A WBC-MAC Framework for Secure Integrity Verification of On-Device AI Models [*]

Han-Bin Lee[1], JunYong Cho[1], Hyeon Park[1], and Taesung Kwak[1] Mee Lan Han[1][‡], and TaeGuen Kim[1][§]

[1] Korea University, 2511 Sejong-ro, Sejong, 30019, Republic of Korea
{2019270119, 0808jim, rainbow00000, blosst, tg$_k$$im$, ss020312}@$korea.ac.kr$
[2] Sejong University, Seoul, 05006, Republic of Korea

Abstract

On-device AI models provide privacy and independence from network connectivity, but their direct deployment exposes them to tampering, redistribution, and reverse engineering under white-box attacks. Conventional CBC-MAC and CMAC methods are insufficient due to structural weaknesses and symmetric key exposure. This paper presents a White-box Cryptography-based MAC framework that enables integrity verification without key distribution, supports public, private, and offline environments, providing a lightweight and practical protection mechanism, including isolated or resource-constrained devices.
Keywords: On-device AI, Integrity Verification, White-box Cryptography, Message Authentication Code (MAC), Embedded Security

1 Introduction

The recent shift from cloud-based computation to on-device AI has improved privacy protection, where inference is performed directly on user devices [**?**]. This approach provides advantages in privacy protection and reduced network dependency, but it also exposes models to white-box threats such as tampering, unauthorized redistribution, and reverse engineering. Existing protection mechanisms can be categorized into three types: (1) model encryption, which prevents parameter exposure but incurs decryption overhead and key management issues; (2) MAC-based integrity verification (e.g., CMAC), which suffers from symmetric key exposure risks; and (3) hardware-based security such as TEEs, which face scalability and compatibility limitations. Although digital signatures can ensure integrity, they rely on certificate authority (CA) infrastructures and are unsuitable for constrained or offline on-device environments. To address these challenges, this paper proposes a White-box Cryptography-based MAC framework that enables integrity verification without directly distributing secret keys, mitigating key exposure risks in device-level deployment environments.

[*]This work was supported by the IITP(Institute of Information Coummunications Technology Planning Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT)(IITP-2025-RS-2022-00164800). This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No. RS-2025-25411243. This work was supported by the Institute of Information Communications Technology Planning Evaluation(IITP) grant funded by the Korea government (MSIT) (No.RS-2025-02215590, Development of AI implementation obfuscation technology to prevent information leakage in On-Device AI))

[†]Proceedings of the 9th International Conference on Mobile Internet Security (MobiSec'25), Article No. P-29, December 16-18, 2025, Sapporo, Japan.
[‡]Corresponding Author
[§]Corresponding Author

Algorithm	Execution Time (μs)
WBC-Chow	37.7566
WBC-Xiao	131.1981
WBC-Full	3791.1668
AES	0.0293
DES	0.3387
TDES	0.8625
ED25519	55.8950
RSA_PSS_2048	28.4540
ECDSA_P_256	66.1290

Figure 1: Comparison of execution performance among White-Box cryptography, conventional cryptography, and digital signature algorithms. The reported average time was obtained from 10 iterations with 32-byte input data.

2 Proposed Framework

This section describes the structure and design principles of the proposed security framework for integrity verification of on-device AI models.

The framework supports three deployment environments: public, private, and offline. In public or private networks, integrity is verified through a CA–based digital-signature infrastructure, with private CAs used for internal verification at the cost of higher administrative overhead. In offline or isolated environments, where external servers or certificate authorities are unavailable, the framework employs a WBC–MAC scheme to perform local integrity verification without key exposure. The WBC-MAC structure addresses the symmetric-key distribution problem inherent in conventional CBC-MAC and CMAC, ensuring lightweight yet secure validation suitable for embedded and industrial devices.

2.1 MAC Generation Algorithm Design

The framework employs a White-box Cryptography (WBC)–based MAC mechanism, inspired by prior studies on MAC-based integrity approaches [**?**], but optimized for on-device deployment. The proposed structure is derived from the CMAC scheme, replacing the symmetric encryption function with a WBC function to prevent key extraction. This design allows encryption to be performed without directly exposing the secret key on the device, thereby providing a more secure integrity verification mechanism against white-box attacks.

3 Cryptographic Performance Comparison

We conducted a benchmark to evaluate the performance of the proposed WBC implementation. The results were compared against conventional cryptographic algorithms, including AES with hardware acceleration, DES, TDES, and software-based AES implementations, as well as digital signature schemes such as RSA-PSS 2048, ECDSA P-256, and Ed25519. For WBC, we evaluated and compared three representative implementations: Chow, Xiao, and Full [**?**, **?**, **?**].

4 Conclusion and Future Works

This paper presented a WBC-MAC framework designed to enhance the integrity verification of on-device AI models. The proposed method serves as an alternative in scenarios where hardware-based security solutions—such as digital signatures using public/private Certificate Authorities (CAs) or Trusted Execution Environments (TEEs)—are unavailable, thereby improving the trustworthiness of on-device AI model deployment. To validate the framework, we conducted a comparative benchmark against conventional symmetric-key encryption algorithms, digital signature schemes, and publicly available WBC implementations. Future work will focus on experimental evaluation under realistic attack scenarios, performance optimization for lightweight deployment, and integration with hardware-assisted protection layers.

A WBC-MAC Framework for Secure Integrity Verificationof On-Device AI Models

Han-Bin Lee, JUN-YOUNG CHO, HYEON PARK, TAESUNG KWAK, and Mee Lan Han†,and TAE GEUN KIM

Dept. of Cyber Security, Korea University

Abstract

- On-device AI enhances privacy and independence from network connectivity, yet direct deployment exposes models to tampering, redistribution, and reverse engineering. Conventional CBC-MAC and CMAC are limited by structural weaknesses and symmetric key exposure.
- This study introduces a White-box Cryptography-based MAC framework that enables integrity verification without key distribution, offering a lightweight and practical protection mechanism for public, private, and offline environments.

- **Keywords:** On-device AI, Integrity Verification, White-box Cryptography, MAC, Embedded Security

Introduction

- The shift from cloud-based AI to on-device inference enhances privacy and reduces network dependency, yet it exposes models to white-box threats such as tampering and reverse engineering.
- Existing protections—model encryption, MAC-based verification, and TEE-based security—face challenges of overhead, key exposure, and limited scalability. Digital signatures also rely on CA infrastructures, making them unsuitable for offline environments.
- To overcome these limitations, this study proposes a White-box Cryptography-based MAC framework that enables integrity verification without secret key distribution, mitigating key exposure risks in on-device deployments.

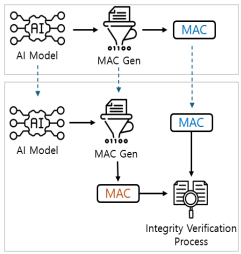

Figure 1: Integrity verification in offline environments using WBC-MAC.

Proposed Framework

- The framework ensures integrity verification of on-device AI models through a **White-box Cryptography (WBC)–based MAC** mechanism, addressing key exposure and scalability issues in conventional methods.
- **Public** – verified via CA-based digital signatures
- **Private** – internal CA for local validation
- **Offline** – WBC-MAC enables verification without key exposure

Algorithm Design:

- WBC-MAC can be used for integrity verification on devices where using or deploying a Certificate Authority (CA) is difficult, or where hardware-based security such as TEE is unavailable.
- WBC-MAC is derived from the CMAC structure but replaces the symmetric encryption process with White-Box encryption, enabling secure integrity verification without exposing key information to end users.

Algorithm	Execution Time (μs)
WBC-Chow	37.7566
WBC-Xiao	131.1981
WBC-Full	3791.1668
AES	0.0293
DES	0.3387
TDES	0.8625
ED25519	55.8950
RSA_PSS_2048	28.4540
ECDSA_P_256	66.1290

Table 1: Comparison of execution performance among White-Box cryptography, conventional cryptography, and digital signature algorithms. The reported average time was obtained from 10 iterations with 32-byte input data.

Conclusion

- This paper proposed a White-box Cryptography-based MAC framework to enhance the integrity verification of on-device AI models. The proposed method modifies the conventional CMAC structure to enable integrity verification without directly distributing secret keys, thereby enhancing the reliability of on-device AI model deployment. Future research will focus on experimental evaluation under realistic attack scenarios, performance optimization for lightweight deployment, and integration with hardware-assisted protection layers.

Acknowledgements
This work was supported by the IITP(Institute of Information & Coummunications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT)(IITP-2025-RS-2022-00164800). This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No. RS-2025-25411243)
This work was supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government (MSIT) (No.RS-2025-02215590, Development of AI implementation obfuscation technology to prevent information leakage in On-Device AI)

Figure 2: PDF 1