УДК 004.056.523

СИСТЕМА ВИЯВЛЕННЯ СПУФІНГ-АТАК НА ГОЛОСОВІ СКУД

Жерновніков О.О.
Науковий керівник – к.т.н., доц. Горелов Д.Ю.
Харківський національний університет радіоелектроніки,
студентський науковий гурток «Біометричні технології контролю доступу»
каф. КРіСТЗІ, м. Харків, Україна
тел. +38(057) 702-14-30, e-mail: oleksandr.zhernovnikov@nure.ua

The structure of the spoofing attack detection system is proposed, which consists of three main components: the training module for training bases, the training module, and the exploitation module. A detailed description of the components of each module, as well as the functions they perform, is given.

Алгоритми підтвердження особистості людини за голосом добре вивчені, зручні у використанні та застосуванні як для безперервної, так і разової аутентифікації. Однак через поширення недорогих пристроїв запису і відтворення звуку вони схильні до спуфінгу, тобто, уразливі до дій зловмисників, спрямованих на видачу себе за іншу людину. У зв'язку з цим розробка та вивчення способів протидії спуфінгу є основним напрямком розвитку систем голосової аутентифікації.

Запропонована система виявлення спуфінг-атак на голосові СКУД складається з трьох модулів (рис. 1):

1. Модуль підготовки баз;
2. Модуль навчання системи детектування спуфінг-атак;
3. Модуль експлуатації системи детектування спуфінг-атак.

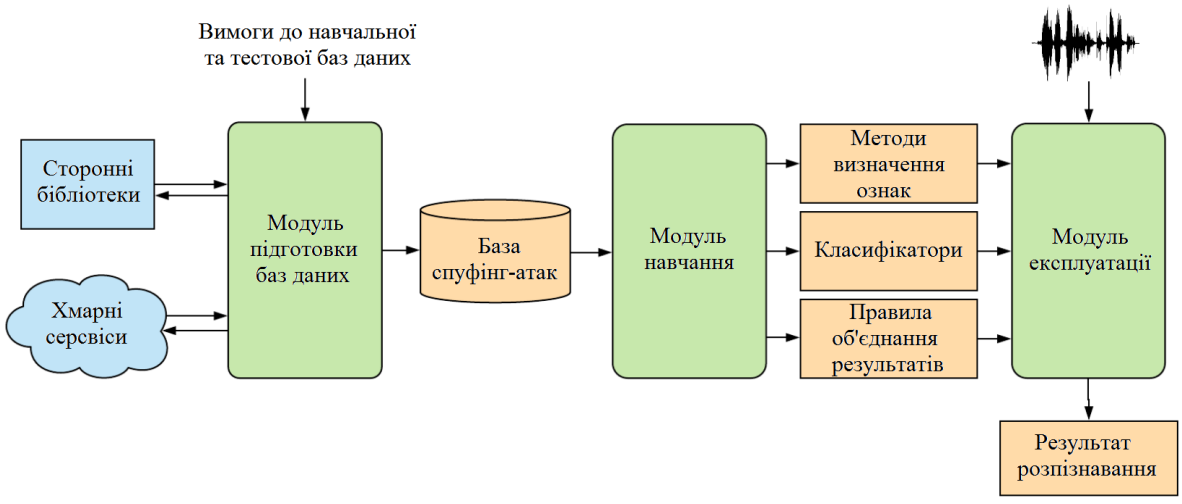

Рисунок 1

Модуль підготовки баз відповідає за автоматизацію процесів, необхідних для підготовки навчальної та тестової баз. Основною функцією модуля є автоматична генерація спуфінг-атак на основі вимог, які одержує модуль. З цією метою модуль використовує зовнішні технології генеру-

вання синтезу. До них відносяться хмарні сервіси синтезу мови Google та IBM. Для генерації синтезованих записів модуль використовує готовий набір текстів, запозичений з кількох творів художньої літератури українською та англійською мовами. Для генерації атак повторного відтворення використовує готовий набір фонограм. Модуль має можливість розширення за рахунок використання нових технологій, мов, а також використання інших текстових корпусів.

Модуль навчання (рис. 2) відповідає за навчання всіх моделей виділення високорівневих ознак та класифікаторів та складається з 5 підмодулів.

Підмодуль обробки вхідних даних здійснює отримання ознак нижнього рівня, наприклад спектрограм, та реалізує алгоритми підготовки вхідних даних для навчання нейронних мереж.

У підмодулі LPCC / PLP відбувається розрахунок кепстральних коефіцієнтів лінійного передбачення (LPCC) та коефіцієнтів перцептивного лінійного передбачення (PLP) для аналізу часових ознак аудіосигналів.

Підмодуль навчання нейронних мереж навчає модулі отримання високорівневих ознак зі спектрального подання акустичних сигналів. Крім того, він навчає мережу з архітектурою CNN+RNN для класифікації спектрограм безпосередньо.

Підмодуль навчання класифікаторів для класифікації високорівневих ознак, отриманих на основі використання нейронних мереж.

Оскільки в системі використовуються декілька незалежних методів розпізнавання спуфінг-атак, результати їх роботи треба об’єднати. Одним з варіантів вирішення цієї задачі є застосування логістичної регресії, яку, застосовують у випадку, коли результат класифікації є бінарним – спуфінг чи природна мова. За цей функціонал відповідає підмодуль навчання, який приймає на вхід результати класифікації кожної індивідуальної системи для тестової множини та навчає на їх основі параметрів загальної системи.

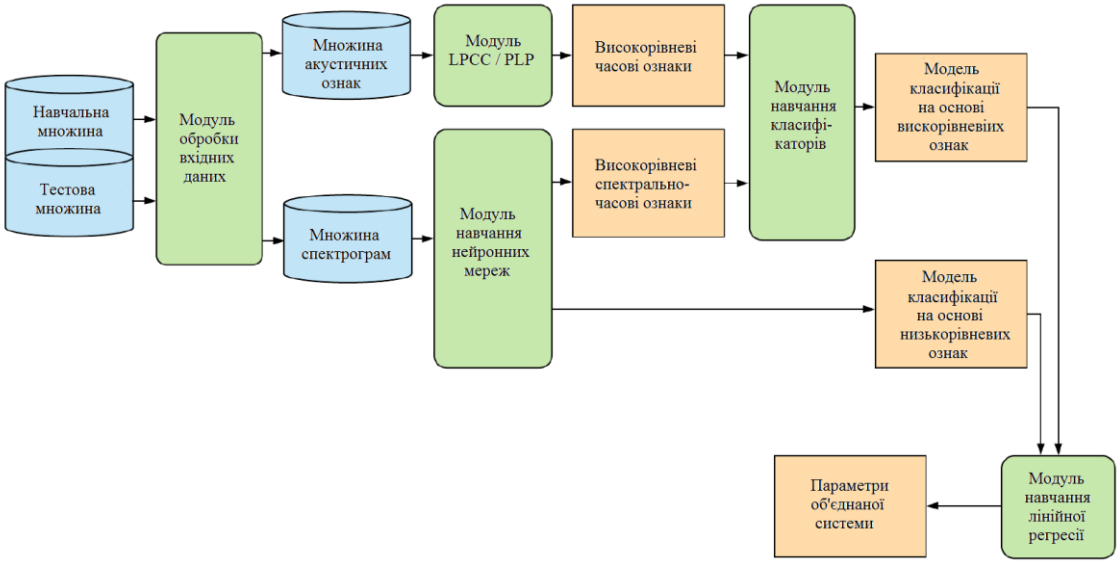

Рисунок 2

Модуль експлуатації (рис. 3) безпосередньо виконує функцію детектування спуфінг-атак. Спершу виконується обчислення ознак нижнього рівня. Далі виконується обчислення ознак високого рівня. На даному етапі ознаки нижнього рівня обробляються двома модулями: модулем вилучення LPCC/PLP та модулем глибокого навчання. Цей етап не виконується в системі на основі CNN+RNN, оскільки ця нейронна мережа виступає одразу в ролі класифікатора. Далі відбувається класифікація за кожним методом. Всі результати класифікації, отримані від індивідуальних систем,надходять на вхід до модуля прийняття рішення, який, використовуючи вагові коефіцієнти, формує фінальну ймовірність приналежності вхідної фонограми до класу спуфінг-атак.

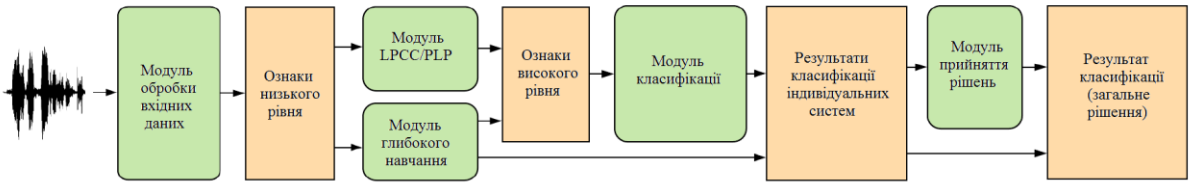

Рисунок 3

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ:

1. Wu Z., Das R. K., Yang J., Li H. Light convolutional neural network with feature genuinization for detection of synthetic speech attacks. In: Proceedings of the 21st Annual Conference of the International Speech Communication Association. 2020, 1101–1105.

2. Chen, Z.; Zhang, W.; Xie, Z.; Xu, X.; Chen, D. Recurrent neural networks for automatic replay spoofing attack detection. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Calgary, AB, Canada, 15–20 April 2018; pp. 2052-2056.

3. Dehak, N. Discriminative and Generative Approaches for Long- and Short- term Speaker Characteristics Modeling: Application to Speaker Verification – 2009.

4. Xiao, X. Spoofi speech detection using high dimensional magnitude and phase features: the NTU approach for ASVspoof 2019 challenge / X. Xiao, X. Tian, S. Du, H. Xu, C. E. Siong, H. Li // INTERSPEECH. – 2015.

5. Todisco, M. A New Feature for Automatic Speaker Verification Anti-Spoofing : Constant Q Cepstral Coefficients / M. Todisco, H. Delgado, N. Evans //. – 2016.

6. Nagarsheth, P. Replay Attack Detection Using DNN for Channel Discrimination / P. Nagarsheth, E. Khoury, K. Patil, M. Garland // INTERSPEECH. – 2017.