Risk Analysis of Adversarial Attacks on Biometric Systems

Seong Hee Park and Youn Kyu Lee*
Department of Computer Engineering
Hongik University
Seoul, Republic of Korea
tjdgml0401@g.hongik.ac.kr, younkyul@hongik.ac.kr

Abstract—With the development of various adversarial attack mechanisms targeting deep learning-based biometric systems, these systems face a threat from adversarial attacks. To analyze the risk of adversarial attacks on biometric authentication systems, we defined the adversarial attack surfaces and adversarial attack scenarios of biometric systems.

Keywords—adversarial attack, biometric authentication, deep learning, adversarial attack surface, attack scenarios

I. INTRODUCTION

Recently, adversarial attacks have been discovered that cause misclassification for deep learning (DL) models by adding imperceptible perturbation to images, which threaten the security of DL-based biometric systems [1]. To analyze the risk of adversarial attack on biometric systems, we defined adversarial attack surfaces where attacks could potentially occur and designed possible adversarial attack scenarios for each attack surface.

II. OUR APPROACH

In general, biometric systems receive the user's biometric trait through a sensor and identify if the user is authenticated through two stages: liveness detection and identity matcher. The liveness detection detects the falsification of the submitted biometric trait, and the identity matcher identifies whether the submitted trait matches the system's authenticated one [2]. We defined the adversarial attack surfaces of the biometric process as follows: (1) Sensor-Liveness detector: the process of submitted biometric trait from the sensor to the liveness detector; (2) Liveness detector: it calculates the liveness score of the biometric trait; (3) Liveness comparator: it compares the calculated score with the liveness threshold; (4) Sensor-Identity matcher: the process of submitted biometric trait from the sensor to the identity matcher; (5) Identity matcher: it calculates the identity score of the biometric trait; and (6) Identity comparator: it compares the calculated score with the threshold.

Adversarial attacks are categorized based on their goals, types, and approaches [3]. The goals of these attacks are *integrity*, *availability*, and *privacy* violation, and the types of attacks are *evasion*, *poisoning*, and *exploratory* attacks. The approaches of attacks are the *training* phase that transforms the training dataset or algorithm, and the *testing* phase that transforms the submitted data to lead to misclassification. The strategies during the *training* phase include *data injection* (*DI*), *data modification* (*DM*), and *logic corruption* (*LC*), while those during the *testing* phase include *white-box* (*WB*) attacks and *black-box* (*BB*) attacks. Therefore, we defined a total of 21 adversarial attack scenarios for each attack surface: (1) Sensor-Liveness detector: integrity/evasion/testing-WB, integrity/evasion/testing-BB; (2) Liveness detector: availability/poisoning/training-DI, availability/poisoning/training-DM, availability/poisoning/training-LC, privacy/exploratory/testing-WB, privacy/exploratory/testing-BB; (3) Liveness comparator: availability/poisoning/training-LC; (4) Sensor-Identity matcher: integrity/evasion/testing-WB, integrity/evasion/testing-BB, availability/evasion/testing-WB, availability/evasion/testing-BB; (5) Identity matcher: availability/poisoning/training-DI, availability/poisoning/training-DM, availability/poisoning/training-LC, integrity/poisoning/training-DI, integrity/poisoning/training-DM, integrity/poisoning/training-LC, privacy/exploratory/testing-WB, privacy/exploratory/testing-BB, and (6) Identity comparator: availability/poisoning/training-LC.

III. CONCLUSION

We defined the adversarial attack surfaces and scenarios of biometric systems to analyze the risk of such attacks in these systems. Our future work includes the implementation of all defined scenarios applying state-of-the-art attack methods and analysis of adversarial attack risk on these systems using evaluation metrics from various perspectives.

ACKNOWLEDGMENT

This work was supported partly by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2022-00165648) and partly by Korea Foundation for Women In Science, Engineering and Technology (WISET) grant funded by the Ministry of Science and ICT(MSIT) under the team research program for female engineering students (No. WISET-2023-187)

[1] Fei, Jianwei, et al. "Adversarial attacks on fingerprint liveness detection." EURASIP Journal on Image and Video Processing 2020 (2020): 1-11.

[2] Biggio, Battista, et al. "Adversarial biometric recognition: A review on biometric system security from the adversarial machine-learning perspective." IEEE Signal Processing Magazine 32.5 (2015): 31-41.

[3] Chakraborty, Anirban, et al. "A survey on adversarial attacks and defences." CAAI Transactions on Intelligence Technology 6.1 (2021): 25-45.