ВІДГУК

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

Актуальність теми дисертації. Життя сучасного суспільства неможливо уявити без використання сучасних інформаційних технологій. Активне впровадження сучасної комп'ютерної техніки та мережевих технологій у будь-яку сферу життя майже кожної людини призвело до того, що величезні обсяги різноманітної інформації в цифровій формі зберігаються в комп'ютерних системах і передаються відкритими каналами зв'язку. Серед всього обсягу інформаційних ресурсів є інформація, що потребує обмеження в доступі. Це може бути інформація, що містить комерційну таємницю, особисті або службові дані тощо. Природньо виникає потреба захистити таку інформацію від несанкціонованого доступу, крадіжки, знищення або інших несанкціонованих злочинних дій. Концентрація інформації в цифровій формі в телекомунікаційних та інформаційних системах примушує все більше приділяти увагу задачам її захисту. Питання безпеки і захисту інформації в телекомунікаційних мережах від несанкціонованого доступу є важливими та актуальними на сьогоднішній день.

Сучасний стан розвитку інформаційного простору характеризується новими потребами в створенні умов для безпечного функціонування його суб'єктів, коли особливо важливими стають проблеми протидії інформаційним війнам, захист власного кіберпростору, забезпечення конфіденційності даних тощо. Одним з специфічних напрямків інформаційної безпеки є захист мовної інформації, на що і спрямовано дисертаційне дослідження здобувача.

Варто також зазначити, що ніякі апаратні, програмні та будь-які інші рішення не зможуть гарантувати абсолютну надійність і безпеку даних у інформаційних системах. Крім того, велика концентрація захисних засобів в інформаційній системі може привести не лише до підвищення її вартості, а й до перевантаження та зниження її продуктивності. Тому вказані фактори необхідно обов'язково враховувати під час удосконалення існуючих і розроблення нових методів захисту інформації.

Згадані вище проблеми і задачі обумовлюють актуальність цієї дисертаційної роботи.

Дисертаційна робота Самойлика Є.О. пов'язана з актуальними науково-практичними розробками, що реалізуються у рамках «Концепції національної інформаційної політики» і «Основними науковими напрямами та найважливішими проблемами фундаментальних досліджень у галузі природничих, технічних і гуманітарних наук НАН України на 2014-2018 роки», з міжнародними програмами, зокрема Horizon 2020 (SU-DS05-2018-2019: Digital security, privacy, data protection and accountability in critical sectors, SU-FCT03-2018-2019-2020: Information and data stream management to fight against (cyber) crime and terrorism), а також у науково-дослідних роботах, що проводяться в Національному авіаційному університеті та інших закладах вищої освіти та наукових установах.

Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації. Дисертаційна робота Самойлика Є.О. є кваліфікаційною науковою працею, написаною ним власноручно.

Обсяг дисертаційної роботи, що містить вступ, чотири розділи, висновки і додатки, становить 157 сторінок.

У вступі наведено актуальність, мету та завдання дослідження, наукову новизну, практичне значення одержаних результатів, відомості про впровадження, зв'язок роботи з науковими програмами, планами, темами.

У першому розділі для визначення задач, які необхідно вирішити, автор послідовно, системно надав узагальнену характеристику інформаційної моделі суб'єкта, що придатна для вирішення завдань семантичної криптографії. Модель відбиває основну життєву функцію суб'єкта – адекватно реагувати на інформацію, що поступає із середовища його існування. Згідно цієї моделі потоки інформації, що сприймаються чутливими елементами сенсорних систем суб'єкта, подаються в реальному часі на обробку засобами інтелекту.

На базі викладеного матеріалу розроблено формальну модель розуміння мовної інформації, відповідно до якої у процесі розумової діяльності суб'єкт генерує детерміновані скінчені дискретні часові ряди смислових образів, що вибираються ним із доступного йому тезаурусу. Чітко визначено поняття тезаурусу інтелекту суб'єкта та показано, що якість тезаурусу визначається як розмірністю бази смислових образів, якими здатен оперувати суб'єкт у процесі розумової діяльності, так і кількістю рівнів абстрактності відображення смислових образів у структурі тезаурусу. Ці твердження покладено в основи подальших наукових досліджень автора.

Формальну модель розуміння мовної інформації можна було б розглядати як одино із пунктів наукової новизни, проте автором цього зроблено не було.

У другому розділі на базі представлених у першому розділі результатів досліджень розроблено формальну структуру тезаурусу смислових образів. Показано, що такий тезаурус має ієрархічну структуру і за ступенем абстрагування відображення смислових образів розподіляється на i рівнів, де $i = 1, 2, \ldots, I$ – кількість рівнів абстрагування відображення смислових образів, якими оперує колективний інтелект носіїв мови взагалі або індивідуальний інтелект суб'єкту розумової діяльності окремо, а I – максимальна кількість рівнів абстрагування відображення SO, що є доступною інтелекту. Показано, що простір смислових образів, доступний суб'єкту (або групі суб'єктів), є дискретним, кінцево-мірним, який має коренево-подібну структуру. Показано також, що структуру тезаурусу в загальному випадку можливо представити у вигляді рекурентної схеми (вираз (2.1)). Автором показано, яку ієрархію смислових одиниць, що відображають логічно завершені думки певного рівня абстрагування доцільно використовувати у різних прикладних сферах застосування.

Для побудови будь-якої практично стійкої лексикографічної системи захисту текстової інформації запропоновано формальні визначення таких смислових відношень між семантичними одиницями, як смислова ідентичність, смислова відмінність, сумнівна смислова ідентичність (або правдоподібність), сумнівна смислова відмінність (або фальшиво-подібність). На основі використання відношень смислової правдоподібності запропоновано дієвий підхід до захисту смислу текстових повідомлень.

Представлено формальні відображення семантики текстової інформації (вирази (2.4) – (2.7)), що дозволили автоматизувати процес створення семантичних словників.

Таким чином, отримані у другому розділі результати можна підсумувати як наступний науковий результат: метод побудови семантичних тезаурусів для лексикографічних криптосистем**.**

Третій розділ дисертаційного дослідження присвячено розробці методу побудови стійкої криптосистеми захисту від порушень конфіденційності текстової інформації. Розроблений метод базується на застосуванні механізму укрупнення алфавіту мови відображення текстової інформації, що використовує словник прикладної області і враховує структуру табличної форми.

Під час проведення досліджень отримано математичні вирази та побудовано відповідні графіки, що визначають залежності відстані єдиності та ентропії ключа шифру від довжини повідомлення. Отримані результати доцільно використовувати для обчислення максимально можливої кількості сеансів зв'язку без зміни ключа шифру. Отримано математичний вираз для

визначення показника виграшу в довжині ключа (за інших рівних умов) у разі застосування в криптосистемі укрупненого алфавіту в порівнянні з методом одноразових блокнотів. На основі цього математичного апарату проведено аналіз ефективності запропонованого методу. У наведеному прикладі показано майже 50-ти кратний виграш за цим критерієм ефективності.

Крім того, в третьому розділі розглянуто питання щодо побудови ГПВП, що можна використати в запропонованому автором методі. Визначено спосіб використання цього ГПВП у лексикографічній системі захисту мовної інформації.

У третьому розділі також визначено стійкість запропонованої криптосистеми із збільшеною відстанню єдиності за різних ресурсних можливостей порушника. Показано, що за умов, коли хоча б один зразок вихідного (відкритого) текстового повідомлення та зразок кореспондованої з ним шифрограми є відомими, стійкість синтезованої криптосистеми співпадає із стійкістю алгоритму шифроперетворень, що використовується. Отримано математичні вирази та надано графік залежності показника стійкості від розрядності ключа шифру. Продемонстровано, що з ростом довжини ключа шифру стійкість системи захисту зростає. Зі збільшенням основи алфавіту мови ключової інформації значно збільшується показник стійкості криптосистеми за сталого значення розрядності ключа шифру.

До результатів, які мають наукову новизну і представлені в третьому розділі, можна віднести безпосередньо метод побудови лексикографічної криптосистеми.

У четвертому розділі розглянуто вирішення завдання з розробки методу побудови стійкої криптосистеми на основі використання прикладних лексикографічних систем захисту текстової інформації, зокрема тезаурусів смислових образів (синтез структури яких виконано в другому розділі), що забезпечує незалежність обсягу текстової інформації, що підлягає шифруванню, від значень відстані єдиності та довжини ключів шифрів.

Згідно цього методу простір заданої прикладної області застосування *i*-ої інформаційної системи відображається на простір смислових образів цієї системи з урахуванням семантичних співвідношень між ними, що визначаються заданим простором обмежувальних умов. Синтезується структура семантичного тезаурусу заданої прикладної області. У структурі тезаурусу кожному смисловому образу із простору смислових образів ставиться у відповідність множина інших смислових образів із цього ж простору, що знаходяться із цим смисловим образом у відношенні смислової правдоподібності. Таким чином, розроблений метод 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.

Показано, що в деяких випадках довжина ключа шифру для крипто-семантичного шифрування має обиратися в залежності від припустимого значення ймовірності прийняття безпомилкових рішень у процесі дешифрування смислового потоку. Надано відповідний вираз для вибору мінімально можливого значення довжини ключа шифру. Після цього проведено ймовірнісний аналіз стійкості захисту смислу текстових повідомлень, що забезпечує лексикографічна криптосистема, яку побудовано згідно запропонованого методу.

У ході проведених досліджень синтезовано алгоритм реалізації запропонованого методу захисту смислу текстової інформації, а на його базі розроблено програмне забезпечення, що здійснює шифрування та розшифрування мовної інформації, що передається відкритим каналом в прикладній області застосування «Радіообмін диспетчер-пілот». Для цього автором розроблено структуру тезаурусу радіообміну. Результати тестування додатка показали, що його можна використовувати на практиці. Покращити ефективність його роботи можна шляхом використання більш досконалих бібліотек для розпізнавання мовної інформації.

Отже, основним наукових результатом, отриманим у четвертому розділі, є метод побудови крипто-семантичної системи захисту мовної інформації.

Достовірність і новизна отриманих результатів, наукових положень, висновків та рекомендацій. Результати дисертаційної роботи викладено послідовно, систематично, а також відповідають поставленим задачам. Достовірність наведених результатів підтверджується збігом теоретичних розрахунків з результатами експериментальних досліджень, а також коректних застосуванням математичного апарату теорії ймовірностей, математичної статистики тощо.

Наукова новизна дисертаційної роботи Самойлика Є.О. наступна:

1. Вперше розроблено модель крипто-семантичного словника, яка за рахунок уведення в прикладну лексикографічну систему показників семантичних зв'язків між смисловими конструкціями мови відображення області застосування дозволяє визначити тезаурус бази захисту інформації у прикладній системі та семантичну структуру словників прикладної області.

2. Вперше розроблено метод побудови лексикографічної криптосистеми, який за рахунок розширення базового алфавіту лінгвістичної формальної системи джерела мовних повідомлень забезпечує збільшення відстані єдиності шифру за ключем та дозволяє збільшити довжину

шифрованих повідомлень відносно довжини ключової інформації та відповідно зменшити частоту зміни ключів шифру.

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 при відсутності будь-яких обмежень на обсяг мовної інформації, що підлягає шифруванню, і, тим самим, усуває необхідність у періодичній зміні ключової інформації.

Практична цінність роботи.

1. Розроблено методику автоматизації розробки тезаурусу бази захисту інформації у прикладній області під час побудови лексикографічної криптосистеми.

2. 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.

3. Розроблено програмне забезпечення крипто-семантичного захисту текстових даних, що засноване на використанні прикладного тезаурусу смислових образів, який здатний забезпечити режим підвищеної стійкості у рамках конкретно визначених прикладних систем

Повнота викладу основних результатів та висновків в опублікованих працях. За матеріалами дисертації опубліковано 13 робіт, які розкривають зміст дисертаційної роботи. Серед них: розділ колективної монографії, фахові статті, статті у періодичному закордонному виданні ЄС, тези доповідей на міжнародних науково-технічних конференціях.

Ідентичність змісту автореферату й основних положень дисертації. Розбіжності між змістом автореферату та змістом представленої дисертаційної роботи відсутні.

Відповідність дисертаційної роботи спеціальності. Дисертаційна робота Самойлика Є.О. відповідає паспорту спеціальності 05.13.21 – Системи захисту інформації.

Зауваження до роботи.

1. З тексту дисертації та, відповідно, автореферату не зовсім зрозуміло, якими правилами необхідно керуватися під час вибору кількості рівнів абстрагування та кількості смислових варіацій на кожному рівні для забезпечення повного відображення предметної області, а й, відповідно, забезпечення більшого рівня секретності переданих повідомлень.

2. Для процесу автоматизації процесу створення тезаурусів прикладної області застосування для розроблених методів захисту мовної інформації доцільно було б використати підходи штучного інтелекту. Крім того, не пояснено, яким саме чином подолано проблему росту тезаурусу прикладної області в залежності від збільшення обсягу текстової інформації, а також яким саме чином опрацьовуються поліморфізми, особливо за відсутності аналізу контекстуального зв'язку.

3. У роботі «смисловий потік розглядається як дискретна невипадкова послідовність смислових образів» (стор. 6 автореферату) і стверджується, що це детермінована послідовність. При цьому не пояснюється зв'язок в цій послідовності. Автором також не розглянуто питання семантичної синтагматики та контекстуального зв'язку, без чого неможливо проводити семасіологічні дослідження тексту та зв'язків ядра семантики з вказівником. Не пояснено, які методи текстуального та інтертекстуального дослідження використано та як саме.

4. У п. 3.3.1 третього розділу стверджується, що для забезпечення випадковості характеру підмін необхідно, щоб період генератора псевдовипадкових чисел T був значно більшим за довжину таблиці n, проте відсутні більш конкретні рекомендації, що могли б бути використані розробниками програмного забезпечення.

5. На рисунку 4.2 представлено базовий варіант реалізації методу побудови лексикографічної криптосистеми, який містить кроки зі створення словника («Статистичний аналіз предметної області» та «Семантичний аналіз предметної області»), проте формалізований математичний опис цих кроків відсутній. Крім того, семантична складова текстової інформації в роботі розглядається як щось абстрактне та не пов'язане з морфологічно-синтаксичною складовою, адже саме на основі цієї складової можливо реалізувати дослідження семантики та уникнути порушення текстуального зв'язку. На стор. 6-7 автореферату продемонстровано ж, що ієрархія смислових одиниць складається зі слова, фрази чи ситуації, і при цьому відсутній морфологічно-синтаксичний аналіз, а тому постає питання, яким саме чином отримуються вхідні дані для семантичного аналізу.

6. У запропонованих автором методах відсутня прив'язка до конкретної мови чи, хоча б, групи мов. У дисертації розглядається обробка як української, так і англійської мов (джерела 26 та 41 дисертації). Для мов різних мовних сімей і груп буде відрізнятися логіка побудови логічно несуперечливої смислової конструкції. Не розглянуто, на чому базується та як детермінується послідовність смислового потоку. Отже, надано лише загальну модель та не визначено закони, за якими буде змінюватись послідовність залежно від мови.

Не зовсім коректно використовувати «мову людського спілкування» (стор. 84 дисертації).

7. З роботи зрозуміло, що корпусний метод є основним для створення електронного словника, тобто застосовується методика аналізу текстів. Однак з роботи неясно, яким чином створювався корпус текстів. Звідси витікає ще одна проблема багатомовної направленості методу.

8. У тексті дисертаційної роботи відсутні результати перевірки розробленого криптографічного алгоритму за допомогою спеціальних тестів, що не дає змогу оцінити його криптографічну стійкість.

9. У роботі, на жаль, відсутні результати аналізу апаратних вимог до обладнання, на якому можуть бути розроблені автором алгоритми, аналізу швидкодії алгоритму тощо. Це не дає змоги чітко окреслити всі потенційні області застосування.

Загальні висновки

Дисертаційна робота «Лексикографічні методи захисту мовної інформації» є завершеною науковою працею, в якій отримано нові науково обґрунтовані результати, що в сукупності вирішують задачу підвищення ефективності захисту мовної інформації під час її передавання відкритими каналами зв'язку. Дисертація відповідає вимогам «Порядку присудження наукових ступенів», затвердженого постановою Кабінету міністрів № 567 від 24 липня 2013 року (зі змінами), а її автор – Самойлик Євген Олександрович – заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.21 – Системи захисту інформації.

Офіційний опонент:
доктор технічних наук, професор
проректор з науково-дослідної роботи
та міжнародних зв'язків
Черкаського державного
технологічного університету Е.В. Фауре

«16» 11 2020 р.

Підпис офіційного опонента
Фауре Еміля Віталійовича
засвідчую:

Учений секретар
Черкаського державного
технологічного університету І.В. Миронець
к.т.н., доцент

«16» 11 2020 р.