РЕЦЕНЗІЯ

доктора технічних наук, професора,
професора кафедри захисту інформації
Національного університету «Львівська політехніка»
Хоми Володимира Васильовича
на дисертацію
Рудої Христини Степанівни
«УДОСКОНАЛЕННЯ МЕТОДІВ ТА ЗАСОБІВ БІОМЕТРИЧНОЇ АВТЕНТИФІКАЦІЇ ЗА ГОЛОСОМ З ЗАСТОСУВАННЯМ ТЕХНОЛОГІЙ МАШИННОГО НАВЧАННЯ»
подану до захисту на здобуття наукового ступеня доктора філософії за спеціальністю 125 «Кібербезпека» (галузь знань 12 «Інформаційні технології»)

Актуальність теми дисертації. В епоху глобальної цифровізації, стрімкого поширення інфотелекомунікаційних технологій та постійного зростання кіберзагроз питання безпечної та надійної автентифікації набуває особливого значення. Одним із найперспективніших напрямів у цій сфері є біометрична автентифікація користувачів за голосом. Цей підхід поєднує у собі природність і зручність, доступність і низькі затрати на реалізацію голосового інтерфейсу, а також можливість простого інтегрування із багатьма існуючими цифровими технологіями.

Проте на цей час голосова автентифікація зіткнулася із серйозними викликами передовсім через загрози до атак синтезованим голосом за допомогою інструментів штучного інтелекту. Крім того, ефективність відомих підходів голосової автентифікції істотно знижується у реальних умовах застосування, для яких є характерними вплив шуму середовища, відмінності у характеристиках застосованих технічних засобів, мовна і міжсесійна варіативність, зростання числа користувачів. Це вимагає переходу до сучасних рішень на основі машинного навчання та штучного інтелекту, які здатні забезпечити достовірність автентифікації і маштабованість рішень у багатокористувацьких сценаріях, стійкість до атак та інваріантність до сторонніх впливів.

Таким чином, тема цього дисертаційного дослідження є актуальною оскільки спрямована на вирішення низки ключових науково-прикладних завдань сучасної кібербезпеки та добре узгоджується із сучасними світовими трендами. Вирішення цих завдань надасть нового поштовху у розвитку прикладних сервісів, таких як голосові асистенти, мобільні додатки, системи інтелектуального будинку, технології автентифікації у банківській справі, медичних застосунках і у військовій сфері.

Зв'язок теми дисертації з державними програмами, науковими напрямами університету та кафедри. Дисертаційне дослідження виконано в межах держбюджетної науково-дослідної роботи: «Дослідження стійкості біометричних систем автентифікації до атак із застосуванням технології клонування голосу на основі глибинних нейронних мереж» (№ держреєстрації 0124U000407). Основні положення та результати дисертаційної роботи впроваджено у навчальний процес кафедри Захист інформації Національного університету "Львівська політехніка".

Наукова новизна основних результатів дисертації полягає у такому:

1) Одержала подальший розвиток концепція удосконалення голосової автентифікації, що базується на широкому і комплексному застосуванні технологій машинного навчання та штучного інтелекту на різних структурних рівнях з урахуванням міжнародних стандартів у сфері біометричної безпеки. Це формує методологічну основу до підвищення стійкості системи до зростаючих кіберзагроз, а також покращення її дискримінаційної здатності в багатокористувацьких сценаріях та в умовах реального застосування.

2) Вдосконалено такі ключові структурні комоненти системи біометричної автентифікації як модулі порівняння, налаштування та реєстрації голосових зразків шляхом застосування нейромережевих і трансформерних голосових моделей і обґрунтованого вибору метричних функцій. Це збільшує інваріантність голосової автентифікації до міжсесійних варіацій та масштабування, а відтак підвищує достовірність верифікаційних рішень біометричної системи в умовах багатокористувацьких сценаріїв.

3) Вперше розроблено та застосовано методику оцінювання стійкості до атак із застосуванням синтетично згенерованого мовлення, зокрема на основі генеративних моделей (RVC, ElevenLabs, XTTS, Tortoise), що дозволяє верифікувати ефективність системи згідно з вимогами ISO/IEC 30107-3:2023 щодо тестування Presentation Attack Detection. Виявлено, що всі протестовані моделі демонструють вразливість до атак з застосуванням клонування голосу, що стало підставою для рекомендації впровадження детекторів синтетичного мовлення як обов'язкового компонента систем біометричної голосової автентифікації.

4) Вперше реалізовано та зіставлено підходи до виявлення синтетичного мовлення на основі генеративної моделі ElevenLabs і авторської моделі MFCC+SVM. Отримані результати покладено в основу рекомендацій щодо впровадження універсальних антиспуфінгових механізмів у архітектуру голосових біометричних систем.

Ступінь обґрунтованості наукових положень дисертації і їх достовірність та новизна ґрунтується на коректному підході до

формулювання дослідницьких завдань, логічному обґрунтуванні вибору перспективних рішень, правильному вибору методології досліджень і прийнятих припущень, використанні новітних методів моделювання та аналітичного аналізу результатів експерименту. Наукові положення дисертації є добре обґрунтованими та підтверджуються результатами експериментів, проведених автором. Використання сучасних методів машинного навчання та технологій штучного інтелекту для удосконалення методів голосової автентифікації є інноваційним підходом, що підтверджує новизну даного дисертаційного дослідження.

Наукове значення виконаного дослідження із зазначенням можливих наукових галузей та розділів програм навчальних курсів, де можуть бути застосовані отримані результати. Наукові результати дисертаційної роботи, що включають в себе розробку та удосконалення низки структурних модулів системи голосової автентифікації, а також застосування моделей машинного навчання, можуть бути використані для створення ефективних та надійних систем контролю доступу. Ці результати можуть бути застосовані у різних наукових галузях, пов'язаних з інформаційними технологіями та кібербезпекою. Розроблені методи та моделі можуть бути інтегровані у навчальні курси з кібербезпеки, машинного навчання та аналізу даних, надаючи студентам практичні навички та знання у сфері виявлення та протидії кіберзагрозам. Це особливо актуально для студентів спеціальностей, пов'язаних з інформаційними технологіями, комп'ютерними науками та інженерією.

Крім того, результати дослідження можуть бути використані як фрейворк для створення комерційних продуктів та послуг у різних сферах, де вимагається розпізнавання за голосом.

Практичне значення одержаних результатів полягає у тому, що:

1. Розроблено та реалізовано архітектуру системи голосової автентифікації на основі вбудованих векторів, яка забезпечує відповідність принципам надійності, масштабованості та обмеження доступу до біометричних шаблонів, визначеним у ISO/IEC 24745:2022 та ISO/IEC 19795-1:2021. Впровадження архітектури в експериментальному середовищі дозволило досягти точності автентифікації до 96%.
2. Імплементовано модель виявлення синтетичного мовлення на основі MFCC+SVM, яка забезпечує точність виявлення атак до 89.75%, при цьому точність класифікації справжніх голосів залишається низькою (25.96%), що свідчить про потребу додаткових рішень. Модель відповідає вимогам ISO/IEC 30107-1:2023 до PAD-механізмів і рекомендована до застосування як додатковий рівень захисту в багатофакторних біометричних системах, що працюють із критичними даними.

3. Створено тестовий корпус голосових даних, за допомогою якого проведено порівняльне оцінювання ефективності мовних моделей у сценаріях підробки голосу з різним лінгвістичним наповненням.

4. Результати порівняльного тестування моделей TitaNet, ECAPA-TDNN, WavLM та PyAnnote можуть слугувати основою для обґрунтованого вибору архітектури в задачах біометричної автентифікації, орієнтованих на великомасштабні користувацькі бази. Зокрема, модель TitaNet продемонструвала найвищу стійкість до масштабування: збільшення числа зареєстрованих користувачів від 10 до 70 зумовило зниження точності в середньому на 3,55%. Оцінювання масштабованості здійснювалася згідно методологічних положень стандарту ISO/IEC 19795-2:2007, який регламентує сценарне тестування з варіативною чисельністю суб'єктів як одного з ключових параметрів експериментального дизайну.

5. Результати оцінювання ресурсної ефективності моделей TitaNet, ECAPA-TDNN та WavLM вказали на їхню придатність для розгортання на обчислювальних платформах з обмеженими ресурсами. Зокрема, модель WavLM-base-plus-sv забезпечує автентифікацію з точністю 93% витрачаючи на зіставлення голосових векторів в середньому 4.34 мс. Це означає, що модель, будучи заімплементованою на периферійних пристроях із обмеженими резурсами задовольняє вимоги до швидкодії необхіді для роботи в реальному часі. Отримані результати відповідають критеріям ефективності, визначеним у ISO/IEC 29167-1:2014, та засвідчують можливість застосування таких моделей у вбудованих біометричних рішеннях, мобільних пристроях і системах контролю доступу.

Розроблені підходи до побудови архітектури систем використано і впроваджено у професійну діяльність компанії UNIDATALAB LTD у контексті виконання прикладних завдань у рамках комерційних проєктів, що підтверджено актами впровадження.

Повнота оприлюднення результатів дисертаційної роботи. Основні результати дисертаційної роботи Рудої Христини викладено у чотирнадцяти наукових публікаціях, а саме: у восьми статтях (із них шість – у фахових наукових виданнях України та дві – у закордонних періодичних виданнях), розділі монографії і п'яти матеріалах наукових конференцій і семінарах. Особистий внесок здобувача у колективно опублікованих працях полягав у формуванні підходів, реалізації експериментів та інтерпретації їх результатів.

Зауваження до дисертаційної роботи.

1. Дисертаційна робота є інноваційною оскільки ґрунтується на найсучасніших досягненнях у сфері машинного навчання та штучного інтелекту. Разом з тим у першому розділі бажано було б подати більш

ґрунтовний аналіз сучасного стану розвитку систем голосової автентифікації. Представлений огляд відомих рішень у цій царині є надто загальним і не містить кількісних оцінок хоча б базових характеристик, а це ускладнює об'єктивне порівняння ефективності здобутих результатів.

2. До порівняльного аналізу метрик обчислення відстані векторними представленнями голосових ознак було розглянуто чотири типові метрики: мангетенську (L1-норму), евклідову (L2-норму), Махаланобісову та косинусну відстані. За результатами експериментального оцінювання обґрунтовано зроблено вибір косинусної відстані як метрики, що найкраще відображає кутову схожість між векторами і зберігає інваріантність до масштабування векторів.

Водночас для підвищення дискримінаційної здатності моделей у системах голосової автентифікації доцільно застосувати не лише метрики подібності, а й спеціалізовані функції втрат, які не лише вимірюють схожість між векторами, але й активно формують структуру векторного простору під час навчання. Наприклад, функція Triplet Loss використовує трійки зразків, цілеспрямовано змінюючи векторний простір з метою зменшити відстань між анкерним (якірним) і позитивним зразком, що належить одній особі, та, навпаки, збільшити відстань між анкерним і негативним зразком, що належить іншій особі. Це забезпечить моделі краще розрізняти користувачів у військовимірному просторі ознак, що є критично важливим для надійної голосової верифікації.

3. Аудіодані, використані в експериментах (табл. 3.2), є відносно одноманітними, оскільки переважно походять від дикторів аудіокниг. Така специфіка джерела обмежує варіативність мовних і акустичних характеристик, що може впливати на узагальнювальну здатність моделі. З метою підвищення репрезентативності експериментальної бази доцільно було б розширити та урізноманітнити набір аудіозаписів, залучаючи дані з альтернативних джерел зокрема з месенджерів соціальних мереж. Це забезпечило б об'єктивнішу оцінку масштабованості запропонованого підходу, його стійкості до акустичних варіацій та ефективність у реальних сценаріях застосування.

4. Одним із важливих результатів досліджень є висновок про потребу калібрування порогових значень одержаних емпірично для кожної моделі. На жаль, у роботі не розкрито як саме обчислювалися ці значення.

5. Чому для оцінювання масштабованості моделей автентифікації за голосом застосовано лише одну метрику – точність, яка, як відомо, має обмежені можливості щодо комплексної оцінки продуктивності автентифікації (табл. 3.4), хоча у попередніх експериментах застосовано ще чотири метрики - FAR, FRR, EER і DFC.

6. Решта зауважень стосуються стилю і оформлення роботи. Передовсім формулювання таких важливих атрибутів як мета, об'єкт і предмет дослідження можуть бути підсилені. Так само можна б покращити виклад деяких пунктів наукової новизни. Висновки доцільно доповнити числовими даними, які є, наприклад, у рубриці Практичне значення одержаних результатів.

Наведені зауваження не знижують загальної позитивної оцінки дисертаційної роботи.

Висновок

На основі критичного аналізу дисертації та праць здобувача, які опубліковані за темою дисертації, встановлено, що дисертаційна робота Рудої Христини Степанівни «Удосконалення методів та засобів біометричної автентифікації за голосом з застосуванням технологій машинного навчання», є завершеним та цілісним науковим дослідженням, в якому одержано вагомі наукові результати, а також має практичну цінність у сфері кібербезпеки. Використання чужих наукових результатів без посилань на авторів у дисертації не виявлено, що свідчить про особистий внесок здобувача в науку. Дисертаційне дослідження відповідає спеціальності 125 «Кібербезпека» та вимогам наказу Міністерства освіти і науки України «Про затвердження вимог до оформлення дисертації», постанові Кабінету Міністрів України «Порядок присудження ступеня доктора філософії...», а тому її автор, Руда Христина Степанівна заслуговує присудження наукового ступеня доктора філософії за спеціальністю 125 – "Кібербезпека".

Офіційний рецензент
доктор технічних наук, професор,
професор кафедри захисту інформації
Національного університету
"Львівська політехніка" Володимир ХОМА

Підпис д.т.н, професора Хоми В.В.
засвідчую:

Проректор Микола ЛОГОЙДА