УДК 519.8: 004.934

А.М. НИЩУК
Науково-дослідний інститут воєнної розвідки, Київ, Україна, e-mail: *svs14@ukr.net*.

С.М. НІКОЛАЄВ
Науково-дослідний інститут воєнної розвідки, Київ, Україна, e-mail: *divan24@i.ua*.

О.М. РОМАНОВ
Науково-дослідний інститут воєнної розвідки, Київ, Україна, e-mail: *rolex@i.ua.*

МЕТОДИКА АНАЛІЗУ БІТОВИХ ПОТОКІВ НА ОСНОВІ ВИКОРИСТАННЯ ВІДСТАНІ ДАМЕРАУ–ЛЕВЕНШТЕЙНА ТА ІНШИХ МЕТРИК

Анотація. Проведено аналіз особливостей різних відстаней (метрик) під час розв'язання задач порівняння бітових потоків. На прикладі розглянуто застосування запропонованої методики для розв'язання задачі ідентифікації бітових потоків цифрового кодування мовлення. Показано, що у цьому випадку найбільш ефективною є метрика Дамерау–Левенштейна.

Ключові слова: бітовий потік, відстань, метрика, цифрове кодування мовлення.

ВСТУП. АКТУАЛЬНІСТЬ ДОСЛІДЖЕНЬ ТА АНАЛІЗ НАЯВНИХ РЕЗУЛЬТАТІВ

Післядетекторне оброблення сигналів зазвичай означає роботу з цифровими даними, представленими у вигляді бітових потоків (bitstreams). Це стосується більшості радіотехнічних та телекомунікаційних систем, які на фізичному рівні обробляють радіосигнали, а на канальному та інших рівнях — бітові потоки. Дослідники вживають такі синоніми терміна «бітовий потік»: цифровий потік; двійкова, дискретна або бітова послідовність; цифрове представлення сигналу; файл тощо. Автори пропонують послуговуватися саме терміном «бітовий потік», який має чіткий відповідник англійською мовою. Завдяки цьому можна коректно розрізнити цифрове оброблення сигналів (signal processing) та оброблення бітових потоків (bitstream processing).

Очевидно, що під час аналізу бітових потоків одним з важливих процесів є пошук. Цей пошук є основою відповідних систем в інтернеті, його також використовують у криптографії для виявлення періодичностей у бітових потоках. На практиці задачу порівняння та визначення статистичних властивостей бітових потоків розв'язують у різних технічних галузях: від радіотехніки і передавання інформації до біомедицини, оброблення зображень тощо. Основою порівняння бітових потоків є пошук періодичностей і повторень окремих фрагментів з використанням різних методів. Наприклад, у криптографії оцінювання бітового потоку здійснюють за допомогою тестів NIST [1]. Їх рекомендовано використовувати, щоб з'ясувати, чи можна вважати бітовий потік випадковим. Слід зазначити, що не завжди на основі цих тестів можна відрізнити архівовані файли від шифрованих [2]. До методики на основі зазначених тестів відносять перевірки бітового потоку на статистичні властивості, виявлення факту оброблення регістрами зсуву тощо. Однак, для порівняння двох та більше бітових потоків тести NIST застосовувати незручно і в багатьох випадках вони не забезпечують потрібного результату. Загалом для розроблення деякої системи ідентифікації або класифікації, яка ґрунтується на порівнянні бітового потоку з еталоном, потрібно аналізувати ці потоки на основі математичних виразів для обчислення певних відстаней (метрик). Аналіз особливостей зазначених метрик і є метою цієї статті.

© А.М. Нищук, С.М. Ніколаєв, О.М. Романов, 2023