

# 블록체인 기반 신원인증 및 인공지능 기술을 활용한 IoT 침입 방지 시스템

방수진, 윤유준, 장성원, 조은희, 이광재\*

상명대학교 정보보안공학과

sujean2002@naver.com, aladin409@naver.com, jjang6251@gmail.com,

3eunhui3@gmail.com, \*beglearn@smu.ac.kr

## An IoT Intrusion Prevention System using Blockchain and Artificial Intelligence Technology

Sujean Pang, YuJoon Yoon, SungWon Jang, EunHui Cho, KwangJae Lee

Dept. of Information Security Engineering, Sangmyung University

### 요약

스마트 홈 확대로 생활이 편리해졌지만, 홈 카메라 해킹 등 보안 취약점을 악용한 사생활 침해 사례가 증가하고 있다. 현재 IoT 보안은 제 3 자의 개입을 통한 신원인증 방식으로 크리덴셜 스테핑에 취약하며, IoT 장치를 위한 인공지능 기술은 다양한 방식으로 개발 중이다. 본 논문은 블록체인 기반 신원인증과 인공지능 기반 침입 탐지를 결합한 IoT 침입 방지 시스템을 제안한다. 이 시스템은 검증가능한 자격증명을 통해 인증을 수행하며, CatBoost 모델을 활용해 이상 징후를 실시간으로 탐지하고 차단한다. 제안한 탐지 모델은 95.8%로 기존 연구보다 3.4% 향상시켜 비정상 트래픽에 대한 탐지 효율성을 크게 높였다. 이를 통해 IoT 환경의 신뢰성과 안전성을 확보할 수 있다.

### Abstract

The expansion of smart home use has made life more convenient, but there has been an increase in privacy invasion cases exploiting security vulnerabilities such as home camera hacking. Currently, IoT security is vulnerable to credential stuffing due to identity authentication through third-party intervention, and artificial intelligence for IoT devices are being developed in various ways. This paper proposes an IoT intrusion prevention system that combines blockchain-based identity authentication and artificial intelligence-based intrusion detection. This system performs authentication through verifiable credentials and detects and blocks abnormal signs in real time using a CatBoost model. The proposed detection model significantly increases the detection efficiency for abnormal traffic by 3.4% to 95.8% compared to previous studies. This can contribute to the reliability and safety of the IoT environment.

### 1. 서론

IoT 기술의 급속한 발전과 보급으로 연결된 장치의 수가 기하급수적으로 증가하면서 IoT 보안의 중요성이 강조되고 있다. 먼저, ThroughTek Kalay 플랫폼의 오디오 및 비디오 데이터 유출 위험 사례는 IoT 기기를 통한 개인정보 유출의 심각성을 보여주는 큰 예시이다 [1]. 또한, Roku 플랫폼에서 발생한 57 만 건 이상의 크리덴셜 스테핑과 TP-Link Archer AX21 Wi-Fi 라우터에서 발견된 Mirai 봇넷 취약점은 IoT 보안의 취약성을 명확히 드러낸다 [2, 3]. 이러한 사례들은 데이터 유출 및 개인정보 침해와 나아가 대규모 분산 서비스 거부(DDoS, Distributed Denial of Service) 공격의 매개체로 악용될 가능성이 높다. 기존의 IoT 보안 연구들은 이러한 문제를 해결하기 위해 침입 탐지 시스템(IDS, Intrusion Detection System)의 로그 분석과 신원인증을 제시하였다. 전통적인 규칙 기반 침입 탐지 시스템은 미리 정의된 패턴과 시그니처를 기반으로 공격을 탐지하였으나, 새로운 유형의 공격에 대한 대응이 어렵다는 한계를 보였다 [4]. 이를 보완하기 위해 대규모 데이터에서 숨겨진 패턴을 학습하고 새로운 공격 유형에도 유연하게 대응할 수 있는 인공지능(Artificial Intelligence, AI) 접근법을 도입하였다. 먼저, 기계학습 접근법으로는 Support Vector Machine 이나 Random

Forest 등의 알고리즘이 제시되었고, 이를 활용하여 새로운 유형의 비정상 트래픽을 탐지하였다. 하지만 대용량 트래픽 처리 시 성능 저하가 발생하는 문제가 있었다 [5]. 또한, 딥러닝 접근법은 고차원적인 특징을 자동으로 학습하고 복잡한 데이터 구조를 효과적으로 처리할 수 있었다. 그러나 실시간 처리가 요구되는 IoT 환경에서는 여전히 성능상의 제약이 존재했다 [6]. 신원인증 측면에서는 인증 서버를 통한 전통적인 사용자 인증 방식이 주로 사용되어 왔다. 이러한 방식에는 사용자 ID 와 비밀번호를 기반으로 한 인증, One-Time Password 인증, 그리고 보안 토큰을 활용한 인증 등이 포함된다. 이러한 중앙화된 서버 구조는 단일 실패 지점(Single Point of Failure)의 위험을 내포하고 있다 [7]. 또한 선행 연구들은 주로 데이터셋을 활용하여 AI 기반 침입 탐지 시스템을 구축하는 데 초점을 맞췄으며, IoT 환경의 특수성을 고려한 신원인증 기술의 필요성을 간과하는 경우가 많았다 [8].

본 연구에서는 AI 모델과 블록체인 기반 분산신원인증 (Decentralized Identity, DID) 기술을 결합한 하이브리드 보안 시스템을 제안한다. 제안된 하이브리드 방식은 DID 기반 인증을 통해 사용자의 신원을 안전하게 확인하고, AI 기반 IoT 침입 탐지 시스템을 통해 네트워크를 실시간으로 보호함으로써 이중 보안 체계를 구현하였다. 이러한 이중

보안 구조는 기존 연구와 달리 단일 실패 지점 문제를 해결하는 동시에 AI 모델의 높은 탐지 정확도를 바탕으로 IoT 보안의 신뢰성을 크게 향상시킬 수 있다.

## 2. 본론

본 논문에서 제안하는 DID 및 AI 기술을 활용한 IoT 침입 방지 시스템은 그림 1 과 같다. 이 시스템은 비정상 트래픽 탐지를 결합하여 IoT 기기의 접근 신뢰성을 보증하고 데이터를 안전하게 관리한다. 사용자는 웹 브라우저를 통해 시스템에 접속하여 블록체인 네트워크에서 검증가능한 자격 증명인 Verifiable Credential(VC)을 발급받고 DID 를 활용해 JSON Web Token(JWT)를 생성한다. JWT 는 사용자 인증 정보를 포함하는 디지털 토큰으로, 이를 통해 사용자는 신원을 인증한다. 인증된 사용자의 요청은 웹 서버를 통해 IoT 기기에 전달되며, IoT 네트워크 트래픽은 AI 서버에서 실시간으로 분석된다. 이 서버는 비정상 트래픽을 탐지해 차단하고, 정상 트래픽만 허용하여 IoT 기기와의 안전한 데이터 교환을 지원한다. 또한, 웹 서버는 IoT 기기로부터 수집된 데이터를 관리하고, 클라이언트 요청에 따라 필요한 정보를 제공한다.

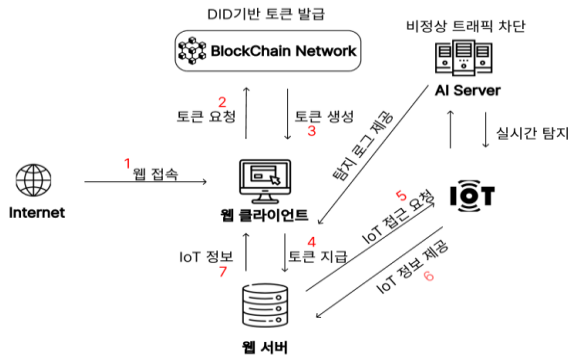


그림 1. 제안하는 DID 및 AI 기술을 활용한 IoT 침입 방지 시스템

### 2.1.2 단계 신원인증

DID 신원인증은 중앙 서버에 의존하지 않고, 블록체인 상에서 사용자의 신원 정보를 관리하는 분산 신원 인증 방식으로, 데이터 위조 및 탈취의 위험을 최소화하는 데 효과적이다. 그러나 DID 인증만으로는 실제 사용자가 인증된 DID 정보와 일치하는지 완벽히 보장하기 어렵다. 이를 보완하기 위해 본 시스템은 카카오톡 로그인과 MetaMask 지갑 주소를 결합하여 사용자 신원 확인의 신뢰성을 높였다. 이러한 2 단계 신원인증 구조는 사용자 경험을 저해하지 않으면서도 보안성을 강화하는 데 중점을 두고 있다 [12]. 추가적으로 사업자는 신원 확인의 신뢰성을 더욱 높이기 위해 신분증 인증 및 위조 검증, 전화번호 인증과 같은 다양한 인증 수단을 선택할 수 있다. 신분증 인증은 사용자의 신분증을 OCR 기술로 스캔하고 위조 여부를 검증하는 방식으로 이루어지며, 전화번호 인증은 SMS 인증 코드를 통해 사용자 소유 여부를 확인하는 방법이다. 이러한 추가 인증 옵션은 시스템의 유연성을 제공하면서도 다양한 상황에 맞게 보안 수준을 강화한다.

### 2.2 IoT 침입 탐지 AI 모델 설계

IoT 침입 탐지 AI 모델은 IoT 환경의 특수성을 고려하므로, 제안하는 탐지 시스템에 적용할 AI 모델의 선정하기 위해서 기계학습과 딥러닝 모델을 모두 고려하였다. CatBoost는 전통적인 기계학습 모델로 빠른 학습, 높은 성능, 그리고 해석 가능성을 제공하여 데이터셋의 구조적 특성을 잘 활용하는 데 적합하다. 반면 MLP는 딥러닝 모델로서 복잡한 데이터 간의 관계를 학습하고 고차원 데이터를 처리하는 데 강점을 가지며, 딥러닝 연구의 기초 모델로 활용되기에 적합하다. 본 논문에서는 이 두가지 모델을 모두 구현하고, 성능평가를 통해 최종 AI 모델을 선정한다.

### 2.2.1 데이터셋

본 논문에서는 캐나다 통신보안기구와 사이버보안 연구소가 공동으로 개발한 CSE-CIC-IDS2018 데이터셋을 사용하였다 [13]. 이 데이터셋은 실제 네트워크 환경을 모사하여 생성된 현대적인 네트워크 침입 탐지용 데이터셋으로, IoT 환경에서 자주 발생하는 Brute-Force 공격, 웹 공격, DDoS 공격, 봇넷 등 다양한 공격 시나리오를 포함하고 있다. 네트워크 플로우 기반으로 구성된 각 레코드는 프로토콜 유형, 패킷 길이, 플로우 지속 시간, 플로우당 패킷 수 등 다양한 네트워크 특성을 포함하고 있어, IoT 장치들의 통신 패턴을 효과적으로 분석하고 특히 이 데이터셋은 최신 공격 패턴을 포함하고 있으며, 레이블링이 잘 되어 있어 침입 탐지 모델 개발에 적합하다는 점에서 데이터셋 활용을 결정하였다.

### 2.2.2 전처리 및 모델 훈련

CatBoost 모델의 데이터 전처리 과정은 Flow ID, IP 주소 등 불필요한 칼럼을 제거하고, 결측값은 평균값으로 대체하였으며, Min-Max Scaling 을 적용하여 정규화하였다. Protocol 과 Dst Port 는 범주형 변수로 변환하였고, 기존 연구가 데이터 처리의 효율성을 위해 일부 정상('Benign') 데이터를 제거한 것과 달리, 본 연구는 모든 원본 데이터를 유지하고 RandomOverSampler 기법으로 클래스 불균형 문제를 해결하였다 [9]. CatBoostClassifier 는 GPU 가속을 활용해 학습률 0.1, 깊이 10, 반복 횟수 1,000 회로 설정하였으며, Early Stopping 20 회를 적용하여 과적합을 방지하였다. 이러한 포괄적 데이터 활용 방식과 최적화된 모델 설정을 통해 95.81%의 정확도를 달성하여 기존 연구의 92.41% 정확도를 크게 상회했다 [9]. 이를 통해 IoT 네트워크 데이터 분석에서 높은 정확도와 안정적인 성능을 확보하였다.

MLP 모델 데이터 전처리 과정에서는 먼저 데이터 품질 향상을 위해 중복 데이터를 제거하고 결측치는 중위수로 대체하였다. Timestamp 등 불필요한 칼럼을 제거하고, IP 주소는 정수형으로 변환하였으며, Flow ID 와 Dst Port, Protocol 은 LabelEncoder 를 사용하여 수치화 했다. 모든 특성은 StandardScaler 를 사용하여 정규화 하였으며, 스케일러 모델을 별도 저장하여 전처리 일관성을 보장하였다. 클래스 불균형 문제와 딥러닝에 필요한 대량의 데이터셋은 오버샘플링 방법 중 하나인 SMOTE 기법을 통해 해결하였다 [14]. 모델은 ReLU 활성화 함수를 사용하는 3 개의 은닉층(100, 75, 50 노드)으로 구성하였으며, 과적합 방지를 위해 드롭아웃 (rate=0.3)과 L2 정규화( $\alpha=0.0001$ )를 적용하였다.

### 2.2.3 성능평가 및 모델 선정

모델의 성능 평가를 위해 다양한 평가 지표를 활용하였다. 표 1 은 각 모델의 클래스별 상세 성능을 보여주며, 표2는 모델의 전반적인 성능을 나타내는 지표들을 보여준다. 여기서 Precision 지표는 특정 클래스로 예측한 결과 중 실제로 해당 클래스인 비율을, Recall 지표는 실제 해당 클래스 중 올바르게 예측한 비율을 의미한다. F1-score 지표는 Precision 과 Recall 의 조화평균으로, 두 지표 간의 균형을 고려한 성능 지표이다. Support 지표는 각 클래스의 샘플 수를 나타낸다. Accuracy 지표는 전체 예측 중 올바른 예측 비율을, Area Under the Curve(AUC) 지표는 ROC 곡선 아래 면적으로 모델의 분류 성능을, Matthews Correlation Coefficient (MCC) 지표는 클래스 불균형을 고려한 지표로, 값이 1 에 가까울수록 성능이 우수함을 의미한다 [15].

표 1. 모델별 상세 성능평가

Model	Class	Precision	Recall	F1-score	Support
Cat Boost	정상	0.9826	0.9689	0.9757	1,027,751
	비정상	0.8116	0.8864	0.8473	155,201
MLP	정상	0.9403	0.9333	0.9368	2,681,768
	비정상	0.9338	0.9407	0.9373	2,682,421



### 3.3 전체 연동 테스트

AI 모델의 비정상 트래픽 탐지 성능을 검증하기 위해, IoT 환경에서 가장 흔히 발생하는 두 가지 취약점을 기반으로 시나리오를 구성하여 테스트를 진행하였다. 첫 번째 시나리오는 DDoS 공격으로, 다수의 클라이언트 특정 서버나 네트워크에 과도한 트래픽을 유발하여 정상적인 서비스를 방해하거나 중단시키는 대표적인 공격 방식이다. DDoS 공격은 서버 과부하로 인한 서비스 중단, 정당한 사용자의 접근 차단, 시스템 성능 저하 및 기업 평판 손상과 같은 심각한 문제를 초래할 수 있다. 이 시나리오에서는 9000 번 포트를 대상으로 대량의 트래픽을 생성하여 공격을 모의하였다. 테스트 결과, 모델은 이 비정상 트래픽을 성공적으로 탐지하였으며, 공격자의 IP 를 차단하고 해당 정보를 사용자에게 전송하였다. 이를 통해 DDoS 공격에 대한 방어 가능성과 대응 체계의 효과성을 입증하였다. 두 번째 시나리오는 Brute-Force 공격으로, 무작위로 비밀번호나 인증 정보를 반복적으로 시도하여 시스템에 접근하려는 대표적인 공격 방식이다. Brute-Force 공격은 인증 서버에 과도한 부하를 발생시키고, 계정 탈취로 이어질 경우 민감한 정보 유출이나 시스템 권한 탈취와 같은 심각한 보안 문제를 초래할 수 있다. AI 가 비정상 트래픽을 효과적으로 탐지하는지 검증하기 위해, 9000 번 포트에 임의의 로그인 기능을 구현한 후 이를 대상으로 Brute-Force 공격을 시도하였다. 또한, 그림 6 과 같이 일부러 ID 와 비밀번호가 맞는 조합을 설정하여 공격이 성공했을 때도 AI 가 이를 탐지하고 차단하는지 확인하였다. 테스트 결과, 모델은 이러한 비정상적인 시도를 성공적으로 탐지하고 차단하였으며, 그림 7 과 같이 공격자의 IP 와 관련 정보를 사용자에게 전송하였다. 이를 통해 Brute-Force 공격에 대한 방어 가능성과 AI 의 대응 체계의 효과성을 확인하였다. 이와 더불어, 테스트 환경에서 웹 서버와 클라이언트 간 통신은 HTTPS 프로토콜을 사용하여 암호화되었으며, 이를 통해 중간자 공격에 의한 데이터 탈취 가능성을 효과적으로 방지하였다. 이는 AI 모델과 DID 기반 인증 구조를 활용한 시스템의 보안을 보완하고, 전송 계층의 신뢰성을 높이는 데 기여하였다.

```
Trying Username: test | Password: guest
[FAILED] Username: test | Password: guest
Trying Username: test | Password: root
[FAILED] Username: test | Password: root
:
:
:
Trying Username: admin | Password: iotping
SUCCESS Username: admin | Password: iotping
Trying Username: user1 | Password: iotping
[FAILED] Username: user1 | Password: iotping
Trying Username: user1 | Password: password123
SUCCESS Username: user1 | Password: password123
Trying Username: guest | Password: iotping
[FAILED] Username: guest | Password: iotping
Trying Username: guest | Password: password123
```

Brute-Force 공격 중 시도했으나 실패한 사용자 ID와 비밀번호

Brute-Force 공격을 통해 얻은 사용자 ID와 비밀번호

그림 6. Brute-Force 공격 시도 결과



그림 7. AI 모델이 전송한 차단된 트래픽 로그 알림

### 4. 결론

본 연구는 AI 모델과 DID 기술을 결합한 하이브리드 IoT 보안 시스템을 제안하였다. 제안된 시스템은 블록체인 기술이 적용된 탈중앙화된 인증 구조로 중앙 서버 의존성을 제거하여 단일 실패 지점 문제를 해결하였으며, 데이터 위변조 및 탈취의 위험을 효과적으로 감소시켰다. 또한, AI 기반 침입 탐지 시스템은 CatBoost 모델을 사용하여 네트워크에서 발생하는 비정상 트래픽을 실시간으로 탐지하고 차단하는 데 성공하였다. 나아가 탐지 모델의 정확도가 95.8%로 기존 연구보다 3.4% 향상시켜 IoT 환경에서의 보안 문제를 효과적으로 해결할 수 있는 실질적인 솔루션을 제공함을 증명하였다. 특히, 본 연구는 기존 연구와 달리 실제 공격 시나리오를 기반으로 실증적 검증을 수행하여 시스템의 실용성을 명확히 입증하였다. 또한, 기존 연구가 블록체인을 IoT 데이터를 저장하거나 검증하는 용도로만 활용한 것과 달리, 본 연구에서는 블록체인 기반 DID 기술을 통해 사용자 인증을 수행하여 IoT 보안을 한층 강화하였다. 향후 연구에서는 새로운 유형의 사이버 공격에 대한 AI 모델의 탐지 성능 고도화와 다양한 IoT 환경에서의 시스템 적용 가능성을 검토할 예정이다.

### [참고문헌]

- [1] [Internet], Available: <https://securityonline.info/millions-of-iot-devices-vulnerable-after-researchers-uncover-flaws-in-thoughtek-kalay-platform/>
- [2] [Internet], Available: <https://www.bleepingcomputer.com/news/security/roku-warns-576-000-accounts-hacked-in-new-credential-stuffing-attacks/>
- [3] [Internet], Available: <https://hackread.com/iot-cameras-exposed-by-chainable-exploits/>
- [4] 신경일 *et al.*, "IDS 알고리즘에 대한 탐지율 연구 비교", 2017 년 춘계학술발표대회 논문집, 제 24 권, 제 1 호, pp. 223-226, 2017.
- [5] 이상호, 정민우, "SVM 과 Random Forest 를 이용한 네트워크 이상탐지 시스템의 성능 평가", 한국통신학회논문지, 제 47 권 제 8 호, pp.1432-1440, 2022.
- [6] 하희리, 백운홍, IoT 환경을 위한 인공지능 기반 네트워크 침입 탐지 시스템, 서울대학교 대학원 학위논문, 전기정보공학부, p.1-16, 2021.
- [7] Nilesh A. Lal, Salendra Prasad, Mohammed Farik, "A Review Of Authentication Methods", *International Journal of Scientific & Technology Research*, vol. 5, no. 11, pp. 246-249, Nov. 2016.
- [8] Sohaib Hanif *et al.*, "Intrusion Detection in IoT Using Artificial Neural Networks on UNSW-15 Dataset", *Proceedings of the 2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT)*, pp. 135-141, 2019.
- [9] 알리카노브 주마베크, 양 승 삼, 노 영 태, "불균형 CIC-IDS-2018 데이터 세트에 대한 CatBoost 기반 네트워크 침입 탐지", 한국통신학회논문지, 제 46 권 제 12 호, pp. 2191-2197, 2021.
- [12] 권준우 *et al.*, "블록체인 기반 분산신원증명의 이해와 서비스 적용 사례", *ACK 2021*, 제 28 권, 제 2 호, pp. 309-312, 2021.
- [13] 김수환 *et al.*, "CSE-CIC-IDS2018 데이터를 활용한 딥러닝 기반 네트워크 침입 탐지 시스템", 2023 년 대한전자공학회 하계학술대회, pp. 2221-2224, 2023.
- [14] 함기봉, 김문석, "데이터 불균형 문제에 따른 네트워크 침입 탐지를 위한 딥러닝 모델의 성능 비교", 한국통신학회 추계종합학술발표회, pp. 1077-1078, 2023.
- [15] Davide Chicco, Giuseppe Jurman, "The Advantages of the Matthews Correlation Coefficient (MCC) Over F1 Score and Accuracy in Binary Classification Evaluation", *BMC Genomics*, vol. 21, no. 6, pp. 1-13, 2020.