Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Факультет комп’ютерної інженерії та управління
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Метод перевірки цілісності аудіоданих
на основі цифрових водяних знаків

(тема)

Виконав:
студент II курсу, групи КСМм-22-2
Бондарчук П.С.
(прізвище, ініціали)

Спеціальність
123 «Комп’ютерна інженерія»
(код і повна назва спеціальності)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма
Комп’ютерні системи та мережі
(повна назва освітньої програми)

Керівник: доц. Мартовицький В.О.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ЕОМ _____ Коваленко А.А.
(підпис) (прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет комп'ютерної інженерії та управління

Кафедра електронних обчислювальних машин

Рівень вищої освіти другий (магістерський)

Спеціальність 123 «Комп'ютерна інженерія»
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Комп'ютерні системи та мережі
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
"____"_____ 20___ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Бондарчуку Петру Сергійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Метод перевірки цілісності аудіоданих на основі цифрових водяних знаків

затверджена наказом по університету від "06" листопада 2023 р. № 1298Ст

2. Термін подання студентом роботи до екзаменаційної комісії 15 січня 2024 р.

3. Вхідні дані до роботи: набір аудіофайлів.

4. Перелік питань, що потрібно опрацювати у роботі

1) розглянути методи перевірки цілісності;
2) розглянути методи стеганографії;
3) реалізувати гібридний метод перевірки цілісності аудіофайлів;
4) висновки.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів)

Слайд презентація – 15 слайдів.

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Огляд автентифікації аудіоданих	07.11.23 – 05.11.23	
2	Огляд механізмів захисту цілісності та	06.11.23 – 12.11.23	
	доброчесності аудіоданих		
3	Огляд перцептивного аудіохешування	13.11.23 – 19.11.23	
4	Розробка комбінованого підходу до	20.11.23 – 03.12.23	
	перцептивного хешування/водяних знаків		
5	Проведення експериментів	04.12.23 – 10.12.23	
6	Оформлення матеріалів кваліфікаційної роботи	11.12.23 – 29.12.23	
7	Подання кваліфікаційної роботи керівникові та	30.12.23 – 04.01.24	
	її попередній захист		
8	Подання кваліфікаційної роботи на рецензу-	05.01.24 – 10.01.24	
	вання		

Дата видачі завдання 06 листопада 2023 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Мартовицький В.О.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 57 с., 11 рис.,1 дод., 23 джерела.

RMAC, CCTV, AUDIO CD, ХЕШ, АВТЕНТИФІКАЦІЯ, ВОДЯНІ ЗНАКИ, ЦИФРОВІ ДАННІ, КОРИСНЕ НАВАНТАЖЕННЯ.

Метою кваліфікаційної роботи є дослідження та розробка методу перевірки цілісності аудіоданих на основі цифрових водяних знаків.

Об'єктом дослідження є процес автентифікаці аудіофайлів.

Предметом дослідження полягає в методи перевірки цілісності аудіоданих.

У ході виконання кваліфікаційної роботи було запропоновано механізм перевірки цілісності та автентичності цифрових звукозаписів. Він розроблений і реалізований таким чином, щоб бути нечутливим до поширених операцій постобробки аудіоданих, які лише незначною мірою впливають на суб'єктивне акустичне сприйняття. Прикладами таких операцій є стиснення з втратами, що зберігає високу якість звуку аудіоносія, або перетворення форматів без втрат. Мета полягає в тому, щоб уникнути фактичних хибних спрацьовувань, які очікувано спостерігаються в стандартних протоколах криптографічної автентифікації за наявності такої легітимної пост-обробки. Для досягнення цієї мети досліджується можлива комбінація методів цифрового водяного маркування та аудіоспецифічного хешування.

Розробляється відповідний алгоритм хешування аудіо, що як код автентифікації повідомлення "rMAC", дозволяє здійснювати перевірку цілісності "на основі сприйняття".

ABSTRACT

Master’s thesis: 57 pages, 11 figures, 1 appendices, 23sources.

RMAC, CCTV, AUDIO CD, HASH, AUTHENTICATION, WATERMARK, DIGITAL DATA, PAYLOAD.

The major goal of this thesis is qualification work is to research and develop a method for verifying the integrity of audio data based on digital watermarks.

The object of study is the process of authenticating audio files.

The subject of the study is the method of verifying the integrity of audio data.

In order to the course of the qualification work, a mechanism for verifying the integrity and authenticity of digital audio recordings was proposed. It is designed and implemented in such a way as to be insensitive to common audio data post-processing operations that only slightly affect subjective acoustic perception. Examples of such operations are lossy compression, which preserves the high sound quality of the audio carrier, or lossless format conversion. The goal is to avoid the actual false positives that are expected to occur in standard cryptographic authentication protocols in the presence of such legitimate post-processing. To achieve this goal, we investigate a possible combination of digital watermarking and audio-specific hashing methods.

An appropriate audio hashing algorithm is developed, which, as a message authentication code "rMAC", allows for "perception-based" integrity verification.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ 7
ВСТУП 8
1 АВТЕНТИФІКАЦІЯ АУДІОДАНИХ 10
1.1 Достовірність цифрових аудіоданих 10
1.2 Проблеми та задачі автентифікації аудіоданих 14
2 ІСНУЮЧІ МЕХАНІЗМИ ЗАХИСТУ ЦІЛІСНОСТІ ЦИФРОВИХ ДАНИХ 21
2.1 Огляд механізмів захисту цілісності та доброчесності 21
2.2 Перцептивний аудіохешування/аудіофінгерпринтинг 26
3 ЗАПРОПОНОВАНА СИСТЕМА ВОДЯНИХ ЗНАКІВ ЦІЛІСНОСТІ НА ОСНОВІ ВМІСТУ 29
3.1 Опис підходу до захисту водяними знаками, вразливими до змін вмісту 29
4 КОМБІНОВАНИЙ ПІДХІД ДО ПЕРЦЕПТИВНОГО ХЕШУВАННЯ/ВОДЯНИХ ЗНАКІВ 40
4.1 Адаптація хешування та вбудовування водяних знаків 40
4.2 Корисне навантаження вбудованого повідомлення 41
ВИСНОВКИ 45
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ 46
ДОДАТОК А Графічний матеріал кваліфікаційної роботи 49

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

CCTV – відеоспостереження (англ., Closed Circuit Television)

BWF – це розширення імені аудіофайлів, збережених у цьому форматі (англ., Broadcast Wave Format)

rMAC – код автентифікації повідомлення (англ., Message Authentication Code)

RIPEMD – алгоритм безпечного (англ., RACE Integrity Primitives Evaluation Message Digest)

HMAC – ключовий код автентифікації повідомлень (англ., Hash-based Message Authentication Code)

ВСТУП

Сьогодні комп'ютерне обладнання та програмне забезпечення надають багато можливостей для створення, запису, редагування, розповсюдження та архівування цифрового мультимедійного контенту. Особливо сучасне програмне забезпечення для редагування цифрових зображень, відео- та аудіоданих дозволяє дуже ефективно та оперативно вносити зміни до цього контенту. Як наслідок, такий контент може бути ненавмисно змінений або навіть підроблений легко і з незначними слідами, якщо взагалі залишиться. Прикладами значного контенту, вразливого до підробки, є петабайти даних з історичних архівів, що зберігають культурну спадщину цивілізації, або контент, створений користувачами в Інтернеті. Сьогодні маніпуляції з контентом у цифрових форматах зберігання можуть бути легше застосовані до мультимедійного контенту, ніж це було в аналогову епоху.

У суспільстві проблема підробки цифрового мультимедійного контенту дуже часто обговорюється насамперед на прикладі фотографій. Загальновідомо, що фотографії підробляли як акт цензури чи пропаганди ще з часів появи аналогової фотографії у 19 столітті. Іноді фотографії для преси редагують для того, щоб прикрасити картинку або драматизувати її. Крім того, нинішня дискусія про так звані "фейкові новини" та ознаки систематичної дезінформації з боку різних суб'єктів нагадують нам про достовірність медіа-даних, що циркулюють у пресі. У певних сферах застосування цифрові аудіозаписи містять чутливий вміст. Прикладами є історичні архівні матеріали в державних архівах, які зберігають нашу культурну спадщину, або цифрові докази в контексті правоохоронної діяльності та цивільного судочинства. Через потужні можливості сучасних інструментів для редагування мультимедіа такі матеріали є вразливими до підробки змісту та фальсифікації їхнього походження зі зловмисними намірами. Також ненавмисна модифікація даних і помилкове походження

можуть бути спричинені людською помилкою. Таким чином, достовірність і походження такого аудіоконтенту, а також впевненість у його походження є критично важливими факторами, що стосуються непідробного і справжнього стану такого аудіоконтенту.

Для вирішення цієї проблеми в роботі запропоновано механізм перевірки цілісності та автентичності цифрових звукозаписів. Він розроблений і реалізований таким чином, щоб бути нечутливим до поширених операцій постобробки аудіоданих, які лише незначною мірою впливають на суб'єктивне акустичне сприйняття (якщо взагалі впливають). Прикладами таких операцій є стиснення з втратами, що зберігає високу якість звуку аудіоносія, або перетворення форматів без втрат. Мета полягає в тому, щоб уникнути фактичних хибних спрацьовувань, які очікувано спостерігаються в стандартних протоколах криптографічної автентифікації за наявності такої легітимної пост-обробки. Для досягнення цієї мети досліджується можлива комбінація методів цифрового водяного маркування та аудіоспецифічного хешування.

Спочатку розробляється відповідний алгоритм хешування аудіо, що залежить від секретного ключа. Він включає та вдосконалює так звану технологію аудіовідбитків, яка є найсучаснішою в галузі аудіоідентифікації на основі контенту. Представлений алгоритм (позначений як код автентифікації повідомлення "rMAC") дозволяє здійснювати перевірку цілісності "на основі сприйняття". Це означає, що порушення цілісності класифікується як таке до того, як воно стає чутним.

Крім того, цей rMAC вбудовується і зберігається безшумно всередині аудіоносія за допомогою технології нанесення водяних знаків на аудіо. Такий підхід дозволяє зберегти код автентифікації під час вищезгаданих допустимих операцій постобробки і зробити його доступним для перевірки цілісності пізніше. Для цього в роботі роботі використано та вдосконалено існуючий алгоритм залежного від секретного ключа аудіо водяного маркування.

1 АВТЕНТИФІКАЦІЯ АУДІОДАНИХ

1.1 Достовірність цифрових аудіоданих

Як і фотографія, історія звукозапису сягає 19 століття, а саме винаходу фонографа де Мартенвілем у 1860 році або фонографа Едісоном у 1877 році. Відтоді було створено незліченну кількість аналогових і цифрових аудіоносіїв. І в певних сценаріях відповідний зміст насправді міститься у фрагментах аудіоінформації, а не в зображеннях (у русі). Очевидно, що це стосується чистого аудіоконтенту, наприклад, записів голосу / передач або радіопередач. Але це також стосується звукових доріжок, що містяться у різного роду "відео" даних.

Вставка, склеювання, видалення, приглушення або обрізка аудіоматеріалів може бути легко застосована і може означати значну зміну того, що чує і, зрештою, розуміє користувач-людина. Наприклад, у голосовому записі навіть додана, змінена або видалена фраза короткої тривалості (наприклад, слова "так" або "ні") або префіксальний склад (наприклад, слова "згоден" або "не згоден") може повністю змінити його зміст. Навіть гірше, повні аудіозаписи можуть бути створені з нуля, що може приховати підроблене походження певного аудіодоказу. Як і у випадку з цифровими зображеннями, таку підробку можна легко застосувати і приховати від пересічного слухача за допомогою програмного забезпечення для редагування аудіо та інших заходів: Відповідно, чути – не означає вірити!

Таким чином, достовірність і походження цих медіа з точки зору перевірки їхнього справжнього і неспотвореного стану є життєво важливими. Прикладами такого важливого контенту є історичні архівні матеріали, що зберігають культурну спадщину, та цифрові докази, про які йтиметься далі.

Чутливий аудіоконтент, який варто захищати, міститься в історичних архівних записах, що зберігають культурну спадщину. Нижче наведені найбільш релевантні класи і кілька мотивуючих прикладів такого контенту:

- історичні промови: Приклади відомих цитат з різних століть, які сприяють формуванню культурної пам'яті, є частиною загального знання. Записи таких подій доступні з кінця 19-го століття. Забезпечення їхньої доступності є природним мандатом державних архівів та подібних культурних інституцій;

- усноісторичні інтерв'ю: Крім того, державні архіви надають мільйони свідчень сучасних свідків з різних епох і культурних кіл. На відміну від попереднього прикладу з відомими публічними виступами, ці свідчення зберігають пам'ять і досвід незліченної кількості "звичайних" людей;

- новини: Новини про поточні події та справи в суспільстві також роблять свій внесок у сучасну історію. Те саме стосується зібраного новинного контенту, навіть якщо він не виходить в ефір, а залишається в інформаційних агентствах. Тут перевірка автентичності та цілісності джерел (запитів, репортерів, інформаторів, фотографів, знімальних груп тощо) завжди була головним обов'язком журналістів і письменників;

- користувацький контент: Будь-який вид створеного користувачами контенту, який циркулює в Інтернеті або в приватних колах, виражає інтереси та занепокоєння суспільства (і може бути, в свою чергу, знову опублікований в "Новинах");

- радіо/телепрограми: Навіть тривіальні повсякденні теле- і радіопрограми і навіть комерційна реклама в теле- і радіоефірі відображають культурне розмаїття сучасної цивілізації;

- музика: Серйозна музика та поп-музика є важливими елементами культурної спадщини також є важливими елементами культурної спадщини. Музика різних жанрів, створена/виконана різними виконавцями, композиторами чи диригентами. диригентами.

Рисунок 1.1 – Приклади культурної спадщини

Усі типи аудіоконтенту, згадані вище, доступні для громадськості у вигляді аналогових або цифрових/оцифрованих записів з архівів, музеїв чи інших культурних організацій, або є комерційними продуктами, або циркулюють в Інтернеті.

Інший важливий клас значущого аудіоконтенту – це докази, що містяться в цифровому аудіо. Приклади наведені нижче.

Допити в поліції та судові засідання: Свідчення свідка чи потерпілого або зізнання підозрюваного під час поліцейських допитів, судових засідань (у тому числі відеозаписів) часто записуються за допомогою електронного відеообладнання або простих диктофонів. У багатьох країнах (але не в Німеччині) це робиться в повному обсязі.

Законне перехоплення / електронне прослуховування: Підслуховування підозрюваних у скоєнні злочину здійснюється правоохоронними органами в ході кримінального розслідування. Тут аудіоспостереження за допомогою електронних жучків на місці злочину або прослуховування телефонного зв'язку чи комп'ютера підозрюваного в ході законного (а також стратегічного) перехоплення є поширеним заходом для збору доказів.

Військові бойові операції: Під час бойових дій бойові підрозділи та їхнє спорядження часто оснащені пристроями для передачі та запису відео в реальному часі. Голоси, шум навколишнього середовища (наприклад, постріли) або радіопереговори можуть бути використані для розслідування інцидентів або навіть воєнних злочинів.

Відеореєстратори: Бортові "відеореєстратори" в автомобілях приватних користувачів, схоже, стають дедалі популярнішими в деяких країнах. Знову ж таки, звукова доріжка в таких відеозаписах, створених користувачами, може бути використана для обґрунтування цивільних позовів у разі дорожньо-транспортних пригод.

Відеоспостереження/CCTV: Нарешті, все більшого поширення набувають системи відеоспостереження за громадськими місцями та приватними приміщеннями. За певних правових умов такі системи відеоспостереження (CCTV) також фіксують звук, який може бути використаний як доказ у кримінальних або цивільних справах.

Телефонний зв'язок: Спілкування бізнесу з клієнтами по телефону вже десятиліттями є звичним явищем у сфері роздрібної торгівлі або домашнього банкінгу. Іноді колл-центри роблять запис телефонної розмови, щоб надати докази в разі скарг клієнтів. Іншими прикладами таких важливих телефонних розмов є записи екстрених дзвінків до служб порятунку або поліції.

Цивільна авіація та судноплавство: Комерційні літаки та судна повинні бути обладнані записуючими пристроями (наприклад, диктофон в кабіні / бортовий самописець). У разі аварії або катастрофи, останні хвилини розмови екіпажу, оголошення для пасажирів та інші звуки/шуми повинні бути доступні для подальшого розслідування. Це також стосується радіопереговорів, записаних на місці розташування повітряних/морських диспетчерів.

Рисунок 1.2 – Приклади цифрових доказів

Наведені раніше приклади цифрових доказів об'єднує те, що життєво важливим є повний ланцюжок доказів. Як наслідок, запропоновані механізми захисту повинні бути інтегровані в (надійний) записуючий пристрій, наприклад, диктофон, камеру або в центр моніторингу.

Приклади аудіоносіїв, що відповідають культурній спадщині та цифровим доказам, візуалізовано на рисунках 1.1 та 1.2. У світлі цих прикладів поняття достовірності аудіоданих буде більш детально розглянуто далі.

1.2 Проблеми та задачі автентифікації аудіоданих

Перш ніж обговорювати і вибирати відповідні технічні підходи для більш детального вивчення, необхідно пояснити найважливіші проблеми досліджуваної системи захисту.

Попередній огляд демонструє, що аудіодані самі по собі або як звукова доріжка відеоконтенту можуть представляти чутливий контент. Загальна мета достовірності цієї роботи полягає в тому, щоб контент не був зловмисно підроблений і була забезпечена впевненість у його походженні. Це безпосередньо відповідає цілям безпеки цілісності та автентичності, які згадуються нижче.

Цілісність аудіоданих. Нагадаємо, що сам термін "Цілісність" походить від латинського виразу "integritas", що буквально означає "integritas (лат.): повнота, цілісність, "непошкоджений або непорушений стан речі" і в переносному значенні відповідає "правильності" або "недоторканності". Воно також відсилає до протилежного поняття "tangere (лат.): торкатися" [1]. Добре відомі стандартні механізми перевірки цілісності, такі як хеш-функції або контрольні суми, можуть дуже точно і чутливо перевіряти "правильність" і "повноту" вхідних даних. Але особливо для аудіоданих існує багато операцій постобробки, які, по суті, змінюють двійкове представлення, але залишають аудіоконтент у його "непошкодженому" стані з точки зору

середньостатистичного слухача. Наприклад, після перетворення формату без втрат або стиснення з незначними втратами під час постобробки аудіоданих, вони все ще забезпечують "звучання" [sic!] з точки зору того, що чує або розуміє людина-користувач.

Автентичність аудіоданих. Щодо поняття "автентичність" слід насамперед зазначити, що цей термін походить від виразу "autenticus (лат.): оригінальний (документ), справжній ... "який походить від автора" з давніх часів [2]. У контексті аудіоданих зазначена автентичність має бути обов'язковою для достовірної ідентифікації фізичної особи, організації або пристрою, які створили певний аудіоносій. Згідно з наведеними вище прикладами, це може стосуватися того, з якого архіву чи прес-служби походить носій, який диктофон чи відеокамеру було використано, який інтерв'юер чи музичний виконавець створив певний твір тощо. Але це також означає, що сам носій має бути справжнім, тобто не бути повністю вигаданим у результаті зловмисних дій, або що він не був помилково переплутаний зі справжнім автентичним носієм.

Для цифрових аудіозаписів не кожен виявлений акт модифікації даних протягом життєвого циклу носія можна вважати зловмисним порушенням цілісності, як видно з наступних прикладів:

1. Деякі поширені модифікації є "прозорими" щодо звукових відчуттів і залишаються нечутними для пересічного слухача. Прикладами є:

- перенесення записів з диктофона у форматі MP3 на Audio CD;
- "копіювання" Audio CD у формати без втрат, такі як лінійні PCM-WAVE, FLAC або MPEG-4 ALS;
- підвищення частоти дискретизації до 48 або 96 кГц;
- кодування без втрат при високих налаштуваннях якості, наприклад, MP3 з бітрейтом 256 кбіт/с або вище;
- перетворення без втрат у формат BWF або RF64 під час вбудовування таких даних в аудіоархіви.

2. Інші модифікації насправді чутні і можуть навіть погіршити якість звуку – але не змінюють того, що слухачі "розуміють" інтелектуально про семантику. Прикладами є перекодування голосового запису в MP3-файл із середньою або нижчою якістю звуку заради економії місця на диску або пропускної здатності каналу зв'язку. Інший приклад – зведення 5.1-канальної звукової доріжки до стерео.

3. Нарешті, для інших чутних нешкідливих модифікацій байдуже, чи означають вони зміну "правильності" чи ні. Прикладом може слугувати фільтр шумозаглушення, який послаблює фонові шуми у записі голосу або телефонній розмові. Залежно від контексту такий фоновий шум може бути як несуттєвим, так і життєво важливим: З одного боку, таке шумозаглушення може покращити розбірливість голосової комунікації і дійсно збільшує/покращує те, що слухач розуміє. Це бажана функція, наприклад, при щоденному використанні гучного зв'язку або телефонних розмовах на вулиці. З іншого боку, фоновий шум може виявити важливу інформацію про контекст або фізичне середовище телефонної розмови. Це може бути важливо при розслідуванні кримінальних справ або інших інцидентів: у цьому випадку така фільтрація шуму може ненавмисно або навіть зловмисно замаскувати докази, що викривають (або виправдовують, відповідно).

Читач повинен мати на увазі, що попередні приклади дуже специфічні для аудіо або інших видів мультимедіа. Поняття "стиснення з втратами", "аналогова передача" або "фоновий/передній шум" не застосовуються до текстових даних (наприклад, у документах), числових даних (наприклад, у фінансових цифрах або банківських переказах) або двійкових даних (наприклад, виконуваний об'єктний код або зашифровані дані). Для цих видів не мультимедійних даних навіть невеликі зміни вмісту майже ніколи не можуть бути допустимими. Отже, поняття цілісності має бути переглянуте і адаптоване до випадку захисту мультимедійних даних від підробки і втручання.

Наведені приклади показують, що в мультимедійному контексті поняття "семантика" або "контент" самі по собі є неоднозначними і часто залежними від контексту. А для того, щоб семантично видобути, обробити та оцінити ідентичність даних за допомогою технічних засобів, потрібні великі зусилля: наприклад, сучасні технології семантичного видобування інформації або видобування знань – це надзвичайно широке поле досліджень, яке, зрештою, виходить далеко за межі цієї тези. Крім того, більшість семантичних технологій зосереджені на текстовому введенні, а не на голосових даних чи навіть загальному звуковому супроводі (музика, шуми тощо).

З іншого боку, "сприйняту ідентичність" загальних аудіоданих дещо легше моделювати і вимірювати з технічної точки зору: хоча це і не тривіально, але людське слухове сприйняття і якість сприйняття глибоко проаналізовано в кодуванні аудіо з втратами, аудіоаналізі та когнітивних науках протягом десятиліть. Були розроблені математичні моделі з психофізики, які можуть передбачити вплив об'єктивних модифікацій звуку на суб'єктивні відчуття слухача. Щоб уникнути необхідності подолання "семантичного розриву", в цій кваліфікаційній роботі перевірка цілісності даних буде проводитися на рівні сприйняття, як це пропонується нижче.

Пропозиція: Запровадити наступну ієрархічну модель різних рівнів цілісності для мультимедіа.

1. Бінарна цілісність: Описує ідентичність даних на рівні їхнього цифрового представлення. Два фрагменти аудіоданих вважаються ідентичними на двійковому рівні, якщо всі слова даних ідентичні, тобто якщо вони ідентичні "біт за бітом".

2. Перцептивна цілісність: Описує ідентичність того, що середньостатистичний користувач сприймає з набору мультимедійних даних. Тут відповідним критерієм є відчуття того, що він/вона чує від даних. Два фрагменти аудіоданих вважаються ідентичними за сприйняттям, якщо вони

звучать однаково. Це означає, що пересічні користувачі навряд чи зможуть відрізнити ці набори аудіоданих, прослухавши їх.

3. Семантична цілісність: Відображає значення (тобто семантику), яке асоціюється з даними. Семантична ідентичність зберігається, коли користувач розуміє одне й те саме з двох порівнюваних наборів даних або, більш загально, коли вони відповідають одному й тому ж поняттю.

Для спрощення різні рівні розглядаються в порядку зростання (див. рисунок 1.3): наприклад, перцептивна ідентичність є достатньою умовою для семантичної цілісності. Це визначення натхненне моделлю так званої м'якої автентифікації, описаної Чжу, Тьюфіком і Свонсоном [3]: автори розрізняють маніпуляції, які зберігають перцептивну якість (позначені авторами як "засновані на якості"), від маніпуляцій, які зберігають семантичне значення (позначені як "засновані на змісті").

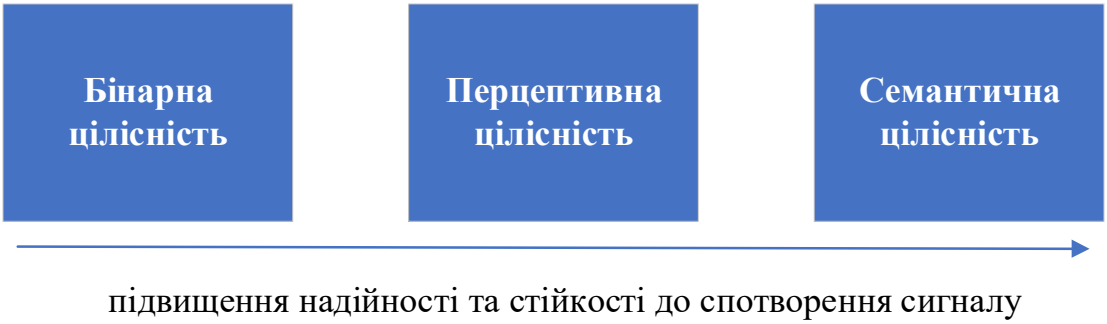

Рисунок 1.3 – Ієрархічна модель цілісності, що використовується в цій роботі

Як правило, для захисту аудіо та інших мультимедійних даних суворі вимоги точної перевірки цілісності, як у випадку з криптографічними хешами тощо, послаблюються: Досліджуваний підхід скоріше забезпечить певний ступінь толерантності до допустимих або, принаймні, нечутних модифікацій аудіо. Перцептивна ідентичність може і буде використовуватися як необхідний критерій для перевірки/фальсифікації семантичної цілісності аудіоданих протягом всієї роботи.

Механізми захисту цілісності даних від криптографії для довільних даних або мультимедійних даних, а також від загальних комунікацій або інформатики відомі давно. Але лише деякі з них відповідають розслабленому моделюванню цілісності сприйняття. У світлі реальних прикладів, наведених до цього часу, ця кваліфікаційна робота буде використовувати, розширювати і розвивати алгоритми з досліджень безпеки мультимедіа для перевірки перцептивної цілісності цифрових аудіозаписів. Для досягнення цієї мети визначено наступні завдання:

Чутливість до звукових модифікацій: Для виявлення зловмисного втручання, очевидно, бажано, щоб система використовувала коди автентифікації, які вказують на те, що два аудіосигнали (або їхні частини) сприймаються по-різному. Це дозволить виявити зловмисне втручання.

Інваріантність до нечутних модифікацій: Система повинна бути інваріантною або, принаймні, мало чутливою до звичайних перетворень аудіоносіїв, які залишають аудіосигнали схожими на слух. На відміну від методів перевірки криптографічної цілісності, допустимі та поширені кроки постобробки повинні бути допустимими у світлі визначеної вище перцептивної цілісності.

Локалізація: Наведені приклади показують, що в деяких сценаріях навіть видалення короткої ділянки тривалістю кілька секунд або менше може спричинити значні зміни семантики в записі голосу. Таким чином, дизайн системи повинен забезпечувати достатню часову локалізацію атаки. Локалізація потенційної атаки на дані може навіть допомогти ідентифікувати наміри зловмисника.

Оцінка: Крім того, запропонований підхід повинен давати градуйовану оцінку серйозності або ймовірності модифікації даних, спричиненої атакою. Це контрастує з простими бінарними результатами перевірки "прийнято/не прийнято", як, наприклад, у звичайних криптографічних примітивах або протоколах.

Доступність/Вбудовування: Як пояснювалося раніше, у великих обсягах мультимедійних даних може бути важко знайти правильний оригінальний запис або його окремі коди автентифікації (наприклад, як у хеш-базах даних). Це особливо актуально, коли доводиться працювати з великими обсягами мультимедійних даних, наприклад, у стокових фото- і відеоагентствах або в публічних аудіоархівах. Інша проблема полягає в тому, що для довготривалого архівування можна очікувати, що час від часу дані будуть перекодовані в інші (тобто новіші) формати файлів. Тому бажано зберігати коди перевірки "всередині" захищених медіа-даних, а не зберігати їх як окремі метадані, як накладні витрати у власному форматі файлу тощо.

Автентифікація походження джерела: Як другорядне завдання кваліфікаційної роботи, бажано перевірити походження та/або оригінального (sic!) творця звукозапису.

Безпека: Система повинна протистояти атакам, спрямованим безпосередньо на відповідні алгоритми. Зусилля, спрямовані на навмисне фальсифікацію шляхом використання слабких місць у механізмах захисту, повинні бути занадто великими, щоб їх можна було здійснити.

2 ІСНУЮЧІ МЕХАНІЗМИ ЗАХИСТУ ЦІЛІСНОСТІ ЦИФРОВИХ ДАНИХ

2.1 Огляд механізмів захисту цілісності та доброчесності

У розділі наведено абстрактну класифікацію та конкретний огляд існуючих механізмів захисту у світлі захисту цілісності аудіоданих. Це дозволяє відрізнити обраний підхід від сучасного рівня розвитку.

Існуючі механізми можна охарактеризувати за такими критеріями:

Модель цілісності (бінарна чи перцептивна ідентичність): Більшість механізмів спрямовані на перевірку або підтримку доброчесності саме на бінарному рівні. Інші підходи є більш "толерантними", оскільки вони іманентно допускають певне відхилення даних від початкового стану.

Модифікація вмісту обкладинки (активний та пасивний захист): Активні механізми спричиняють модифікацію даних у момент часу, з якого захист починає діяти. Модифікація може бути зворотною або незворотною. У пасивних підходах дані залишаються незмінними під час захисту.

Робочий процес (попередня обробка проти пост-обробки): Деякі механізми вимагають етапу підготовки або попередньої обробки, перш ніж захист набуде чинності. Інші підходи дозволяють застосовувати їх до відповідних даних у будь-який довільний момент часу без попередньої підготовки. Очевидно, що така постобробка можлива лише для пасивних механізмів.

Мета захисту (перевірка проти запобігання проти реконструкції): Деякі механізми технічно не запобігають модифікації даних, але дозволяють перевірити їх цілісність у більш пізній момент часу. Інші механізми підтримують цілісність, активно запобігаючи модифікації даних або, принаймні, дозволяючи реконструювати початковий стан даних.

Нижче розглядаються і порівнюються найпоширеніші класи алгоритмів захисту цілісності, які підходять для аудіоданих. Чисто апаратні рішення,

такі як WORM-носії ("записати один раз, прочитати багато [разів]", як оптичні носії CD-R), обладнання для блокування запису (як у криміналістичному обладнанні) або квантовий розподіл ключів, не обговорюються детально заради простоти.

Криптографічні хеші ("криптохеші") – це стандартний підхід у криптології для перевірки цілісності. Вони визначаються як алгоритми, які відображають двійковий набір вхідних даних c0 довільної довжини в компактний дайджест двійкового повідомлення H(c0) фіксованої довжини, зазвичай від 128 до 512 біт. Криптохеші за своєю конструкцією є надзвичайно чутливими: навіть найменша зміна вхідних даних спричиняє значні зміни вихідного хешу через передбачений алгоритмічний "лавинний ефект". Згідно з роботою Преніла "Аналіз та проектування криптографічних хеш-функцій" [4], перше формальне визначення було дано Дамгордом не раніше 1988 року в [5].

Особливо стійкі до колізій односторонні хеш-функції мають ряд властивостей, які є життєво важливими для досягнення цілей криптографічної безпеки. Як пояснює Преніл або Менезес та ін. [6], найбільш важливими з них є:

- перша властивість – стійкість до модифыкацыъ зображення/ односторонність: Хеш-функція повинна бути односторонньою, тобто для будь-якого довільного значення h_0 має бути "важко" знайти відповідний набір даних у вигляді попереднього зображення c_0 так, щоб $h_0 = H(c_0)$;

- друга стійкість до попереднього сигналу зображення: Для відомого набору даних c_0 з його хешем $h_0 = H(c_0)$, має бути "важко" знайти інший набір даних $c`_0 \neq c_0$ в якості попереднього образу так, щоб $H(c`_0) = H(c0)$. Така пара позначається як хеш-колізія;

- сильна стійкість до колізій: Має бути "важко" знайти/створити пару з двох різних наборів даних (c_0 ,c_1), хешування яких призведе до (довільного) ідентичного результату $H(c_0) = H(c_1)$.

Термін "складний" відображає те, що пошук 1-го попереднього зображення або створення колізій ("атака на день народження") вимагає значних обчислювальних витрат, і це завдання не може бути вирішене розумними зусиллями в світлі конкретного застосування або бажаного рівня безпеки.

Широко використовуваними підходами до криптохешування є MD5 [7], алгоритм безпечного хешування SHA-1 87] (досі) та його більш безпечні наступники SHA-2 і SHA-3, або сімейство RACE Integrity Primitives Evaluation Message Digest (RIPEMD) [9].

Поза контекстом криптографічної перевірки цілісності, хеш-функції є широко використовуваними допоміжними алгоритмами для загальної мети ідентифікації даних (тобто вказівки на ідентичність наборів даних). Наприклад, вони широко застосовуються в алгоритмах сортування та зіставлення, для обробки запитів до баз даних або для дедуплікації даних.

На додаток до крипто-хешів, алгоритм побудови деяких хеш-алгоритмів дозволяє включати залежність від секретного ключа K. Такий вид залежного від ключа хешу H(c0 , K) зазвичай позначається як (ключовий) код автентифікації повідомлень (MAC). Прикладом є CMAC, який є похідним від функцій шифрування на основі блоків [10]. Іншим прикладом є сімейство HMAC (наприклад, HMAC-SHA1, HMAC-MD5), яке алгоритмічно походить від криптографічних хеш-функцій, як пояснювалося вище [11]. Вони використовуються в протоколі TLS для перевірки цілісності.

Категоризація: MAC-адреси можна класифікувати так само, як і криптографічне хешування.

Нове сімейство хеш-функцій дозволяє ідентифікувати набори даних, які є лише майже ідентичними. Це так зване нечітке хешування реалізується шляхом кускового/часткового хешування ретельно відібраних підмножин даних. На відміну від криптохешування, значення нечітких хешів майже ідентичних наборів даних майже однакові (наприклад, з точки зору відстані Хеммінга).

Підхід і термінологію ввів Корнблум [12]. Нечітке хешування добре підходить для пошуку, зіставлення та криміналістичних застосувань, таких як зіставлення виконуваного двійкового коду або текстових рядків, що читаються людиною. Часто цитованими алгоритмами є ssdeep1,2 [12], алгоритм F2S2 [13] Вінтер, Шнайдер та Яннікос або saHash [14] Брайтінгера та Байєра.

Технологія перцептивного хешування – це метод ідентифікації цифрових мультимедійних даних. Як і в кожному хеш-алгоритмі, набір вхідних даних довільної довжини відображається в компактний дескриптор на основі вмісту або дайджест повідомлень дуже короткої довжини. У випадку перцептивного хешування процес вилучення хешу нечутливий або навіть інваріантний до нечутних/допустимих перетворень вхідного сигналу. Еквівалентними позначеннями є робастне хешування, а також аудіодактилоскопія.

Це досягається шляхом вилучення та постобробки релевантних для сприйняття аудіо ознак, так званих робастних 4 ознак. Робастне хешування чутливе до помірних спотворень, наприклад, спричинених стисненням MP3/AAC чи AVS з втратами [15] або перетворенням DA/AD під час аналогової передачі (динаміки, мікрофон тощо).

Перцептивне хешування дозволяє ідентифікувати аудіодані, які звучать подібно один до одного. І навпаки, воно дозволяє ідентифікувати модифікації даних, як тільки вони стають чутними. Таким чином, перцептивне хешування буде використано в контексті цієї тези для виявлення чутних модифікацій як індикатора для редагування аудіо.

Цифрові водяні знаки – це метод прихованого вбудовування секретного двійкового повідомлення в мультимедійні дані. Вбудовування досягається шляхом навмисного внесення незначних змін до вмісту, які представляють символи повідомлення, що приховується. Вбудовування повідомлення відбувається без зміни формату або розміру файлу. При нанесенні звукових водяних знаків також зберігається якість звуку, що сприймається. Для цього

ступінь модифікації вбудовування контролюється і зводиться до мінімуму за допомогою вмілого використання математичних психофізичних моделей ("психоакустики"), які відображають людське сприйняття.

При ретельному виконанні, вбудоване повідомлення звукового водяного знаку є нечутним і водночас може допускати помірне спотворення маркованого контенту – в той самий час. Остання властивість позначається як робастність також в області досліджень водяних знаків: виявлення символів вбудованого повідомлення пізніше можливе, навіть якщо маркований вміст піддається певній постобробці, наприклад, стисненню з втратами в форматі MP3/AAC або передачі в аналоговому вигляді.

Захист цілісності на основі водяних знаків може бути реалізований наступним чином: Вбудоване повідомлення являє собою нечутну "цифрову печатку", яка ламається, якщо до частин носія, захищеного водяним знаком, застосовано зловмисне виправлення.

Контрольні суми – це добре відомі коди виявлення помилок в обчислювальних і комунікаційних технологіях, які дозволяють вказати на помилки передачі. Як і в хешах та MAC-адресах, вхідне повідомлення довільної довжини відображається в дайджест коротких повідомлень дуже короткої довжини.

Дуже поширеною є циклічна перевірка надлишковості (CRC), запропонована Петерсоном і Брауном [16] у 1960-х роках, яка ґрунтується на модулі/залишку двійкового полінома від ділення двійкових чисел. Типова довжина CRC-кодів становить 32 біти, як у стандарті CRC-32. Іншими поширеними типами контрольних сум є перевірка на парність, коди Флетчера, оптимізовані за швидкістю коди Адлера або (зважене) перехресне підсумовування, як у книжкових ідентифікаторах ISBN, серійних номерах банкнот, посвідченнях особи/паспортах тощо.

2.2 Перцептивний аудіохешування/аудіофінгерпринтинг

Перцептивне аудіохешування – це метод ідентифікації, спеціально розроблений для цифрових аудіоданих. З технічної точки зору, перцептивне аудіохешування визначається як механізм відображення вхідного аудіосигналу довільної довжини в компактний дескриптор на основі вмісту або дайджест повідомлень дуже короткої довжини на виході. Формально кажучи, набір аудіоданих c_0 відображається у двійкову послідовність $H = H(c_0)$. На відміну від криптографічного хешування, перцептивне хеш-значення є стійким до незначних перетворень вхідного сигналу. Набір аудіоданих вважатиметься "подібним" з точки зору перцептивного хешування, навіть якщо дані піддаються, наприклад, стисненню MP3/AAC з помірними втратами або перетворенню DA/AD під час запису з мікрофону.

Як не дивно, ця технологія відома під ключовими словами як синоніми:

- "аудіо дактилоскопія";
- "надійне (аудіо) хешування" або;
- "підписи на основі вмісту";
- у численних публікаціях на сучасному рівні.

Для досягнення своєї стійкості алгоритми перцептивного хешування виділяють з даних акустичні особливості, які є релевантними для людського сприйняття і можуть бути використані для ідентифікації. Таким чином, Доетс і Лагендейк у своїй публікації "Теоретичне моделювання надійної системи аудіодактилоскопії" [17] чітко визначають вилучений аудіо "відбиток" як "...компактне представлення релевантних для сприйняття частин аудіоконтенту, яке можна використовувати для ідентифікації аудіофайлу, навіть якщо він сильно погіршився внаслідок стиснення або інших видів обробки сигналу".

Також кавер-версії або ремікси однієї і тієї ж оригінальної поп-пісні повинні мати відмінні один від одного аудіовідбитки. Дослідження з

ідентифікації таких нових версій пісень проводяться, але вони не є предметом цієї [18].

Це також виражається в тому, що хеші зазвичай витягують не тільки з повних аудіофайлів, але й з їхніх фрагментів. "Напівунікальність" визначається авторами як властивість, що для двох сигналів, які представляють різний зміст, їхні перцептивні хеші дозволять розрізнити ці сигнали. Щоб розвинути цю думку, термін "неперервний" припускає, що людське сприйняття і оцінка "акустичної схожості" можуть бути змодельовані і виміряні математичними функціями з неперервною областю визначення.

Як пояснює Ланчіні [19], типове вилучення перцептивного хешу виконується в наступні етапи обробки (див. рисунок 2.1):

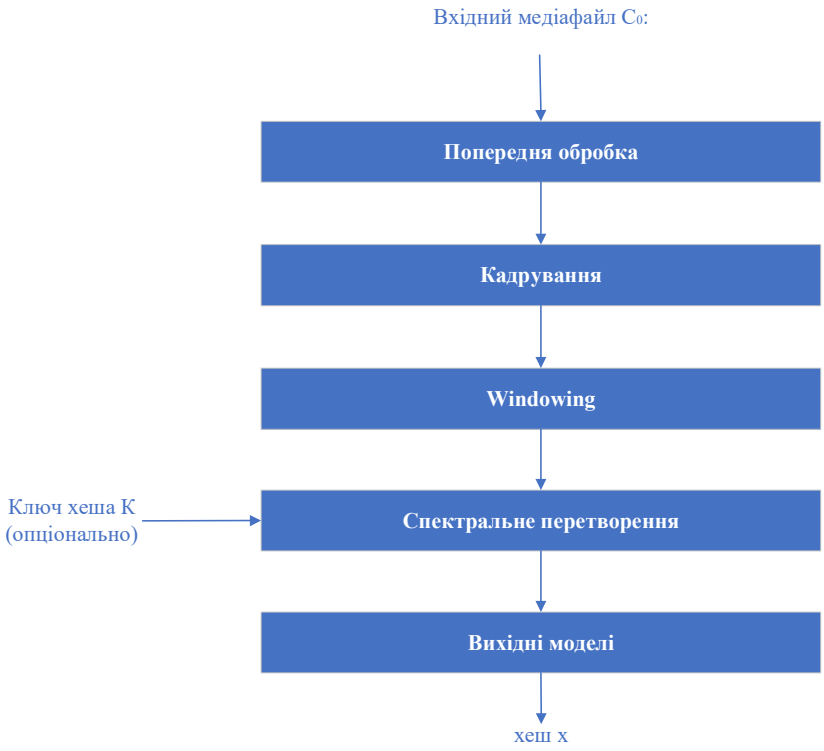

Рисунок 2.1 – Етапи обробки для вилучення хешу

1. Попередня обробка: Спочатку вхідні дані готуються для підвищення ефективності або робастності. Для прикладу цієї кваліфікаційної роботи

вхідні дані PCM можуть бути передискретизовані до стандартної частоти дискретизації та змішані до моно.

2. Кадрування: Типова довжина кадру в цій кваліфікаційній роботі становить 1024 або 2048, що дорівнює 23 або 46 мілісекунд, для яких стаціонарність є достатньою для більшості музичного та голосового контенту.

3. Віконне перетворення: Якщо виконується спектральне перетворення (на наступному кроці), кадрированные аудіодані часто перемножуються вибірка за вибіркою з відповідною віконною функцією, наприклад, вікном Геммінга або вікном фон Ганна.

4. Спектральне перетворення: У більшості підходів до віконних кадрів застосовується відповідне спектральне перетворення. Часто спектр є більш показовим для розпізнавання звуку, ніж часове представлення. Типові спектральні представлення використовують коефіцієнти ДПФ, (модифіковані) дискретні косинусні коефіцієнти (ДКП/MDCT) або так звані самочастотні цепстральні коефіцієнти (MFCC). У цій кваліфікаційній роботі буде використано оптимізовану за продуктивністю реалізацію ДПФ.

5. Виділення ознак: З представлення в перетвореному спектральному базисі вибираються відповідні ознаки для вилучення. Вмілий вибір ознак на цьому кроці обробки є життєво важливим для надійності та розрізнювальної здатності всього алгоритму. У деяких алгоритмах етап виділення ознак залежить від секретного ключа K як додаткового вхідного параметра.

6. Моделювання вихідних даних: Виділені ознаки аналізуються відповідно до попередньо визначеного алгоритму. Зазвичай, витягнутий робастний хеш є двійковою послідовністю. Однак у деяких підходах це n-кортеж з дійсними, тобто неперервними елементами.

3 ЗАПРОПОНОВАНА СИСТЕМА ВОДЯНИХ ЗНАКІВ ЦІЛІСНОСТІ НА ОСНОВІ ВМІСТУ

3.1 Опис підходу до захисту водяними знаками, вразливими до змін вмісту

Захист та перевірка здійснюється у два етапи (рисунки 3.1 та 3.2).

1. Етап вбудовування: Передбачається, що вхідним аудіоданим можна довіряти з точки зору цілісності та автентичності. Тепер вхідне аудіо розбивається на коротші сегменти, зазвичай від 10 до 20 секунд. З кожного аудіосегмента обчислюється його хеш. Витягнутий хеш потім вбудовується як надійний цифровий водяний знак в аудіодані, див. рисунок 3.1.

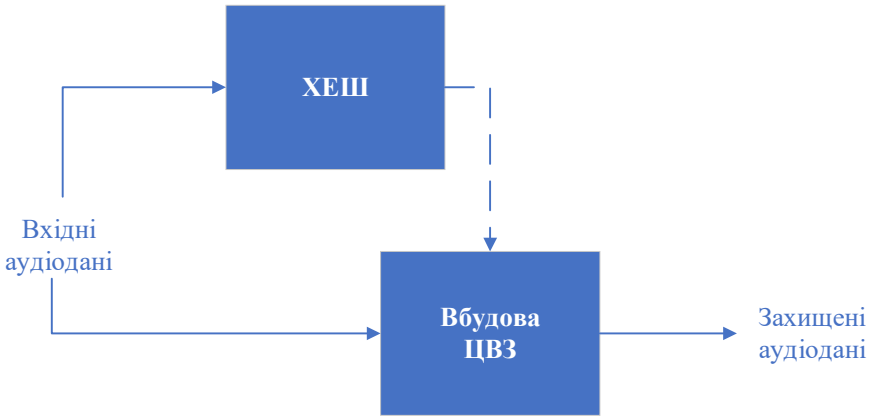

Рисунок 3.1 – Модель вразливого до вмісту водяного знаку – етап захисту/вбудовування

2. Етап виявлення: У будь-який момент часу повідомлення водяного знаку, тобто перцептивний хеш попереднього автентичного стану аудіоданих, може бути виявлено та вилучено із захищених ділянок аудіозапису. Порівняння попереднього та поточного стану перцептивного хешу дозволяє виявити порушення цілісності, див. рисунок 3.2.

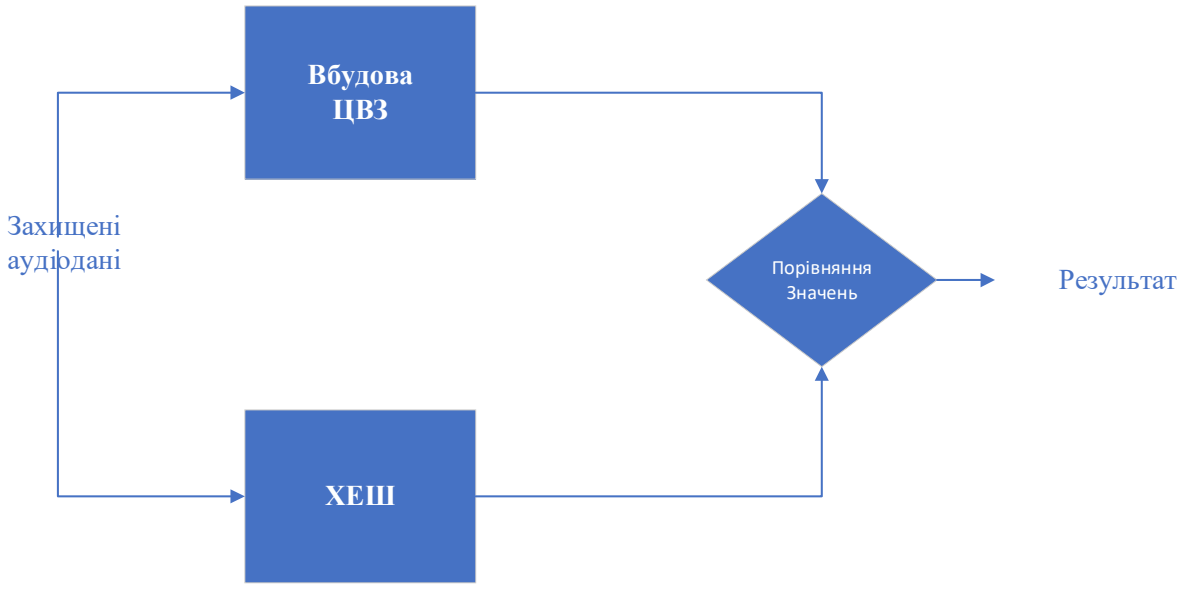

Рисунок 3.2 – Модель вразливого до зміни вмісту водяного знаку – етап перевірки/виявлення

Для цього пропонуються нові або вдосконалені підходи до перцептивного хешування, надійного нанесення водяних знаків на аудіо, перцептивного моделювання та їх відповідної комбінації. Перед цим необхідно запровадити низку технічних вимог.

1. Прийнятна якість звуку контенту з водяними знаками: Вбудовування водяного знаку має бути якомога непомітнішим, тобто погіршення акустичної якості, спричинене вбудовуванням, має бути мінімальним. Крім того, вбудовування має бути прозорим для цифрового представлення сигналу, тобто вбудовування має здійснюватися без зміни формату файлу.

Відповідність перцептивного хешування водяним знакам (і навпаки): У підході до захисту контенту водяними знаками перцептивне хешування аудіо має відображати технічні властивості та обмеження задіяного алгоритму водяних знаків, і навпаки, як показано нижче:

2. Стійкість аудіохешу до водяних знаків: Оскільки стійкий хеш є частиною системи автентифікації з водяними знаками, очевидно, що він також повинен бути стійким до спотворень, внесених водяними знаками, щоб уникнути негайних помилкових спрацьовувань. У цій кваліфікаційній роботі

буде запропоновано відповідне узгодження, щоб водяні знаки і перцептивне хешування не заважали один одному.

Потенціал водяних знаків: Звукові хеші, як і будь-який алгоритм хешування, обов'язково повинні мати достатню довжину (в бітах), щоб уникнути помилкових спрацьовувань і бути захищеними від атак грубої сили або цілеспрямованих атак на день народження. Більшість сучасних надійних підходів до створення водяних знаків, як правило, пропонують дуже низьку швидкість вбудовування – декілька бітів повідомлення, приблизно десяток бітів на секунду. З іншого боку, приблизно кожні 10 секунд необхідно вбудовувати мінімум 150 біт корисного навантаження. Таким чином, помірне збільшення можливостей вбудовування буде розроблено в ході цієї роботи.

Компактність аудіохешу: Враховуючи згадані обмеження ємності, аудіохеш повинен надавати дескриптор вмісту, який, хоча і відокремлює допустимі акти модифікації даних від зловмисних, є достатньо компактним. Для порівняння: 10,0 секунд необробленого аудіо у форматі PCM (стерео, 16 біт на відлік, частота дискретизації 44,1 кГц) становлять приблизно 1,5 МБ вхідних даних. Тим не менш, цей обсяг даних необхідно хешувати в компактний вихідний хеш, скажімо, 128 біт (тобто в 100 000 разів коротший, ніж вхідні), який все ще дозволяє відокремити допустимі від зловмисних актів модифікації даних.

Доступність: Важливою вимогою є те, що захист водяними знаками повинен бути забезпечений для всіх можливих вхідних аудіоданих, якщо це можливо. У контексті нанесення водяних знаків це може бути складним завданням, оскільки деякі види аудіоматеріалів не дуже добре підходять для нанесення вбудованих водяних знаків. Це часто трапляється, коли вхідний аудіоматеріал має лише невелику енергію сигналу на деяких ділянках своєї тривалості або в своєму спектрі. Прикладами є голосовий контент з дуже "чистим/сухим" звучанням голосу (як в аудіокнигах), контент з обмеженою частотною смугою (як у телефонії, радіозв'язку), а також деякі приклади сучасної електронної та/або експериментальної музики. Тут (часткова)

відсутність енергії сигналу ускладнює приховування аудіо водяного знаку, щоб він не став помітним або навіть дратівливим для слухача. Принаймні, повинні бути визначені обмеження щодо застосовності.

Приклад: Очевидно екстремальним прикладом критичного аудіоконтенту є композиція 4' 33" авангардного композитора-експериментатора Джона Кейджа: вона "містить" чисту тишу тривалістю 4:33 хвилини. Записи живих виступів (солістів або навіть симфонічних оркестрів) відрізняються повною тишею, за винятком навколишньої акустичної "атмосфери" концертного залу та віртуального напруження аудиторії46 .

4. Безпека: Водяний знак і перцептивний хеш повинні витримувати атаки, спрямовані безпосередньо на відповідний алгоритм. Що стосується водяного знаку, то неавторизований користувач не повинен мати змоги прочитати, змінити або видалити вбудоване повідомлення без того, щоб зробити обкладинку непридатною для використання. Що стосується перцептивного хешу, необхідно унеможливити зловмисне втручання, яке може обійти перевірку на основі хешу.

5. Ефективність: Нарешті, запропоноване рішення повинно забезпечувати достатньо швидкий час обробки. Можливість роботи в реальному часі може бути важливою вимогою, особливо для пристроїв запису в реальному часі, наприклад, відеокамери або записуючого пристрою на базі ПК.

Для кращого розуміння наступного пояснення, слід згадати визначення обчислення хеш-біт з оригінального підходу "Phillips Hash", запропонованого Хайтсма, Калкером.

Одне з найбільш часто цитованих і обговорюваних досліджень в області хешування аудіо було представлено Хайтсма, Калькером і Оостевеном [20], проведене в Phillips Research, Ейндховен, Нідерланди. Цей "хеш Філліпса" пояснюється насамперед тим, що багато інших сучасних публікацій - і ця робота в тому числі – були написані під впливом цієї роботи. Дизайн вражає своєю простотою і бажаною надійністю.

Надійне вилучення ознак використовує спектрально-часовий аналіз абсолютного значення коефіцієнтів Фур'є аудіосигналу. Для вилучення так званого субвідбитка з кожного аудіокадру виконується наступна обробка:

- в оригінальному алгоритмі аудіосигнал подається в цифровому вигляді за допомогою відліків PCM і аналізується в досить великих кадрах L = 2^{14} =16384 відліків, які перекриваються на 31/32. Великі фрейми забезпечують хорошу частотну роздільну здатність, але погану часову роздільну здатність. Останнє компенсується великим перекриттям кадрів;

- дані кадру спочатку зважуються за допомогою вікна Хеммінга [S+1997], щоб уникнути ефекту спектрального витоку. Потім обчислюється перетворення Фур'є;

- потім коефіцієнти Фур'є $\{r_k\}$ передискретизуються відповідно до логарифмічно масштабованої частотної осі: лінійна частотна вісь з індексами k відображається на дециміновану шкалу відповідно до 33 музичних півтонів з індексами i, що відповідають 300-2000 Гц. Децимований спектр виражається через величину $\{e_i,t\}$, яка є частковою енергією сигналу в i-му півтоні на t-му часовому кроці. При цьому враховується, що людське сприйняття висоти звуку можна моделювати переважно на логарифмічній частотній шкалі, як це відомо для музичних тонів.

Експериментально встановлено, що взаємні спектрально-часові відмінності між цими енергетичними коефіцієнтами є досить робастною ознакою для довільного музичного та голосового аудіоматеріалу. Робастна ознака виводиться шляхом порівняння різниць сусідніх енергетичних коефіцієнтів (на частотних індексах i та i +1) в заданий момент часу t з такими ж на наступному часовому кроці t +1 згідно з наступним визначенням:

$$d_{i,t} := e_{i,t} - e_{i,t+1} - \left[e_{i+1,t} - e_{i+1,t+1} \right],\ i = 1...32\ . \qquad (3.1)$$

Біти звукового хешу Hi,t визначені авторами як знак величини $d_{i,t}$.

Згідно з рівнянням (3.1) оригінального алгоритму, коефіцієнти ДПФ "вибираються" в регулярно зростаючому порядку за часовими та частотними індексами: послідовні індекси бітів хешу відповідають послідовним/суміжним частотним індексам.

Візуалізація вибору ознак наведена на рисунку 3.3. Зауважимо, що графік дає лише символічну візуалізацію невеликого діапазону з області 12 на 10 частотно-часових показників. У реальній реалізації алгоритму використовуються наступні налаштування.

32-бітний так званий субвідбиток витягується в момент часу t0 шляхом оцінки 33 енергетичних коефіцієнтів на кадр відповідно до рівняння (3.1).

Це повторюється 256 разів, кожні 11,6 мілісекунд з перекриттям 31/32. Повний блок відбитків $\{H_{i',t'}\}$ визначається як конкатенація цих 256 послідовних субвідбитків, тобто t'=1...32, i'=1...256.

Таким чином, корисне навантаження блоку відбитків становить 256 - 32 біт = 8 кбіт. Це приблизно 256 – 11,6 мілісекунд ≈ 3 секунди аудіо. Отже, швидкість передачі даних відбитків пальців становить приблизно 8 кбіт/3 с ≈ 2,6 кбіт/с. Ця досить висока швидкість передачі даних в основному спричинена перекриттям 31/32, яке компенсує обмежену часову роздільну здатність одиночних перетворень Фур'є.

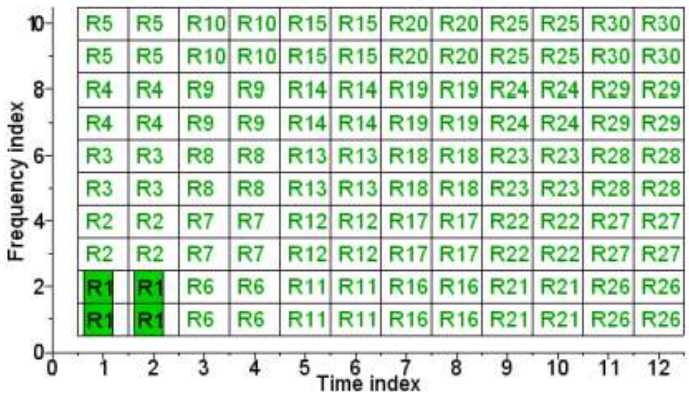

Рисунок 3.3 – Схема виділення ознак в аудіохеші Філліпса

Тобто, для кожного біта хешу обчислюються чотири енергетичні коефіцієнти з сусідніх частотних індексів k і k + 1, і оцінюються послідовні часові інтервали t і t + 1.

Візуалізація наведена на рисунку 3.4 (лівий рисунок): Кожна "клітинка" на рисунку відповідає одному часово-частотному індексу (k, t) у віртуальній "області" в часово-частотній області. Чітко видно закономірний розподіл "вибраних" індексів вздовж часової та частотної осей. Зауважте, що рисунок дає лише символічне представлення часово-частотної області. Дуже обмежений частотний діапазон у 10 індексів є спрощеним представленням для демонстрації. У цій роботі фактична "висота" "області" можливих частотних індексів може досягати 1024, що відповідає (половині) довжини кадру ДПФ.

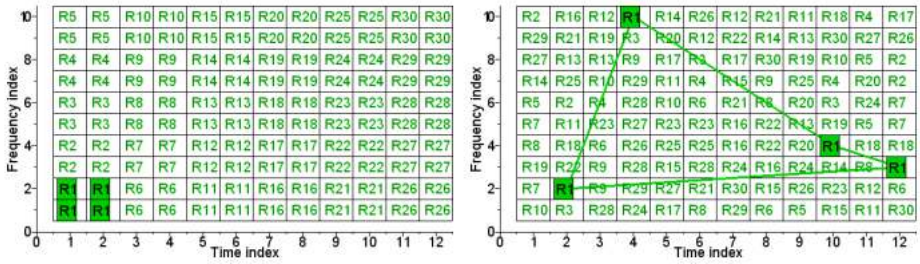

Рисунок 3.4 – Запропонована схема виділення ознак rMAC – символічне представлення індексів ДПФ у часово-частотній області

Кожен набір з чотирьох відповідних коефіцієнтів, позначених як "Rm", робить внесок у m-й біт хешу.

Коефіцієнти R1, які відповідають першому біту хешу, виділено для демонстрації; Ліворуч: вибір коефіцієнтів ДПФ у рівномірному порядку згідно з рівнянням (3.1); Праворуч: вибір у псевдовипадковому порядку згідно з рівнянням (3.2).

Таким чином, кожен "субвідбиток" (використовуючи оригінальне позначення) тепер буде вилучатися з аудіосегмента, що складається з M послідовних кадрів з кроком у кілька сотень разів, тобто з аудіозапису

тривалістю в кілька секунд. Стерео або багатоканальні вхідні PCM-сигнали змішуються до моно перед обчисленням спектру ШПФ.

Далі пропонується отримати біти хешу з чотирьох коефіцієнтів ШПФ на довільних часових кроках t_1 , t_2 , t_3 , t_4 = 1...L та індексах смуг k_1 , k_2 , k_3 , k_4 у часово-частотній області:

$$d'_{k,t} := e_{k_1,t_1} - e_{k_2,t_2} - \left[e_{k_3,t_3} - e_{k_4,t_4} \right], \tag{3.2}$$

Вибір часових та частотних індексів здійснюється псевдовипадково в залежності від секретного ключа K_2 , замість того, щоб використовувати сусідні часові та частотні індекси. Такий підхід приховує від зловмисника, які частотні компоненти впливають на певний хеш-біт (а які ні). Дивіться рисунки 3.4 (правий рисунок) або рисунок 3.5 для візуалізації: там чітко видно псевдовипадковий вибір часово-частотних індексів.

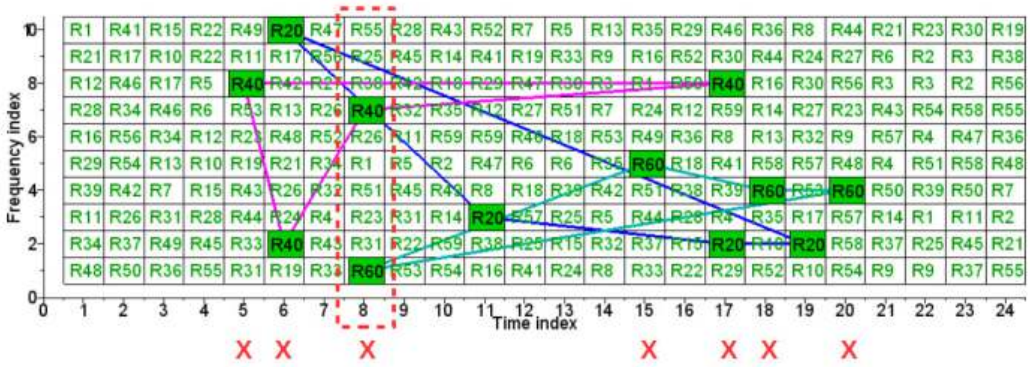

Рисунок 3.5 – Приклад: покриття rMAC на різних часових кроках ("розсіяний" режим)

Для прикладу з демонстраційною метою виділено квадрати коефіцієнтів, які відповідають 20-му, 40-му та 60-му бітам хешу (позначені як "R20", "R40" та "R60"); пунктирна рамка: передбачається атака на часовому індексі 8. Маркери "X" – часові індекси, що відповідають атакованим бітам R40 та R60.

Оскільки запропонований аудіохеш залежить від спільного секрету K2, його можна розглядати як код автентифікації повідомлення (MAC), а не як хеш. Як буде показано в наступному розділі, він витримує ряд перетворень сигналу.

Далі описано запропоновані адаптації алгоритму накладання водяних знаків на аудіо, які стали необхідними для цієї роботи з метою досягнення кращої продуктивності та стійкості водяних знаків.

Для вбудовування перцептивного аудіохешу як повідомлення водяного знаку в якості технічної основи використовується існуючий підхід водяного маркування зі сліпим спектром [19]. Тут вбудовування здійснюється в основному в області Фур'є відповідно до підходу до нанесення водяних знаків Patchwork.

В оригінальному підході кожен кадр містить один біт повідомлення водяного знаку. Зазвичай для кожного кадру псевдовипадково вибирають кілька десятків із кількох сотень можливих коефіцієнтів ШПФ. Щоб підвищити стійкість, кожен біт повідомлення вбудовується з надлишком у групу послідовних кадрів (зазвичай від двох до десяти кадрів). Це збільшує популяцію статистики виявлення Patchwork у рівнянні. Це дозволяє виявляти біти повідомлення більш суттєво, навіть за наявності спотворень маркованого сигналу. Нарешті, наступні біти повідомлення водяного знаку вбудовуються в наступні групи кадрів у послідовному порядку.

Рисунок 3.6, ліворуч, демонструє вбудовування бітів повідомлення в послідовному порядку. Для спрощення візуалізації загальна довжина повідомлення становить лише M=4 біти, а в кожному кадрі вибирається лише 2N=6 частотних індексів з W=10 загальних (смуга пропускання). Кожен біт повідомлення вбудовується з надлишком тричі, тобто в три послідовних кадри. Тобто, загальна тривалість становить 3 – 4 кадри = 12 кадрів, а для кожного біта повідомлення використовується 3 – 6 коефіцієнтів = 18 коефіцієнтів. Всього віртуальна область в часо-частотній області охоплює 12 – 10 = 120 можливих коефіцієнтів.

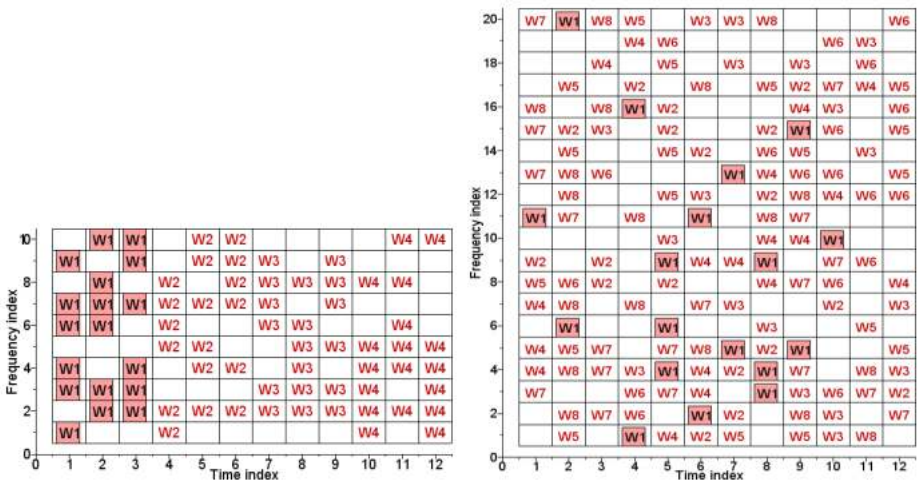

Рисунок 3.10 – Запропоновані схеми вбудовування водяних знаків

Біти повідомлення вбудовуються у послідовному часовому порядку (лівий рисунок) та у псевдовипадковому часовому порядку (правий рисунок). Кожні 9+9 відповідних коефіцієнтів Wn дають внесок у n-й біт повідомлення згідно з рівнянням (2.6). Відповідні коефіцієнти першого біта повідомлення W1 виділено з демонстраційною метою.

Для достатньо тонкого виявлення знімання бажано, щоб захищеними були одиниці часу тривалістю в декілька секунд. Таким чином, достатня здатність вбудовування звукових водяних знаків є життєво важливою з точки зору вимог. Дотепер існуючий підхід, як пояснювалося раніше, був недостатнім через його досить негнучке призначення коефіцієнтів Фур'є водяних знаків до бітових індексів повідомлення. Більш детальний аналіз показує, що необхідно покращити пропускну здатність у 5-10 разів. Більш детальну інформацію можна знайти в числовому прикладі, наведеному нижче.

Фіксоване розміщення відкидається заради більшої гнучкості та покриття, як показано на рисунку 3.6, праворуч. "Хмароподібне" розсіювання місця вбудовування дозволяє легко збільшити потужність вбудовування порівняно з фіксованим відображенням за рахунок збільшення ширини підсмуги водяних знаків (вертикальна вісь).

Як реалістичну цифру, можна припустити, що фактична загальна довжина повідомлення становить приблизно 450 біт. Це включає сам rMAC, додаткові метадані, короткий CRC-код і те, що використовується пряма корекція помилок. Для досягнення достатньої стійкості використовується 180 коефіцієнтів для вбудовування кожного з бітів у спектр. Для вбудовування використовується діапазон частот від 1 кГц до 9 кГц (тобто смуга пропускання 8 кГц). Ця смуга відповідає W = 8 кГц – 2048/44,1кГц = 371 частотному індексу. Це візуалізується як "висота" віртуальної "області" в часово-частотній області, як показано на рисунку 3.6.

Загалом, вбудовування/виявлення Patchwork виконується у наступні етапи обробки: Спочатку, для вбудовування кожного окремого біта повідомлення використовуються дві незв'язні підмножини.

4 КОМБІНОВАНИЙ ПІДХІД ДО ПЕРЦЕПТИВНОГО ХЕШУВАННЯ/ВОДЯНИХ ЗНАКІВ

4.1 Адаптація хешування та вбудовування водяних знаків

Як зазначалося в розділі 2, у системі захисту від вразливого вмісту важливо, щоб спотворення аудіоданих, спричинені вбудовуванням будь-якого повідомлення водяного знаку, не змінювали суттєво значення rMAC. Для алгоритму rMAC, абсолютно нечутливого до послідовного вбудовування водяних знаків, можна було б очікувати, що відстань Хеммінга rMAC до і після застосування захисту дорівнюватиме нулю. Насправді, на більш ранній стадії дослідження, опублікованій в [18], відстані Хеммінга не зникали повністю.

Для експериментів, опублікованих у [16], rMAC довжиною M=128 біт було вилучено з одиниць аудіоданих тривалістю 10 секунд (6,5 годин загальної тривалості). За розумної сили вбудовування спотворення процесу вбудовування водяного знаку призвели до того, що деякі біти rMAC позначили зміну значення біта, як видно з гістограми на рисунку 4.1.

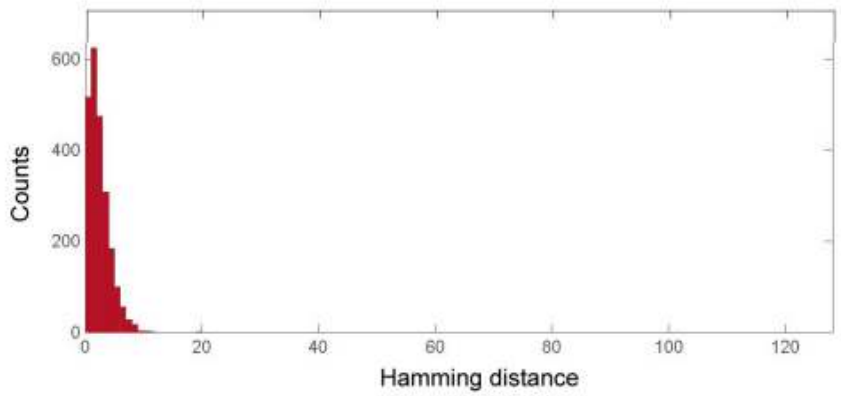

Рисунок 4.1 – Гістограма відстані Хеммінга rMAC для аудіоданих з водяними знаками та оригінальних аудіоданих

Під час попереднього аналізу було виявлено, що вилучення rMAC і вбудовування водяного знаку в конкретний аудіокадр не були повністю розділені через ефекти спектрального витоку: вбудовування водяного знаку в певний коефіцієнт Фур'є має вплив і на сусідні коефіцієнти в спектрі. Якщо такий коефіцієнт обрано під час вибору функції rMAC, то значення rMAC, до якого цей коефіцієнт впливає, може змінитися. Це часто спостерігається для згаданих "слабких" бітів, для яких неперервна міра d', як у рівнянні (3.1), має абсолютне значення, близьке до нуля. Таким чином, процес вилучення ознак rMAC і відповідну схему робастного нанесення водяних знаків пропонується належним чином узгодити.

Для цього аудіодані внутрішньо розділяються вздовж часової осі, спочатку розбиваючи аудіокадри в кожній захищеній аудіосекції на два класи:

- більшість аудіокадрів використовується виключно для вбудовування повідомлення з водяним знаком;
- меншість аудіокадрів використовується виключно для вилучення rMAC.

Мультиплексування вздовж часової осі допомагає уникнути небажаних ефектів витоку під час вбудовування водяного знаку. Мультиплексування полегшується завдяки тому, що під час попереднього обчислення rMAC кадри ШПФ не перекриваються.

4.2 Корисне навантаження вбудованого повідомлення

Повідомлення про вбудований водяний знак визначається наступними елементами.

Очевидно, що rMAC є важливою частиною корисного навантаження вбудованого повідомлення з водяним знаком. Використовується rMAC довжиною M=128 біт, що означає основну частину загального корисного навантаження.

Як обговорювалося в розділі 2, необхідно запобігати цілеспрямованим атакам, які дозволяють непомічену модифікацію шляхом видалення або заміни повних аудіосегментів. Для цього, як незначне вдосконалення, до rMAC кожного аудіосегмента додано часовий код. Для простоти використовується індекс аудіосегмента: перший сегмент позначається цілочисельним часовим кодом "0", наступний – "1" і так далі. Це дозволяє в подальшому перевірити, що порядок аудіосегментів правильний і жодного сегмента не пропущено. Атаки шляхом видалення або заміни повних аудіосегментів можна виявити за допомогою "атаки Холлімана". До повідомлення з водяним знаком додається часовий код від 8 до 10 біт. Це досить очевидне, просте, але дійсно ефективне вдосконалення також можна знайти в роботі Парка [21]. Це відрізняється від роботи Чен Фана та ін. [22], в якій пропонується реалізувати пов'язаний ланцюжок вилучення ознак і вбудовування ознак: автори описують, що звукові ознаки певного аудіосегмента вбудовуються в наступний сегмент, і так далі. На відміну від цього, в даній кваліфікаційній роботі часовий код вбудовується у відповідний сегмент. Це має ту перевагу, що кожен сегмент всередині аудіофайлу є самодостатнім щодо всіх метаданих, необхідних для його автентифікації, навіть якщо доступні лише фрагменти файлу.

До повідомлення з водяним знаком додається CRC-16. Це дозволяє перевірити, чи правильно було отримано водяний знак

Кожен аудіосегмент, що містить повідомлення водяного знаку, починається з послідовності синхронізації водяного знаку. Формально це теж повідомлення водяного знаку, але воно має фіксоване значення, як стартовий код. Детектор водяних знаків спочатку намагається знайти і зафіксувати/включити відому(!) послідовність "Sync". Якщо результат виявлення перевищує заздалегідь визначений поріг, запускається подальший пошук повідомлення та перевірка rMAC. У досліджуваному підході можна регулярно досягати точності часової синхронізації на стороні детектора на рівні ±1 зразка (зразків) ПКМ. Це перевершує точність оригінального

алгоритму Штайнебаха, який синхронізує з точністю ±10 відліків ПХМ. Деталі в цій роботі не розглядаються.

Таким чином, корисне навантаження вбудованого повідомлення в кінцевому підсумку може бути записано як

$$\{Sync + Message_1\} \{ Sync + Message_2 \} \{ Sync + Message_3 \} \ldots (4.1)$$

З

$$Message_i = TurboCoding\,(rMAC_i + TimeCode_i + \\ + CRC(\, rMAC_i + TimeCode_i \,)\,) \quad . \qquad (4.2)$$

В експериментальній оцінці в цій кваліфікаційній роботі приблизна тривалість послідовності "Синхронізація" зазвичай становить 2 секунди, в той час як повідомлення водяного знаку вимагають від 10 до 15 секунд тривалості кожне [23].

Далі буде коротко описано програмну реалізацію алгоритму. Видобування rMAC та послідовне вбудовування відбувається наступним чином (спрощену блок-схему див. на рисунку 4.1).

1. Вхідними даними для програмної реалізації є файл даних обкладинки та текстовий файл, що містить секретні ключі (позначені як "Звукові дані обкладинки" та "Секретний ключ" на рисунку 3.14).

2. Крім того, на вхід потрібно передати конфігураційний файл, що контролює технічні налаштування як для вилучення rMAC, так і для нанесення водяних знаків (позначений як "Технічні конфігураційні дані"). Найбільш важливими технічними параметрами є:

- смуга частот, що використовується для вилучення rMAC і накладання водяного знаку в Герцах (що символічно визначає "висоту" віртуальної "області" в часо-частотній області, як показано на рисунку 3.13);

- кількість частотних коефіцієнтів, яка використовується на один біт повідомлення (що впливає на стійкість водяного знаку та загальну тривалість, необхідну для вбудовування);
- відсоток часових інтервалів, з яких дозволено вилучення rMAC (контролює, наскільки щільно здійснюється вилучення rMAC);
- відсоток частотних коефіцієнтів на кожному часовому кроці в "області", які дозволено використовувати для вбудовування (що впливає на прозорість і загальну тривалість, необхідну для вбудовування);
- запланована довжина rMAC та часового коду в бітах тощо.

Виходячи з цих параметрів, можна обчислити простими арифметичними діями, скільки кадрів потрібно для вбудовування повного повідомлення водяного знаку (що символічно визначає "довжину" "області" в часо-частотній області).

ВИСНОВКИ

У кваліфікаційній роботі було вирішено наступні завдання:

- проведено аналіз сучасних методів перевірки цілісності аудіофайлів;
- проведено аналіз методів стеганографії для аудиофайлів;
- реалізовано алгоритм перцептивного аудіохеша;
- реалізовано алгоритм Patchwork вбудови ЦВЗ;
- реалізовано комбінований метод перевірки цілісності аудіоданих на основі алгоритму перцептивного аудіохеша та алгоритму Patchwork вбудови ЦВЗ.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Charlton T. Lewis and Charles Short. Latin Dictionary: Founded on Andrews' edition of Freund's Latin dictionary. Oxford University Press. Clarendon Press, Oxford, UK,1879. 480p.

2. P. G. W. Glare, editor. Oxford Latin Dictionary. Oxford University Press? 1982. 355p.

3. Bin B. Zhu, Mitchell D. Swanson, and Ahmed H. Tewfik. When Seeing Isn't Believing (Multimedia Authentication Technologies). IEEE Signal Processing Magazine, 21:40–49, March 2004

4. art Preneel. Analysis and Design of Cryptographic Hash Functions. PhD thesis, KU Leuven, Belgium, 1993. 516p.

5. I.B. Damgård. Collision free hash functions and public key signature schemes. In: D. Chaum and W.L. Price, editors, Advances in Cryptology, EurocryptŠ87, LNCS 304, Springer, 1988. P.203–216.

6. lfred J. Menezes, Scott A. Vanstone, and Paul C. Van Oorschot. Handbook of Applied Cryptogra phy. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 1996.

7. Ronald L. Rivest. RFC 1321: The MD5 Message-Digest Algorithm. Number 1321 in Request for Comments (RFC). Internet Engineering Task Force (IETF), April 1992

8. D. Eastlake, 3rd and P. Jones. RFC 3174: US Secure Hash Algorithm 1 (SHA1). Number 3174 in Request for Comments (RFC). Internet Engineering Task Force (IETF), September 2001.

9. H. Dobbertin, A. Bosselaers, and B. Preneel. RIPEMD-160, a strengthened version of RIPEMD. Fast Software Encryption, 1996. P. 71–82.

10. Morris J. Dworkin. SP 800-38B. Recommendation for Block Cipher Modes of Operation: The CMAC Mode for Authentication. Technical report, National Institute of Standards and Technology (NIST), Gaithersburg, MD, United

States, 2005.

11. . Krawczyk, M. Bellare, and R. Canetti. RFC 2104: HMAC – Keyed-Hashing for Message Authen tication. Number 2104 in Request for Comments (RFC). Internet Engineering Task Force (IETF), February 1997.

12. Jesse Kornblum. Identifying almost identical files using context triggered piecewise hashing. Digital Investigation, 3:91–97, September 2006.

13. Christian Winter, Markus Schneider, and York Yannikos. F2S2: Fast forensic similarity search through indexing piecewise hash signatures. Digital Investigation, 2013. Vol.10, No 4. P 361–371

14. Frank Breitinger, Georg Ziroff, Steffen Lange, and Harald Baier. Similarity Hashing Based on Lev enshtein Distances. In: Gilbert Peterson and Sujeet Shenoi, editors, Advances in Digital Forensics X, volume 433 of IFIP Advances in Information and Communication Technology. Springer Berlin Heidelberg, 2014. P133–147

15. Wen Gao. AVS standard - Audio Video Coding Standard Workgroup of China. In: 14th Annual International Conference on Wireless and Optical Communications, 2005. WOCC 2005, pages 54–, April 2005.

16. W.W. Peterson and D.T. Brown. Cyclic codes for error detection. Proceedings of the IRE, , January 1961. Vol. 49, No 1. P. 228– 235

17. Peter Jan O. Doets and R.L. Lagendijk. Theoretical modeling of a robust audio fingerprinting system. In: Fourth IEEE Benelux Signal Processing Symposium, 2004. P. 228– 235

18. Ning Chen, J. Stephen Downie, Hai-dong Xiao, and Yu Zhu. Cochlear pitch class profile for cover song identification. Applied Acoustics, 2015.

19. . Lancini, F. Mapelli, and R. Pezzano. Audio content identification by using perceptual hashing. In: Multimedia and Expo, 2004. ICME ’04. 2004 IEEE International Conferenc, 2004. P 739–742.

20. J.A. Haitsma, J.C. Oostveen, and A.A.C. Kalker. Robust audio hashing for content identification. In: Content based multimedia Indexing (CBMI) 2001, Brescia Italy, 2001.

21. Chang-Mok Park, Devinder Thapaa, and Gi-Nam Wang. Speech authentication system using digital watermarking and pattern recovery. Pattern Recognition Letters, June 2007. Vol. 28. P. 931–938.

22. Fan Chen, Hong-Jie He, and Hong-Xia Wang. A fragile watermarking scheme for audio detection and recovery. In: Image and Signal Processing, 2008. CISP ’08. 2008. Vol. 5. P. 135 –138.

23. Бондарчук П. С., et al. "Аналіз вучасних методів перевірки цілісності аудиофайлів." InterConf (2023): 44-52.