DOI https://doi.org/10.15407/usim.2020.01.035
УДК 003.26:004.056.55

Р.С. ОДАРЧЕНКО, доктор технічних наук, старший науковий співробітник, Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, просп. Академіка Глушкова, 40, Київ, 03187, Україна, odarchenko.r.s@ukr.net

Є.О. САМОЙЛИК, пошукач, факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, просп. Любомира Гузара, 1, Київ, 03058, Україна, sea110913@gmail.com

В.М. СІМАХІН, аспірант, м.н.с., Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, просп. Академіка Глушкова, 40, Київ, 03187, Україна, sima@irtc.org.ua

В.О. БОРОВИК, аспірант, м.н.с., Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, просп. Академіка Глушкова, 40, Київ, 03187, Україна, dep185@irtc.org.ua

Р.М. ТИМЧИШИН, аспірант, м.н.с., Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, просп. Академіка Глушкова, 40, Київ, 03187, Україна, dep185@irtc.org.ua

КРИПТОСЕМАНТИЧНА СИСТЕМА ЗАХИСТУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Запропоновано метод побудови абсолютно стійкої криптосистеми захисту текстової інформації. Метод базується на застосуванні механізму укрупнення абетки мови, в результаті якого збільшується відстань єдиності. Система захисту складається з двох частин: семантичного тезаурусу, розробленого для конкретної прикладної області, та засобів реалізації операторів шифрування/розшифрування, що використовують створений тезаурус.

Ключові слова: *технічний захист інформації, абсолютно стійкі симетричні криптосистеми, лексикографічні системи, відстань єдиності, семантичний тезаурус.*

Вступ

Проблема захисту інформації на цей час є ключовою в багатьох застосуваннях. За деяких обставин повідомлення, які необхідно передавати, містять чутливу або секретну інформацію, яку не повинна перехопити та розпізнати третя сторона. Криптографія та системи захисту інформації постійно розвиваються, вдосконалюючи наявні алгоритми та розроблюючи нові. Стійким криптографічним алгоритмом вважаться такий алгоритм, який для розшифрування закодованої інформації вимагає необмеженої кількості обчислювальних ресурсів або часу розшифрування, за який інформація стане неактуальною. Абсолютна стійкість симетричних криптосистем, як показав К. Шеннон [1], забезпечується лише за умови, коли так

звана відстань єдиності за ключем не перевищується під час шифрування. Тобто необхідно, щоб поточна сумарна довжина зашифрованих текстових повідомлень в процесі шифрування не перевищувала довжину ключа шифру, інакше виникає теоретична можливість розкриття ключа шифру.

Аналіз останніх досліджень і публікацій

Багато досліджень [2–6] вказують на відносно невеликі значення відстані єдиності під час шифрування повідомлень, складених із символів абетки будь-якої із природних мов. Це призводить до необхідності часто змінювати інформацію про ключі шифрування, що може стати проблемою для багатьох застосувань. Крім цього, необхідно забезпечити випадковість й однакову ймовірність вибору варіантів реалізації ключа. Тому актуальним завданням є створення методів побудови абсолютно стійких криптосистем, які забезпечують більші значення відстані єдиності або, в найкращому разі, забезпечують формальну можливість шифрування якнайбільших обсягів текстових повідомлень, зокрема голосових, незалежно від значень відстані єдиності та довжини ключів шифрів.

Постановка завдання

Метою цього дослідження є позбування необхідності періодичної зміни ключів криптографічних шифрів у задачах абсолютно стійкого захисту текстової інформації. Основними завданнями дослідження є аналіз результатів теорії лексикографічних систем і розробка методу побудови абсолютно стійкої криптосистеми. За отриманими результатами аналізу теорії визначається можливість використання семантичних характеристик текстової інформації для створення абсолютно стійких криптосистем, що не ставлять суворих вимог до систем управління ключами шифрування. Запропонована абсолютно стійка криптосистема заснована на використанні прикладних лексикографічних систем захисту текстової інформації, зокрема тезаурусів смислових образів [7–8], які забезпечують незалежність обсягу текстової інформації, що підлягає шифруванню, від значень відстані єдиності та довжини ключів шифрування.

Об'єктом дослідження є процеси технічної підтримки абсолютно стійкого захисту текстової інформації. Предметом дослідження є лексикографічні методи абсолютно стійкого захисту текстової інформації, спрямовані на уникнення необхідності періодичної зміни ключів шифрування.

Формальний опис методу побудови лексикографічної криптосистеми

Приймемо наступні позначення:

S — суб'єкт, що синтезує зразок вихідного голосового повідомлення D в процесі вирішення прикладних задач заданої області застосувань;

D — зразок необробленого вихідного відкритого (незашифрованого) текстового повідомлення, смислове значення якого потребує захисту ($S{:}D$, де «:» — знак візуального або голосового сприйняття);

S_F — оператор обробки вихідного зразка за правилами граматики мови (природної або штучної), прийнятий для відображення текстових повідомлень заданої сфери застосувань, з використанням засобів спеціально розробленого лінгвістичного корпусу;

F_D — зразок коректно сформованого за правилами граматики вихідного відкритого текстового повідомлення та додаткова інформація про структру цього повідомлення: прийня та система текстових одиниць, локалізація текстових одиниць у дискретній послідовності цих одиниць, результати маркування (розмітки) вихідного повідомлення за лінгвістичними характеристиками тощо;

S_C — оператор аналізу зразка відкоригованого текстового повідомлення на відповідність елементам семантичного тезаурусу прикладної області та визначення параметрів локалізації цього повідомлення у структурі тезауруса;

C_D — зразок відкритого текстового повідомлення, який безпосередньо відображає зміст вихідного текстового повідомлення та відповідає прийнятим граматичним правилам та семантичним обмеженням;

S_Z — оператор шифрування, який перетворює вихідний зразок відкритого текстового повідомлення на зашифрований, що має правдоподібний, але, найімовірніше, інший смисловий зміст (чим забезпечується неоднозначність у сприйнятті змісту зашифрованого повідомлення);

Z_D — зразок зашифрованого текстового повідомлення (синтезований з семантичних одиниць використаного тезауруса), смисловий зміст якого, ймовірно, відрізняється від початкового реального змісту вихідного відкритого текстового повідомлення;

S^0_Z — оператор розшифрування, що забезпечує перетворення неоднозначного зашифрованого повідомлення, наприклад, тексту, в однозначне істинне за змістом вихідне відкрите текстове повідомлення;

Z^0_D — зразок розшифрованого текстового повідомлення, який повністю співпадає з C_D, тобто $Z^0_D \leftrightarrow C_D$, де символ $\leftrightarrow$ означає збіжність форми і змісту повідомлень Z^0_D і C_D.

Метод побудови лексикографічної криптосистеми має відтворювати послідовність операцій з обробки текстових повідомлень, зображеної на рис. 1 у вигляді діаграми форм представлення оброблюваного зразка текстового повідомлення та операторів перетворення цих форм засобами системи захисту.

Конфіденційність текстових повідомлень відповідно до цього методу забезпечується у такий спосіб. Зразок незашифрованого вихідного текстового повідомлення *D* за допомогою засобів спеціально розробленого лінгвістичного корпусу обробляється відповідно до правил *SF* граматики природної або штучної мови, прийнятих для відображення мовних повідомлень заданої області прикладних застосувань.

Як результат, отримуємо коректно сформований за правилами граматики зразок F_D відкритого вихідного текстового повідомлення та додаткову інформацію про структуру цього повідомлення: прийнята система мовних одиниць, локалізація в дискретній послідовності цих одиниць, результати маркування (розмітки) вихідного повідомлення за лінгвістичними характеристиками тощо.

Далі виконується аналіз (оператор S_C) зразка F_D на відповідність елементам семантичного тезаурусу, спеціально розробленого для відображення змісту мовних повідомлень і семантичних зв'язків між ними у вигляді відповідних семантичних співвідношень для заданої прикладної області. Внаслідок цього отримуємо зразок вихідного повідомлення C_D, що без спотворень відображає істинний зміст вихідного текстового повідомлення та відповідає прийнятим граматичним правилам і семантичним обмеженням.

Після чого безпосередньо виконується шифрування з використанням будь-якого відомого симетричного шифру S_Z, заснованого на застосуванні генератора псевдовипадкових послідовностей (ГПВП) [6, 9]. Початковий стан ГПВП встановлюється згідно з паролем і далі шляхом гамування (тобто виконання операції *XOR* — підсумовування за модулем 2) елементи повідомлення C_D замінюються на правдоподібні елементи з тезауруса. Вибір методів заміни справжніх за змістом повідомлень на правдоподібні залежить від синтезованої структури тезауруса, зокрема від кількості врахованих рівнів абстрагування його семантичних одиниць. Як результат, отримується зразок зашифрованого текстового повідомлення Z_D, що має правдоподібний, але, найімовірніше, інший зміст (чим забезпечується неоднозначність у сприйнятті змісту зашифрованого повідомлення). Для розшифрування повідомлення Z_D застосовується оператор зворотного перетворення . Зокрема, за допомогою відомого пароля ГПВП встановлюється той самий початковий стан, який був під час шифрування, і далі повторно виконується операція *XOR,* тобто елементи неоднозначного за змістом зашифрованого повідомлення Z_D шляхом відповідного вибору з тезауруса замінюються на елементи повідомлення Z^0_D, що однозначно відображає істинний зміст вихідного мовного

повідомлення C_D і відповідає прийнятим граматичним правилам та семантичним обмеженням заданої предметної області.

Стисло суть методу можна визначити як включення в симетричну систему криптографічного захисту інформації засобів лексикографічної системи в такий спосіб, щоб у процесі шифрування забезпечувалася семантична неоднозначність зашифрованих зразків повідомлень. При цьому використовувані засоби лексикографічної системи мають бути здатні здійснювати граматичний і семантичний аналіз повідомлень у рамках заданої предметної області.

Базова схема реалізації крипто-семантичного методу побудови системи захисту

Розгляньмо основний варіант технічної реалізації крипто-семантичного методу забезпечення абсолютної конфіденційності змісту текстових повідомлень. Можливі й інші варіанти реалізації цього методу захисту [4, 10–13].

Вважається, що текстові повідомлення можуть подаватися будь-якою мовою спілкування (природною людською або формально визначеною штучною), для якої створено відповідну лексикографічну систему [14] (зокрема, так званий лінгвістичний корпус [15]), у складі якої розроблено прикладний семантичний тезаурус (тобто, у відповідний спосіб структурована система семантичних словників), який вповні охоплює предметну область використання цього методу захисту.

Базовий варіант цього методу матиме поширене застосування в системах зберігання та передавання текстових повідомлень через відкриті канали зв'язку, не захищені від перехоплення інформації, у тому разі, коли порушення конфіденційності змісту цих повідомлень може призвести до неприйнятних негативних наслідків для власників. Використання цього варіанту реалізації крипто-семантичного методу може виявитися безальтернативним технічним рішенням у прикладних задачах, де необхідно забезпечити абсолютну гарантованість захисту текстових повідомлень в умовах, коли немає довіри до будь-яких структур і суб'єктів.

У процесі розробки цього варіанту реалізації крипто-семантичного методу було визначено завдання: за допомогою попередньої лінгвістичної обробки зразків текстових повідомлень, смисловий зміст яких підлягає шифруванню, забезпечити можливість функціонування симетричної криптографічної системи захисту з будь-яким визначеним формально обґрунтованим рівнем стійкості до крипто-аналітичних атак, при цьому необхідно забезпечити абсолютну гарантію захисту змісту текстових повідомлень від порушень конфіденційності і під час зберігання їх у комп'ютерах, і під час передачі їх через незахищене середовище транспортування інформації. При цьому абсолютна гарантованість захисту має забезпечуватися і з теоретичного, і з практичного погляду, і не залежати від умови неперевищення відстані єдиності.

Поставлена задача вирішується так: в рамках формально встановлених обмежень обраної сфери прикладних застосувань розробляється лексикографічна система, реалізована у вигляді прикладного лінгвістичного корпусу, і далі, безпосередньо перед шифруванням текстових повідомлень, згідно з будь-яким обраним способом симетричної криптографії здійснюється семантична структуризація цієї інформації з використанням засобів побудованого лінгвістичного корпусу так, щоб зразки зашифрованих текстових повідомлень були семантично правдоподібними образами в контексті обраної предметної області, а будь-який результат застосування можливих способів криптоаналізу приводив до отримання семантично правдоподібних текстових повідомлень, які, однак, можуть відрізнятися за змістом.

Лінгвістичний корпус тут визначається як програмний засіб автоматичного розбиття потоку текстових повідомлень, смисловий зміст яких необхідно захистити, на «мікроконтексти» — фрагменти потоку, які групуються навколо лінгвістичних одиниць (зокрема, слів, фраз, сценаріїв тощо), є об'єктами тлумачення [14–16]. Під тезаурусом мови розуміється лек-

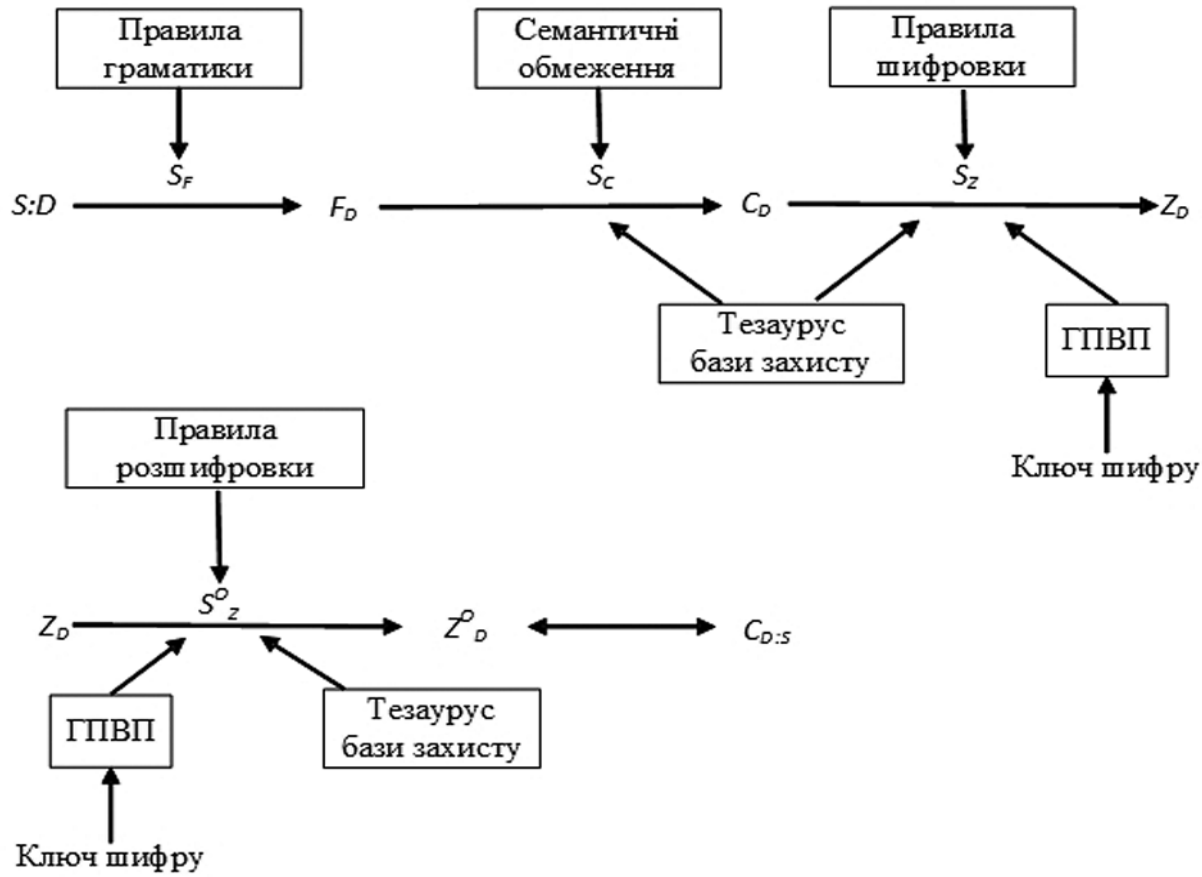

Рис. 1. Алгоритм реалізації лексикографічної криптосистеми

сика мови з певними семантичними зв'язками між лінгвістичними одиницями. У цьому разі тезаурус — це ієрархічна структура семантичних словників, яка визначає семантичні зв'язки між лінгвістичними одиницями мови, прийнятими для відображення змісту текстових повідомлень заданої сфери застосування.

Суть такого варіанту технічної реалізації крипто-семантичного методу захисту можна визначити як включення засобів лексикографічної системи в симетричну криптографічну систему так, щоб в процесі шифрування забезпечувалася семантична неоднозначність зашифрованих зразків текстових повідомлень. При цьому використовувані засоби лексикографічної системи повинні мати можливість здійснювати граматичний і семантичний аналіз вихідного потоку текстових повідомлень у рамках заданої предметної області.

Послідовність операцій з обробки текстових повідомлень, її умови та результати на кожному з технологічних етапів реалізації цього методу захисту зображено на рис. 1. Для реалізації оператора S_F, крім вихідних текстових повідомлень, необхідно використовувати дані про структуру потоку текстових повідомлень, їхні лінгвістичні характеристики, зокрема правила граматики мови, прийняті для відображення текстових повідомлень заданої сфери застосування. Засоби відтворення цих даних складають так званий лінгвістичний корпус, із використанням якого здійснюється маркування (розмітка) потоку текстових повідомлень за його лінгвістичними характеристиками [14]. За допомогою оператора S_F забезпечується коректність граматичної структури вихідних повідомлень, що створюються людиною або комп'ютерною програмою, згідно з правилами

та обмеженнями, що задаються засобами лінгвістичного корпусу. У разі невідповідності цим правилам голосові повідомлення повертаються їхнім авторам на доопрацювання.

Як вихідні дані, необхідні для реалізації оператора S_C, крім граматично коректних вихідних текстових повідомлень використовуються семантичні конструкції задіяного тезауруса мови. Оператор S_C забезпечує відповідність текстових повідомлень елементам тезауруса, що гарантує потенційну можливість шифрування змісту цих повідомлень. Оператор шифрування S_Z здійснює заміну вихідного (скоригованого операторами S_F і S_C) зразка відкритого текстового повідомлення на лінгвістичну конструкцію, елементи якої обрано з тезауруса у випадковий спосіб. Характер випадковості визначається властивостями генератора псевдовипадкових чисел, що має функціонувати в складі задіяної схеми шифрування.

На рис. 2 зображено блок-схему базової моделі системи, принцип дії якої відтворює крипто-семантичний метод захисту текстових повідомлень.

Така система містить усі головні елементи симетричної криптографічної системи та забезпечує захист змісту текстових повідомлень за умови синхронізації ГПВП, розташованих на передавальній і приймальній сторонах каналу секретного обміну інформацією. Розшифрування здійснюється з використанням ключа шифру та пароля. Окрім цього, у блок-схему моделі до складу лінгвістичного аналізатора текстових повідомлень додатково включено засоби лінгвістичного корпусу, тезауруса смислових образів предметної області, в рамках якої планується використовувати систему захисту, та програмного комплексу маркування (розмітки) вихідних потоків текстових повідомлень. Лінгвістичний аналізатор контролює відповідність вихідних зразків текстових повідомлень до граматичних і семантичних правил та обмежень, прийнятих у рамках заданої предметної області.

Тезаурус створюється за результатами статистичного та семантичного аналізів предметної області й має містити всі лінгвістичні одиниці, які потенційно можуть бути включені до складу будь-якого зразка (фрагмента повідомлення) потоку текстових повідомлень [8, 13, 17], має бути захищений від втрати конфіденційності. Система функціонує в такий спосіб. Користувач прикладної системи формує вихідний потік відкритих голосових повідомлень, зміст яких потребує захисту, та подає його на обробку засобами лінгвістичного корпусу. Засоби лінгвістичного корпусу здійснюють лексикографічну обробку потоку вихідних зразків голосових повідомлень у два етапи.

На *першому* етапі, здійснюється розмітка структури потоку голосових повідомлень за граматичними характеристиками й перевіряється його відповідність правилам граматики використовуваної мови. За невідповідності цим правилам зразки голосових повідомлень повертаються на доопрацювання. Далі, на *другому* етапі, потік голосових повідомлень, сформований відповідно до прийнятих граматичних правил, подається на семантичний аналізатор, який перевіряє відповідність лінгвістичних конструкцій цього потоку елементам тезауруса, що використовується. Тобто, здійснюється розмітка потоку за семантичними характеристикам. Для маркування зразків лінгвістичних конструкцій у потоці голосових повідомлень може бути використана довільна з відомих уніфікованих систем розмітки документів — *SGML, HTML, XML*, тощо [18]. Якщо лінгвістична конструкція містить елементи, не відображені в структурі тезауруса, то вона повертається на доопрацювання. В іншому разі робиться висновок, що оброблена лінгвістична конструкція вповні відображає зміст вихідного голосового повідомлення, відповідає прийнятим граматичним правилам і семантичним обмеженням та є придатною для шифрування її смислового змісту.

Для захисту обробленої лінгвістичної конструкції від порушень конфіденційності її змісту користувач (або відповідна комп'ютерна програма) повинен подати символьну послідовність, що відображає цю конструкцію, на вхід шифрувального пристрою (або програмного шифратора) та ввести ключову

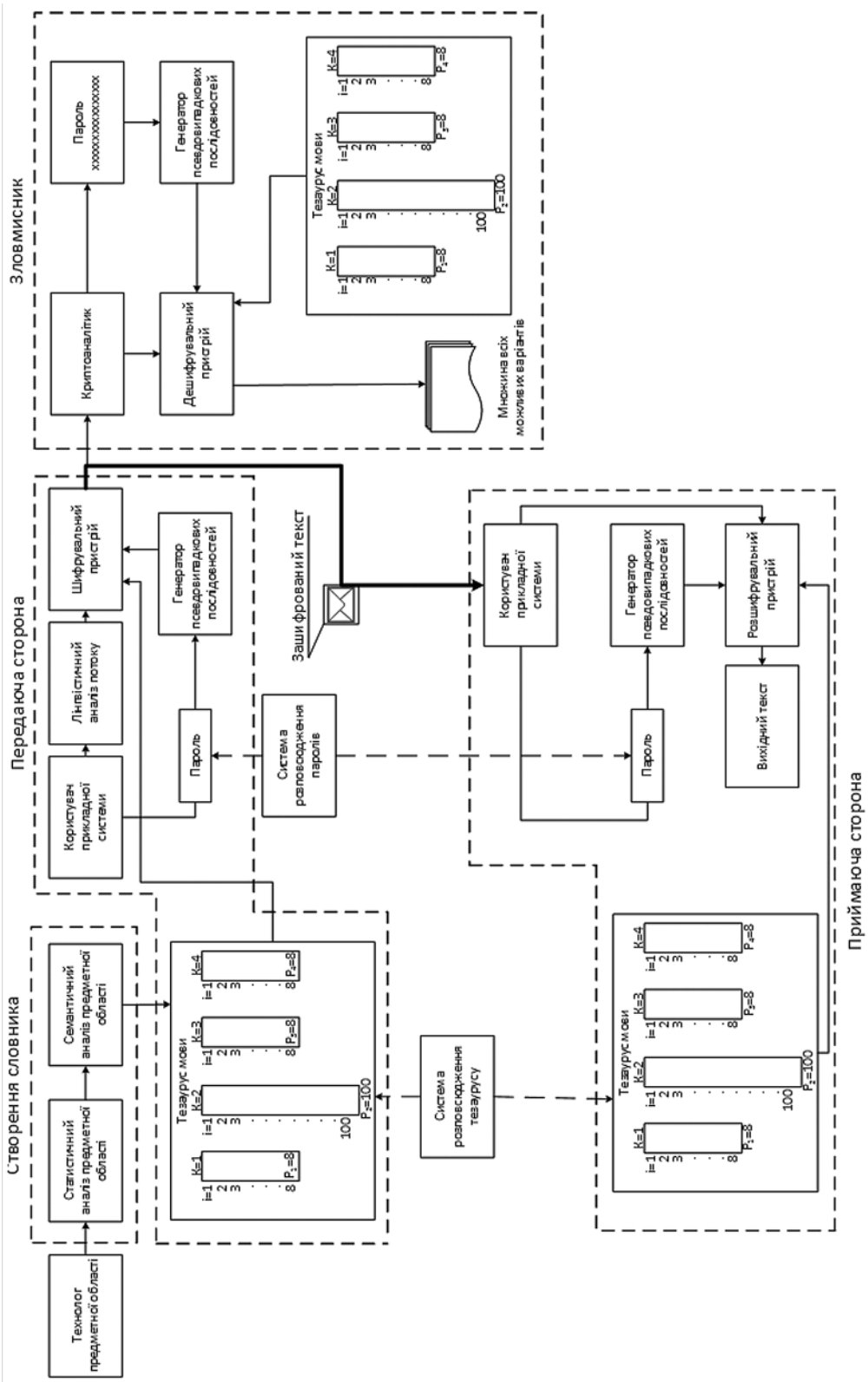

Рис. 2. Базовий варіант побудови лексикографічної криптосистеми

інформацію (пароль). За допомогою ключової інформації генератор псевдовипадкових послідовностей (ГПВП) ставиться у певний початковий стан. Шифратор обробляє символьні послідовності завдяки реалізації будь-якого відомого алгоритму шифрування, робота якого базується на використанні випадкових чисел, що генеруються ГПВП. Суть шифрування полягає у випадковій заміні лінгвістичних конструкцій, узятих із вихідного потоку голосових повідомлень на елементи задіяного тезауруса. Результатом роботи шифрувального пристрою є потік зашифрованих відображень голосових повідомлень, що має властивість семантичної неоднозначності. У разі нелегального перехоплення повідомлення, криптоаналітик не матиме можливості визначити істинний зміст зразків вихідних голосових повідомлень, оскільки навіть під час безпосереднього перебору ключів шифру в умовах повної поінформованості про задіяний лінгвістичний корпус і прийняту систему захисту, він щоразу отримуватиме правдоподібні зразки голосових повідомлень (складені з елементів тезауруса), які із заздалегідь заданою вірогідністю не відображають істинний сенс повідомлень. Навіть якщо криптоаналітик правильно набере парольну послідовність, він не зможе виявити сам факт злому шифру, оскільки не зможе відрізнити за семантичною ознакою вихідний зразок голосового повідомлення, який було зашифровано, від інших правдоподібних зразків голосових повідомлень.

Зашифрований потік відображень голосових повідомлень може зберігатися в запам'ятовуючих пристроях або бути переданий через будь-яке незахищене фізичне середовище, наприклад, через радіоканал зв'язку. У будь-якому разі зашифрований потік відображень голосових повідомлень може бути розшифрований на будь-якому комп'ютері, де встановлено відповідні засоби лінгвістичного корпусу та криптографічної системи захисту інформації. Щоб здійснити розшифрування, необхідно подати символьну послідовність, яка відображатиме зашифрований потік, на розшифрувальний пристрій (або відповідну комп'ютерну програму розшифрування) й за допомогою відомого ключа шифру задати початковий стан ГПВП, що має бути ідентичним початковому стану ГПВП — того, що був під час шифрування вихідного потоку голосових повідомлень. Розшифратор обробляє символьні послідовності завдяки реалізації будь-якого відомого алгоритму розшифрування, робота якого базується на використанні ГПВП. Суть розшифрування полягає в заміні лінгвістичних елементів зашифрованого потоку на елементи тезауруса, які є ідентичними елементам вихідного потоку. Результатом роботи розшифрувального пристрою є потік голосових повідомлень, ідентичний за змістом і формою його відображення потоку вихідних голосових повідомлень.

Висновки

Уперше на формальному рівні здійснено синтез методу побудови лексикографічної криптосистеми. Відповідно до цього методу простір заданої предметної області застосування інформаційної системи відображається на простір смислових образів цієї системи з урахуванням семантичних співвідношень між ними, що визначаються заданим простором обмежувальних умов. Тобто, синтезується структура семантичного тезауруса заданої предметної області. У структурі тезауруса кожному смисловому образу із загального простору образів ставиться у відповідність безліч інших смислових образів із цього ж простору, що перебувають із ним у співвідношенні смислової правдоподібності. Така структура тезауруса дає змогу під час шифрування замінювати повідомлення, що підлягають шифруванню, на інші правдоподібні повідомлення, які відображають істинний зміст вихідних повідомлень. Під час розшифрування, якщо знати пароль, уможливлюється здійснення зворотної заміни правдоподібних фальшивих повідомлень на справжні за змістом. Цей метод передбачає створення системи захисту в два етапи: спочатку створюється семантичний словник (тезаурус) прикладної області, а потім розробляються програмно-технічні засоби реалізації

операторів шифрування/розшифрування, в складі яких використовується створений тезаурус. Показано також, що іноді довжина ключа шифру для крипто-семантичного шифрування має обиратися залежно від допустимого значення ймовірності прийняття безпомилкових рішень у процесі дешифрування смислового потоку. Надано відповідний вираз для вибору мінімально можливого значення довжини ключа шифру.

ЛІТЕРАТУРА

1. *Shannon, C. E.* Communication Theory of Secrecy Systems: Bell System Technical Jour-nal, vol. 28, iss. 4, 1949, pp. 656–715.
2. *Darwish A., El-Gendy M.M., Hassanien A.E.* A New Hybrid Cryptosystem for Internet of Things Applications: Multimedia Forensics and Security. Intelligent Systems Reference Library, vol 115. 2017.
3. *Hu D., Su B., Zheng Sh.* Security and privacy protocols for perceptual imagehashing: Int. J. Sensor Networks, vol. 17, N 3, 2015.
4. *Oppliger R.* Contemporary Cryptography, Second Edition: Artech House, 2011. 571 p.
5. *Wei J., Zheng X., Yu J. et al.* Application of unicity distance in a cryptosystem based on chaos: 7th International Conference on Computer Science & Education (ICCSE), 2012.
6. *Яремчук Ю.* Алгебраїчні моделі асиметричних криптографічних систем: Захист ін-формації, т. 16, № 1, 2014. С. 68–80.
7. *Пономаренко В.С.* Сучасні методи та моделі обробки даних в інформаційних сис-темах Монографія Харків. Вид. ХНЕУ ім. Семена Кузнеця, 2013.
8. *Одарченко Р.С., Самойлик Є.О., Абакумова А.О.* Метод побудови семантичного словника у складі досконало стійкої криптосистеми захисту текстової інформації: Наукоємні технології № 3(39), 2018.
9. *Lokhande U., Gulve A.K.* Steganography using Cryptography and Pseudo Random Numbers: International Journal of Computer Applications, vol. 96, N 19. 2014.
10. *Mao Y., Wu M.* Unicity Distance of Robust Image Hashing: IEEE Transactions on Infor-mation Forensics and Security, vol. 2, iss. 3, 2007.
11. *Klimchuk V., Samoylik E., Gnatyuk V. et al.* Synthesis of Quite Proof Cryptosystem with Increased Unicity Distance for Cloud Computing: ICTERI Workshops, 2018.
12. *Goyal R., Khurana M.* Cryptographic Security using Various Encryption and Decryption Method: International Journal of Mathematical Sciences and Computing, vol.3, N 3, pp. 1–11. 2017
13. *Aparna R., Chithra Dr.P.L.* A Review on Cryptographic Algorithms for Speech Signal Security: International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol. 5, iss. 5. 2016.
14. *Широков В.А.* Феноменологія лексикографічних систем: Монографія. К.: Наукова думка, 2004, 326 с.
15. *Широков В.А., Бугаков О.В., Грязнухіна Т.О.* та ін. Корпусна лінгвістика: Монографія. К.: Довіра, 2005, 471 с.
16. *Белявская Е.Г.* Семантическая структура слова в номинативном и коммуникативном аспектах (когнитивные основания формирования и функционирования семантической структуры слова): Монографія. М., 1992.
17. *Nechiporuk O.P., Odarchenko R.S., Potapov V.G.* et al. Speech transfer in digital communication systems: Electronics and Control Systems, vol. 4. 2011.
18. *XML.* https://www.w3.org/People/Raggett/Drafts/xml.html.

Надійшла 31.10.2019

REFERENCES

1. *Shannon, C.E.*, 1949. "Communication Theory of Secrecy Systems". Bell System Technical Journal, 28(4), pp. 656–715.
2. *Darwish, A., El-Gendy, M.M., Hassanien, A.E.*, 2016. "A New Hybrid Cryptosystem for Internet of Things Applications: Multimedia Forensics and Security". Intelligent Systems Reference Library, 115, pp. 365-380. DOI: https://doi.org/1136/10.1007/978-3-319-44270-9_16.
3. *Hu, D., Su, B., Zheng, Sh.*, 2015. "Security and privacy protocols for perceptual imagehashing". Int. J. Sensor Networks, 17(3), pp. 146-162.

4. *Oppliger, R.*, 2011. Contemporary Cryptography, Second Edition: Artech House. 571 p.
5. *Wei, J., Zheng, X., Yu, J. et al.*, 2012. "Application of unicity distance in a cryptosystem based on chaos". 7th International Conference on Computer Science & Education (ICCSE), DOI: 10.1109/ICCSE.2012.6295088.
6. *Yaremchuk, Yu.*, 2014. Alhebrayichni modeli asymetrychnykh kryptohrafichnykh system: Zakhyst informatsiyi, 16(1), pp. 68– (In Ukrainian).
7. *Ponomarenko, V.S.,* 2013. Suchasni metody ta modeli obrobky danykh v informatsiynykh systemakh Monohrafiya Kharkiv. Vyd. KHNEU im. Semena Kuznetsya, 540 p. (In Ukrainian).
8. *Odarchenko, R.S., Samoylyk, Ye.O., Abakumova, A.O.*, 2018. "Metod pobudovy semantychnoho slovnyka u skladi doskonalo stiykoyi kryptosystemy zakhystu tekstovoyi informatsiyi". Naukoyemni tekhnolohiyi, 3(39), pp. 355-361. (In Ukrainian).
9. *Lokhande, U., Gulve, A.K.*, 2014. "Steganography using Cryptography and Pseudo Random Numbers". International Journal of Computer Applications, 96(19), pp. 40-45.
10. *Mao, Y., Wu, M.*, 2007. "Unicity Distance of Robust Image Hashing". IEEE Transactions on Information Forensics and Security, 2(3), pp. 462-467.
11. *Klimchuk, V., Samoylik, E., Gnatyuk, V.* et al., 2018. "Synthesis of Quite Proof Cryptosystem with Increased Unicity Distance for Cloud Computing". ICTERI Workshops.
12. *Goyal, R., Khurana, M.*, 2017. "Cryptographic Security using Various Encryption and Decryption Method". International Journal of Mathematical Sciences and Computing, 3(3), pp. 1–11.
13. *Aparna, R., Chithra, Dr.PL.*, 2016. "A Review on Cryptographic Algorithms for Speech Signal Security". International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 5(5). pp. 84-88.
14. *Shyrokov, V.A.*, 2004. Fenomenolohiya leksykohrafichnykh system: Monohrafiya. Kyiv: Naukova dumka. 326 p. (In Ukrainian).
15. *Shyrokov, V.A., Buhakov, O.V., Hryaznukhina, T.O.* et al., 2005. Korpusna linhvistyka: Monohrafiya. Kyiv: Dovira. 471 p. (In Ukrainian).
16. *Belyavskaya, Ye.G.,* 1992. Semanticheskaya struktura slova v nominativnom i kommunikativnom aspektakh (kognitivnyye osnovaniya formirovaniya i funktsionirovaniya semanticheskoy struktury slova). Monograf ya. (In Russian).
17. *Nechiporuk, O.P., Odarchenko, R.S., Potapov, V.G.* et al., 2011. Speech transfer in digital communication systems. Electronics and Control Systems, 4 (30).
18. *XML*, [online] Available at: https://www.w3.org/People/Raggett/Drafts/xml.html [Accessed 31 Okt. 2019].

Надійша 31.10.2019

R.S. Odarchenko, Doctor of Technical Sci., Senior Research Associate, International Research and Training Center for Information Technologies and Systems of the NAS of Ukraine and MES of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine,
odarchenko.r.s@ukr.net
E.O. Samoilik, Applicant, Faculty of Aeronavigation, Electronics and Telecommunications, National Aviation University, Lubomyr Husar ave., 1, Kyiv, 03058, Ukraine,
sea110913@gmail.com
V.M. Simakhin, Postgraduate, Junior research associate, International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine,
sima@irtc.org.ua
V.O. Borovik, Postgraduate, Junior research associate, International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine,
dep185@irtc.org.ua
R.M. Tymchyshyn, Postgraduate, Junior research associate, International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine,
dep185@irtc.org.ua

CRYPTO-SEMANTIC SYSTEM FOR TEXT INFORMATION PROTECTION

Introduction. Information security is one of the most important areas of computer science. Often it is necessary to transmit messages containing sensitive or secret information that a third party must not intercept and recognize. A crucial task is to create the methods for constructing cryptographically strong cryptosystem.

Purpose. A method of constructing cryptographically strong cryptosystem, which provides a formal opportunity to encrypt unlimited volumes of text messages.

Methods. Lexicographic methods of information security

Results. Usage of the lexicographic mechanism for enlarging the alphabet of the textual information's language allows increasing the unicity distance, which is the main threshold indicator for the cryptosystem to belong to the class of cryptographically strong security systems with theoretically proven ideal information-theoretic stability. Images' space of a given subject area of an information system application is mapped onto the space of semantic images of this system, taking into account the semantic relationships between them, which are determined by a given restrictive conditions. This is the basis of the thesaurus structure, which allows during encryption to replace the message to be encrypted with other plausible messages that reflect the true meaning of outgoing messages.

Conclusion. For the first time at a formal level, a synthesis of the method of constructing a lexicographic cryptosystem with theoretical absolute stability has been carried out. A security system has been built, which consists of two parts: a semantic thesaurus developed for a specific application area, and software and hardware tools for implementing encryption/decryption operators using the created thesaurus.

Keywords: *technical information protection, perfectly persistent symmetric cryptosystems, lexico-graphic systems, unicity distance, semantic thesaurus.*

Р.С. Одарченко, доктор технических наук, старший научный сотрудник, Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, просп. Академика Глушкова, 40, Киев, 03187, Украина,
odarchenko.r.s@ukr.net
Е.А. Самойлик, соискатель, Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, просп. Любомира Гузара, 1, Киев, 03058, Украина,
sea110913@gmail.com
В.М. Симахин, аспирант, м.н.с., Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, просп. Академика Глушкова, 40, Киев, 03187, Украина,
sima@irtc.org.ua
В.А. Боровик, аспирант, м.н.с., Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, просп. Академика Глушкова, 40, Киев, 03187, Украина,
dep185@irtc.org.ua
Р.М. Тимчишин, аспирант, м.н.с., Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, просп. Академика Глушкова, 40, Киев, 03187, Украина,
dep185@irtc.org.ua

КРИПТОСЕМАНТИЧЕСКАЯ СИСТЕМА ЗАЩИТЫ ТЕКСТОВОЙ ИНФОРМАЦИИ

Введение. Защита информации является одним из важнейших направлений компьютерных наук. Зачастую необходимо передавать сообщения, содержащие чувствительную или секретную информацию, которую не должна перехватить и распознать третья сторона. Актуальной задачей является создание методов построения абсолютно устойчивых криптосистем.

Цель. Метод построения абсолютно устойчивой криптографической системы, которая обеспечивает формальную возможность шифрования неограниченных объемов текстовых сообщений.

Методы. Лексикографические методы защиты информации.

Результаты. Применение лексикографического механизма укрупнения алфавита языка для отображения текстовой информации позволяет увеличить так называемое расстояние единственности, которое является основным пороговым показателем принадлежности криптосистемы к классу абсолютно устойчивых систем защиты с теоретически доказанной идеальной теоретико-информационной устойчивостью. Пространство образов заданной предметной области применения информационной системы отображается на пространство смысловых образов этой системы с учетом семантических соотношений между ними, которые определяются заданным пространством ограничительных условий. Это является основой структуры тезауруса, который позволяет во время шифрования заменять сообщение, подлежащие шифрованию, на другие правдоподобные сообщения, отражающие истинный смысл исходящих сообщений.

Выводы. Впервые на формальном уровне осуществлен синтез метода построения лексикографической криптосистемы с теоретической абсолютной устойчивостью. Построена система защиты, состоящая из двух частей: семантического тезауруса, разработанного для конкретной прикладной области, и программно-технических средств реализации операторов шифрования/дешифрования, использующих созданный тезаурус

Ключевые слова: *техническая защита информации, совершенно стойкие симметричные криптосистемы, лексикографические системы, расстояние единственности, семантический тезаурус.*