ВІДГУК

офіційного опонента
на дисертаційну роботу Самойлика Євгена Олександровича
на тему «Лексикографічні методи захисту мовної інформації»,
представлену на здобуття наукового ступеня кандидата технічних наук
за спеціальністю 05.13.21 – Системи захисту інформації

Актуальність теми дисертації, зв'язок з науковими програмами, планами, темами. Проблема забезпечення інформаційної безпеки є на сьогодні однією з найгостріших не лише в Україні, але і в розвинених країнах світу. Досвід експлуатації інформаційних систем і ресурсів в різних сферах життєдіяльності показує, що існують різні і вельми реальні загрози пошкодження інформації, що приводять до матеріальних і інших збитків. При цьому забезпечити на 100% безпеку інформації практично неможливо. Тому в цих умовах одним із головних напрямків розвитку сучасних телекомунікацій є удосконалення існуючих і створення нових методів захисту критично важливої інформації.

При цьому, величезне значення для забезпечення конфіденційності інформації мають криптографічні системи захисту даних. Їх застосування забезпечує конфіденційність, наприклад, текстових даних (мовної інформації) навіть у разі потрапляння до рук сторонньої особи. Але будь-який криптографічний алгоритм має таку властивість як крипостійкість, тобто його захист має певну межу. Можна констатувати, що практично немає шифрів, які не можна було б зламати — це лише питання часу і коштів. Ті алгоритми, які ще кілька років тому вважалися надійними, сьогодні вже можуть бути скомпрометованими.

Таким чином, існує важлива наукова задача удосконалення існуючих і створення нових методів забезпечення ефективного захисту мовної інформації під час її передавання відкритими каналам зв'язку. Це і обумовлює **актуальність** даного дисертаційного дослідження.

Наукова робота за темою дисертації пов'язана з реалізацією положень «Стратегії розвитку інформаційного суспільства в Україні» (затверджена Кабінетом Міністрів України від 15 травня 2013 року), «Основними науковими напрямами та найважливішими проблемами фундаментальних досліджень у галузі природничих, технічних і гуманітарних наук НАН України на 2014-2018 роки», Стратегією національної безпеки України від 26 травня

2015 року № 287/2015, Стратегією кібербезпеки України від 15 березня 2016 року №96/2016 і Рамковою програмою ЄС з досліджень та інновацій «Горизонт 2020» та науково-дослідними роботами, що проводилися в Національному авіаційному університеті та ДП «Антонов».

Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації. Наукові положення, висновки і рекомендації, сформульовані у дисертаційній роботі Самойлика Є.О. «Лексикографічні методи захисту мовної інформації», є обґрунтованими. Робота характеризується глибоким аналізом сучасних методів криптографічного захисту текстових даних; методів передавання інформації; лексикографічних методів захисту інформації; використанням теорії випадкових процесів, теорії телетрафіка для вирішення поставлених в дисертаційному дослідженні задач; коректність виконаних у дисертаційній роботі теоретичних досліджень не викликає сумніву, оскільки вони базуються на фундаментальних наукових теоріях.

Оцінка змісту та структури дисертації. Дисертаційна робота складається із вступу, чотирьох розділів, висновків і додатків, загальним обсягом у 157 сторінок.

У **вступі** автором обґрунтована актуальність проблеми, означені мета, задачі, об’єкт та предмет досліджень, визначена наукова новизна та практична значимість результатів роботи, наведені дані про їх впровадження.

У **першому розділі** автором проведено дослідження шляхів підвищення стійкості сучасних систем захисту мовної інформації. Автором виокремлено напрямки подальших наукових досліджень, зокрема проведено визначення характеристик лексикографічних систем з позицій технічного захисту текстової інформації. Автором запропоновано основоположну ідею, яка спрямована на досягнення мети даного дослідження, що базується на використанні крипто-семантичного підходу до створення стійкої системи захисту текстової інформації. Автором запропоновано використовувати основні результати теорії лексикографічних систем та методів побудови відповідних лінгвістичних корпусів. Проаналізовано нові введені автором поняття, які використані в дисертації. Проведено змістовний аналіз поняття «Інформація», її видів, методів інтерпретації. Після цього запропоновано інформаційну модель суб'єкта, що призначена для використання в завданнях криптосемантики, а також інформаційну модель інтелекту. При розгляді представленої моделі розглянуто питання про існування інтегрального інтелекту, тобто про

реальність існування в просторі деякого суб'єкта, інтелект якого включає інтелекти усіх раніше існуючих, нині існуючих і створених в майбутньому суб'єктів матеріального світу. На основі вищенаведеного автором зроблені висновки про те, що розумова діяльність суб’єкта заснована на використанні первинної інформації, яка містить відомості про форму відображення у пам’яті суб’єкта структури співвідношень між характеристиками властивостей об’єктів середовища існування цього суб’єкту.

У дисертаційній роботі досліджуються найбільш важливі з точки зору сфер застосування два види текстової інформації – мовна усна та мовна письмова. Перший вид відповідає голосовій інформації, що сприймається інтелектом через слуховий апарат, а другий вид сприймається через зір. В обох випадках автором розглядається первинна інформація, що закодована з використанням правил побудови будь-якої із відомих природних або штучних мов, що зрозумілі суб’єкту. При цьому, у результаті інтелектуальної обробки цих зразків може виявитися їхній смисл, тобто, утворитися смислові образи.

Таким чином, розглянута в дисертаційному дослідженні семантична криптографія заснована на використанні такої властивості мовної інформації, як правдоподібність. Для формального визначення показників цієї властивості автору необхідно було розробити модель розуміння мови, яка була б придатна для синтезу крипто-семантичних методів захисту смислу текстової інформації. Це завдання також було вирішено в першому розділі дисертаційного дослідження.

Таким чином, за результатами вищенаведеного аналізу та за допомогою розроблених моделей у кінці розділу коректно здійснена постановка задачі дослідження, що спрямована на розробку та удосконалення лексикографічних методів захисту мовної інформації. Крім того, можна стверджувати, що представлені в першому розділі результати можна розглядати як базові відомості, необхідні для розуміння роботи крипто-семантичних систем захисту смислового змісту інформації, застосування яких є можливим на практиці у ряді областей.

У **другому розділі** розроблено метод побудови семантичних тезаурусів для лексикографічних криптосистем. Спочатку показано, що невід’ємним елементом будь-якої практично стійкої криптосистеми, яка заснована на застосуванні певним чином побудованої лексикографічної системи, є тезаурус бази захисту інформації у прикладній системі, де ця криптосистема використовується. Після цього автором розроблено формальну модель

тезауруса бази захисту. Для неї у загальному випадку доцільно задати ієрархію її семантичних одиниць, що наведена у другому розділі. Тезаурус TZ_{PS} містить усі можливі семантичні образі на усіх заданих рівнях абстрагування, що у сукупності складають мову відображення цієї прикладної області. Так що, семантичні відношення між семантичними образами будь-якого одного рівня абстрагування можуть бути семантично відмінні, правдоподібні або фальшиво-подібні.

Тому для побудови будь-якої крипто-семантичної системи захисту мовної інформації, було надано формальні визначення смисловим відношенням між семантичними одиницями: смислова ідентичність, смислова відмінність, сумнівна смислова ідентичність (або правдоподібність), сумнівна смислова відмінність (або фальшиво-подібність).

Також у другому розділі дисертаційного дослідження для захисту смислу текстових повідомлень було запропоновано методологічний підхід, що ґрунтується на використанні відношень смислової правдоподібності. Таким чином, за рекурентною схемою можливо записати вирази щодо відношень між смисловими одиницями для будь-яких інших більш деталізованих рівнів абстрагування представлення смислових образів. Також вагомим результатом є те, що автор представив формальні відображення семантики текстової інформації, що дозволяє автоматизувати процес створення семантичних словників.

Третій розділ дисертації присвячений розробці методу побудови лексикографічної криптосистеми.

Автором розглянуто практично стійку криптосистему захисту текстової інформації, яка поміщена у задану табличну форму за умови, що ця текстова інформація береться із семантичного словника наперед визначеної прикладної області. Показано, що перш ніж здійснювати шифрування, використовуючи будь-який відомий досконало стійкий шифр, необхідно створити семантичний словник, лінгвістичні одиниці якого відображають змістовний простір цієї прикладної області. Тому запропоновано механізм збільшення відстані єдиності у рамках цієї стійкої криптографічної системи шляхом синтезу штучної мови відображення прикладної області, де ця система використовується. Також автором визначено умови забезпечення режиму досконалої стійкості у криптосистемах із збільшеною відстанню єдиності.

Ефективність методу захисту текстової інформації з укрупненням алфавіту мови відображення цієї інформації було порівняно із ефективністю

шифру одноразових блокнотів. Показано, що при застосуванні стійкої криптосистеми з укрупненим алфавітом довжина ключа шифру може бути у *46,2* рази менша за довжину ключа шифру при застосуванні шифру одноразових блокнотів.

У **четвертому розділі** розроблено метод побудови крипто-семантичної системи захисту мовної інформації, заснованої на використанні прикладних лексикографічних систем захисту текстової інформації, зокрема тезаурусів смислових образів, що забезпечує незалежність обсягу текстової інформації, що підлягає шифруванню, від значень відстані єдиності та довжини ключів шифрів. У цьому випадку відпадає необхідність дотримання умови неперевищення відстані єдиності і відкривається можливість вибору довжини ключів шифру незалежно від обсягів інформації, що потребують шифрування.

Спочатку проведено формальне обґрунтування можливості створення досконало стійких незалежних від довжини ключу шифру криптосистем. Потім виконано синтез крипто-семантичного методу побудови системи захисту. Цей метод забезпечує захист смислу інформації, що передається, оброблюється та зберігається засобами інформаційної системи, від порушень конфіденційності на рівні надання абсолютних гарантій неможливості злому системи захисту як з теоретичної (формальної), так і з практичної точок зору. При цьому обсяг інформації, що підлягає захисту, за будь-якої довжини ключа шифру, що використовується, має не обмежуватися довжиною цього ключа, а довжина ключа обиратися в залежності від припустимого значення ймовірності прийняття безпомилкових рішень в процесі дешифрування смислового потоку.

Для виконання поставленого завдання використано крипто-семантичний підхід щодо функціонування інтелекту. Таким чином, запропонований метод захисту передбачає створення системи захисту у два етапи: спочатку створюється семантичний словник (тезаурус) прикладної області**,** а потім розроблюються програмно-технічні засоби реалізації операторів шифрування/розшифрування, у складі яких використовується створений тезаурус.

Найбільш поширеного застосування базовий варіант даного методу матиме у системах зберігання та передавання текстових повідомлень через відкриті канали зв’язку, що незахищені від перехоплення інформації, у випадках, коли порушення конфіденційності цих повідомлень може призвести до неприйнятних негативних наслідків для її власників. Використання цього варіанту реалізації крипто-семантичного методу може виявитися

безальтернативним технічним рішенням у прикладних задачах, де необхідно забезпечити гарантованість захисту текстових повідомлень в умовах, коли відсутня довіра до будь-яких структур та суб'єктів.

Достовірність і новизна отриманих результатів, наукових положень, висновків та рекомендацій. Висновки та результати дисертації викладені змістовно, в логічній послідовності, у відповідності зі структурою задач, поставлених і вирішених у дисертаційній роботі. Достовірність отриманих результатів, наукових положень підтверджена результатами статистичного моделювання на ЕОМ, коректністю теоретичних і математичних викладок.

Наукова новизна дисертаційної роботи Самойлика Є.О. полягає в наступному:

1. Вперше розроблено модель крипто-семантичного словника, що дозволяє визначити тезаурус бази захисту інформації у прикладній системі та семантичну структуру словників прикладної області.

2. Вперше розроблено метод побудови лексикографічної криптосистеми, який дозволяє збільшити довжину шифрованих повідомлень відносно довжини ключової інформації та відповідно зменшити частоту зміни ключів шифру. При застосуванні стійкої криптосистеми з укрупненим алфавітом довжина ключа шифру може бути у *46,2* рази менша за довжину ключа при застосуванні шифру одноразових блокнотів.

3. Вперше запропоновано лексикографічний метод захисту текстової інформації, що за рахунок випадкової заміни первинного смислового образу повідомлення на інший правдоподібний елемент, узятий із семантичного тезаурусу бази захисту прикладної області, дозволяє забезпечити підвищену стійкість захисту.

Практична цінність роботи.

Отримані в дисертаційної роботі результати дозволяють підвищити ефективність захисту мовної інформації, яка призначена для передачі по відкритим каналам зв'язку.

Практичні результати дисертаційної роботи:

- розроблено методику автоматизації створення тезаурусу бази захисту інформації у прикладній області;
- запропоновано схему технічної реалізації методу побудови криптосистеми із збільшеною відстанню єдиності за ключем шифру для стійкої системи передавання текстових даних, представлених у вигляді табличних форм;

– розроблено алгоритмічне та програмне забезпечення крипто-семантичного захисту текстових даних.

Результати дисертаційної роботи упроваджено у виробничий процес ДП «Антонов» та у навчальний процес Національного авіаційного університету.

Повнота викладу основних результатів та висновків в опублікованих працях. За матеріалами дисертації опубліковано 13 наукових праць, що відображають положення дисертації, з них 5 статей у фахових виданнях, 1 розділ колективної монографії, 1 праця, яка включена до науково-метричної бази Scopus (у закордонному періодичному виданні ЄС), 6 матеріалів доповідей на міжнародних науково-технічних конференціях.

Кількість та якість наукових робіт здобувача з теми дисертації відповідають вимогам до дисертацій на здобуття наукового ступеня кандидата технічних наук.

Ідентичність змісту автореферату й основних положень дисертації. Зміст автореферату повністю відповідає змісту дисертаційної роботи та відображає основні положення, що виносяться на захист.

Відповідність дисертаційної роботи спеціальності. Дисертаційна робота Самойлика Є.О. за змістом і отриманими результатами повністю відповідає паспорту спеціальності 05.13.21 – Системи захисту інформації.

Зауваження до дисертаційної роботи.

1. У першому розділі дисертаційної роботи можна було би виділити один підрозділ та присвятити його змістовному критичному аналізу сучасних криптографічних алгоритмів, які широко використовуються в телекомунікаційних системах. Це б дозволило здійснити більш ґрунтовну постановку завдань, а також окреслити коло тих алгоритмів захисту мовної інформації, з якими необхідно порівнювати характеристики розробленого рішення.

2. Абстрактна модель розуміння текстової інформації (рис. 1.3) представляє доволі спрощений процес. В моделі є відсутніми опис можливих результатів сприйняття смислу досліджуваного зразка текстового повідомлення, альтернативні можливості по відображенню форми досліджуваного зразка текстового повідомлення на основі його безпосереднього сприйняття суб’єктом. Конкретизація вище наведених пунктів могла би відкрити шлях для створення більш детальної моделі, що могла би бути використана для вирішення задач в галузі захисту інформації.

3. У другому розділі розроблено ієрархічну структуру тезаурусу (рис. 2.2), який можна використовувати для створення крипто-семантичної системи захисту мовної інформації, проте не наведено жодного конкретного прикладу створення дерева тезаурусу для певної прикладної області. Це дещо ускладнює розуміння запропонованої автором структури.

4. В роботі не розглянуто саме процес автоматизації створення тезаурусів, а лише стверджується, що розроблена модель надає можливість автоматизувати процес створення тезаурусів прикладної області застосування. Доцільно було б вказати, за рахунок яких засобів можливо реалізувати автоматичну побудову тезаурусів для прикладної області, а також навести приклад подібної реалізації.

5. На рис. 3.5 наведено архітектуру лексикографічної системи захисту мовної інформації, що базується на синхронізації генераторів псевдовипадкових послідовностей, які розташовані на передавальній та приймальній сторонах каналу секретного обміну інформацією, за допомогою відомого обом сторонам секретного ключа шифру. Але залишається відкритим питання щодо передачі цього секретного ключа.

6. У дисертаційній роботі відсутній аналіз генераторів псевдовипадкових послідовностей, що можуть бути використані за своїми характеристиками у складі лексикографічних систем захисту мовної інформації.

7. За результатами аналізу графіку залежності виграшу в ефективності застосування криптосистеми з укрупненим алфавітом від довжини повідомлення було зроблено висновок про те, що виграш у ефективності застосування криптосистеми з укрупненим алфавітом у порівнянні з криптосистемою, що реалізує шифр одноразових блокнотів, лінійно залежить від довжини повідомлення. Це звичайно підтверджує ефективність запропонованого рішення в порівнянні із шифром одноразових блокнотів, проте відсутні жодні інші порівняння із сучасними криптографічними алгоритмами, що утруднює прийняття однозначного рішення щодо можливостей використання розробленого метода на практиці.

8. В четвертому розділі дисертації наведено інформацію щодо розробленого програмного забезпечення для захисту мовної інформації на основі розроблених методів, проте відсутній детальний аналіз результатів роботи даного програмного забезпечення.

Загальні висновки

Викладене дозволяє зробити висновок, що дисертаційна робота Самойлика Є.О. «Лексикографічні методи захисту мовної інформації» є

завершеною працею, що виконана автором самостійно на високому науковому рівні. В роботі отримані нові науково обґрунтовані теоретичні та практичні результати, які є суттєвими для розвитку систем захисту мовної інформації.

Дисертація задовольняє вимогам «Порядку присудження наукових ступенів», які висуваються до кандидатської дисертації, як кваліфікаційної наукової праці, а її автор, Самойлик Євген Олександрович, заслуговує присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.21 – Системи захисту інформації.

Офіційний опонент,
директор навчально-наукового інституту
Кібербезпеки, комп'ютерних і радіо технологій
Одеської національної академії зв'язку ім. О. С. Попова,

доктор технічних наук, професор Є.В. Василіу

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ * ОДЕСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ ЗВ'ЯЗКУ ім. О.С. ПОПОВА * Код в ЄДРПОУ

ПІДПИС ЗАВІРЯЮ:
УЧЕНИЙ СЕКРЕТАР
ОНАЗ ім. О.С. Попова
Руда Г.В.