

# A Comparison of Information Security Certification Architectures

Seong-Kyu Kim\*

\*Assistant Professor (Tenure Track) of Department of Information Security, Joongbu University, Gyeonggi-do Goyang-Si 10279, Republic of Korea.

Department of Public Policy and Information Technology, Seoul National University of Science and Technology, Seoul 01811, Korea.(e-mail : [skkim@joongbu.ac.kr](mailto:skkim@joongbu.ac.kr) or [guitara77@gmail.com](mailto:guitara77@gmail.com) )

Jun-Ho Huh\*\*

\*\* Assistant Professor (Tenure Track) of Department of Data Informatics, (National) Korea Maritime and Ocean University, 727, Taejong-ro, Yeongdo-gu, Busan, Republic of Korea (e-mail : [72networks@pukyong.ac.kr](mailto:72networks@pukyong.ac.kr) or [72networks@kmou.ac.kr](mailto:72networks@kmou.ac.kr) )

## Abstract

This paper needs to have a standardized legal and management system as the development of AI, artificial intelligence, IOT, etc. developed through the 4th industrial era, and the use of personal information is increased in the era of global connection. Therefore, through this paper, I would like to compare and analyze the privacy management system in the U.S. and Korea, which emphasized technical protection measures, to find a more advanced direction of protection measures.

## I. Introduction

In this paper, the growing importance of information security and personal information has led to an increase in crimes exploiting it. Previously, crimes for self-disclosure were prevalent, but as financial transactions through personal information became possible and the number of cases of personal privacy exposure increased, and more accidents caused corporate and individual damage due to personal mental damage, corporate image damage, and legal litigation costs [1].

To address these risks, countries have developed laws on privacy and management systems for privacy protection, and technologies to de-identify specific personal information to promote business activation have also developed. Global countries have enacted laws based on OECD 8 principles

and protect personal information based on related laws, but as some different legal systems between countries have increased the number of cases of excessive fines on companies, global businesses have frequently been closed. It will be necessary to look at and operate this globalized personal information in various fields of view [2]. In this paper, we want to compare and analyze the operation method of personal information operated abroad compared to the US personal information management system, which leads the personal information management system, and find ways to expand and use Korea's personal information management system in global countries [3].

## II. Related Work

Representative compliance and management systems related to information protection and personal information include HIPAA(Health Insurance Portability and Accountability), the US Privacy Framework CSF-P, the International Privacy Standards ISO27701, the European Union General Privacy Act GDPR(General Data Protection Regulation), and Korea's ISMS-P. The relevant management systems for personal information protection are structured based on laws enacted in each country. The basic principles of privacy in each country are mostly similar. However, there are some differences in the operation of the US and Korea's privacy management system, so we want to compare and analyze relevant protection items to find ways to improve privacy [4].

### 2.1. ICT and Cybersecurity Status

The International Telecommunications Union (ITU) compared ICT(Information & Communications Technology) development and cybersecurity indicators between countries through measurements of ICT development index (IDI) by weighting measurements such as Internet use, wired and wireless high-speed Internet subscribers, and ICT utilization. When comparing the ICT development index of the U.S. and South Korea[5], it can be confirmed that South Korea ranks second in the global ICT rankings, far superior to the U.S. It can be seen that South Korea is actively investing in network infrastructure and ICT businesses in global countries[6].

### 2.2. Understand of Cyber Security Framework

The United States issued the Cyber Security Framework (CSF) by President Obama in 2013 and NIST issued a final security framework and roadmap in response to cybersecurity threats to its major infrastructure, finance, health and energy, under executive order 13636. After February 12, 2014, the core areas were supplemented to implement an information security roadmap applicable to the marketplace by combining private, public and academic circles[7].

Initially, CSF version 1.0 established a framework to reduce the security risk of major national infrastructure, and on 16 April 2018, version 1.1 was revised to add protection measures for supply chains. CSF implements a life cycle with five functions: identification, protection, detection, response, and recovery, and guides the organization to select and perform targets appropriate to the organization's situation among four levels of partial response, risk information utilization, iterative action, and continuous improvement.

## III. Information Security and Privacy Framework

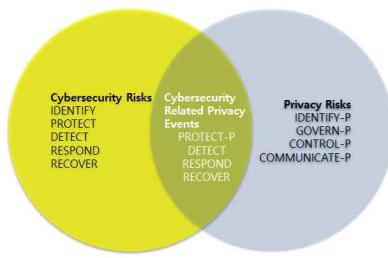
### 3.1. CSF

The CSF is structured in stages of systematic operations based on a set of information protection activities, from identification of assets to recovery, and is structured in a structure that is common to major infrastructures [8]. The basic security control of an organization is to establish policies, processes, and procedures to share the entire organization's activities with its members to operate and manage them systematically. CSF can operate and maintain

the five-step capabilities of identification, protection, detection, response, and recovery simultaneously or continuously, and can also perform high-level risk strategies through life cycles such as asset identification, evaluation, response and maintenance.

### 3.2. CSF-P

NIST (National Institute of Standards and Technology) has released PRIVACY FRAMEWORK through transparent and agreed-upon procedures, including private and public stakeholders, to encourage autonomous compliance with the privacy management system. It was used as a tool for personal information protection through risk management through the corporate privacy framework and developed into a technology for personal information protection for organizations and individuals. The privacy framework supports five functions. The five functions defined below, Identification-P, Gov-P, Control-P, Communicate-P and Protect-P, can be used to manage the risk of privacy arising from privacy data processing. Protect-P is particularly focused on risk management. CSF is intended to address all types of cybersecurity incidents but has been utilized to further support risk management related to cybersecurity-related privacy events using Detect, Response, and Recover Functions. The organization has comprehensively addressed privacy and cybersecurity risks using all five cybersecurity framework features, along with Identification-P, Gov-P, Control-P and Communicate-P. [Fig. 1] shows how to manage various aspects of privacy and cybersecurity risks by using the features of the two frameworks in various combinations. The five privacy framework functions are as follows.



[Fig. 1] Using Functions to Manage Cybersecurity and Privacy Risks

[Tab. 1] Using Functions to Manage

Function Unique Identifier	Function	Category Unique Identifier	Category
ID-P	Identify-P	ID.IM-P	Inventory and Mapping
		ID.BE-P	Business Environment
		ID.RA-P	Risk Assessment
		ID.DE-P	Data Processing Ecosystem Risk Management
		GV.PO-P	Governance Policies, Processes, and Procedures
GV-P	Govern-P	GV.RM-P	Risk Management Strategy
		GV.AT-P	Awareness and Training
		GV.MT-P	Monitoring and Review
CT-P	Control-P	CT.PO-P	Data Processing Policies, Processes, and Procedures
		CT.DM-P	Data Processing Management
		CT.DP-P	Disassociated Processing
CM-P	Communicate-P	CM.PO-P	Communication Policies, Processes, and Procedures
		CM.AW-P	Data Processing Awareness
PR-P	Protect-P	PR.PO-P	Data Protection Policies, Processes, and Procedures
		PR.AC-P	Identity Management, Authentication, and Access Control
		PR.DS-P	Data Security
		PR.MA-P	Maintenance
		PR.PT-P	Protective Technology
DE	Detect	DE.AE	Anomalies and Events
		DE.CM	Security Continuous Monitoring
		DE.DP	Detection Processes
RS	Respond	RS.RP	Response Planning
		RS.CO	Communications
		RS.AN	Analysis
		RS.MI	Mitigation
RC	Recover	RS.IM	Improvements
		RC.RP	Recovery Planning
		RC.IM	Improvements
		RC.CO	Communications

Identification-P develops organizational understanding to manage personal privacy risks arising from data processing. The activity of the Identification-P function is a fundamental step in the effective use of the privacy framework. By investigating the environment in which data is processed,

understanding the privacy of individuals directly or indirectly provided or affected by the organization, and conducting risk assessment, organizations can understand the business environment in which data operates, identify and prioritize privacy risks [Tab. 1].

### 3.3. ISMS-P

Personal information in ISMS-P consists of five areas of certification, including personal information collection, privacy measures, privacy measures, privacy measures, and privacy rights protection, which emphasize the life cycle analysis of personal information only applied in Korea.

## IV. Future research and conclusions

Unlike the U.S. management system, the domestic privacy management system is designed to manage the flow of personal information by preparing flow charts and flow charts. It analyzes and schematizes the flow of personal information when it is introduced or changed. Based on this schematic document, the risk of personal information is quickly recognized and maintained to check the risk area. On the other hand, CSF-P in the U.S. is actually establishing a countermeasure step by step to check the actual response attack flow in preparation for hacking attacks or attacks by external intruders. Therefore, it is possible to make practical judgments and respond to dangerous behavior is possible. Therefore, it is necessary to operate a personal information management system in an easy-to-access manner in identifying and responding to risks in small and medium-sized enterprises as a management procedure for identifying, managing, controlling, communicating,

detecting, blocking, responding and recovering assets required by CSF-P. And depending on the size, a response system will be needed to understand and analyze risks through the management of detailed personal information flow through the preparation of Korea's personal information flow chart.

## [Reference]

- [1] J. L. Hennessy and D. A. Patterson, "Instruction-level parallelism and its exploitation," in Computer Architecture: A Quantitative Approach, 4th ed., San Francisco, CA: Morgan Kaufmann Pub., pp.66–153, 2007.
- [2] S. Y. Hea, E. G. Kim, "Design and implementation of the differential contents organization system based on each learner's level," The KIPS Transactions: Part A, Vol.18, No.6, pp.19–31, 2011.
- [3] S. Russell, P. Norvig, "Artificial Intelligence: A Modern Approach," 3th ed., New York: Prentice Hall, 2009.
- [4] D. B. Lenat, "Programming artitical interlligence," in Understanding Artificial Intelligence, Scientific American, Ed., New York: Warner Books Inc., pp.23–29, 2002.
- [5] Paul Dunphy, and Fabian A. P. Petitcolas. "A First Look at Identity Management Schemes on the Blockchain," IEEE Security and Privacy Magazine special issue on "Blockchain Security and Privacy", August 2018.
- [6] A. Stoffel, D. Spretke, H. Kinnemann, D.A. Keim, "Enhancing document structure analysis using visual analytics," Proceedings of the ACM Symposium on Applied Computing, 2010, pp .8–12.
- [7] J. Y. Seo, "Text driven construction of discourse structures for understanding descriptive texts," Ph.D. Dissertation, University of Texas at Austin, TX, USA, 1990.
- [8] Park, Kyeong-tae, "A Study on the Obstacle Factors during ISMS Certification: Focused on SMEs," KAIST, 2015.

# Multi Authority Attribute-Based Encryption 시스템에서 탈중앙화된 보증금 관리 프로토콜을 이용한 공모공격의 방지

노시완\*, 장설아\*\*, 이경현\*\*\*

\*부경대학교 일반대학원 정보보호학과

\*\*부경대학교 일반대학원 인공지능융합학과

\*\*\*부경대학교 IT융합응용공학과

nosiwan@pukyong.ac.kr, seolahh1020@gmail.com, khrhee@pknu.ac.kr

## A Security Deposit Protocol to Prevent Collusion Attacks in the Multi-Authority Attribute-Based Encryption System

Siwan Noh\*, SeolAh Jang\*\*, and Kyung-Hyune Rhee\*\*\*

\*Department of Information Security, Graduate School,  
Pukyong National University

\*\*Department of Interdisciplinary Graduate Program of Artificial  
Intelligence on Computer, Pukyong National University

\*\*\*Department of IT Convergence and Application Engineering,  
Pukyong National University

### 요약

속성기반암호는 특유의 유연성으로 다양한 환경에서 활용되고 있는 암호 기술이다. 다중기관 속성기반암호 시스템은 여러 기관이 사용자의 속성을 발급 및 관리하는 모델로 사용자는 여러 기관에 속하면서 각 기관으로부터 속성을 발급받아 시스템에서 사용한다. 전통적인 다중기관모델에서의 공모공격 방지는 중앙관리기관으로부터 사용자마다 고유한 정보를 발급받아 키 생성에 사용하여 다른 사용자 간의 공모를 원천적으로 차단하고 있으나 이로 인해 중앙관리기관에 대한 의존성이 문제가 된다. 본 논문에서는 중앙기관에 의존하지 않고 사용자 스스로 보증금을 등록하고 이 보증금을 출금하기 위한 정보가 공모공격에서 노출되도록 하여 합리적인 사용자로 하여금 공모공격의 동기를 잃도록 하는 방법을 제안한다.

## I. 서론

속성기반암호(Attribute-Based Encryption, ABE)[1]는 암호화 통신을 위해 수신자를 특정 할 필요가 없어 기존의 일대일 암호통신 기법 보다 유연하게 활용이 가능한 암호기술이다. 메시지 송신자는 수신자의 고유한 정보 대신 수신자의 속성(소속, 직위 등)을 사용하여 암호화를 수행하고 해당 속성을 가진 사용자들은 누구나 암호문을 복호화할 수 있다. 사용자의 속성은 사용자가 소속된 기관에서 속성을 검증하고 사용자가 보유하고 있는 속성의 집합에 해

당하는 비밀키를 생성함으로서 관리되어진다. 단일기관(Single Authority) ABE에서는 하나의 기관이 모든 사용자의 속성을 관리하지만 다중 기관(Multi Authority) ABE에서는 각 기관이 각자의 속성들을 관리하고 사용자에게 발급한다. 즉, 사용자는 여러 기관에 속해서 각 기관들로부터 서로 다른 혹은 동일하지만 발급 주체가 다른 속성과 이에 해당하는 비밀키를 발급받아 사용하게 된다.

공모공격(Collusion Attack)은 여러 사용자 혹은 기관들이 공모하여 개인이 보유한 속성만

으로는 복호화할 수 없는 암호문을 복호화하거나 검증되지 않은 속성의 비밀키를 생성하는 등 악의적인 목적으로 자신이 가진 키를 공유하는 공격이다. 현재 사용되는 대부분의 다중기관 ABE 모델에서는 사용자 간의 공모공격을 방지하기 위해서 사용자 비밀키의 생성과정에서 사용자마다 서로 다른 고유한 비밀 값을 사용하여 공모공격을 방지한다[2]. 하지만 비밀 값을 결정하는 중앙기관(CA)의 단일 장애점 문제(Single point of failure) 및 사용자와 CA가 공모하여 다른 사용자의 비밀키를 생성할 수 있는 등 많은 한계가 존재한다[3].

본 논문에서는 기존 CA가 사용자마다 고유하게 결정하던 비밀값을 사용자가 스스로 결정하게 하되 이로 인해 발생할 수 있는 사용자의 악의적인 행동을 방지하기 위한 보증금 프로토콜을 제안한다.

## II. 제안 모델

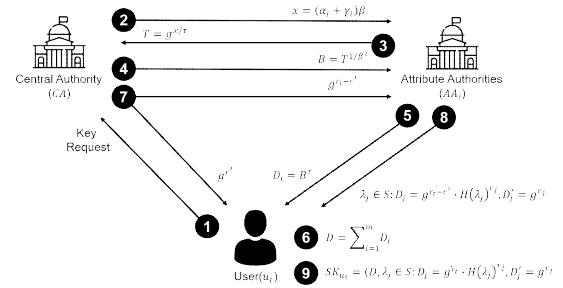
### 2.1 시스템 모델

Hur는 [4]에서 CA와 각 속성 발급기관들이 공동으로 사용자의 키생성을 알고리즘을 수행하는 분산된 속성기반암호 모델을 제시하였다. 이장에서는 Hur의 모델에 보증금 프로토콜을 적용하여 사용자의 공모공격 참여를 회피할 수 있는 모델을 제안한다. 제안 모델에서 고려하는 위협 모델은 다음과 같다.

- ◆ **사용자들 간의 공모:** 사용자들은 자신이 가진 속성키를 대가를 받고 공격자에게 공유할 수 있다. 공격자는 공유받은 속성키로 현재 가진 속성으로는 복호화할 수 없는 암호문을 복호화할 수 있다.
- ◆ **사용자와 CA의 공모:** CA는 이미 비밀키를 발급받은 사용자로부터 키를 공유받아 제3의 다른 사용자의 비밀키를 속성발급기관의 참여 없이 생성할 수 있다.

제안 모델의 핵심 개념은 Hur의 모델(그림 1)의 키생성 과정에서 CA가 사용자마다 고유하게 결정하던 값인  $r_t = \sum_{i=1}^m \gamma_i$ 를 사용자가 스스로

선택하는 것이다. 선택한 값을 CA가 알지 못하는 환경에서 사용자, CA, 그리고 각 속성 발급기관들이 협력하여 사용자의 비밀키를 생성한다. 하지만 사용자가 자신이 가진 값  $r_t$ 를 공모공격에 사용하지 않음을 보장할 수 없다는 문제가 발생한다.



[그림 1] [4]의 키 생성 프로세스

### 2.2 제안 모델

본 논문에서는 2.1절의 두 가지의 공모공격에 대한 저항성을 보장하기 위한 방법으로 블록체인 스마트계약을 사용한 보증금 프로토콜을 사용한다. 사용자는 시스템에 참여하기 전에 스마트계약에 보증금을 납부한다. 보증금은 사용자가 지정한 어떤 값에 대한 지식을 보임으로서 출금할 수 있다. 만약 사용자가 공모공격에 참여할 경우 공격과정에서 상대방에게 노출하는 값으로 상대방은 공모자의 보증금을 출금할 수 있게 된다. 시스템의 모든 참여자가 합리적인 선택을 하고 공모공격의 참여자들이 서로 신뢰하는 관계가 아니라고 가정할 때 제안 방법을 통해 공모공격을 회피하는 것이 가능하다. 자세한 보증금 프로토콜은 다음과 같다.

- (1) 사용자는 CA에게 키생성을 요청한다.
- (2) CA는 보증금 관리를 위한 스마트계약의 주소를 사용자에게 전달한다.
- (3) 사용자는 랜덤한  $r_t = \sum_{i=1}^m \gamma_i$ 를 선택하고  $hvalue = H(r_t)$ 를 계산한다( $H$ 는 해시함수).
- (4) 계산한  $hvalue$ 와 함께 보증금을 스마트계약에 제출한다(스마트계약은 그림 2의 출금 알고리즘과 같이  $hvalue$ 의 Pre-image를 출금을 위한 조건으로 설정한다).

- (5) 스마트계약에 보증금이 등록된 후 사용자, CA, 그리고 각 속성발급기관은 다자간계산(Multi-party computation)을 사용하여 Hur 모델의 키생성 과정을 수행한다(여기서 사용자의 입력값은  $r_t$ , CA와 각 속성발급기관의 입력은 각자의 마스터키이다).

```
withdraw( $\mathcal{D}_u$ , SC, proof, addrx)
```

---

```
If SC.hvalue ≠ H(proof) return 0  
SC.setstate(closed)  
return Send( $\mathcal{D}_u$ , addrx)
```

[그림 2] 보증금 출금 알고리즘

### 2.3 공모공격 저항성

- ◆ **사용자들 간의 공모:** 공격자 A와 B는 속성키를 결합하기 위해 동일한 값  $r_A = r_B$ 를 각각 CA와의 키 생성에 사용할 수 있다. 현실적인 가정에서 공모공격을 원하는 사용자들이 공격에 필요한 속성을 모두 가지고 있기는 어렵기 때문에 필요한 속성을 지닌 사용자에게 대가를 지불하고 속성키의 공유를 요청할 것이다. 하지만 이 경우 두 사용자의 보증금을 출금하기 위한 조건이 동일한 값으로 설정되어 상대에게 보증금이 노출될 수 있어 공격이 성사되지 않을 것이다.
- ◆ **사용자와 CA의 공모:** Hur 모델에서 CA는 속성발급 기관의 참여없이 임의 사용자의 비밀키를 생성하기 위해서 시스템의 사용자로부터 사용자의 비밀키를 전달받아 다음과 같이 키생성에 필요한 속성발급기관의 비밀정보( $g^\alpha$ )를 추출할 수 있다. 제안 모델에서는 CA가 선택하던  $r_t$ 를 사용자의 비밀정보로 설정하여 CA와 공모하기 위해서는 이 비밀정보를 CA에게 노출해야하므로 보증금이 노출될 수 있어 공격이 성사되지 않을 것이다.

$$\begin{aligned} D^\beta / g^{r_t} &= \left( g^{\frac{(\alpha_1 + \dots + \alpha_m) + r_t}{\beta}} \right)^\beta \times g^{-r_t} \\ &= g^{(\alpha_1 + \dots + \alpha_m) + r_t + (-r_t)} \\ &= g^{(\alpha_1 + \dots + \alpha_m)} \end{aligned}$$

## III. 결론

본 논문에서는 다중기관 속성기반암호 시스템에서 공모공격을 방지하기 위한 새로운 방법을 제시하였다. 전통적인 방법은 신뢰하는 CA에 의존하였으나 제안 방법은 사용자 개인의 이익을 위해 스스로 정직하게 행동하게 하는 방법으로 탈중앙화라는 기존 연구의 방향에 적합하다.

### [사사표기]

본 연구는 과학기술정보통신부 및 정보통신 기획평가원의 글로벌핵심인재양성지원사업의 연구결과로 수행되었음(2020-0-01596)

### [참고문헌]

- [1] A. Sahai and B. Waters, “Fuzzy identity-based encryption,” in Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, 2005, pp. 457
- [2] M. Chase, “Multi-authority attribute based encryption,” in Theory of cryptography conference. Springer, 2007, pp. 515 - 534.
- [3] E. Meamari, H. Guo, C.-C. Shen, and J. Hur, “Collusion Attacks on De-centralized Attributed-Based Encryption: Analyses and a Solution,” arXiv preprint arXiv:2002.07811, 20
- [4] J. Hur and K. Kang, “Secure data retrieval for decentralized disruption-tolerant military networks,” IEEE/ACM transactions on networking, vol. 22,no. 1, pp. 16 - 26, 2012, publisher: IEE

# NIST PQC Round 3 암호에 관한 성능 분석\*

이명훈<sup>1</sup>, 전찬호<sup>1</sup>, 허동희<sup>1</sup>, 김수리<sup>2</sup>, 홍석희<sup>3</sup>

고려대학교 (<sup>1</sup>대학원생, <sup>2</sup>박사 후 연구원, <sup>3</sup>교수)

## Performance Comparison of Round 3 candidates in the NIST PQC Standardization Project

MyeongHoon Lee<sup>1</sup>, Chanho Jeon<sup>1</sup>, Donghoe Heo<sup>1</sup>, Suhri Kim<sup>2</sup>, Seokhie Hong<sup>3</sup>

Korea University (<sup>1</sup>Graduate student, <sup>2</sup>Post Doc, <sup>3</sup>Professor)

### 요약

양자컴퓨터가 개발됨에 따라 기존에 사용 중인 공개키 암호화 알고리즘 및 전자 서명 방식은 Shor's Algorithm에 의해 더 이상 안전하지 않다는 사실이 알려져 있다. 미국 국립 표준 연구소 (NIST) 에서는 양자 컴퓨팅 환경에서 안전한 양자 내성 암호 표준화 작업을 진행 중이다. 표준화 작업은 현재 round 3이 진행 중이며 후보 알고리즘은 7종으로 KEM (Key Encapsulation Mechanism) Classic McEliece, CRYSTALS-Kyber, NTRU, Saber 4종, 전자 서명 알고리즘 CRYSTALS-Dilithium, Falcon, Rainbow 3종이 있다. 본 논문에서는 NIST security category I 에 해당하는 보안 강도에서 round 3에 올라온 양자 내성 암호 후보 7종에 대해 Reference implementation 코드와 AVX2 implementation 코드의 알고리즘별 속도를 비교하고 향후 연구 방향을 제시한다.

## I. 서론

컴퓨터의 계산능력이 향상됨에 따라, 향상된 컴퓨팅 환경에서도 안전하게 사용할 수 있는 암호가 요구되고 있다. 그 중, 양자컴퓨터 개발이 가시화된 현재 상황에서 양자컴퓨터가 실용화되면 기존에 사용 중인 공개키 암호화 및 전자 서명 방식은 Shor's Algorithm [1]에 의해 더 이상 안전하지 않다. 이에 대응하기 위해 양자 컴퓨팅 환경에서도 안전한 공개키 암호인 양자 내성 암호 (post-quantum cryptography, PQC) 가 필요해졌고, NIST에서는 2017년에 PQC 표준화 공모전을 시작하여, 현재 round 3이 진행 중이다.

양자 내성 암호는 기반 문제에 따라 크게 코드 기반 (Code-based), 격자 기반 (Lattice-based), 다변수 이차식 기반 (Multivariate-based), 아이소제니 기반 (Isogeny-based), 4가지로 분류할 수 있다. 이 중, round 3에는 1종의 코드 기반 암호인 Classic McEliece [2] 와 5종의 격자 기반 암호인 CRYSTALS-Kyber [3], NTRU [4], Saber [5], CRYSTALS-Dilithium [6], Falcon [7] 과 마지막으로 1종의 다변수 이차식 기반 암호인 Rainbow [8]

로 총 7종의 후보가 존재한다. 7종의 암호 중 KEM은 Classic McEliece, CRYSTALS-Kyber, NTRU, Saber로 4종, 전자 서명 알고리즘은 CRYSTALS-Dilithium, Falcon, Rainbow로 3종이다.

NIST에서는 위 7종의 알고리즘 중 KEM과 전자 서명 알고리즘 각각에서 양자 내성 암호의 표준 알고리즘을 선정할 예정이라고 공표했다.

본 논문에서는 NIST PQC round 3에 올라온 7종의 알고리즘 중 KEM 4종의 파라미터와 키 생성, 암호화, 복호화 수행 속도를 비교하고 3종의 전자 서명 알고리즘의 파라미터와 키 생성, 서명, 검증 수행 속도를 비교한다.

본 논문의 구성은 다음과 같다. 2장에서는 round 3에 올라온 7종의 알고리즘을 간략히 소개한다. 3장에서는 실험 결과를 통해 round 3에 올라온 KEM 4종과 전자 서명 알고리즘 3종의 파라미터 분석 및 수행 속도를 비교한다. 4장에서는 결론 및 알고리즘 개선 방향을 제안한다.

## II. Round 3 후보 7종

본 장에서는 NIST PQC round 3에 올라온 후보 7종을 간략히 소개한다.

\* 본 연구는 삼성전자의 지원(과제번호IO201209-07857-01)을 받아 수행된 결과임

## 2.1 KEM 4종

- *McEliece* : 선형 코드를 올바르게 디코딩 하는 것의 어려움에 기반하는 코드 기반 암호로, 이항 고파코드를 사용한다. 코드 기반 암호의 단점인 느린 복호화 속도를 보완한 알고리즘이다.
- *Kyber* : LWE (Learning With Errors) 문제에 대수적 구조를 더해 MLWE (Module-LWE) 의 어려움에 기반하는 격자 기반 암호로, 키 생성과 암복호화 과정이 빠르다는 장점이 있다.
- *NTRU* : 주어진 다항식을 작은 계수의 다항식 2개간의 나눗셈으로 표현하는 것의 어려움에 기반하는 격자 기반 암호로, 1996년에 제안돼 다른 격자 기반 암호에 비해 오랜 시간 암호학적으로 안전함이 검증되었다는 장점이 있다.
- *Saber* : LWR (Learning With Rounding) 문제에 대수적 구조를 더해 MLWR (Module-LWR) 의 어려움에 기반하는 격자 기반 암호로, 라운딩 과정에서 얻어지는 오차를 통해 안전성을 확보하므로 LWE기반 알고리즘에 비해 요구되는 연산량이 작은 알고리즘이다.

## 2.2 전자 서명 알고리즘 3종

- *Dilithium* : Kyber와 마찬가지로 LWE문제에 대수적 구조를 더한 MLWE의 어려움에 기반하는 격자 기반 암호로, Fiat-Shamir with abort방식을 사용하며 단순한 구현이 가능하고 키 생성, 서명, 검증과정이 빠르다는 장점이 있다.
- *Falcon* : NTRU격자 위에서의 SIS (Short Integer Solution) 문제의 어려움에 기반하는 격자 기반 암호로, hash-and-sign방식을 사용하며 다른 두 서명 알고리즘에 비해 공개키와 서명값의 크기 합이 작다는 장점이 있다.
- *Rainbow* : UOV (Unbalanced Oil and Vinegar) 문제의 어려움에 기반하는 다변수 이차식 기반 암호로, 공개키 크기는 크지만 서명의 크기가 작다는 장점이 있다.

## III. 실험 결과

본 장에서는 NIST PQC round 3에 올라온 KEM 4종과 전자 서명 알고리즘 3종의 성능을 비교한다.

실험에 사용된 PC의 CPU는 Intel(R) Core (TM) i7-6700 CPU @ 3.40 GHz이며, 운영체제는 Ubuntu 20.04.1 LTS, 컴파일러는 GNU GCC version 9.3.0를 사용했다.

Table 1은 NIST PQC round 3에 올라온 알고리즘 7종을 KEM 4종과 전자 서명 알고리즘 3종으로 나눈 표이다. 단위는 파라미터 항목에는 바이트를 사용했고, 속도는 ms를 사용했으며 팔호 안에 적힌 숫자는 클럭 사이클을 의미한다. 실험한 알고리즘들은 NIST security category I 기준이며 Dilithium에 대해서는 category I 기준의 파라미터가 존재하지 않아

category II 기준에서 진행했다. 마지막으로, 각 알고리즘에 대해 구현 방법에 따라 reference implementation 환경과 AVX2 implementation 환경에서 실험을 진행했다.

## 3.1 KEM 4종 비교

- **파라미터** : 격자 기반 암호인 Kyber, NTRU, Saber의 암호문 크기는 768, 699, 736바이트인 반면, 코드 기반 암호인 McEliece는 암호문의 크기가 128바이트로 격자 기반 암호들에 비해 약 5배 이상 짧지만, 공개키, 개인키가 격자 기반 암호들에 비해 크다. 특히, 공개키의 크기가 261,120바이트로 가장 짧은 NTRU의 공개키 크기인 699바이트에 비해 약 373배 크다. 격자 기반 암호 내에서는 NTRU의 파라미터가 전반적으로 작은 경향을 보인다.
- **속도** : 키 크기가 큰 McEliece는 키 생성 속도가 다른 암호들에 비해 현저히 떨어진다. 특히, AVX2 implementation 환경에서도 키 생성 시간이 13.883ms이 소요되어 나머지 3종 중 가장 오래 소요된 0.203ms (NTRU의 키 생성 속도)에 비해 약 68배 느린다. 격자 기반 암호 내에서는 Kyber의 키 생성 속도는 0.008ms이고 Saber의 키 생성 속도는 0.013ms로 0.203ms가 소요되는 NTRU보다 AVX2 implementation 환경에서도 10배 이상 빠름을 알 수 있다. 한편, 암복호화 속도는 4종의 알고리즘 모두 AVX2 implementation 환경에서는 0.05ms 이하의 속도를 보이는데 이 중 가장 느린 알고리즘은 NTRU이며 가장 빠른 알고리즘은 Kyber이다.

## 3.2 전자 서명 알고리즘 3종 비교

- **파라미터** : 격자 기반 암호인 Falcon, Dilithium에 비해 다변수 이차식 기반 암호인 Rainbow의 공개키는 161,600바이트로 Falcon의 897바이트에 비해 약 180배 크며 개인키는 103,648바이트로 Falcon의 1,281바이트에 비해 약 80배 크다. 반면, Rainbow는 서명의 크기가 66바이트로 Falcon의 666바이트에 비해 약 10배, Dilithium의 2,420바이트에 비해 약 36배 짧다는 장점을 가지고 있다. 전자 서명 알고리즘 중 파라미터 크기의 합이 가장 작은 알고리즘은 Falcon이다.
- **속도** : 키 생성 속도는 Dilithium이 AVX2 implementation 환경에서 0.03ms로 가장 빠르며 Falcon의 7.501ms에 비해 약 250배, Rainbow의 2.746ms에 비해 약 91배 빠른 속도이다. 하지만, 서명 속도는 Rainbow가 0.02ms로 Dilithium의 0.093ms에 비해 약 4배, Falcon의 0.229ms에 비해 약 11배 빠르며, 검증 속도는 Rainbow가 0.014ms로 Dilithium의 0.035ms에 비해 약 2배 Falcon의 0.044ms 비해 약 3배 빠른다. 파라미터가 상대적으로 작은 격자 기반 암호 Falcon과 Dilithium을 비교하면 Dilithium이 더 빠른 속도를 보여준다.

Table 1. Round 3 알고리즘 7종 파라미터 및 속도 실험 결과. 파라미터의 단위는 바이트이며. 속도의 단위는 ms이다. 단, 팔호 안의 숫자는 클럭 사이클을 의미한다.

구현 방법		Reference				AVX2			
KEM		McEliece	Kyber	NTRU	Saber	McEliece	Kyber	NTRU	Saber
파라미터	공개키	261,120	800	699	672	261,120	800	699	672
	개인키	6,492	1,632	935	1,568	6,492	1,632	935	1,568
	암호문	128	768	699	736	128	768	699	736
속도	키 생성	97,644 (347,594,363)	0.033 (109,697)	24,014 (81,705,608)	0.021 (70,965)	13,883 (46,643,366)	0.008 (28,411)	0.203 (692,708)	0.013 (44,423)
	암호화	0.049 (169,560)	0.041 (139,305)	0.719 (2,453,306)	0.027 (92,961)	0.015 (53,926)	0.011 (37,389)	0.051 (173,456)	0.015 (53,007)
	복호화	19,445 (66,057,189)	0.049 (165,877)	1,879 (6,366,401)	0.031 (104,615)	0.035 (123,266)	0.008 (27,631)	0.032 (111,629)	0.014 (51,136)
전자 서명		Dilithium	Falcon	Rainbow	Dilithium	Falcon	Rainbow		
파라미터	공개키	1,312	897	161,600	1,312	897	161,600		
	개인키	2,544	1,281	103,648	2,544	1,281	103,648		
	서명	2,420	666	66	2,420	666	66		
속도	키 생성	0.094 (306,687)	16,077 (55,331,852)	6,020 (20,561,806)	0.030 (102,199)	7,501 (24,239,478)	2,746 (8,804,507)		
	서명	0.436 (1,421,964)	4,589 (15,627,886)	0.075 (257,876)	0.093 (313,917)	0.228 (791,943)	0.020 (74,602)		
	검증	0.111 (363,506)	0.047 (156,081)	0.014 (47,324)	0.035 (119,420)	0.044 (151,819)	0.014 (53,823)		

## IV. 결론

본 논문에서는 NIST PQC round 3에 올라온 알고리즘 7종의 NIST security category I 기준의 파라미터 및 속도를 분석했다.

KEM은 Kyber가 다른 알고리즘에 비해 전반적으로 작은 파라미터와 빠른 속도를 가지며 전자 서명 알고리즘은 Falcon이 작은 파라미터를 가지고 속도는 Dilithium이 가장 빠르다는 것을 확인했다.

알고리즘별로 안전성 및 최적화에 관한 연구가 끊임없이 진행되고 있으므로, 현 상황에서 단순히 속도가 빠르다고 하여 해당 알고리즘이 선택될 것이라는 보장은 없다.

향후에는, round 3에 올라온 7종의 알고리즘에 대해 어떤 연산이 코드에서 가장 큰 비중을 차지하는지 확인하고 안전성을 유지하며 해당 연산에서의 효율적인 연산을 통해 알고리즘 속도를 개선하는 추가 연구를 진행할 계획이다.

## [참고문헌]

- [1] Shor, Peter W. "Algorithms for quantum computation: discrete logarithms and factoring." Proceedings 35th annual symposium on foundations of computer science. Ieee, 1994.
- [2] M.R. Albrecht, et al., "Classic McEliece: conservative code-based cryptography," NIST PQC round 3 submission, Nov. 19, 2020.
- [3] R. Avanzi, et al., "CRYSTALS-Kyber:

Algorithm Specifications And Supporting Documentation," NIST PQC round 3 submission, Oct. 1, 2020.

- [4] C. Chen, et al., "NTRU: Algorithm Specifications And Supporting Documentation," NIST PQC round 3 submission, Sep. 30, 2020.
- [5] A. Basso, et al., "SABER: Mod-LWR based KEM(round 3 Submission)," NIST PQC round 3 submission, Oct. 21, 2020
- [6] S. Bai, et al., "CRYSTALS-Dilithium: Algorithm Specifications and Supporting Documentation," NIST PQC round 3 submission, Oct. 1, 2020.
- [7] P.-A. Fouque, et al., "Falcon: Fast-Fourier Lattice-based Compact Signatures over NTRU: Specification v1.2," NIST PQC round 3 submission, Oct. 1, 2020.
- [8] J. Ding, et al., "Rainbow," NIST PQC round 3 submission, Oct. 1, 2020.

# 경량 블록 암호 PIPO에 대한 상관관계 전력분석 공격

심민주\*, 김현준\*\*, 서화정\*†

\*한성대학교 IT융합공학부 (대학원생)

\*\*한성대학교 정보컴퓨터공학과 (대학원생)

\*† 한성대학교 IT융합공학부 (교수)

## Correlation Power Analysis Attack on Lightweight Block Cipher PIPO

Min-Joo Sim\*, Hyun-Jun Kim\*\*, Hwa-Jeong Seo\*†

\*Hansung University, Department of IT Convergence Engineering  
(Graduate student)

\*\*Hansung University, Department of Information Computer Engineering  
(Graduate student)

\*† Hansung University, Department of IT Convergence Engineering  
(Professor)

### 요약

다양한 통신환경에서 보안 요소와 성능을 충족시키기 위해서 경량 블록 암호 알고리즘이 필요하다. 이론적인 안전성이 증명된 국산 블록 암호 알고리즘으로는 SEED, ARIA, HIGHT, LEA가 있다. 최근 ICISC 2020에서 처음 국산 경량 블록 암호 PIPO 가 발표되었다. 본 논문에서는 PIPO에 대한 상관관계 전력분석 공격을 시도하여 PIPO가 부채널 공격에 대한 취약성을 갖고 있음을 증명한다.

### I. 서론

사물인터넷(IoT) 환경에서 다양한 장비에서 인터넷 통신이 가능하다. 다양한 통신환경에서 보안 요소와 성능을 충족시키기 위한 경량 블록 암호 알고리즘이 필요하다. 다양한 환경에서의 적절한 알고리즘의 필요성이 요구되어 여러 경량 블록 알고리즘이 제안이 되고 있다. 이론적인 안전성이 증명된 국산 블록 암호 알고리즘으로는 SEED, ARIA, HIGHT, LEA가 있다. 하지만 블록 암호에서 부채널 분석에 취약점이 존재한다[1,2]. 최근 ICISC 2020에서 처음 발표된 국산 경량 블록 암호 PIPO가 발표되었다[3]. 본 논문에서는 새로 발표된 PIPO-64/128에 대한 전력분석 공격 실험을 통하여 마스터키 값을 획득할 수 있음을 확인하

였다. 해당 실험은 XMEGA 보드에서 동작하는 PIPO 알고리즘을 대상으로 전력분석을 수행하였다.

본 논문의 구성은 다음과 같다. 2장에서 PIPO에 대한 소개 및 전력분석 공격 방법에 대해 간략하게 서술한다. 3장은 PIPO-64/128이 갖는 보안 취약점을 이용한 공격 방법을 최초로 제시한다. 마지막으로 4장에서 본 논문에 대한 결론을 내린다.

### II. 관련 연구

#### 2.1 경량 블록 암호 PIPO 알고리즘

ICISC 2020에서 발표된 PIPO는 Plug-In과

Plug-Out의 약자로, 각각 부채널 보호 환경과 비보호 환경에서 사용된다. Table 1은 키 사이즈에 따른 PIPO의 설명이다.

Cipher	Size of blocks	Key length	Number of rounds
PIPO-128	64	128	13
PIPO-256	64	256	17

Table 1. Specification of PIPO

단순한 형태로 더 작은 S-box를 조합한 Unbalanced-Bridge 구조의 S-Layer로 Fig. 1과 같이 구성된다.

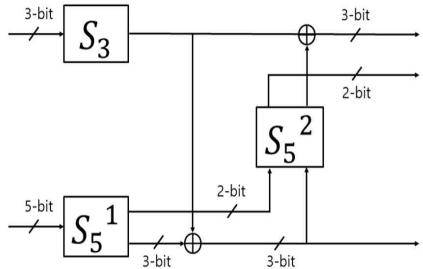


Fig. 1. Unblanced-Bridge

PIPO의 전체적인 구조는 Fig. 2에서 확인할 수 있다. PIPO의 각 라운드는 S-Layer로 표현된 비선형 레이어, R-Layer로 표현된 선형 레이어, 라운드 키와 상수의 XOR 연산으로 구성된다. 라운드 키도 단순하게 키를 평문의 길이처럼 64bit로 나눠 반복된 키를 사용한다.

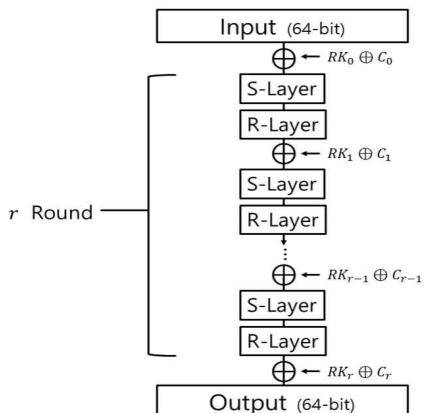


Fig. 2. PIPO overall structure

여기서  $RK_0$ 은 whitening 키이며,

$RK_1, RK_2, \dots, RK_r$ 은 라운드 키이다. S-Layer는 8개의 동일한 8bit s-box(S8)를 병렬로 실행한다. R-Layer는 각 행의 비트를 주어진 오프셋으로 회전시킨다[3].

## 2.2 상관관계 전력분석 공격

전력분석 공격은 보통 데이터 수집 단계와 데이터 분석 단계 총 2단계를 거쳐 공격이 진행된다. 먼저 데이터 수집 단계에서는 랜덤하게 선택된 평문을 이용하여 암호화 연산을 수행한 후, 해당 연산에 대한 소비전력 파형을 수집한다. 이후 데이터 분석 단계에서 비밀키 일부분에 대한 그 값을 예측한 후, 예측값과 입력된 평문을 이용하여 내부 연산 값을 계산한다. 이렇게 얻은 계산 값의 유효성을 수집된 전력 소비 파형을 이용하여 검증하는 과정을 반복하면서 비밀키 전체 값을 복구한다[4].

상관관계 전력분석 공격(correlation power analysis, CPA)은 암호 모듈을 계속 동작시키고 고정된 비밀키에 다른 평문을 입력으로 넣어 암호문을 얻는 동시에 파형 수집 장치를 통해 파형들을 수집한다. 수집된 파형, 평문, 암호문을 통해 설계된 분류 함수를 기준으로 파형 통계 처리하여 분석하는 방식이다[5].

## III. 실험 결과

실험은 PIPO-64/128를 대상으로 8bit 프로세스 XMEGA 보드에 Chipwhisperer를 사용하여 13라운드의 S-box의 파형을 10,000개 수집하였다. 실험 결과의 효율성을 위해 마스터키값을 고정 값으로 놓고 진행하였다. CPA 공격을 진행하기 위해서 비트 단위 병렬로 S-Layer가 구성된 특성으로 인해 S-box 연산 후의 값을 중간값으로 사용하여 공격을 진행하였다.

중간값 8bit에서 1bit씩 공격을 수행하였을 때, 각각의 비트마다 공격 결과가 다르게 나온 것을 확인하였다. 가장 높은 상관계수 값이 1개였기 때문에 얻고자 하는 키값을 얻을 수 있었다. Fig. 3처럼 최상위 비트, 네 번째, 다섯 번째, 여섯 번째 비트에서는 획득하고자 하는 1라운드의 라운드 키값을 얻었다.

Fig. 3.에서 가장 큰 상관계수 값

(0.33129548988263025)이 나타났다. 이는 상위 네 번째 비트가 공격이 가장 잘 되는 비트였음을 확인할 수 있었다.

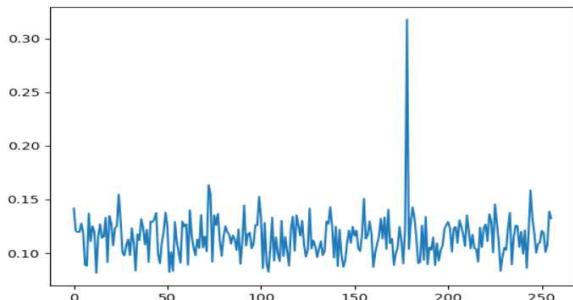


Fig. 3. The 4<sup>th</sup> high-order bit is the best attack. The highest correlation coefficient value is represented at (0xb2).

반면, Fig. 4처럼 상관계수 값이 같은 한꺼번에 여러 개의 키값이 나온 비트 공격 결과도 확인하였다. 이와 같은 결과는 두 번째 상위 비트, 세 번째, 일곱 번째, 최하위 비트가 해당되는 것을 확인하였다. 각각 같은 상관계수 값은 4개, 2개, 2개, 2개를 갖는다. 이와 같은 결과를 얻게 되면 해당 비트에서는 완벽한 라운드 키값을 얻을 수는 없다.

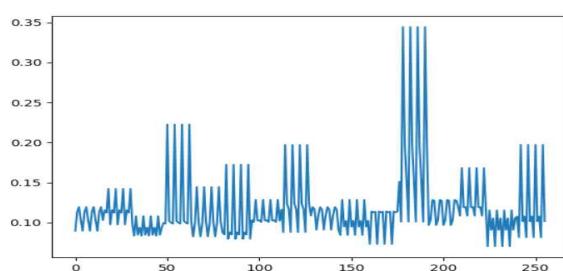


Fig. 4. The second high-order bit has 4 correlation coefficients. The highest correlation coefficient value is represented at (0xb2, 0xb6, 0xba, 0xbe).

하지만, Fig. 4에 나타난 두 번째 상위 비트 값의 결과값으로 분석한 결과, 각각 0xb2, 0xb6, 0xba, 0xbe의 8비트 중 6개의 모두 같은 비트 값을 지니고 있다는 것을 확인하였다. 이는 얻고자 하는 키값은 총 8비트에 해당하여 정확한 키값을 얻을 수는 없지만, 해당 공격으로 총 6개의 비트를 얻을 수 있음을 의미한다.

결과적으로, 각각 공격 결과가 달랐음에도

불구하고, 0~7 사이의 비트마다 모든 공격한 결과값에서 획득하려는 값인 0xb2가 모두 포함된 것을 확인하였다.

#### IV. 결론

본 논문에서는 경량 블록 암호 알고리즘 PIPO-64/128을 대상으로 저사양 8bit AVR 프로세서상에서의 상관관계 전력분석 공격에 대한 취약성을 확인하였다.

#### V. Acknowledgment

이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478).

#### [참고문헌]

- [1] T. Kim, Y. Won, J. Park, H. An, D. Han, “Side Channel Attacks on HIGHT and Its Countermeasures” Journal of the Korea Institute of Information Security & Cryptology 25(2), 2015.4, 457-465(9 pages), 2015.
- [2] D. Hong, J. Sung, S. Hong, J. Lim, “HIGHT : a new block cipher suitable for low-resource device”, CHES 2006, LNCS 4249:46~59, Oct. 2006.
- [3] <https://eprint.iacr.org/2020/1582.pdf>
- [4] 백유진. (2020). 전력분석공격에 대한 하드웨어 마스킹 대응기법 동향. 정보보호학회지, 30(1), 23-33.
- [5] J. Kim, K. Oh, Y. Choi, T. Kim, D. Choi, “Side channel analysis system technology trend”, Electronic Communication Trend Analysis, 28(3), 2013.
- [5] D. Kwon, S. Jin, H. Kim, S. Hong, “Improving Non-Profiled Side-Channel Analysis Using Auto-Encoder Based Noise Reduction Preprocessing”, Journal of The Korea Institute of Information Security & Cryptology, Vol.29(3), pp. 491-501 , Jun. 2019.

# 사물인터넷 장치 통합 제어용 게이트웨이 인증 알고리즘 개선 연구 - A제품을 기반으로

정소영\* 박준용\*\* 오인수\*\* 김찬민\*\* 임강빈\*\*\*

\*순천향대학교(대학생), \*\*순천향대학교(대학원생), \*\*\*순천향대학교(교수)

Study on the Improvement of Authentication Algorithm of Gateway for Integrated Control of IoT Devices-Based on Product A

Jung So Young\* JunYoung Park\*\* Insu Oh\*\* Chanmin Kim\*\* Kangbin Yim\*\*\*

\*Soonchunhyang University(Undergraduate student)

\*\*Soonchunhyang University(Graduate student)

\*\*\*Soonchunhyang University(Professor)

## 요약

사물인터넷 장치는 경량화 및 소형화의 조건을 충족하면서도 적은 전력 소모와 원활한 무선 통신 성능이 보장함으로써 사용자에게 편리한 서비스 환경을 제공한다. 하지만 이러한 조건들을 충족하기 위해 설계와 개발 단계에서의 보안 요구사항에 대한 인지 결여가 시스템 내외부적으로 심각한 보안 취약점으로 이어진다. 본 논문에서는 시중에 판매 중인 사물인터넷 제품 중 게이트웨이 제품군 중 A 대상을 선정하여 이에 대해 발생할 수 있는 취약점을 분석하며 특히 인증 과정에서 취약성을 보완하고 암호화 기법 등을 적용하여 보안성을 개선한 알고리즘을 제안한다.

## I. 서론

사물인터넷은 사물이라 불리는 다양한 형태의 사물 장치들이 상호 연결되어 인터넷 네트워크상에서 데이터 수집과 교환, 공유와 같은 기능을 적절히 활용할 수 있는 서비스 환경을 제공하여 사용자에게 편의를 제공하는 기술이다. 통신 네트워크 기술 발전에 따른 무선 네트워크 인프라의 확장과 생산 기술 고도화에 따른 센서 생산에 필요한 비용 절감 등 다양한 요인이 상호작용한다. 이에 따라 사용자가 사물인터넷 기술을 이용함에 받는 혜택이 줄었으며 다른 산업 기술과 융합되어 일상에서 사용하는 가전 소품부터 산업 분야까지 기술에 대한 접근성 및 확장성이 좋아졌다[1]. 하지만 사물인터넷의 활용 범위가 확장됨에 따라 사물 간의 통신에 취급되는 사용자 정보와 사용 환경 정보의 양 또한 축적되므로 이를 목표로 한 사이버 공격과 위협이 지속해서 발생하는 추세이다[2].

임베디드 운영체제 기반의 사물인터넷은 경량화와 소형화의 조건을 충족하면서 일반 가전 장치보다 전력 소모량이 적고 무선 통신 기술 성능의 보장을 요구한다. 때문에 장치와 보드의 설계 및 개발 단계에서 보안 지식 부족으로 인해 보안 요구사항이 충족하지 못하는 경우가 발생할 수 있다. 이로 인해 시스템 내외부적으로 공격자에게 무방비하게 노출될 수 있다[3].

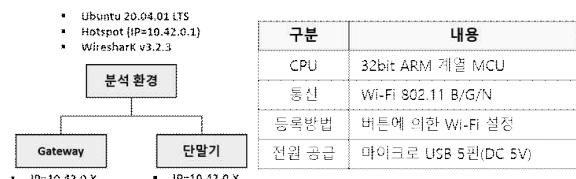
본 논문은 사물인터넷 장치 중 네트워크에 연결된 다수의 사물의 시스템 상태 정보 수집과 제어가 가능한 게이트웨이 제품군을 선정하여 분석한다. 사물인터넷 게이트웨이는 특히 사물 간의 모든 통신에 대한 네트워크 게이트웨이 역할을 수행하므로 이에 대한 공격 발생 시 정보 유출, 제어권 상실 등의 피해가 발생할 수 있다[4]. 따라서 본 논문에서는 사물인터넷 게이트웨이 A 제품을 선정하여 사용자의 인증 과정에 발생할 수 있는 취약점에 대해 분석하고 이에 대한 개선된 인증 알고리즘을 제안하고자 한다.

## II. 사물인터넷 게이트웨이 인증 알고리즘 취약점 분석

### 2.1 분석 대상

사물인터넷 게이트웨이는 위한 사물과의 연결 수립 과정에서 무결성을 보장하는 인증 과정을 요구한다. 또한 연결 수립 이후 게이트웨이와 사물 간의 신뢰성 검증과 보장을 위한 암호화 과정 등이 게이트웨이의 보안 요구사항에 해당된다. 게이트웨이에서 보안 요구사항의 불충족으로 인한 취약점 노출은 사용자 정보와 사용 환경에 대한 정보 노출에서부터 사생활 침해, 제3자에 의한 제어권 상실 등의 피해가 발생한다. 실제 게이트웨이 장치 A를 선정하여 시중에 판매 중인 제

품에 대한 취약점 존재 여부에 대해 분석했다.

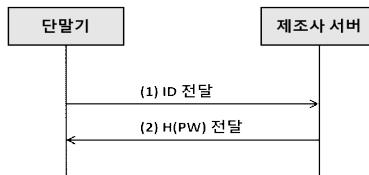


[그림 1] 분석 대상 정보 및 분석 환경

위 [그림 1]은 분석 대상의 성능과 분석 환경의 도식화이다. 게이트웨이 제어를 위해 제조사에서 제공하는 어플리케이션과 게이트웨이 장치 사이에 발생하는 네트워크 통신 패킷을 수집하여 발생 가능한 취약점을 분석했다.

## 2.2 로그인 과정 분석

\* 로그인 과정



[그림 2] 알고리즘 - 로그인

게이트웨이 제어는 제조사의 모바일 어플리케이션으로 수행하며 위 [그림 2]와 같다. (1) 서버로 사용자의 ID를 전달하고 PW는 해시 연산을 위해 임시 저장한다. (2) ID의 해시 값인 H(PW)를 어플리케이션으로 전달한다. H(PW)는 어플리케이션 내부에서 PW 평문과 검증 알고리즘 연산을 통해 일치 여부를 판단한다.

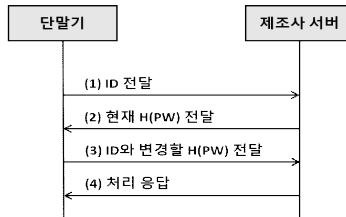


[그림 3] 로그인 패킷

위 [그림 3]는 로그인 과정에서 수집한 (1)과 (2)의 패킷이다. 로그인 과정에서 서버의 통신 시 UDP 페이로드 내 사용자의 ID와 H(PW)가 평문으로 노출되는 취약한 구조임을 알 수 있다.

## 2.3 패스워드 변경 과정 분석

\* 패스워드 변경 과정



[그림 4] 알고리즘 - 패스워드 변경

위 [그림 4]는 어플리케이션 내 패스워드 변경 과정의 도식화이다. (1) 서버로 ID를 전달한다. (2) ID의

H(PW)를 전달한다. (3) 패스워드 검증 연산 후 변경할 PW를 해시 연산하여 ID와 전달한다. (4) 패스워드 변경 처리에 대해 응답한다.



[그림 5] 패스워드 변경 패킷 - 과정 (2)



[그림 6] 패스워드 변경 패킷 - 과정 (3)

위 [그림 5, 6]에서 패스워드 해시 값과 변경할 패스워드 해시 값 또한 노출됨을 확인했다. 해당 취약점에 대해 패스워드 임의 변조 실험을 수행했다. [그림 6]은 변경할 H(PW)를 전달하는 패킷이며 해당 H(PW)의 내용 변조 후 재전송하여 수행했다. 기존 패스워드 [12345678]의 해시 값을 [bbbbbbbb]의 해시 값으로 대체했다.



[그림 7] 패스워드 변조 실험

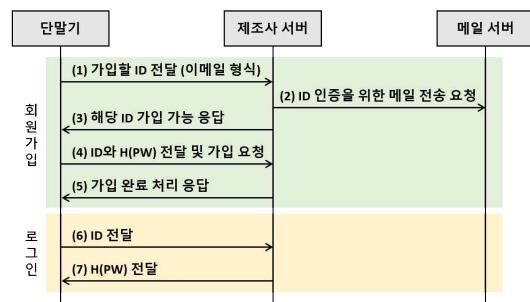
위 [그림 7]은 패스워드 [bbbbbbbb]로 로그인이 성립했으며 본 실험을 통해 사용자 인증 과정이 취약함을 보여준다.

## III. 사물인터넷 게이트웨이 인증 알고리즘 개선 방안 제안

### 3.1 기존 인증 알고리즘의 문제점

게이트웨이 A는 어플리케이션을 통해 제어하며 해당 어플리케이션은 사용자의 인증을 과정에 사용되는 해시 연산에 대하여 Bcrypt 알고리즘을 사용한다[5].

\* 기존의 회원가입과 로그인 알고리즘



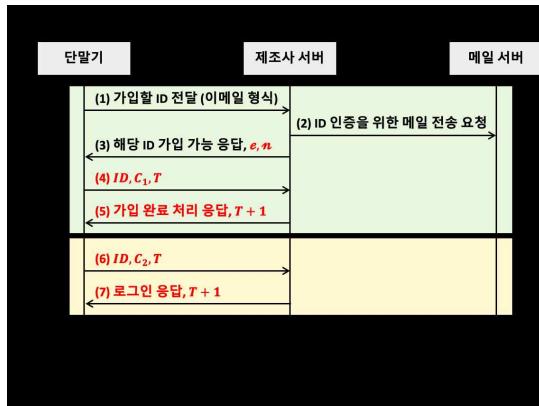
[그림 8] 기존 인증 알고리즘

위 [그림 8]은 어플리케이션 상에서 회원가입과 로그인을 처리하는 알고리즘을 정리한 과정의 도식화이다. (4)(6)(7)의 과정에서 회원가입과 로그인 시 평문으로 사용자의 ID와 H(PW)이 노출되며 이를 이용한 패스워드 임의 변조의 가능성을 앞의 2장에서 확인했다. 따라서 패킷을 암호화하여 보안성이 개선된 인증 알고리즘을 제안한다.

### 3.2 보안성이 개선된 인증 알고리즘 제안

[표 2] 알고리즘 내 수식 정리

$n$	법위 내 가장 큰 두 정수의 곱
$e$	$\gcd(e, \phi(n)) = 1$ 가 1인인 공개키
$d$	$d \cdot e = 1 \bmod \phi(n)$ 인 개인키
$K$	대칭적 AES 알고리즘의 비밀키
$H(PW)$	Bcrypt 알고리즘으로 연산된 패스워드 해시 값
$T$	타임스탬프



[그림 9] 알고리즘 제안

위 [그림 9]는 보안성을 개선한 인증 알고리즘을 도식화한 그림이다. 인증 알고리즘의 기밀성과 사용자 인증을 보장을 위해 AES[6]과 RSA[7]으로 암호화하고 타임스탬프로 메시지 무결성을 보장한다.

(1) 사용자가 가입할 아이디를 서버로 전달한다. (2) 서버는 이메일 형식으로 전달받은 아이디의 중복 여부와 유효 메일 확인을 위해 메일 서버에 인증을 요청한다. (3) 아이디의 가입 가능 여부 응답과 공개키 RSA 연산의에 이용할 공개키  $e$ 와 두 소수의 곱  $n$ 을 전달한다.

$$C = (E(H(PW), K))^e \bmod n$$

[수식 1] 암호문 구조

(4) 사용자의 아이디와 패스워드를 입력받은 어플리케이션은 아이디와 위 [수식 1]에 해당하는 암호문 그리고 무결성 인증 목적의 타임스탬프를 서버로 전달한다. (5) 서버는 개인키  $d$ 로 암호문을 복호화하여 사용자의 패스워드 해시 값을 서버 내 데이터베이스에 저장하며 응답 처리와 함께 한 단계 증가한 타임스탬프를 전달한다. (6) 평문 노출의 문제가 있었던 로그인 과정에서도 역시 패스워드 해시 값을 위 [수식 1]과 같이 전달한다.

위와 같은 기밀성과 송수신자 인증, 메시지 무결성이 보장되는 알고리즘을 제안함으로써 사물인터넷과 어플리케이션 사이 통신 속도 측면에서의 성능 저하가 예상된다. 하지만 기존의 인증 알고리즘과 비교 시 암호화와 타임스탬프를 적용함으로써 메시지 변조와 재전송의 방지가 가능하여 사용자 안전성이 보장된다.

### IV. 결론

본 논문은 사물인터넷 제품 중 게이트웨이 제품군 A를 선정하여 취약점을 분석했다. 로그인 과정과 패스워드를 변경하는 과정 각 과정에서 패스워드 해시 연산 알고리즘 파악이 가능했고 평문 문자열의 노출됨을 확인했다. 또한 패스워드 변경 과정에서 발생한 패킷의 변조 및 재전송 결과로 패스워드 변조 성립을 확인했다.

인증 구조상의 취약점으로 판단하여 인증 알고리즘을 분석하고 보안성이 추가된 알고리즘을 제안한다. 이 외에도 암호화되지 않은 통신에 대한 다양한 패킷 변조 및 재전송 공격이 가능함으로 보이며 앞으로 사물인터넷 장치를 보다 안전하게 이용하기 위한 보안 알고리즘 및 사물인터넷에 적합한 경량 암호화 기술에 대한 연구가 지속될 것으로 예상한다.

### 참고문헌

- [1] 황원식, “산업 패러다임에 따른 미래 제조업의 발전전략 (2) 사물인터넷(IoT)이 가져올 미래의 산업변화 전망”, KIET 산업경제 2016년 3월, pp. 15–22, 2016.
- [2] “IoT 취약점 5년새 54배 늘었다.” 아이뉴스24, <http://www.inews24.com/view/1185014>.
- [3] “IoT 보안 취약 신고, 5년간 1400여건”. 이뉴스투데이, <http://www.enewstoday.co.kr/news/articleView.html?idxno=1343147>.
- [4] 신승혁. “빅 데이터 처리를 위한 개방형 플랫폼 기반 IoT 센서 미들웨어의 설계.” 박사학위, 금오공과대학교 대학원, 2015.
- [5] Bcrypt, <https://www.npmjs.com/package/bcrypt>
- [6] Simon Heron, “Advanced Encryption Standard (AES)”, Network Security, Volume 2009, Issue 12, 2009, Pages 8–12, [https://doi.org/10.1016/S1353-4858\(10\)70006-4](https://doi.org/10.1016/S1353-4858(10)70006-4).
- [7] R. Rivest, A. Shamir, L. Adleman, “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems”, Communications of the ACM, Vol. 21 (2), 1978, pages 120 - 126.

# 해시함수 LSH 양자 회로 최적화를 통한 그루버 알고리즘 적용 자원 추정

송경주\* 장경배\* 서화정\*\*†

\*한성대학교 (대학원생)

\*\*한성대학교 (교수)

Resource Estimation Applying Groover Algorithm through Hash Function LSH Quantum Circuit Implementation

Gyeong-Ju Song\*, Kyung-Bae Jang\*, Hwa-Jeong Seo\*\*†

\*Hansung University(Graduate student)

\*\*Hansung University(Professor)

## 요약

그루버 알고리즘은  $n$ -bit의 보안 레벨의 대칭키 암호와 해시 함수를  $2/n$ -bit 보안 레벨까지 낮출 수 있는 양자 알고리즘이다. 그루버 알고리즘은 양자 컴퓨터상에서 동작하기 때문에 적용 대상이 되는 대칭키 암호와 해시함수는 양자 회로로 구현되어야 한다. 이러한 연구 동기로, 최근 들어 대칭키 암호 또는 해시 함수를 양자 회로로 구현하는 연구들이 활발히 수행되고 있다. 본 논문에서는 국산 해시함수 LSH를 양자 회로로 최적화하여 구현하였다. 큐빗 재활용, 사전 연산 그리고 Mix, Final 함수와 같은 핵심 연산들을 효율적으로 구현하여 국산 해시함수 LSH를 양자 회로로 설계하는데 필요한 양자 자원을 평가하였다.

## I. 서론

그루버 알고리즘은  $n$ -bit 의 보안레벨을 가지는 대칭키 알고리즘과 해시 함수에 대하여  $n/2$ -bit 보안레벨까지 낮출 수 있는 가장 효과적인 양자 공격 방법이다[1]. 양자 알고리즘인 그루버 알고리즘을 활용하기 위해서는 적용 대상 암호나 해시 함수를 양자 회로로 구현해야만 한다. 현재 양자 컴퓨터는 아직 초기 개발 단계이기 때문에 사용 가능한 큐빗이 매우 적고, 복잡한 양자 연산에서 생기는 오류를 제어하기 매우 까다롭다. 때문에 양자 회로를 최적화 구현하여 효율적인 자원을 사용하는 것이 매우 중요하다. 이러한 연구 동기에 따라 최근에는 대칭키 암호와 해시 함수를 양자 회로로 최적화 구현하는 연구들이 많이 진행되고 있다. [2]는 블록암호 AES를 양자 회로로 구현하여 그루버 알고리즘을 적용하는데 필요한 양자 자원을 추정하였고, [3,4]는 앞선 연구보다 최적화

된 AES 양자 회로를 제안하였다. [5]는 미국 NSA(National Security Agency)에서 개발한 경량 블록암호 SIMON을 양자 회로로 구현하였으며 [6]에서는 SPECK을 양자 회로로 구현하였다. 마지막으로 [7]은 국산 경량 블록 암호 HIGHT, LEA, CHAM을 양자 회로로 구현하였다.

본 논문에서는 해시함수 LSH를 양자회로로 최적화하여 구현하였다. 해시함수는 임의의 길이의 비트 열을 입력 받아 고정 길이의 해시 값을 출력하는 함수이다. LSH는 2014년 국가보안기술연구소에서 개발한 해시함수로, 높은 안전성과 우수한 효율성을 제공한다. 안정성으로는 충돌 쌍 공격, 역상 공격 등 해시 함수와 관련한 공격에 대하여 안전하게 설계되었으며, 국외 전문 연구기관인 벨기에 COSIC(Computer Security and Industrial Cryptography) 연구소로부터 안정성에 대한 객관적 검증을 받았다. 효율성은

국제 표준해시함수인 SHA-2와 SHA-3와 비교하여 3배 이상의 우수한 성능을 갖는다.[8]

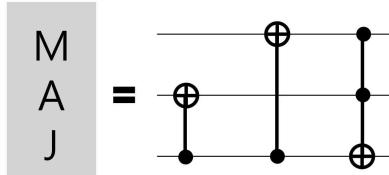
w 비트 워드 단위로 동작하며 n 비트 출력값을 가지는 해시함수 LSH-8w-n으로 구성된 해시함수 군(Hash Function Family)이다. LSH는 w는 32 또는 64이며 n은 1과 8w 사이의 정수이다. 본 논문에서는 LSH-256/256을 양자회로로 최적화 구현하여 이에 그루버 알고리즘을 적용하기 위한 양자 자원들을 추정하였다.

## II. 관련연구

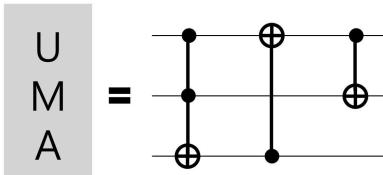
### 2.1 quantum ripple-carry addition[9]

본 논문에서는 ripple-carry addition 양자회로를 사용하였다. rippl-carry addtion 양자회로는 덧셈의 대상이 되는 2개의 입력 값 중 한 개의 입력 값에 덧셈 결과를 저장하기 때문에 큐빗을 최소화 할 수 있다.

그림 1은 MAJ 양자 회로이다. 두 개의 CNOT 게이트와 한 개의 Toffoli 게이트가 사용된다.



(그림 1) MAJ quantum circuit



(그림 2) UMA quantum circuit

그림 2 는 UMA 양자 회로이다. 각 큐비트에서 동일한 계산하는 함수를 두가지 제공한다. 2 개의 CNOT 게이트와 한 개의 Toffoli 게이트가 사용된다.

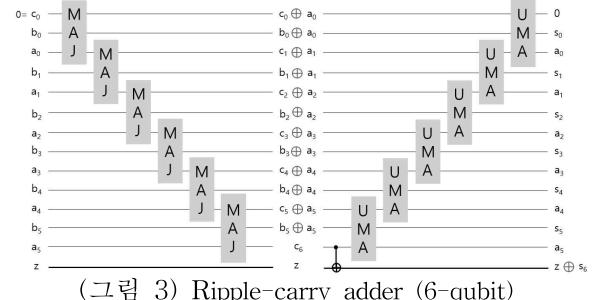
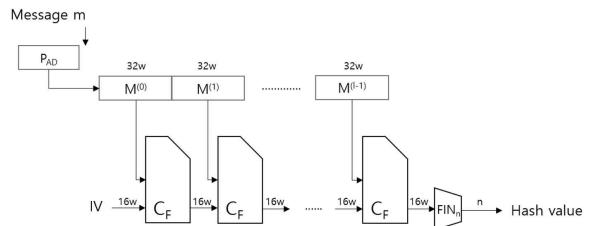


그림 3은 6-qubit ripple-carry addition 양자 회로이다. MAJ 및 UMA 의 조합으로 구성되어 있다.

### 2.2 해시함수 LSH[10]

LSH를 구성하는 각 해시함수는 그림 4와 같은 전체 구조를 가지며, 입력 메시지에 대해 초기화, 압축, 완료 세 가지 단계를 거쳐 해시값을 출력한다.



#### 2.2.1 완료 함수 $\text{FIN}_n$

완료 함수  $\text{FIN}_n : W16 \rightarrow \{0,1\}^n$ 은 압축 과정의 결과로 생성된 마지막 연결 변수값  $CV(t) = (CV(t)[0], \dots, CV(t)[15])$ 에 적용되어 n 비트 길이의 해시값 h를 생성한다.  $H = (H[0], \dots, H[7])$ 를 8 워드 배열,  $hb = (hb[0], \dots, hb[w-1])$ 는 w 바이트 배열이라고 하면, 함수  $\text{FIN}_n$ 은 수식 1을 수행한다.

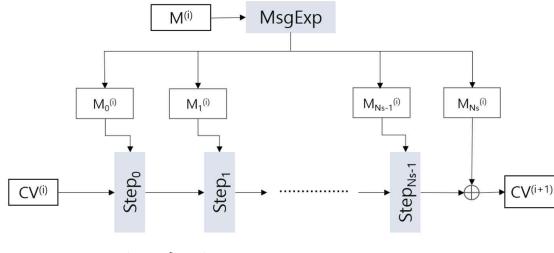
$$H[i] = CV^{(t)}[i] \oplus CV^{(t)}[i+8] \quad (0 \leq i \leq 7), \\ h_b[s] = H[(8s/w)] \gg (8s \bmod w) [7:0] \quad (0 \leq s \leq (w-1)), \\ h = (h_b[0] \mid \mid \dots \mid \mid h_b[w-1])_{[0:n-1]}$$

(수식 1)  $\text{FIN}_n$  function

#### 2.2.2 압축 함수

압축 함수의 입력값 중 메시지 블록  $M(j)$ 는 메시지 확장 함수  $\text{MsgExp}$ 를 거쳐  $(N_s + 1)$  개의 16 워드 배열  $M_j(i)$  ( $0 \leq j \leq N_s$ )로 확장된다. 이어서 16 워드 배열 크기의 임시 변수  $T = (T[0], \dots, T[15])$ 에 초기값을  $CV(i)$ 로 할당한 후, 순차적으로 단계 함수  $\text{Stepj}$ 를 통해

$M_j(i)$ 를 처리하면서  $T$ 를 갱신한다. 단계 함수를 거쳐  $T$ 에 저장된 값은 MsgAdd 함수를 통해 처리된 후  $(i + 1)$  번째 단계 변수  $CV(i+1)$ 에 입력된다.



(그림 5) Compression function

### III. 제안기법

LSH-256/256를 대상으로 양자 회로를 구현하였으며, 입력 메시지를 패딩 한 1024비트 평문  $M$ 을 위한 1024큐빗, 라운드 함수의 입력 값이 되는  $CV$ 를 위한 1024 큐빗, ripple-carry 덧셈의 캐리 값을 저장하기 위한 1큐빗, 마지막으로 최종 해시 값을 저장하는 256큐빗이 사용되었다.

초기 할당된 메시지  $M$ 을 재활용 하였으며 라운드 함수에서 사용되는 SC값을 위한 큐빗을 할당하지 않고 사전 연산 값에 따라 라운드 함수를 구현하였다. 이를 통해 SC를 위한 큐빗들을 할당하지 않았으며 SC를 업데이트 하는 양자 게이트 또한 사용되지 않았다. 제안하는 기법을 통해 앞서 언급한 큐빗 외 추가 큐빗을 할당하지 않으며, 양자 게이트의 사용 또한 최적화 하였다.

LSH-256/256에서는 1024비트의 메시지를 1 word (word=32bit) 씩 16개의  $M[t]$  ( $0 \leq t \leq 15$ )로 나눠 Step(단계함수)을 진행한다.  $M$ 은 업데이트 되어 매 라운드에 사용되는데 제안하는 기법에서는  $M$ 의 업데이트 값을 위한 큐빗을 할당하지 않고 기존 사용된  $M$ 에 새로운 값을 생성함으로써 큐빗을 절약하였다. 연결 변수  $T[i], T[i+8]$  ( $0 \leq i \leq 7$ )는 MsgExp, Mix, WordPerm 함수를 진행하며  $T$ 를 갱신하고 최종적으로 Final함수를 통해 해시 값을 얻는다. 양자회로를 이용한 Mix 함수 구현은 알고리즘 1과 같다.

---

#### Algorithm 1 : Quantum circuit implementation of Mix

---

**Input:**  $T[i], T[i+8], SC[i]$  ( $0 \leq i \leq 7$ )  
**Output:**  $T = \{T[0] \dots T[15]\}$

- 1: **ripple\_carry\_add** ( $T[i], T[i+8]$ )
- 2: **a\_rotation**( $T[i]$ )
- 3: Applying **X gate** to  $T[i]$  according to  $SC[i]$
- 4: **ripple\_carry\_add**( $T[i], T[i+8]$ )
- 5: **b\_rotation**( $T[i+8]$ )
- 6: **ripple\_carry\_add**( $T[i+8], T[i]$ )
- 7: **c\_rotation**( $T[i+8]$ )
- 8: **return**  $T = \{T[0] \dots T[15]\}$

---

#### (알고리즘 1) Quantum circuit implementation of Mix

두 개의 워드 쌍  $T[i], T[i+8]$  ( $0 \leq i \leq 7$ ) 으로 한번의 Mix를 동작하며 총 8번의 Mix 함수가 사용된다. 알고리즘 1의 2, 5, 7 번째 라인 a, b, c 로테이션 함수에서의 순환 값은 각  $Step_j$  의  $j$ 값이 짹수/홀수에 따라 a, b 로테이션의 비트 순환량  $\alpha_j, \beta_j$  가 다르며 c 로테이션의 비트 순환량은 단계 변수의 값  $T[k]$  ( $k = 8, 9, 10, 11, 12$ )에 따라 바뀌기 때문에 각 경우를 주의하여 순환량을 설정해 두고 Swap 연산을 통해 비트를 로테이션 시켰다.

알고리즘 1의 3 라인은 CNOT-gate를 통해 SC(단계 상수)와  $T$ 를 XOR 연산하였다. Mix 함수를 통해 얻은  $T = \{T[0] \dots T[15]\}$  은 WordPerm 함수에서 Swap을 통해 치환된다. 비트 로테이션 및 WordPerm에서 사용된 Swap 연산은 서로 비트 위치만을 주는 연산이므로 별도의 게이트 비용이 들지 않는다. 제안하는 기법에서는 모든 라운드에 대한  $SC[i]$ 를 사전 연산하여 이에 따라 라운드마다  $T[i]$ 에 X 게이트를 수행한다. 때문에 큐빗이 사용되지 않았으며 기존의  $SC[i]$  값을 업데이트하기 위한 양자 게이트를 전혀 사용하지 않았다.

$N_s$  ( $N_s = 26$ )번의 Step을 마치고 얻은  $CV$ 를 이용하여 Final 함수에서 최종 해시 값을 얻을 수 있다. 양자 회로를 이용한 Final 함수 구현은 알고리즘 2와 같다.

---

#### Algorithm 2 : Quantum circuit implementation of Final

---

**Input:**  $CV[i], CV[i+8]$  ( $0 \leq i \leq 7$ )  
**Output:**  $h = \{ h[0], \dots, h[256] \}$

- 1: **for**  $k = 0$  to 7
- 2:     **for**  $i = 0$  to 31
- 3:         **CNOT** ( $CV_{K+8}[i], CV_K[i]$ )
- 4: **for**  $k = 0$  to 7

---

```

5:   for i = 0 to 7
6:     CNOT (CVK[7-i], hK[i])
7:   for i = 0 to 0
8:     for j = 0 to 31
9:       Swap (CVK[i], CVK[i+1])
10:    for i = 0 to 7
11:      CNOT (CVK[7-i], h[8*(k*4)+i])
12:    for i = 0 to 0
13:      for j = 0 to 31
14:        Swap (CVK[i], CVK[i+1])
15:    for i = 0 to 7
16:      CNOT (CVK[7-i], h[16*(k*4)+i])
17:    for i = 0 to 0
18:      for j = 0 to 31
19:        Swap (CVK[i], CVK[i+1])
20:    for i = 0 to 7
21:      CNOT (CVK[7-i], h[24*(k*4)+i])

```

(알고리즘 2) Quantum circuit implementation of  
Final

알고리즘 2의 2번째 줄은  $CV_{K+8}$ ,  $CV_K$  ( $0 \leq k \leq 7$ ) 쌍으로 짹지어 총 7번의 CNOT-gate를 수행하여 XOR 연산한 값  $CV_0, \dots, CV_7$ 을 얻는다.  $CV_K$  ( $0 \leq k \leq 7$ )를 1비트씩 로테이션하며  $CV_K[7], \dots, CV_K[0]$  순서로 8비트씩을 0으로 세팅된 256비트 큐빗에 순서대로 저장한다. 이때 CNOT-gate를 이용한 XOR 연산을 통해 대입 연산을 대신한다. Final 함수를 통한 256개의 비트  $h$ 는 곧 해시 값이 된다.

#### IV. 구현 결과

LSH-256/256을 양자 회로로 구현하는데 필요한 자원은 표 1과 같다.

	Qubits	Toffoli gates	CNOT gates	X gates
LSH-256/256	1,918	63,488	145,408	3,495

(표 1) Quantum resources for LSH-256/256

LSH는 메시지를 입력 받으면 블록 메시지 길이의 배수만큼 패딩한다. 본 논문에서는 블록 메시지 길이의 1배수인 1024 비트 블록을 해시할 때 사용되는 양자 자원을 평가하였다.

$M$  값을 업데이트 하는데 있어 기존  $M$ 에 할당 된 큐빗을 재활용하였으며 SC 값을 XOR 연산하기 위해 라운드마다의 SC 값을 모두 사전에 연산하고 이에 따라 X 게이트를 T에 수행하였다. 또한 Mix, Final 등의 연산들을 최적화함으로써 LSH의 양자회로에 사용되는 큐빗 수와

양자 게이트 수를 최소화 하였다.

#### V. 결론

본 논문에서는 국산 해시 함수 LSH-256/256을 양자 회로로 최적화 구현하여, 최종적으로 필요한 자원을 추정하였다. 해시 함수를 양자 회로로 구현하여 그루버 알고리즘 오라클에 적용하기 위한 자원을 확인함으로써 양자 컴퓨터의 공격에 대한 안전성의 지표로 활용할 수 있다. 제시한 LSH 해시 함수의 양자 회로 구현을 잡재적으로 그루버 알고리즘에 활용할 수 있을 것이라 기대된다.

#### VI. Acknowledgment

이 논문은 부분적으로 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (<Q|Crypton>, No.2019-0-00033, 미래컴퓨팅 환경에 대비한 계산 복잡도 기반 암호 안전성 검증 기술개발) 그리고 이 성과는 부분적으로 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478).

#### [참고문헌]

- [1] Grover, L.K, “A fast quantum mechanical algorithm for database search.” Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, pp. 212–219, - 1996.
- [2] Grassl. M, Langenberg. B, Roetteler. M, Steinwandt. R, “Applying Grover’s algorithm to AES: quantum resource estimates.” Post-Quantum Cryptography.” Springer, pp. 29 – 43, 2016.
- [3] Langenberg. B, Pham. H, Steinwandt. R, “Reducing the cost of implementing AES as a quantum circuit.” Technical report, Cryptology ePrint Archive, Report 2019/854, 2019.

- [4] Jaques. S, Naehrig. M, Roetteler. M, Virdia. F, "Implementing Grover oracles for quantum key search on AES and LowMC." Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, pp. 280 - 310, 2020.
- [5] Anand, R.; Maitra, A.; Mukhopadhyay, S. "Grover on SIMON." 2020.
- [6] K.B. Jang, S.J. Choi, H.D. Kwon, H.J. Seo, "Grover on SPECK : Quantum Resource Estimates." ePrint Archive, Report 2020/640, 2020.
- [7] K.B. Jang, S.J. Choi, H.D. Kwon, H.J. Seo, H.J. Kim, J.H. Park, H.J. Seo, "Grover on Korean Block Ciphers", Appl. Sci. 10, 6407. 2020.
- [8] Korea Cryptography Forum, "High Speed Hash Function LSH", 2015 LSH Implementation Contest, Korea Cryptography Forum, 2015
- [9] Cuccaro, Steven & Draper, Thomas & Kutin, Samuel & Moulton, David. (2004). A new quantum ripple-carry addition circuit.
- [10] Kim DC., Hong D., Lee JK., Kim WH., Kwon D. (2015) LSH: A New Fast Secure Hash Function Family. In: Lee J., Kim J. (eds) Information Security and Cryptology – ICISC 2014. ICISC 2014.

# 확장 이진 유한체 제곱근 연산의 최적 구현 기법에 관한 연구

박민진 오진석 인재휘 전창열 김동찬

국민대학교

A Study on Optimal Implementation of Square Root in Extended Binary Field

MinJin Park, JinSeok Oh, JaeHui In, ChangYeol Jeon, Dong-Chan Kim

Kookmin University

## 요약

확장 이진 유한체의 제곱근 연산은 사전계산 테이블 참조 방식의 지수승 연산으로 계산할 수 있다. Takuya Sumi 등은 이진 유한체의 성질을 이용하여 보다 효율적으로 제곱근을 계산하는 방식을 사용하였다. 본 논문에서는 두 제곱근 연산 방식인 지수승 기반 방식과 Takuya Sumi 등의 방식을 소개하고, 두 연산을 C언어로 구현하여 시간과 메모리 관점에서 성능을 비교 분석한다.

## I. 서론

확장 이진 유한체의 원소의 제곱근은 항상 존재하고, 사전계산 테이블 참조 방식의 지수승 연산으로 계산할 수 있다. Takuya Sumi 등은 이진 유한체의 성질을 이용하여 보다 효율적으로 제곱근을 계산하였다[2]. 해당하는 두 가지 제곱근 연산에 대해 [1]에서는 FLINT로 구현한 후 연산 시간과 메모리를 비교하였다. 구현 시 FLINT를 이용했기 때문에 최적화 과정에서 한계를 보였다.

본 논문에서는 [1]에서 사용한 두 제곱근 연산인 지수승 기반 제곱근 연산과 Takuya Sumi 등의 제곱근 연산을 소개한다. 그리고 두 연산을 C언어로 구현 시 진행한 최적 구현 기법에 대해 설명한다. 이후 구현한 연산을 시간과 메모리 관점에서 [1]과 비교한다. 비교는 Classic McEliece에서 제안한 4개의 파라미터에 대해 이루어진다. C언어로 구현 시 FLINT를 이용한 [1]에 비해 지수승 기반 제곱근 연산 기준 최대 122배, Takuya Sumi 등의 제곱근 연산 기준 최대 33배 빠르게 동작하였다.

본 논문의 구성은 다음과 같다. II절에서는 기호를 정의한다. III절에서는 [1]에서 사용한 두 알고리듬에 대해 소개한다. IV절에서는 두 알고리듬에 적용한 최적 구현 기법에 대해 설명한다. V절에서는 C언어로 구현한 두 연산의 시간과 메모리를 측정하고 [1]의 결과와 비교한다.

## II. 기호

본 논문에서 사용하는 기호는 다음과 같다.

$-F_{2^m}$	$2^m$ 개의 원소를 가진 유한체
$-f(s) (= \sum_{i=0}^m z_i s^i)$	유한체 $F_{2^m}$ 의 생성 기약 다항식 ( $z_i \in F_2$ )
$-F_{2^m}[X]$	$F_{2^m}$ 으로 정의한 다항식 환
$-g(X) \in F_{2^m}[X]$	$t$ 차 기약다항식
$-F_{2^m}[X]/(g(X))$	$g(X)$ 로 정의한 $2^{mt}$ 개의 원소를 가진 유한체
$-\sqrt{\alpha}$	$\alpha (\in F_{2^m})$ 의 제곱근
$-\sqrt{A(X)}$	$A(X) (\in F_{2^m}[X]/(g(X)))$ 의 제곱근

## III. 제곱근 연산

$F_{2^m}[X]/(g(X))$ 의 원소  $A(X) = \sum_{i=0}^{t-1} a_i X^i$ 는 다음 두 연산으로 제곱근을 계산할 수 있다.

$$\sqrt{A(X)} = A(X)^{2^{mt-1}} \quad (\text{식 } 1)$$

$$= \sum_{i=0}^{t-1} \sqrt{a_i X^i}. \quad (\text{식 } 2)$$

(식 1) 이용 시 다항식  $A(X)$ 에 대한 제곱 모듈러 연산을  $mt - 1$ 회 하는 것으로 제곱근을 구할 수 있다. 이때 사전계산 테이블  $T, S$ 를 이용한다.  $T$ 에는  $F_{2^m}$ 의 모든 원소를 제곱한 결과를 저장하고,  $S$ 에는  $\lfloor (t+1)/2 \rfloor \leq i \leq t-1$ 에 대해  $X^{2^i} \bmod g(X)$ 를 저장한다. 이 방식의 수행과정은 알고리듬 1과 같다.

#### 알고리듬 1: (식 1)을 이용한 제곱근 연산

입력:  $A(X) = \sum_{i=0}^{t-1} a_i X^i \in F_{2^m}[X]/(g(X))$ ,  
사전계산 테이블  $T, S$   
출력:  $\sqrt{A(X)}$

1.  $k \leftarrow \lfloor (t-1)/2 \rfloor$
2. **for**  $j = mt-2$  **downto** 0 **do**
3.  $\sum_{i=0}^{t-1} a_i X^i \leftarrow \sum_{i=0}^k T[a_i] X^{2^i} + \sum_{i=k+1}^{t-1} T[a_i] S[i]$
4. **return**  $\sum_{i=0}^{t-1} a_i X^i$

(식 2)는 Takuya Sumi 등이 이용한 제곱근 연산 방법에서 이용한다. 이때 사전계산 테이블  $\tilde{T}, \tilde{S}$ 를 사용한다.  $\tilde{T}$ 에는  $F_{2^m}$ 의 모든 원소의 제곱근을 저장하고,  $\tilde{S}$ 에는  $0 \leq i \leq \lfloor t/2 \rfloor - 1$ 에 대해  $X^i \sqrt{X} \bmod g(X)$ 를 저장한다. 이 방식의 수행과정은 알고리듬 2와 같다.

#### 알고리듬 2: (식 2)를 이용한 제곱근 연산

입력:  $A(X) = \sum_{i=0}^{t-1} a_i X^i \in F_{2^m}[X]/(g(X))$ ,  
사전계산 테이블  $\tilde{T}, \tilde{S}$   
출력:  $\sqrt{A(X)}$

1.  $A'(X) \leftarrow \sum_{i=0}^{\lfloor (t-1)/2 \rfloor} \tilde{T}[a_{2i}] X^i$
2.  $A''(X) \leftarrow \sum_{i=0}^{\lfloor t/2 \rfloor - 1} \tilde{T}[a_{2i+1}] \tilde{S}[i]$
3. **return**  $A'(X) + A''(X)$

## IV. 구현

알고리듬 구현을 위해 다항식을 구조체로 저장한다. 구조체는 다항식의 계수와 최고차수로 구성된다. 이때 다항식의 최고차수는 int형 변수로 설정하고, 다항식의 계수는 int형 배열을 이용하여  $F_{2^m}$ 의 원소 하나를 4바이트에 저장한다.

두 알고리듬은 사전계산을 통한 테이블 참조 이외에 유한체 곱셈, 다항식 덧셈으로 구성된다.

유한체 곱셈에서는 사전계산 테이블  $R$ 을 사용한다.  $m \leq i \leq 2m-1$ 에 대해  $s^i \bmod f(s)$ 를 계산하여  $R$ 에 저장한다. 이후 과정에서 유한체 곱셈 시 발생하는 모듈러 연산은 테이블 참조로 처리하였다.

다항식 덧셈은 동일한 차수에 대해  $F_{2^m}$ 상에서의 덧셈이다. 해당 연산은 입력받은 다항식의 계수에 대해 XOR 연산을 하는 것으로 처리하였다.

## V. 측정 결과

실험 환경은 다음과 같다.

- 하드웨어 환경	1.4GHz 쿼드 코어 Intel Core i5, 8GB RAM, macOS Big Sur 버전 11.0.1
- 컴파일러	Apple clang version 12.0.0 (clang-1200.0.32.27) 컴파일 옵션 -O2

Classic McEliece에서 제안한 파라미터에 대해 사용한 테이블의 메모리, 연산 수행 횟수(테이블 참조 횟수, XOR 연산 횟수), 연산 시간을 측정하였다.

사용한 사전계산 테이블은  $T, \tilde{T}, S, \tilde{S}, R$ 이며 구현에서 사용한 총 메모리는 다음과 같다.  $T, \tilde{T}$  테이블은 각각  $F_{2^m}$ 의 원소를  $2^m$  개 저장한다.  $F_{2^m}$ 의 각 원소는 4바이트의 공간을 차지하므로 각 테이블은  $2^{m+2}$  바이트의 메모리를 사용한다.  $S, \tilde{S}$  테이블은 각각  $F_{2^m}[X]/(g(X))$ 의 원소를  $\lfloor t/2 \rfloor$  개 저장한다. 각 원소는 곱셈을 고려하여 계수를 최대 2t개로 설정한다. 따라서 테이블 당  $2t \cdot \lfloor t/2 \rfloor \cdot 2^2$  바이트의 메모리를

사용한다.  $R$ 테이블은  $F_{2^m}$ 의 원소를  $m$ 개 저장한다. 그러므로  $2^2m$  바이트의 메모리를 사용한다. 최종적으로 각 알고리듬에서 사용하는 메모리는 다음과 같다.

$$(2^m + 2t \lfloor t/2 \rfloor + m) \cdot 2^2$$

연산 수행 횟수는  $A(X)$ 에 따라 달라진다. 그러므로 임의의  $A(X)$ 를 1000회 생성하여 측정한 후 평균값으로 나타내었다. 메모리와 연산 수행 횟수는 [표 1]과 같다. 사용한 메모리의 단위는 바이트(byte)이다. 또한 XOR 횟수는 4바이트에 대한 연산 횟수이다.

[표 1] 알고리듬 연산 수행 횟수 및 메모리

파라미터 ( $m, t$ )	Mceliece348864 (12, 64)	Mceliece460896 (13, 96)		
알고리듬	1	2	1	2
참조횟수	1,330,541	12,376	4,716,419	30,051
XOR횟수	25,422,382	88,217	98,334,238	214,094
메모리		32,816		69,684
파라미터 ( $m, t$ )	Mceliece6960119 (13, 119)	Mceliece6688128 (13, 128)		
알고리듬	1	2	1	2
참조횟수	9,260,570	45,817	17,623,528	53,504
XOR횟수	186,319,780	329,093	270,897,271	380,841
메모리		88,988		98,356

FLINT로 구현한 [1]과 C언어로 구현한 알고리듬의 연산 시간은 [표 2]와 같다. 1000회 동작 시 걸린 시간의 평균으로 나타내었고, 사용한 시간 단위는 밀리초(ms)이다.

[표 2] 알고리듬 연산 시간 측정결과 (단위: ms)

파라미터 ( $m, t$ )	알고리듬	FLINT [1]	C(본 논문)
Mceliece348864 (12, 64)	1	2,984.971	24.294
	2	2.878	0.086
Mceliece460896 (13, 96)	1	6,320.881	100.715
	2	4.182	0.218
Mceliece6960119 (13, 119)	1	11,400.386	173.928
	2	6.472	0.284
Mceliece6688128 (13, 128)	1	14,497.782	292.432
	2	7.904	0.393

연산 시간에 대한 신뢰성을 확인하기 위해 C 언어로 구현한 두 알고리듬의 수행 횟수와 연

산 시간에 대한 비율을 확인하였다. 이는 [표 3]과 같다. 비는 수행 횟수 비율을 연산 시간 비율로 나눈 값이고, 각각의 비율은 알고리듬 1에 대한 측정값을 알고리듬 2에 대한 측정값으로 나눈 값이다.

[표 3] 파라미터별 수행 횟수 및 연산 시간 비율

파라미터 ( $m, t$ )	수행 횟수	연산 시간	비
Mceliece348864 (12, 64)	265	282	0.93
Mceliece460896 (13, 96)	422	461	0.91
Mceliece6960119 (13, 119)	521	612	0.85
Mceliece6688128 (13, 128)	664	744	0.89

## VI. 결론

본 논문에서는 두 알고리듬을 C언어로 구현한 결과와 FLINT로 구현한 [1]의 결과를 비교하였다. 결과적으로 연산 시간이 알고리듬 1 기준 최대 122배, 알고리듬 2 기준 최대 33배 줄었다. 또한 파라미터 별로 연산 수행 횟수 비율과 연산 시간 비율을 확인한 결과 두 비율이 유사함을 확인하였다.

추후에는 SIMD를 이용한 병렬연산으로 제곱근 연산의 고속 구현 기법에 대해 연구할 예정이다.

## 【참고문헌】

- [1] 전창열, 박민진, 오진석, 인재휘 and 김동찬. “확장 이진 유한체 제곱근 연산의 효율적 구현에 관한 연구.” 한국통신학회 동계종합 학술발표회. 2021.
- [2] Sumi Takuya, Morozov Kirill, and Takagi Tsuyoshi. Efficient implementation of the McEliece cryptosystem. In computer security symposium 2011, volume 2011, pages 582–586, oct 2011.

# Blockchain-based Decentralized Incentive Mechanism for Trusted Data Management on Internet of Vehicle

Muhammad Firdaus\* and Kyung-Hyune Rhee\*\*

\*Department of Artificial Intelligence Convergence  
Pukyong National University

\*\*Department of IT Convergence and Application Engineering  
Pukyong National University

e-mail: mfirdaus@pukyong.ac.kr, khrhee@pknu.ac.kr

## Abstract

The Internet of Vehicles (IoV), which combines smart vehicles and the internet, has emerged to enhance traffic safety and efficiency. Moreover, it also is promoted to advance the future of intelligent transportation system (ITS) applications. However, The existing IoV is still facing challenges in providing a trusted management system and information security protection. Hence the participant's vehicle may be unwilling to share their data since the transaction system still relies on a centralized server approach with the potential risk of data leakage and privacy security. Also, vehicles have difficulty evaluating the credibility of the messages they received because of untrusted environments. To address these challenges, we propose consortium blockchain and smart contracts to accomplish a decentralized trusted data sharing management system in IoV. This system allows vehicles to validate the credibility of messages from their neighboring by generating a reputation rating. Moreover, the incentive mechanism is utilized to trigger the vehicles to store and share their data honestly; thus, they will obtain certain rewards from the system.

## 1. Introduction

The IoV allows vehicles to share road-related information messages with their neighbors, e.g., road conditions, traffic congestions, accident information, and safety warnings. Consequently, vehicles can be more aware of traffic situations, as well as contribute to improving the system transportation safety and efficiency [1]. However, the conventional architecture of IoV with a centralized approach has crucial challenges related to user's data security and privacy. In this sense, the potential exposure of user information with Single Point of Failure (SPoF) challenges still will reasonably occur since the IoV framework's data is centralized on a central server. Hence, the IoV network participants might be hesitant in the data sharing process that contains private information such as customer identities, vehicle numbers, and

driving preference. Moreover, the risk of selfish behaviors might diminish participants' enthusiasm to cooperate with each other in the system. This problem becomes more serious when there exists a malicious vehicle in the network. The various adversarial actions may threaten the privacy security to gather the user's private information for personal benefit as well as endanger traffic safety and efficiency by giving the incorrect report to the system.

In order to address these issues, in this paper, we utilize a consortium blockchain and leverage the smart contracts to develop a decentralized trusted data sharing management system in IoV. Moreover, we propose an appropriate incentive mechanism based on the vehicle's contribution by leveraging smart contracts' self-execution nature. This scheme aims to motivate vehicles to

participate positively in maintaining trusted data sharing activities and ensuring the system's security and sustainability.

The rest of this paper is organized as follows: Section 2 describes problem definition based on traditional vehicular ad-hoc networks and centralized incentive mechanism. Then, we present the design architecture of the IoV-blockchain, including its detailed procedures in Section 3. We demonstrate the proposed design by analyzing security and evaluating performance in Section 4. Finally, Section 5 concludes this paper.

## 2. Background

### 2.1 Data Management in Vehicular Networks

In the vehicular networks (VNs) environment, the primary entities in the data sharing process are vehicles and roadside units (RSUs), which form two types of communication, namely vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). Vehicles interact with other neighboring vehicles by using onboard units (OBUs) equipped with several sensing devices with simple computation and communication capabilities. OBUs are also used to automatically recognize traffic-related information and facilitate the vehicles to send notification messages to others using V2V communication standards to improve traffic safety and efficiency. On the other hand, V2I facilitates a single or multi-hop communication between vehicle and RSU, supported by dedicated short-range communication (DSRC) standards [2]. RSUs, as the roadside infrastructure, provide wireless communications along the road to vehicles. Thus, RSUs are prepared to aggregate the traffic data in a particular coverage area in the VN system.

### 2.2 Blockchain-based Incentive Mechanism

Previous work considers that vehicle may share their data voluntarily [3]. Unfortunately, it might cause the low participation of vehicles in the data sharing process and then affects the system's reliability in the future. Moreover, due

to self-interest characteristics, vehicles may also be unenthusiastic to share the data because they do not obtain particular benefits or compensation from the system. Therefore, the incentive mechanism is aimed to motivate honest vehicle (as data owner) to give a relevant contribution with long term participation for system reliability and sustainability. Incentive mechanism provides rewards to the vehicles that contribute to the data sharing process to form the trusted data management system in IoV.

Blockchains have recently received increasing attention as a promising technology for providing distributed and secure solutions. They are open databases that guarantee data security by enabling anonymous and trustworthy transactions on an immutably distributed ledger without the help of a central intermediary [4]. Each transaction is recorded with a timestamp to be validated by the consensus mechanism before it is stored on a blockchain network. Moreover, by leveraging its smart contracts, blockchains have been utilized to form fairness incentives by providing a decentralized approach to overcome the risks of any single point of failure on a centralized incentive approach. Therefore, a decentralized incentive might effectively encourage user participation and contributions, creating a secure framework for users to share their data to improve system reliability and sustainability.

## 3. The Framework of IoV-Blockchain

This section explains our proposed model, blockchain-based secure and trusted data sharing management in IoV networks. Inspired by [5], our architecture model relies on consortium blockchain, as shown in Figure 1. In our scenario, vehicles represent the user network that communicates with other vehicles and RSUs, using V2V and V2I communication to improve traffic safety and efficiency. The sending vehicle ( $V_m$ ) equipped with OBUs and their sensing devices automatically collect road-related messages ( $M_i$ ) to the system, while the neighboring vehicle

$(V_n)$  evaluate the message  $M_i$  by uploading the message credibility rating ( $R_i$ ) to nearby RSU.

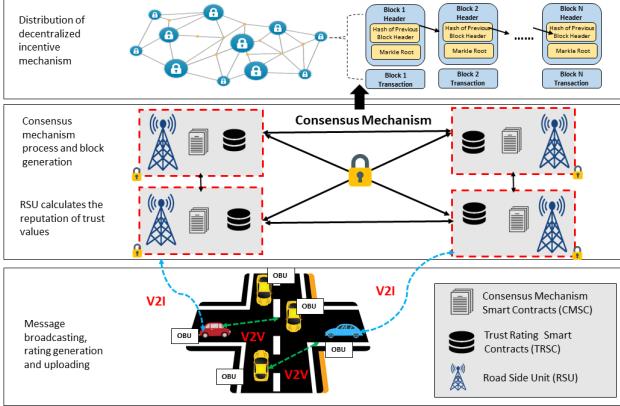


Figure 1. The framework of IoV-blockchain

Then, RSU as the roadside infrastructure and traffic handler, aggregate  $M_i$  based on  $R_i$  using the trust rating smart contract (TRSC) to generate the trust value rating of  $M_i$  ( $Tr_M$ ).

After that,  $Tr_M$  will be validated in the consensus mechanism smart contract (CMSC) using consensus mechanism. Here, only authorized RSUs are eligible to be the nodes participants in the consensus mechanism with more extensive storage and computation capability compared to the OBUS. Once the consensus process is finished, a new block is uploaded to the blockchain network. Thus, the distributed RSUs automatically obtain the log and authentication of block.  $V_m$  and  $V_n$ , which correctly provide the road-related message and assess the message credibility, respectively, will obtain the proportional incentive.

#### 4. Simulation and Results

Using the OSMWebWizard package provided in the simulation of urban mobility (SUMO), we modeled a highway traffic scenario to prototype and evaluate IoV networks' efficiency. Here, a discrete-event network simulator is used to verify the result, analyzing a trace file for vehicle mobility and message credibility. We use an optimized link-state routing protocol (OLSR) as one of the protocol standards in the wireless

access for the vehicular environments (WAVE). In this scenario, there are 26 vehicles in the user network layer.

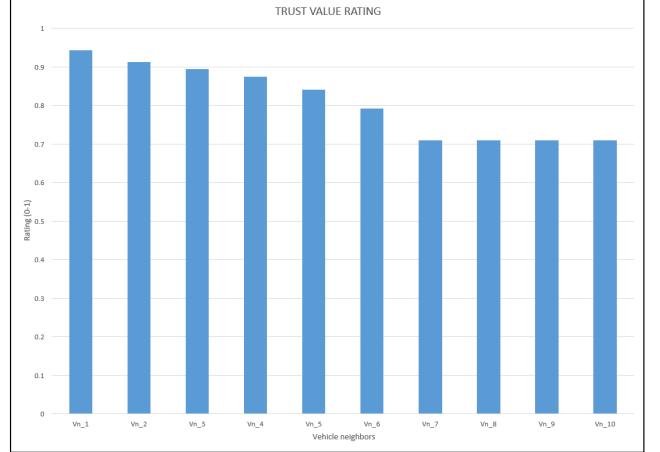


Figure 2. Trust value rating aggregation

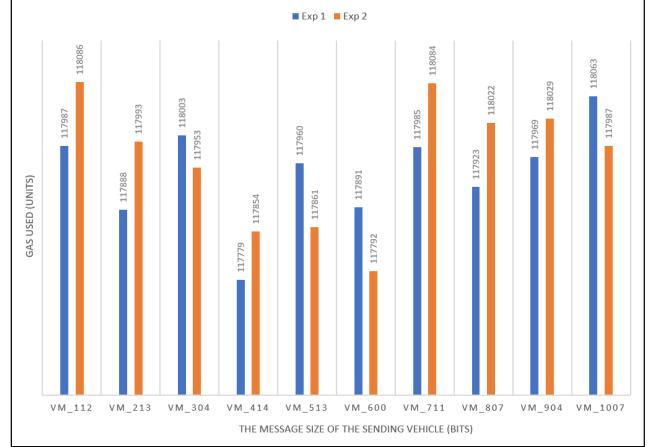


Figure 3. The information on gas usage by RSU

However, only 10 vehicles are proposed to be neighboring vehicles  $V_n$  and placed 50 meters apart from the occurred event. Once  $V_m$  broadcasts the road-related message  $M_i$  via V2V communication,  $V_n$  is allowed to evaluate the message credibility by generating the message rating to RSU. Figure 2 shows the trust value rating aggregation based on message credibility assessment over 10  $V_n$  on various separations. The  $V_{n1}$  obtains the highest value of the message's credibility due to the nearest distance with  $M_i$ . Contrary, the  $V_{n10}$  obtains the lowest value of the message's credibility due to the farthest distance with  $M_i$ .

To support an adequate incentive for the information provider  $V_m$ , we implement Ethereum smart contracts as a decentralized and tamper-proof incentive mechanism. We utilize a smart contract feature in the Ethereum platform through Ganache Truffle (v.2.4.0) graphical user interface (GUI). Figure 3 illustrates the information on gas usage by RSU in distributing Ether for the contributed vehicles. Even though Figure 3 shows the amount of gas usage is significant from one another, the amount of gas difference between transactions is relatively the same by using units notion. Smart contracts store the address information of the requester and provider, while Ethereum network only stores arbitrary values of related information.

## 5. Conclusions

We have introduced a consortium blockchain and smart contracts to achieve a decentralized trust data management system in IoV. In this paper, smart contracts are exploited to accomplish efficiency, reliability, and secure data storage and sharing. Here, two smart contracts, TRSC and CMSC, are deployed on distributed RSUs to gather the trust value rating and conduct the consensus mechanism. Moreover, this framework permits vehicles to validate the credibility of messages from their neighboring vehicles by generating a reputation rating. Additionally, we utilize an incentive mechanism to motivate and propel the vehicle to contribute and sincerely share their data to obtain certain rewards from the system.

## Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2018R1D1A1B07048944) and partially was supported by the Republic of Korea's MSIT (Ministry of Science and ICT), under the High-Potential Individuals Global Training Program (2020-0-01596) supervised by the IITP (Institute of Information and Communications Technology Planning

& Evaluation).

## [References]

- [1] Yang, Z., Yang, K., Lei, L., Zheng, K., & Leung, V. C. (2018). Blockchain-based decentralized trust management in vehicular networks. *IEEE Internet of Things Journal*, 6(2), 1495–1505.
- [2] Kenney, J. B. (2011). Dedicated short-range communication standards in the United States. *Proceedings of the IEEE*, 99(7), 1162–1182.
- [3] Xu, C., Wang, Y., Zhou, Z., Gu, B., Frascolla, V., & Mumtaz, S. (2018, December). A low-latency and massive-connectivity vehicular fog computing framework for 5G. In *2018 IEEE Globecom Workshops (GC Wkshps)* (pp. 1–6). IEEE.
- [4] Firdaus, M., & Rhee, K. H. (2020). Empowering Blockchain For Secure Data Storing in Industrial IoT. In *Proceedings of the Korea Information Processing Society Conference* (p. 231–234). Korea Information Processing Society.
- [5] Firdaus, M., & Rhee, K. H. (2021). On Blockchain-Enhanced Secure Data Storage and Sharing in Vehicular Edge Computing Networks. *Applied Sciences*, 11(1), 414.

# Security for Blockchain-based DID Key Generation

Seong-Kyu Kim\*

\*Assistant Professor (Tenure Track) of Department of Information Security, Joongbu University, Gyeonggi-do Goyang-Si 10279, Republic of Korea.(e-mail : [skkim@joongbu.ac.kr](mailto:skkim@joongbu.ac.kr) or [guitara77@gmail.com](mailto:guitara77@gmail.com) )

Jun-Ho Huh\*\*

\*\* Assistant Professor (Tenure Track) of Department of Data Informatics, (National) Korea Maritime and Ocean University, 727, Taejong-ro, Yeongdo-gu, Busan, Republic of Korea (e-mail : [72networks@pukyong.ac.kr](mailto:72networks@pukyong.ac.kr) or [72networks@kmou.ac.kr](mailto:72networks@kmou.ac.kr) )

## Abstract

This study has recently been actively researched on blockchain-based identity authentication systems. In addition, DID, which refers to a digital identity verification system that allows users to select only the information needed for proof purposes and provide it to verification institutions to strengthen privacy protection, has been studied a lot. In addition, as the non-face-to-face economy is expected to accelerate after Corona19, the proliferation of DID services, an online identity verification technology, is in full swing, and in this study, we study major architectures for DID key generation and mobile-based privacy certification.

## I. Introduction

Blockchain-based distributed IDs (DIDs) will enable digital financial transactions anytime, anywhere by embedding digital IDs and identification cards in smartphones. In addition to resident registration cards and passports and driver's licenses, these digital identification cards will be developed into self-sustaining identification using biometric distributed IDs. Domestic and foreign mobile carriers also use biometric information such as fingerprint authentication and face recognition of smartphones [1].

It uses an app for authentication and financial account number by biometric authentication ID(Identification) and electronic signature. In this paper, we would like to propose a plan to utilize a blockchain DID-based biometric key generation device.

Specifically, we design DID(Decentralized Identifier) generation and authentication modules for the storage of distributed IDs. As such a user-certified node, the USB(Universal Serial Bus) module intends to propose a virtual currency recommendation in the key-store through online real-time authentication by DID-based FIDO(Fast IDentity Online). In addition, we propose a digital identity verification system that allows users to manage and control their own identity information online, just as they manage identity verification in real life as a distributed identity management system [2].

## II. Related Research

### 2.1 Blockchain Security

Blockchain technology includes concepts

beyond distributed ledger technology, and blockchain definitions can be considered as 'global trust computers'. Figure 1 summarizes five ways to define blockchain, and the most appropriate definitions can be seen as the third "smart contract execution platform" and the fourth "global trust computer." Here, smart contracts can be considered software running on blockchain computers. In addition, blockchain/blockchain-based security technologies are classified in various ways depending on the source core technologies, platform technologies [3], and security service technologies. Along with the development of blockchain technology, it is a technology field that is gradually expanding its application to all industrial areas.

In existing centralized transaction systems, each transaction must be validated through a trusted central server (e.g., a central bank), resulting in a cost and performance bottleneck on the central server. Unlike centralized systems, blockchain does not require a third trust organization. In blockchain, consensus techniques are used to maintain consistency of data in distributed networks.

## 2.2. Characteristics of blockchain (invariant)

Transactions on the chain can be quickly validated and invalid transactions are not accepted by nodes in the blockchain [4]. Once a transaction is recorded in the blockchain, it is almost impossible to delete or revert the content. Blocks on nodes containing invalid transactions can be found immediately.

## 2.3. Characteristics of blockchain (anonymous)

Each user interacts with the blockchain with their similar anonymous address. This address does not include user entity identification information [5]. Thus, user

anonymity can be preserved. However, transaction anonymity cannot be guaranteed because blockchain discloses all transaction content [6].

## 2.4. Characteristics of Blockchain (Traceability)

All transactions are referenced to previous unused transactions assigned to nodes on the block [7]. When the current transaction is written to the blockchain, the state of the referenced unused transaction transitions from unused to used. Therefore, all transactions can be easily identified and tracked.

# III. Research method

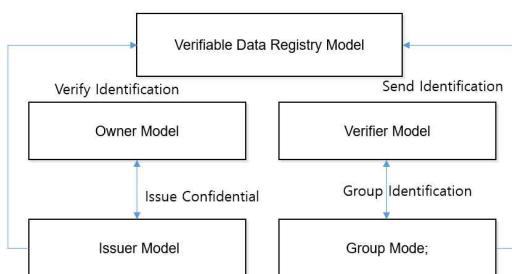
## 3.1. Study Design

With the advent of distributed ledger technology, a completely new identity management system is being developed. Entities that exist in distributed identity management systems are free to use shared trust roots. Globally distributed ledger (or distributed P2P(Peer to Peer) networks that provide similar functions) provides a means of managing IDs without using centralized privileges. The combination of distributed ledger technology and distributed identifiers allows all entities to be distributed to create and manage unique identifiers in independent and reliable roots. In addition, entities in distributed ID management systems are identified as distributed identifiers and can be authenticated through proofs. Distributed identifiers are a new type of identifier for verifiable autosovereignty IDs, and provide a standard way to create globally unique and cryptographically provable permanent identifiers. Distributed identifiers are managed under the control of DID subjects, independent of centralized based identity providers. The distributed identifier refers to

DID Documents, which briefly describes how to use them, and each DID Document provides at least proof points, verification methods, and service endpoints information. Include, DID Document describing sample information and distributed identifiers that represent distributed identifiers. The proof purpose, combined with the verification method, provides a mechanism for proving things. DID Document can specifically specify that proofs generated for authentication purposes can be validated using specific verification methods, such as public keys or anonymous biometric protocols. Service endpoints enable reliable interaction with DID controllers.

### 3.2. Research Architecture

This research architecture works with the Verifiable Credentials Data Model 1.0 standard in W3C's Verifiable Climes Working Group, which provides cryptographically secure, privacy-preserving, and machine-verifiable ways to represent a wide variety of Credentials on the web. This allows verifiable Credentials to represent all the same information as physical Credentials, and by adding techniques such as digital signatures, information can be tampered with, making it more reliable than physical Credentials. We also describe the role of ecosystems and key actors for verifiable CREDITALS and their relationships [Fig. 1]



[Fig. 1] Architecture for verifiable DIDs

#### 3.2.1 Issuer Model

Produces verifiable credentials for a particular object and verifies them by serving to deliver them to Holder.

#### 3.2.2 Owner Model

Owns one or more verifiable certificates and is responsible for generating and delivering them in the form of Presentation when submitting them to Verifier.

#### 3.2.3 Verifier Model

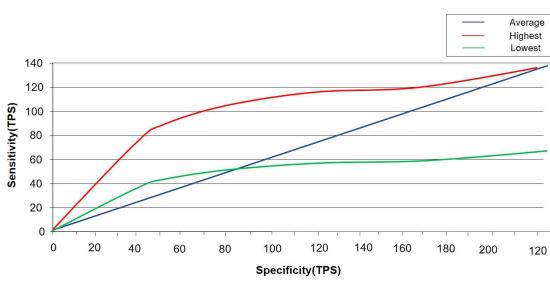
Holder has a role in verifying that it has adequate verifiable credentials.

#### 3.2.4 Verifiable Data Registry Model

Distinguish distributed DB(Database), government ID DB, distributed ledger, etc. from the role of mediating the generation and verification of identifiers, keys, and other relevant data.

### 3.3 Experimental Results

The demonstration objective of the experiments in the study, combined with the verification method, provides a mechanism for proving things. For example, DID Document can specifically specify that proofs generated for authentication purposes can be verified using certain verification methods, such as public keys or anonymous biometric protocols. Service endpoints enable reliable interaction with DID controllers. The DID performance test was also performed as shown in Figure 2. The average value was 80 for TPS(Transaction per Second), 120 for the best and 60 for the lowest.



[Fig. 2] DID Performance Specific Results

#### IV. Future research and conclusions

In this paper, we describe ID management techniques using blockchain technology and distributed identifiers and verifiable credentials for self-sovereign ID management. Blockchain-based ID management technology emphasizes self-sovereignty to manage and control one's own information. Distributed identifiers and verifiable credentials that are essential to constructing a self-sustaining ID system are now actively standardized through W3C, but have so far either been a group report or a draft statement.

It is a stage that is being announced in the form of a book. Currently, the study of blockchain-based ID management techniques is focused on aspects of sharing and utilizing information. Given its application in various fields in the future, research on analysis of constraints for information sharing between heterogeneous domains and update and deletion processing of shared information should also be carried out. It is also necessary to focus on future research on identity management and new forms of authentication technology for a wide variety of entities, such as the ID of groups such as corporations and organizations, and the ID of smart devices, beyond identity management technology for individuals.

In addition, standards that are required for key technologies on DID's technology roadmap in advance or that require simultaneous development need to be derived as a key standardization item in the standardization strategy map.

#### [Reference]

- [1] Manohar, Arthi, Jo Briggs, "Identity Management in the Age of Blockchain 3.0," HCI for Blockchain - CHI2018 workshop, 22nd, April 2018.
- [2] Rohan Pinto, "How Blockchain Can Solve Identity Management Problems", Forbes, July 2018.
- [3] Microsoft Blog, "Decentralized digital identities and blockchain: The future as we see it," February 2018.
- [4] Huh , J.-H.; Kim, S.-K. The blockchain consensus algorithm for viable management of new and renewable energies. Sustainability 2019, 11, 3184.
- [5] Paul Dunphy, and Fabian A. P. Petitcolas. "A First Look at Identity Management Schemes on the Blockchain," IEEE Security and Privacy Magazine special issue on "Blockchain Security and Privacy", August 2018.
- [6] Mingoo, Kang et al, "Blockchain system for authorized recommendation of cryptocurrency based on context-aware smart kisok ,," Korea patent(10-2019-0137060), 2019.10.31.
- [7] Nakhoon Choi, Heeyoul Kim, "A Blockchain-based User Authentication Model Using MetaMask," Vol. 20, No. 6, pp. 119-127, Dec.2019.

# 블록체인을 활용한 카페 이용시간 관리 시스템

박재훈\*\*, 강예준\*, 김원웅\*, 서화정\*\*†

\*\*한성대학교 IT융합공학부 (대학원생)

\*한성대학교 IT융합공학부 (학부생)

\*\*† 한성대학교 IT융합공학부 (교수)

Implementation of Management System of Customer Spending Time  
in Cafe with Blockchain

Jae-Hoon Park\*\*, Yea-Jun Kang\*, Won-Woong Kim\*, Hwa-Jeong Seo\*\*†

\*\*Division of IT convergence engineering, Hansung University  
(Graduate student)

\*Division of IT convergence engineering, Hansung University  
(Undergraduate student)

\*† Division of IT Convergence Engineering, Hansung University  
(Professor)

## 요약

COVID-19가 발생한 이후로 보건복지부로부터 여러 방역수칙이 제정되었으며, 최근 확진자의 계속 증가함에 따라 2.5단계가 시행되어 카페 매장 내에서의 취식은 불가하였다. 하지만 COVID-19가 다시금 누그러짐에 따라 카페 매장 내 취식이 1시간에 한해 허용되었으나, 이 수칙을 사람들이 잘 지키지 않을뿐더러 마땅히 잡아낼 수단도 존재하지 않는다. 본 논문에서는 블록체인을 활용한 카페 출입 관리 시스템을 통해 이러한 수칙 위반자들에 대한 검출을 가능하게 한다. 또한 향후에는 이 시스템에 백엔드 및 프론트엔드 구성을 통해 하나의 애플리케이션으로서의 개발도 해볼 것이다.

## I. 서론

최근 COVID-19가 누그러짐에 따라 수도권 지역은 사회적 거리두기 2.5단계임에도 카페 매장 내 취식이 가능하게 되었다. 그에 따라 카페 내 취식 가능 시간은 21시까지, 매장을 1시간만 이용할 수도록 권고되고 있다.

식당·카페(무인카페 포함)	<ul style="list-style-type: none"><li>• 2인 이상이 커피음료류, 디저트류만을 주문했을 경우 매장 내 머무르는 시간을 1시간으로 제한(강력 권고).</li><li>• 식당·카페 모두 21시~익일 05시 까지 포장·배달만 허용.</li><li>• 테이블 또는 좌석 한 칸을 띄워 매장 좌석의 50%만 활용하여 이를 준수하기가 어려울 경우 ① 고정폭 빙 칸 테이블 간 1m 거리 간 칸막이/가림막 설치 중 한 가지 준수(시설·면적 50㎡ 이상).</li><li>• 뷔페의 경우 공용침대접사수저 등 사용 전후 손소독제 또는 비닐장갑 사용 음식 담기 위한 대기 사용자 간 간격 유지.</li></ul>
----------------	---

(그림 1) 카페 이용제한 권고[1]

하지만 대부분의 사람들이 이 1시간 이용제한을 잘 지키지 않고 있으며, 이것에 대한 확인 조차 명확히 되지 않고 있다. 이러한 상황에 대해 명확한 해결 방안이 요구된다. 본 논문에서

는 블록체인을 활용하여 이러한 문제에 대해 1시간 이용제한 수칙을 잘 지킬 수 있도록 확인 및 통제가 가능한 시스템을 구현하였다.

## II. 관련 연구

### 2.1 블록체인

블록체인이란 기존의 거래 시스템 방식처럼 중앙 서버에서 장부를 관리하는 방식이 아닌, 암호화된 전자 장부를 거래에 참여한 모든 사람의 컴퓨터에 분산화 시켜 보안성을 높이는 방식이다[2]. 즉, 장부를 은행이 독점적으로 관리하는 방식에서 벗어나, 동일한 장부를 모든 사람이 가지고 있는 탈중앙화 형태의 방식이다. 거래에 참여한 모든 사람이 가지고 있는 장부를 계속해서 비교하기 때문에, 한두 명의 장부

를 조작하더라도 이는 조작된 장부라는 것을 바로 알 수 있다.

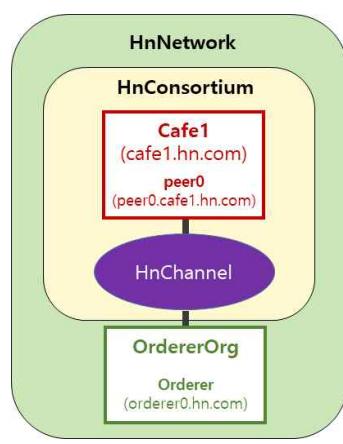
## 2.2 하이퍼레저 패브릭

하이퍼레저 패브릭은 리눅스 재단이 주도하는 오픈소스 블록체인 프로젝트이며 Private Blockchain의 형태를 띤다[3]. Private Blockchain이란 신뢰할 수 있는 기관인 CA(Certificate Authority)에 의해 허가를 받은 사용자만 네트워크에 참여할 수 있는 형태를 말한다. 이러한 과정에 의해 네트워크에 참여하는 노드들은 자연스럽게 신뢰를 가진 노드로 판별되기 때문에 복잡한 합의 알고리즘을 필요로 하지 않게 된다.

하이퍼레저 패브릭 내의 네트워크에 있는 참여자들은 체인코드(Chaincode)를 실행하여 원장에서 기존의 내용을 읽어오거나 새로운 내용을 업데이트하게 된다[4].

## III. 카페 이용 제한 시스템

본 시스템은 하이퍼레저 패브릭 2.3.0을 통해 구현되었으며, 체인코드에는 typescript를, 데이터베이스는 CouchDB를 사용하였다. 네트워크는 기본적으로 1개의 오더러와 1개의 조직으로 구성되어 있으며, 각 조직에는 1개의 피어가 구성되어 있다. 하나의 카페가 조직을 구성하며, 피어는 이용객의 수에 따라 조절될 수 있다.



(그림 2) 네트워크 구성

네트워크는 쉘 스크립트를 이용하여 실행시키게 되며, 실행 시 그림 2와 같은 형태의 네트워크가 생성 및 실행된다.

본 시스템에서는 카페 이용 고객의 출입 정

보를 기록하기 위해 IO 클래스를 생성하였다. IO 클래스의 필드는 표 1과 같다.

필드	형식	설명
key	string	키값
cafe	string	카페명
phone	string	핸드폰 번호
in_ms	number	입장시간
out_ms	number	퇴장시간
agreeToOfferInfo	boolean	정보제공동의여부

(표 1) IO 클래스 필드

체인코드는 입장할 때 이용되는 gettingIn, 퇴장할 때 이용되는 gettingOut, 1시간 이상 머무르는 고객을 확인하는 checkDisobeyed 메소드로 구성된다.

```
async gettingIn(context: Context, cafe: string, phone: string, agreeToOfferInfo: boolean) {
    if (!agreeToOfferInfo) {
        console.log('not available without agreeing to offer the information');
        return;
    }
    try {
        const io: IO = {
            key: this.generateKey(),
            cafe: cafe,
            phone: phone,
            in_ms: new Date().getTime(),
            out_ms: -1,
            agreeToOfferInfo: agreeToOfferInfo
        };
        await context.stub.putState(io.key, stateValue(io));
        console.log('new IO generated');
    } catch(err) {
        console.log(err);
    }
}
```

(그림 3) gettingIn 메소드

gettingIn 메소드는 카페명과 핸드폰 번호를 입력받아 새로운 IO 객체를 생성하여 DB에 추가하는 메소드이다. 객체를 생성할 때 in\_ms 값에 현재의 밀리세컨드 값을 입력함으로서 퇴장 시 검증에 이용한다. 또한 out\_ms 값은 -1으로 설정함으로서 향후 out\_ms가 -1인 데이터는 아직 나가지 않은 손님으로 구분할 수 있게끔 한다. 만약 고객이 정보제공동의를 하지 않았을 경우에는 입장이 거부된다.

```

async gettingOut(context: Context, phone: string) {
  try {
    const res = context.stub.getQueryResult(JSON.stringify({
      selector: {
        phone: phone
      },
      limit: 1
    }));
    if (!res) throw 'could not find the given phone number';
    await (const {key, value} of res) {
      const io = toItemValue();
      io.out_ms = new Date().getTime();

      if (io.out_ms - io.in_ms > 60 * 60 * 1000) {
        console.log(`the customer '${io.phone}' stayed in the cafe '${io.cafe}' over 1 hour`);
      }
      await context.stub.putState(io.key, stateValue(io));
    }
  } catch(err) {
    console.log(err);
  }
}

```

(그림 4) `gettingOut` 메소드

`gettingOut` 메소드는 사용자의 핸드폰 번호를 입력받아 그 사용자의 퇴장 처리를 진행하는 메소드이다. 메소드가 실행되면 DB에서 핸드폰 번호를 조회하여, `out_ms`에 현재의 밀리세컨드 값을 입력한 뒤 저장하게 된다. 만약 입장 시간과 퇴장 시간이 1시간 이상 차이날 경우, 알림을 띄우게 된다.

```

async checkDisobeyed(context: Context) {
  const ts = new Date().getTime();
  const res = context.stub.getQueryResult(JSON.stringify({
    selector: {
      $and: [
        {
          out_ms: { $lt: 0 },
        },
        {
          in_ms: { $lt: ts - (60 * 60 * 1000) }
        }
      ]
    }
  }));
  const disobeyeds: IO[] = [];
  for await (const {key, value} of res) {
    disobeyeds.push(toItem(value));
  }
  console.log(`${disobeyeds.length} customer(s) existing`);
  return disobeyeds;
}

```

(그림 5) `checkDisobeyed` 메소드

`checkDisobeyed` 메소드는 DB로부터 `out_ms` 값이 0 미만인, 즉 아직 퇴장하지 않은 손님 및 `in_ms` 값이 현재의 밀리세컨드보다 1시간 이상 차이나는, 즉 입장한지 1시간이 초과된 손님에 해당하는 쿼리를 통해 방역수칙을 지키지 않은 손님의 목록을 얻어낼 수 있는 메소드이다. 이 메소드를 통해 어떠한 손님이 1시간을 넘겼는지, 그 수는 몇명인지 알 수 있다.

## IV. 결론 및 향후 연구

본 논문에서는 블록체인을 활용하여 카페에 1시간 이상 머무르는 고객을 검출할 수 있는 시스템을 구현하였다. 이 시스템을 통해 COVID-19에 대한 좀 더 강경한 대응이 가능할 것으로 기대된다.

현재 시스템에서는 도커를 통한 네트워크 및 체인코드 구동만이 구현되어 있는데, 향후에 이 네트워크를 애플리케이션으로서 활용할 수 있도록 백엔드 및 프론트엔드 또한 구현할 예정이다.

## V. Acknowledgment

이 성과는 부분적으로 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478) 그리고 이 성과는 부분적으로 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.2018-0-00264, IoT 융합형 블록체인 플랫폼 보안 원천 기술 연구).

## [참고문헌]

- [1] “사회적 거리 두기 조정안 내일(1.18일)부터 시행”, 보건복지부 [Internet]. Available: <https://bit.ly/39IjK6l>
- [2] Satoshi Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”
- [3] Androulaki, Elli, et al. “Hyperledger fabric: a distributed operating system for permissioned blockchains.” Proceedings of the thirteenth EuroSys conference. 2018.
- [4] Beckert, Bernhard, et al. “Formal specification and verification of hyperledger fabric chaincode.” 3rd Symposium on Distributed Ledger Technology (SDLT-2018) co-located with ICFEM. 2018.

# 이더리움 이클립스 공격 분석

황보규민\*, 황선진\*, 최윤호\*\*

\*부산대학교 (대학원생), \*\*부산대학교 (교수)

## Analysis of Ethereum Eclipse Attack

Gyu-Min Hwang-Bo\*, Hwang Seon-Jin\*, Yoon-Ho Choi\*\*

\*Pusan National University(Graduate student),

\*\*Pusan National University(Professor)

### 요약

블록체인은 최근 여러 어플리케이션에 적용되어 활용되고 있다. 퍼블릭 블록체인인 이더리움은 분산화 어플리케이션(DApps)을 최상위 계층에서 구동하며 다양한 서비스를 제공하고 있다. 이러한 서비스는 2016년 DAO 계약 해킹으로 전체 이더리움 화폐의 10%가 도난당한 사건이 있고 이에 따라 어플리케이션 계층의 보안연구가 활발히 이루어지고 있다. 하지만 대부분의 보안 연구는 어플리케이션 계층에서 많고 네트워크 계층의 연구는 아직 미비한 상황이다. 따라서 본 논문은 대표적인 이더리움 네트워크 공격인 이클립스의 개념과 공격 방식을 분석하고 현재 대응책이 적용된 부분과 이더리움이 가지고 있는 취약점을 설명한다. 이를 통해 네트워크 공격의 특징을 이해하고 클라이언트 구현 시 위협요소들을 보완한다면 더욱 안전한 플랫폼을 구축할 수 있다.

## I. 서론

블록체인 기술은 분산 컴퓨팅 기반의 원장 관리 기술로써 최초의 블록체인인 비트코인을 사토시 나카모토가 제안했다. 이후 튜링 언어로 제작된 프로그래밍 가능한 스마트계약을 융합해 다양한 서비스를 가능하게 하는 이더리움이 등장한다. 퍼블릭 블록체인인 이더리움은 중앙 기관 없이 합의 알고리즘[1]을 통해 불변성과 일관성을 인정받는다.

이더리움은 분산 어플리케이션(DApps)을 이용하여 개발자들의 필요와 목적에 따라 다양한 서비스 제공이 가능하다. 이러한 분산 기술의 발전에 따라 안정적인 서비스 제공을 위해 블록체인 기반 어플리케이션 계층의 보안 연구가 활발히 이루어지고 있다.

하지만 대부분 어플리케이션 계층의 보안 위협 대응 방안 연구가 이루어지고 있고 네트워크 계층에서는 아직 미비한 상황이다.

본 논문에서는 2장에서 이클립스 공격의 개

념 및 공격 방식을 설명하고 3장에서는 결론을 서술한다.

## II. 이클립스 공격

이클립스 공격은 단일 노드를 대상으로 하는 공격으로써 희생자 노드의 네트워크를 다수의 악의적인 노드들로 독점하는 공격이다. 다음은 대표적인 이클립스 공격 방식이다.

### 2.1 Permanent Eclipse Attack

Permanet Eclipse Attack[2]은 블록 전파 방식의 취약점을 이용한 공격으로 공격 방법은 아래와 같다.

- 1) 공격자 노드는 최초 블록부터 정상 체인 보다 블록이 256개 많은 체인을 난이도를 낮추어 생성한다.
- 2) 희생자 노드는 공격자 노드에게 GetBlockHeaders를 전송한다.
- 3) 공격자 노드는 가장 높은 블록부터 제네

- 시스 블록까지 희생자 노드의 요청에 따라 헤더만 전송한다.
- 4) 희생자 노드가 모든 블록 헤더를 받으면 공격자 노드에게 블록 정보를 요청한다.
  - 5) 공격자 노드는 블록을 한 번에 보내지 않고 하나씩 전송함으로써 블록 전파를 지연시킨다.

## 2.2 Synchronising to longer chain with lower total difficulty

블록체인이 자신보다 낮은 전체 난이도를 가지고 있는 체인과는 블록 전파 과정을 가지지 않음을 이용한 공격[2]이다. 공격 방법은 아래와 같다.

- 1) 공격자 노드는 최초 블록부터 정상 체인 보다 블록이 256개 많은 체인을 난이도를 낮추어 생성한다.
- 2) 새롭게 참가한 노드와 연결된다.
- 3) 희생자 노드가 정상적인 다른 체인보다 256개 블록 이상 길다면 연결이 독점화된다.

## 2.3 Increasing the Minimum Difficulty Parameter in geth

공격자 노드가 최소 난이도로 채굴하여 희생자 노드에게 전송하면 희생자 노드의 최소 난이도가 높아진다. 이때, 정상 노드보다 최소 난이도가 높다면 희생자 노드가 정상 노드와 동기화를 하지 않는 취약점을 노린 공격[2]이다. 공격 방법은 아래와 같다.

- 1) 공격자 노드는 최소 난이도로 다른 체인 보다 더 긴 체인을 생성한다.
- 2) 공격자 노드는 최소 난이도로 긴 체인을 생성한다.
- 3) 희생자 노드는 공격자 노드 각 블록의 난이도를 확인한다.
- 4) 희생자 노드는 난이도를 확인할수록 최소 난이도 파라미터가 점차 증가하며 결국 최근 블록체인의 난이도보다 높아져 다른 정상 블록체인과 동기화가 불가능해진다.

## 2.4 Monopolize Connection Eclipse Attack

이 공격[3]은 희생자 노드가 다른 노드에게

outgoing 연결을 하기 이전에 공격자 노드로 빠르게 TCP 연결을 생성하는 공격이다. 공격 방법은 아래와 같다.

- 1) 공격자는 maxpeers보다 더 많은 수의 노드를 미리 생성한다.
- 2) 미리 생성한 공격자 노드에서 희생자 노드 리부팅 시 TCP 연결 메시지를 전송해 테이블을 빠르게 채운다.

위의 방법은 Geth 1.8 이전에 적용이 가능했고 1.8 이후에는 outbound와 inbound로 연결 가능한 Peer 수를 각각  $1/3 \times \text{maxpeers}$ ,  $2/3 \times \text{maxpeers}$ 로 제한하여 들어오는 TCP 연결로만 테이블이 다 채워지지 않도록 적용됐다.

## 2.5 Table Poisoning Eclipse Attack

이더리움은 노드 탐색 프로토콜을 통해 생성된 Table을 참조하여 연결을 설정한다. Table Poisoning은 테이블을 악의적인 노드로 채워 넣어 연결을 독점화 시키는 공격이다. 이 방법은 outbound와 inbound 연결 제한을 우회할 수 있다. 대표적인 공격 방식은 두 가지가 있다.

첫 번째는 inbound 연결을 생성하는 UDP listener이 outbound 연결을 보장하는 테이블 seeding 과정 보다 더 먼저 실행되는 점을 이용한 공격[3]이다. 이는 하나의 IP로도 공격이 가능하고 방법은 아래와 같다.

- 1) 공격자는 미리 많은 수의 노드를 생성한다.
- 2) 희생자 노드에게 ping을 전송해서 db에 공격 노드를 삽입한다.
- 3) 희생자 노드 리부팅 시, ping을 다량 전송하여 테이블을 공격자 노드로 포화시켜 db에 있는 노드들이 테이블로 들어오지 못하게 한다.

이러한 공격 방식은 Geth 1.8 이후부터는 테이블에 IP subnet을 적용하여 테이블이 하나의 IP로 다 채워지지 않도록 수정됐다.

하지만 여전히 IP subnet을 우회할 수 있는 이클립스 공격[4]이 가능하다. 공격 방법은 아래와 같다.

- 1) 공격자 노드는 미리 많은 수의 노드를 생성한다.

- 2) 희생자 노드에게 지속적인 TCP 연결을 신청한다.
- 3) 희생자 노드에게서 FIND\_NODE 패킷이 왔을 때, 미리 생성해둔 노드ID들 중 타깃과 가장 가까운 노드를 탑장한다.

Inbound 연결은 특정 노드 탐색 프로토콜 버전에서 여전히 지속적인 TCP 연결로 장악이 가능하다.

Outbound 연결의 절반은 ReadRandomNodes 과정으로 선택하고 나머지 절반은 lookup buffer 과정으로 선택하며 악의적인 노드로 연결이 가능하다.

ReadRandomNodes는 테이블 내에서 무작위 노드를 선택하는데 노드 탐색 프로토콜의 버전에 따라 다른 랜덤 한 방식을 선택한다. 버전 4에서는 테이블 내 존재하는 모든 노드를 랜덤하게 선택하여 연결을 설정하며 테이블을 독점화하는 데에 다양한 IP가 필요하다. 하지만 버전 5에서는 테이블 내 노드가 아닌 베킷 선택만 랜덤하게 하고 이후 베킷의 헤더 노드만 추출하여 피어를 연결한다. 따라서 버전 5에서는 IP subnet 개념이 적용되었더라도 실제 Peer로 연결되는 각 베킷의 헤더만 채우면 되어 적은 수의 IP로도 outbound 연결의 절반을 장악할 수 있다. Ping을 지속적으로 보내면 노드가 베킷의 헤더로 이동함으로 베킷 내 공격 노드가 들어가 있다면 헤더로 올리는 방법은 간단하다.

Lookup-buffer은 희생자 노드가 FIND\_NODE 패킷을 보낼 때, 미리 생성해둔 노드 ID 중 타깃과 가장 가까운 노드들을 전송하면 공격 노드로 채울 수 있다. 따라서 outbound 연결 중 나머지 절반을 공격 노드로 마저 채울 수 있다. 이와 같은 공격이 가능하게 하는 근본적 이더리움 네트워크 취약점 중 하나는 노드 ID가 비용 없이 다량 생성이 가능하다는 점이다. 즉, 이더리움 네트워크 참여가 비용이 거의 발생하지 않아 공격 노드를 다량 생성할 수 있어 이클립스 공격에 취약하다.

또한 이더리움은 노드 ID와 테이블이 공개되어 있고 노드 ID 간 거리를 구하는 방식이 모두 같다는 점이 취약할 수 있다. 공개된 노드 ID와 공격 노드 ID의 거리를 구한 뒤 희생자

노드의 테이블 내 원하는 베킷에 공격 노드를 삽입할 수 있기 때문이다.

### III. 결론

다양한 블록체인 플랫폼 기술이 발전하고 이를 활용하는 서비스가 증가함에 따라 DAO와 같은 보안 위협 또한 증가하고 있다. 하지만 대부분 최상위 계층에서의 보안연구가 이루어지고 있고 네트워크 계층에서의 보안은 미비하다. 본 논문에서 대표적인 네트워크 공격인 이클립스 공격의 개념을 살펴보고 다양한 공격 방식에 대해 분석하였다. 이를 통해 이더리움 네트워크의 취약한 부분을 이해하여 위협요소들을 보완한다면 더욱 안전한 서비스를 구축할 수 있을 것이다.

### Acknowledgement

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(2019-0-013 43,융합보안핵심인재양성사업) 및 4단계 BK21, 동남권4차산업혁명리더양성사업단에 의하여 지원되었음.

### [참고문헌]

- [1] Go Ethereum, Retrieved Feb., 3, 2021, from <https://github.com/ethereum/ethas>
- [2] K. Wüst and A. Gervais, "Ethereum eclipse attacks," ETH Zurich, Tech. Rep., 2016.
- [3] Y. Marcus, E. Heilman, and S. Goldberg, "Low-resource eclipse attacks on Ethereum's peer-to-peer network," IACR, vol. 246, 2018
- [4] Sebastian Henning sen, Daniel Teunis, Martin Florian, and Björn Scheuermann. Eclipsing ethereum peers with false friends. In 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pages 300 - 309. IEEE, 2019.

# 코로나-19 역학조사를 위한 블록체인 기반 QR코드 전자출입명부 관리 시스템

한찬희\*, 김문선\*\*, 이만희\*\*\*

\*한남대학교(학부생), \*\*한남대학교(대학원생), \*\*\*한남대학교(교수)

Hyperledger fabric-based QR code electronic access list management system for Corona-19 epidemiological investigation

Chan-Hee Han\*, Moon-Sun Kim\*\*, Man-Hee Lee\*\*\*

\*Hannam University(Undergraduate student),

\*\*Hannam University(Graduate student), \*\*\*Hannam University(Professor)

## 요약

최근 코로나-19로 인한 팬데믹 상황에 대응하기 위해 출입 명부 관리 시스템이 도입되었다. 해당 시스템은 초기에는 수기로 작성되어 허위 기재 및 개인 정보 유출로 인한 문제가 발생하였다. 향후 도입된 QR코드 기반 전자 출입 명부 시스템은 사용 방법이 상대적으로 복잡하고 중앙에서 모든 정보가 관리되기 때문에 출입 정보에 대한 보안성이 부족하다. 또한 번거로운 본인 인증 과정을 거쳐야 하므로 사용성이 불편하다. 본 논문은 전자 출입 명부에 대한 보안성을 강화하고 사용 편의성을 개선한 블록체인 기반 전자 출입 명부 관리 시스템을 제안한다. 이 시스템은 별다른 인증 없이 휴대전화를 암호화한 정보를 바탕으로 개인을 식별하며, RSA 암호화를 바탕으로 개인정보 및 출입 기록을 체인 서버에 저장한다. 제안한 시스템은 기존 전자 출입 명부의 단점을 개선하였으며, 보다 효과적으로 코로나-19 역학조사에 기여할 수 있을 것으로 기대한다.

## I. 서론

코로나-19사태가 장기화함에 따라 질병관리청에서는 확산 방지 및 방역을 위하여 카페, 식당, 공공시설 등 모든 다중 이용 시설은 출입 명부 작성 의무 시행하도록 공공시설 출입 명부 작성 시스템을 도입하였다. 하지만 수기로 작성하는 출입 명부는 누구든지 쉽게 열람할 수 있어, 개인 정보 유출의 가능성이 존재한다. 또한 일부 방문자의 허위정보 기재로 인해 환자 발생 시 코로나-19 역학조사에 혼선을 야기하고 막대한 경제적인 피해가 발생한다[1][2].

이를 방지하기 위해 보건복지부는 QR코드를 통한 전자출입명부 기술을 구축했다. 이는 허위 기재 문제를 효과적으로 해결할 수 있지만, 본인인증 절차가 복잡하다[3]. 또한, 사회보장정보원의 중앙서버에 모든 데이터가 저장되는 문제

점이 있다. 이는 데이터센터에 문제가 발생하면 출입 명부 기록이 모두 소실될 수 있는 위험성이 있으며, 코로나-19 역학조사가 불가능해지는 치명적 위험성을 가지고 있다[4].

본 논문은 이러한 문제점을 해결하기 위해 블록체인 기반 분산저장 전자 출입 명부 시스템을 제안한다. 이 시스템은 QR코드 생성 시, 본인인증 과정을 거치지 않으며 오직 전화번호와 출입 시작을 활용하여 QR코드를 생성한다. 생성한 QR코드는 중앙에서 제공하는 RAS 공개키로 암호화되어 체인 서버로 전송된다. 블록체인 서버는 허가형(private) 블록체인인 하이퍼레저 패브릭(hyperledger fabric)으로 구축하여 출입 기록을 더욱 안전하게 저장할 수 있다. 해당 시스템은 하이퍼레저 패브릭과 안드로이드 앱으로 실제 프로토타입이 구현되었다.

본 논문의 구성은 다음과 같다. 먼저 2장에

서는 체인 서버 구축을 하기 위한 블록체인 플랫폼으로 알려진 하이퍼레저 패브릭(Hyperledger fabric)과 관련 연구에 대해 살펴본다. 3장에서는 전체적인 시스템 구조에 관해 설명한 다음, 시스템의 프로토콜에 대하여 설명한다. 마지막 4장에서는 논문의 간단한 결론을 맺는다.

## II. 관련 연구

### 2.1 하이퍼레저 패브릭

하이퍼레저 패브릭(이하 하이퍼레저)은 리눅스 재단에서 관리하는 오픈소스 블록체인 플랫폼이다[5]. 하이퍼레저는 누구나 자유롭게 참여 가능한 기존의 퍼블릭 블록체인이 아닌, 인증 관리 시스템에 의해 허가된 사용자만이 블록체인 네트워크에 참여할 수 있는 허가형 프라이빗 블록체인이다. 따라서 여러 허가된 사업장 간에 데이터가 동기화하여 분산 저장하기 때문에 공격자가 한쪽 노드에 공격을 가하더라도 나머지 노드에서 지속해서 동기화하고 검증하여 더욱 안전한 데이터 관리가 가능하다.

국내에서는 Ye-Jin Choi[6] 등이 하이퍼레저 기반 헬스케어 데이터 공유 플랫폼을 개발하였다. 이 시스템은 블록체인 기반 저장 방식이 수집이 허가된 익명 의료 데이터를 기존의 방식 보다 더 안전하게 보관할 수 있음을 보였다. Jin-Suk Bong[7] 등 또한 하이퍼레저를 이용한 의료 데이터를 더욱 안전하게 저장하는 시스템을 제안했다.

### 2.2 DApp

DApp(Decentralized Application)는 탈중앙화 애플리케이션으로 블록체인 기술을 사용하여 중앙서버 없이 네트워크상에서 정보를 분산하고 저장한다. 구동 방식은 스마트 컨트랙트(Smart Contract)를 통해 명령을 수행하고 정보를 불러오고 저장할 수 있다. 스마트 컨트랙트는 블록체인에서 미리 정해진 조건을 달성 시 자동으로 계약의 내용을 수행하는 일련의 소프트웨어 코드이다.

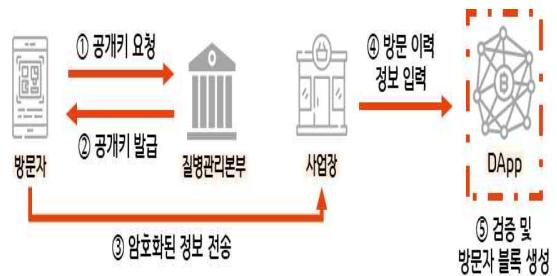
본 논문은 기존 연구와 같이 하이퍼레저가

허가형 프라이빗 블록체인이라는 장점을 활용하여 국가적 재난 속에서 바이러스의 확산을 최소화하기 위한 전자출입명부 관리 시스템을 소개한다.

## III. 제안된 시스템 설계 및 구현

본 절에서는 평상시 다중 이용 시설에 출입하는 방문자들을 관리하고 확진자 발생 시 확진자의 동선을 파악하는 일련의 프로토콜을 소개한다.

### 3.1 평상시



[그림 1] 다중 이용 시설 출입 시

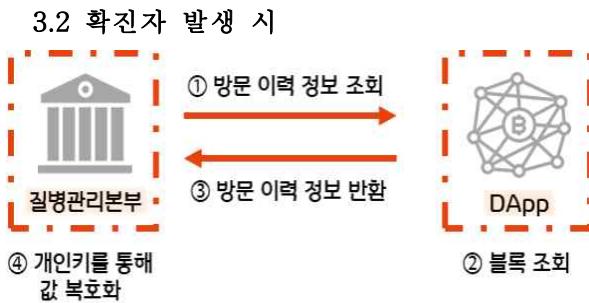
그림 1은 다중 이용 시설을 출입할 때의 상황이다. 방문자는 스마트폰에 본 연구에서 개발한 앱을 사용하여 QR코드를 생성한다. 또한, 사업장도 개발된 앱을 사용하여 방문자의 정보를 검증하고 블록을 생성하여 체인 서버에 저장한다. 프로토콜 순서는 다음과 같다.

①, ② 생성 날짜의 공개키를 질병관리본부에 요청하여 발급받는다.

③ 받은 공개키를 방문자의 휴대전화 번호와 방문 시간에 대한 문자열을 공개키 암호화 알고리즘을 사용하여 암호화를 한다. 암호화한 정보를 바탕으로 QR코드를 생성한다.

④ 사업장에서 QR코드 스캔 시, 방문자 앱에서 생성한 QR코드에 담긴 정보와 사업장 정보(사업자등록번호, 대표자명, 사업장 주소, 휴대전화 번호 등)를 함께 블록체인 서버에 전송한다.

⑤ DApp을 통해 암호화한 방문자의 정보와 사업장 정보를 검증한 후 블록을 생성하여 블록체인 서버에 저장한다.



[그림 2] 확진자 동선 조사 시

그림 2는 확진자 발생 시 동선을 조사하는 과정을 나타낸다.

① 관리자는 체인 서버에 방문 이력 정보를 조회한다.

② 암호화된 확진자의 정보를 받아 해당하는 블록을 조회한다.

③ 질병관리본부로 방문 이력 정보를 반환한다.

④ 질병관리본부에서 받은 개인 키를 사용하여 암호화되어 있는 정보를 복호화한다.

복호화된 정보를 통해 확진자가 방문했던 사업장을 조회하고 해당 사업장에 방문한 방문자의 명단을 조회한다.

#### IV. 결론

본 논문에서는 코로나-19로 인해 의무로 시행하게 된 출입 명부 작성에 대해 허위 정보기재와 개인정보 도난을 방지하기 위한 블록체인 기반 QR코드 전자출입명부 관리 시스템을 제안하였다. 기존의 복잡한 인증 절차를 거쳐 QR코드를 생성하는 방식에서 앱만 켜면 바로 QR코드로 체크인을 할 수 있는 서비스로 개발하였다. 또한, 휴대전화 번호와 방문 시간을 암호화했던 공개키는 14일 이후 자동으로 폐기하도록 하여 휴대전화 번호를 제외한 다른 정보(이름, 성별 등)를 제공하지 않기 때문에 개인정보 유출을 최소화한다. 향후 연구로 현재 개발한 앱은 기존의 네이버나 카카오에서 만든 QR코드와 호환이 불가능하여 추후 호환이 가능하도록 연구 및 개발할 예정이다.

블록체인의 특징을 활용하여 분산 저장된 데이터를 동기화하고 검증하여 기존의 중앙통제형 서버 시스템보다 안전하게 관리할 수 있다. 따라서 본 시스템은 방문자의 개인정보를 안전하게 보호하고 확진자 발생 시 정확하고 무결한 데이터를 통한 동선 파악을 제공하는데 효과적일 것으로 기대된다.

#### [참고문헌]

- [1] Nam-E Kim, “내가 적었던 ‘식당 출입자 명부’ 진짜로 불법 거래됐다”, <https://news.mt.co.kr/mtview.php?no=202012208430663386>, November, 2020.
- [2] So-Yeong Kim, “명부 적으면서도 불안했는데”… 개인정보 노출 우려 현실화되나”, <https://news.joins.com/article/23927707>, November, 2002.
- [3] Han-Gyeol Joung, Young-Sang Kim, Kang-Jun Lee and Kyong-Hun Jeong, “큐알 머시기가 뭐여? 헬스장 당황한 70대, 그 뒤로 줄줄이…”, <https://news.mt.co.kr/mtview.php?no=2020070113080196534>, July, 2020.
- [4] 중앙방역대책본부, 중앙사고수습본부, June, 2020, ”전자출입명부 활용 안내(안) (이용자 및 시설관리자용)”,
- [5] HYPERLEDGER, “Hyperledger Blockchain Performance Metrics”, October, 2018.
- [6] Ye-Jin Choi and Kyoung-jin Kim, “Secure Healthcare Data Management and Sharing Platform Based on Hyperledger Fabric”, Journal of Internet Computing and Services(JICS), February, 2020.
- [7] Jin-Suk Bong, “A Personal Health Information Sharing Platform based on Hyperledger Fabric Blockchai”, June, 2019.

# 탈중앙형 신원 식별 서비스 기반 안전한 접근 제어 기법에 관한 고찰

조강우\*, 정병규\*, 신상욱\*\*

\*부경대학교 대학원 정보보호학과 (대학원생)

\*\*부경대학교 IT융합응용공학과 (교수)

## A Study on Secure Access Control Management Based on Decentralized Identification Service

Kang Woo Cho\*, Byeng-Gyu Jeong\*, Sang Uk Shin\*\*

\*Dept. of Information Security, Graduated School,

Pukyong National University (Graduate Student)

\*\*Dept. of IT Convergence and Application Eng.,

Pukyong National University (Professor)

### 요약

급증하는 차세대 신원 식별 서비스 수요에 따라 안전한 탈중앙형 ACM 기법에 대한 요구가 상승하였다. 이에 따라 탈중앙형 신원 식별 서비스에 기반하는 다양한 ACM 기법이 제안되었으나 대부분 연구 도입기에 해당한다. 본 논문에서는 탈중앙형 신원 식별 서비스에 대한 개요에 대하여 논하고 안전한 ACM 기법의 요구사항 및 현행 연구 동향에 관한 고찰을 수행하며, 결과적으로 탈중앙형 신원 식별 서비스 기반 안전한 접근 제어 기법에 대한 연구 시사점을 제공한다.

### I. 서론

다양한 신원 식별 기법이 탈중앙 서비스 형태로 변화함에 따라, 기존 중앙 집중 형태의 신원 식별에서 활용하던 접근 제어 관리(Access Control Management, ACM) 기법의 변화가 요구되었다. 중앙 집중형 ACM의 경우 일괄적으로 저장되는 신원 데이터에 대한 효율적인 정책 관리가 가능하며, 모든 네트워크 참여자가 신뢰하는 정책에 따른 접근 권한 분배가 용이했다. 하지만 탈중앙 형태의 신원 식별은 신원 데이터를 분산된 원장에 기록하며, 상호 비신뢰 관계의 P2P(Peer-to-Peer) 네트워크를 형성하기 때문에 이러한 기존 ACM 기법을 적용하는 것에 다수의 제약이 잇따른다.

탈중앙형 신원 식별 기법 연구는 단순한

DID(Decnetrnlized IDentity) 형태에서 발전하여 DIDs(Decentralized IDentifiers)를 포함한 SSI(Self-Sovereign Identity) 기술로 변화한다. 현행 연구의 초점은 탈중앙 기술에 범용적으로 적용 가능한 ACM 기법을 개발하는 것에 있다.

본 논문에서는 2장에서 탈중앙형 신원 식별에 대한 배경지식을 제시한다. 3장에서는 탈중앙형 신원 식별에 대한 안전한 ACM 기법의 요구사항을 도출하고 이를 바탕으로 기존 제안된 탈중앙형 신원 식별 기반 ACM 관련 연구를 논하며, 4장에서 결론을 제시한다.

### II. 탈중앙형 신원 식별

탈중앙형 신원 식별은 식별자와 신원으로 분류된다. 그 중 식별자 기술에 해당하는 DIDs는

실질적인 탈중앙형 동작을 가능하게 하는 기술이다[1]. 이는 지시된 신원 데이터에 식별자 테깅을 통해 정확한 신원 식별을 제공하며, 나아가 안전한 로컬 스토리지에 저장되는 신원 데이터에 탈중앙 속성을 부여한다. 일반적인 DIDs Resolver는 다음의 형식을 따른다.

**did : [method] : [identifier]**

한편, DIDs를 사용하는 SSI는 검증 가능한 자격 증명인 VC(Verifiable Credential)을 포함한 탈중앙형 신원 식별 솔루션이다. 신원 검증을 위한 방법으로 VC를 VP(Verifiable Presentation)로 2차 가공하는 것을 통해 정보 주체의 강력한 선택권, 삭제권 등의 자기 주권형 동작을 보장하는 것이 특징이며, 영지식 증명(Zero-Knowledge Proof) 등을 통해 신원 데이터를 안전하게 거래할 수 있도록 고안되었다. 또한, 데이터 거래 환경에 Steward-Ownership 기반 프라이빗 블록체인 네트워크를 도입하는 것으로 탈중앙형 동작을 달성하며 네트워크에 참여하는 각 노드가 상호 비신뢰 관계에서도 DID 검증을 통해 증명인-발행인-검증인으로 구분되는 역할을 유동적으로 수행할 수 있다.

현행 SSI 연구로는 Linux 재단의 Hyperledger Indy 프로젝트 일종인 SOVRIN, Etherium 기반 uPort, Jolocom 등을 비롯하여 ShoCard, OmniOne 등 다수 존재하며, 이들은 상이한 자체 개발 ACM 기법을 채택하였다[2].

### III. DIDs/SSI 기반 안전한 접근 제어

#### 3.1 요구사항

일반적인 ACM의 요구사항은 다음과 같다[3].

1. 최소 권한 : 사용자가 수행할 작업에 필요한 권한만이 역할에 부여된다.
2. 직무 분리 : 접근 제어 관리자가 접근 권한을 스스로 부여하는 것을 방지하기 위해 상호 배타적 역할을 호출할 수 있어야 한다.
3. 컨텍스트 기반 권한 : 접근 제어 결정을 계산하는데 반드시 필요한 정보만을 저장하고 처리한다.

하지만 SSI는 ID의 관리 권한을 개인에게 부여하기 때문에 개인 중심형 ID 기술로 분류된다. 따라서 이는 중앙 집중식 기관이 존재하지 않기 때문에 기존의 ID 관리와는 대조적인 개념으로, 다음과 같은 탈중앙 형태의 요구사항을 추가적으로 지닌다[4].

1. 참여자 선택 : 접근 제어 정책에 합당한 VC를 보유한 사용자만이 권한 부여의 자격을 지닌다.
2. 데이터 기밀성 : 접근 제어 엔진은 ZKP를 통해 최소한의 정보만을 기반으로 의사 결정을 수행한다.
3. 책임 및 부인 방지 : 발행인은 발행한 VC에 대한, 검증인은 검증한 VC에 대한 책임을 지닌다. 동시에 상호 비신뢰 관계의 탈중앙 네트워크에서 신원 식별과 검증에 대한 송수신 및 서비스 종단점에서의 부인이 불가능하여야 한다.

#### 3.2. SSIBAC

SSIBAC은 인증 및 자격 증명 발급을 위한 SSI 모델에서 사용자 데이터 프라이버시 및 주권에 초점을 맞춘 ACM 기법 제안이다[4]. 이는 사용자가 데이터에 접근하기 위해 컨텍스트 기반 권한 VC를 PRP(Policy Retrieval Point)에 저장하는 것으로 기존 중앙 집중형 ACM에서 사용하던 접근 제어 정책을 블록체인 기반 SSI 솔루션에 이식하는 것을 목표로 한다.

이는 발행인에 의해 발행된 VC를 증명인이 VP로 생성하기 위해 분산 원장으로부터 Schema 요소를 획득하는 시점에서 접근 제어를 수행한다. Schema는 검증인이 증명인에게 요구하는 VC 속성을 포함하고 있으며, 증명인은 이를 기반으로 검증에 필요한 최소한의 VC 속성 정보를 조합하여 VP를 생성할 수 있다. SSIBAC은 검증인이 Schema에 PRP의 접근 제어 정책 VC를 포함하여 증명인에게 반환한다. 즉, 검증인은 Schema에 접근 제어 정책을 포함한 Challenge를 생성하며, 증명인은 이에 응답할 수 있어야 한다. 증명인은 보유한 VC의 신원 속성과 접근 제어 정책에 대한 Challenge 응

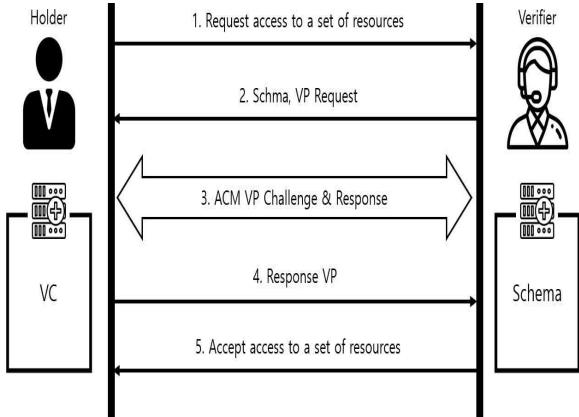


Fig 1. SSIBAC Architecture

답을 VP의 형태로 구성하여 검증인에게 전송하며, 검증인은 이를 검증하는 것으로 접근 권한을 부여할 수 있다.

### 3.3. Jolocom

Jolocom은 계층형 결정적(Hierachial Deterministic, HD) 키 기반의 ID 관리 시스템을 제안하였다[5]. SOVRIN 등의 오픈 소스 프로젝트와 마찬가지로 정보 주체에 대한 제한적 정보 제공, 명백한 동의, 삭제권 등 강력한 자가 주권형 동작을 보장하는 것을 목표로 한다.

Jolocom은 알려진 시드(Seed)에서 생성되어 사용자에 의해 직접적으로 제어되는 HD 키를 사용한다. 이는 계층형 결정적 특성으로 인해 계층적인 복수 파생키를 생성할 수 있으며 같은 시드를 공유한다. HD 키의 각 파생키는 Jolocom 환경에서 페르소나(Personas)로 정의된 하위 ID를 생성하며, 원본 HD 키를 DID와 결합한다. 이후 파생키 및 시드를 기반으로 하여 개인의 페르소나를 식별하고, ACM을 위한 각 페르소나에 IPFS의 해시 맵핑을 수행한다. 결과적으로 IPFS 구조 하에서 이더리움의 스마트 컨트랙트를 통한 접근 제어를 달성한다.

## IV. 결론

본 논문은 SSI의 탈중앙형 동작 개념 하에서 ACM을 제공하기 위한 현행 연구 동향의 고찰을 진행하였다. 그 결과, SSI 솔루션의 ACM 기준이 제안되지 않았기 때문에 기존 SSI 솔루

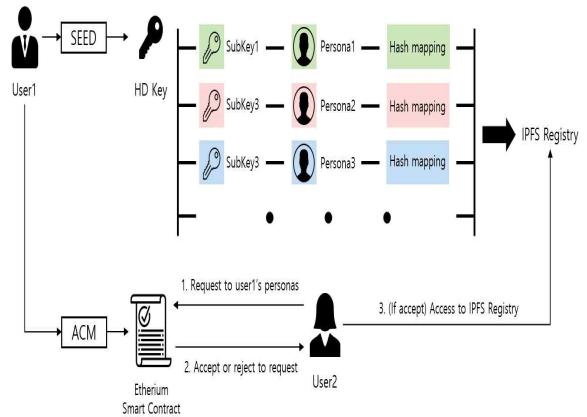


Fig 2. Jolocom ACM Architecture

션들은 독립적인 ACM 기술을 구현하는 것으로 분석되었다. 이에 본 논문은 관련 연구 분석을 통해 탈중앙형 신원 식별 서비스 기반 안전한 접근 제어 기법에 대한 연구 시사점을 활기하였으며, 추후 다양한 관점에서의 ACM 기법 연구가 제안될 것임을 기대할 수 있다.

## Acknowledgement

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구 사업임 (No. 2019R1I1A3A01060652).

## [참고문헌]

- [1] D. Reed et al., “Decentralized Identifiers(DIDs) v1.0”, W3C Working Draft, (2020, Nov. 08) [Online]. Available: <https://www.w3.org/TR/did-core/>
- [2] Roos, Julian. "Identity management on the blockchain." Network 105 (2018).
- [3] S. Sandhu, E. Coyne, H. Feinstein, and C. Youman, "Role Based Access Control Models," Computer, vol. 29, no. 2, pp. 38 - 47, 1996.
- [4] Belchior, Rafael, et al. "SSIBAC: Self-Sovereign Identity Based Access Control." (2020).
- [5] Fei, Ch, et al. "Jolocom: Self-sovereign and decentralised identity by design." White paper (2018).

# 하이퍼레저 패브릭 블록체인 기반 간단검증 탈중앙 신원 증명 시스템

조욱\*, 김도훈\*\*, 김호원†

\* , \*\* 부산대학교 (대학원생), † 부산대학교 (교수)

## Lightweight Verification Decentralized Identifier System based Hyperledger Fabric Blockchain

Uk Jo\*, Dohun Kim\*\*and Howon Kim†,

\*, \*\* Pusan National University (Graduate student)

† Pusan National University (Professor)

### 요약

탈중앙 신원 증명 시스템(Decentralized Identifier, DID)은 사용자 신원 증명을 중앙 시스템에서 사용자 중심으로 변화시켰다. 탈중앙 신원 증명 시스템은 사용자의 신원 증명시 개인정보 및 제 공을 사용자가 직접 관리할 수 있다. 현재 DID 증명은 발급기관이 발급한 VC(Verifiable Credential)를 사용자가 VP(Verifiable Presentation)로 가공하여 서명한 후 검증기관에 제출하고 검증기관은 두 번의 서명을 검증하여 신원 증명을 완료한다. 본 논문에서는 두 번의 서명검증 단계를 VP ID 등록 검증과 VP검증 단계로 성능을 높임을 보였다. VP ID 등록 검증을 통해 불 필요한 서명검증 단계의 계산 시간을 줄이는 효과를 보였다.

## I. 서론

정보의 디지털화가 가속화되며 인터넷을 통해 유통되는 개인정보는 급속도로 증가하고 있다. 무분별한 개인정보 제공으로 발생하는 개인정보의 유출을 방지하기 위해 개인은 자신의 개인정보에 대해 주체적으로 정보를 소유하며 통제할 수 있는 방법이 필요하다.

블록체인은 4차 산업혁명의 발전과 함께 다양한 도메인(금융, 물류, 부동산)과 융합되어 이용되고 연구되고 있다.[1] 특히 탈중앙 신원 증명 시스템은 블록체인으로 자기 주권 신원 (Self-Sovereign Identity, SSI)를 구현한 시스템 [2]으로 기존의 중앙 시스템이 사용자의 개인정보를 보관하고 관리하는 방식에서 벗어나 사용자가 자신의 개인정보를 전적으로 소유하고 관리한다.[3]

본 논문에서는 블록체인 기반의 탈중앙 신원 증명 시스템의 성능을 높이는 것을 목표로 한

다. 특히 검증단계에서 검증기관이 사용자가 제출한 VP의 두 번의 서명을 검증하는 과정을 query transaction으로 확인하는 VP ID 등록 검증과 VP 검증 단계로 나눔으로써 성능을 높이는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 하이퍼레저 패브릭과 탈중앙 신원 증명 시스템을 설명한다. 3장에서 하이퍼레저 패브릭 기반 간단검증 탈중앙 신원 증명 시스템을 제안하고 간단자격검증 단계를 소개하여 성능 향상됨을 보인다. 4장에서 결론을 맺는다.

## II. 배경지식

### 2.1 하이퍼레저 패브릭(Hyperledger Fabric)

리눅스 재단에서 진행중인 오픈소스 기반 블록체인 프로젝트인 Hyperledger의 Fabric은 기존의 누구나 참여 가능한 비허가형 블록체인(비트코인, 이더리움)이 아닌, 시스템 관리자가 허

가한 참여자만 참여 가능한 허가형 블록체인 네트워크 구조를 가지고 있다.[4]

탈중앙화, 데이터의 무결성과 같은 블록체인의 장점을 활용하기 위해 다양한 블록체인 플랫폼이 연구·개발되고 있지만, 네트워크의 참여자와 트랜잭션 처리가 증가하며 처리속도에 대한 효율성 문제가 발생하고 있다. 이러한 문제점을 해결하기 위해 기존 블록체인의 장점을 가지며 필요한 참가자만 사용 가능한 허가형 블록체인인 Hyperledger Fabric이 개발되었다.

## 2.2 탈중앙 신원 증명

기존의 중앙기관 중심의 신원인증 방식이 아닌 정보 주체가 신원인증에 필요한 정보를 직접 소유하며 블록체인 플랫폼에 저장된 인증정보를 통해 사용자의 신원을 검증한다.[5] 탈중앙 신원 증명을 통해 발급받은 유일한 식별자를 통해 사용자를 구분하며 불필요한 추가 개인정보 노출의 위험없이 필수정보만을 제출하여 사용할 수 있다.

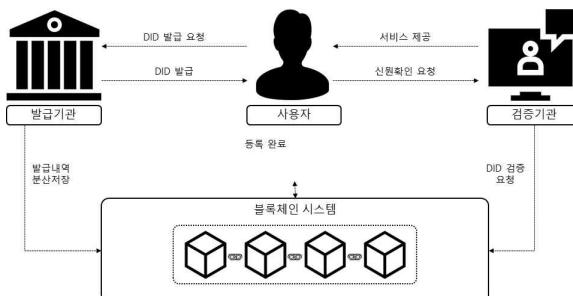


그림 1. DID 기반 신원인증 절차

## III. 본론

하이퍼레저 패브릭 기반 간단검증 탈중앙 신원 증명 시스템 아키텍처는 그림1에서 보이듯이 발급기관, 검증기관, 사용자, 리졸버, 블록체인 시스템으로 구성된다. 블록체인 시스템은 전체 참여자 공개키와 발급기관이 VP ID를 등록하는 채널로 나뉜다. 발급기관과 검증기관 그리고 사용자는 리졸버를 통해 블록체인 시스템과 통신한다. 사용자는 발급기관을 통해 자격을 검증받고, 사용자가 검증기관의 서비스를 이용하고자 할 때 발급기관을 통해 받은 자격을 통해 이용할 수 있다. 시스템의 과정은 자격 발급 단

계, 자격 검증 단계로 나뉜다.

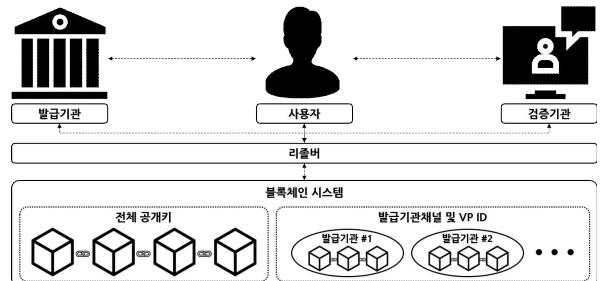


그림 2. 탈중앙 아이디 시스템 아키텍처

그림2는 자격발급단계의 전체 흐름을 보여준다. 자격발급단계 전에 모든 참가자는 자신의 공개키를 블록체인 시스템에 저장한다. 자격발급단계는 발급기관과 사용자 사이에서 일어난다. 사용자가 발급기관에게 자격 발급 요청을 하면 발급기관은 사용자의 자격을 자체적으로 검증하고 검증이 완료되면 VC를 사용자에게 전달하고 리졸버에게 블록체인 시스템에 VC ID 등록 요청을 한다. 리졸버는 발급기관의 해당 채널에 사용자의 VC ID를 등록한다.

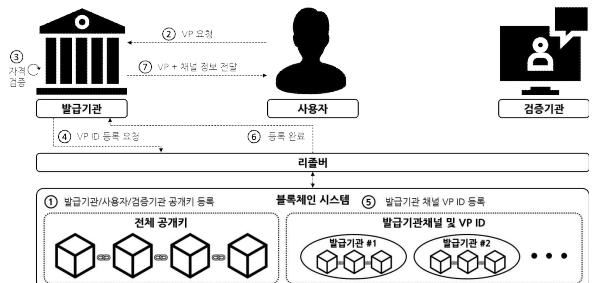


그림 3. 자격 발급 단계

그림3의 자격검증단계는 검증기관과 사용자 사이에서 발생한다. 제안된 검증 단계는 2단계로 나뉘며, 전체자격검증과 간단자격검증으로 구성된다. 전체자격검증은 기존 DID 검증 단계와 같이 VP를 사용자 서명과 발급기관 서명을 각각 검증한다. 간단자격검증은 검증기관이 리졸버에게 VP ID와 검증기관 채널을 리졸버에게 전달하고 해당 채널에 VC ID가 등록되어 있는지 검증 요청한다. 리졸버는 해당 채널에 사용자 공개키가 있는지 확인하고 등록되어 있는 경우 발급자 결과와 발급자 공개키를 전달하고 등록되어 있지 않은 경우 검증 실패 결과를 전달한다. 검증기관은 VP의 발급자 서명을 검증

함으로써 사용자 VP의 자격 검증을 완료한다.

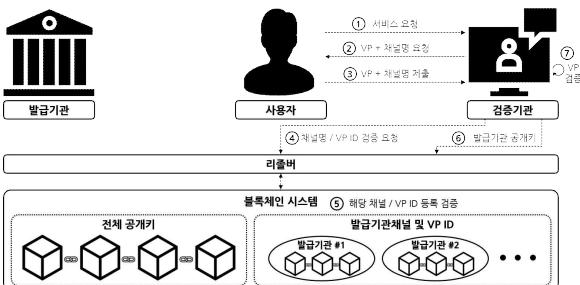


그림 4. 간단 자격 검증

제안한 간단검증단계는 사용자가 제출한 VP의 서명을 검증하기 전 VP의 등록 여부를 확인함으로써 불필요한 검증 과정을 생략할 수 있다. 하이퍼레저 패브릭 블록체인은 state db를 사용하여 가장 최신의 key/value 값을 query transaction으로 간단하게 확인 가능하다. 보증(endorsement)-오더링(ordering)-검증(validation) 단계를 거치는 invoke-trasaction 과는 다르게 그림4와 같이 피어내에서 체인코드 시뮬레이션만을 통해 해당 값의 존재 여부만을 간단하게 확인 가능하다. 등록 여부가 확인되면 발급기관의 공개키로 VP 서명을 검증한다. 기존의 검증 단계가 ECDSA(Elliptic Curve Digital Signature Algorithm)을 두 번 계산하는 것에 반해 간단검증단계는 한 번의 query transation 호출과 ECDSA 한 번 계산하는 것으로 가능함을 보임으로 성능 개선이 가능함을 보였다.

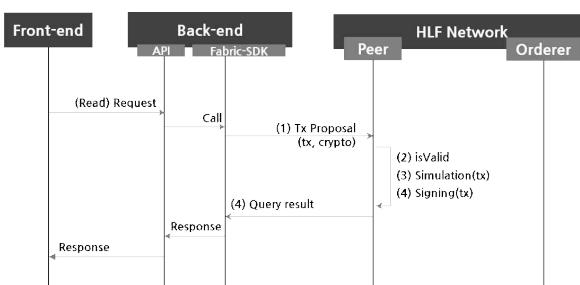


그림 5. query-transaction flow

#### IV. 결론

본 논문에서는 하이퍼레저 패브릭 블록체인을 기반 간단검증 탈중앙 신원 증명 시스템에 대해서 제안했다. 특히 검증단계에서 발급기관 서명만을 검증하는 간단검증단계를 제안했다.

간단검증단계는 query transaction을 통해 블록체인내에 블록을 생성하지 않고도 피어내 시뮬레이션만을 통해 간단하게 등록여부를 확인할 수 있으며 등록 여부가 확인 되었을 경우에만 VP 서명을 검증한다. 기존 VP의 두 번의 ECDSA 계산 서명검증 방식을 한 번의 query transaction 호출과 한 번의 ECDSA 계산 방식으로 가능함을 보였다. 향후 제안한 인증 방식에 대한 구현을 통해 기존 방식들과의 성능차이에 대해서 분석하고자 한다.

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-01343, 융합보안핵심인재 양성사업)

#### [참고문헌]

- [1] Mohamed, N., & Al-Jaroodi, J. (2019, January). Applying blockchain in industry 4.0 applications. In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0852–0858). IEEE.
- [2] Zbinden, F., & Kondova, G. (2019). Economic development in Mexico and the role of blockchain. Advances in Economics and Business, 7(1), 55–64.
- [3] EU Blockchain Observatory and Forum, “Blockchain and Digital Identity,” 2019. <https://www.eublockchainforum.eu/reports>
- [4] Hyperledger Fabric: Read-the-Docs, <https://hyperledger-fabric.readthedocs.io/en/latest/>,
- [5] World Wide Web Consortium (W3C), “Decentralized Identifiers (DIDs) v1.0: Core Data Model and Syntaxes,” 2021. <https://w3c.github.io/did-core/#did-document>

# NAC솔루션을 활용한 효과적인 망혼용 탐지 기법

박종현\*, 김창훈\*

\*대구대학교 컴퓨터공학

johpark74@naver.com

## Effective Multi-homed host Detection Method Using Nac Solution

Jong-hyun Park\*, Chang Hoon Kim\*

\*Department of Computer Engineering, Daegu University.

### 요약

업무 전산망분리는 금융·공공기관을 시작으로 내부에서 운영 중인 중요 정보자산을 외부의 사이버 위협으로부터 보다 완벽하게 보호하기 위하여 도입·운영 중인 네트워크 보안의 한 분야이다. 하지만 물리적 또는 논리적 방법에 의해 전산망을 분리하여 운영하는 기관에서도 PC 사용자의 부주의 또는 의도적인 망혼용 시도 등으로 전산망 분리의 근본적인 취지를 저해하는 사례가 빈번히 발생하고 있다. 그리고 대부분의 기관·학교에서는 비인가 단말기에 대한 네트워크 접근 통제를 위해 업무망과 인터넷망에 독립적으로 정보보호 솔루션을 설치하여 IP와 MAC Address 기반의 통제를 실시하고 있다. 하지만 이런 Stand Alone환경에서는 내부의 인가된 단말기의 업무망과 인터넷망을 혼용 사용에 대한 통제가 불가능하다. 본 고에서는 현재 도입·운영 중인 NAC(네트워크 접근 제어)솔루션을 사례로 전산망 분리환경에서 발생할 수 있는 망혼용 단말기에 대해서 보다 효과적인 탐지 및 대응 방안에 대해서 제안하고자 한다.

### I. 서론

우리나라 전산망 분리는 2006년 “해외발 국가 기관 해킹 실태 및 대처방안” 일환으로 국가기관 업무전산망과 인터넷 분리 방침이라는 대통령 보고 자료에서 처음 언급되기 시작했으며 2007년 국가/공공기관 업무전산망 분리 실무매뉴얼을 제작하여 방통위, 기재부 등에서 시범적으로 망 분리 사업을 추진하였다. 2008년부터 1, 2차 국가기관 망 분리 사업을 진행하였고, 2010년에는 지자체 및 산하기관으로 확대되어 2020년 기점으로 대부분의 공공기관에서 구축 완료하였다.[1]

또한 개인정보의 기술적·관리적 보호조치를 위하여 관련 법령에서는 매출액 등 일정 조건 이상의 기업에 전산망 분리를 의무화하고 있다.

전산망 분리 초기에는 대규모 예산 투입 또는 사용자 불편 등의 여러 가지 어려움으로 전산망 분리 사업이 난항을 겪기도 하였으나 10년 이 지난 현재는 법 준수 및 사용자 인식 개선 등으로 안정화 되었다.

하지만 이후 분리된 전산망을 운영함에 있어 정보시스템 서비스 연속성 및 PC사용자 불편

해소를 위한 다양한 불법적인 행위가 계속되고 있다. 본 고에서는 분리된 전산망을 운영하고 있는 기관에서 공통적으로 발생하는 유형별 망 혼용 사례를 살펴보고 효과적인 해결방안을 제시한다.

### II. 망혼용 유형 및 대응 방안

#### 1.1 MAC Clone(Spoofing) 사례 및 대응

윈도우 환경에서는 랜카드의 속성정보 변경을 통하여 MAC주소를 변경할 수 있도록 허용한다.

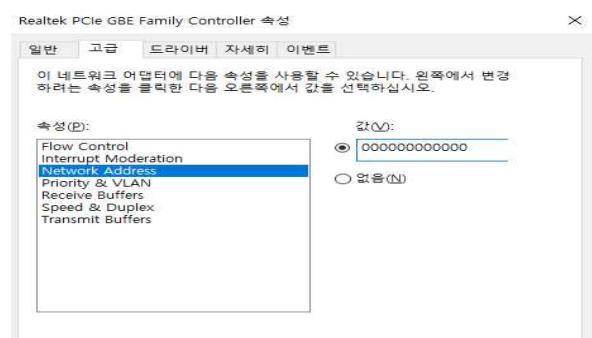


그림 1 MAC 주소 변경

PC 사용자는 임의로 MAC 주소를 변경함으로써 중요 정보를 처리하는 정보 시스템 관리자 계정에 대한 권한을 획득할 수 있다.

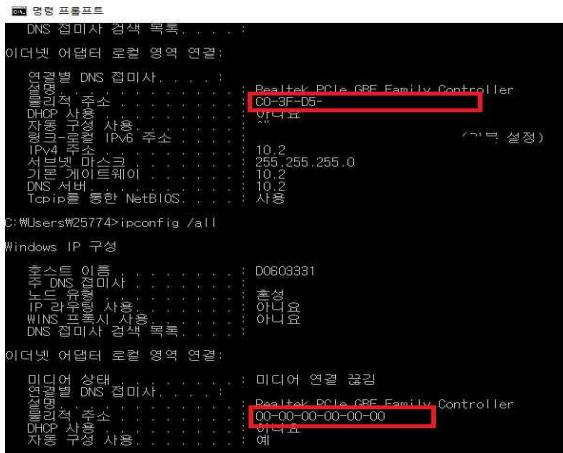


그림 2 MAC 주소 변경(전,후)

이 사례에 대한 해결책은 사전에 관리자를 포함한 PC단말기의 IP 및 MAC주소를 사전에 수집하고 데이터베이스화하여 관리함으로써 새로운 PC추가 또는 변경으로 인한 기존 네트워크 자원(IP,MAC등)에 대한 중복 사용을 차단할 수 있다. 이를 위하여 네트워크 접근제어 솔루션에서는 MAC주소에 대한 Clone을 방지하기 위한 기능을 제공하고 있다.

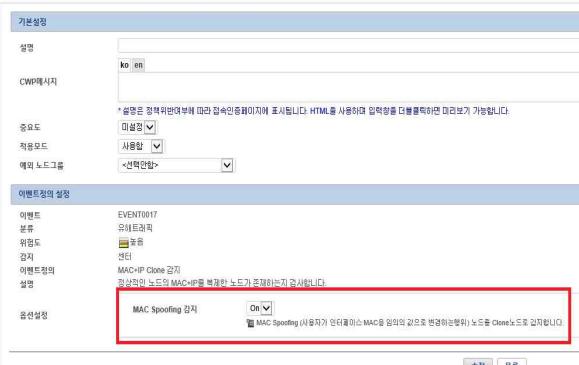


그림 3 MAC Spoofing 설정 화면

이와 함께 고려해봐야 할 문제점으로는 관리자의 랜카드를 탈취하여 공격자의 PC에 설치한 후 정보시스템에 접근을 시도하는 경우이다. 이 경우에는 앞에서 살펴본 해결안으로는 보안사고를 방지할 수 없다.

랜카드 변경에 대한 권한 우회공격을 사전에 방지하기 위해서는 PC단말기의 플랫폼 정보

즉, CPU나 메인보드 등 사전에 PC 단말기에 설치된 다양한 정보를 수집 관리함으로써 새로운 하드웨어 추가로 인한 정보 변경이 발생할 경우(ex, 메인보드 교체, HDD의 용량 증가 등)에도 보다 효과적인 대응이 가능하다. 즉 권한 우회에 대한 네트워크 서비스의 차단을 수행하게 함으로써 관리자 랜카드 획득의 경우에도 충분히 권한 획득을 제어할 수 있다.



그림 4 단말기 HW 구성 변경 감지 및 차단

### 1.2 외장 랜카드 망우회 사례 및 대응

전산망 분리 환경에서 가장 손쉽게 전산 망 우회(업무망PC에서 인터넷망 연결)를 시도할 수 있고 가장 빈번하게 발생하는 망우회 방법은 외장형태(USB, 블루투스, 테더링 등)의 무선랜카드를 연결하는 방법이다.

이 문제에 대한 가장 효과적인 대응 방법은 PC로 연결되는 모든 경로의 외장형태의 기기를 사전에 차단하는 것이다. 이를 위해 네트워크 접근제어 솔루션에서 매체 제어 기능을 활성화하여 외부 매체(USB, SD카드, CD-ROM 등) 제어를 통해 외부망 연결을 차단한다.



#### 그림 5 매체 제어 기능 설정 화면

### 1.3 정상 행위를 통한 망혼용 사례

앞에서 살펴본 사례들은 비정상적인 행위를 수행하는 패턴 분석을 통해 사전에 감지 또는 차단하는 경우를 살펴보았다. 다음은 정상적으로 네트워크 접근에 대한 승인을 획득함으로써 전산망을 우회할 수 있는 사례를 설명하고자 한다. 업무망에 접근하기 위하여 사전에 정상적인 관리적 절차를 통해 접근을 시도한 후 동일한 PC단말기를 이용하여 인터넷망 접근 신청 및 승인을 획득한 경우 업무망 및 인터넷망 연결을 동시에 사용 가능하게 된다.

이는 전산망 분리에 따라 업무망과 인터넷망에서 별도 네트워크 접근제어 솔루션을 독립적으로 운영함에 따른 부작용(Side Effect)이다.

이런 망혼용을 효과적으로 방지하기 위한 방안은 업무망 PC를 인터넷망에 연결할 경우 자동으로 차단하는 것이다. 업무망에서 운영 중인 네트워크 접근제어 솔루션의 PC단말기에 대한 정보(IP, MAC주소 등)를 인터넷망의 네트워크 접근제어 솔루션과 상호 연동·교환함으로써 해결할 수 있다.

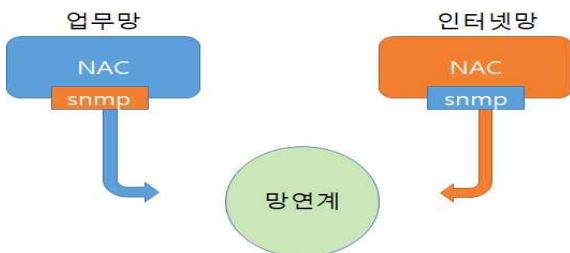


그림 6 업무망·인터넷망 SNMP 교환

위에 그림을 통해 알 수 있듯이 업무망과 인터넷망에 운영 중인 네트워크 접근제어 솔루션은 SNMP프로토콜을 이용하여 PC단말기의 정보(IP, MAC주소)를 상호 송수신함으로써 업무망·인터넷망용 네트워크 접근제어 솔루션에서 운영 중인 PC단말기 정보를 상호 비교함으로써 동일 PC단말기 망혼용 여부를 감지 및 차단한다.

### 1.4 실시간 모니터링 체계 구축

위에서 제안한 네트워크 접근제어 솔루션을 활용한 다양한 망혼용에 대한 대응 방안들은 실제 적용을 통하여 효과성이 검증되었으며, 실

제 전산망 혼용에 대한 시도를 현저히 감소시켜 주었다. 하지만 이런 대응 방안에 대한 적용과 함께 고려해야 할 중요한 사항은 전산망 혼용에 대한 위반 사항을 실시간으로 모니터링하고 주기적으로 점검할 수 있는 업무절차 마련이 필요하다. 그리고 보안담당자는 정기적인 점검 절차를 통해 비정상적인 망혼용에 대한 접근 시도를 탐지·차단하고 망혼용에 대한 위규자 교육 및 처벌 등 관리적인 요소와 앞에서 제안한 다양한 기술적인 요소들을 병행하여 구현함으로써 전산망 혼용을 방지할 수 있다.

## III. 결론

전산망 분리는 외부의 모든 사이버위협으로 원천적으로 기업의 중요자산을 보호할 수 있는 대안으로 대두되어 막대한 비용을 투자하여 금융 또는 공공분야를 필두로 일반 기업까지 확대 해나가고 있다. 하지만 전산망 분리 사업이 완료된 이후에도 안정된 전산망 분리 운영을 저해하는 여러 가지 취약요인이 생겨나고 있으며, 그 대표적인 예로써 사용자 PC단말기에서 업무망과 인터넷망을 혼용해서 사용하려는 시도가 자주 발생하고 있다. 기관이나 기업에서는 제한된 예산과 인력으로 인하여 새로운 보안 취약요소에 대해서 능동적으로 대처하기 쉽지 않은 환경에 놓여 있다. 본고에서는 예산·인력·가용자원의 제한적인 기업 환경에서 가장 효과적으로 전산망 혼용을 대응하기 위해 별도의 예산 투입 없이 현재 운영 중인 정보보호시스템 자원을 활용하여 각 기업에서 발생할 수 있는 망혼용 사례를 제시하고 적용 가능한 다양한 보안 대응 방안을 제안하였다.

## [참고문헌]

- [1] 안랩연구소, 이용진, 망 분리의 필요성 및 방식

# 논리적 블록 이동을 적용한 Fixed-Gimli Permutation

권혁동\* 엄시우\* 서화정\*†

\*한성대학교 IT융합공학부(대학원생)

\*† 한성대학교 IT융합공학부(교수)

Fixed-Gimli Permutation with logical block movement

Hyeok-Dong Kwon\* Si-Woo Eum\* Hwa-Jeong Seo\*†

\*Hansung University IT Convergence Engineering(Graduate student)

\*† Hansung University IT Convergence Engineering(Professor)

## 요약

Gimli는 NIST 경량 암호화 표준 공모전에 제출된 Permutation 알고리즘으로, 암호화와 해시 함수에 사용할 수 있는 알고리즘이다. Gimli는 384-bit의 평문을 입력받아 24라운드 동안 Permutation을 진행하며, 라운드 함수 도중에는 하나의 열을 대상으로 해당 열의 행 간 값을 교체하는 Swap 단계가 존재한다. 본 논문에서는 Swap 단계를 생략하여 물리적으로 값들의 위치를 고정하는 대신, 값의 호출 순서를 변경한 Fixed-Gimli를 제안한다. Fixed-Gimli는 Swap 연산을 생략하지만 호출 순서를 변경하는 것으로 논리적으로는 블록 이동이 진행한 것으로 설계하여 Swap에 소요되는 시간 차원을 줄인 알고리즘이다.

## I. 서론

최근 다양한 기기들이 사물인터넷 기술로 연결되어 통신을 하며 사용되어 오고 있다. 이러한 다양한 기기들이 안전한 연결과 통신을 위해 암호기술이 사용되어야 한다. 하지만 기존에 사용하는 암호기술은 데스크톱, 서버 환경용으로 설계되어 제한된 환경(낮은 CPU 성능, 작은 용량의 메모리, 낮은 전력 등)에서 작동하는 사물인터넷 기기에서는 사용이 적합하지 않다. 이러한 제한된 환경에게 사용하기 위해 HIGHT, LEA, CHAM 등 경량암호가 개발되고 있다[1].

Gimli는 ‘NIST 경량 암호화 표준화 프로세스’에서 2차 후보로 올라온 암호로 제한된 환경 뿐만 아니라 다양한 환경에서 높은 성능의 높은 보안을 달성하도록 설계된 384-bit 순열로, 암호화와 해시함수에 사용된다.

본 논문에서는 8-bit AVR 프로세서 상에서 Gimli Permutation을 최적 구현하기 위해 논리적 블록 이동을 사용한 Fixed-Gimli

Permutation을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 Gimli Permutation에 대해 설명한다. 3장에서는 논리적 블록 이동을 적용한 Fixed-Gimli Permutation을 제시한다. 4장에서는 기존의 Gimli Permutation과 Fixed-Gimli Permutation의 성능을 비교한다. 마지막으로 5장에서 본 논문에 대한 결론을 내린다.

## II. Gimli Permutation

### 2.1 라운드 함수[2]

Gimli는 입력 값을 [그림 1]과 같이 32-bit의  $3 \times 4$  형태로 간주하며 이를 S로 칭한다. Gimli는 24라운드가 진행되며, 각 라운드 함수는 다음 세 가지의 계층의 조합으로 구성된다.

첫 번째는 SP-box를 통과하는 비선형 계층으로 각각의 열이 혼합되는 연산을 진행한다. 두 번째는 선형 혼합 계층으로, Small Swap과 Big Swap으로 나뉜다. 세 번째는 라운드 상수

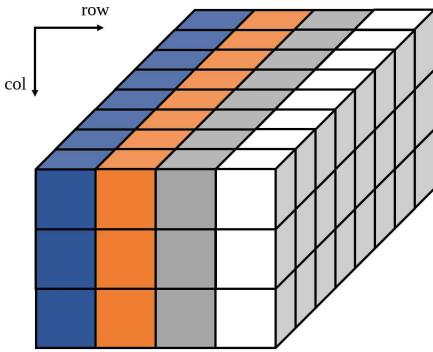


Fig. 1. State representation

추가 계층이다.

## 2.2 비선형 계층

비선형 계층의 SP-box는 세 가지 단계의 작업을 거친다. 첫 번째는  $S_0$ 열과  $S_1$ 열의 블록에 각각 24-bit, 9-bit rotation 연산을 취해준다. 두 번째는  $S_0$ ,  $S_1$ ,  $S_2$ 열 간에 혼합 연산을 취해준다. 마지막에서는  $S_0$ 와  $S_2$ 열의 값을 교체한다.

## 2.3 선형 계층

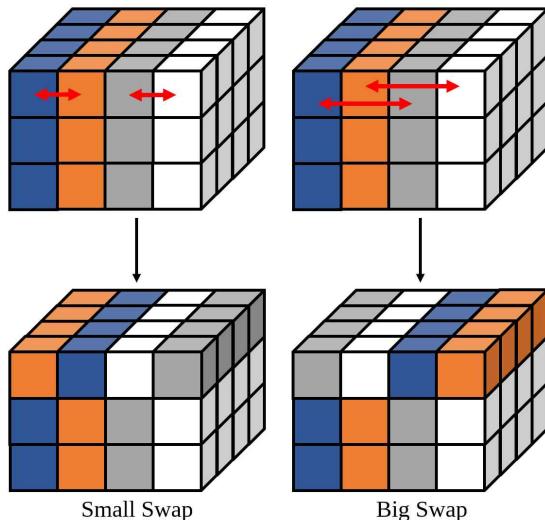


Fig. 2. Left: Small Swap operation

Right: Big Swap operation

선형 계층은 [그림 2]와 같이 행 간 값 교체가 이루어지며 두 종류로 나뉜다. Small Swap은 1라운드부터 4라운드마다, Big Swap은 2라운드부터 4라운드마다 진행한다. 값 교환은 첫 번째 열에서만 발생한다.

## 2.4 라운드 상수 추가 계층

4의 배수 라운드마다 현재 라운드  $r$ 과 라운드 상수 값  $0x9E377900$ 을 XOR 한다. 이렇게 획득한 값을  $S_{0,0}$ 과 XOR하여 라운드 상수를 반영한다.

라운드 함수의 전체 구조를 의사코드 형태로 나타내면 [그림 3]과 같다.

---

### Algorithm

**Input:** state =  $(state_{col, row}) \in W^{3 \times 4}$   
**Output:** Gimli(state) =  $(state_{col, row}) \in W^{3 \times 4}$

---

```

for round = 24 to 1
  for row = 0 to 3                                // SP-box
    X ← ROL24(state0, row)
    Y ← ROL9(state1, row)
    Z ← state2, row
    state2, row ← X ⊕ LSL1(Z) ⊕ LSL2(Y & Z)
    state1, row ← Y ⊕ X ⊕ LSL1(X | Z)
    state0, row ← Z ⊕ Y ⊕ LSL3(X & Y)
  end for
  if round mod 4 == 0                           // Small Swap
    state0,0 state0,1 state0,2 state0,3 ← state0,1 state0,0 state0,3 state0,2
  else if round mod 4 == 2                      // Big Swap
    state0,0 state0,1 state0,2 state0,3 ← state0,2 state0,3 state0,0 state0,1
  end if
  if round mod 4 == 0                           // Add Constant
    state0,0 = state0,0 ⊕ 0x9e377900 ⊕ round
  end if
end for
Return(statecol, row)

```

---

Fig. 3. Pseudocode for Gimli permutation

## III. 제안 기법

본 논문에서는 Small swap과 Big swap을 생략한 기법을 제안한다. 선형 계층의 Swap 연산은 첫 번째 열에 대해서만 진행된다. 따라서 Swap에 소요되는 시간만큼 연산 시간을 단축 시킬 수 있다.

하지만 Swap이 생략되기 때문에 각 라운드 별로 특정 행의 값을 로드해야 한다. 이 사안은 비선형 계층과 라운드 상수 추가 계층에 해당된다.

Swap을 진행한 경우, [그림 4]와 같이  $S_{0,0}$ ,  $S_{0,1}$ ,  $S_{0,2}$ ,  $S_{0,3}$ 의 값을  $X_{0 \sim 4}$ 에 로드할 때,  $S_{0,0}$ ,  $S_{0,1}$ ,  $S_{0,2}$ ,  $S_{0,3}$  순서로 값을 호출하는 것을 알 수 있다. 이는 일반적인 Gimli Permutation의 형태이다. 선형 계층을 통과하면서 첫 번째 열 값에 Swap이 적용되기 때문에 특별한 호출 구조가 필요 없다.

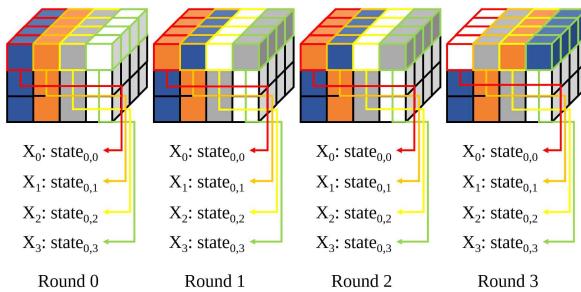


Fig. 4. Original Gimli permutation flow

하지만 Swap을 생략하게 될 경우, 첫 번째 열 호출을 다르게 구성해야 한다. Swap이 생략된다면 행 간에 값 이동이 발생하지 않는다. 하지만 각 라운드에는 값 이동이 발생한 값이 입력되어야 하므로 값을 로드하는 순서가 바뀌게 된다. [그림 5]는 이를 묘사한 그림으로, 실제로 행 간 값의 이동은 발생하지 않지만 입력되는 값은 Swap 연산이 적용된 값과 동일한 값이 입력됨을 알 수 있다.

즉, 물리적으로 Swap이 발생하지는 않기에 각 값들이 저장된 위치는 고정되지만, 논리적으로 Swap 연산을 진행하는 것으로 취급이 가능하다.

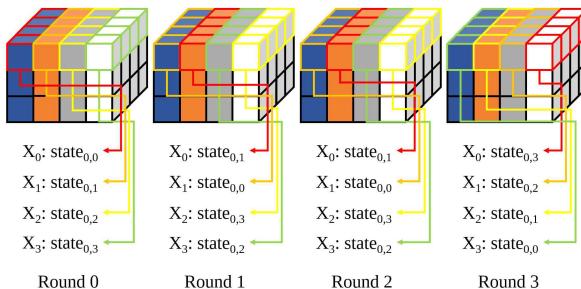


Fig. 5. Fixed-Gimli permutation flow

#### IV. 성능 비교

제안하는 기법인 Fixed-Gimli와 기존 Gimli의 성능 비교를 진행한다. 구현은 ATmega128 프로세서를 대상으로 구현하였으며, Microchip Studio를 사용하여 구현하였다. 구현 결과는 [표 1]에서 확인이 가능하다. 기존 구현물은 Tiny와 Fast 버전으로 나뉘며 각각 413cpb, 213cpb의 동작 속도를 지닌다. 제안 기법을 구현하기 위해서 우선 기본 구현물을 작성하고, 이를 토대로 제안 기법을 적용한 구현물을 작성한다.

기본 구현물은 266cpb로 구현되었고, 해당 구현물에 제안 기법을 적용한 Fixed-Gimli는 248cpb의 성능을 지닌다. 제안하는 기법은 기존 Tiny 기법에 비해 약 40% 향상된 속도를 지니나, Fast 구현물에 비해 느린 성능을 보인다.

이는 기존 구현물이 AVR Assembly만을 사용한 것이 아닌, Python 코드를 사용하여 Swap을 진행하기 때문에 이와 관련한 부분에서 동작 속도의 이득이 발생한다. 따라서 이와 완전한 비교는 다소 어렵다. 기존 Fast 구현물을 토대로 AVR Assembly만을 사용한 기본 구현물과 비교한다면 제안하는 기법이 성능이 우수한 것을 확인할 수 있다.

Table. 1. Comparison result table (Unit: cpb)

[2] Tiny	[2] Fast	Proposed Normal	<b>Proposed Fixed</b>
413	213	266	<b>248</b>

#### V. 결론

본 논문에서는 Gimli Permutation의 Swap 구조를 변형하여, 실제적인 값의 이동은 발생하지 않는 Fixed-Gimli Permutation을 제안하였다. 제안 기법은 Swap 연산이 생략된 만큼 성능 향상을 확인할 수 있었다. 이후로 AVR 환경의 특성을 더욱 활용하여 추가적인 개선을 진행하는 것을 후속 연구로 제시한다.

#### VI. Acknowledgment

이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478).

#### [참고문헌]

- [1] Kwon H.D. et al. “Designing a CHAM Block Cipher on Low-End Microcontrollers for Internet of Things,” *Electronics*, Sep, 2020.
- [2] Bernstein D.J. et al., “Gimli : A Cross-Platform Permutation,” *Cryptographic Hardware and Embedded Systems*, Aug, 2017.

# 소프트웨어 불법복제 및 역공학 취약점 분석: A 소프트웨어를 기반으로

이재혁\*, 이경률\*

\*대구가톨릭대학교 컴퓨터소프트웨어학부

## Vulnerability Analysis of Software Piracy and Reverse Engineering: Based on A Software

JaeHyuk Lee\* Kyungroul Lee\*

\*School of Computer Software, Daegu Catholic University

### 요약

정보통신 기술의 발전으로 인하여, 다양한 소프트웨어들이 개발되고 있으며, 소프트웨어 시장 및 산업이 활성화되었다. 시장 및 산업이 활성화됨에 따라, 소프트웨어의 소스코드와 같은 저작권을 보호하기 위한 기술들이 등장하였다. 그럼에도 불구하고, 소프트웨어 시장 및 산업에 큰 결림들이 되는 소프트웨어 불법복제가 지속해서 시도되는 실정이다. 따라서 소프트웨어 저작권 보호 기술의 안전성 향상을 위한 방안이 요구됨에 따라, 불법복제 소프트웨어의 사용을 탐지하고 방지하기 위하여, A 소프트웨어를 중심으로 소프트웨어 저작권 보호 기술이 가지는 취약점과 보안 위협을 분석하고 실증한다.

**키워드:** 소프트웨어 저작권 보호 기술, 소프트웨어 불법복제, 역공학, 라이선스 인증

## I. 서론

IT (Information Technology) 기술이 비약적으로 발전함에 따라, 패키지 소프트웨어, IT 서비스, 게임 소프트웨어와 같이 사용자 개인의 목적에 초점을 맞춘 소프트웨어가 많이 개발되는 실정이다 [1].

소프트웨어 시장 및 산업이 활성화됨에 따라, 상용 소프트웨어 (Commercial Software), 프리웨어 (Freeware), 셰어웨어 (Shareware)를 비롯한 다양한 소프트웨어 종류가 등장하였다 [2]. 그중, 소프트웨어 저작권 보호를 위하여, 프리웨어를 제외한 대부분 소프트웨어는 소프트웨어 저작권자에게 정당한 비용을 지급함으로써 사용하며, 이러한 기술이 소프트웨어 저작권 보호 기술이다 [4].

소프트웨어 저작권 보호 기술이 적용되었음에도 불구하고, 소프트웨어 역공학을 통하여, 정품 미보유, 라이선스 위반, 기간 초과 사용과

같은 소프트웨어 불법복제 및 불법 사용이 지속해서 시도되는 현실이다 [3]. 이러한 불법복제 소프트웨어의 사용은 소프트웨어 개발자의 개발 의욕을 감소시킬 뿐만 아니라, 전반적인 소프트웨어 시장 및 산업에 큰 결림들이 된다 [4]. 따라서 불법복제 소프트웨어의 사용을 탐지하고 방지하기 위하여, 소프트웨어 저작권 보호 기술이 가지는 취약점 및 보안 위협을 분석할 필요성이 있으며, 그 결과를 기반으로 소프트웨어 저작권 보호 기술의 안전성 향상 방안이 요구된다.

이러한 요구사항을 만족하기 위하여, 본 논문에서는 소프트웨어 저작권 보호 기술 중 라이선스 인증이 적용된 A 소프트웨어를 기반으로, 라이선스 인증 방식을 분석하고 불법복제 취약점을 도출한다.

논문의 구성은 다음과 같다. 2장에서는 A 소프트웨어의 라이선스 인증 취약점을 분석하기

위하여 인증과정을 분석하며, 3장에서는 분석한 결과를 기반으로 신규 발굴한 A 소프트웨어 취약점을 서술한다. 마지막으로 결론 및 향후 계획을 4장에 서술한다.

## II. A 소프트웨어의 소프트웨어 저작권 보호 기술 분석

일반적으로 소프트웨어 저작권 보호를 위한 기술은 라이선스 키를 기반으로 정품을 인증하는 라이선스 인증기술 [5], 소프트웨어에 특정 정보를 삽입하여 무결성을 탐지하는 워터마킹 기술 [6], 마지막으로 평문을 해독 불가능한 형태로 변형하는 난독화 및 암호화 기술이 있다 [4, 7]. 본 논문에서의 분석 대상인 A 소프트웨어는 라이선스 인증기술을 사용하여 소프트웨어 저작권을 보호한다.

A 소프트웨어의 라이선스 인증기술 취약점을 분석하기 위하여, 라이선스 인증과정을 분석하였으며, 이를 그림 1에 나타내었다. A 소프트웨어의 라이선스 인증 방식은 크게 소프트웨어 사용 단계, 라이선스 인증 단계, 라이선스 인증 결과 단계인 총 3단계로 구성된다.

첫 번째 단계인 소프트웨어 사용 단계는, 사용자가 소프트웨어를 설치하고, 실제로 사용하는 단계이다. 이 단계에서 라이선스가 인증되지 않으면, 특정 기능이나 사용 기간이 만료되어 소프트웨어를 정상적으로 사용하지 못한다.

두 번째 단계인 라이선스 인증 단계는 특정 기능이나 사용 기간의 제한을 해제하기 위하여 라이선스 키를 인증하는 단계이다. 라이선스 키를 인증하기 위하여 사용자는 정당한 비용을 지불하여 라이선스 키를 발급받으며, 발급된 키를 인증함으로써 제한 기능 및 기간 만료 없이 정상적으로 프로그램을 사용한다.

A 소프트웨어의 제한 기능 및 기간은 표 1과 같으며, 제한 기간은 없고 특정 기능만 제한한다. 특정 기능으로는 첫 세션을 기준으로 200회의 명령을 사용할 수 있으며, 그 이후부터는 20 회의 명령마다 라이선스 구매 유도 메시지를 출력한다. A 소프트웨어의 최대 명령 횟수 제한은 2,500회이며, 그 이후부터는 1회의 명령마다 라이선스 구매 유도 메시지가 출력된다. 따라서 2,500회를 초과하는 경우에는 소프트웨어를 구매하지 않고서는 정상적으로 소프트웨어를 사용하기 어렵다.

마지막 단계인 라이선스 인증 결과 단계는 사용자가 입력한 라이선스 키를 기반으로 인증 결과를 출력하며, 인증에 성공한 경우에는 최대 명령 횟수 제한이 없는 정식 버전을 사용할 수 있고, 인증에 실패한 경우에는 계속 평가 버전을 이용한다.

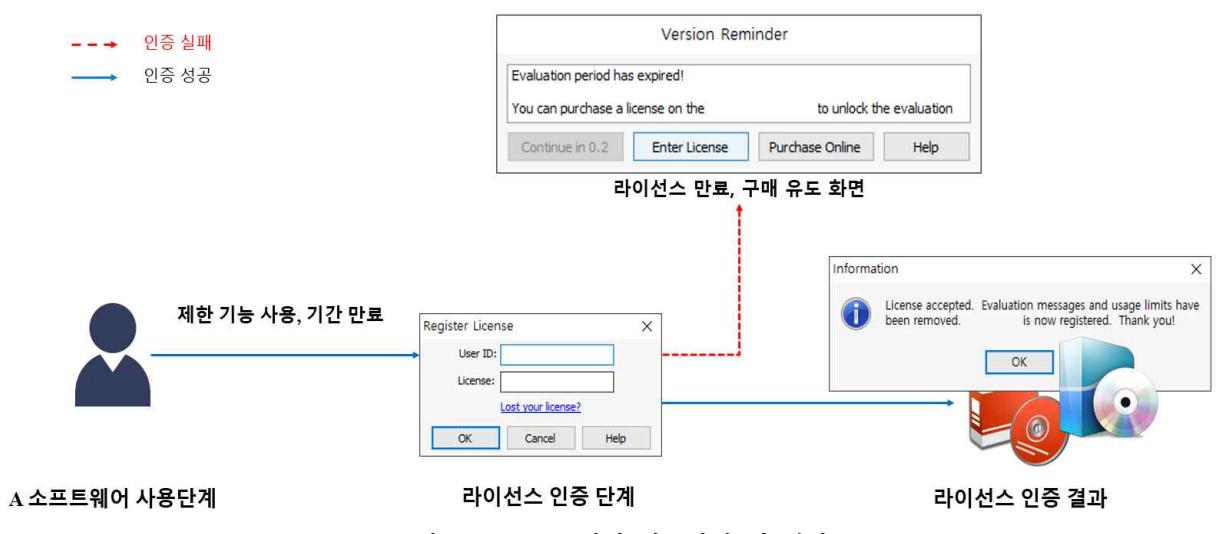


그림 1 A 소프트웨어 인증과정 및 결과

표 1. A 소프트웨어 기능 및 기간 제한

제한 항목	설명
기능 제한	최대 입력 가능 횟수 제한 (총 2,500회 제한, 첫 세션 당 200회, 이후 20회)
기간 제한	없음

### III. A 소프트웨어 불법복제 취약점 분석

#### 3.1. 라이선스 인증 취약점

A 소프트웨어에 적용된 소프트웨어 저작권 보호 기술은 라이선스 인증기술이며, 이 기술은 소프트웨어 내부에 유효한 라이선스 키와 관련된 정보가 노출되는 근본적인 문제점이 존재한다. 즉, 사용자가 입력한 키를 검증하기 위해서는 유효한 라이선스 키가 내부에 존재하여야만 하며, 해당 키들을 비교할 수밖에 없는 구조로 인하여 유효한 라이선스 키가 노출된다.

이러한 문제점으로 인하여, 불법복제가 가능한 취약점 및 보안 위협이 발생한다. 따라서 라이선스 키를 비교하는 코드를 분석한다면, 유효한 라이선스 키를 탈취할 수 있으며, 탈취된 키를 기반으로 소프트웨어를 구매하지 않고도 정당한 사용자로 인증하도록 우회하는 것이 가능하다.

A 소프트웨어의 라이선스 인증 방식 역시 이러한 구조적인 문제점으로 인하여 소프트웨어 저작권 보호 기술의 취약점이 발생하고, 불법적인 사용이 가능하다. 본 논문에서는 A 소프트웨어를 대상으로 라이선스 인증기술의 취약점을 분석한다.

라이선스 키를 비교하는 코드를 분석한 결과, 입력한 라이선스 키를 인자로 전달되는 함수를 확인하였고, 이를 그림 2에 나타내었다. 입력한 User ID는 1234이고, License는 QWER이다. 이러한 구조는 해당 함수가 User ID와 License를 비교하기 위한 함수일 가능성성이 매우 높다.

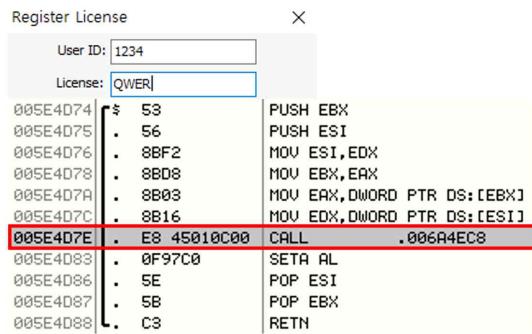


그림 2. 분석한 라이선스 키 비교 함수

해당 함수가 라이선스 키를 비교하는 함수임을 검증하기 위하여, 해당 함수가 호출될 때의 인자를 확인하였으며, 입력한 User ID와 License 정보가 그림 3과 같이 EAX에 저장되는 것을 확인하였다.

(1)	<table border="1"> <tr> <td>006A44E00</td><td>&gt; 0FB608</td><td>MOVZX ECX, BYTE PTR DS:[EAX]</td></tr> <tr> <td>006A44E03</td><td>. 2A0A</td><td>SUB CL, BYTE PTR DS:[EDX]</td></tr> <tr> <td>006A44E05</td><td>v 75 25</td><td>JNZ SHORT 006A44EFC</td></tr> <tr> <td>006A44E07</td><td>. 53</td><td>PUSH EBX</td></tr> </table> <table border="1"> <tr> <td>Registers (FPU)</td><td>Registers (FPU)</td></tr> <tr> <td>ERX 01580004 ASCII "1234"</td><td>ERX 02F0FFB4 ASCII "1234"</td></tr> <tr> <td>ECX 00000001</td><td>ECX 00000001</td></tr> <tr> <td>EDX 0326683C ASCII "GN8P95AK9"</td><td>EDX 02F55F1C ASCII "ZUGREDDB94"</td></tr> </table>	006A44E00	> 0FB608	MOVZX ECX, BYTE PTR DS:[EAX]	006A44E03	. 2A0A	SUB CL, BYTE PTR DS:[EDX]	006A44E05	v 75 25	JNZ SHORT 006A44EFC	006A44E07	. 53	PUSH EBX	Registers (FPU)	Registers (FPU)	ERX 01580004 ASCII "1234"	ERX 02F0FFB4 ASCII "1234"	ECX 00000001	ECX 00000001	EDX 0326683C ASCII "GN8P95AK9"	EDX 02F55F1C ASCII "ZUGREDDB94"	(4)
006A44E00	> 0FB608	MOVZX ECX, BYTE PTR DS:[EAX]																				
006A44E03	. 2A0A	SUB CL, BYTE PTR DS:[EDX]																				
006A44E05	v 75 25	JNZ SHORT 006A44EFC																				
006A44E07	. 53	PUSH EBX																				
Registers (FPU)	Registers (FPU)																					
ERX 01580004 ASCII "1234"	ERX 02F0FFB4 ASCII "1234"																					
ECX 00000001	ECX 00000001																					
EDX 0326683C ASCII "GN8P95AK9"	EDX 02F55F1C ASCII "ZUGREDDB94"																					
(2)	<table border="1"> <tr> <td>Registers (FPU)</td><td>Registers (FPU)</td></tr> <tr> <td>ERX 02F1005C ASCII "1234"</td><td>ERX 02F1002C ASCII "QWER"</td></tr> <tr> <td>ECX 00000001</td><td>ECX 00000001</td></tr> <tr> <td>EDX 0324640DC ASCII "57ZXPABRL"</td><td>EDX 02F5313C ASCII "V1WLOZPBF"</td></tr> </table>	Registers (FPU)	Registers (FPU)	ERX 02F1005C ASCII "1234"	ERX 02F1002C ASCII "QWER"	ECX 00000001	ECX 00000001	EDX 0324640DC ASCII "57ZXPABRL"	EDX 02F5313C ASCII "V1WLOZPBF"	(5)												
Registers (FPU)	Registers (FPU)																					
ERX 02F1005C ASCII "1234"	ERX 02F1002C ASCII "QWER"																					
ECX 00000001	ECX 00000001																					
EDX 0324640DC ASCII "57ZXPABRL"	EDX 02F5313C ASCII "V1WLOZPBF"																					
(3)	<table border="1"> <tr> <td>Registers (FPU)</td><td>Registers (FPU)</td></tr> <tr> <td>ERX 02F0FFB4 ASCII "1234"</td><td>ERX 02F1002C ASCII "QWER"</td></tr> <tr> <td>ECX 00000001</td><td>ECX 00000001</td></tr> <tr> <td>EDX 02F55F1C ASCII "W4YPP4E7T"</td><td>EDX 02F5313C ASCII "V1WLOZPBF"</td></tr> </table>	Registers (FPU)	Registers (FPU)	ERX 02F0FFB4 ASCII "1234"	ERX 02F1002C ASCII "QWER"	ECX 00000001	ECX 00000001	EDX 02F55F1C ASCII "W4YPP4E7T"	EDX 02F5313C ASCII "V1WLOZPBF"													
Registers (FPU)	Registers (FPU)																					
ERX 02F0FFB4 ASCII "1234"	ERX 02F1002C ASCII "QWER"																					
ECX 00000001	ECX 00000001																					
EDX 02F55F1C ASCII "W4YPP4E7T"	EDX 02F5313C ASCII "V1WLOZPBF"																					

그림 3. 분석한 하드코딩된 유효한 User ID 및 License 정보

상기 인증정보를 비교하는 함수는 총 5번 호출되고, 첫 번째 호출부터 네 번째 호출까지는 EAX에 User ID가 저장되었으며, 마지막 다섯 번째 호출에서는 EAX에 License가 저장되었다. 같은 호출과정에서 EDX에는 임의의 9자리 문자열이 저장되는 것을 확인하였다. 따라서 이 함수를 인증정보를 비교하는 함수라고 가정한다면, EDX에 저장된 정보는 유효한 User ID와 License라고 판단할 수 있다.

이러한 가정을 바탕으로, 노출된 유효한 User ID와 License를 검증하기 위하여, 다섯 번째 호출에서 EDX에 저장된 문자열을 그림 4와 같이 입력한 결과, 정상적으로 라이선스가 인증되었으며, 제한 기능이 제거된 것을 확인하였다.

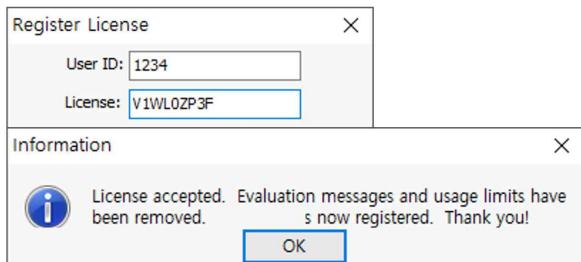


그림 4. 라이선스 인증 우회 결과

본 논문에서 분석한 결과를 통하여, A 소프트웨어는 소프트웨어 내부에 라이선스 인증정보를 하드코딩한다는 것을 실증하였으며, 노출된 라이선스 인증정보를 통하여 인증을 우회하는 것을 검증하였다.

하지만, 이 정보를 활용하는 경우, 라이선스가 정상적으로 등록되었다는 메시지가 출력되지만, 실제로 제한된 기능이 제거되지는 않는다. 다시 말하면, 상기 취약점을 활용하여 라이선스 인증을 우회하더라도, 최대 명령 횟수 제한인 2,500회를 초과하면, 평가 기간이 만료되었다는 메시지가 출력되며, 기능 사용에 제약이 따른다.

### 3.2. 기능 제한 취약점

상기 라이선스 인증을 우회하더라도 제한이 제거되지 않은 기능을 무력화하기 위하여, 기능 제한 취약점을 분석하였다. A 소프트웨어는 기능 제한을 위하여 라이선스 구매 유도 메시지를 반복적으로 출력한다. 이는 명령 횟수를 계산하고, 기준 횟수를 초과할 경우, 메시지를 출력하는 것으로 판단된다. 따라서 기준 횟수를 초과할 경우, 메시지를 출력하기 위하여 호출되는 함수가 존재할 것으로 가정하고, 해당 함수를 분석하였다.

분석 결과, 그림 5와 같이 0x004EBCA7 위치에서 특정 함수를 호출하는 것을 확인하였고, 해당 함수에서 라이선스 구매 유도 메시지가 출력되었다. 자세하게는, 해당 함수 내부에서 출력되는 라이선스 구매 유도 메시지에 포함되는 문자열인 “Evaluation period has expired!”와 “Session limit exceeded!”가 노출된 것을 확인하였다. 즉, 해당 함수는 평가 기

```

CPU - main thread, module
004EBCA7 FF92 28010001 CALL DWORD PTR DS:[EDX+128]
004EBCAD . 66:C745 A8 31 MOV WORD PTR SS:[EBP-58],30
004EBCB3 . 83F8 01 CMP EAX,1
004EBCB6 . 75 0B JNZ SHORT .004EBCC3
004EBCB8 . 33D2 XOR EDX,EDX
004EBCBA . 8BC3 MOV EAX,EBX

ASCII "Your
license has expired.

ASCII "Evaluation period has expired! "

```

그림 5. 분석한 라이선스 구매 유도 메시지 출력 함수

간이 만료되었을 때, 라이선스 구매 유도를 위한 함수인 것으로 가정하였다.

따라서 본 논문에서는 해당 함수가 라이선스 구매 유도 메시지를 출력하는 함수이며, 기능을 제한하는 코드를 실행하는 함수로 가정하여, 해당 함수가 호출되지 않도록 수정하였다. 그 결과, 그림 6과 같이 최대 입력 가능한 명령 횟수를 초과하더라도, 라이선스 구매 유도 메시지가 출력되지 않고 정상적으로 사용하는 결과를 실증하였다.



그림 6. 기능 제한 우회 결과

본 논문에서 분석한 A 소프트웨어의 소프트웨어 저작권 보호 기술 취약점을 통하여, 소프트웨어 내부에 하드코딩된 라이선스 정보가 존재한다는 것뿐만 아니라, 특정 코드를 수정함으로써 기능 제한을 우회하는 취약점을 도출하였다. 도출된 취약점을 통하여 소프트웨어 불법복제와 같은 불법적인 사용이 가능한 것으로 판단된다.

## IV. 결론

본 논문에서는 소프트웨어 저작권 보호 기술의 안전성 향상을 위하여, 소프트웨어 저작권 보호 기술 중 라이선스 인증이 적용된 A 소프

트웨어를 기반으로, 라이선스 인증 방식 및 과정을 분석하고 불법복제 취약점을 도출하였다. 도출한 취약점으로는 라이선스 인증 취약점과 기능 제한 취약점이 있으며, 실질적인 취약점인 기능 제한 취약점을 실증함으로써 신규 취약점을 발굴하였다. 라이선스 인증 취약점은 라이선스 키가 하드코딩되어 코드상에 그대로 노출되는 취약점이지만, 실질적으로 라이선스 인증이 우회되지는 않고 기능이 제한되었다. 이에 특정 코드를 수정함으로써 기능 제한을 우회하여 소프트웨어를 구매한 사용자와 동일하게 제한되었던 기능의 사용이 가능한 취약점을 검증하였다.

향후 연구로는, 본 논문에서 발굴한 취약점의 근본적인 문제점 해결을 위하여, 하드코딩된 라이선스 키를 노출하지 않는 방안 및 개인화된 라이선스 인증 방안을 연구할 계획이다.

회지, 제11권, 제2호, pp. 23-32, 2013년 12월.

- [5] 이상렬, “인터넷을 통한 소프트웨어 불법사용 방지시스템 설계,” 한국컴퓨터정보학회 논문지, 제6권, 제4호, pp. 110-118, 2001년 12월.
- [6] Y. Wang, D. Gong, B. Lu, F. Xiang, and F. Liu, “Exception Handling-Based Dynamic Software Watermarking,” IEEE ACCESS, vol. 6, pp. 8882-8889, Feb. 2018.
- [7] J. Katz and Y. Lindell, “Introduction to Modern Cryptography,” 3rd Ed., CRC press, Dec. 2020.

## 감사의 글

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음”(2019-0-01056)

## [참고문헌]

- [1] 한국정보통신진흥협회, “2019 ICT 실태조사 보고서,” [https://www.kait.or.kr/notice/stat\\_board\\_view.jsp](https://www.kait.or.kr/notice/stat_board_view.jsp), 2020년 9월 10일 등록, 2020년 2월 2일 접속.
- [2] 이진천, “소프트웨어 라이센스 종류,” 대한설비공학회 설비저널, 제43권, 제6호, pp. 102-103, 2014년 6월.
- [3] 소프트웨어정책연구소, “국가별 SW 불법복제율,” [https://stat.sppri.kr/posts/view/22306?code=stat\\_sw\\_illegal\\_copy](https://stat.sppri.kr/posts/view/22306?code=stat_sw_illegal_copy), 2019년 6월 28일 등록, 2020년 2월 2일 접속.
- [4] 조성제, 김동진, 박민규, “소프트웨어 저작권 보호 기술 동향,” 한국정보기술학회 학

# 소프트웨어 지적 재산권 보호 기술의 역공학 취약점 분석: B 소프트웨어를 기반으로

정원태\*, 이경률\*

\*대구가톨릭대학교 컴퓨터소프트웨어학부

Vulnerability Analysis of Intellectual Property Rights using Reverse Engineering: Based on B Software

Wontae Jung\*, Kyungroul Lee\*

\*School of Computer Software, Daegu Catholic University

## 요약

소프트웨어의 불법적인 사용을 예방하기 위하여, 다양한 소프트웨어 지적 재산권 보호 기술이 등장하였지만, 역난독화 및 역공학과 같은 기술을 악용함으로써 소프트웨어의 지적 재산권을 보호받지 못하는 한계점이 존재한다. 이러한 한계점이 가지는 근본적인 원인을 규명하기 위하여, 소프트웨어의 불법적인 사용을 방지하는 지적 재산권 보호 기술 중 라이선스 인증 기술이 적용된 B 소프트웨어를 중점으로, 발생 가능한 취약점을 분석한다. 분석 결과, 단순히 인증 결과를 조작함으로써 라이선스 인증을 우회하는 취약점은 파일 개수를 제한하는 기능을 사용하지 못함으로써 실질적인 취약점으로 판단할 수 없지만, 분석한 파일 개수와 비교 명령어를 기반으로 파일 개수를 조작한 결과, 라이선스를 인증하지 않더라도 B 소프트웨어가 제한하는 기능을 우회하는 취약점을 실증하였다. 본 논문에서 규명한 라이선스 인증 기술이 가지는 근본적인 취약점을 제거한다면, 더욱 효과적으로 소프트웨어의 지적 재산권을 보호할 수 있을 것으로 판단된다.

**키워드:** 소프트웨어 저작권 보호 기술, 취약점 분석, 역공학, 라이선스 인증

## I. 서론

사용자가 컴퓨터 하드웨어를 더욱 편리하게 사용하기 위하여, 프리웨어, 애드웨어, 세어웨어와 같은 다양한 종류의 소프트웨어가 등장하였다 [9]. 그중, 세어웨어는 사용자가 정당한 비용을 지불하여야만 소프트웨어가 제공하는 모든 기능을 사용할 수 있다. 하지만 많은 사용자들은 비용을 지불하지 않고 소프트웨어를 사용하기를 원하였으며, 이로 인하여 소프트웨어 불법 복제가 등장하였고 지금까지 지속적으로 악용되는 실정이다.

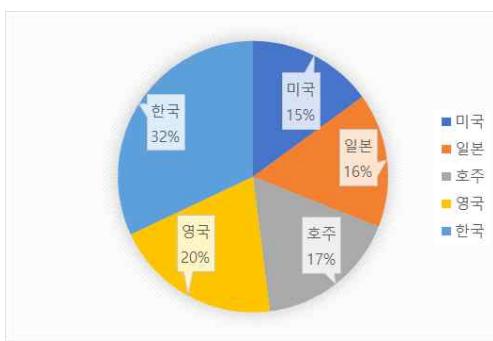
<그림 1>과 같이 국내의 소프트웨어 불법 복제 비율은 32%이며, 세계적으로는 평균 약 37%로 높은 수치를 가진다. 이와 같은 불법 복제로 인하여, 소프트웨어 업계들은 약 7,200억 원의 경제적 손실이 발생하는 심각한 문제점이 발생한다고 보고되었다 [1, 2].

소프트웨어의 불법적인 사용을 예방하기 위하여, 라이선스 인증, 소스코드 난독화, 그리고 역공학 방지, 템퍼링 방지, Anti-Piracy 도구들을 활용함으로써 소프트웨어의 지적 재산권을 보호하기 시작하였다.

그럼에도 불구하고, 역난독화 및 역공학과 같은 기술을 악용함으로써 소프트웨어의 지적 재산권을 보호받지 못하는 한계점이 존재한다. 따라서 이러한 한계점이 가지는 근본적인 원인을 규명하기 위하여, 현재 적용된 기술이 가지는 취약점 분석하는 연구가 요구되며, 분석 결과를 기반으로 더욱 효과적으로 소프트웨어의 지적 재산권을 보호할 수 있을 것으로 판단된다. 이러한 요구를 만족하기 위하여, 본 논문에서는 소프트웨어의 불법적인 사용을 방지하는 지적 재산권 보호 기술 중 라이선스 인증 기술이 적용된 B 소프트웨어를 중점으로, 해당 기술

의 취약점을 분석하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 소프트웨어 지적 재산권 보호 기술을 소개하고, 3장에서는 B 소프트웨어의 지적 재산권 보호 기술 및 취약점 분석 결과를 서술하며, 결론 및 향후 연구를 4장에 나타내었다.



<그림 1> 2017년 소프트웨어 불법 복제 비율

## II. 관련 연구

소프트웨어 지적 재산권 보호 기술은 라이선스 인증, 소스코드 난독화, 불법 복제 방지 기술이 있다. 라이선스 인증 기술은 저작권자로부터 소프트웨어의 사용을 허락받는 것으로, 인증을 위하여 문자열 형태의 유일한 라이선스 키를 활용한다 [3]. 소프트웨어 난독화 기술은 소스코드의 추가 및 수정을 통하여 분석가로 하여금 분석을 방해하거나 시간이 많이 소요되도록 가독성을 낮추는 기술이다 [4, 5]. 불법 복제

방지 기술은 역공학 방지, 템퍼링 방지, Anti-Piracy를 지원하는 도구들을 사용하여 프로그램의 분석 및 복제를 방지하는 기술이다 [7].

상기와 같이 소프트웨어의 지적 재산권을 보호하기 위한 다양한 기술이 존재하며, 본 논문의 분석 대상인 B 소프트웨어는 라이선스 인증 기술을 사용한다. 이 기술에서 인증과 관련된 핵심 정보는 문자열 형태의 유일한 라이선스 키이며, 키의 노출 및 탈취, 인증 우회와 같은 취약점을 집중적으로 분석함으로써 지적 재산권을 보호받지 못하는 근본적인 원인을 분석하고자 한다.

## III. 분석 결과

### 3.1. B 소프트웨어 지적 재산권 보호 기술 분석

B 소프트웨어는 라이선스 인증 기술을 기반으로 지적 재산권을 보호하며, <그림 2>와 같이 라이선스 인증 과정을 통하여 소프트웨어의 사용을 허가받는다. 라이선스를 인증받지 못하면, 30개를 초과하는 파일을 편집하지 못하며, 라이선스 인증 요구 메시지가 지속적으로 출력된다. 즉, B 소프트웨어는 라이선스 인증을 통하여 30개를 초과하는 파일을 편집할 수 있도록 사용을 허가받는다. 따라서 라이선스를 구매하여 올바른 라이선스 키를 입력한 사용자만 B 소프트웨어가 제공하는 모든 기능을 사용할 권한을 부여받는다.



<그림 2> B 소프트웨어 라이선스 인증과정

하지만, 이처럼 문자열을 비교하는 인증은 구조적으로 올바른 라이선스 키와 입력한 라이선스 키를 비교할 수밖에 없으며, 비교 결과를 기반으로, 인증 성공 및 실패 메시지를 출력한다. 이 과정에서 인증 결과를 조작한다면, 인증에 실패하더라도 인증에 성공한 사용자로 위장이 가능한 취약점이 존재한다.

### 3.2. 취약점 분석 결과

상기 가정한 취약점을 기반으로 B 소프트웨어의 라이선스 인증 우회를 시도하였다. B 소프트웨어는 라이선스 인증을 위하여, <그림 2>와 같이 사용자의 이름과 라이선스 키에 해당하는 일련번호를 입력받는다. 입력된 일련번호는 0x00501756번지의 call에서 비교하며, 인증 결과에 따라, 0x0050175F번지에서 코드의 흐름이 달라진다.

<그림 3>에 표시된 문자열과 같이 키가 올바른 경우에는 0x00501761번지의 코드를 실행하며, 올바르지 않은 경우에는 0x00501795번지로 분기한다. 이는 코드상에 노출된 문자열인 “Valid Licence. Thank you for registering our product.”와 “Invalid Licence Data.”를 통하여 쉽게 코드를 분석할 수 있다.

```

00501756 E8 25010000 CALL .00501800
0050175D 8BD8 MOV EBX,EAX
0050175E 84DB TEST BL,BI
0050175F v 74 34 JE SHORT .00501795
00501761 6A 04 PUSH 4
00501763 804D E4 LEA ECX,DWORD PTR SS:[EBP-1C]
00501766 BA 28185000 MOV EDX, .00501828
00501768 BB86 14030000 MOV EAX,DWORD PTR DS:[ESI+314]
00501771 F8 2600FFFF CALL .004E0E9C
ASCII "Valid Licence. Thank you for registering our product."
00501779 50 PUSH EAX
0050177A 8BC6 MOV EAX,ESI
0050177C E8 7309EFFF CALL .003F20F4
00501781 33C9 XOR ECX,ECK
00501783 5A POP EDX
00501784 E8 C79FFCCF CALL .004CB750
00501789 C786 4C020000 MOV DWORD PTR DS:[ESI+24C],1
00501793 v EB 28 JMP SHORT .005017BD
00501795 6A ASCII "Invalid Licence Data."
00501797 8D 4C200000 R SS:[EBP-20]
0050179A BA 68185000 MOV EDX, .00501868

```

<그림 3> 라이선스 인증 함수 및 인증 결과 코드 일부

따라서 강제적으로 올바른 라이선스를 인증하는 주소인 0x00501761번지를 실행하도록 0x0050175F번지에서 실행 흐름을 조작한다면,

올바르지 않은 비밀번호를 입력하더라도, 라이선스를 정상적으로 인증하였을 때 출력되는 문자열인 “Valid Licence. Thank you for registering our product.”가 출력된다.

상기 코드 분석을 통하여, 라이선스 인증의 우회가 가능하지만, 인증을 우회하더라도, 30개를 초과하는 파일을 편집할 수 있는 제한 기능을 사용하지 못하는 한계점이 존재한다.

B 소프트웨어에서 제한하는 기능을 우회하기 위하여, 라이선스를 인증하지 않더라도, 30개를 초과하는 파일을 편집하는 취약점을 분석하였다. B 소프트웨어가 특정 코드에서 편집을 원하는 파일의 개수를 비교함으로써 기능을 제한하는 것으로 가정하였고, 10진수 30과 비교명령어를 포함하는 코드를 중점적으로 분석하였다.

분석 결과, <그림 4>와 같이 특정 함수 호출 후, 10진수 30에 해당하는 0x1E와 비교하는 코드를 확인하였고, 비교 결과에 따라, 실행 흐름이 달라진다. 다시 말하면, 0x002A514E번지의 call은 업로드한 파일의 개수를 계산하여 EAX에 저장하며, EAX에 저장된 파일의 개수와 0x1E (30) 을 비교하여 작을 경우, 0x002A51AE번지로 분기하여 파일 업로드 기능을 정상적으로 실행한다. 하지만, 업로드한 파일의 개수가 30보다 큰 경우에는 0x002A5158번지를 실행하여 라이선스가 등록되지 않았다는 문자열인 “Not registered. Trimming to 30 files”를 참조하였다.

```

002A514E E8 79DA0000 CALL .002B2BCC
002A5153 83F8 1E CMP EAX,IE
002A5156 v 7E 56 JLE SHORT .002A51AE
002A5158 B9 01000000 MOV ECX,1
ASCII "Not registered. Trimming to 30 files"

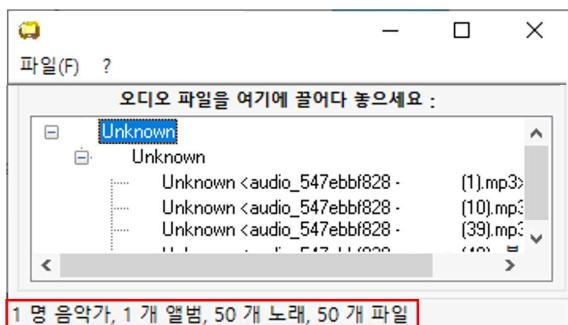
```

파일 업로드 기능 실행

<그림 4> 업로드하는 파일의 개수를 비교하는 코드 일부

따라서 0x002A5153번지의 0x1E를 30보다 더 큰 수, 예를 들면, 50으로 변경한다면, 제한 기능을 우회하여 50개의 파일을 업로드하고 편집할 수 있을 것으로 가정하였다. 이러한 가정

을 검증하기 위하여, 총 50개의 파일을 업로드한 후, 해당 코드의 0x1E를 0x32 (50)로 수정하였으며, 그 결과, 그림 5와 같이 라이선스를 인증하지 않더라도 B 소프트웨어가 제한하는 기능을 우회하여 50개의 파일이 편집 가능한 취약점을 실증하였다.



<그림 5> B 소프트웨어 제한 기능 우회 결과

#### IV. 결론

소프트웨어의 지적 재산권을 보호하기 위한 다양한 기술들이 등장하였지만, 역난독화 및 역공학과 같은 기술을 악용함으로써 소프트웨어의 지적 재산권을 보호받지 못하는 한계점이 존재한다. 이러한 한계점이 가지는 근본적인 원인을 규명하기 위하여, 본 논문에서는 라이선스 인증 기술이 적용된 B 소프트웨어를 대상으로 취약점을 분석하였다.

분석 결과, 라이선스 인증 결과에 출력되는 문자열과 라이선스 키인 일련 정보를 비교하는 코드를 분석함으로써 인증을 우회하였지만, B 소프트웨어에 적용된 제한 기능을 사용하지 못하는 한계점이 존재하였다. 이를 우회하기 위하여, 파일 개수와 비교 명령어를 기반으로 제한 기능을 우회함으로써 라이선스 인증 기술을 무력화하는 것을 실증하였다.

본 논문에서 규명한 라이선스 인증 기술이 가지는 근본적인 취약점을 기반으로, 제한 기능과 관련된 정보를 하드코딩하지 않는 방안 및 라이선스 키와 연동함으로써 제한 기능과 관련된 정보를 노출시키지 않는 방안을 연구할 예정이다.

#### 감사의 글

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학 지원사업의 연구결과로 수행되었음"(2019-0-01 056)

#### [참고문헌]

- [1] 소프트웨어 정책연구소(SPRI), "국가별 SW 불법 복제율," [https://stat.spri.kr/posts/view/22306?code=stat\\_sw\\_illegal\\_copy](https://stat.spri.kr/posts/view/22306?code=stat_sw_illegal_copy), 2019년 6월 28일 등록, 2021년 2월 5일 접속.
- [2] 한국저작권위원회, "정정당당 브로슈어," <https://www.copyright.or.kr/kcc/itsam/licensedata/view.do?brdctsno=2058>, 2021년 2월 5일 접속.
- [3] 한국저작권위원회, "소프트웨어 관리가이드," <https://www.copyright.or.kr/information-materials/publication/education-and-promotion/view.do?brdctsno=40228&list.do?pageIndex=1&brdctsstatecode=&brdclasscode=&servicecode=06&nationcode=&searchText=&searchTarget=ALL#>, 2017년 3월 7일 등록, 2021년 2월 5일 접속
- [4] 서재훈, 유성민, 박유경, "소프트웨어 보호를 위한 기술 분석," 한국정보기술학회 학회지, 제13권, 제1호, pp. 33-39, 2015년 6월.
- [5] 이경률, 육형준, 임강빈, 유일선, "소프트웨어 보안을 위한 난독화 기술 동향," 한국정보과학회 학회지, 제34권, 제1호, pp. 22-27, 2016년 1월.
- [6] 조성제, 김동진, 박민규, "소프트웨어 저작권 보호 기술 동향," 한국정보기술학회 학회지, 제11권, 제2호, 2013년 12월
- [7] 장혜영, "역공학 공격에 대한 소프트웨어 보호 기법," 단국대학교, 박사학위논문, 2010년 8월

- [8] 강기봉, “컴퓨터프로그램의 리버스 엔지니어링에 관한 법정책적 소고”, [https://www.moleg.go.kr/mpbleg/mpblegInfo.mo?mid=a10402020000&mpb\\_leg\\_pst\\_seq=133303](https://www.moleg.go.kr/mpbleg/mpblegInfo.mo?mid=a10402020000&mpb_leg_pst_seq=133303), 한양대학교 법학연구소, 2014년 3월 25일 등록, 2021년 2월 5일 접속
- [9] 유성민, “소프트웨어 저작권 보호를 위한 정부의 정책방향,” 한국정보기술학회 학회지, 제13권, 제1호, pp. 41-47, 2015년 6월

# 원격근무 환경에 대한 악성코드 보안 기술 동향

홍승표\*, 이훈재\*\*

\*동서대학교 (대학원생)

\*\*동서대학교 (교수)

Malware security technology trends in remote work environments

Seoung-Pyo Hong\*, Hoon Jae Lee\*\*

\*DongSeo University(Graduate student)

\*\*DongSeo University(Professor)

## 요약

최근 코로나 19를 통한 우리 생활의 패턴은 많은 변화를 주었다. 기업 근무환경은 원격근무가 표준으로 잡힐 만큼 많은 변화가 있었다. 코로나 19 확산에 따라 사이버 공간은 관련된 위협이 더 증가할 것으로 보인다. 공격자들은 이러한 코로나 19 확산을 이용하여 피싱, 스미싱 등 사회 공학적 기법으로 악성코드 및 악성 앱 유포, 개인정보 유출 등 다양한 공격을 수행한다. 본 논문에서는 이에 따른 보안 기술 동향을 분석하고, 보안 위협에 대응하는 해결방안을 알아보려고 한다.

## I. 서론

2019년 12월을 기점으로 코로나 19가 퍼진 아래 2021년 현재도 아직 큰 피해를 보고 있다. 기업 근무환경도 국내·외 기업에서는 원격근무가 표준으로 잡힐 만큼 많은 변화가 있었다. 비대면을 일컫는 ‘언택트(Untact)’가 급격히 증가했기 때문이다. 코로나 19 이전에는 원격 진료, 온라인 쇼핑, 재택근무 등 스마트폰 앱을 이용하여 비대면 일들이 부분적으로 이루어졌지만, 앞으로는 온라인 일 처리가 더욱 심화하고 일상화되는 세상이 되었다. 이런 상황에서 사이버 공간은 관련된 위협이 지속되고 있다. 공격자들은 코로나 19 관련 피싱, 스미싱 등 사회공학적 기법으로 악성코드 및 악성 앱 유포, 개인정보 유출 등 다양한 형태의 공격을 수행하고 있다. 많은 기업은 이러한 사이버 보안 위협 속에서 대응할 수 있는 보안 솔루션이 취약해 있다. 이에 따른 보안 기술 동향을 분석하고, 보안 위협

에 대응하는 해결방안을 알아보려고 한다.

## II. 관련 연구

### 2.1 원격근무

원격근무 개념은 1973년 미국 캘리포니아대학 미래연구센터의 ‘Jack Nilles’가 보험회사의 원격근무 시범프로젝트를 수행하면서 최초로 사용되었다. [1] 개인용 컴퓨터나 통신기기를 이용해서 사무실 이외의 장소에서 작업을 수행하는 근무를 이야기한다.

### 2.2 원격근무에 대한 위협

많은 국내·외 기업들은 코로나 19 확산에 따라 비대면 근무를 하고 있다. 변화된 근무형태에 따라 새로운 보안 위협이 발생할 수 있다. 원격근무에 대한 보안 위협을 [시스템 OS, 네트워크, 애플리케이션, 정보보호 일반/사회공학]의 4가지의 기준으로 [표 1]과 같이 분류 및 분

석한다. [2]

시스템 / OS	네트워크
Mac 사용자 대상 공격	안전하지 않은 네트워크로의 접속
엔드포인트 장비 및 관련 기기에 대한 위협	다수의 원격 접속 지점으로 인한 취약점
취약한 접근제어기술에 대한 위협	종단(Edge, Endpoint) 보안에 대한 취약점
애플리케이션	정보보호일반
취약한 FTP 서비스의 사용, 그에 대한 위협	근무자의 보안 부주의
취약한 이메일 서비스의 사용	임직원 계정을 노리는 크리덴셜 스터핑/피싱
웹캠 관련 해킹	중요 데이터 유출
웹, DNS에 대한 위협	원격접속 로그 기록위협

표 1 원격근무에 대한 보안 위협

### III. 사이버 보안 위협 동향

#### 3.1 사회공학적 기법

사회공학적 기법 공격은 크게 2가지로 인간기반과 컴퓨터 기반으로 나누어지는데, 사이버 공간에서는 컴퓨터 기반 방법으로 피해를 본다. 컴퓨터 기반은 공격 대상에게 악성코드, 컴퓨터 프로그램 혹은 웹 사이트 등의 수단을 이용하여 접근하는 경우이다. [3] [그림 1]은 사회공학적 공격 분류를 나타내는 그림이다.

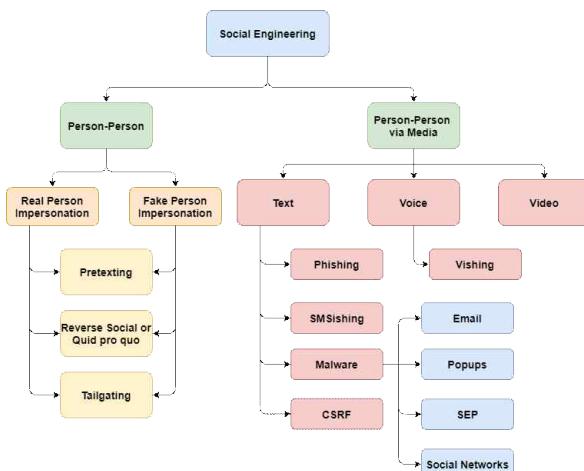


그림 1 사회공학적 공격 분류

#### 3.2 사회공학적 공격 피해사례

현재 국내에서는 코로나 19 확산에 따른 재난지원금이 지급된다. 1차 재난지원금 같은 경

우 전 지역에서 지원이 되었다. 최근에는 2차 재난지원금을 경기도에서 지원하게 되는데, 일부 포털사이트에서 재난지원금 관련 검색을 하면 공식 홈페이지가 아닌 광고·유사·가짜사이트로 위장한 피싱범죄가 일어나고 있다. 악성코드가 설치된 해당 사이트는 지급을 위해 접속을 하게 되면 소액결제를 하여 돈을 빼가는 사이트로 확인이 되었다.

#### 3.3 사회공학적 공격 대응방안

정보기기의 종류와 새로운 정보기술서비스의 증가로 인해 사회공학적 공격을 할 수 있는 통로가 증가하고 있다. 미리 그 통로를 예측하고, 사용자들에게 미리 위험성을 경고해야 한다. 일반 사용자 측면에서 보았을 때 사회공학적 공격에 대한 대응은 어떻게 대응할 것인지에 초점이 맞춰질 수 있다. 사회공학적 공격에 대한 대응방안은 여러 논문에서 제시가 되었지만 큰 효과를 볼 수가 없었다. 정보보안의 가장 큰 위협은 ‘사람’이다. 보안의 위협 ‘사람’을 막기 위해서는 기업 내부를 통제하기 위한 정보보호 매뉴얼이 필요하다. 기초적인 매뉴얼은 아래와 같다.

- 1) 인터넷 브라우저의 차단 기능
- 2) 광고 흥보, 스팸 메일 차단 기능
- 3) 피싱, 스미싱 사례 및 통계 제공
- 4) 일반적인 보안 준수 사항 고지

#### 3.4 첨부파일 악성코드 유포

대부분의 피싱 공격은 SMS, 메일을 통해 진행된다. SMS 같은 경우 코로나 19 관련 내용을 통해 사용자의 클릭을 유도해 개인정보를 탈취해 간다. 메일 같은 경우 사회공학적 기법을 통해 사용자가 클릭하게끔 유도하여 첨부파일을 실행하게끔 한다. 첨부파일의 내용은 기업 업무 관련 문서파일, 재난 본부 등으로 속여 파일첨부를 하게 된다. 첨부한 실행 파일은 RAT 계열 및 개인정보 유출형 악성코드로 확인이 된다. 업무 관련 문서파일의 경우 MS 오피스의 취약점을 악용하여 추가 악성 파일을 다운로드하게 한다. 추가 악성 파일은 사용자의 시스템

을 장악하고, 다양한 악성 행위를 할 수 있다.

#### IV. 결론

코로나 19 이슈로 인해 현재 사이버 공간에서는 많은 보안 위협들이 존재한다. 그중에서도 이를 악용하는 사회공학적 기법 공격은 인간의 심리를 이용하여 많은 사람의 개인정보를 탈취해 간다. 기업들의 가장 치명적인 위협은 사회공학 해킹 기법이 주를 이룬다. 사회공학 기법은 보안 담당자를 포함해서 모든 직원이 인지하고 대비해야 사전 예방, 피해를 최소화할 수 있다.

#### [참고문헌]

- [1] “Recommended Guide to Information Protection for Smart Work Activation” KISA, pp. 40–107, December 2011.
- [2] 김소연, 하영민, 김성율, 최상용, 이종락 (2020). 원격근무 환경에서의 사이버 보안 위협 분석. 한국컴퓨터정보학회 학술발표논문집, 28(2), 97–98.
- [3] “사회공학적 해킹의 변화양상”, KISA, 2008.

# 자율주행 자동차 내부 CAN 네트워크 공격 탐지를 위한 프레임워크 개발

김수빈\*, 고예지\*, 김세진\*, 이예솔\*, 석지은\*, 임강빈\*\*

\*순천향대학교(대학생), \*\*순천향대학교(교수)

Development of a Framework for Detection of CAN Network Attacks inside Autonomous Vehicles

Jieun Seok\*, Subin Kim\*, Yeji Koh\*, Sejin Kim\*, Yesol Lee\*, Kangbin Yim\*\*

\*Soonchunhyang University(Undergraduate student)

\*\*Soonchunhyang University(Professor)

## 요약

자율주행 기술은 주변 환경과 통신하여 위험 분석과 운행 경로를 계획하는 기술로 운전자의 안전성과 효율성, 편의성을 향상시킨다. 자율주행 차량의 내부에서는 모듈과 컨트롤러가 CAN 네트워크를 사용하여 제어된다. 이러한 네트워크는 Broadcast 형식으로 통신하여 송신자를 확인하지 않는다는 점에서 메시지 주입과 같은 공격을 야기할 수 있다. 본 논문에서는 실제 자율주행기술이 탑재된 차량을 기반으로 자동차 내부 네트워크인 CAN 프로토콜의 데이터 수집을 통하여 위협 가능성을 분석한다. 또한, 수집한 CAN 프로토콜을 기반으로 실제 주행 상황의 재연과 실시간 모니터링 및 비정상 메시지 탐지가 가능한 동적 분석 프레임워크를 개발한다.

## I. 서론

현대의 자동차는 이동수단으로서의 의미뿐만 아닌 다양한 기능이 결합하여 자율주행이라는 새로운 분야로서 발전해가고 있다. 자율주행이란 운전과 관련된 기술적인 간섭 없이 시스템 자체적으로 판단하여 차량이 주행하는 것이다. 차량에 부착된 레이더, 카메라, 초음파 센서 등으로 차간거리 제어, 차선 유지 제어, 충돌 회피기술과 같은 위험 요소를 판단하고 주변 환경을 인식하여 경로를 계획한다.[1]

본 논문에서는 자율주행 시스템이 탑재된 차량을 기반으로 수집된 CAN 메시지 주입 및 재실행을 위한 프레임워크를 구축한다. 구축한 프레임워크를 이용하여 각 모듈에서 추출한 CAN 메시지를 수집 및 분석하는 과정을 진행한다. 이후, 실제 주행을 통해 수집한 CAN 메시지 데이터를 프레임워크에 주입 및 재실행한다. 동시에 프로그램을 통해 아이디의 경우 프레임워크에서 모듈별로 수집하였던 CAN 메시지 아이

디와의 비교 분석을 진행하며, 메시지 데이터 값의 경우 일정 값 이상의 오차 발생을 감지한다. 감지된 이상 데이터는 웹 페이지를 이용해 시각화하여 CAN(Controller Area Network) 프로토콜의 실시간 모니터링이 가능하도록 하였으며 해당 과정에서 특정 CAN 메시지에 비정상적인 메시지가 탐지될 경우, 그와 관련된 공격 탐지 및 경고 알림이 발생하도록 개발했다.

## II. CAN 네트워크 분석

### 2.1 CAN 네트워크

주행 데이터의 송·수신 과정에서 차량의 각 모듈은 차량 내 통신 네트워크 형식인 CAN 프로토콜을 사용하여 통신한다[2]. CAN 버스를 통해, 병렬로 연결된 다수의 ECU 간 데이터를 주고받는[3] CAN 네트워크 종류에는 C-CAN, B-CAN, M-CAN 등이 존재한다.

C-CAN은 속도, RPM, 브레이크 등 전반적

인 차량 운행과 관련된 데이터가 통신하는 네트워크이다. B-CAN은 차량 차체 부분의 제어 메시지가 통신하는 네트워크이며, M-CAN은 내비게이션과 같은 차량 내 멀티미디어 데이터 관련 메시지가 통신하는 네트워크이다[4].

## 2.2 CAN 메시지 수집

아래 [그림 1]은 실제 자율주행 차량의 ECU에서 수집한 CAN 데이터가 CAN ID, Type, DLC, Data 순서로 시각화된 화면이다. 프로그램을 통해, 미리 수집한 데이터뿐만 아니라 CAN 수집기와 각 모듈을 연결하여 수집한 모듈이 동작할 때 생성되는 CAN 데이터 또한 실시간으로 확인할 수 있다.

Time	Type	ID	DLC	Data
14.9781	Data	129	4	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
14.9881	Data	657	4	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
14.9881	Data	110	8	E0 2C 29 09 00 00 00 00 00 00 00 00 00 00 00 00
15.0281	Data	050	4	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.0479	Data	4F2	8	00 00 00 00 39 00 00 00 00 00 00 00 00 00 00 00
15.0482	Data	018	8	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.0482	Data	4F2	8	2C 29 09 00 00 00 00 00 00 00 00 00 00 00 00 00
15.0782	Data	034	8	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.0879	Data	88	8	20 00 00 00 38 00 00 00 00 00 00 00 00 00 00 00
15.0982	Data	550	8	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.0982	Data	019	8	00 00 00 00 00 20 00 00 00 00 00 00 00 00 00 00
15.0985	Data	120	4	00 00 10 20 00 00 00 00 00 00 00 00 00 00 00 00
15.1279	Data	4F2	8	20 00 00 00 38 00 00 00 00 00 00 00 00 00 00 00
15.1279	Data	110	8	E0 2C 29 09 00 00 00 00 00 00 00 00 00 00 00 00
15.1862	Data	030	8	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.1864	Data	120	4	00 00 10 20 00 00 00 00 00 00 00 00 00 00 00 00
15.1479	Data	4F2	8	20 00 00 00 38 00 00 00 00 00 00 00 00 00 00 00
15.1782	Data	4F2	8	2C 29 09 00 00 00 00 00 00 00 00 00 00 00 00 00
15.1782	Data	018	8	00 00 00 00 00 20 00 00 00 00 00 00 00 00 00 00
15.1784	Data	120	4	00 00 00 00 20 00 00 00 00 00 00 00 00 00 00 00
15.1879	Data	557	8	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.1879	Data	4F2	8	20 00 00 00 38 00 00 00 00 00 00 00 00 00 00 00
15.1882	Data	610	8	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.2181	Data	4F2	8	2C 29 09 00 00 00 00 00 00 00 00 00 00 00 00 00
15.2279	Data	018	8	00 00 00 00 00 20 00 00 00 00 00 00 00 00 00 00
15.2279	Data	4F2	8	2C 29 09 00 00 00 00 00 00 00 00 00 00 00 00 00
15.2382	Data	110	8	E0 2C 29 09 00 00 00 00 00 00 00 00 00 00 00 00
15.2382	Data	030	8	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
15.2384	Data	120	4	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

[그림 1] 수집된 CAN 메시지

실제 자율주행 차량의 ECU에 해당하는 ICU, IBU, ECM, CLUSTER, AVN, MDPS, DATC의 모듈별 단위 테스트를 통해 개별 CAN 데이터를 수집하였으며, ECU를 두 개 이상 연결할 경우 발생하는 CAN 데이터 또한 동일한 방법을 이용하여 수집하였다. 실제 주행 데이터의 경우, 연구 대상인 차량을 일정 거리를 주행하며 실시간으로 발생하는 데이터를 수집하였다. 이처럼 수집한 두 가지의 데이터를 비교하여 중복되는 CAN 데이터를 분석하였다.

## 2.3 CAN 메시지 분석

전반적인 차량 운행에 관련된 C-CAN 데이터 중 차량의 속도 값과 RPM, 기어, 좌우 방향 등, 비상등, 브레이크 및 충돌 경고음을 집중적으로 분석하였다. 다음 [표 1]은 집중적으로 분석이 이뤄진 C-CAN의 메시지 데이터 값이다.

분석을 통해, 실제 주행 중 수집한 CAN 데이터 값과 분석한 내용을 기반으로 프로그램에 직접 조작하여 주입한 CAN 데이터 값의 결과가 일치함을 확인 할 수 있었다.

CAN ID	기능	DATA
386h	속도	홀수 바이트의 첫 번째 비트 값에 따라 속도 값의 범위 변화
366h	RPM	3byte에서 (상위 4비트 * 0.065 + 하위 4비트) * 1000 연산 수행 시, RPM 값 예측 가능
367h	기어	4byte에서 x5이면 D/ x0이면 P
541h	좌측 방향 지시등	2byte에서 41이면 On/ 49이면 Off
541h	우측 방향 지시등	7byte에서 4D이면 On/ 4D이면 Off
541h	비상등	2byte와 7byte에서 41이면 On/ 4D이면 Off
4a9h	브레이크	7byte가 브레이크를 끊으면 40/ 끊지 않으면 00
58bh	충돌 경고음	3byte에서 0x이면 Off

[표 1] 분석한 CAN 메시지 데이터

## III. CAN 네트워크 공격 탐지 방안

### 3.1 공격 탐지를 위한 Database 구축

데이터 관리 기능 및 보안성을 위해 수집된 데이터의 Database를 구축하였다. CAN의 종류에 따라 테이블을 생성하고 각각의 테이블에는 CAN ID, 길이, 데이터 필드를 생성한다. CAN 수집기를 통해 수집된 데이터는 파이썬 코드 내의 ReadData 함수에서 Json 포맷 형태로 변환하고, ProcessData 함수 부분에서 16진수로 표현되어있는 데이터 부분을 가공하여 일반 사용자도 쉽게 의미를 알아차릴 수 있도록 한다.

아래의 [그림 4]는 수집된 차량 속도 데이터를 가공하는 코드 중 일부분이다. 수집된 데이터(예: 01 CA 01 05 01 4F 01 C0)를 가지고 데이터를 가공하여 70이라는 가공된 값을 데이터베이스에 전송하고 동시에 이상 데이터 탐지를 수행한다. 가공된 속도 값과 이전의 속도 값의 차이가 일정 값 이상 발생하거나 단위 테스트 과정에서 수집되지 않은 CAN ID가 탐지될 경우 이상 데이터로 간주한다.

```
if msg.ID == 902:
    if Data[3] == "0" or Data[3] == "4" or Data[3] == "8" or Data[3] == "c":
        n = 0
        while (n < 23):
            temp = float(int(Data[4 + n], 16))
            value = value + temp * 2
            temp = float(int(Data[9 + n], 16))
            value = value + temp * 0.15
            n = n + 6
        Data = str(value)
```

[그림 2] 데이터 가공 코드 - 속도

이와 마찬가지로 수집된 RPM 데이터도 가공 과정을 거쳐 이해하기 쉬운 데이터로 변환되어 데이터베이스에 저장된다. 이 부분에서도 이전 RPM 값과 비교하여 일정 값 이상 차이가 나면 이상 데이터로 간주하는 이상 탐지 기능이 실행된다. 탐지 과정에서 이상 데이터로 간주 될 경우, 속도 및 RPM 값이면 -1, CAN ID 부분이면 0xffffffff이 데이터 필드 값에 저장된다.

가공된 데이터는 SendData 함수 내의 SQL 쿼리문을 통해 약 0.5초마다 데이터베이스에 저장된다. 다음 [그림 5]는 가공된 데이터가 C\_CAN 테이블에 저장된 모습이다.

mysql> select * from c_can;		
ID	Length	Data
0x1	4	00 00 00 04
0x130	8	d8 7f 7f 80 00 00 0b 81
0x140	8	33 7f 00 71 20 00 0a 16
0x153	8	20 80 10 ff 00 ff 70 1e
0x162	9	00 8e 15
0x164	4	00 08 16 28
0x220	8	ff 03 7e 00 00 f0 0f b6
0x222	8	60 00 03 00 00 00 00 00

[그림 3] 저장된 가공된 CAN 데이터

## IV. 테스트 프레임워크 개발

### 3.1 테스트베드 구축

아래 [그림 6]은 실제 차량 내부를 그대로 재현한 프레임워크의 모습이다. 전체적인 전원은 파워를 사용하여 모듈에 전원을 공급하고 아두이노 릴레이를 통해 모듈들의 ON/OFF 제어가 가능하다. 아크릴 박스 최하층 칸에는 B-CAN, C-CAN, M-CAN 별로 CAN 버스를 배치하여 데이터를 수집한다. 프레임워크에 위치한 모니터 중 좌측은 실제 주행 데이터 수집 과정에서 녹화된 영상이 재생되고, 우측에는 실시간으로 주입되는 CAN 메시지가 분석된 결과가 웹 페이지를 통해 실시간으로 시각화된다.



[그림 4] 테스트 프레임워크 전체 모습

### 3.2 CAN 네트워크 공격 탐지

웹 페이지는 가공된 CAN 데이터가 저장되어 있는 DB로부터 결과를 받아와 보기 쉽게 UI로 차량의 속도, RPM, 방향, 브레이크의 실시간 상태를 시각화하였다.



[그림 5] CAN메시지 조작 탐지 화면

수집된 CAN 데이터의 실시간 모니터링이 가능한 웹 페이지는 [그림 7]과 같다. CAN 메시지 데이터 이상 탐지 시, 웹 페이지 좌측 상단에 메시지 조작 탐지 문구와 함께 알림을 생성하여 공격 탐지 결과를 알려준다.

## V. 결론

본 논문에서는 실제 자율주행 시스템이 탑재된 차량을 기반으로 테스트 프레임워크를 구축하였으며 수집한 CAN 메시지 데이터를 분석하여 실시간 공격 탐지 모니터링이 가능한 웹 페이지를 개발하였다.

테스트 프레임워크는 실제 자율주행 차량에서 수집한 데이터와 모듈별 데이터를 이용하여 차량 내의 메시지 조작, 비정상 전달 등의 시스템 공격 탐지 기술개발에 활용할 수 있다.

## [참고문헌]

- [1] 이병윤, “국내외 자율주행자동차 기술개발 동향과 전망”, 2016.
- [2] 이병수, 박민규, 성금길, “CAN을 기본으로한 전기자동차용 차량 네트워크 교육용 시스템 개발”, 2006.
- [3] Youngho An, Junyoung Park, Insu Oh, Myoungsu Kim, Kangbin Yim, “Design and Implementation of a Novel Testbed for Automotive Security Analysis”
- [1] 커넥티드 카 침해사고 시나리오 모델 구현 및 분석방법 연구(KISA-WP-2018-002)

# CNN 네트워크 물리채널 보안 분석을 위한 하드웨어 가속기 설계

이진재\* 박종욱\*\* 이준호\*\*\* 김호원\*\*\*\*

\*부산대학교 (대학원생) \*\*부산대학교 (학부생) \*\*\*동의대학교 (학부생)

\*\*\*\*부산대학교 (교수)

## Hardware Accelerator Design for CNN Network Physical Channel Security Analysis

Jin-Jae Lee\* Jong-Uk Park\*\* Jun-Ho Lee\*\*\* Ho-Won Kim\*\*\*\*

\*Pusan National University(Graduate student),

\*\*Pusan National University(Undergraduate student)

\*\*\*Dong-eui University(Undergraduate student)

\*\*\*\*Pusan National University(Professor)

### 요약

본 연구에서는 CNN 네트워크 물리채널 보안 분석을 위한 하드웨어 가속기를 구현하였다. 실제 딥러닝 가속기에서 사용되는 기법을 적용하여 상용 딥러닝 가속기의 구조와 유사하도록 구현하였고 물리 채널 분석이 용이하도록 구현하여 딥러닝 가속기 보안 분석의 접근성을 높일 수 있을 것으로 기대된다.

I. 서론 주제어 : CNN, 딥러닝 가속기, 물리채널 보안 분석  
은 진입장벽으로 인해 활발한 연구가 이루어지지 못하고 있다.

최근 딥러닝 기술은 의료, 산업, 가전 등의 다양한 분야에서 활용되고 있으며 이러한 딥러닝 기술에 대한 하드웨어 가속화 연구 또한 활발하게 진행되고 있다. 하지만 딥러닝 네트워크의 학습이나 추론 과정에서는 개인의 생체정보나 위치정보와 같은 민감한 데이터를 다루며, 이를 악용하여 생체정보 탈취를 통한 금융서비스 인증 우회, 개인정보 유출을 통한 프라이버시 침해 등의 악의적인 공격이 이루어질 수 있다. 또한, 학습된 딥러닝 네트워크 모델을 물리 채널 분석으로 획득하여 도용하는 방식을 통해 지적재산권을 침해하는 사례도 발생할 수 있다.

딥러닝 가속기에 대한 분석 및 공격기법에 대한 연구는 아직 초기 단계이며, 악의적인 공격을 예방하거나, 공격이 이루어졌을 경우에 대한 대응방안은 딥러닝 가속기 분석에 대한 높

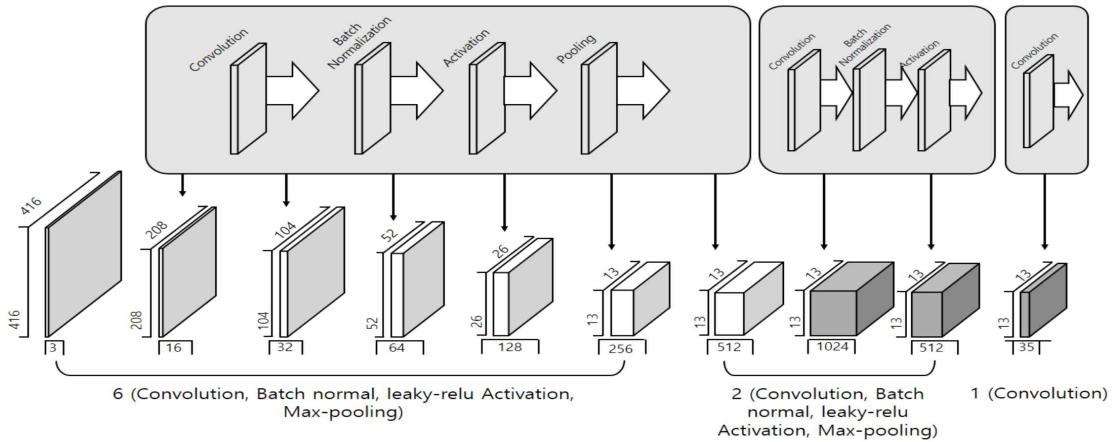
이에 본 연구에서는 물리채널 분석을 위한 딥러닝 가속기를 구현하여 딥러닝 가속기 공격에 대한 대응기술 연구의 접근성을 높이고자 한다.

### II. 관련 연구

본 절에서는 물리채널 분석을 통한 CNN 공격과 가속기 설계 관련 연구에 대해 기술한다.

#### 2.1 CNN 설계

가속기 설계에서 가장 중요한 것은 한정된 하드웨어 자원에서 효율적인 연산을 하는 것이다. 또한 가속화를 위해서는 외부 메모리의 접근을 줄이고 on-chip 메모리 사용량을 높여야 한다. 이를 위한 다양한 연산 스케줄링 방안에 대한 연구가 진행 중이다. 기본적인 개념은 Co



[그림 1] Tiny Yolo 네트워크 구조

nvolution 연산을 할 때, Row base로 partial product들을 이용하여 연산의 파이프라인화, 병렬화 및 메모리 접근을 최소화하는 Row pass[1] 방안 등이 있다.

## 2.2 물리채널 분석을 통한 CNN 가속기 공격

학습 및 추론 과정에서 사용되는 CNN 가속기는 첫 번째 계층은 많은 이미지를 처리해야 하므로 주로 버퍼를 이용하여 구현이 된다. 이러한 버퍼가 삽입된 첫 레이어 로직의 전력 트레이스를 측정하여 이미지를 재구성하는 방법에 대한 연구가 이뤄지고 있다[2]. 또한 로직의 전력 측정뿐만 아니라 메모리 접근 패턴을 활용하여 각 계층의 이미지 크기, 레이어 수 등 네트워크 구조를 분석함은 물론 가중치 값을 복구 할 수 있는 연구가 활발히 진행 중이다[3]. 이처럼 물리채널 분석을 통한 CNN 가속기 공격은 목표에 따라 다양한 접근법으로 연구되고 있다.

## III. 배경지식

### 3.1 CNN

CNN(Convolutional Neural Network)은 인간의 시신경을 모방하여 만든 딥러닝 구조 중 하나이며 이미지 처리에 높은 성능을 보이는 신경망이다. CNN은 크게 Convolution Layer와 Pooling Layer로 이루어져 있으며 Convolution Layer에서는 이미지의 각 Tensor Value(Height, Width, Channel)를 사용된 kernel에 맞게 합

성 합(convolution) 연산을 통해 축소된 행렬의 결과로 데이터를 도출해내게 된다. Pooling Layer에서는 도출된 합성 곱 데이터의 공간적인 특성을 유지하면서 크기를 줄이는 작업을 실행한다. Pooling에는 Max Pooling, Average Pooling이 있으며 Max Pooling은 kernel과 겹치는 영역 안의 최댓값을 추출하고 Average Pooling은 영역 안의 평균값을 계산해 추출한다.

### 3.2 Tiny Yolo v2

Tiny Yolo는 기존 Yolo의 알고리즘을 축소하여 저성능 프로세서에서도 원활하게 구동할 수 있게 설계된 알고리즘이다. Yolo는 네트워크의 최종 출력단에서 바운딩 박스의 위치 찾기와 클래스 분류가 동시에 이루어지며 다른 계열(R-CNN)의 딥러닝 네트워크 모델에 비해 간단하고 빠르다는 특징을 가진다.

Tiny Yolo v2 네트워크의 구조는 [그림1]과 같이  $3 \times 416 \times 416$ 의 입력 데이터를 CNN 기반 알고리즘을 거쳐  $1 \times 125 \times 13 \times 13$ 의 텐서로 출력시킨다. 여기서 출력된 텐서 내에는 CNN 알고리즘을 통과하면서 계산한 바운딩 박스와 각 바운딩 박스의 정보가 포함된다.

### 3.3 하드웨어 딥러닝 가속기

하드웨어 딥러닝 가속기(Hardware Deep Learning Accelerator)는 딥러닝 연산의 주를 이루는 병렬 단순 연산(곱셈, 덧셈)을 고속으로 처리하기 위해 사용되는 하드웨어이다. 주로 GPU(Graphic Processing Unit)가 사용되고 있으며 FPGA(Field Programmable Gate Array)와

ASIC(Application-Specific Integrated Circuit)의 형태도 존재한다. ASIC 형태의 하드웨어 딥러닝 가속기에는 NPU(Neural Processing Unit)와 Google의 TPU(Tensor Processing Unit) 등이 있다.

## IV. CNN 가속기 구조

본 절에서는 Tiny Yolo v2 딥러닝 네트워크 모델의 가속기 설계 구조에 대해서 기술한다.

CNN 네트워크의 데이터 양자화와 같은 전처리 연산, 가속기의 연산결과를 사용하여 객체 탐지 결과를 생성하는 후처리 연산은 하드웨어 가속기에서 이루어지지 않으며, 반복이 많고 대용량 연산이 필요한 부분만 하드웨어 가속기상에 구현된다. 본 연구에서 구현하는 CNN 가속기는 16-bit의 fixed point 연산을 지원한다.

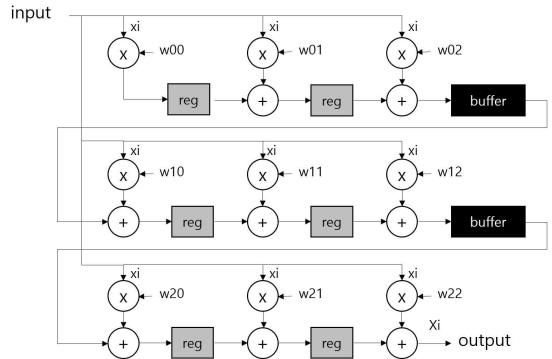
### 4.1 Convolution 모듈

본 연구에서 사용하는 CNN 네트워크 모델인 Tiny Yolo v2는 최대 3x3 크기의 kernel 사이즈를 가지는 filter를 이용하여 convolution 연산을 수행하고, 마지막 fully connected layer에서만 1x1 크기의 kernel 사이즈를 가지는 filter를 사용한다. 때문에 본 연구에서는 filter의 최대 kernel 크기인 3x3을 기본적으로 지원하고 fully connected layer를 위한 kernel 크기의 제어를 컨트롤 유닛에서 수행하는 방식을 가진다. 또한, 대용량 연산이 고속으로 수행되어야 하기 때문에 한 클럭에 하나의 feature map 원소가 출력이 되도록 9개의 MAC(Multiply and Accumulate) 모듈을 사용하는 파이프라인 구조를 가진다.

MAC 모듈은 16-bit fixed point 곱셈과 덧셈을 지원하는 모듈로 파이프라인의 결과 출력 cycle을 단축시키기 위해 단일 clock에 두 연산이 모두 수행되도록 구성된다.

convolution 연산 모듈은 [그림2]와 같은 파이프라인 구조를 가진다. filter의 weight 값은 하나의 채널에 대해서 고정으로 사용되기 때문에 하나의 채널에 대한 연산이 끝날 때까지 register에 저장되며, 매 clock마다 입력 이미지의 pixel 값이 입력되면 해당 자리의 weight 값

과의 fixed point 곱셈 연산을 거친 뒤에 이전 값과의 덧셈 연산으로 값을 누적하여 register에 저장한다. 파이프라인이 가득 차게 되면 다음 clock부터 convolution 연산의 결과값이 출력되게 된다.



[그림 2] convolution 연산 모듈 구조

본 연구에서는 연산 코어간의 부채널 신호 간섭을 최소화하여 물리채널 분석이 용이하도록 하기 위해서 convolution 모듈을 하나만 사용한다.

### 4.2 Batch normalization 모듈

추론용 Batch normalization 연산 모듈은 아래 (a), (b)식과 같은 연산 과정을 가진다.

$$(a) \hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}}$$

$$(b) y_i = \gamma \hat{x}_i + \beta$$

$\mu_\beta$ (mini-batch mean),  $\sigma_\beta^2$ (mini-batch variance),  $\gamma$ (scale),  $\beta$ (bias)는 미리 학습된 파라미터 값을 이용하여 연산을 수행하고 (a), (b) 연산을 수행하기 위해서는 fixed point에 대한 square root 연산 모듈이 추가적으로 필요하다. Batch normalization 연산 모듈은 convolution 연산 모듈의 결과값을 입력으로 받아서 연산을 수행하기 때문에 전체 연산 과정이 파이프라인과 같이 동작하기 위해서는 Batch normalization 연산 또한 단일 clock에 연산 결과값을 출력할 수 있어야 한다.

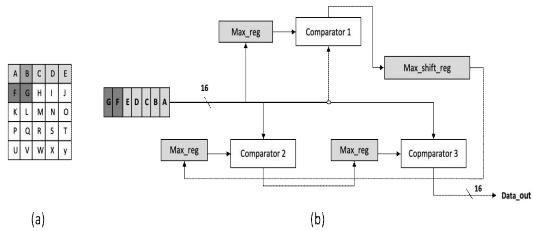
### 4.3 Activation function 모듈

CNN 관련 Activation function은 하드웨어 상에서 MSB의 부호 비트만을 이용하여 효율적인 구현이 가능한 ReLU 및 leaky ReLU function이 주로 사용된다. 기본적인 ReLU의 경우에

는 Multiplexer가 사용되며 leaky ReLU의 경우에는 Scale 값을 곱하는 과정이 있어 DSP(Digital Signal Processor)를 활용하거나 Multiplication 모듈이 추가로 필요하다.

#### 4.4 Pooling 모듈

Max Pooling Layer는 이전 Layer의 출력 데이터의 크기에 따라 유동적으로 설계가 가능하다. 현재 구조는 3개의 Comparator, 3개의 Max Register, (Max image size - kernel size) 크기만큼의 Variable Shift Register를 가진다. 순차적으로 데이터를 입력받기 때문에 kernel의 위치를 기억하고 출력하기 위해 row, column counter가 사용된다. stride의 크기가 3이상이 될 경우 손실되는 데이터가 많기 때문에 주로 1과 2를 사용하는 구조를 가진다.



[그림 3] Max Pooling Layer

(a) 5\*5이미지 입력 예시, (b) Max Pooling 구조

[그림3]에서 B 데이터가 3개의 comparator에 입력될 때, Max Register에 들어있는 A와 비교를 한 후 큰 데이터가 Variable Shift register로 이동하여 F 데이터가 comparator에 입력되기 직전에 Max reg에 B의 데이터가 저장된다. 이러한 설계는 순차적인 입력 데이터의 효율적인 처리를 위한 구조이며 외부 메모리의 접근을 최소화한다.

## V. 구현

본 연구에서는 CNN 하드웨어 가속기 구현을 위해 Xilinx Artix-7 XC7A200T-1SBG484C FPGA 칩을 사용하는 Digilent Nexys Video board를 사용한다.

CNN 네트워크의 전처리와 후처리 연산은 Intel Core i7-9700K 프로세서를 탑재한 외부의 PC를 사용하여 이루어지고 FPGA와의 연결은 CON-FMC USB인터페이스 보드를 사용하며, BFM(Bus Functional Module)을 사용하여

FPGA 내부의 AXI 버스와 통신이 이루어진다.

CNN 네트워크의 전처리 연산이 수행되면 입력 이미지와 filter값이 FPGA 보드상의 DDR3 메모리에 쓰여지고, 각 레이어의 연산을 위한 파라미터 값은 CNN 가속기 모듈의 CSR(Control and Status Register)에 저장된다. CNN 가속기의 연산 중간값은 고속 연산을 위해 On-chip BRAM(Block RAM)에 저장되고 On-chip BRAM의 크기를 초과하는 레이어 연산 결과의 일부분에 대해서는 외부의 DDR3 메모리에 저장한다.

모든 레이어의 연산이 종료되면 PC에서 CNN 가속기의 연산 결과값을 읽어들여 후처리 연산을 수행한 뒤에 객체 탐지의 결과를 얻을 수 있다.

구현에 있어서 CNN 하드웨어 가속기의 convolution 연산 모듈의 MAC과 같이 고속화가 가능한 부분은 FPGA의 DSP를 사용하여 구현한다.

## VI. 결론

본 논문에서는 딥러닝 가속기의 물리채널 보안 분석을 위한 CNN 추론용 하드웨어 가속기를 설계 및 구현하였다. 실제 딥러닝 하드웨어 가속기에서 사용되는 파이프라인 구조를 사용하여 연산 모듈을 구현하였으며, 단일 코어로 구현하여 코어 간의 신호 간섭을 배제하여 물리채널 분석이 용이하도록 구현하였다. 기존의 딥러닝 가속기와 비교하여 분석 난이도를 낮추어 물리채널 보안 분석에 대한 접근성을 높일 수 있을 것으로 기대된다.

## ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신 기획평가원의 대학ICT연구센터지원사업의 연구 결과로 수행되었음 (IITP-2020-0-01797)

## [참고문헌]

- [1] Ding, Caiwen, et al. "REQ-YOLO: A resource-aware, efficient quantization

- framework for object detection on FPGAs." Proceedings of the 2019 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays. 2019.
- [2] Lu, Liqiang, et al. "Evaluating fast algorithms for convolutional neural networks on FPGAs." 2017 IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). IEEE, 2017.
- [3] Yin, Qiao, et al. "FPGA-based High-performance CNN Accelerator Architecture with High DSP Utilization and Efficient Scheduling Mode." 2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS). IEEE, 2020.
- [4] Wei, Lingxiao, et al. "I know what you see: Power side-channel attack on convolutional neural network accelerators." Proceedings of the 34th Annual Computer Security Applications Conference. 2018.
- [5] Weizhe Hua, Zhiru Zhang, and G. Edward Suh. 2018. Reverse engineering convolutional neural networks through side-channel information leaks. In Proceedings of the 55th Annual Design Automation Conference, DAC 2018, San Francisco, CA, USA, June, 2018.

# 독립적인 트랜잭션 연합학습에 관한 연구

산디 라마디카, 이경현

부경대학교 정보보호학협동과정

부경대학교 IT융합응용공학과

sandika@pukyong.ac.kr, khrhee@pknu.ac.kr

## Dissacosiation Transactions in Federated Learning

Sandi Rahmadika and Kyung-Hyune Rhee

Interdisciplinary Program of Information Security, Graduate School PKNU

Department of IT Convergence and Application Engineering

Pukyong National University, Republic of Korea

### Abstract

Segregated learning is disclosed publicly that enables multiple users to improve a global deep learning model gradually. The objective of segregated learning is to preserve privacy for users during training without having to expose their data to the public. However, the system lacks an adequate incentive scheme. A blockchain smart contract can be a credible solution to provide distributed incentives for users since it is self-executing contracts with immutable data records that resistance to failure. Straightforwardly adopting smart contracts in a segregated learning system might break users' privacy. Malicious users can infer the properties of training resources. Therefore, in this paper, we investigate in which case malicious users are likely performing the inference attacks. The preliminary design scheme for dropping the risk of such attacks is also elaborated.

## I. Introduction

The decentralized approaches in managing computer commands over the internet have been extensively researched lately by academia, developers, and industries. The key motivation of this approach is to tackle the communication bottleneck issues and memory usage of the conventional centralized system [1]. Therefore, the paradigm of a centralized system for a various implementations also shifted towards decentralized manners such as financial technology, medical records, intellectual property, and to name a few.

Segregated learning is a breakthrough in the deep learning environment. It turns up the from centralized users' raw data for training to a distributed form. The raw data owned by clients are never leaving the

devices [2]. Thus, the issues of privacy in centralized learning can be addressed naturally by design.

The contributed parties in the segregated learning system should be incentivized proportionally. The parties use their resources (valuable data and computing power) to improve the global model. Ethereum smart contracts are also well-known as self-executing contracts can be a plausible solution to preserve distributed incentives that also resistance to failure.

A *two-phase commit* protocol is being implemented in order to accommodate more conventional communication between users and the aggregation server. However, during protocol implementation, the information about aggregation values might be inferred by malicious users. By adopting the inference

attacks, the observer with certain assumptions can infer and link the presence of specific data features of the users' private dataset. Hence, we outline the segregated learning model along with a decentralized incentive mechanism. We also elaborate on the potential inference attacks that can infer the knowledge between output and the owner.

## II. Core System Components

Decent design on cloud services for training a deep learning model provides tangible benefits for providers and clients in the real world. Several convenience features are accomplished with the existence of cloud-based training services. Therefore, the other well-known technologies such as internet-of-things (IoT), virtual machines, artificial intelligence (AI), fifth-generation (5G) mobile networks, and blockchain can be performed remotely in a commercial providers' data center.

**Conventional Learning Approaches.** In the AI environment, the model providers can offer their original algorithm publicly with certain conditions. Thus, the researchers and developers can use the model directly through cloud services with varied data possessed. Cloud computing is considered essential in guaranteeing the quality of services in the AI model and reducing energy consumption since the workload in many applications is always evolving.

**Blockchain in Segregated Learning.** One of the most tangible benefits of utilizing blockchain technology is the absence of intermediaries in handling the transactions. This feature is also useful for collaborative intelligence where data is not concentrated. The users' data in a segregated system are unique to each other. In short, there is a set of users with a distinguished dataset in

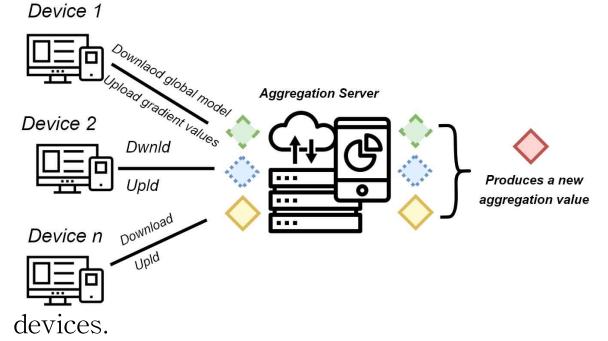


Figure 1. High-level of segregated learning

We notice that the aggregation server nor other devices within the maximum waiting time  $T_{max}$  having no knowledge to the local computing data of all parties (AFL is global accuracy). Hence, it preserves data privacy. The computation for upper bound in the general iteration is denoted to be (1). The total execution time respects the number of iterations. It is upper bounded by  $O(\log(1/\theta))$ .

$$K(A_{FL}, \theta) = \frac{\mathcal{O}(\log(1/A_{FL}))}{1 - \theta} \quad (1)$$

$$\min_{\theta} \Omega(\theta) + \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y_i, f_{\theta}(x_i)) \quad (2)$$

The training algorithm  $T$  has a set of parameters  $\theta \rightarrow f_{\theta} = x \rightarrow y$ . It is applied to minimize a loss function  $L$  which penalizes the mismatches between true labels  $y$  and predicted labels produced by  $f_{\theta}(x)$ . With  $\Omega(\theta)$  as a regularization term that penalizes model complexity. It helps preventing models from overfitting as shown in (2).

Ethereum platform preserves a decentralized ecosystem for developers to create products using the Ethereum Virtual Machine (EVM) which is powerful and embedded within each full blockchain network as shown in Figure 2. The smart contracts bytecodes are executed through EVM. Interacting with the EVM via smart contracts is likely more costly than traditional servers. However, numerous use cases are favored using EVM rather than conventional servers.

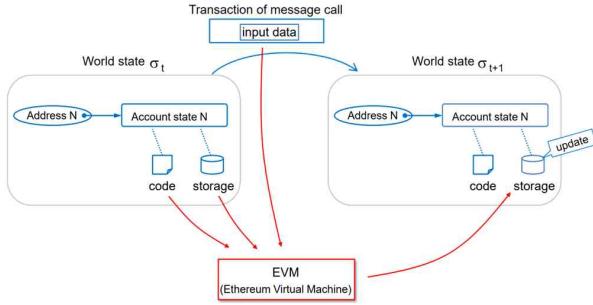


Figure 2. Ethereum virtual machine

### III. Secure Collaborative Learning with Blockchain Scheme

#### A. Segregated Learning Activities and Decentralized Revenue

**Segregated Learning Activities.** The users are data owners who hold a considerable amount of individual training data, while the AI provider stores the deep learning model that can be accessed by authorized users. The updated gradient are collected by the provider regularly, which later to be employed to calculate the final aggregation value. This process is repeated as long as necessary for improving the global model.

At the beginning of the process for each round, the aggregation server stores a global model in the cloud server. Then, the server roughly mapping the users available in the network along with the dynamic rules.

Eventually, Figure 3 describes the performance of loss over time training from multiple users with the scattered datasets. The loss for the lowest number of devices at 1st epoch is 1.882 and continued to decrease at the 50th epoch by 0.191. The same thing occurred in 15 devices with losses of 1.071 (1st epoch) and 0.206 (50th epoch). Moreover, the average value of loss for all devices is 0.618. The loss decreases as the number of epochs increases for all devices.

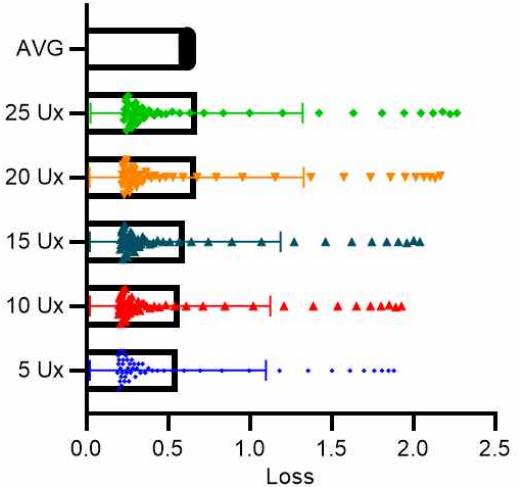


Figure 3. The performance of the variance number of segregated learning users

It can be concluded that most of the losses occurred in the early stages of training per cycle. The differences might be significant if the number of devices implemented is extremely large.

Figure 4 presents the distribution points of an average loss of SL. In short, the average loss for every device is almost identical. The number of devices involved in our scheme slightly affects the loss. Whilst, Fig. 5 depicts the comparison amount of Ethereum gas used between smart contract manager SC6 and users C6. The provider consumes gas stably for every transaction, while the clients produce gas depend on their contribution which is stated in the smart contracts.

#### B. Traceability and Privacy Concerns with Potenial Solutions

In the segregated learning, when the users deploy their transaction that consists of a cipher to encrypt the information to the AI provider, the observer can impose the dataset knowledge by adopting active and passive inference attacks. Yet, the performance of the adversary decreases with an increasing number of users.

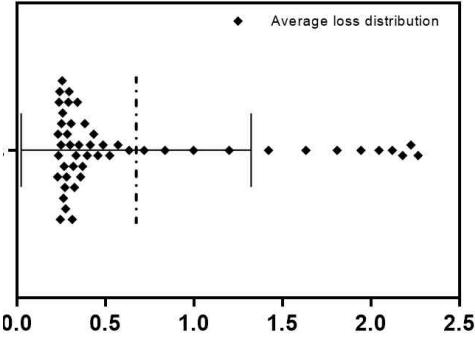


Figure 4. Average loss distribution of SL

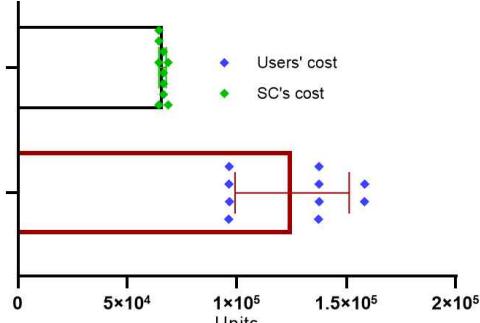


Figure 5. The gas used by users and SC

In short, as batch sizes increase, this type of attack produces more false positives. More promptly, the size of the batch influences the precision of the adversary. The larger the size of a batch, the lower the precision from the adversary's side.

**Solutions.** In the segregated learning system, every users and model provider have a pair of public keys ( $A_n, B_n$ ). The  $A_n$  is used to generate a one-time public key for transaction, and  $B_n$  is attached to the transaction as a tracking pointer.

We assume User  $A$  ( $U_a$ ) wants to conduct transactions with User  $D$  ( $U_d$ ). In concurrent User  $C$  ( $U_c$ ) and User  $B$  ( $U_b$ ) want to make transactions as well. User  $U_a$  then wants to use *Mi6er* services to hide his identity. The *Mi6er* then combines the  $A$   $U_a$  transaction  $TX_a = T(U_b \oplus U_c \oplus U_d \oplus U_a)$ .

There are two kinds of pub keys  $A_n$  is used to create a stealth address, and  $B_n$  is used to searching the transaction as shown in the following equations:

$$P = Hs(rA\alpha)G + B\alpha$$

$$P' = Hs(a\alpha R)G + B\alpha, \text{ then}$$

$$a\alpha R = a\alpha rG = rA\alpha; P' = P$$

The provider uses  $P$  as a destination key for the output and attaches the new value  $R = rG$  into the transaction. The data with attachments to  $P$  and  $R$  values are stored into shared storage after being validated by the miner. The recipient later checks every transaction using his private key  $(a\beta, b\beta)$  and calculates the new  $P'$ , and compares the value  $P$  received with the value  $P'$ . Finally, an unlinkability transaction is preserved.

## IV. Conclusion

Dissacotiation transactions in segregated learning with Ethereum smart contract as an incentive mechanism has been presented in this paper. The system goals are to preserve users' privacy by cutting off the role of a centralized machine to do the training. The provided scheme brings privacy challenges, yet it can be overcome with the solutions that have been outlined in this paper.

## Acknowledgment

This research was supported by the Republic of Korea's MSIT (Ministry of Science and ICT), under the High-Potential Individuals Global Training Program (2020-0-01596) supervised by the IITP (Institute of Information and Communications Technology Planning & Evaluation) and partially supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2018R1D1A1B07048944).

## [References]

- [1] Koloskova, A., Stich, S. U., & Jaggi, M. (2019). Decentralized stochastic optimization and gossip algorithms with compressed communication. arXiv preprint arXiv:1902.00340.
- [2] Weng, Jiasi, et al. "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive." IEEE Transactions on Dependable and Secure Computing (2019).
- [3] Ezhilchelvan, P., Aldweesh, A., & van Moorsel, A. (2020). Non blocking two phase commit using blockchain. Concurrency and Computation: Practice and Experience, 32(12), e5276.
- [4] Zhang, Q., Yang, L. T., Yan, Z., Chen, Z., & Li, P. (2018). An efficient deep learning model to predict cloud workload for industry informatics. IEEE Transactions on Industrial Informatics, 14(7), 3170-3178.

# FakeText:FastText 기반의 한국어 가짜 뉴스 분류 모델

강효은, 심준석, 김용수, 홍윤영, 김호원\*

부산대학교

## *FakeText:FastText-based Korean Fake News Classification Model*

Hyoew Kang, Junseok Shim, Yongsu Kim, Yoonyoung Hong, Howon Kim\*

Pusan National University

### 요약

최근 주식시장의 열기가 뜨거워짐에 따라 주식시장의 변동성을 예측할 수 있는 뉴스의 영향력도 커지고 있다. 동시에 기업가치를 왜곡하는 가짜 정보 확산이 사회적으로 큰 영향을 미치고 있다. 해외의 경우 가짜 뉴스 유포 문제를 해결하기 위해 구글과 페이스북 등 유명 IT기업들에서 솔루션을 제안하고 있으며, 국내의 경우에도 텍스트 마이닝, 머신러닝/딥러닝을 활용한 가짜뉴스 판별에 대한 연구가 활발히 진행 중이다. 본 논문은 한국어 가짜 뉴스를 판별하기 위해 형태소 단위로 학습된 워드 임베딩 모델인 FastText과 딥러닝 모델을 활용한 한국어 가짜뉴스 판별 모델을 제안한다.

## I. 서론

가짜뉴스는 정치 및 경제적 이익을 위해 의도적으로 언론 보도의 형식을 한 거짓 정보이다. 가짜뉴스는 의도적으로 만들어지고 오해를 부를 수 있는 허위정보, 고의로 조작한 정보를 사실을 가장하는 거짓정보, 근거없이 퍼지는 소문인 루머 등의 형태로 퍼지기도 한다[1].

2016년 미국 45대 대선 과정에서 페이크 뉴스가 광범위하게 파급되어 선거에 영향을 미쳤으며[1], 미국뿐만 아닌 우리나라에서도 최근 코로나19 재확산에 따른 방역 당국이 정치적인 목적으로 진단 검사 결과를 조작하고 있다는 가짜뉴스 사례가 유포되는 등 사회적으로 큰 영향을 미치고 있다. 최근 국내 주식시장의 열기가 뜨거워지면서 주가조작을 위한 가짜뉴스도 확산되고 있다. 주로 뉴스를 통해 기업의 내재가치를 분석하는 투자자들은 언제든지 주가 조작 피해 위험에 노출될 수 있다.

가짜뉴스의 유형은 기사 제목과 본문이 부정

합한 경우와 본문 중 맥락에 관계가 없는 내용으로 나눌 수 있다. 본 연구에서는 본문의 흐름에서 벗어나는 왜곡된 문장을 탐지하여 가짜뉴스 여부를 판별하는 모델을 제안한다.

본 논문에서는 성능이 우수하다고 알려진 공개용 한국어 형태소 분석기 중 하나인 Mecab-ko[7]와 FastText[5]을 이용하여 뉴스 데이터의 언어 표현을 학습한다. 최종적으로 딥러닝 모델을 이용하여 가짜뉴스 여부 판별을 진행한다.

### 1.2 관련 연구

구글은 가짜 리뷰 탐지를 위해 SVM을 사용해 리뷰어의 행동 데이터와 리뷰를 결합하여 분석하는 방법을 소개하였으며[2], 페이스북의 FiB[8]는 자바스크립트 기반으로 링크, 포스트, 이미지 정보를 추출하고 이를 딥러닝을 통해 기사의 원출처를 확인하여 가짜뉴스를 판별한다.

국내의 경우 Yun Tae-Uk[3]는 토퍽 모델

링 기법을 사용하여 가짜뉴스를 분류하다. Ye-Chan Ahn[4]는 한국어 BERT 사전 학습 모델에 Fine tuning을 적용한 가짜뉴스를 판별한다.

### 1.3 워드 임베딩

워드 임베딩(Word embedding)이란 단어를 벡터로 표현하는 방법으로, 단어를 희소 표현과 밀집 표현으로 구분된다. 희소 표현에는 대표적으로 표현하고자 하는 단어의 인덱스의 값만 1로 표현하고 나머지 인덱스를 0으로 표현하는 원-핫 벡터 표현 방법이 있다. 해당 방법을 사용하면 단어의 개수가 늘어남에 따라 벡터의 차원이 한없이 커진다는 단점이 있다.

한편 밀집 표현에 해당하는 방식에는 대표적으로 FastText[5]가 있다. FastText는 주변 단어와 단어의 부분 단어(subword)를 학습하여 학습 말뭉치에 없는 Out of Vocabulary(OOV)에 해당하는 단어의 임베딩 값을 얻을 수 있다. 본 연구에서 사용하는 딥러닝 모델은 FastText의 임베딩 결과 값을 입력으로 받아 가짜뉴스 여부를 판별한다.

## II. 데이터 전처리

### 2.1 데이터 세트

본 연구는 실제로 국내에서 이슈가 되었던 가짜뉴스를 판별하는 모형을 만들기 위해 다수의 선행 연구에서 활용되었던 SNU Factcheck와 뉴스톱, 팩트체크넷 사이트에서 총 2,066개의 데이터세트를 구축하였다. 진짜뉴스는 1,056개, 가짜뉴스는 1,010개를 수집하여 실험 데이터세트를 완성하였다. 실험을 위한 데이터 분할은 Train 1,446개, Validation 310개, Test 310개를 사용하였다.

### 2.2 데이터 전처리

본 연구에서는 형태소 분석을 이용해 어휘에서 의미적(semantic)인 부분이 아닌 것을 제거하였다. 형태소 분석에는 오픈 한국어 형태소 분석기인 Mecab-ko을 사용하였다.

## III. FastText 기반 가짜뉴스 판별 모델

### 3.1 FastText

한국어는 교착어에 속하기 때문에 다양한 접사들이 하나의 어근에서 약 60여 가지의 단어로 파생될 수 있다[6]. 한국어와 같은 교착어는 어절 단위로 그대로 학습할 경우 개별 단어의 빈도가 대부분 낮기 때문에 유의미한 정보를 얻기 어렵다. 본 연구에서는 n-gram 알고리즘을 적용하여 형태학적(morphological) 특성을 벡터 값에 반영할 수 있는 FastText를 사용하였다. 오탈자 및 오류가 거의 없는 한국어 위키 피디아 문서를 기반으로 2백만개의 단어에 대하여 300차원으로 한국어 어휘 임베딩을 학습한 사전학습 모델을 사용하였다.

### 3.2 지도학습 기반 가짜뉴스 판별 딥러닝 모델

본 연구에서는 본문의 맥락에서 벗어나는 문장에 대하여 가짜뉴스 판별을 하기 위해 랜덤 포레스트, CNN, Bidirectional LSTM 3가지 분류 모델을 구현하였다.

## IV. 실험 결과

본 논문에서 제안한 가짜뉴스 판별 모델의 성능 검증을 위해 표 1과 같이 정확도와 FPR, FNR을 사용하였다. FPR은 정상 뉴스를 가짜 뉴스로 오분류한 비율을 나타내며 FNR은 가짜 뉴스를 정상 뉴스로 판단한 비율을 나타낸다. FastText와 BiLSTM을 결합한 모델이 우수한 성능을 보임을 확인할 수 있다.

모델 성능	랜덤포레스트	CNN	BiLSTM
Acc (%)	92.80	97.10	<b>98.90</b>
FPR	0.1245	0.0390	<b>0.0160</b>
FNR	0.3317	0.0544	<b>0.0213</b>

표 1. 가짜뉴스 판별 모델 평가 결과

## V. 결론

본 연구는 FastText 워드임베딩 모델을 이용하여 한국어 뉴스의 어휘 특징을 추출하고, 딥러닝 모델을 적용하여 가짜뉴스 여부를 판별하였다. 본문이 전반적으로 정상적인 정보를 전달하고 일부 문장이 거짓 정보를 의미하는 데이

터에 대하여 가짜뉴스 판별 모델을 설계하였다. 추후 본 연구에서 구현한 모델을 사용하여 구독자 또는 정보 수용자들이 뉴스의 맥락을 통해 사실 관계를 합리적으로 이해할 수 있는 플랫폼으로 확장하여 구현하고자 한다.

## 감사의 글

본 연구는 과학기술정보통신부 및 정보통신 기획평가원의 대학ICT연구센터지원사업의 연구 결과로 수행되었음 (IITP-2021-0-01797), 블록체인 기반 및 플랫폼 분야 핵심기술 개발 및 미래 혁신인재 양성

- [6] 김한샘.현대국어사용빈도조사.Vol.2.국립국어원, 2005
- [7] 은전한닢 Mecab-ko, <https://bitbucket.org/eunjeon/mecab-ko/src/master/>
- [8] FiB, “FiB: Lets stop living a lie,” 2016.

## [참고문헌]

- [1] 윤성옥. 가짜뉴스의 개념과 범위에 관한 논의. 언론과법, 2018, 17.1: 51-84.
- [2] MUKHERJEE, Arjun, et al. Fake review detection: Classification and analysis of real and pseudo reviews. UIC-CS-03-2013. Technical Report, 2013.
- [3] YUN, Tae-Uk; AHN, Hyunchul. Fake News Detection for Korean News Using Text Mining and Machine Learning Techniques. Journal of Information Technology Applications and Management, 2018, 25.1: 19-32.
- [4] AHN, Ye-Chan; JEONG, Chang-Sung. Natural language contents evaluation system for detecting fake news using deep learning. In: 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2019. p. 289-292.
- [5] BOJANOWSKI, Piotr, et al. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 2017, 5: 135-146.

# Information Security Hacking Defense Intelligence and Misinformation Protection Algorithms

Seong-Kyu Kim\*

\*Assistant Professor (Tenure Track) of Department of Information Security, Joongbu University, Gyeonggi-do Goyang-Si 10279, Republic of Korea.

Department of Public Policy and Information Technology, Seoul National University of Science and Technology, Seoul 01811, Korea.(e-mail : [skkim@joongbu.ac.kr](mailto:skkim@joongbu.ac.kr) or [guitara77@gmail.com](mailto:guitara77@gmail.com))(e-mail : [skkim@joongbu.ac.kr](mailto:skkim@joongbu.ac.kr) or [guitara77@gmail.com](mailto:guitara77@gmail.com) )

Jun-Ho Huh\*\*

\*\* Assistant Professor (Tenure Track) of Department of Data Informatics, (National) Korea Maritime and Ocean University, 727, Taejong-ro, Yeongdo-gu, Busan, Republic of Korea (e-mail : [72networks@puskyong.ac.kr](mailto:72networks@puskyong.ac.kr) or [72networks@kmou.ac.kr](mailto:72networks@kmou.ac.kr) )

## Abstract

This paper has seen a rapid increase in services and users over the Internet. As a result, cyberattacks are increasing, and information leakage and financial damage are occurring. The government, public institutions, and companies use signature-based detection rules to respond to known malicious codes during these rapid cyber attacks, but it takes a long time to generate and verify signature-based detection rules. To address this problem, in this paper, we propose and develop a signature-based detection rule generation and verification system using signature extraction and traffic analysis techniques through latent Dirichlet allocation algorithms. Experiments on the developed system have shown that we generate and validate detection rules much faster and more accurately than existing ones.

## I. Introduction

This paper introduces and operates various network security systems such as firewall, intrusion detection system (IDS), intrusion prevention system (IDS), and web firewall (WAF) to cope with cyber attacks through the Internet. Despite the active study of various AI-based and behavior-based malware detection techniques, most security systems use signature-based detection methods that boast high detection performance in known malware detection [1]. Signature-based detection methods inevitably imply an abstract definition of malicious code

called detection rules. The effective definition of detection rules requires professional knowledge of networks, security, operating systems, etc. The operation of inappropriate detection rules can cause numerous false positives, causing security system performance degradation, and even paralyzing the entire network with network security systems installed. In this paper, we propose a system that allows not only security experts but also quasi-experts to generate and validate detection rules quickly and accurately to efficiently respond to rapidly increasing malware [2]. The proposed technique is to automatically generate snort detection rules

using research from existing signature extraction and traffic analysis results, and to validate detection rules generated by deploying IDS servers in virtual environments [3].

## II. Related Research

### 2.1 Signature-based Detection Rules

Among the signature-based detection rules, we analyze snort detection rules, which are the most commonly used globally and are adopted as TTA standards and are mostly supported by domestic network security equipment. In this paper, we define the requirements needed to design a system that automatically generates and validates snort detection rules based on the adopted TTA standard. Snort detection rules are logically divided into detection rule headers and detection rule options. In the detection rule header, each configuration is separated by a space, the detection rule header and options are separated by square brackets, and the options and options of the detection rule are separated by semicolons [4]. Options consist of optional keys and optional values, separated by colons. And as a result of analyzing the detection rules, the use of many options may increase detection performance depending on the traffic it detects, but vice versa, it may increase the false detection rate. Therefore, we minimize the automatically generated options and implement them so that users can add them arbitrarily.

### 2.2 Malicious Signature Extraction Study

Many studies have been conducted to automatically extract malicious signatures from malicious code or malicious traffic [5]. built two honeypods to extract malicious

signatures from the traffic generated by the worms generated using Metasploit, and then used the Longest Common Substring (LCS) algorithm in the traffic sent and received. Recent work has extracted signatures using Voting Experts algorithm and Ranking algorithm using unsupervised learning-based segmentation algorithm from silver malicious traffic and malicious code analysis using Denising Auto-Encoder algorithm from deep trust neural networks (DBNs).

## III. Research method

### 3.1 Study Design

Topic modeling algorithms have been used to extract keywords from social networks to track changing issues, analyze papers published in papers to identify topics that draw attention from time to time, extract topics from newspaper articles, and analyze changes in issues from time to time. As such, topic modeling techniques are algorithms that have been used to analyze trends for a particular field and have proven their performance [6]. Topic modeling algorithms include VectorSpace Model (VSM), Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (pLSA), and Latent Dirichlet Allocation (LDA). In this paper, we propose a system that leverages LARGen techniques developed based on latent Dirichlet allocation algorithms to perform signature extractions necessary for generating detection rules, and uses them to automatically generate snort detection rules and validate generated snort detection rules. The latent Dirichlet allocation (LDA) algorithm is a topic modeling algorithm designed to infer topics inherent in text documents. Recently, it has been widely used as a statistical topic model for finding topics in very large data

(BigData). This section briefly summarizes the LDA [Fig. 1].

doc1		doc1		doc1		doc2		doc2		doc2	
word	apple	banana	apple	dog	dog	word	apple	banana	apple	dog	dog
topic	B	B	???	A	A	topic	B	B	???	A	A
word	cute	book	king	apple	apple	word	cute	book	king	apple	apple
topic	B	B	B	B	B	topic	B	B	B	B	B

[Fig. 1] How to proceed with the learning

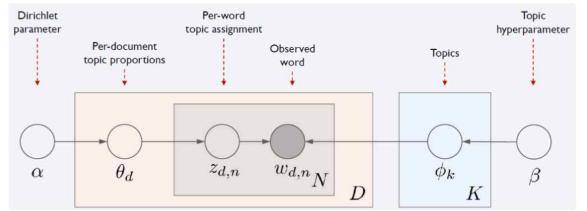
### 3.2 Research Architecture

The detection rule generation system provides an easy environment to analyze malicious traffic that requires generating detection rules, automatically recommends candidate signatures, and generates detection rules according to snort grammar. In the Recommended Detections string window, select the flow you want to extract the signature from the Malicious Packet File Management window and click the "Extract Malicious Signature" button to ensure that the program has successfully entered the database[7].

We show a list of malicious signatures extracted using . The Edit Detection Rule window also displays the recommended detection string window in the Recommended Detection String window.

When selected, the extracted signature automatically analyzes the corresponding flow information and shows the results of the snort detection rule, which is generated to fit the detection rule format. It can be modified by the user arbitrarily. In addition, the Detection Rule Management window contains a list of detection rules saved after completing the detection rule editing in the Detection Rule Editing window and a list of detection rules applied to real-world information security systems. The initial detection rules generated here can be applied

to security equipment after verification is performed [Fig. 2].



[Fig. 2] Dirichlet Distribution Inference Process

## IV. Future research and conclusions

In this paper, we propose a system that allows not only experts but also quasi-experts to generate and validate detection rules quickly and accurately, and even experimentally through real-world development. The proposed and developed system can generate detection rules regardless of the type of malicious code if only payload exists in malicious traffic that wants to generate detection rules. System experiments have allowed us to generate and validate detection rules normally, and with this system, there are differences depending on system users and detection rules, but we have been able to generate and validate detection rules within an hour. With the developed detection rule generation and verification system, experts will be able to generate and validate detection rules much faster than existing ones, and semi-professionals who are not yet skilled will be able to generate and validate detection rules more easily than existing ones.

## [Reference]

REliability Enginerring Workshops, Oct, 2017.

- [1] Myung-Hoon Kang, "Completion of information security monitoring system with complete control and security of the IDS pattern matching techniques discussed in Big Data Analytics", pp 40–41. 5 2013.
- [2] Jae Chan Yoo, "A Study on the Protection for Corporation Information Using Scenario Technique," The Graduate of SungKyunkwan University, pp. 14–16, August 2012.
- [3] Kelly M, Mark Nicolett, Oliver Rockford, "Magic Quadrant for Security Inofrmation and Event Management", Gartner Group, pp. 2–8, June 2014.
- [4] DongSung im, YoungMin Kim, "Enhancement of internal Control by expanding Security Information Event Management System", Korea Society of Computer and Information, pp 36–37. 8 2015.
- [5] Zhuo. Zhang, Zhibin Zhang, Patrick P.C.Lee, Yunjie Liu and Gaogang Xie "ProWord: An unsupervised approach to protocol feature word extraction", in INFOCOM, 2014 Proceedings IEEE, pp. 1393–1401, July, 2014.
- [6] Omid E. David and Nathan S. Netanyahu, "DeepSign: Deep learning for automatic malware signature generation and classification", International Joint Conference on Neural Networks, July, 2015.
- [7] Fabrizio Biondi, Francois Dechelle and Axel Legay "MASSE: Modular Automated Syntactic Signature xtraction", IEEE International Symposium on Software

# Rethinking Edge AI Architecture

Elizabeth Nathania Witanto, Yustus Eko Oktian, Sang-Gon Lee\*

Dongseo University

\*Corresponding Author

## Abstract

The proliferation of 5G connections and Internet-of-Things (IoT) drives to the data explosion generated by the massive number of IoT devices (e.g., sensors, cameras, etc.) and end-devices (e.g., smartphones, tablets, etc.). Rapidly increasing data volume brings more advantages and challenges to Artificial Intelligence (AI) development. Data is the heart of AI. The conventional way to process generated data is to transfer it over the internet to the data center. Sending bulks of data from the IoT devices to the cloud data center causes high financial cost, transmission delay, and privacy leakage. It is about time that the technology trend is shifting to so-called Edge AI. In this paper, we explain and evaluate about three possible types of architecture for the Edge AI training process. There are centralized, decentralized, and distributed. In addition, we present the pros and cons of each architecture.

## I. Introduction

We are living in an era of rapid technological and communication development. Most recently, the proliferation of 5G connections and Internet-of-Things (IoT) drives to the data explosion generated by the massive number of IoT devices (e.g., sensors, cameras, etc.) and end devices (e.g., smartphones, tablets, etc.). Research from International Data Corporation (IDC) shows that the amount of data generated by connected IoT devices, forecast to grow to 41.6 billion by 2025, is expected to generate 79.4 zettabytes (ZB) of data [1]. This amount of data brings more advantages and challenges to Artificial Intelligence (AI) development. Data is the heart of AI. More training data will increase the accuracy results of the AI algorithm.

The conventional way to process generated data is to transfer it over the internet to the data center. However, there is a problem due to the concerns of

performance, cost, and privacy. Send bulks of data from the IoT devices to the cloud data center is highly non-trivial even with fast connections. The financial cost and transmission delay can be prohibitively high, and privacy leakage is a crucial concern.

It is about time that the technology trend is shifting to so-called Edge AI. Edge AI is a system that uses machine-learning algorithms to process data generated by a hardware device at the local level [2]. Thus, it will conduct the process closer to the IoT devices and data sources. Since the data does not need to be transferred through the internet, the device can make a real-time decision in a matter of milliseconds. According to Vector ITC Group, the latency of cloud computing would be seconds; with Edge AI, the times are reduced to less than 400ms [2]. The combination of edge computing and AI comes with several benefits compared to traditional cloud computing, including low-latency, less bandwidth consumption, privacy protection,

scalability, and adaptability [3].

In this paper, we explain and evaluate about three possible types of architecture for the Edge AI training process. There are centralized, decentralized, and distributed. Besides, we present the pros and cons of each architecture.

## II. Edge AI Architecture

The workflow in the AI environment is divided into two parts, training, and inference. In the training process, the algorithm gives and calculates the value of weights of the model. The output is the task result, and there is a loss function to evaluate the correctness of the result by calculating the error rate. The inference process happens after training. It tests the model by giving the input and shows the predictions. In the scope of our paper, we show possible types of architecture for Edge AI training only.

It is crucial to design and choose the suitable architecture to gain optimized advantages from Edge AI. Based on references [4] and [5], we conclude three types of architecture, centralized, decentralized, and distributed. We evaluate and present the pros and cons of each type of architecture. It is worth noting that there is some value to trade-off for each architecture. Figure 1 shows three types of architecture that we describe as follows:

1. *Centralized*: training of the model is happening in the cloud while edge devices such as surveillance cameras, traffic lights, smart watches, and smart phones send the training data to the cloud.

Pros:

*Low-computation latency.* The central server has enough resources (memory,

storage, energy) to compute the training algorithm. Therefore, it will reduce the computation-latency compare to constraint device with constraint resources.

*High-accuracy.* The computation that happens in a central server with enough resources (memory, storage, power) can train more data. More data will increases the accuracy of the results.

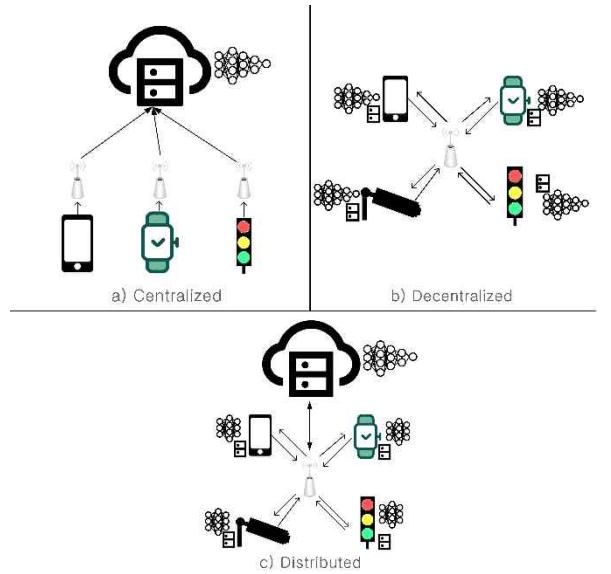


Figure 1. Edge AI Training Architecture.

Cons:

*High-communication latency and cost.* To send bulks of data from edge-devices to the central server will take time, not to mention the data's size. As a result, it will increase the communication latency alongside the cost. Another consequence is increased energy consumption.

*Data privacy problem.* To train the data in the central server, the data need to be transferred across the network. It is unavoidable the privacy issues.

2. *Decentralized*: The model's training is happening directly in each edge-device locally without sending data to the cloud. Between

edge-devices will communicate to exchange their local model.

Pros:

Since the computation happens locally, *the communication latency and cost will be reduced*. It also *decreases energy consumption*. However, it depends on the specification of each edge device. Another advantage is decentralized type will preserve data privacy.

Cons:

*High-computation latency.* The edge-devices are not designed with many resources (e.g., memory, storage, power). Due to these limited resources, it takes more time until the training result converged.

*Less accuracy.* The resource's limitation also limits the amount of training data. The edge-devices can not train as much as the central server can. Therefore, it affects the accuracy of the training results.

*Data redundancy.* The training data gathered by edge-devices might be the same or partially the same. It leads to data redundancy. The further effect is the waste of computation power. The devices might train the same data repeatedly.

*Distributed.* each edge-device trains the model locally, and periodically the central server will aggregate the local update from each device.

Pros:

Since the training happens locally as a decentralized type, distributed type inherits the advantages from it, such as less-energy consumption, less-communication latency, and cost. Besides, the accuracy will be higher since there is a central server that aggregates the local model update from

edge-devices in the network.

Cons:

The distributed type also inherits the cons from decentralized type such as data redundancy and high-computation latency. For the data redundancy problem, reference [6] suggested a solution called edge-caching. The training data will be stored in the cache. Later, when the edge-device captures the same data as that in the cache, it will not store the data to avoid redundancy and save computation power.

### III. Future Research Directions

As Edge AI proliferating, it comes with challenges and future research directions for fellow developers and researchers. We describe as follows:

- **Training data curation problem and reliability.** In Edge AI, we get the training data directly from many edge-devices distributed across the edge-network. It is different from the cloud-based AI technique that uses an available dataset that is already being curated and labeled. Therefore, it raises the problem of reliability of the training data. Besides, the devices will have a different environment and high device heterogeneity.
- **Training data completeness.** The training data distributed across the edge-devices. In some cases, such as distributed architecture, the data will be sent to the central server. There is a chance that there are stagger devices and some data not arrived in the central server. Therefore, it needed to do further research to ensure all the training data from edge-devices were not missing.

## IV. Conclusions

Architecture is the foundation of a system. It is vital to design and choose suitable architecture according to each system's needs for gaining maximized advantages from technology. In this paper, we evaluate three types of architecture, centralized, decentralized, and distributed. Besides, we present the pros and cons of each architecture. In the last section, we present Edge AI's future research directions that will be the next mission to accomplish for developers and researchers.

## Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant Number: 2018R1D1A1B07047601).

## [References]

- [1] Eden Estospace, “IDC forecasts connected IoT devices to generate 79.4ZB of data in 2025,” Jun. 22, 2019. <https://futureiot.tech/idc-forecasts-connecte-d-iot-devices-to-generate-79-4zb-of-data-in-2025/> (accessed Feb. 01, 2021).
- [2] Vector ITC, “Edge AI: The Future of Artificial Intelligence,” Aug. 12, 2020. <https://www.vectoritcgroup.com/en/tech-magazine-en/artificial-intelligence-en/edge-a-i-el-futuro-de-la-inteligencia-artificial/> (accessed Feb. 01, 2021).
- [3] Maximilian Bischoff, Johannes M. Scheuermann, Christoph Kiesl, and Julian Hatzky, “The Edge is Near: An Introduction to Edge Computing!,” Jun. 03,

<https://www.inovex.de/blog/edge-computing-introduction/> (accessed Feb. 01, 2021).

- [4] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, “Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing,” Proc. IEEE, vol. 107, no. 8, pp. 1738 - 1762, Aug. 2019, doi: 10.1109/JPROC.2019.2918951.
- [5] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, “Communication-Efficient Edge AI: Algorithms and Systems,” IEEE Commun. Surv. Tutorials, vol. 22, no. 4, pp. 2167 - 2191, 2020, doi: 10.1109/COMST.2020.3007787.
- [6] D. Xu, T. Li, Y. Li, T. Jiang, J. Crowcroft, and P. Hui, “Edge Intelligence: Architectures, Challenges, and Applications,” arXiv:2003.12172, p. 53.

# Towards Ethic-Friendly AI Architecture

Yustus Eko Oktian, Elizabeth Nathania Witanto, Sang-Gon Lee\*

Dongseo University

\*Corresponding Author

## 요 약

The state-of-the-art AI systems pose many ethical issues ranging from massive data collection to bias in algorithms. In response, this paper proposes a more ethic-friendly AI architecture by combining Federated Learning and Blockchain. We first discuss the requirements for an ethical AI system then show how our solutions can achieve more ethical paradigms. By committing to our design, adopters can perform AI services more ethically.

## I. Introduction

We long time believe that users only have constrained machine that is unable to train the machine-learning model. They also have a limited data size, which is not enough to produce a highly accurate training model. Therefore, many data from different users must be aggregated to a high-performance server, where the training takes place. In consequence, users lose control of their data once the data is transferred out from their devices. This data collection practice sometimes does not explicitly request user consent. Many companies use an “opt-out” mechanism instead of “opt-in”, which puts users on surprise when they realize such a data collection setting exists. Even worse, there is no regulation for companies when they conduct AI practices. Until recently, the public became aware of the importance of user privacy with the introduction of the GDPR law [1].

Even though massive data collection can be compelling, it is challenging to adjust the trade-off between the benefits and user privacy. For example, China’s social credit

system [2] can shape the business and citizen’s behavior towards better goals (in the view of the government). However, the citizen is at a disadvantage by losing freedom over this mass surveillance program. Moreover, AI is a black box system (in the current form), making it very tough to be debugged. This problem leads to many biases in AI algorithms. For instance, South Korea AI persona, Lee Luda [3], makes a controversy because she used offensive language targeting a minority community. Amazon AI recruitment tools also being shut down because it prefers men over women in selecting candidates [4]. As a result, researchers and AI practitioners must conduct AI services with ethics-in-mind, which always preserve human values.

This paper aims to seek solutions towards more ethic-friendly AI architecture by combining Federated Learning (FL) [5] and Blockchain [6]. FL preserves user privacy by training private user data on user local machines instead of sending them to the server. Meanwhile, the blockchain serves as a trusted platform to conduct the overall FL

process so that FL participants can collaborate in a secure, transparent, and fair manner. We also discuss the requirements for an ethical AI system and show that our solutions tackle the necessary components. By committing to our design, adopters can realize an ethic-friendly AI architecture.

## II. Requirements for Ethcial AI

Floridi and Taddeo [7] divides ethics of AI into three spheres: ethics of data, ethics of algorithms, and ethics of practices.

*Ethics of Data:* The ethics of data focuses on the ethical problems related to data, including generation, curation, processing, dissemination, sharing, and usage [7]. Tranberg et al. [8] recommends five principles to enforce data ethics: **R1**) human being at the center, **R2**) individual data control, **R3**) transparency, **R4**) accountability, and **R5**) equality.

*Ethics of Algorithms:* The ethics of algorithms addresses issues posed by the increasing complexity and autonomy of the AI algorithms [7]. High-Level Expert Group on Artificial Intelligence, which is an independent expert group that was set up by the European Commission, mentioned that AI algorithm must follow these ethical principles [9]: **R6**) respect for human autonomy, **R7**) prevention of harm, **R8**) fairness, and **R9**) explicability.

*Ethics of Practices:* The ethics of practice focuses on the pressing questions about the responsibilities and liabilities of people and organizations in charge of data, strategies, and policies of AI system [7]. Google provides a recommendation practices for AI [10], which includes **R10**) use a human-centered design approach, **R11**) rigorous

testing, and **R12**) continuous monitoring and updates.

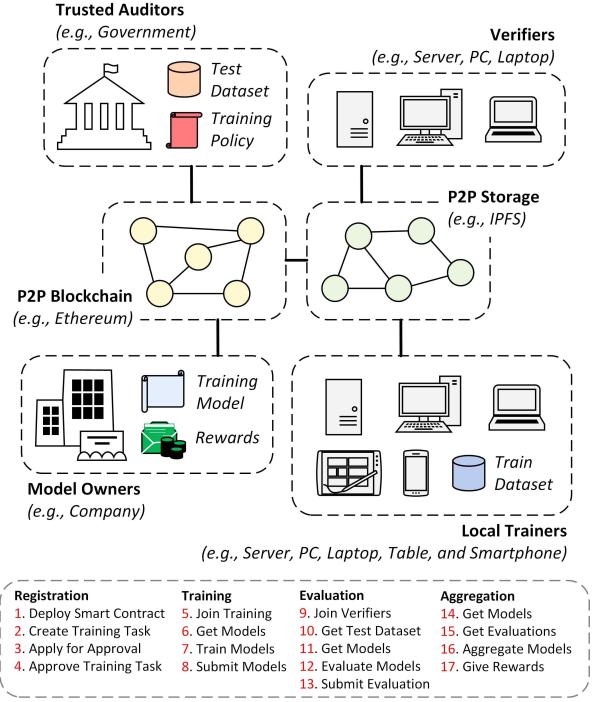


Fig 1. Our ethic-friendly AI system.

## III. Proposed Architecture

Using the previously mentioned ethic requirements as our foundation, we propose an ethic-friendly AI architecture as depicted in Fig 1. The proposed system comprises six components: model owners (e.g., AI companies), local trainers (e.g., users), verifiers (e.g., users or government personnel), trusted auditors (e.g., the government), peer-to-peer (P2P) blockchain, and P2P storage. All participants are authenticated and endorse the use of a reputation system in our system. The AI workflow is described as follows.

*Registration:* The government makes the digital representation of the training policy in smart contracts (Step 1). AI companies, as model owners, create an initial global model and prepare rewards for trainers. They then

create a training task in the smart contract (Step 2). After that, the companies request approvals from the government (Step 3). Before approving a task, the government must make sure that the proposal provides enough incentives for trainers. They also create a standardized test dataset suitable for the proposal (Step 4). The model parameters and the test dataset will be distributed to trainers and verifiers through the P2P storage. Meanwhile, the hash of the model and dataset is stored in the blockchain.

*Training:* Users can join the training as trainers by registering themselves in the smart contract (Step 5). They can then get the model from P2P storage (Step 6) and begin training using their local data (Step 7). When the training is complete, users submit the trained model through P2P storage while the hash is logged in the blockchain (Step 8).

*Evaluation:* Users or government personnel can register themselves as verifiers in the smart contract (Step 9). At each global epoch, the verifiers must get the test dataset (Step 10) and the trained local models (Step 11) from P2P storage. They then verify the accuracy of the trained models using the test dataset (Step 12). Once the evaluation finishes, the evaluation result is submitted to the smart contract (Step 13).

*Aggregation:* When a particular global epoch finishes, the companies get all of the trained local models from the P2P storage (Step 14). They then retrieve all of the associated evaluation scores from the smart contract (Step 15). Using the evaluation scores as a guideline, the companies aggregate the models according to their contributions (Step 16). For example, they may skip models with low accuracy as they are most probably trained with poisoned data

or low-quality data. During evaluations, verifiers use adversarial defense techniques to check if the model is trained with adversarial examples. Therefore, the companies must also skip models, which contains malicious flag from the verifiers. Once the aggregation is completed, the companies distribute the reward to all trainers and verifiers through the smart contract (Step 17).

## IV. Ethic-Friendliness Analysis

*Training distributedly using Federated Learning:* Users train their data locally on their devices and only send the model parameters instead of the private data to the server. The server then combines the trained local models into a single global model using an aggregation algorithm (e.g., Federated Averaging [5]). Using this approach, the user data do not leave the devices, and users still have control over their data (i.e., solving **R2**).

*Rigorous evaluation and auditing:* To ensure the quality of the trained models, they must be evaluated. For this purpose, we employ the government and volunteers as our verifiers.

The government must first create a standardized training policy for AI companies in the form of federal or international law (e.g., GDPR [1]). With this law, we can hold malicious persons or organizations accountable (i.e., solving **R4**). We can also ensure that the AI models will always benefit humans (i.e., solving **R1**, **R6**, and **R10**). Moreover, the government must produce a generalized test dataset to be used during the evaluation stage. Assuming that this standardized test dataset has a high variance to cope with all possible classes, then this test should mitigate the AI bias that may

happen during training (i.e., solving **R5** and **R8**).

The group of verifiers evaluates the submitted local models from users to detect potential poisoning attacks on each epoch. Attackers can intentionally train the local model with bad or low-quality data to reduce the global model's overall accuracy. Moreover, the attackers can also train the model with adversarial examples to make the global model misclassify particular targets. Once detected, the attackers will be punished economically or by law (i.e, solving **R7**, **R11**, **R12**)

*Logging training processes using the blockchain:* In our architecture, all of the training processes are logged in the blockchain (e.g., Ethereum [6]). Because of the chain-of-hashes introduced in the blockchain, the stored data in the blockchain becomes hard-to-tamper. All nodes must also include their digital transactions when storing data to the blockchain. Hence, malicious entities can be detected easily. Finally, all data in the blockchain is open for all the blockchain nodes. Hence, solving **R3** and **R9**.

*Sharing through distributed storage:* Because storing in the blockchain is quite expensive to perform, we can use distributed storage system (e.g., IPFS [11]) to store massive data (e.g., model parameters). Meanwhile, the corresponding metadata (e.g., the hash of the model) can be stored efficiently in the blockchain. Note that we refrain from using a centralized database due to trust issues that may persist in such a system.

## V. Conclusion

This paper proposed a more ethical AI

architecture through a combination of federated learning and blockchain technologies. The federated learning yielded promising solutions towards AI ethics in terms of data collection and training transparency. Meanwhile, the blockchain enhanced the AI ethics with its secure, transparent, and fair collaborative auditing platform. However, our proposal still does not solve AI's fundamental issues regarding its "black box" properties. More research towards "explainable AI" is still required in the future so that we as humans and AI supervisors can make a better decision on how to use AI.

## Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant Number: 2018R1D1A1B07047601).

## [참고문헌]

- [1] M. Goddard, "The EU general data protection regulation (gdpr): European regulation that has a global impact," International Journal of Market Research, vol. 59, no. 6, pp. 703 - 705, 2017.
- [2] K. Munro. (2018) China's social credit system could interfere in other nations' sovereignty. [Online]. Available: <https://bit.ly/3qSmFz8> [Accessed: 27-Jan-2021].
- [3] J. McCurry. (2021) South Korean AI chatbot pulled from facebook after hate speech towards minorities. [Online]. Available: <https://bit.ly/2NwoZgM> [Accessed: 27-Jan-2021].
- [4] Reuters. (2018) Amazon ditched ai

recruiting tool that favored men for technical jobs. [Online]. Available: <https://bit.ly/3sTehBc> [Accessed: 27-Jan-2021].

- [5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Artificial Intelligence and Statistics. PMLR, 2017, pp. 1273 - 1282.
- [6] G. Wood et al., “Ethereum: A secure decentralised generalised transaction ledger,” Ethereum project yellow paper, vol. 151, no. 2014, pp. 1 - 32, 2014.
- [7] L. Floridi and M. Taddeo, “What is data ethics?” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 374, no. 2083, p. 20160360, dec 2016.
- [8] P. Tranberg, G. Hasselbalch, B. K. Olsen, and C. S. Byrne. (2018) Dataethics - principles and guidelines for companies, authorities, and organisations. [Online]. Available: <https://bit.ly/3canj6V> [Accessed: 26-Jan-2021].
- [9] AI HLEG. (2019) Ethics guidelines for trustworthy AI. [Online]. Available: <https://bit.ly/3iTIRGs> [Accessed: 26-Jan-2021].
- [10] Google. (2020) Responsible AI practices. [Online]. Available: <https://bit.ly/3opdd4Q> [Accessed: 26-Jan-2021].
- [11] J. Benet, “Ipfs-content addressed, versioned, p2p file system,” arXiv preprint arXiv:1407.3561, 2014.

# Related-key Neural Distinguisher on Lightweight Block Ciphers SPECK-32/64, HIGHT, SIMECK-32/64 and CHAM-64/128\*

Erzhena Tcydenova\*, Byoungjin Seok\*, Changhoon Lee\*\*†

Dept. of Computer Science and Engineering,  
Seoul National University of Science and Technology  
(Graduate Student\*, Professor\*\*)

## Abstract

Neural networks have shown an excellent results in various areas such as image classification and natural language processing. Application of neural networks in cryptography has been around for many years, but results have not been significant until recently. At the CRYPTO'19 a new deep learning-based distinguisher on SPECK-32/64 was proposed, which was extended to a practical key recovery attack. The proposed distinguisher was constructed using the differential characteristics of the cipher and neural distinguisher performed better than classical one. Recently most researches are conducted on extending attack to other ciphers and constructing different attack scenarios. In this paper, we extend our method of constructing a neural distinguisher using related key characteristics to lightweight block ciphers SPECK, HIGHT, SIMECK and CHAM.

## I. Introduction

Security of lightweight cryptography is a big issue since it has to be resistant to all cryptanalytic attacks while providing efficient performance. A large number of lightweight block ciphers have been proposed so far, and standardization of lightweight block ciphers is still in progress. Lightweight block ciphers are designed with a relatively simpler structure and because of this they might become a target of new cryptanalytic attacks. One of the new attack methods has recently attracted interest in the field of cryptography which is Neural cryptanalysis [1]. Neural cryptanalysis showed possibility to analyze a cipher without a great amount of time and cryptographic knowledge [2]. In particular, a

research presented on CRYPTO'19 by Aron Gohr [3] proposed a key recovery attack conducted by constructing a differential neural distinguisher to the lightweight block cipher SPECK, and it achieved better results than conventional differential cryptanalysis. Further researches in this field are mostly conducted on applying attacks to other algorithms and building new attack scenarios.

In this paper, we extend our study on *Related-key neural distinguisher* [1] which is method of constructing neural distinguisher using relation between keys. We apply it to lightweight block ciphers SPECK-32/64, HIGHT, SIMECK-32/64 and CHAM-64/128. Also we apply Gohr's neural distinguisher on HIGHT, SIMECK-32/64 and CHAM-64/128 block ciphers and compare results of Gohr's distinguisher and Related-key neural distinguisher.

---

\* This paper was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1F1A1076468).

† Corresponding Author: chlee@seoultech.ac.kr

## II. Background

### 1.1. Related works

Neural Distinguisher proposed by Gohr on CRYPTO'19 is based on convolutional neural network model ResNet, the winning model of the image classification competition ILSVRC'15. Input of the distinguisher consist of two concatenated ciphertexts  $C_0||C_1$  converted to binary, where  $C_0 = \text{SPECK}_K(P_0)$  and  $C_1 = \text{SPECK}_K(P_0 \oplus (0x0040, 0x0000))$ . Output of the d is a result of determining whether the input data is a ciphertext or a random data. Their proposed neural distinguisher on Speck achieved better results than existing differential distinguisher and results are shown in the Table 1.

R	Model	Accuracy
5	Neural distinguisher	0.929
5	Differential distinguisher	0.911
6	Neural distinguisher	0.788
6	Differential distinguisher	0.758
7	Neural distinguisher	0.616
7	Differential distinguisher	0.591

**Table 1.** Differential Distinguisher on SPECK-32/64

Inspired by this work several studies were conducted. Anubhab Baksi et al. [4] proposed new neural network distinguisher scenario using the concept of an all-in-one differential that performs an attack using multiple differences. Tarun Yadav et al. [5] constructed distinguisher by extending classical differential distinguisher with high probability of  $r$ -round by using the  $s$ -round machine learning distinguisher and it distinguishes  $r+s$  rounds.

### 1.2. Related-key attack.

A related-key attack is a type of cryptographic attack that exploits some mathematical relation between keys. Relation is a some function  $F$  known or selected by an attacker where  $K_1 = F(K_0)$ . One of the its

forms is differential relation between keys such  $K_1 = K_0 \oplus \Delta$ . This relation exploits properties of differential distribution when plaintexts are encrypted with related keys [6].

### 1.3. Lightweight block ciphers

SPECK-32/64 is block cipher that has a Feistel structure with 32-bit block, 64-bit key and 32 rounds. Round function consists of simple operations: bitwise XOR, addition modulo  $2^{16}$  and circular shifts [7]. Block cipher HIGHT has ARX based Feistel structure. It consists of operations such as XOR, addition modulo  $2^8$ , and left bitwise rotation. HIGHT has 32 rounds, 64-bit block size and 128-bit key size [6]. SIMECK-32/64 is block cipher with Feistel structure. It consists of circular shift, bitwise AND and bitwise XOR operations and it has 32-bit block, 64-bit key and 32 rounds [8]. CHAM-64/128 is a block cipher with 64-bit block, 128-bit key, 80 rounds and it has Feistel structure. It consist of bitwise XOR, addition modulo  $2^{16}$  and circular shifts [9].

## III. Related-key Neural Distinguisher

Neural distinguisher proposed by Gohr was constructed by exploiting differential characteristics of the cipher and it had better performance than classical differential distinguisher. New method of constructing neural distinguisher which uses differential relation between keys - Related-key neural distinguisher was proposed in [1].

Input of the model consist of ciphertext pair  $(C||C')$  where  $C = \text{CIPHER}(P_K)$  and  $C' = \text{CIPHER}(P_{K \oplus \Delta})$  encrypted using related key with input differential  $\Delta$ . Output of the model is a result of distinguishing ciphertext

of target cipher from random data.

Dataset for our distinguisher is generated using *Algorithm 1*. The algorithm generates labeled ciphertext pair ( $C||C'$ ) converted to binary which is result of encryption with related keys.

---

**Algorithm 1**

---

```

Input: Data size:  $m$ , number of rounds:  $n$ , input differential:  $\Delta$ 
Output: Binary data:  $BD$ , labels:  $X$ 
1: Generate random sequences  $P = (P_0, \dots, P_m)$ ,  $K = (K_0, \dots, K_m)$ ,  $X = (X_0, \dots, X_m)$ 
2: for  $i = 0; i < m; i \leftarrow i + 1$  do
3:   if  $X_i == 0$  then
4:     Generate random  $C_i, C'_i$ 
5:   else if  $X_i == 1$  then
6:      $C_i = \text{CIPHER}_{K_i}^n(P_i)$ 
7:      $C'_i = \text{CIPHER}_{K_i \oplus \Delta}^n(P_i)$ 
8:   end if
9: end for
10:  $BD \leftarrow \text{Binary}(C||C')$ 
11: return  $BD, X$ 
```

---

Using generated dataset we train neural network and get accuracy of distinguisher for target number of rounds by *Algorithm 2*. Neural network model used is SE-ResNet which is ResNet model with SE block from SENet. SENet is winning model of classification challenge ILSVRC'17.

---

**Algorithm 2**

---

```

Input: Number of rounds:  $n$ , input differential:  $\Delta$ , epochs:  $e$ 
Output: Best validation accuracy:  $acc$ 
1: Train data size =  $m$ , validation data size =  $m'$ ,
2: Number of rounds =  $n$ ,  $tmp = 0$ 
3: Train data:  $Data_{train} \leftarrow$  Algorithm 1 ( $m, n, \Delta$ )
4: Validation data:  $Data_{val} \leftarrow$  Algorithm 1 ( $m', n, \Delta$ )
5: for  $i = 0; i < e; i \leftarrow i + 1$  do
6:    $acc \leftarrow \text{SE-ResNet}(Data_{train}, Data_{val})$ 
7:   if  $acc > tmp$  then
8:      $tmp \leftarrow acc$ 
9:   end if
10: end for
11:  $acc \leftarrow tmp$ 
12: return  $acc$ 
```

---

Choice of input differential is an important part of differential related-key attack. In most cases the differential that shows good probability is a one-bit difference. In the experiments differences were chosen by exhaustive search. *Algorithm 2* was run for every one-bit difference and the differential that had the best result was chosen as a final input differential.

## IV. Experiments and results

4.1. Differential neural distinguisher on HIGHT, SIMECK and CHAM.

We applied Gohr's differential neural disting uisher on block cipher HIGHT, SIMECK-32/64 and CHAM-64/128. Ciphertext pairs for input of the model are generated as follows:  $C_0 = \text{CIPHER}_K(P_0)$  and  $C_1 = \text{CIPHER}_K(P_0 \oplus \Delta)$ .

Input differentials are one-bit differentials that had best accuracy:

- SPECK: (0x0040, 0x0000)
- HIGHT: (0x00800000, 0x00000000)
- SIMECK: (0x0400, 0x0000)
- CHAM: (0x8000, 0x0000, 0x0000, 0x0000)

Data size for training =  $10^6$ , for validation =  $10^5$ . Results are shown in the Table 2.

R	SPECK	HIGHT	SIMECK	CHAM
1	1.0	1.0	1.0	1.0
...	...	...	...	...
5	0.9065	1.0	1.0	1.0
6	0.7540	1.0	0.9997	0.9999
7	0.5078	0.9999	0.9970	1.0
8	-	0.9990	0.9720	0.9999
9	-	0.7472	0.7888	1.0
10	-	-	0.6125	0.9999
...	-	-	-	...
27	-	-	-	0.5593
28	-	-	-	0.5603

**Table 2.** Results of Gohr' neural distinguisher on HIGHT, SIMECK and CHAM

4.2. Related-key neural distinguisher on SPECK, HIGHT, SIMECK and CHAM.

Ciphertext pairs are generated as follows:  $C = \text{CIPHER}(P_K)$  and  $C' = \text{CIPHER}(P_{K \oplus \Delta})$ . Input differentials:

- SPECK: (0x0040, 0x0000, 0x0000, 0x0000)
- HIGHT: (0x0000000080000000, 0x0000000000000000)
- SIMECK: (0x0000, 0x0000, 0x0000, 0x1000)
- CHAM: (0x0000, 0x0000, ..., 0x0000, 0x4000)

Training data size -  $10^6$ , validation -  $10^5$ .

Results are shown in the Table 3.

R	SPECK	HIGHT	SIMECK	CHAM
1	1.0	1.0	1.0	1.0
...	...	...	...	...
6	1.0	1.0	1.0	1.0
7	0.9773	0.9999	0.9999	0.9999
8	<b>0.8467</b>	0.9999	0.9998	1.0
9	<b>0.5920</b>	0.9998	0.9957	1.0
10	-	<b>0.9991</b>	0.9706	1.0
11	-	<b>0.7493</b>	<b>0.7818</b>	1.0
12	-	-	<b>0.6088</b>	0.9999
...	-	-	-	...
27	-	-	-	0.6496
28	-	-	-	0.5348

**Table 3.** Results of Related-key neural distinguisher

Related-key neural distinguisher on SPECK was able to improve a number of attacked rounds by 2 compared to Gohr's results. Similarly to results of SPECK, proposed distinguisher on HIGHT and SIMECK attacked 2 more rounds compared to Gohr's distinguisher. Results of proposed related-key distinguisher on CHAM did not show improvement of attacked rounds. By these results, we can assume that distribution of related-key characteristics have higher non-random behavior than differential characteristics for SPECK, HIGHT and SIMECK and similar for CHAM.

## V. Conclusion

In this paper we extended our study [1] on neural distinguishers. We applied Related-key neural distinguisher on lightweight block ciphers SPECK, HIGHT, SIMECK and CHAM. Application of this method which uses differential relation between keys showed to have better results than Gohr's distinguisher for SPECK, HIGHT, SIMECK and similar for CHAM. These results show that we can improve performance of neural distinguisher by constructing new attack scenarios using different properties of ciphers.

## [Reference]

- [1] E. Tcydenova, Cryptanalysis of lightweigh t block ciphers based on neural distinguisher, Master's thesis, 2021.
- [2] E. Tcydenova, M. Cho, B. Seok, C. Lee, Application of Neural Differential Distinguisher on SIMON-32/64, CISC-S, 2020
- [3] A. Gohr, Improving attacks on round-reduced speck32/64 using deep learning, Annual I nternational Cryptology Conference, Springer, Cham, 2019.
- [4] A. Baksi, J. Breier, X. Dong, C. Yi, Machi ne Learning Assisted Differential Distinguishe rs For Lightweight Ciphers, IACR Cryptol. eP rint Arch, 2020
- [5] T. Yadav, M. Kumar, Differential-ML Dis tinguisher: Machine Learning based Generic E xtension for Differential Cryptanalysis.
- [6] B. Koo, D. Hong, D. Kwon, D, Related-ke y attack on the full HIGHT, In International Conference on Information Security and Crypt ology, Springer, Berlin, Heidelberg.
- [7] R. Beaulieu, D. Shors, J. Smith, S. Treat man-Clark, B. Weeks, L. Wingers, The SIMO N and SPECK lightweight block ciphers, Proc eedings of the 52nd Annual Design Automatio n Conference, 2015.
- [8] G. Yang, B. Zhu, V. Suder, M.D. Aagaard, G. Gong, The simeck family of lightweight bl ock ciphers, International workshop on crypto graphic hardware and embedded systems, Spr inger, Berlin, Heidelberg, 2015.
- [9] B. Koo, D. Roh, H. Kim, Y. Jung, D.G. Le e, D. Kwon, CHAM: a family of lightweight block ciphers for resource-constrained device s, International Conference on Information Se curity and Cryptology, Springer, Cham, 2017.

# 딥페이크 데이터셋을 이용한 딥페이크 영상 검출 방법의 성능 비교

Rafiu1 Hasan Khan\*, 이석환\*\*, 권기룡\*

\*부경대학교 IT융합응용공학과, \*\*동아대학교 컴퓨터공학과

## *Performance Comparison of Deepfake Image Detection Methods using Deepfake Dataset*

Rafiu1 Hasan Khan\*, Suk-Hwan Lee\*\*, Ki-Ryong Kwon\*

\* Dept. of IT Convergence Engineering, Pukyong National University,

\*\* Dept. of Computer Engineering, DongA University,

### **Abstract**

The inception of artificial image generation was groundbreaking in the field of image processing. However, with mass popularity and difficulty in detection, this innovation is posing threats and concerns towards society. In this paper, we have comprised a study on the recent techniques that have been working on fake image detection. These techniques are based on deep learning methods and they are being regarded as the most efficient methods till now. For this research, we have used deep learning methods i.e. Mesonet, EfficientNet, GoogleNet, VGG-16, and VGG-19 to detect deep fake images. For this research, we have used the Deepfake Dataset of Mesonet article. This Deepfake Dataset is made up of fake images collected from 175 forged videos and real face images from various internet sources. We have implemented all those deep learning methods onto the Deepfake Dataset and compared them based on their total accuracy rate. Our study shows that VGG-19 outperformed the other methods with an overall accuracy rate of 97.2%. Although VGG-19 is computationally more expensive, the accuracy puts it at the top.

## I. Introduction

With the advent of social mediums and available smart gadgets, images and videos have become far more accessible than at any time in history. According to the statistics of Facebook, more than 300 million photos get uploaded every day. All of these photos are not original rather they are a mix-up of tempered or altered images. Many freeware software and enthusiastic research have paved the way towards the rise of fake images. The field of digital image forensics

research is dedicated to the detection of image forgeries to regulate the circulation of such falsified contents [1]. Recently, deep learning methods have been established as successful methods for digital image forensics. Thus, in this research, we have compiled a comparative study on deep fake image detection based on the recent deep learning methods. There are thousands of deep fake videos to analyze but we have chosen MesoNet's [1] Deepfake dataset so that we can have a common platform to

compare. We believe, this study will give a brief insight into digital image forensics and will guide future research.

## II. Methods

### 2.1 MesoNet

MesoNet [1] automatically and efficiently detects face tampering in videos. Traditional image forensics techniques struggle to detect fake image because the fake videos get degraded due to the high data compression. Mesonet follows a deep learning approach and presents two networks, both with a low number of layers to focus on the mesoscopic properties of images. The proposed two architectures namely “Meso-4” and “MesoInception-4” to solve these problems while producing 27,977 and 28,615 trainable parameters respectively.

### 2.2 EfficientNet-b0

EfficientNet-b0 [2] is the base network of the EfficientNets group. Although EfficientNet-b7 is the strongest network among them, however, they were built on EfficientNet-b0. Moreover, computationally EfficientNet-b0 is the least expensive network among them as it produces 5.3 million parameters whereas, EfficientNet-b7 produces 66 million parameters. EfficientNets [2] are the most promising networks in the field of digital image forensics.

### 2.3 GoogleNet

GoogleNet [3] is one of the most efficient networks for classification tasks. With features such as 1x1 convolution or modified inception module, it successfully applies dimensionality reduction as well as does not compromise with the performance. GoogleNet [3] produces 7 million parameters making it one of the least expensive networks in the least.

### 2.4 VGG

VGG or Visual Geometric Group is a series of the convolution network model starting from VGG11 to VGG19. The main intention behind it was to understand how the depth of convolutional networks affects the accuracy of the models of large-scale image classification and recognition. The VGG-16 has 13 convolutional layers and 3 fully connected layers while VGG-19 has 16 convolutional layers and 3 fully connected layers. The overall structure includes 5 sets of convolutional layers, followed by a MaxPool. The difference between all the VGGS is the increase in the depth as we move from VGG11 to VGG19 more and more cascaded convolutional layers are added in the five sets of convolutional layers. Both the VGG-16 and VGG-19 produce 138 million parameters making them the most computationally expensive networks on our list

## III. Results and Discussion

### 3.1 Dataset

Table 1: Deepfake Dataset.

Set	Forged Class	Real class
Deepfake training	5103	7250
Deepfake testing	2845	4259

The Deepfake dataset was created using Deepfake technique. Deepfake is a technique that aims to replace the face of a targeted person with the face of someone else in a video [1]. After some modifications, Deepfake technique was created into a user-friendly application called FakeApp. Deepfake images were created from 175 rushes of forged videos. All videos are compressed with different compression levels. All the faces

have been extracted using the Viola-Jones detector [5] and aligned using a trained neural network for facial landmark detection [6]. Then, the dataset has then been doubled with real face images, also extracted from various internet sources and with the same resolutions. Precise numbers of the image count in each class as long as the separation into a set used for training and model evaluation can be found in Table 1.

### 3.2 Experimental Results

We did transfer learning with all the above-mentioned deep learning methods using the Deepfake dataset. The dataset was divided into eighty percent to twenty percent for training and validation/testing. After the division, we applied image augmentation to the training dataset. For all the simulations, we used the Stochastic Gradient Descent with Momentum (SGDM) and set an initial learning rate of 0.0003. We set the max epoch to 30 and the mini-batch size to 16. Also, we applied shuffle after every epoch. Finally, we trained them on a single NVIDIA GeForce RTX 2070 16GB GPU.

Table 2: Classification Score.

<b>Network</b>	<b>Deepfake Classification Score</b>	
<b>Class</b>	<b>Forged</b>	<b>Real</b>
Meso-4	0.882	0.910
MesoInception-4	0.934	0.900
EfficientNet-b0	0.913	0.934
GoogleNet	0.950	0.965
VGG-16	0.955	0.9685
VGG-19	0.964	0.9765

Our simulation was done on individual frame / image. Table 2 displays the deepfake classification score of all the networks. Apparently, Meso-4 is the least performing network whereas, VGG-19 outperformed others.

Table 3 displays the overall accuracy of all the networks. The VGG-19 topped the chart with an accuracy rate of 97.2% whereas. Meso-4 achieved the least accuracy rate of 89.1%.

Table 3: Accuracy Rates.

<b>Network</b>	<b>Accuracy</b>
Meso-4	0.891
MesoInception-4	0.917
EfficientNet-b0	0.922
GoogleNet	0.958
VGG-16	0.963
VGG-19	0.972

## IV. Conclusion

According to our study, VGG-19 proved as the most accurate network. However, VGG-19 is also the most computationally expensive network. Since our concern is about deepfake detection, VGG-19 fits to our choice. In the future, we will extend our study by experimenting on more networks i.e. EfficientNet-b2 to EfficientNet-b7 to find a least computationally expensive method for deepfake detection.

## Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1F1A1069124 and 2020R1I1A3066594) and by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2020-0-01797) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation).

## [Reference]

- [1] D. Afchar, V. Nozick, J. Yamagishi, and

- I. Echizen, “MesoNet: a Compact Facial Video Forgery Detection Network,” *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1 - 7, Dec. 2018, doi: 10.1109/WIFS.2018.8630761.
- [2] M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *International Conference on Machine Learning*, pp. 6105 - 6114, May 2019, Accessed: Feb. 04, 2021.
- [3] C. Szegedy et al., “Going Deeper With Convolutions,” *arXiv:1409.4842*, pp. 1 - 9, 2015, <https://arxiv.org/abs/1409.4842>,
- [4] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *The 3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 2015,
- [5] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, p. I - I, Dec. 2001, doi: 10.1109/CVPR.2001.990517.
- [6] D. E. King, “Dlib-ml: A Machine Learning Toolkit,” *Journal of Machine Learning Research*, vol. 10, no. 60, pp. 1755 - 1758, 2009.

# 심층신경망에 대한 적대적 공격과 탐지 방법 조사

박태우, 최석환, 최윤호

부산대학교

A Survey of Adversarial Attack and Detection Methods for DNN

Taeu Bahk, Seok-Hwan Choi, Yoon-Ho Choi

Pusan National University

## 요약

심층신경망(DNN) 기술은 최근 다양한 작업에서 놀라운 발전을 하고 있다. 하지만 심층신경망은 노이즈나 적대적 섭동(Adversarial Perturbation)에 취약하다. 심층신경망을 기만하기 위하여 데이터에 적대적 섭동을 추가한 산출물이 적대적 예제이다. 이 과정을 적대적 공격(Adversarial Attack)이라고 한다. 적대적 공격을 통해 생성되는 적대적 예제는 심층신경망이 잘못된 예측을 출력하도록 야기한다. 이는 실제로 심층신경망 기반 시스템을 도입 및 적용하는 데 제한 요인인 된다. 따라서 적대적 공격과 탐지 기술은 중요하다. 본 논문에서는 적대적 공격과 탐지 방법을 조사하고자 한다.

## I. 서론

심층신경망(DNN, Deep Neural Network)은 최근 몇 년 동안 다양한 작업에서 매우 발전하여 널리 사용되고 있다. 하지만 내재적으로 불확실성을 가진 심층신경망은 노이즈나 적대적 섭동(Adversarial Perturbation)에 취약하다. 이로 인해 데이터에 적대적 섭동을 명시적으로 생성하여 심층신경망을 기만 및 회피하는 적대적 공격(Adversarial Attack)이 등장하였다. 이 과정에서 생성되는 적대적 공격의 산출물이 적대적 예제(Adversarial Example)이다. 적대적 예제는 심층신경망이 높은 신뢰도(Confidence)로 잘못된 예측을 출력하게 만들 수 있다. 즉, 적대적 예제는 심층신경망의 오작동을 야기한다. 적대적 예제는 현실에서도 문제를 초래할 수 있다. 예컨대, 딥 페이크(Deep Fake)[1]와 같이 개인의 신원을 나타내는 영상을 조작하거나, 정지를 의미하는 표지판을 최저속도로 제한하는 표지판[2]으로 기만하여 자율주행 자동차가

위험한 행동을 수행하도록 만드는 범죄 등의 악용 사례가 발생할 수 있다. 적대적 공격과 탐지 방법에 대해 많은 연구가 있다. 하지만, 특히 적대적 탐지에 대해 조사하거나 연구한 국내 논문은 드물다. 따라서 본 논문에서는 대표적인 적대적 공격과 적대적 탐지 방법에 대해 조사하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 적대적 공격 방법에 대해 소개한다, 3장에서는 적대적 탐지 방법에 대해 기술한다. 마지막으로, 4장에서는 본 논문의 결론을 맺는다.

## II. 적대적 공격 방법

본 장에서는 적대적 공격 방법들에 대해 소개한다. 적대적 공격 방법은 대표적으로 Fast Gradient Sign Method (FGSM)[3]와 Basic Iterative Method (BIM)[4], Projected Gradient Descent (PGD)[5], C&W[6]가 있다.

### 2.1 Fast Gradient Sign Method (FGSM)

Ian Goodfellow 등[3]은 적대적 예제를 생성하기 위해 이미지에 추가할 적대적 섭동을 찾는 방법을 2014년에 최초로 제안하였다. FGSM은 인공신경망의 학습 방법인 경사 하강법(Gradient Descent)을 역방향으로 수행하는 방법이다. 즉, 심층신경망의 손실함수의 경사를 부호(sign) 방향이 가장 가파른 방향으로 증가시킨다.

### 2.2 Basic Iterative Method (BIM)

Kurakin 등[4]은 FGSM[3] 방법을 확장하여, 경사를 반복적으로 개선하는 방법을 제안하였다. 각 반복마다 손실함수의 값을 계산하여 경사를 이동시킨다. 또한, 각 반복에서 변경 가능한 적대적 예제의 픽셀 값을 제한한다.

### 2.3 Projected Gradient Descent (PGD)

Madry 등[5]은 FGSM 알고리즘을 사용하여  $\|\delta\|_\infty \leq \epsilon$  을 만족하는 제약조건상에서 손실함수 경사의 방향으로 적대적 섭동인  $\delta$ 를 반복적으로 개선하는 방법을 제안하였다. 다시 말해  $\delta$ 의 값 범위를  $\epsilon$ (적대적 섭동의 크기)값 안으로 제한시킨 채로, 손실함수의 경사를 반복적으로 개선하여 적대적 예제를 생성한다.

### 2.4 C&W

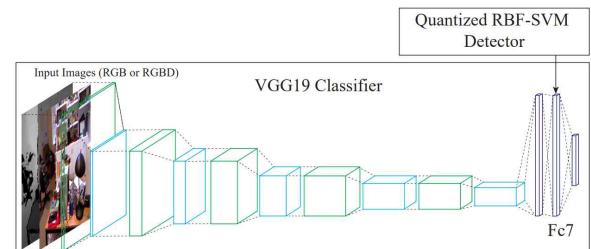
Carlini와 Wagner 등[6]은 적대적 섭동의 크기를 최소로 만들기 위해 세 개( $l_\infty$ ,  $l_0$ ,  $l_2$ )의 거리 메트릭(Distance metric)을 이용하는 적대적 공격을 제안하였다. 적대적 예제를 생성하기 위해, 적대적 섭동 계산을 반복하고 적대적 크기가 최소인 값 하나를 최종적으로 선택한다.

## III. 적대적 탐지 방법

본 장에서는 적대적 공격에 대응하기 위해, 입력 이미지를 정상 이미지와 적대적 예제 중 하나로 분류하는 적대적 탐지(Adversarial Detection) 방법을 소개한다. 적대적 탐지 방법은 대표적으로 Safetynet[7], [8]이 있다.

### 3.1 Safetynet

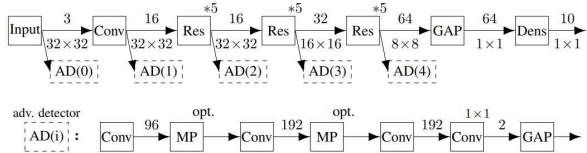
Lu 등[7]은 주어진 입력 이미지를 적대적 예제와 합법적(정상) 예제 중 하나로 이진 분류하기 위해, 기존 심층신경망에 적대적 예제 탐지 네트워크를 추가로 불린 아키텍처를 제안하였다. 이 논문의 저자는 소프트맥스(Softmax) 함수의 직전에 위치한 활성화 함수가 적대적 예제와 합법적 예제에 대하여 서로 다른 결과(패턴)를 나타낼 것이라고 가정하였다. 여기서 VGG19 분류 네트워크를 통해 최종 활성화 함수의 출력을 양자화한다. 그 다음, RBF-SVM 기반 적대적 예제 탐지 네트워크는 양자화를 통해 요약된 표현의 특징을 입력받아서 적대적 예제와 합법적 예제의 분포 차이를 학습한다. 다음 그림 1은 Safetynet의 아키텍처이다.



[그림 1] Safetynet 아키텍처

### 3.2 On detecting adversarial perturbations

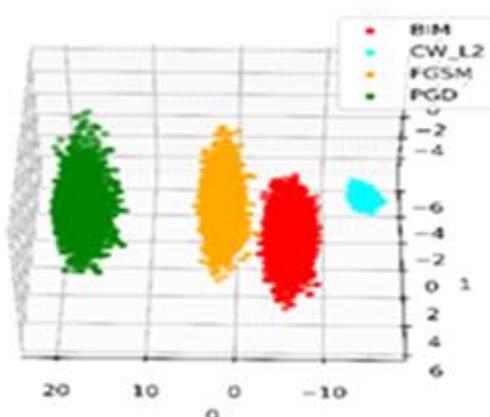
Metzen 등[8]은 주어진 입력 이미지가 적대적 예제일 확률을 계산하여, 입력 이미지를 적대적 예제와 합법적 예제 중 하나로 이진 분류하는 적대적 예제 탐지 네트워크를 제안하였다. 이 논문의 아키텍처도 Safetynet[7]와 마찬가지로, 기존 심층신경망에 적대적 예제 탐지 네트워크를 추가하였다. Safetynet과의 가장 큰 차이점은 심층신경망의 최종 활성화 함수의 특징이 아닌, 중간 활성화 함수의 특징을 사용한다. 적대적 예제 탐지 네트워크는 심층신경망 중간 계층 특징을 사용하여 입력 이미지가 적대적 예제일 확률을 학습한다. 다음 그림 2는 이 논문에서 제안한 모델의 아키텍처이다.



[그림 2] On detecting adversarial perturbations  
논문에서 제안한 아키텍처:  
(위) 심층신경망(ResNet);  
(아래) 적대적 예제 탐지 네트워크(adv. detector)

#### IV. 결론

본 논문에서는 적대적 공격과 탐지 방법들을 조사하였다. 적대적 공격을 통해 생성된 적대적 예제는 심층신경망 기반 시스템을 기만하고 무력화 시킬 수 있다. 따라서 많은 연구자들의 관심이 필요하다. 본 논문을 확장하여, CIFAR-10 데이터셋으로 생성된 적대적 예제가 4가지 유형의 적대적 공격 방법(FGSM[3], BIM[4], PGD[5], C&W[6]) 중 어떠한 공격에 의해 생성되었는지 탐지할 수 있는 연구를 다음 그림 3과 같이 수행하고 있다.



[그림 3] 적대적 공격 탐지 실험 결과

#### ACKNOWLEDGMENT

본 논문은 한국연구재단 논문연구과제 (NRF-2018R1D1A3B07043392) 및 4단계 BK21, 동남권4차산업혁명리더양성사업단에 의하여 지원되었습니다.

#### [참고문헌]

- [1]ROSSLER, Andreas, et al. Faceforensics++: Learning to detect manipulated facial images. In: Proceedings of the IEEE International Conference on Computer Vision. 2019. p. 1–11.
- [2]LU, Jiajun, et al. No need to worry about adversarial examples in object detection in autonomous vehicles. arXiv preprint arXiv:1707.03501, 2017.
- [3]GOODFELLOW, Ian J.; SHLENS, Jonathon; SZEGEDY, Christian. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572, 2014.
- [4]A. Kurakin, I. Goodfellow, and S. Bengio, “Adversarial examples in the physical world,” arXiv preprint arXiv:1607.02533, 2016.
- [5]MADRY, Aleksander, et al. Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083, 2017.
- [6]CARLINI, Nicholas; WAGNER, David. Towards evaluating the robustness of neural networks. In: 2017 ieee symposium on security and privacy (sp). IEEE, 2017. p. 39–57.
- [7]LU, Jiajun; ISSARANON, Theerasit; FORSYTH, David. Safetynet: Detecting and rejecting adversarial examples robustly. In: Proceedings of the IEEE International Conference on Computer Vision. 2017. p. 446–454.
- [8]METZEN, Jan Hendrik, et al. On detecting adversarial perturbations. arXiv preprint arXiv:1702.04267, 2017.

# 오토인코더 및 컨볼루션 네트워크를 활용한 사용자 인증에서의 개인정보 보호

김현지\*, 임세진\*\*, 양유진\*\*, 서화정\*\*\*†

\*한성대학교 IT 융합공학부 (대학원생)

\*\*한성대학교 IT 융합공학부 (대학생)

\*\*\*† 한성대학교 IT 융합공학부 (교수)

Privacy protection in user authentication using autoencoder and convolutional neural network

Hyun-Ji Kim\*, Se-Jin Lim\*\*, Yu-Jin Yang\*\*, Hwa-Jeong Seo\*\*\*†

\*Hansung University, Department of IT Convergence Engineering.  
(Graduate student)

\*\*Hansung University, Department of IT Convergence Engineering.  
(Student)

\*\*\*† Hansung University, Department of IT Convergence Engineering.  
(Professor)

## 요약

최근, COVID-19로 인해 특정 장소 출입 시 QR code 인증이 필수적으로 수행되고 있다. 그러나 QR code 기반의 인증 방식은 디코딩 과정을 필요로 하며, 저장 및 인증 과정에서 사용자의 개인정보가 모두 노출된다. 이러한 보안 취약점을 극복하기 위해 본 논문에서는 디코딩 과정이 필요하지 않은 생체인증 방안을 제안한다. 사용자의 지문 정보가 오토인코더에 입력되면, 지문의 특징 정보만이 추출되며 차원변경을 통해 인증에 사용될 이미지가 생성된다. 검증자는 해당 정보를 인식하여 컨볼루션 네트워크를 통해 사용자 인증을 수행한다. 즉, 검증자는 사용자의 실제 지문 정보를 학습 및 저장할 필요가 없으므로, 기존 방법의 보안 취약점을 극복할 수 있다. 또한, 비콘 등의 BLE 단말과 함께 사용할 경우 다른 장소에서 인증을 수행하는 경우를 방지할 수 있다.

## I. 서론

COVID-19가 전 세계적으로 유행함에 따라, 특정 시설 출입 시 방문 기록을 남기기 위해 QR-code 인증 방안이 시행되고 있다. 그러나 QR-code 사용으로 인한 개인 정보 노출, 악성 코드 삽입 등의 보안 취약점이 존재한다.

## II. 관련 연구

### 2.1 QR code 인증의 보안 취약점

QR코드로 인증을 할 때 암호화된 정보를 해독 해주는 디코딩(decoding) 과정이 필수적이다. 인증 절차를 수행하기 위하여 디코딩 과정을 거쳐 나온 개인 정보를 인증 서버로 전송하여야 하는

데, 이 과정에서 사용자의 정보가 모두 노출된다 는 문제점이 있다.

### 2.2. 오토인코더 (Auto Encoder)

오토인코더는 데이터 레이블 없이 학습시키는 비지도 학습 방법이다[1]. 인코더는 입력된 데이터의 핵심 특징만을 추출하고, 은닉층에서 이 추출한 특징을 학습시키면 디코더에서 추출된 특징 값을 바탕으로 원본 데이터와 근사한 값이 나오도록 재구성해준다. 이 때, 인코더와 디코더에 들어가는 노드수보다 은닉층에 들어가는 노드수가 더 적은 순실 압축 방법을 사용한다.

## III. 시스템 제안

본 논문에서는 QR code 기반 인증 방법의 보안 취약점을 극복하기 위해 오토인코더 및 컨볼루션 네트워크 (Convolutional Neural Network, CNN) 기반의 본인 인증 방안을 제안한다. 제안 기법을 통해 인증 정보에 대한 디코딩 과정 없이 본인 인증을 수행할 수 있다. 시스템 구성은 그림 1과 같다. 제안 시스템에는 인증 대상자인 사용자와 사용자에 대한 인증을 수행하는 기관인 검증자가 참여한다. 해당 시스템은 tensorflow lite 모델로 변환되어 디바이스 상에서 수행되며, 디바이스에 저장된 지문 정보를 활용하여 인증에 사용될 이미지를 생성한다. 검증자는 생성된 해당 이미지를 통해 사용자 인증을 수행한다.

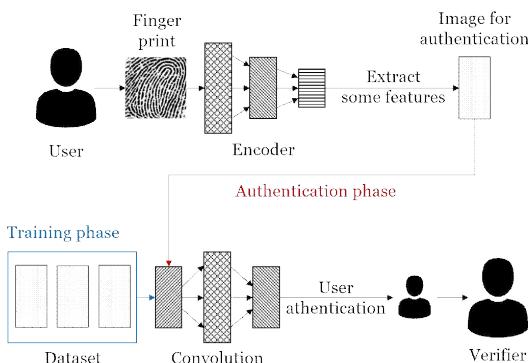


그림 1. 시스템 구성도

### 3.1 사용자

사용자는 본인 인증에 사용할 지문 정보에 대한 이미지를 생성해야한다. 이를 위해 오토인코더를 사용하여 자신의 지문 정보로부터 특징점을 추출해낸다. 오토인코더는 인코더를 활용하여 주요 특징 정보들을 추출한 후, 디코더를 통해 해당 정보를 기반으로 다시 원본 데이터로 복구한다. 이 과정에서 노이즈가 제거된다. 그러나 지문 정보를 이미지와 같은 원본 데이터의 형태로 재생산하는 것이 아니라, 해당 지문이 갖는 주요 특징만을 가지고 인증 정보를 생성하는 것으로 제안하는 시스템에서는 디코더 모델이 필요하지 않다. 따라서 학습 시에는 각각 설계된 인코더와 디코더를 결합한 모델을 사용하고, 이후 인증 정보 생성 시에는 학습된 인코더만을 사용하여 추론한다. 이러한 구조를 통해 지문 전체에 대한 정보 노출 없이도 본인인증이 가능하며, 모바일 장치 상에서의 추론 시 배포되는 모델의 용량을 줄일 수 있다.

훈련에 사용되는 오토인코더의 구조 및 하이퍼파라미터는 다음과 같다. conv2D + Maxpooling2D 구조를 두 번 반복한 후 출력층으로 conv2D를 사용한다. max pooling에서 pool size를 2로 설정하였으므로 입력 이미지의 가로, 세로 길이가 총 4배 줄어든다. 또한, 특징 벡터의 차원을 나타내는 latent vector 값을 128로 설정하여 인코더의 출력은 (56,56,128)의 모양을 갖게 된다. 인코더의 출력은 디코더의 입력으로 사용된다. 데이터의 크기를 늘려주는 upsampling 과정을 통해 가로, 세로의 크기를 인코더의 입력 데이터의 형태로 복원한다. 학습을 위해 Mean Square Error 손실 함수를 사용하였고, 최적화 함수는 RMSprop, 활성화 함수는 은닉층에 ReLu를, 출력층에 Sigmoid를 사용하였다.

### 3.2 검증자

검증자는 CNN을 기반으로 하여 인증을 원하는 사용자로부터 인식한 정보를 통해 본인 인증을 수행한다. 학습에 사용되는 데이터는 인코더를 통해 생성되는, 인증을 위한 이미지이다. 즉, 사용자들의 실제 지문 정보를 저장하지 않고, 특징점 추출 후 변형된 상태의 이미지만을 저장하여 학습에 사용한다. 따라서 기존의 QR code 인증과 달리 디코딩 과정에서의 정보 노출 없이 인코딩된 정보 그대로 인증이 가능하다. 또한, 비콘과 같은 BLE 단말을 사용하여 사용자가 해당 위치에 존재함을 증명하고 그 이후에 본인 인증 단계를 수행할 수 있게 할 경우, 사용자가 현재 위치를 속이고 다른 장소에서 인증하는 상황을 방지할 수 있다.

사용자의 본인 인증을 위해 훈련된 모델을 통한 추론을 진행한다. 먼저, 사용자가 생성한 인증 데이터를 훈련된 모델에 입력한다. 다중 클래스 분류 문제이기 때문에, 입력 데이터는 각각의 클래스에 속할 확률 값을 가진다. 해당 값들 중 가장 높은 값이 사전에 설정한 인증 임계값을 넘을 경우 본인으로 인증한다. 만약 A 사용자의 인증 정보가 A 사용자로 분류되었다고 해도 임계값을 넘지 못 할 경우 인증되지 않는다. 또한, B 사용자의 인증 정보를 A 사용자로 분류할 경우, 임계값 초과 여부와 상관없이 인증에 실패하게 된다. 즉, 인증이 가능한 경우는 본인이 본인으로 분류되며, 해당 클래스로 분류될 확률이 임계값을 넘는 경우뿐이다. 사용자 인증을 위한 CNN 모델로는 사전 학습된 모델인 Inception v3 model를 사용하였다.

## IV. 성능 평가

본 실험을 위해 클라우드 기반 서비스인 Google Colaboratory를 활용한다. Ubuntu 18.04.3 LTS에서 실행되며, 12GB RAM의 Nvidia GPU로 구성되고 Python 3.6.9, tensorflow 2.2.0-rc, Keras 2.3.1이 사용된다. 실험 결과는 딥러닝 평가 기준 중 하나인 재현율 (recall)과 정밀도 (precision)의 조화평균인 F-measure로 평가 한다. 또한, 동일 오류율 (Equal Error Rate, EER)을 통해 본인인증에 대한 성능을 확인한다.

인증 정보를 생성하는 오토인코더의 학습 결과는 그림 2와 같다. 훈련 및 검증 손실이 비슷한 수준으로 감소한 것을 볼 수 있다.

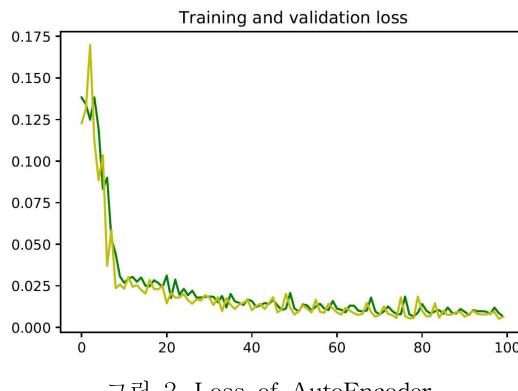


그림 2. Loss of AutoEncoder

그림 3은 사용자 인증을 위한 CNN의 학습 결과를 나타낸다. 또한, 그림 4와 같이 동일 오류율을 계산하고, 결과 값을 정상 서명과 위조 서명을 구분할 임계값으로 설정한다. EER은 타인 수락률 (FAR)과 본인 거부율 (FRR)이 같아지는 지점이며, 타인 수락률과 본인 거부율은 같은 과정을 통해 계산한다. 표 1은 제안 시스템의 본인 인증에 대한 성능, EER 및 임계값을 정리한 표이다.

$$FAR = \frac{FP}{FP+TP}, \quad FRR = \frac{FN}{FN+TN} \quad (1)$$

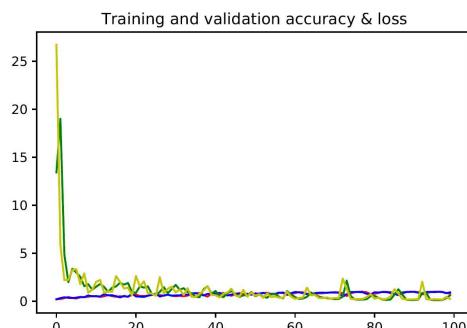


그림 3. Loss of CNN

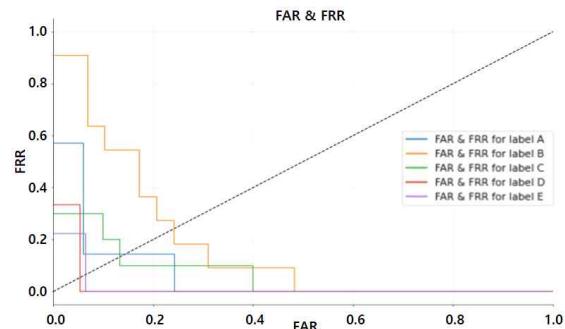


그림 4. 동일 오류율

표 1. F-measure, EER and threshold

F-measure	0.68
EER(average, max)	0.1, 0.225
threshold	0.225

## V. 결론

본 논문에서는 오토인코더 및 컨볼루션 네트워크를 기반으로 디코딩 과정이 필요하지 않은 생체인증 방안을 제안하였다. 본 시스템이 위치인증과 함께 사용될 경우 보안성을 더욱 향상시킬 것으로 기대된다.

## VI. Acknowledgment

이 성과는 부분적으로 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478) 그리고 이 성과는 부분적으로 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.2018-0-00264, IoT 융합형 블록체인 플랫폼 보안 원천 기술 연구).

## [참고문헌]

- [1] SBaldi, Pierre. "Autoencoders, unsupervised learning, and deep architectures." Proceedings of ICML workshop on unsupervised and transfer learning. JMLR Workshop and Conference Proceedings, 2012.

# 정보주체의 개인정보 통제권 강화를 위한

## 개인정보보호법 쿠키 동의 개정안

전주현\* 이경현\*\*

\*부경대학교 대학원 정보보호협동과정

\*\*부경대학교 IT융합응용공학과

### *A Proposal for Amended Cookie Consent of Personal Information Protection Act to Strengthen Data Subjects' Right to Control Personal Information*

Ju-Hyun Jeon\* Kyung-Hyune Rhee\*\*

\* Interdisciplinary Program of Information Security, The Graduate School,  
Pukyong National University

\* Department of IT Convergence & Application Engineering, Pukyong  
National University

### 요약

현재 개인정보처리방침에 공개된 쿠키 설정은 텍스트적인 공개로 실제 정보주체의 개인정보 자기결정권에 영향을 주기는 어렵다. 이는 개인정보처리자보다 정보주체에게 책임을 전가하는 형태로 명시되고 있으며 보다 구체적이고 명확한 기술적 구현을 통해 정보주체가 쉽게 개인정보자기결정권을 행사 가능하도록 개선해야 한다. 본 논문에서는 해외 쿠키 정책을 비교 분석해 웹 사이트 접속 시 브라우저상에서 사용자가 직접 통제 가능한 기술적 개선을 통해 개인정보처리자의 책임을 부각시키고 정보주체가 직관적으로 개인정보처리 접근성에 대한 선택권을 부여하여 개인정보 보호를 강화하는 기술적 구현을 GDPR과 비교 분석해 제안하였다..

키워드 개인정보, GDPR, 개인정보처리방침, 쿠키 동의, 브라우저

### I. 서론

지난 2020년 2월 크롬, 엣지, 파이어폭스 등 사용자가 많은 브라우저에 교차사이트와 동일 사이트 쿠키에 대한 정책이 변경 되었다. 웹 브라우징의 개인정보보호 및 보안을 개선하기 위한 노력이 지속되고 있다. 현재 개인정보처리방침에 따라 공개된 쿠키 설정에 대한 고지사항이 있지만 사용자에게 단순 정보공개에 의존한 텍스트상의 내용에 불과해 실질적 정보주체 보안에 영향을 주는지는 의문이다. 본 논문에서는 개인정보처리방침에 공개하는 쿠키 정책을 보다 정보주체 친화적이고 기술적 구현을 통해 해외 사례와 개인정보 추적 기술에 대응하는 정보주체 권리 대한 기술적 연구를 국내 개인정보보호법과 유럽의 GDPR(General Data Protection Regulation)을 비교 분석해 보았다.

### II. 브라우저 쿠키 정책

프랑스의 개인정보보호 기구가 사용자 동의 없이 ‘쿠키’를 설치해 광고에 활용한 구글과 아마존에 각각 1억유로(약1천317억원), 3천500만 유로(약461억원)의 과징금을 부과했다[1].

사용자가 많은 엣지, 크롬, 모질라, 오페라 브라우저 등 최근 브라우저에서 쿠키 관련 보안을 강화하고 있는 흐름을 <그림-1>에서 보이고 있다.

<그림-1> 브라우저 적합성 메트릭스

The table displays the following data:

	Chrome	Edge	Firefox	Internet Explorer	Opera	Safari	Android	Chromium-based	Firefox-based	Internet Explorer-based	Opera-based	Safari-based	Android-based
set-cookie	Yes	12	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HttpOnly	Yes	12	3	9	11	5	37	Yes	4	Yes	4	Yes	Yes
max-age	Yes	12	Yes	8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SameSite	51	16	60	No	39	13	★	51	51	60	41	13	50
SameSite=Lax	51	16	60	No	39	12	51	51	60	41	13	50	50
Defaults to Lax	60	60	60	No	67	No	60	60	No	No	No	No	No
SameSite=None	51	16	60	No	39	13	★	51	51	60	41	13	50
SameSite=Strict	51	16	60	No	39	12	51	51	60	41	12.2	5.0	5.0
Secure context required	60	60	60	No	67	No	60	60	No	No	No	No	No
Cookie prefixes	49	79	50	No	36	Yes	49	49	50	36	Yes	5.0	5.0

Legend:  
□ Full support  
□ Partial support  
■ No support  
★ See implementation notes.  
■ User must explicitly enable this feature.

## 2.1 교차사이트와 동일사이트 쿠키

웹 사이트는 일반적으로 개인화된 광고, 추천, 위젯, SNS 삽입 등 외부 기능을 통합하고 있으며 웹을 탐색할 때 외부 서비스는 사용자 브라우저에 쿠키를 저장한 후 개인화된 서비스를 제공하거나 행태(behavior) 측정이 가능하다. 모든 쿠키는 연결된 도메인이 있으며 사용자가 사용하는 웹사이트 도메인과 연결되지 않은 외부 서비스와 일치하는 경우 ‘교차사이트’ 또는 ‘제3자 사이트’라고 한다.

반면에 동일한 사이트에서 쿠키 액세스는 쿠키 도메인이 사용자 주소 표시줄의 도메인과 일치하는 경우 ‘동일 사이트’로 인식하고 개별 웹 사이트에 로그인한 상태를 유지하고 개인행태 분석을 지원한다[2].

## 2.2 사이트간 위조공격(CSRF)

OWASP TOP 10에 빠지지 않고 웹 공격기법으로 언급되고 있는 사이트간 교차 공격에 대한 사항으로 이는 쿠키를 이용한 웹 공격의 대표적인 공격기법이다. 공격자가 웹 응용 프로그램의 출력에 스크립트를 삽입하여 브라우저가 페이지의 일부라고 판단하여 스크립트가 실행되게 하는 공격이다[3]. 즉 공격자는 스크립트가 실행 가능한 링크를 타겟에게 보내 클릭하게 함으로써 피해자의 웹 브라우저에서 쿠키, 세션, 사용자 자격증명 등을 악용하는 공격이다. 사용자가 사이트에서 인증이 되면 사이트는 합법적인 요청과 위조된 요청을 구분하기 어렵

다.

## 2.3 개인정보처리방침 및 공개방법

개인정보보호법 제30조에 따르면 개인정보처리방침을 수립·공개 하도록 되어있다. 동법 제30조제1항7호에는 ‘인터넷 접속정보파일 등 개인정보를 자동으로 수집하는 장치의 설치·운영 및 그 거부에 관한 사항(해당하는 경우에만 정한다)’을 공개하도록 법률에 명시하고 있으며 실제 대부분의 사이트에서 공개된 개인정보처리방침 현황을 살펴보면 정보주체에게 브라우저상에서 각 개인이 설정하도록 안내만 되어 있다. 즉, 브라우저에서 설정을 하지 않은 사용자의 경우 쿠키를 통한 각 개인의 행태정보 등이 모두 수집 가능하다고 해석이 되어 정보주체에게 불리하게 포괄적 규정만 명시해 놓은 상황이다.

## III. GDPR에서 요구하는 쿠키 정책

GDPR에서는 사이트 이용의 경우 이용자의 동의를 구할 때 다음 사항이 반드시 포함되어야 한다[6].

- 정보수집하는 기업과 파트너사의 구체적이고 명확한 정보를 제공해야 한다.
- 개인정보 수집의 목적을 구체적이고 명확하게 밝혀야 한다.
- 언제든지 동의를 철회할 권리 등에 관한 구체적인 정보를 제공해야 한다.
- 동의란 정보주체의 진술 또는 적극적인 행위로 이루어져야 하며 모호하지 않아야 한다. (자동체크 안됨)
- 동의 거부에 대한 불이익이 없어야 하며 이용약관으로부터 동의를 분리해야 한다.
- 복수의 목적과 다른 유형의 처리에 대해서는 개별적으로 동의하는 세부 옵션을 제공해야 한다.

실제로 GDPR은 IP주소, MAC주소, 온라인 쿠키 등을 통해 정보주체 식별이 가능한 경우에는 해당 정보를 개인정보로 간주하고 있다[4].

### 3.1 EU e-privacy Directive

2002년에 제정된 EU e-privacy Directive가 2009년에 개정되어 ‘인터넷 쿠키를 이용자 PC에 저장하기 위해서는 해당 이용자의 명시적인 동의(explicit consent)를 구해야 한다’는 내용이 포함되었다. 2년의 유예기간을 거쳐 2011년5월 26일부터 시행 되었으며 ‘쿠키법’이라고도 불린다. 2011년5월 이전에는 현재 국내 개인정보보호법 제30조 개인정보처리방침에서 명시하듯이 개괄적으로 웹사이트에 게재하는 것을 법률적 요구사항으로 하였으나 2011년5월 이후부터는 쿠키 관련해 명시적인 동의를 받고 있다[5].

### 3.2 EU GDPR 시행에 따른 쿠키정책

GDPR 시행 이전 쿠키는 e-privacy 법령에 따라 통신전송의 목적에 따라 사용했으나 쿠키 적용에 동의가 요구되지는 않았다. 하지만 GDPR 시행 이후에는 쿠키 식별값도 개인정보의 한 종류이기 때문에 처리를 위해 적법한 근거가 필요하였다. 대표적인 방법으로 쿠키에 대한 ‘동의’를 사용하고 있다.

### 3.3 개인정보보호법 동의철회

개인정보보호법 제39조의7에 ‘정보통신서비스제공자 등은 제1항에 따른 동의철회, 법 35조에 따른 개인정보 열람, 법 제36조에 따른 정정을 요구하는 방법을 개인정보의 수집 방법보다 쉽게 하여야 한다’고 명시하고 있다. 이는 정보통신서비스제공을 하는 이용자로 한정되어 있어 특례조항으로 규정하고 있다. 실제 한번 동의한 사항은 다시 철회나 정정을 요구하는 절차가 복잡해 정보주체가 적극적으로 개인정보 자기결정권을 행사하기 위한 기술적 방법이 제시되어야 한다.

#### IV. 개인정보보호법 개정 제안

#### 4.1 개인정보 보호법 개정

<sup>10</sup> 본 논문에서는 개인정보보호법 제35조 개인

정보처리방침 개정을 제안한다. 개인정보처리방침 공개시에 쿠키정책에 대한 조항을 개정해 사이트를 방문 시에는 의무적으로 ‘쿠키 동의’를 얻도록 한다. 개인정보처리방침 제1항제7호에 명시되어 있는 자동접속기술에 대한 사항을 이용자에게 브라우저 설정 정도로만 안내하는 것이 아닌 이용자나 정보주체가 직접 관련 조항에 대한 사항을 선택적으로 사용 가능하게 옵션을 제공하도록 다음 <표 1>과 같이 법 개정을 제안한다.

<표 1> 개정 제안안

## 4.2 쿠키 동의 기술적 제안

해외 사례에서도 보면 “당신의 향상된 서비스를 위해 쿠키를 수집하고 있습니다. 웹사이트를 계속 이용하려면 쿠키에 동의한 것으로 간주 하겠습니다.”라는 알림 팝업을 보게 된다. 이는 유럽의 GDPR 시행 이후 모두 불법이다. 정보주체의 적극적인 행위의 동의가 이루어지지 않았기 때문이다. 국내 현행법 하에서는 정보주체의 쿠키 동의에 대한 적극적 행위가 제도적으로 반영돼 있지 않다.

본 논문에서 정보주체(정보통신서비스는 이용자)가 적극적인 개인정보처리에 따른 쿠키 동의에 대한 방법을 제안한다.

<그림-2>와 <그림-3>에서처럼 이용자가 웹사이트에 방문하면 자동으로 수집하는 개인정보 장치 설치·운영에 대한 동의를 받도록 한다. 메뉴는 4가지로 구성된다. 첫 번째는 모든 쿠키에 대해 동의한다. 두 번째는 현재 방문하고 있는 동일 사이트에만 허용한다. 세 번째는 접속 사이트 서비스 이용에 직접적인 저장과 접근 외에 쿠키 적용 거부한다. 네 번째는 모든 쿠키에 대해 거부한다.



<그림-2> 쿠키동의 제안모델



<그림-3> 쿠키 옵션별 제안모델

기본값은 ‘동일 사이트 쿠키허용’과 ‘거부(OFF)’로 되어 있어 사이트를 접속하고 이용하면 정보주체의 명시적이고 적극적인 행위를 하도록 제안한다. 두 번째와 세 번째 차이점은 두 번째는 일상적인 사이트 이용자의 행태분석 까지 가능한 허용범위를 나타내고 세 번째는 사이트 이용함에 있어 가장 기본적인 쿠키 기술에 대한 접근과 저장을 의미하며 성향이나 행태분석은 거부하는 옵션이다.

## V. 결론 및 향후 방향

데이터 3법 개정 이후 국내 개인정보에 대한 법 제도는 규제 중심에서 활용도를 높이도록 완화되었다. GDPR시행은 이런 측면에서 동의 만능주의에 있는 국내 개인정보보호 관련 법률에 참고 할 사항이 많다. 4차산업혁명 기술의 중심에 있는 인공지능과 빅데이터 기술을 제대로 활용하려면 자동화된 기술의 개인정보처리 과정에서 침해하는 행위가 있어서는 안된다. 최근 언론에서 이슈화된 인공지능 데이터 관련 문제도 정보주체의 명시적인 동의 없이 사용했다는 점에서 많은 시사점을 보여주고 있다. 본 논문에서 제안하고 있는 자동화 기술 설치 운영에 따른 ‘쿠키 동의’에 대한 법률 개정 제안과 구현 기술은 향후 인공지능, 빅데이터 활용에 정보주체(이용자)의 적극적인 행위로 동의를 반영한 컴플라이언스 리스크를 해소하는데 기여할 것으로 기대한다. 또한, 향후 동의 철회나 3자 제공에 대한 정보주체의 철회를 개선하는 제도적·기술적 연구를 지속할 예정이다.

## [참고문헌]

- [1] 프랑스 당국 ‘동의없는 쿠키광고’ 구글에 1천317억원 과징금, 연합뉴스. 2020.12.10
- [2] 접근제어를 이용한 교차 사이트 스크립트 필터링, 202 한국정보과학회 봄 학술집. vol29. no1
- [3] [https://cheatsheetseries.owasp.org/cheatsheets/Cross-Site\\_Request\\_Forgery\\_Prevention\\_Cheat\\_Sheet.html](https://cheatsheetseries.owasp.org/cheatsheets/Cross-Site_Request_Forgery_Prevention_Cheat_Sheet.html)
- [4] 한국인터넷진흥원, 2020년 GDPR상담사례집
- [5] Bond Robert, The EU E-Privacy Directive and Consent to Cookies, The Business Lawyer. 68(1):215–223. 2012
- [6] 행정자치부, 우리기업을 위한 GDPR안내서, PIWIK.pro. 영국감독기구 동의 가이드라인

# 정적 분석 도구를 활용한 오픈소스의 보안 취약점 탐지 성능 비교 및 분석

정지인\*, 이경률\*

\*대구가톨릭대학교 컴퓨터소프트웨어학부

Vulnerability Detection Performance Comparison and Analysis of Open Source Software using Static Analysis Tools

Jiin Jeong\*, Kyungroul Lee\*

\*School of Computer Software, Daegu Catholic University

## 요약

현재 전 세계적으로 많은 오픈소스가 공개되는 상황에서, 개별적으로 작성하는 소스코드에 취약점이 내포되는 가능성이 존재한다. 본 논문에서는 취약점을 탐지하는 정적 분석 도구인 Cppcheck, Yasca, Flawfinder를 활용하여 공개된 오픈소스의 취약점 탐지 결과를 기반으로 성능을 비교하고 분석한다. 이를 위하여, 취약점을 포함하는 샘플 소스코드와 암호화 과정을 포함하는 실제 오픈소스를 대상으로 각 도구의 취약점 탐지 결과를 비교함으로써 성능을 분석하였다. 탐지된 소스코드 개수 및 정확도를 기준으로 탐지 성능을 분석한 결과, Flawfinder가 가장 성능이 높은 것으로 분석되었으며, Cppcheck가 그다음, Yasca가 가장 성능이 낮은 것으로 분석되었다. 하지만 각 도구가 탐지한 CWE가 중복되지 않고 상이하며, 이는 CWE ID를 기준으로 성능을 정량적으로 평가하기 어려운 한계점이 존재할 것으로 판단된다. 이러한 한계점을 극복하기 위하여, 공개된 CWE의 탐지 정확도에 대한 연구 및 동적 분석을 통하여 탐지된 취약점을 검증하는 연구를 진행할 예정이다.

**키워드:** 정적 분석 도구, 취약점 분석, 오픈소스 소프트웨어, 성능 분석

## I. 서론

전 세계 많은 개발자들이 자신이 작성한 오픈소스를 공개하며, 공개된 오픈소스를 실제 프로그램 및 시스템에 탑재함으로써 빠른 개발이 가능한 장점이 있다 [1]. 그럼에도 불구하고, 개별적으로 작성하는 소스코드는 취약점을 포함할 가능성이 있으며, 이로 인하여 보안 위협이 발생한다. 이러한 문제점을 해결하기 위하여, 소스코드에 내포된 취약점을 자동으로 분석하는 다양한 정적 도구들이 등장하였다.

하지만 방대한 정적 분석 도구들이 공개되어 있고, 각 도구마다 취약점 분석 및 탐지 방법이 상이한 문제점이 존재한다. 그뿐만 아니라, 도구들의 취약점 탐지와 관련된 성능의 비교 및 분석에 대한 연구가 미비한 실정이다 [2, 3].

따라서 본 논문에서는 취약점 분석을 위하-

여, 오픈소스 정적 분석 도구인 Cppcheck, Yasca, Flawfinder의 취약점 탐지 성능을 비교하고자 한다. 성능 비교를 위하여, 암호화 과정을 포함하는 오픈소스를 선정하여, 각 도구의 성능 평가 결과를 비교 및 분석한다.

## II. 관련 연구

본 논문에서는 취약점 분석을 위하여, 기존의 성능 평가가 연구되었고, 접근성 및 편의성이 높은 C/C++ 언어가 대상인 오픈소스 정적 분석 도구를 선정하여 성능을 평가한다 [4].

### 2.1. Cppcheck

Cppcheck는 c/c++ 코드를 위한 정적 분석 도구로, 코드 분석을 통하여 버그, 정의되지 않은 동작 및 위험한 코딩 구조를 집중적으로 탐지

한다 [5]. 이 도구는 2007년 5월 8일에 처음 공개되었으며, 최근 버전인 2.3은 2020년 12월 5일에 업데이트되었다.

## 2.2. Yasca

Yasca는 보안 취약성, 코드 품질, 성능 및 프로그램 소스코드의 모범사례에 대한 적합성을 검사하는 오픈소스 프로그램이다. Findbugs, PMD, JLint, JavaScirpt Lint, PHPLint, Cppcheck, ClamAV, Pixy 및 RATS와 같은 외부 오픈소스 프로그램을 활용하여 특정 파일 형식을 스캔하며, 지정 스캐너도 포함할 수 있다 [6]. 이 도구는 2007년에 처음 공개되었으며, 최근 버전은 2017년에 업데이트되었다.

## 2.3. Flawfinder 도구

Flawfinder는 c/c++ 소스코드를 검사하며, 발생 가능한 보안 취약점을 위험 수준별로 분류하여 보고하는 도구이다. 취약점 탐지를 위하여, 버퍼 오버플로우 위험 함수, 형식 문자열 문제 함수, 경쟁 조건 문제 함수, 잠재적인 헬메타 문자 위험 및 잘못된 난수 획득 함수와 같은 잘 알려진 문제가 있는 c/c++ 함수의 내장 데이터베이스를 사용한다 [7]. 이 도구는 2001년에 처음 공개되었으며, 최근 버전인 2.0.15는 2021년에 업데이트되었다.

## 2.4. 기준 성능 평가 연구

[8]에서는 NIST의 취약점 분석 테스트 코드인 Juliet Test Suite for C/C++를 기반으로 정적 분석 도구인 Cppcheck, Yasca, Flawfinder의 성능을 평가하였다.

이 연구에서는 특정 CWE (Common Weakness Enumeration) 인 78, 79, 89, 99, 121, 122, 134, 170, 244, 251, 259, 362, 367, 391, 401, 411, 412, 415, 416, 457, 468, 476, 489를 대상으로, 정적 분석 도구인 Cppcheck, Yasca, Flawfinder의 탐지 정확도를 None, Low, Medium, High로 분류하였다. 탐지 방법으로는, 동일한 취약점을 가지는 테스트 코드를 다수 분석하여 정확도를 도출하였으며, 70% 이상 탐지할 경우 High, 40~69%를 탐지한 경우 Medium, 40% 미만인

경우 Low, 0%인 경우 None으로 정의하였다.

이 연구는 공개된 테스트 코드를 대상으로 정적 분석 도구들의 정확도에 대한 성능 평가를 연구하였지만, 오픈소스와 같이 실제 사용되는 소스코드를 대상으로 성능을 평가하지는 않았다. 따라서 본 논문에서는 소스코드에서 특히 중요한 부분인 암호화를 사용하는 공개된 오픈소스를 대상으로 정적 분석 도구들의 성능을 비교하고 분석하고자 한다.

## III. 정적 분석 도구의 취약점 탐지 성능 비교 및 분석

정적 분석 도구의 취약점 탐지 성능을 비교하고 분석하기 전에, 각 도구의 탐지 결과를 확인하기 위하여, 버퍼 오버플로우와 포맷 스트링 취약점이 명백하게 존재하는 소스코드를 대상으로 정적 분석 도구인 Cppcheck, Yasca, Flawfinder의 탐지결과를 도출하였으며, 샘플 소스코드 [9]를 그림 1, 탐지 결과를 표 1에 나타내었다.

```
//buffer overflow                                //format string
#include <stdio.h>                            #include<stdio.h>

int main(int argc, char *argv[]) {               main() {
    char buffer[10];                           char *buffer = "w!shfreem\n\x00";
    strcpy(buffer, argv[1]);                  printf(buffer);
    printf("%s\n", &buffer);                   }
}

A. 버퍼 오버플로우 소스코드
B. 포맷 스트링 소스코드
```

그림 1. 버퍼 오버플로우와 포맷 스트링 취약점을 포함하는 샘플 소스코드

표 1. 샘플코드 분석 결과표

도구명	버퍼 오버플로우 예제	포맷 스트링 예제
Cppcheck (v2.3)	X	O
Yasca (v2.0.15)	O	X
Flawfinder (v2.21)	O	O

샘플 소스코드의 취약점 탐지 결과, Cppcheck는 버퍼 오버플로우 취약점을 탐지하지 못하고, 포맷 스트링 취약점만 탐지하였다. Yasca는 버퍼 오버플로우 취약점만 탐지하고, 포맷 스트링 취약점은 탐지하지 못하였다. 마지

막으로 Flawfinder는 유일하게 버퍼 오버플로우와 포맷 스트링 취약점 모두를 탐지하였다.

이와 같이, 간단한 샘플 소스코드임에도 불구하고, 각 도구의 결과가 상이한 문제점이 있으며, 이러한 문제점은 성능 평가에서 오탐 및 미 탐과 같은 탐지율과 밀접한 관계를 가진다. 따라서 본 논문에서는 실제 사용되는 소스코드 중, 표 2와 같이 암호화 과정을 포함하는 오픈소스 5개를 선정하여 각 도구의 성능을 비교하고 평가하였다.

표 2. 선정한 오픈소스

오픈소스명	설명
Cryptsetup	<ul style="list-style-type: none"> <li>DM 커널 모듈을 기반으로 디스크 암호화를 편리하게 설정하는 도구 [10]</li> </ul>
Gnupg	<ul style="list-style-type: none"> <li>RFC 4880 (OpenPGP Message Format)에 정의된 OpenPGP 표준 구현</li> <li>데이터 및 통신 암호화 제공, 키 관리 시스템 및 공개키 딕레토리를 위한 접근 모듈 제공 [11]</li> </ul>
Linux-pam	<ul style="list-style-type: none"> <li>시스템에서 애플리케이션이나 서비스의 인증을 처리하는 라이브러리 [12]</li> </ul>
OpenSSH	<ul style="list-style-type: none"> <li>SSH 프로토콜을 사용한 원격 로그인을 위한 도구</li> <li>도청, 하이재킹 및 다른 공격을 제거하기 위하여 모든 트래픽 암호화 [13]</li> </ul>
OpenSSL	<ul style="list-style-type: none"> <li>TLS (Transport Layer Security) 및 SSL (Secure Sockets Layer) 프로토콜들을 위한 모든 기능을 제공하는 범용 암호화 라이브러리 [14]</li> </ul>

선정한 오픈소스는 많은 소스코드를 포함하여 모든 결과를 도출하기에는 한계가 있어, 각 오픈소스의 암호화와 관련된 모듈인 Gnupg의 g10, Linux-pam의 modules, Cryptsetup의 crypto\_backend, OpenSSH의 cipher, OpenSSL의 aes의 소스코드를 선정하였으며, 탐지된 전체 CWE ID와 개수를 표 3에 나타내었다.

표 3는 선정한 모듈에서의 암호화 기능을 제공하는 일부 소스코드를 대상으로 취약점 탐지 결과를 탐지된 전체 CWE ID 및 개수, 기존 연구에 포함되지 않은 CWE ID 및 개수, 기존 연구에 포함된 CWE ID 및 개수로 정리하였다. 이와 같이 분류한 목적은 탐지된 전체 CWE ID 중 기존 연구에서 탐지 정확도를 포함하지 않은 ID와 정적 분석 도구의 성능을 평가하기 위하여 기존 연구에 포함된 ID를 비교하기 위함이다.

비교 결과, Cryptsetup은 소스코드마다 약 5~10개의 취약점이 탐지되었고, 전반적으로 CWE-119와 120이 가장 많이 탐지되었다. 기존 연구에 포함된 CWE-476는 crypto\_cipher\_kernel.c에서 탐지되었다. Gnupg는 다른 오픈소스보다 많은 개수의 취약점이 탐지되었으며, 그중 CWE-119, 120, 126, 398이 가장 많이 탐지되었다. 기존 연구에 포함된 CWE-476는 keyid.c에서 탐지되었고, CWE-362는 keyring.c에서 탐지되었다. Linux-pam은 다양한 취약점의 종류와 개수가 탐지되었으며, 그중, CWE-120와 362가 가장 많이 탐지되었다. 특히, 이 오픈소스는 탐지된 대부분의 취약점이 기존 연구에 포함되었다. OpenSSH는 다른 오픈소스보다 취약점이 탐지된 소스코드의 개수가 적으며, CWE-120과 398이 가장 많이 탐지되었다. 이 CWE는 cipher.c에서 탐지되었다. OpenSSL 역시 OpenSSH와 비슷하게 다른 오픈소스보다 취약점이 탐지된 소스코드의 개수가 적으며, CWE-398과 457이 탐지되었다. 기존 연구에 포함된 CWE-457은 aes\_ige.c에서 탐지되었다.

표 3의 비교 결과를 토대로 탐지된 모든 CWE의 성능을 비교하고 평가하기 위한 기준이 연구되지 않은 한계점이 존재한다. 따라서 탐지 정확도의 기준을 포함하는 CWE를 대상으로 표 4과 같이 정적 분석 도구의 성능을 비교하였다.

비교 결과를 살펴보면, 각 도구에서 탐지된 CWE ID가 중복되는 항목이 하나도 없으며, Cppcheck와 Yasca는 탐지 정확도가 Low이거나 None인 CWE-401, 457, 476만 탐지되었다.

표 3. 탐지된 전체 CWE ID와 개수 및 기존 연구 포함 및 미포함 CWE ID와 개수 비교

오픈소스명	소스코드명	탐지된 전체 CWE		기존 연구 미포함 CWE		기존 연구 포함 CWE	
		ID	개수	ID	개수	ID	개수
Cryptsetup (v2.3.4)	crypto_kernel.c	20, 119, 120, 398	5	20, 119, 120, 398	5	-	-
	crypto_cipher_ kernel.c	20, 119, 120, 476	10	20, 119, 120	9	476	1
	crypto_qcrypt.c	119, 120, 563, 571	7	119, 120, 563, 571	7	-	-
	crypto_nettle.c	120, 398, 563	6	120, 398, 563	6	-	-
	crypto_nss.c	119, 120, 398	6	119, 120, 398	6	-	-
	crypto_openssl.c	119, 120, 398	6	119, 120, 398	6	-	-
	crypto_storage.c	119, 120, 563	5	119, 120, 563	5	-	-
gnupg (v2.3)	keydb.c	119, 120, 126, 398, 477, 571, 732	29	119, 120, 126, 398, 477, 571, 732	29	-	-
	keyid.c	119, 120, 126, 398, 476	30	119, 120, 126, 398	29	476	1
	encrypt.c	119, 120, 126, 563	10	119, 120, 126, 563	10	-	-
	keylist.c	119, 120, 126, 190, 398, 563, 686	37	119, 120, 126, 190, 398, 563, 686	37	-	-
	keyring.c	119, 120, 126, 362, 398, 477, 732	11	119, 120, 126, 398, 477, 732	10	362	1
	gpgcompose.c	119, 120, 126, 398, 786	47	119, 120, 126, 398, 786	47	-	-
	keyedit.c	119, 120, 126, 190, 398, 563, 571, 686, 758	71	119, 120, 126, 190, 398, 563, 571, 686, 758	71	-	-
Linux-pam (v1.5.1)	pam_access.c	119, 120, 126, 362, 398, 563, 571, 758	37	119, 120, 126, 398, 563, 571, 758	35	362	2
	pam_filter.c	78, 120, 126, 362, 398, 476, 590	29	120, 126, 398, 590	15	78, 362, 476	14
	pam_keyinit.c	134, 398	5	398	3	134	2
	pam_securetty.c	119, 120, 126, 362	12	119, 120, 126	9	362	3
	pam_xauth.c	20, 78, 119, 120, 126, 362, 367, 377, 398, 477, 807	40	20, 119, 120, 126, 377, 398, 477, 807	36	78, 362, 367	4
openssh (v8.4)	cipher.c	120, 126, 398, 401, 476, 477	10	120, 126, 398, 477	8	401, 476	2
	cipher-aes.c	120, 398	8	120, 398	8	-	-
openssl (v1.1.1)	aes_ige.c	119, 120, 398, 457, 563	29	119, 120, 398, 563	26	457	3
	aes_core.c	398, 563	2	398, 563	2	-	-

표 4. 도구별 오픈소스 취약점 탐지 정확도 (C: CWE ID, F: Flow, D: Detection accuracy)

오픈 소스명	소스 코드명	정적 분석 도구명										
		Cppcheck			Flawfinder			Yasca				
		C	F	D	C	F	D	C	F	D		
crypt setup	crypto_ cipher_ kernel.c	476	Null Pointer Dereference	Low	-			-				
gnupg	keyid.c	-			-			476	Null Pointer Dereference	Low		
	keyring.c	-			362	Race Condition	Medium	-				
Linux -pam	pam_ access.c	-			362	Race Condition	Medium	-				
	pam_ filter.c	476	Null Pointer Dereference	Low	78	OS Command Injection	High	-				
					362	Race Condition	Medium	-				
	pam_ keyinit.c	-			134	Uncontrolled Format String	High	-				
	pam_ securetty.c	-			362	Race Condition	Medium	-				
	pam_ xauth.c	-			78	OS Command Injection	High	-				
		-			362	Race Condition	Medium	-				
		-			367	TOUTOU Race Condition	High	-				
Open ssh	cipher.c	401	Memory Leak	Low	-			-				
		476	Null Pointer Dereference	Low	-			-				
Open ssl	aes_ige.c	457	Use of Uninitialized Variable	None	-			-				

탐지된 CWE ID의 소스코드 개수를 기준으로 각 도구의 성능을 비교하면, Yasca는 오직 keyid.c로 가장 성능이 낮으며, 그다음으로 Cppcheck가 crypto\_cipher\_kernel.c, cipher.c, aes\_ige.c로 3개, Flawfinder는 keyring.c, pam\_access.c, pam\_filter.c, pam\_keyinit.c, pam\_securetty.c, pam\_xauth.c는 6개로 가장 성능이 높은 것으로 나타났다.

탐지된 CWE ID의 탐지 정확도를 기준으로 각 도구의 성능을 비교하면, cppchek는 CWE-401, 457, 476가 탐지되었고, CWE-401

과 476의 탐지 정확도는 Low, CWE-457의 탐지 정확도는 None이다. Flawfinder는 CWE-78, 134, 362, 367을 탐지하였으며, 각각의 탐지 정확도는 High, High, Medium, High로 높은 정확도를 가지는 CWE ID를 가장 많이 탐지하였다. Yasca는 gnupg의 다른 도구들에서 탐지한 대부분의 CWE를 탐지하지 못하고 keyid.c에서 CWE-476만 탐지하였다. 따라서 탐지 정확도가 높고, 가장 많은 CWE를 탐지한 Flawfinder가 가장 성능이 높은 것으로 나타났다.

마지막으로 탐지된 CWE ID를 기준으로 각 도구의 성능을 비교하면, Cppcheck와 Yasca는 CWE-476인 Null Pointer Dereference를 가장 많이 탐지하였으며, Flawfinder는 CWE-362인 Race Condition을 가장 많이 탐지하였다. 따라서 각 도구가 탐지한 CWE가 중복되지 않고 상이하며, 이는 CWE ID를 기준으로 성능을 평가하기에는 한계가 있을 것으로 판단된다.

상기 각 도구의 성능 비교 결과를 토대로 성능 평가 결과를 분석하면, 탐지된 CWE ID의 소스코드 개수 및 탐지 정확도를 기준으로 Flawfinder가 가장 성능이 높은 것으로 판단된다.

#### IV. 결론

본 논문에서는 오픈소스 기반의 취약점 정적 분석 도구인 Cppcheck, Yasca, Flawfinder의 탐지 성능을 비교하고 분석하였다. 성능 비교를 위하여, 베패 오버플로우와 포맷 스트링 취약점이 명백하게 존재하는 샘플 소스코드를 대상으로 각 도구의 탐지 성능을 비교한 결과, 각 도구가 모두 다른 결과가 도출되었으며, Flawfinder만 존재하는 취약점을 모두 탐지하였다. 이를 통하여 각 도구의 결과가 상이한 문제점을 도출하였으며, 오픈소스를 대상으로 각 도구의 성능을 비교하고 분석하였다.

암호화 과정을 포함하는 오픈소스 5개를 선정하여 탐지된 CWE ID 및 개수를 비교하고 분석한 결과, 탐지된 모든 CWE의 성능을 비교하고 평가하기 위한 기준이 연구되지 않은 한계점이 존재하였다. 이에 따라, 탐지 정확도의 기준을 포함하는 CWE를 대상으로, 정적 분석 도구의 성능을 탐지된 CWE ID의 소스코드 개수 및 탐지 정확도를 기준으로 탐지 성능을 비교하고 분석하였다. 그 결과, Flawfinder 가 가장 성능이 높은 것으로 분석되었으며, Yasca가 가장 성능이 낮은 것으로 분석되었다.

이러한 결과가 도출되었음에도 불구하고, 각 도구가 탐지한 CWE가 중복되지 않고 상이하며, 이는 CWE ID를 기준으로 성능을 정량적으로 평가하기 어려운 한계점이 존재할 것으로

판단된다. 그뿐만 아니라, 표 2에 나타낸 것과 같이 기존 연구에서 모든 CWE-ID의 탐지 정확도가 연구되지 않은 한계점, 각 도구가 탐지한 CWE가 중복되지 않은 한계점도 포함하는 것으로 나타났다.

이러한 한계점을 극복하기 위하여, 향후 연구로, 공개된 CWE-ID의 탐지 정확도에 대한 연구를 진행할 예정이다. 그뿐만 아니라, 정적 분석 도구의 한계점인 오탐 및 미탐을 실증하기 위하여, 본 논문의 결과를 기반으로 동적 분석을 통하여 탐지된 취약점을 검증하는 연구를 진행할 예정이다.

#### 감사의 글

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음"(2019-0-01056)

#### [참고문헌]

- [1] 류원옥, 강신각, "주요 오픈소스 라이선스 현황," 한국통신학회 학회지 (정보와 통신), 제36권, 제11호, pp. 32-41, 2019년 10월.
- [2] 이용준, 안준선, 최진영, "보안약점 종류에 따른 정적분석 도구의 탐지 성능 연구," 한국정보과학회 학술발표논문집, pp. 272-274, 2019년 12월.
- [3] 서현지, 박영관, 김태환, 한격숙, 표창우, "Juliet과 STONESOUP 테스트 모음을 사용한 C/C++ 프로그램의 보안약점 탐지를 위한 정적 분석기 평가," 한국컴퓨터정보학회 논문지, 제22권 제3호, pp. 17-25, 2017년 3월.
- [4] C. L. Blackmon, D. F. Sang, and C. S. Peng, "Performance Evaluation of Automated Static Analysis Tools," GSTF Journal on Computing (JoC), vol. 2, no. 1, pp. 214-219, Apr. 2012.
- [5] Cppcheck, "A tool for static C/C++ code analysis", <http://cppcheck.sourceforge.net>,

2021년 2월 5일 접속.

- [6] Yasca, “Yet Another Source Code Analyzer,” <http://scovetta.github.io/yasca>, 2021년 2월 5일 접속.
- [7] Flawfinder, <https://dwheeler.com/flawfinder>, 2021년 2월 5일 접속.
- [8] NIST, “Test Suites,” <https://samate.nist.gov/SRD/testsuite.php>, 2021년 2월 5일 접속.
- [9] 양대일, “시스템 해킹과 보안: 정보 보안 개론과 실습 (3판),” 한빛아카데미, 2018년 11월 12일 출판.
- [10] GitLab, “cryptsetup,” <https://gitlab.com/cryptsetup/cryptsetup>, 2021년 2월 5일 접속.
- [11] GnuPG, “The Gnu Privacy Guard,” <https://gnupg.org>, 2021년 2월 5일 접속.
- [12] Linux Documentation, “pam.d(8) – Linux man page,” <https://linux.die.net/man/8/pam.d>, 2021년 2월 5일 접속.
- [13] OpenSSH, <https://www.openssh.com>, 2021년 2월 5일 접속.
- [14] OpenSSL, “Cryptography and SSL/TLS Toolkit,” <https://www.openssl.org>, 2021년 2월 5일 접속.

# 차분 프라이버시의 노이즈 메커니즘에 대한 분석

우타리예바 아셈\*, 신진명\*, 최윤호\*\*

\*부산대학교 (대학원생), \*\*부산대학교 (교수)

## A Survey on Noise Mechanisms of Differential Privacy

Utaliyeva Assem\*, Jinmyeong Shin\*, Yoon-Ho Choi\*\*

\*Pusan National University(Graduate student),

\*\*Pusan National University(Professor)

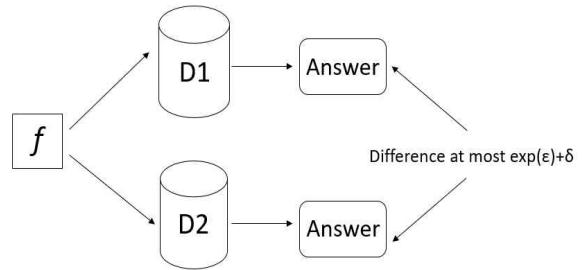
### Abstract

Since the threat to data privacy is increasing, preserving data privacy is an important issue of data publishing and mining tasks. Recently, Differential Privacy emerged as a state-of-art concept that provides strong mathematical guarantees. However, understanding Differential Privacy is not an easy task. In this paper, to help our readers understand Differential Privacy, we introduce concept of Differential Privacy and major noise mechanisms to achieve Differential Privacy.

## I.Introduction

Since the threat to data privacy is increasing, preserving data privacy is an important issue of data publishing and mining tasks.  $k$ -anonymity,  $l$ -diversity, and  $t$ -closeness are widely used to preserve data privacy. However, according to many research[3], such privacy models are not enough to guarantee data privacy.

Different from such conventional models, Differential Privacy(DP), which is introduced by C. Dwork, is resistant to existing data privacy attacks and guarantees data privacy mathematically. The intuitive concept of DP is to add randomness to all data which is published to third party. As shown in Fig. 1, Answers from database D1 and D2, which are different, but satisfy DP, are similar in



**Fig. 1** An Conceptual Overview of Differential Privacy

bound of  $\exp(\epsilon)+\delta$ . This characteristic makes it impossible to recognize the difference between two databases.

Because of such an advantage, DP is widely studied and applied in many fields. However, since the way of achieving DP is a mathematically hard problem, it is not an easy task to understand the detailed concept of DP. To handle such problem, we introduce a detailed concept of DP and summarize

noise mechanisms to achieve DP.

The paper is organized as follows. First, we explain the concept of DP in section 2. Then, representative noise mechanisms are described in section 3. In section 4, we compare the pros and cons of each noise mechanism. Then, we summarize and conclude the paper in section 5.

## II. Differential Privacy

In this section, we describe key concepts of DP, which are DP's definition and sensitivity.

**Definition 1** ( $\epsilon$ ,  $\delta$ - differential privacy) A randomized algorithm  $K$  gives  $\epsilon$ ,  $\delta$ - DP if for all data sets  $D_1$  and  $D_2$  differing on at most one element, and all  $S \subseteq Range(K)$ ,

$$\frac{\Pr[K(D_1) \in S]}{\Pr[K(D_2) \in S]} \leq \exp(\epsilon) + \delta \quad (1)$$

Equation 1 means that the presence or absence of a user in the dataset can cause at most  $\exp(\epsilon) + \delta$  change. Here,  $\epsilon$  is a privacy parameter, and  $\delta$  is a relaxation parameter. A strict version is called  $\epsilon$ -DP when  $\delta$  is 0.

**Definition 2** Sensitivity of a function  $f: D \rightarrow R^w, w \in N^+$  is the maximum difference that absence of one individual in the data set can change.

$$\Delta f = \max \|f(D_1) - f(D_2)\|_1 \quad (2)$$

$$\Delta f^2 = \max \sqrt{\|f(D_1) - f(D_2)\|^2} \quad (3)$$

The equation 2 is definition L1-sensitivity and the equation 3 is definition of L2-sensitivity.

## III. Noise Mechanisms

### 3.1 Laplace Mechanism

Laplace Mechanism is the most common mechanism to satisfy  $\epsilon$ -DP. It perturbs the

output by adding random noise to the true result of the query. The noise can be drawn using the probability density function(PDF) in equation 4.

$$Lap(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) \quad (4)$$

The amount of noise is calibrated according to the privacy parameter  $\epsilon$  and the sensitivity of the query  $\Delta f$ . The result of query is same as equation 5.

$$\tilde{f}(x) = f(x) + Lap\left(\frac{\Delta f}{\epsilon}\right) \quad (5)$$

### 3.2 Gaussian Mechanism

Gaussian Mechanism is another output perturbation mechanism that uses the noise from Gaussian Normal Distribution. The PDF of Gaussian Noise is described in equation 6.

$$Gaussian(\mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \quad (6)$$

Random noise can be calculated by substituting values of mean  $\mu = 0$  and  $\sigma^2 = (2\ln(1.25/\delta) \times \Delta f^2)/\epsilon^2$ . Therefore, the result of query is same as equation 7.

$$\tilde{f}(x) = f(x) + Gaussian(0, \sigma) \quad (7)$$

### 3.3 Exponential Mechanism

The main idea behind the Exponential Mechanism is the utility function  $u: N^{|D|} \times R \rightarrow R$ . The utility function  $u(D, r)$  represents how good output  $r$  is for database  $D$ . As shown in equation 8, the mechanism selects and outputs an element  $r \in R$  with probability.

$$\Pr(r) \propto \exp\left(\frac{\epsilon u(D, r)}{2\Delta u}\right) \quad (8)$$

In this equation,  $\Delta u$  is L1-sensitivity of utility function  $u$ .

Mechanism	Laplace	Gaussian	Exponential
Differential Privacy	$\epsilon$ -DP	$\epsilon, \delta$ -DP	$\epsilon$ -DP
Data Type	Numeric	Numeric	Categorical
Mechanism Type	Output Perturbation	Output Perturbation	Scoring function
Additive Noise	$Lap(\frac{\Delta f}{\epsilon})$	$Gaussian(0, \sigma)$	-
Sensitivity	L1 norm	L2 norm	-

**Table 1.** Comparison of Differentially Private Mechanisms

#### IV. Comparison of Mechanisms

Mechanisms to achieve DP can be categorized into output perturbation and scoring function. The output perturbation type adds additive noise to the query output. Differently, the scoring function type returns the stochastic best element which satisfies the query.

One major difference between two output perturbation mechanisms, which are Laplace and Gaussian, is the sensitivity. Since two mechanisms are using different norms as sensitivities, the Laplace mechanism shows better data utility when the sensitivity is low. On the contrary, the Gaussian mechanism shows better data utility when the sensitivity is high.

For instance, if the user will affect just one statistics the both L1 and L2 sensitivities of the query will be equal to 1. In this case, the Laplace mechanism adds  $Lap(1/\epsilon)$  amount of noise. This is much less than the noise generated by the Gaussian mechanism. However, when one user affects more than one statistic, the value of L2-sensitivity becomes less than L1-sensitivity. In the case of the 50 counting queries, the L1-sensitivity of a query is 50, whereas the L2-sensitivity of a query is  $\sim 7.07$ . Thus, the Gaussian mechanism is more appropriate to use in cases with multiple complex statistics in order to prevent high amounts of additive noise and degradation of the data utility.

Since output perturbation types are suitable

only for numeric queries, the Exponential mechanism, which is the scoring function type, is designed to handle categorical data. Exponential mechanism does not add any noise directly to the answer but returns the best element that satisfies the condition. To satisfy DP, it sometimes returns the element with not the highest score. The mechanism provides higher utility since no additive noise was used.

#### V. Conclusion

In this paper, we reviewed the basic noise generating mechanisms of DP and their features. All previous work was focused on designing tailored mechanisms for specific data analysis. Since the application of DP is expanding to different areas, the regular noise generating mechanisms are not practical enough.

#### Acknowledgement

This research was supported by the MSI T, Korea, under the ITRC support program(II TP-2020-0-01797) and BK21 Four, Korean Southeast Center for the 4th Industrial Revolution Leader Education.

#### [References]

- [1] C.Dwork, A.Roth, The Algorithmic Foundations of Differential Privacy, 2014
- [2] F.McSherry, K.Talwar, Mechanism Design via Differential Privacy, 2007
- [3] C.Clifton, T.Tassa, On Syntactic Anonymity and Differential Privacy, 2013