT 서비스 품질 개선에 필수적이다[6]. 이러한 플랫폼에서는 개인 OTT 정보보호 등의 여러 이슈로 이용자가 시청하는 콘텐츠의 시청 시간을 공개하지 않고 있어 이해관계들의 제삼자 방식의 이용 행태 조사를 하고 있는 실정이여 객관적이고 신뢰성 있는 제삼자 방식의 이용행태조사 방법이 필요하다[7]. 이러한 이용행태조사를 위해 시청률 조사에서 필요한 정보는 어떤 콘텐츠가 얼마나 이용되었는지 만 필요하며 이용자 개인을 특정하는 개인정보는 필요하지 않다[8].

본 논문에서는 OTT 플레이어 에이전트 환경에서 생성되는 이용 정보 데이터를 대상으로, 데이터 민감도에 따라 차등적 비식별화 방법인 OTT 플레이어 에이전트 환경의 콘텐츠 이용 정보 민감도 분류에 따른 비식별화 프로토콜 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구로 비식별화 기술 및 영지식 증명 기술을 기술한다. 3장에서는 OTT 콘텐츠 이용 정보 비식별화 과정을 설명하고 4장에서 본 논문에서 제안하는 영지식 증명 기반 OTT 콘텐츠 이용 정보 비식별화 방법을 제안한다. 5장에서 결론으로 마무리한다.

2. 관련 연구

2.1 비식별화 기술

OTT 이용정보 데이터의 비식별화는 개인정보 보호를 위한 필수적 과정[9]로, 기존 연구에서 다양한 비식별화 기술이 제안되어왔다. 대표적인 기술은 다음과 같다

- 난수화(Randomization): 민감 데이터를 무작위 값으로 대체하여 개인 식별 가능성을 줄이는 방법이다. 예를 들어, 이용자의 이름을 임의의 값으로 변환하거나 시청 시간을 범위 값으로 표현할 수 있다.

- 토큰화(Tokenization): 민감 데이터를 고유한 토큰 값으로 치환하여 원본 데이터에 대한 접근을 제한한다. 이를 통해 데이터 분석 목적으로만 사용 가능하도록 보장한다[9].

- 암호화(Encryption): 비밀키를 사용해 데이터를 암호화하고, 복호화 키를 소유한 이용자만 데이터에 접근할 수 있도록 보안성을 확보한다.

- 가짜 데이터 생성(Synthetic Data Generation): 실제 데이터를 기반으로 유사한 형태의 가짜 데이터를 생성하여 분석 목적에 활용하면서도 원본 데이터의 노출을 방지한다.

2.2 영지식 증명

영지식 증명(Zero-Knowledge Proof, ZKP)은

www.kci.go.kr

민감 정보를 노출하지 않으면서 데이터의 유효성과 무결성을 증명할 수 있는 강력한 암호학적 기술이다[10]. ZKP는 검증자에게 민감한 데이터 자체를 공개하지 않고도 해당 데이터가 조건을 충족함을 증명할 수 있다. 주요 특징은 다음과 같다:

- 데이터 비공개성: 데이터의 민감 정보를 노출하지 않으면서도 검증 과정을 진행할 수 있다.
- 무결성 보장: 데이터가 변경되지 않았음을 증명할 수 있으며, 검증 과정에서 데이터의 원본 내용이 필요하지 않다.
- 재식별 위험 감소: 데이터를 분석 가능한 형태로 유지하면서도 개인정보 보호를 강화할 수 있다.

3. OTT 콘텐츠 이용 정보 비식별화 과정

3.1 이용 정보 생성 데이터

OTT 콘텐츠 이용 정보 생성 과정은 그림 1과 같다.

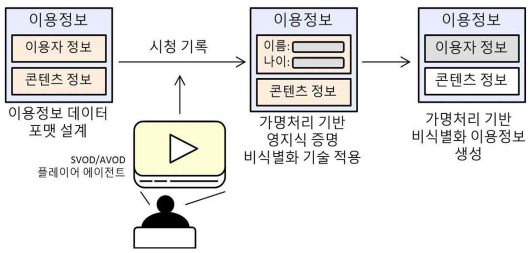

그림 1. OTT 콘텐츠 이용 정보 생성 과정
Fig. 1. OTT Content Usage Information Generation Process

이용자가 SVOD(구독 기반)/AVOD(광고 기반) OTT 플랫폼에 가입하고 콘텐츠를 시청하는 과정에서 다양한 데이터가 생성된다. 이 데이터는 크게 이용자 정보와 이용 정보로 구분할 수 있다.

이용자 정보는 서비스 가입 시 수집되는 정보로서, 플랫폼에 따라 약간의 차이가 있지만 주로 개인 식별을 위한 최소한의 정보로 구성된다. 대표적으로 이메일 주소(ID 대용), 나이(성인 콘텐츠 이용을 위한 연령 확인 목적), 이름(플랫폼 프로필에서 주로 사용됨), 성별 등이다.

이용 정보는 이용자가 실제 콘텐츠를 시청할 때 생성되는 정보로, 콘텐츠 시청 이력, 시청 시간, 시청에 사용된 플랫폼 정보 등을 포함한다. 이 정보는 콘텐츠 추천 알고리즘 및 시청률 분석을 위한 핵심 데이터로 활용된다. 이용 정보의 구체적인 항목은 표 1과 같다.

표 1. OTT 콘텐츠 이용 정보 생생 데이터
Table. 1. OTT Content Usage Data Generation

이용자 정보	이름
	나이
	성별
	이메일 주소
이용 정보	시청 콘텐츠
	시청 시간
	플랫폼

이와 같이 OTT 플랫폼에서 수집되는 데이터는 개인 식별 가능 여부에 따라 민감도가 다르게 나타난다. 따라서 본 논문에서는 데이터 민감도 분류를 기반으로 차등적인 비식별화 프로토콜을 적용하여 개인정보를 보호하면서도 필요한 분석 데이터를 제공하는 방안을 제시한다.

3.2 OTT 콘텐츠 이용 정보 생성

본 논문에서 제안하는 플레이어 에이전트 기반 OTT 영지식 증명 프로토콜을 이용한 이용 정보 비식별화 방법은 그림 2와 같다.

OTT 콘텐츠를 이용할 때 발생되는 이용자 정보 및 콘텐츠 이용 정보는 시청자 OTT 콘텐츠

www.kci.go.kr

를 재생할 때 OTT 콘텐츠의 시청 정보가 기록된다. 이에 이용자 정보는 제삼자가 식별 불가능하도록 비식별화 영역으로 지정한다. 이에 따라 민감도를 분류하면, 표 2와 같다.

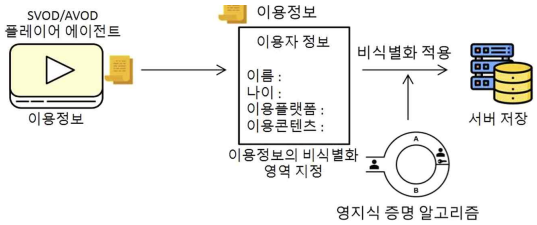

그림 2. OTT 콘텐츠 이용 정보 데이터 비식별화 방법
Fig. 2. OTT Content Usage Data Anonymization Method

표 2. OTT 콘텐츠 이용 정보 생생 데이터 민감도 분류
Table. 2. OTT Content Usage Data Sensitivity Classification

고위험	이름, 이메일 주소, IP 주소 (개인 직접 식별 가능)
중위험	나이, 성별, 시청 시간 (결합을 통해 간접 식별 가능)
저위험	콘텐츠 유형, 플랫폼 종류 (개인과 직접 연관되지 않음)

비식별화된 이용자 정보와 콘텐츠 이용 정보를 합하여 최종적으로 가명 처리된 비식별화 이용 정보를 생성한다.

4. 민감도 분류에 따른 영지식 증명 기반 이용 정보 비식별화 방법

4.1 가명처리 기반 비식별화 영역 지정 및 적용방안 도출

OTT 플랫폼에서 생성되는 이용 정보 데이터에는 이름, 나이, 성별, 시청 콘텐츠 등 개인을 식별할 가능성이 있는 다양한 정보가 포함되어 있다. 이를 위해, 이용 정보 데이터를 민감도에 따라 고위험, 중위험, 저위험 데이터로 분류하고, 각각 적합한 비식별화 기법을 적용하며, 표 3과 같다.

표 3. 민감도에 따른 가명처리 적용 기법
Table. 3. Pseudonymization Techniques Based on Sensitivity Levels

민감도	적용 기법
고위험	토큰화, 암호화
중위험	난수화, 가짜 데이터 생성
저위험	데이터 유지

고위험 데이터에는 이름, 이메일 주소, IP 주소와 같은 직접 식별 가능 정보가 포함되며, 중위험 데이터는 나이, 성별, 시청 패턴과 같이 결합 시 간접 식별 가능성이 있는 정보를 포함한다. 저위험 데이터는 콘텐츠 유형, 플랫폼 종류와 같이 개인과 직접적인 연관이 없는 정보를 포함한다.

이에 따른 가명처리 적용 방안을 도출하며, 그림 3과 같다.

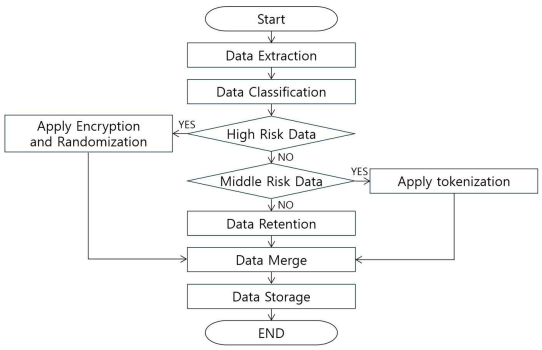

그림 3. 가명처리 과정 흐름도
Fig. 3. Pseudonymization Process Flowchart

www.kci.go.kr

- 데이터 추출(Data Extraction): OTT 플랫폼에서 콘텐츠 이용 정보 원시 데이터를 추출한다.

```
{
    "user_id": "TKN-12345",
    "name": "홍길동",
    "age": 30,
    "viewing_history": [...],
    "platform": "Mobile"
}
```

- 민감도 분류(Data Classification): 데이터를 고위험, 중위험, 저위험 영역으로 구분한다.

· 고위험: name, user_id
· 중위험: age, viewing_history
· 저위험: platform

- 가명처리 기법 적용: 분류된 민감도 수준에 따라 각각의 비식별화 기법을 적용한다.

· 고위험 필드: 암호화 및 난수화 적용
이용자의 이름을 SHA-256 해시(hash) 값을 생성하여 원본 이름을 복구할 수 없도록 처리하고, 이메일 주소는 암호화하여 데이터베이스에 저장한다.
예: name → "홍길동" → "TKN-XXXXX"
· 중위험 필드: 토큰화 적용
이용자의 나이를 정확한 숫자 대신 나이 범위(예: 20대, 30대)로 변환하거나, 성별 데이터를 일정 비율로 변형하여 개인 식별 가능성을 줄인다.
예: age → "30" → "AGE-TKN"
· 저위험 필드: 데이터 유지
개인과 직접적으로 연결되지 않으며 식별 가능성이 거의 없는 저위험 데이터는 데이터의 유용성을 극대화하기 위해 원본 형태 그대로 유지하여 활용한다.
예: platform → "Mobile"

- 데이터 병합(Merge): 비식별화된 데이터를 하나의 최종 데이터로 통합한다.
- 데이터 저장: 비식별화 완료된 데이터를 서버에 안전하게 저장한다.

```
{
    "user_id": "TKN-12345",
    "name": "TKN-12345",
    "age": "AGE-TKN",
    "viewing_history": [...],
    "platform": "Mobile"
}
```

4.2 이용정보 데이터 포맷에 영지식 증명 암호 알고리즘 적용 및 프로토콜 설계

OTT 플랫폼에서 생성되는 이용정보 데이터는 개인 식별이 가능한 중요한 정보를 포함한다. 이러한 데이터를 비식별화하여 안전하게 활용하면서도 데이터의 유효성과 무결성을 보장하기 위해 영지식 증명 기반 암호화 알고리즘을 적용한다. 이를 위해 암호화 대상 데이터를 지정하며 표 3과 같다.

표 3. OTT 콘텐츠 이용 정보 생성 데이터
Fig. 2. OTT Content Usage Information Generation Data

민감도	적용 기법
민감 데이터	user_id, name, age와 같은 개인 식별 가능 정보
비민감 데이터	viewing_history, platform과 같은 분석에 필요한 정보

민감 데이터는 비식별화 후 암호화(ZKP 적용)하며, 비민감 데이터는 무결성 검증만 적용(ZKP를 사용하여 데이터 변경 여부 확인)할 수 있도록 한다.

이를 기반으로 영지식 증명 암호 알고리즘 적용 및 프로토콜을 도출하며 그림 4와 같다.

www.kci.go.kr

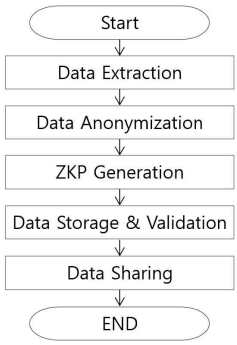

그림 4. ZKP 기반 암호화 및 검증 프로토콜 흐름도

Fig. 4. KP-Based Encryption and Verification Protocol Flowchart

- 데이터 추출(Data Extraction): OTT 플랫폼에서 콘텐츠 이용 정보 원시 데이터를 추출한다. 데이터를 추출한 후, 민감한 정보(name, user_id, age)를 분리한다.

```
{
    "user_id": "U12345",
    "name": "John Doe",
    "age": 30,
    "viewing_history": [
        {"content_id": "C001", "duration": 120},
        {"content_id": "C002", "duration": 45}
    ],
    "platform": "Mobile"
}
```

- 민감도 분류(Data Classification): 데이터를 고위험, 중위험, 저위험 영역으로 구분하고 토큰화 및 암호화 등의 기법을 적용한다.

```
{
    "user_id": "TKN-ABCDE",
    "name": "TKN-XYZ123",
    "age": "30",
    "viewing_history": [
        {"content_id": "C001", "duration": 120},
        {"content_id": "C002", "duration": 45}
    ],
    "platform": "Mobile"
}
```

- ZKP 생성(ZKP Generation): 비식별화된 데이터에 대한 영지식 증명 정보를 생성하여 데이터의 무결성과 진위 여부를 검증할 수 있도록 한다. zk-SNARK 알고리즘을 통해 데이터 증명 생성한다.

```
{
    "user_id": "TKN-ABCDE",
    "proof": {
        "zk_snark": "proof_data_string",
        "hash": "SHA256_HASH"
    },
    "viewing_history": [
        {"content_id": "C001", "duration": 120},
        {"content_id": "C002", "duration": 45}
    ],
    "platform": "Mobile"
}
```

- 데이터 저장 및 검증(Data Storage & Validation): 영지식 증명 정보를 포함하여 데이터를 저장하고, 저장된 데이터가 변경되지 않았음을 주기적으로 검증하여 데이터의 무결성을 유지한다.

```
{
    "user_id": "TKN-ABCDE",
    "proof": "proof_data_string",
    "data": {
        "viewing_history": [
            {"content_id": "C001", "duration": 120},
            {"content_id": "C002", "duration": 45}
        ],
        "platform": "Mobile"
    }
}
```

- 데이터 공유(Data Sharing): 저장된 데이터를 제삼자에게 제공할 경우, 영지식 증명을 통해 데이터가 변조되지 않았음을 확인시켜주어 데이터의 신뢰성을 확보한다.

www.kci.go.kr

5. 결 론

본 논문에서는 OTT 플랫폼에서 생성되는 이용 정보 데이터를 대상으로 영지식 증명 기반의 비식별화 및 암호화 방법을 제안하였다. 제안한 방법은 민감 데이터를 비식별화하고 암호화하여 개인정보를 보호함과 동시에 데이터 무결성을 검증할 수 있다. 이용 정보는 이용자 정보와 콘텐츠 이용 정보로 구성되어 있어 권리관리 행사에 직접적으로 필요한 콘텐츠 이용 정보는 식별화하고 필요하지 않는 이용자 정보는 비식별화하는 방법으로 제안하였다. 이러한 이용자 정보 비식별화 방법은 최근 개인정보 보호의 필요성이 높아짐에 따라 개인정보를 수집 및 처리하는 다양한 플랫폼에서 효과적으로 개인정보 보호가 가능할 것으로 기대된다. 추후 연구로써 환경의 이용자 정보 및 콘텐츠 이용 정보의 메타데이터 연구 및 영지식 증명 프로토콜 적용 방법이 필요하다.

This work was supported by Ministry of Science and ICT and Institute for Information & communication Technology Planning & evaluation (IITP) (2022-0-00510)

참 고 문 헌

[1] I, Hwang, “TMitigation of Lock-in and Privacy Concerns for OTT(Over The Top) Users: The Role of Service Diversity, Technical Support, and Social Interaction”, Journal of Digital Contents Society, Vol. 25, No. 5, pp. 1205-1216, 2024, DOI: https://doi.org/10.9728/dcs.2024.25.5.1205

[2] Korea Copyright Commission, Copyright Issue Trends, Biweekly Report, Vol. 32, No. 7, 2024. Available at: https://www.copyright.or.kr/information-materials/trend/tmis/view.do?brdctsno=53132#.

[3] S. Lee, J. Jeon, J. Oh, “Enhancing Consumer Awareness and Privacy Protection in the Era of Over-the-Top(OTT) Services: Focused on Behavioral Information Collection and Personalized Content”, Journal of The Korea Institute of Information Security & Cryptology, Vol. 34, No. 3, pp.505-513, 2024,
DOI: https://10.13089/JKIISC.2024.34.3.505

[4] A. Park, M. Oh, “Statistical Approaches to Data Privacy and Related Issues,” Issues in Statistical Approaches to Data Privacy, Vol. 262, No. 5, pp. 39-50, 2018. DOI: 10.23062/2018.08.5.

[5] D. Noh, “Enhancing Competitiveness through Data Utilization in Video OTT Platforms,” Media Issue & Trend, Vol. 51, pp. 30-42, 2022. Available at: https://www.kca.kr/Media_Issue_Trend/vol51/download/KCA_Media_Issue_Trend_vol51_featured_report_03.pdf.

[6] J. Kim, H. Ha, S. Kim, Y. W. Jung, “User Experience Analysis of OTT Service Content Recommendation -Focused on Netflix Case”, Journal of Integrated Design Research, pp.73-87, Vol. 20, No. 2, 2021. DOI: 10.21195/jidr.2021.20.2.005

[7] B. Park, S. Jang, S. Kim, Y. Kim, Proposition of Viewing Information De-identification and Verification Method in OTT Player Agent Environment Based on Zero-Knowledge Proof, Journal of Software Assessment and Valuation, Vol. 19, No. 4, pp. 107-113, 2023, DOI: http://dx.doi.org/10.29056/jsav.2023.12.11

[8] K. Min, “Prospects of Audience Rating Surveys According to Changes in Viewing Behavior,” Broadcasting and Media, Vol. 27, No. 4, pp. 35-39, 2022. Available at: https://scienceon.kisti.re.kr/commons/util/originalView.do?cn=JAKO202234060791023&oC

www.kci.go.kr

n=JAKO202234060791023&dbt=JAKO&journal=NJOU00565331.
[9] Financial Supervisory Service, Guidelines on Pseudonymization and Anonymization in the Financial Sector, Financial Services Commission, Financial Supervisory Service. Available at: https://www.fsec.or.kr/bbs/detail?menuNo=246&bbsNo=6484.
[10] C, Ju, H, Lee, H. Chung, J. H. Seo, “Analysis of Zero-Knowledge Protocols for Verifiable Computation and Its Applications”, Journal of the Korea Institute of Information Security & Cryptology, Vol. 31, No. 4, pp. 675-686 2021, DOI: https://doi.org/10.13089/JKIISC.2021.31.4.675
[11] C. Park, J. Kim, D. H. Lee, “Privacy-Preserving Credit Scoring Using Zero-Knowledge Proofs”, Journal of the Korea Institute of Information Security & Cryptology, Vol. 29, No. 6, pp. 1285-1303, 2019, DOI: https://doi.org/10.13089/JKIISC.2019.29.6.1285

저 자 소 개

박병찬(Byeongchan Park)

2015.2 학점은행제 졸업
2018.2 숭실대학교 컴퓨터학과 석사
2023.8 숭실대학교 컴퓨터학과 박사
2023.9-현재 숭실대학교 초빙교수
<주관심분야> 저작권 보호 및 이용활성화

장세영(Seyoung Jang)

2018.2 평생교육원 학점은행 졸업
2021.6 숭실대학교 컴퓨터학과 석사
2023.2-현재 숭실대학교 컴퓨터학과 박사 과정
<주관심분야> 저작권 보호 및 이용활성화

유인재(Injae Yoo)

2017.8 고려사이버대학교 소프트웨어공학과 졸업
2021.02 숭실대학교 컴퓨터학과 석사
2023.02 ~ 현재 숭실대학교 컴퓨터학과 박사 수료
<주관심분야>저작권 보호 및 이용활성화

www.kci.go.kr

김석윤(Seok-Yoon Kim)

1980.2 서울대학교 전기전자 졸업
1990.2 University of Taxas at Austin Dept. of ECE 석사
1993.2 University of Taxas at Austin Dept. of ECE 박사
1982-1987 ETRI 연구원
1993-1995 모토로라 책임 연구원
1995-현재 : 숭실대학교 교수
<주관심분야> 저작권 보호 및 이용활성화

김영모(Youngmo Kim)

2003.2 대전대학교 컴퓨터공학과 졸업
2005.2 대전대학교 컴퓨터공학과 석사
2011.2 대전대학교 컴퓨터공학과 박사
2012-현재 : 숭실대학교 교수
<주관심분야> 저작권 보호 및 이용활성화

www.kci.go.kr