МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ АЕРОНАВІГАЦІЇ,
ЕЛЕКТРОНІКИ ТА ТЕЛЕКОМУНІКАЦІЙ
КАФЕДРА ТЕЛЕКОМУНІКАЦІЙНИХ ТА РАДІОЕЛЕКТРОННИХ СИСТЕМ

ДОПУСТИТИ ДО ЗАХИСТУ
Завідувач кафедри

_____ Роман ОДАРЧЕНКО
“_____”_____2023 р.

КВАЛІФІКАЦІЙНА РОБОТА

(ПОЯСНЮВАЛЬНА ЗАПИСКА)

ВИПУСКНИКА ОСВІТНЬОГО СТУПЕНЯ БАКАЛАВР

Тема: «Криптосемантичний метод захисту текстової інформації»

Виконавець: _____ Данило ГАВРЮШИН
(підпис)

Керівник: _____ Володимир ЧУПРИН
(підпис)

Нормоконтролер: _____ Денис БАХТІЯРОВ
(підпис)

Київ 2023

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет аеронавігації, електроніки та телекомунікацій
Кафедра телекомунікаційних та радіоелектронних систем
Спеціальність 172 «Телекомунікації та радіотехніка»
Освітньо-професійна програма «Телекомунікаційні системи та мережі»

ЗАТВЕРДЖУЮ
Завідувач кафедри

Роман ОДАРЧЕНКО
“_____”_____2023 р.

ЗАВДАННЯ
на виконання кваліфікаційної роботи

Гаврюшина Данила Юрійовича
(прізвище, ім'я, по батькові випускника в родовому відмінку)

1. Тема кваліфікаційної роботи: «Криптосемантичний метод захисту текстової інформації»

затверджена наказом ректора від «29» березня 2023 р. № 421/ст

2. Термін виконання роботи: з 22.05.2023 р. по 25.06.2023 р.

3. Вихідні дані до роботи: інформаційна модель суб'єкта, інформаційна модель інтелекту

4. Зміст пояснювальної записки: вступ, сутність криптографічного перетворення інформації, аналіз сучасного стану проблеми захисту текстової інформації, криптосемантичних метод захисту текстової інформації, висновки, список використаних джерел.

5. Перелік обов'язкового графічного (ілюстративного) матеріалу: варіант реалізації методу побудови лексикографічної криптосистеми, інформаційна модель інтелекту

6. Календарний план-графік

№ пор.	Завдання	Термін виконання	Відмітка про виконання
1	Розробити деталізований зміст розділів кваліфікаційної роботи	12.04.2023-14.04.2023	Виконано
2	Вступ	15.04.2023	Виконано
3	Сутність криптографічного перетворення інформації	17.04.2023-04.05.2023	Виконано
4	Аналіз сучасного стану проблеми захисту текстової інформації	05.05.2023-17.05.2023	Виконано
5	Криптосемантичних метод захисту текстової інформації	18.05.2023-29.05.2023	Виконано
6	Усунення недоліків та захист кваліфікаційної роботи	02.06.2023-09.06.2023	Виконано

7. Дата видачі завдання: “19" травня 2023 р.

Керівник кваліфікаційної роботи _____ Володимир ЧУПРИН
(підпис керівника) (П.І.Б.)

Завдання прийняв до виконання _____ Данило ГАВРЮШИН
(підпис випускника) (П.І.Б.)

РЕФЕРАТ

Кваліфікаційна робота «Криптосемантичний метод захисту текстової інформації» містить 49 сторінок, 8 рисунків, 1 таблиця, 11 використаних джерел.

Захист текстової інформації, семантичний тезаурус, лексикографічна система

Об’єкт дослідження – процес захисту мовної інформації.

Предмет дослідження – методи підвищення ефективності шифрування текстової інформації.

Мета кваліфікаційної роботи – підвищення ефективності шифрування текстової інформації за рахунок використання особливостей семантичних характеристик.

Метод дослідження – розробка методів виконана на основі результатів теорії випадкових процесів, теорії телетрафіка, теорії криптографічних систем та теорії лексикографічних систем.

Матеріали кваліфікаційної роботи рекомендується використовувати при вивченні криптографічних систем.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ 6

ВСТУП 7

1. СУТНІСТЬ КРИПТОГРАФІЧНОГО ПЕРЕТВОРЕННЯ ІНФОРМАЦІЇ 8

1.1 Цілі та завдання криптографії 8

1.2 Симетричні та асиметричні шифри 11

1.3 Загальна характеристика потокових шифрів 15

1.4 Гібридні шифри 18

2. АНАЛІЗ СУЧАСНОГО СТАНУ ПРОБЛЕМИ ЗАХИСТУ ТЕКСТОВОЇ ІНФОРМАЦІЇ 23

2.1 Огляд основних методів криптографічного захисту 23

2.2 Аналіз проблем криптосемантичного методу захисту тексту 26

2.3 Принципові властивості криптосемантичних методів 28

2.4 Підходи до реалізації криптосемантичного методу захисту інформації ... 29

3. КРИПТОСЕМАНТИЧНИХ МЕТОД ЗАХИСТУ ТЕКСТОВОЇ ІНФОРМАЦІЇ 36

3.1 Схема реалізації криптосемантичного методу 36

3.2 Синтез тезаурусу та схема структурування 40

3.3 Забезпечення випадковості процедури заміни 44

ВИСНОВКИ 47

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 49

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ

DES (Data Encryption Standard) – симетричний алгоритм шифрування

IDEA (International Data Encryption Algorythm) – міжнародний алгоритм шифрування даних

RSA (Rivest, Shamir, Adleman) – криптографічний алгоритм з відкритим ключем

ГПВЧ – генератор псевдовипадкових чисел

AES (Advanced Encryption Standard) – симетричний алгоритм блочного шифрування

CSM (Crypto-Semantic Method) – криптосемантичним метод

NLP (Natural language processing) – обробку природної мови

SHA (Secure Hash Algorithm) – алгоритм безпечного хешування

SSL (Secure Sockets Layer) – рівень захищених сокетів

TSL (Transport Layer Security) – протокол захисту транспортного рівня

ВСТУП

Актуальність теми. Наразі відомі криптографічні системи не гарантують абсолютного захисту від порушення конфіденційності, через що існує проблема недовіри до надійності цих систем в задачах передавання інформації, що характеризуються високими рівнями секретності. Тому тема цієї кваліфікаційної роботи спрямована на створення методу, що дозволяє відносно мало затратним шляхом забезпечити ефективний захист текстової інформації.

Мета і завдання дослідження. Підвищення ефективності шифрування текстової інформації за рахунок використання особливостей семантичних характеристик.

Для досягнення поставленої мети вирішуються такі наукові завдання.

1. аналіз ефективності сучасних методів захисту мовної інформації;
2. розробка моделі крипто-семантичного словника;
3. розробка методу побудови лексикографічної криптосистеми;
4. розробка лексикографічного методу захисту текстової інформації.

Об’єктом дослідження – процес захисту мовної інформації.

Предметом дослідження – методи підвищення ефективності шифрування текстової інформації.

Методи дослідження. Розробка методів виконана на основі результатів теорії випадкових процесів, теорії телетрафіка, теорії криптографічних систем та теорії лексикографічних систем.

Практичне значення отриманих результатів. Матеріали кваліфікаційної роботи рекомендується використовувати при вивченні криптографічних систем.

Апробація отриманих результатів. Основні положення роботи доповідалися та обговорювалися на таких конференціях:

Науково-практична конференція «проблеми експлуатації та захисту інформаційно-комунікаційних систем», м. Київ, 2023 р.

РОЗДІЛ 1
СУТНІСТЬ КРИПТОГРАФІЧНОГО ПЕРЕТВОРЕННЯ ІНФОРМАЦІЇ

1.1. Цілі та завдання криптографії

Передача інформації (даних) здійснюється між двома абонентами, званими джерелом повідомлення (ДжП) та одержувачем повідомлення (ОдП). Джерелом та одержувачем можуть бути люди або технічні засоби. ДжП та ОдП обмінюються інформацією за допомогою каналу передачі. Зазначено також, що і в системах зберігання інформації завжди можна виділити ДжП та ОдП. У разі каналом передачі тут виступає пристрій зберігання інформації (пам'ять). Наприклад, при записі даних в ОЗП (оперативно запам'ятовуючий пристрій) комп'ютера в якості ДжП і ОдП може виступати процесор (відповідно при записі і читанні даних). Таким чином, найпростіша інформаційна система складається із трьох перерахованих елементів. Її узагальнена структурна схема наведено на рис. 3.1 (тут параметр k означає кількість символів у повідомленні).

Рис. 1.1. Узагальнена структурна схема інформаційної системи

Під визначення ІС (інформаційна система) підпадає будь-яка система обробки інформації. Далі розглядатимемо ІС як сукупність апаратно-програмних засобів, задіяних для вирішення деякого прикладного завдання. Математичною основою опису та аналізу процесів в ІС у широкому сенсі є теорії інформації. Виникнення теорії інформації пов'язують зазвичай із появою фундаментальних робіт К. Шеннона.

Кодування у сенсі – перетворення повідомлення на сигнал.

Кодування у вузькому значенні – уявлення вихідних знаків, званих символами, у іншому алфавіті.

Воно здійснюється з різними цілями: підвищення надійності передачі, зниження фізичного обсягу повідомлення, підвищення рівня конфіденційності або безпеки. Зазначені цілі відповідають трьом базовим способам перетворення

повідомлення (X_k) до його передачі каналом зв'язку:

- кодування або завадостійке кодування;
- стиснення або архівування повідомлень;
- криптографічне перетворення.

Подальша частина цього посібника спрямовано вивчення та аналіз методів криптографії.

Протягом останніх 20 років спостерігається бурхливий розвиток відкритих академічних досліджень у цій галузі. Поки звичайні громадяни використовували «класичну» криптографію, «комп'ютерна» криптографія ще з часів Першої світової війни застосовувалася виключно у військових цілях.

Говорячи про історичні аспекти наукових досліджень у галузі криптографії, слід зазначити той факт, що весь період із давніх часів до 1949 року можна назвати донауковим періодом, коли засоби «закриття» інформації не мали суворого математичного обґрунтування. Поворотним моментом, який надав криптографії науковість і виділив її в окремий напрямок математики, стала публікація вже згадуваної роботи К. Шеннона «Теорія зв'язку в секретних системах».

Сучасна криптографія широко використовується за стінами військових відомств. Пересічні користувачі інформаційних технологій отримали можливість захиститися від всемогутніх шкідників (не тільки хакерів або кракерів). Нижче проаналізуємо основні особливості та оцінимо ефективність криптографічних методів перетворення (захисту) інформації.

Як зазначалося вище, криптографія є однією з двох гілок загального наукового спрямування – криптології. Другою гілкою криптології є криптоаналіз. Цілі криптографії та криптоаналізу прямо протилежні.

Криптографічні методи знайшли широке застосування у практичній інформатиці на вирішення численних проблем інформаційної безпеки. У проблематиці сучасної криптографії можна виділити такі три типи основних завдань:

1) забезпечення конфіденційності (секретності);

2) забезпечення анонімності (невідстежуваності);

3) забезпечення аутентифікації інформації та джерела повідомлення.

Перший тип завдань належить захисту інформації від несанкціонованого доступу по секретному ключу. Доступ до інформації (інформаційних ресурсів) мають лише власники ключа. Другий і третій типи завдань зобов'язані своєю постановкою масового застосування електронних засобів обробки та передачі (банківська сфера, електронна комерція, канали міжособистісної комунікації та інших).

Криптографічне перетворення, як і два інших з вищезгаданих (перешкодостійке кодування і стиск), складається з двох етапів: прямого і зворотного. Пряме перетворення називають шифруванням або зашифруванням (відповідно до стандарту ISO 7492-2 – зашифрування, encrypt), зворотне – дешифруванням або розшифруванням (розшифрування, decrypt). Процес передачі зашифрованих повідомлень ілюструє рис. 1.2.

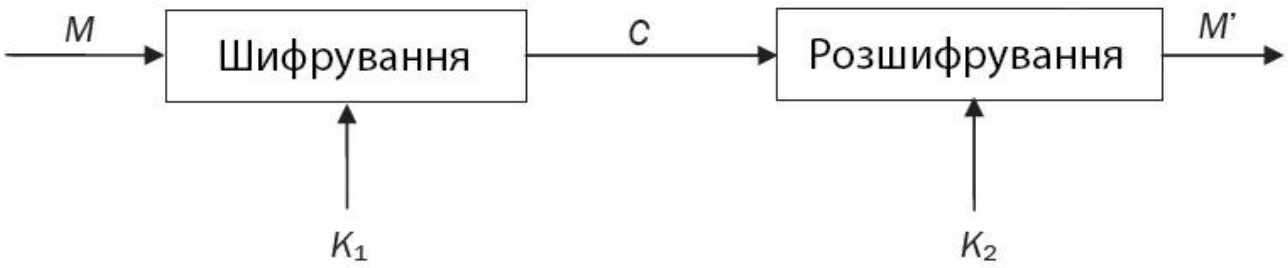

Рис. 1.2. Загальна структурна схема криптосистеми

Вихідне повідомлення називається відкритим текстом (M, від англ. Message). Зашифроване повідомлення - шифртекст або шифрограмою (C, від англ. cipher). Після зворотного перетворення отримуємо вихідний (або наближений до нього) документ (M'). Таким чином, канал передається шифртекст C.

Якщо скористатися символьними позначеннями, введеними нами раніше (X_k – дані до прямого перетворення, X_n – дані після прямого перетворення), т. е. покласти $X_k = M$ і $X_n = C$, порівняльною характеристикою криптографічних методів буде рівність $k = n$. Це означає, що довжина відкритого та зашифрованого повідомлень не

змінюється (виняток становлять методи шифрування з використанням електронного цифрового підпису).

Функція зашифрування E в математичному вигляді представляється так:

$$E(M) = C \tag{1.1}$$

У зворотному процесі функція розшифрування D відновлює M:

$$D(C) = M \tag{1.2}$$

Оскільки зміст зашифрування та подальшого розшифрування повідомлення полягає у відновленні вихідного відкритого тексту, справедлива така рівність:

$$D\big(E(M)\big) = M \tag{1.3}$$

У цьому аналізі прямого і зворотного процесів перетворення шифр ототожнюється з криптографічним алгоритмом, що є математичною функцією, яка використовується для зашифрування і для розшифрування інформації.

До появи комп'ютерів криптографія ґрунтувалася саме на таких алгоритмах, які називають текстовими.

Основою були операції заміни одних символів іншими чи перестановка символів місцями. Перші алгоритми відносяться до класу підстановочних, інші – перестановочні. Сучасні криптосистеми використовують як підстановки, і перестановки символів.

1.2. Симетричні та асиметричні шифри

Сучасна криптографія вирішує проблему за допомогою ключа, що на рис. 1.2 позначений буквою K. Такий ключ може бути будь-яким значенням, вибраним з величезної кількості. Безліч можливих ключів називають простором ключів.

Ключ – секретний параметр, який керує перебігом перетворення. Ключ визначає конкретний варіант перетворення. Ключ використовується в обох операціях як зашифрування, так і розшифрування. Таким чином, тепер функції зашифрування та розшифрування набувають наступного вигляду:

функція зашифрування E:

$$E_k(M) = C \tag{1.4}$$

функція розшифрування D:

$$D_k(C) = M \tag{1.5}$$

Або

$$D_k\big(E_K(M)\big) = M \tag{1.6}$$

Співвідношення вище описують криптосистеми, що використовують однакове значення ключа для зашифрування та розшифрування (K1 = K2 = K). Такі криптосистеми називаються симетричними.

Найпростішим прикладом симетричного криптоперетворення є додавання по модулю двох повідомлень M з ключем K для отримання шифртексту. Розшифрування досягається виконанням такої операції над C і K.

Симетричні криптосистеми називаються також криптосистемами з таємним ключем, оскільки значення ключа має бути відоме лише відправнику та одержувачу повідомлень.

Інший клас сучасних криптоалгоритмів для зашифрування та розшифрування використовує різні ключі ($K1 \neq K2$). У цьому випадку формальне уявлення аналізованих перетворень представляється наступним чином:

функція зашифрування E:

$$E_{k1}(M) = C \tag{1.7}$$

функція розшифрування D:

$$D_{k2}(C) = M \tag{1.8}$$

Або

$$D_{k2}\left(E_{K1}(M)\right) = M \tag{1.9}$$

Ключі K1 і K2 є різними, але взаємозалежними (один із них таємницею не є). Тому асиметричні криптосистеми називають криптосистемами з відкритим або публічним ключем.

Симетричні алгоритми поділяються на два підкласи. Одні алгоритми обробляють відкритий текст побітово (іноді побайтово). Такі алгоритми називають потоковими. Інші алгоритми обробляють групи (блоки) бітів відкритого тексту. Ці алгоритми називають блоковими.

У сучасних комп'ютерних алгоритмах типовий розмір блоку становить 64 біти. Більш детальний розгляд блокових і потокових (іноді в літературі їх називають потоковими) шифрів буде дано нижче.

Коротко охарактеризуємо деякі з відомих блокових алгоритмів симетричної криптографії. Алгоритм Lucifer. Запропоновано У. Тачменом та Х. Файстелом. Алгоритм є послідовністю перестановок і підстановок. Основні блоки нагадують алгоритм DES. Згадаймо, що в алгоритмі DES результат функції f складається операцією XOR із входом попереднього раунду, утворюючи вхід наступного раунду (див. рис. 5.1). S-блоки алгоритму Lucifer мають 4-бітові входи та виходи. Причому вибору одного з двох можливих блоків використовується біт ключа.

Алгоритм IDEA (International Data Encryption Algorythm – міжнародний алгоритм шифрування даних). Працює з 64-бітними блоками даних та 128-бітними ключами. Є підстановочно-перестановним алгоритмом.

Блок даних ділиться на чотири 16-бітових блоки, що є входами першого раунду. Весь алгоритм складається із 8 раундів. У кожному раунді 4 блоки піддаються

операціям XOR, додаванням та множенням один з одним і шістьма 16-бітними підключами. Між раундами другий та третій блоки даних змінюються місцями. У кінцевій операції чотири 16-бітові блоки даних об'єднуються з 16-бітними підключами.

Програмна реалізація IDEA виконується приблизно вдвічі швидше, ніж DES.

Для порівняльної оцінки розглянутих алгоритмів у табл. 1.2 наведено швидкість шифрування (в умовних одиницях) кожного з них.

Алгоритм Blowfish Працює із 64-бітними блоками даних.

Алгоритм складається зазвичай з 16 раундів (можливо 12 або 20), у кожному з яких виконуються операції підстановки та перестановки, що залежать від ключа. Використовуються лише додавання та XOR над 32-бітними словами. Кожен із 4 S-блоків зберігає по 256 чисел.

Таблиця 1.2

Швидкість шифрування (в умовних одиницях) блокових симетричних шифрів

Алгоритм	Швидкіксть шифрування
Blowfish (12 раундов)	182
Blowfish (12 раундов)	135
Blowfish (12 раундов)	110
DES	35
IDEA	70
Lucifer	52
3DES	12

Дві проблеми, пов'язані з практичним використанням симетричних криптосистем (зберігання та обмін ключовою інформацією), стали важливими спонукальні мотиви для розробки принципово нового класу методів шифрування: криптографії з відкритим ключем, або асиметричної криптографії.

Концепція нового підходу запропонована Вітфілдом Діффі (Whitfield Diffie) та Мартіном Хеллманом (Martin Hellman), а також, незалежно, Ральфом Мерклом (Ralph Merkle). В основу покладено ідею використовувати ключі парами: один – для

зашифрування (відкритий або публічний ключ), інший – для розшифрування (таємний ключ). Таким чином, на відміну від симетричних систем, ключі K1 та K2 (див. рис. 3.2) відрізняються. Несекретний ключ може передаватися відкритим каналом. Його знання не дає зловмиснику можливості отримати доступ до інформації, що міститься у повідомленні.

У. Діффі та М. Хеллман у 1976 році концепцію трансформували в алгоритм (у літературі відповідний алгоритм називається алгоритмом Діффі – Хеллмана). Починаючи з цієї події, було створено безліч алгоритмів, що базуються на концепції відкритих ключів. Як правило, ці алгоритми ґрунтуються на вирішенні важких завдань, які ми розглядали вище: обчислення дискретного логарифму та розкладання чисел на множники.

Деякі зі створених алгоритмів з різних причин придатні для шифрування, інші – для генерації електронного цифрового підпису (ЕЦП). І лише три із створених алгоритмів використовуються в обох випадках: алгоритми RSA, Ель-Гамалю та Рабіна. Алгоритм Діффі – Хеллмана для розподілу ключів: абоненти А та В можуть скористатися цим алгоритмом для обміну ключовою інформацією на відкритих каналах. Попередньо сторони вибирають великі прості числа n та g.

1.3. Загальна характеристика потокових шифрів

Як ми переконалися, у блоковому шифрі з двох однакових блоків відкритого тексту виходять однакові блоки шифрованого тексту, що, безумовно, одна із недоліків алгоритмів. Уникнути цього дозволяють потокові шифри, в яких шифруюче перетворення символу відкритого тексту змінюється від одного елемента до іншого.

Потоковий шифр – це, по суті, симетричний шифр, у якому кожен символ відкритого тексту перетворюється на символ шифртексту З залежно як від використовуваного ключа, а й його розташування на потоці відкритого тексту М.

Класичний приклад потокового шифру – шифр Вернама, чи одноразовий блокнот. Вже зазначалося, що з гами послідовність бітів вибирається випадково і довжина гами дорівнює, по крайнього заходу, довжині повідомлення, то зламати

шифр неможливо. Але в даного режиму шифрування є і негативна особливість - проблеми з передачею і зберіганням ключів, адже ключі, порівняні по довжині з повідомленнями, що передаються, важко використовувати на практиці. Тому основна ідея сучасних потокових шифрів – реалізувати концепцію одноразового блокнота, використовуючи секретний ключ меншої довжини, з якого для гами генерується псевдовипадкова числова послідовність, схожа на випадкову.

Можна сказати, що зашифрування здійснюється накладенням гами (шифрування гамуванням). А сама гама є ключем шифрування. У практичних потокових шифрах довжина шифртексту набагато більша за довжину секретного ключа, а ключова послідовність є псевдовипадковою і має деякий період. Вочевидь, що й послідовність бітів гами немає періоду і вибирається випадково, то зламати шифр неможливо.

Якби генератор видавав нескінченну послідовність бітів, у якій кожен біт породжувався незалежно і з ймовірністю 1/2 приймав значення 0 чи 1, ми отримали б стійкий шифр. Але мати ключ, рівний за розміром даних, що шифруються, представляється проблематичним. Тому потокові шифри виробляють вихідну гаму на основі деякого секретного ключа невеликого розміру (наприклад, 128 бітів). З його допомогою генерується псевдовипадкова послідовність, що гамує (ПСГП). Вона має задовольняти постулатам С. Голомба (всього три), основні з яких можна сформулювати наступним чином:

• кількість «1» у кожному періоді ПСГП має відрізнятись від кількості «0» не більше, ніж на одиницю;

• у кожному періоді ПСГП половина серій (з однакових символів) повинна мати довжину один, одна чверть повинна мати довжину два, одна восьма повинна мати довжину три тощо; для кожної з цих довжин має бути однакова кількість серій із «1» і «0».

Загалом схема потокового шифру зображена на рис. 1.3.

Рис. 1.3. Схема потокового шифру

Вочевидь, якщо у разі всі символи, прийняті без спотворення, будуть розшифровані правильно. Відбудеться лише одна втрата символ тексту. Якщо ж один із символів шифртексту під час передачі каналом зв'язку було втрачено. Це призведе до неправильного розшифрування всього тексту, який слідує за втраченим символом.

Практично у всіх каналах передачі для потокових систем шифрування присутні перешкоди. Таким чином для запобігання втраті інформації вирішують проблему синхронізації процедури зашифрування та розшифрування тексту. За способом вирішення цієї проблеми шифрсистеми поділяються на синхронні та самосинхронізуються (асинхронні).

При зашифруванні генератор потоку ключів видає біти потоку ключів, які ідентичні біт потоку ключів при розшифруванні. Втрата символу шифртексту призведе до порушення синхронізації між цими двома генераторами і неможливості розшифрування частини повідомлення, що залишилася. Очевидно, що у цій ситуації відправник та одержувач повинні повторно синхронізуватися для продовження роботи [1].

Головна властивість СПШ – нерозповсюдження помилок. Помилки відсутні, поки працюють синхронно шифрувальний та дешифрувальний пристрій відправника та одержувача інформації.

Перебої у роботі системи називають розсинхронізацією. Її може викликати розбіжність швидкостей зашифрування/розшифрування на різних кінцях каналу зв'язку, випадання знаків при передачі та ін. Якщо так сталося, треба відновити

синхронність у роботі генераторів гами – почати повторне шифрування з реініціалізацією ключа обома користувачами. Один із методів боротьби з розсинхронізацією – розбити відритий текст на відрізки, початок та кінець яких виділити вставкою контрольних міток (спеціальних маркерів). В результаті цього пропущений під час передачі символ призводить до неправильного розшифрування лише доти, доки не буде прийнято одного з маркерів.

Сформулюємо як результат короткого опису СПШ такі властивості:

- вимоги щодо синхронізації: при використанні СПШ одержувач і відправник повинні бути синхронізовані, тобто повинні виробляти однакові значення ключового потоку для відповідних символів потоку даних, що передається. Якщо синхронізація порушиться (наприклад, через втрату символу під час передачі), процес розшифрування не дасть коректного результату;

- відсутність розмноження помилок: зміна символу шифртексту під час передачі не викликає помилок при розшифруванні інших символів шифртексту;

- властивість активної атаки: будь-яка вставка або видалення символу в шифртекст активним противником призводить до порушення синхронізації та виявляється одержувачем, що розшифровує повідомлення (наслідок першої властивості); активний противник може змінювати символи шифртексту, і ці зміни призведуть до відповідних змін у відкритому тексті, який отримується при розшифруванні (наслідком другої властивості).

1.4. Гібридні шифри

Гібридне шифрування – це підхід до кодування та декодування даних, який поєднує швидкість і зручність загальнодоступної асиметричної схеми шифрування з ефективністю приватної схеми симетричного шифрування.

У цьому підході до криптографії відправник генерує приватний ключ, шифрує ключ за допомогою алгоритму відкритого ключа, а потім шифрує все повідомлення (включаючи вже зашифрований приватний ключ) оригінальним симетричним ключем. Закодований шифр можна розшифрувати, лише якщо одержувач знає

закритий ключ, який спочатку згенерував відправник.

Дослідники безпеки шукають способи використання гібридного шифрування як альтернативи більш традиційним схемам шифрування в квантових обчисленнях. Однак доки стандарти не будуть запроваджені, гібридний підхід може супроводжуватися підвищеним ризиком недоліків реалізації, які можуть звести нанівець корисність схеми шифрування.

Гібридна схема шифрування – це така, яка поєднує зручність асиметричної схеми шифрування з ефективністю симетричної схеми шифрування.

Гібридне шифрування досягається шляхом передачі даних за допомогою унікальних сеансових ключів разом із симетричним шифруванням. Шифрування з відкритим ключем реалізовано для шифрування випадковим симетричним ключем. Потім одержувач використовує метод шифрування з відкритим ключем, щоб розшифрувати симетричний ключ. Після відновлення симетричного ключа він використовується для розшифровки повідомлення.

Переваги та недоліки гібридного шифрування

Поєднання методів шифрування має ряд переваг. Один полягає в тому, що канал з'єднання встановлюється між наборами обладнання двох користувачів. Тоді користувачі мають можливість спілкуватися за допомогою гібридного шифрування.

Асиметричне шифрування може уповільнити процес шифрування, але з одночасним використанням симетричного шифрування покращуються обидві форми шифрування. Результатом є додаткова безпека процесу передачі разом із загальною покращеною продуктивністю системи.

Гібридні шифри поєднують в собі переваги симетричних та асиметричних шифрів. Вони використовують симетричні шифри для шифрування самої інформації, а асиметричні шифри - для безпечного обміну симетричним ключем. Це дозволяє поєднати високий рівень безпеки симетричних шифрів з перевагами асиметричних шифрів у відношенні до обміну ключами.

Порівняння шифрів в криптосемантиці показує, що кожен з них має свої переваги та обмеження. Вибір певного шифру залежить від конкретних потреб захисту інформації, таких як рівень безпеки, швидкість обробки, розмір ключа та інші

фактори. Симетричні шифри надають швидкість та ефективність, а асиметричні шифри забезпечують безпеку та можливість безпечного обміну ключами. Гібридні шифри поєднують переваги обох підходів. Розуміння особливостей різних шифрів допоможе вибрати найкращий метод захисту інформації для конкретної ситуації.

Криптографія в сучасному світі відіграє важливу роль у забезпеченні безпеки даних та інформації. Одним з найбільш ефективних підходів до шифрування є використання гібридних шифрів. Гібридні шифри поєднують в собі переваги симетричного та асиметричного шифрування, що дозволяє забезпечити високий рівень безпеки та ефективності в обробці даних.

Гібридні шифри базуються на поєднанні симетричного та асиметричного шифрування. Вони використовують симетричний шифр для шифрування самої інформації, а асиметричний шифр - для безпечного обміну симетричним ключем. Основний принцип полягає в тому, що симетричний ключ, який використовується для шифрування даних, генерується випадково та шифрується за допомогою публічного ключа отримувача. Цей зашифрований симетричний ключ передається разом з зашифрованою інформацією. Отримувач, використовуючи свій приватний ключ, розшифровує отриманий симетричний ключ і використовує його для розшифрування самої інформації [2].

Переваги гібридних шифрів:

- Висока безпека: Гібридні шифри поєднують в собі переваги симетричного та асиметричного шифрування. Використання симетричного шифру для шифрування самої інформації забезпечує високий рівень безпеки, оскільки симетричні шифри є швидкими та стійкими. Асиметричне шифрування використовується лише для обміну симетричним ключем, що забезпечує безпечний обмін ключами між сторонами.

- Ефективність: Гібридні шифри поєднують високу безпеку симетричного шифрування з ефективністю асиметричного шифрування. Оскільки симетричне шифрування є швидким процесом, використання симетричного шифру для шифрування даних дозволяє забезпечити швидку обробку великих обсягів інформації. Асиметричне шифрування, яке використовується лише для обміну симетричним ключем, не потребує великих обчислювальних ресурсів.

- Гнучкість: Гібридні шифри є гнучкими і можуть адаптуватись до різних потреб та вимог безпеки. Розмір симетричного ключа може бути налаштований для досягнення потрібного рівня безпеки, враховуючи швидкість обробки та розмір передаваної інформації.

Гібридні шифри в криптосемантиці є потужним інструментом для забезпечення безпеки інформації. Вони поєднують в собі переваги симетричного та асиметричного шифрування, що дозволяє забезпечити високий рівень безпеки, ефективність та гнучкість. Гібридні шифри є основою багатьох сучасних протоколів захисту даних, таких як SSL/TLS у веб-безпеці та PGP у електронній пошті. Розуміння принципів та переваг гібридних шифрів допомагає розробникам та користувачам вибрати найкращий метод захисту інформації залежно від конкретних потреб та вимог безпеки.

ВИСНОВКИ ДО РОЗДІЛУ 1

У цьому розділі були розглянуті основні цілі та завдання криптографії, криптографічні методи знайшли широке застосування у практичній інформатиці на вирішення численних проблем інформаційної безпеки.

Було порівняні симетричні та асиметричні алгоритми криптографії. У симетричних системах відправник і одержувач використовують один і той самий ключ, який має бути відомий тільки їм. Симетричні алгоритми поділяються на два підкласи. Одні алгоритми обробляють відкритий текст побітово (іноді побайтово). Такі алгоритми називають потоковими. Інші алгоритми обробляють групи (блоки) бітів відкритого тексту. Ці алгоритми називають блоковими. У сучасних комп'ютерних алгоритмах типовий розмір блоку становить 64 біти. Детально розглянуто блокові та потокові (іноді в літературі їх називають поточні) шифри. Таким чином, основним завданням потокових шифрів є вироблення деякої послідовності (вихідної гами) для шифрування, тобто вихідна гама є ключовим потоком для повідомлення.

Синхронні потокові шифри вразливі до атак на основі зміни окремих біт шифртексту, коли зловмисник може змінити ці біти таким чином, що шифртекст розшифрується так, як йому вигідно.

Асиметрична криптографія вирішує дві проблеми, що пов'язані з практичним використанням симетричних криптосистем (зберігання та обмін ключовою інформацією), ці проблеми стали важливими спонукальними мотиви для розробки принципово нового класу методів шифрування: криптографії з відкритим ключем, або асиметричної криптографії.

РОЗДІЛ 2
АНАЛІЗ СУЧАСНОГО СТАНУ ПРОБЛЕМИ ЗАХИСТУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

2.1. Огляд основних методів криптографічного захисту

У цьому підрозділі роботи проводиться огляд основних методів криптографічного захисту, які використовуються для забезпечення конфіденційності, цілісності та аутентифікації текстової інформації. Розглядаються такі методи:

1. Симетричне шифрування:

- Опис принципу симетричного шифрування та його основні переваги та недоліки.
- Популярні алгоритми симетричного шифрування, такі як DES, AES, ідея Feistel, режими шифрування ECB, CBC, CTR тощо.
- Аналіз стійкості симетричних алгоритмів до атак, включаючи криптоаналітичні методи, лінгвістичний аналіз, вибіркову атаку тощо.

2. Асиметричне шифрування:

- Опис принципу асиметричного шифрування (з використанням ключа шифрування та ключа розшифрування) та його переваги та недоліки.
- Популярні алгоритми асиметричного шифрування, такі як RSA, DSA, ElGamal, Елліптична криптографія тощо.
- Методи забезпечення конфіденційності та цілісності інформації, використовуючи асиметричне шифрування.

3. Хеш-функції:

- Опис принципу роботи хеш-функцій та їх використання для забезпечення цілісності даних.
- Популярні хеш-функції, такі як MD5, SHA-1, SHA-256 тощо.
- Використання хеш-функцій для цифрового підпису, валідації даних та інших застосувань.

4. Цифрові підписи:

- Опис принципу роботи цифрових підписів та їх викор

истання для аутентифікації даних та забезпечення неденіабельності.

- Популярні алгоритми цифрових підписів, такі як RSA, DSA, ECDSA тощо.

- Аналіз безпеки цифрових підписів та методів підробки.

5. Протоколи захисту інформації:

- Огляд протоколів, які використовуються для забезпечення безпеки комунікацій та обміну даними, таких як SSL/TLS, SSH, IPsec, PGP/GPG тощо.

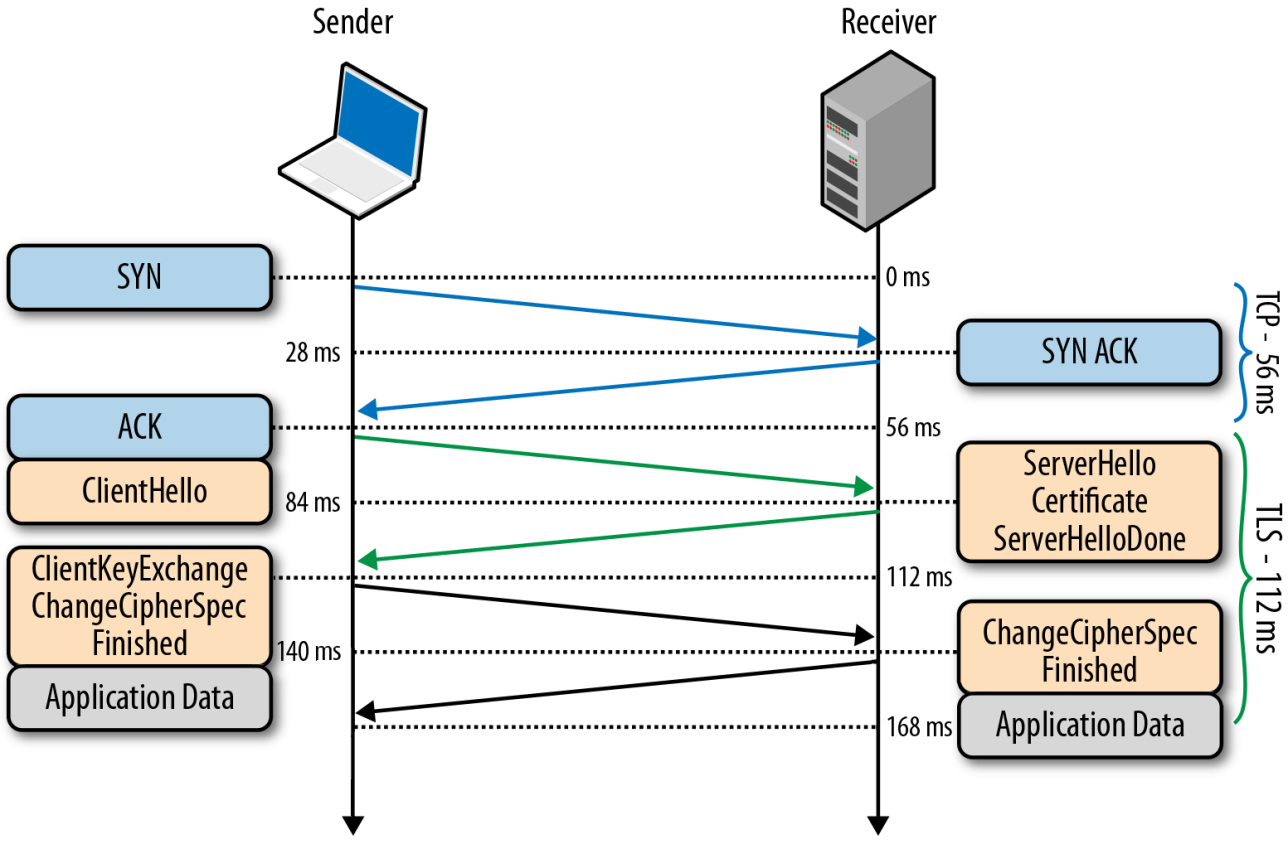

Рис. 2.1. Схема роботи протоколу SSL/TLS

- Аналіз протоколів з точки зору безпеки та стійкості до атак.

Далі робота буде зосереджена на криптосемантичному методі захисту, який розглядається в подальших підрозділах.

При дослідженні звертається увага на наступні аспекти:

1) Ключовий розподіл:

- Аналіз проблем, пов'язаних з безпечним обміном секретних ключів між комунікуючими сторонами.

- Розгляд різних методів для забезпечення безпеки ключового розподілу, таких як використання асиметричних алгоритмів для обміну симетричними ключами, протоколи обміну ключами на основі довірених сторін або використання схеми диффі-хелмана.

- Аналіз вразливостей, що виникають при використанні слабких ключів або неправильних методів генерації ключів.

- Розгляд недоліків в механізмах зберігання та управління ключами, включаючи проблеми з управлінням, резервуванням та обновленням ключів.

2) Криптоаналітичні атаки:

- Детальний огляд різних типів криптоаналітичних атак, таких як переборний пошук, атака з відомим текстом, атака з вибірковим текстом, атака з вибірковим шифртекстом.

- Розгляд вразливостей, які можуть призводити до компрометації ключа або отримання оригінального тексту без додаткової інформації.

- Аналіз різних методів проти-атак, таких як стійке хешування, використання падаючих ключів, застосування аутентифікації повідомлень, криптографічного заміщення тощо.

3) Обчислювальна складність:

- Дослідження важкості обчислювальних операцій, які використовуються в шифрувальних алгоритмах, зокрема шифрування, розшифрування та підписування.

- Аналіз сучасних обчислювальних ресурсів та їх впливу на ефективність застосування шифрування в реальних умовах.

- Розгляд використання розподіленого обчислення, спеціалізованих пристроїв або обчислювання в хмарних середовищах для покращення обчислювальної складності шифрування.

4) Розмір ключа та швидкість:

- Дослідження залежності між розміром ключа та швидкістю шифрування/розшифрування.

- Розгляд оптимального розміру ключа, який забезпечує потрібний рівень безпеки, при цьому зберігаючи прийнятну швидкість обробки даних.

- Аналіз впливу зростання розміру ключа на вимоги до ресурсів (пам'яті, обчислювальної потужності) та величину шифрованих блоків.

Цей детальний аналіз допомагає виявити проблеми та недоліки, пов'язані з методами шифрування, і надає підґрунтя для подальшого розвитку криптосемантичного методу захисту текстової інформації.

2.2. Аналіз проблем криптосемантичного методу захисту тексту

Криптосемантичний метод захисту текстової інформації є широко використовуваним і ефективним способом забезпечення безпеки даних. Однак, жоден метод не є ідеальним, і криптосемантичний метод також має свої проблеми та недоліки. У цій роботі ми розглянемо деякі з них та проаналізуємо можливі шляхи вирішення цих проблем.

1) Проблема безпеки ключа:

Однією з найбільших проблем криптосемантичного методу є безпека ключа. Правильний і безпечний обмін ключами між відправником і отримувачем є критично важливим для забезпечення конфіденційності даних. Якщо зловмисники отримають доступ до ключа, вони можуть розшифрувати повідомлення та отримати доступ до конфіденційної інформації.

Один з можливих шляхів вирішення цієї проблеми - використання криптографічних протоколів для безпечного обміну ключами, таких як протокол Діффі-Хеллмана або RSA. Ці протоколи забезпечують захист ключів під час їх передачі через незахищені канали.

2) Вразливість до криптоаналітичних атак:

Криптосемантичний метод не є імунним до криптоаналітичних атак, які спрямовані на розкриття шифрування і отримання доступу до повідомлення. Наприклад, атаки типу "чоловік посередині" можуть використовувати проміжні вузли для перехоплення та зміни переданих даних.

Для подолання цих проблем необхідно застосовувати сильні криптографічні алгоритми шифрування та підпису, такі як AES (Advanced Encryption Standard) або

SHA (Secure Hash Algorithm). Також важливо використовувати додаткові механізми захисту, такі як цифрові підписи та аутентифікація, для перевірки цілісності даних та підтвердження автентичності відправника [3].

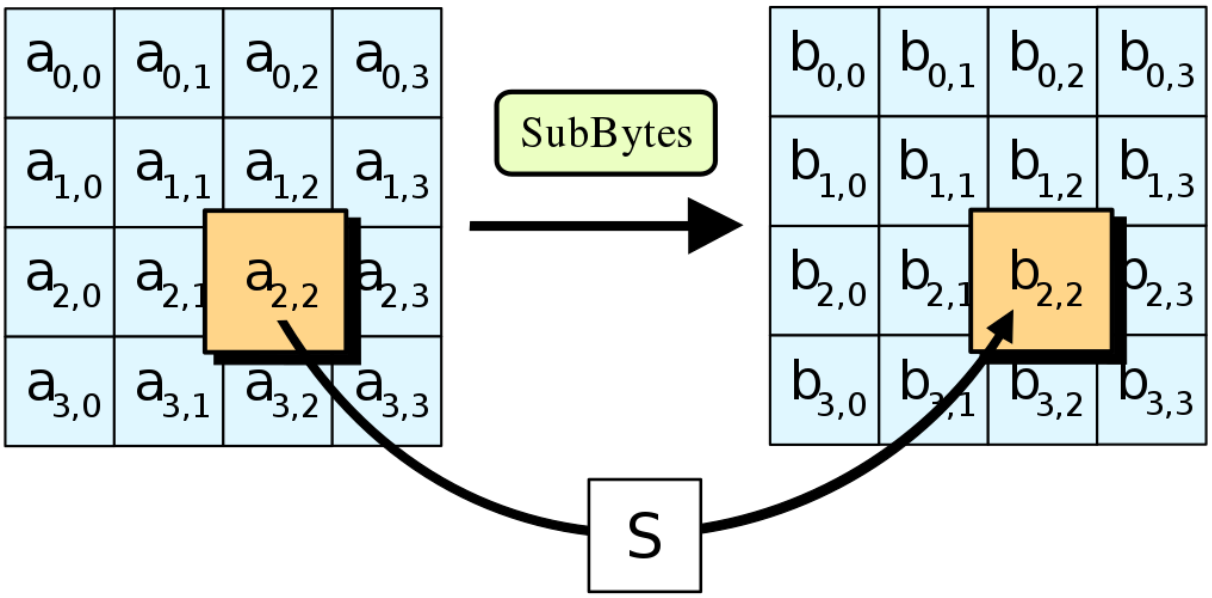

Рис. 2.2. Стандарт шифрування AES

3) Обмежена масштабованість:

Криптосемантичний метод може мати обмежену масштабованість при обробці великого обсягу текстової інформації. Це може бути проблемою в сучасному світі, де великі об'єми даних передаються і обробляються щодня.

Для розв'язання цієї проблеми можна використовувати паралельні обчислення та дистриб'ютивні системи, які дозволяють розподілити обробку даних між багатьма обчислювальними вузлами. Також варто розглянути використання більш ефективних алгоритмів шифрування та компресії даних для зменшення обсягу інформації, яка потребує обробки.

Криптосемантичний метод захисту текстової інформації є потужним інструментом для забезпечення безпеки даних. Однак, він також має свої проблеми, такі як безпека ключа, вразливість до криптоаналітичних атак та обмежена масштабованість.

Шляхи вирішення цих проблем включають використання безпечних протоколів

обміну ключами, сильних криптографічних алгоритмів та масштабованих систем обробки даних. При належному застосуванні та постійному оновленні метод може забезпечити високий рівень безпеки для текстової інформації.

2.3. Принципові властивості криптосемантичних методів

Криптосемантичні методи є важливими засобами захисту конфіденційної інформації, особливо у сучасному цифровому світі, де передача даних відбувається швидко і постійно. У цій роботі ми розглянемо принципи та принципові властивості криптосемантичних методів, які забезпечують їх ефективність і безпеку.

Один з основних принципів криптосемантичних методів полягає в забезпеченні конфіденційності даних. Шифрування повідомлень забезпечує їхню нерозбірливість для несанкціонованих осіб, що гарантує, що тільки правильний отримувач зможе отримати доступ до зрозумілого тексту.

Цілісність є ще одним важливим принципом криптосемантичних методів. Це означає, що при передачі даних повідомлення не можуть бути змінені без виявлення таких змін. Використання криптографічних хеш-функцій та цифрових підписів дозволяє виявляти будь-які зміни в повідомленнях і гарантує їхню цілісність.

Принцип аутентичності визначає, що отримувач може підтвердити автентичність відправника повідомлення. Це досягається за допомогою цифрових підписів, які підтверджують, що повідомлення було створено відправником і не було змінено під час передачі.

Принципові властивості криптосемантичних методів:

1) Алгоритмічна складність

Криптосемантичні методи базуються на алгоритмах шифрування, які є обчислювально складними для виконання. Це означає, що зловмиснику буде дуже складно розкрити шифрування без знання правильного ключа. Алгоритмічна складність гарантує ефективність криптосемантичних методів і забезпечує високий рівень безпеки.

2) Ключова довіреність

У криптосемантичних методах ключ використовується для шифрування та розшифрування повідомлень. Ключова довіреність вимагає, щоб тільки відправник і отримувач мали доступ до правильного ключа. Це гарантує, що лише вони зможуть прочитати та розшифрувати повідомлення.

2) Розширюваність

Криптосемантичні методи повинні бути розширюваними для використання в різних ситуаціях та забезпечення безпеки різних видів даних. Вони повинні мати можливість шифрувати і розшифровувати текстову інформацію різного обсягу та формату.

4) Резистентність до атак

Криптосемантичні методи повинні бути стійкими до різних видів атак, таких як перехоплення повідомлень, аналіз частоти, атаки типу "чоловік посередині" та інші. Вони повинні забезпечувати високий рівень безпеки навіть у вразливих сценаріях атак.

Криптосемантичні методи є важливими інструментами для захисту конфіденційної інформації. Вони базуються на принципах конфіденційності, цілісності та аутентичності, а також володіють принциповими властивостями, такими як алгоритмічна складність, ключова довіреність, розширюваність та резистентність до атак. Розуміння цих принципів та властивостей допоможе в розробці та використанні ефективних криптосемантичних методів для забезпечення безпеки інформації. При належному застосуванні ці методи забезпечують високий рівень безпеки та конфіденційності даних.

2.4. Підходи до реалізації криптосемантичного методу захисту інформації

Криптографія є галуззю науки, що досліджує методи захисту інформації шляхом шифрування та розшифрування даних. Одним з новітніх розробок в цій галузі є криптосемантичний метод захисту інформації. У цьому творі ми детально розглянемо математичні моделі та підходи до реалізації цього методу, що дозволить зрозуміти його принципи та потенціал для захисту інформації.

Криптосемантичний метод базується на математичних моделях, які описують процес шифрування та розшифрування даних. Один з найвідоміших підходів до криптосемантичного шифрування - це використання блочних шифрів. Вони дозволяють розбити дані на блоки та застосовувати до них операції шифрування та розшифрування. При цьому використовуються різні алгоритми та ключі для забезпечення безпеки даних.

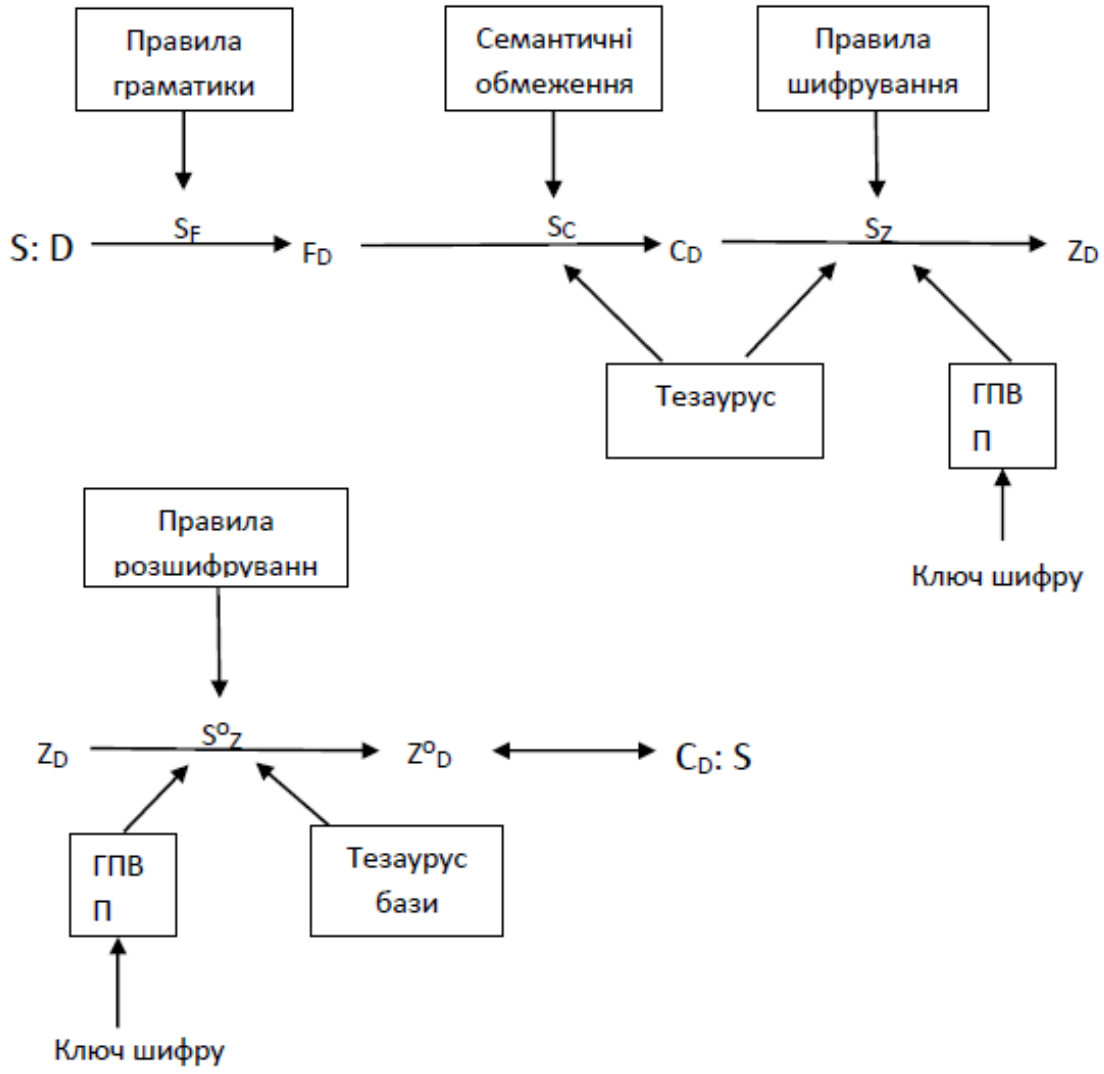

Рис. 2.3. Алгоритм реалізації лексикографічної криптосистеми

Для реалізації криптосемантичного методу захисту інформації існує кілька підходів. Один з них - використання симетричних алгоритмів шифрування, таких як AES або DES. Ці алгоритми базуються на однаковому ключі для шифрування та розшифрування даних. Інший підхід – використання асиметричних алгоритмів, таких

як RSA або ECC. Вони використовують пару ключів - публічний та приватний - для шифрування та розшифрування даних.

Криптосемантичний метод відрізняється від інших методів захисту інформації своїми особливостями. Він забезпечує не тільки конфіденційність даних, але і зберігання семантики, тобто смислу та значення даних. Це дозволяє зберігати цілісність та коректність інформації навіть під час її шифрування та передачі.

Криптосемантичний метод захисту інформації має широкий потенціал в різних сферах застосування. Він може бути використаний для захисту електронної пошти, онлайн-комерції, хмарних систем, мобільних додатків та багатьох інших. Цей метод дозволяє забезпечити безпеку даних навіть в умовах активних кібератак та загроз.

Математичні моделі та підходи до реалізації криптосемантичного методу захисту інформації відіграють важливу роль у забезпеченні безпеки даних. Розуміння цих моделей та підходів дозволяє розробляти ефективні системи захисту, які забезпечують не тільки конфіденційність, але і цілісність та доступність даних. Криптосемантичний метод відкриває нові можливості для безпечного обміну інформацією в цифровому світі та захисту наших даних від кіберзагроз.

Криптосемантичні алгоритми являють собою важливий компонент криптосемантичного методу, який поєднує поля криптографії та семантики для забезпечення підвищеної безпеки та конфіденційності текстової інформації. Ці алгоритми використовують криптографічні методи та семантичний аналіз для захисту конфіденційних даних від несанкціонованого доступу та забезпечення їх цілісності під час передачі та зберігання.

1) Методи шифрування:

Шифрування відіграє фундаментальну роль у криптосемантичних алгоритмах, перетворюючи відкритий текст на зашифрований, роблячи його нечитабельним для неавторизованих осіб. Можна використовувати різні методи шифрування, включаючи шифрування з симетричним ключем і шифрування з асиметричним ключем.

1.1) Шифрування з симетричним ключем:

Шифрування з симетричним ключем, також відоме як шифрування з секретним

ключем, використовує один ключ як для процесів шифрування, так і для дешифрування. Ключ зберігається конфіденційно та використовується між уповноваженими сторонами. Алгоритми симетричного ключа, такі як Advanced Encryption Standard (AES) і Data Encryption Standard (DES), зазвичай використовуються в криптосемантичних методах для захисту текстової інформації.

1.2) Шифрування з асиметричним ключем:

Шифрування з асиметричним ключем, яке також називають шифруванням з відкритим ключем, використовує пару ключів: відкритий ключ для шифрування та закритий ключ для дешифрування. Відкритий ключ можна поширювати відкрито, а закритий ключ надійно зберігається. Алгоритми асиметричного ключа, такі як RSA (Rivest-Shamir-Adleman) і криптографія еліптичної кривої (ECC), використовуються в криптосемантичних алгоритмах для встановлення безпечних каналів зв'язку та захисту текстових даних.

У криптосемантичних алгоритмах генерація ключів шифрування включає семантичні елементи для підвищення безпеки системи. Замість того, щоб покладатися виключно на випадкові чи заздалегідь визначені ключі, методи генерації ключів на основі семантики використовують невід'ємне значення та структуру текстових даних для створення надійних і безпечних ключів шифрування.

Алгоритми генерації семантичних ключів використовують обробку природної мови (NLP) і семантичний аналіз для вилучення семантичних особливостей із відкритого тексту. Потім ці функції перетворюються на криптографічні ключі за допомогою математичних функцій. Включаючи семантичну інформацію в процес генерації ключів, криптосемантичні алгоритми можуть генерувати ключі, стійкі до атак грубої сили, і забезпечують додатковий рівень безпеки [4].

Обробка природної мови (NLP) — це підгалузь штучного інтелекту, яка фокусується на інтерпретації комп'ютерної лінгвістики. Ця сфера охоплює кілька областей інтерпретації тексту та аудіо шляхом інтеграції методів машинного навчання, які поводяться статистично. Він також охоплює сферу прагматичного дослідження комп'ютерної лінгвістики, яка стала дуже широкою та потужною завдяки застосуванню різноманітних методів. Зростання доступності та можливостей

технік NLP, які покращують точність обчислювальної мови та вдосконалюються з кожним днем. NLP і машинне навчання є найбільш фокусованими сферами досліджень. На NLP в основному впливають інші галузі, такі як психологія, когнітивна наука, лінгвістика та багато інших областей. Це стосується обчислювальних моделей інженерії, які обробляються для вирішення людської взаємодії та розуміння людської мови. Для цього розроблено декілька програмних пакетів для областей моделювання мови для інтерпретації обчислювальної мови, яку людина може легко інтерпретувати.

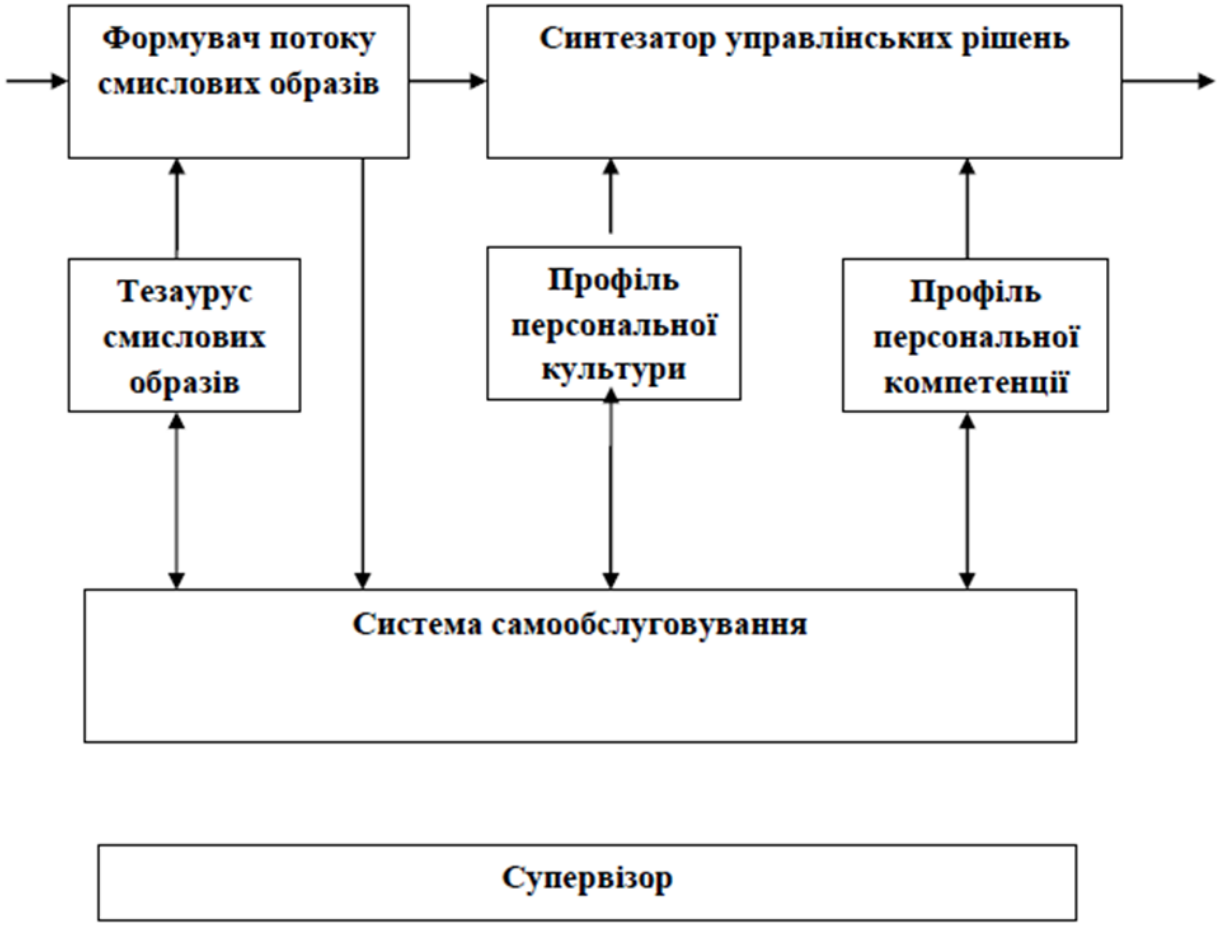

Рис. 2.4. Інформаційна модель інтелекту

Криптосемантичні алгоритми також вирішують проблему безпечного пошуку інформації, гарантуючи, що авторизовані користувачі можуть отримати доступ до захищеної текстової інформації, одночасно запобігаючи неавторизованим особам від

отримання доступу. Методи семантичного аналізу використовуються для безпечного індексування та отримання зашифрованих даних.

Алгоритми пошуку інформації на основі семантики використовують схеми індексування, які фіксують семантичні зв’язки між різними текстовими елементами. Ці алгоритми дозволяють авторизованим користувачам виконувати пошук за ключовими словами в зашифрованих даних без шкоди для їх безпеки. Процес пошуку включає зіставлення зашифрованих ключових слів запиту із зашифрованим індексом, що забезпечує безпечний і ефективний пошук інформації.

ВИСНОВКИ ДО РОЗДІЛУ 2

Підсумовуючи, криптосемантичні алгоритми відіграють важливу роль у криптосемантичному методі, поєднуючи криптографічні методи та семантичний аналіз. Ці алгоритми забезпечують безпечне шифрування, генерацію семантичного ключа та безпечний пошук інформації, забезпечуючи ефективний підхід до захисту текстової інформації від несанкціонованого доступу та забезпечення її конфіденційності в різних програмах.

Криптосемантичні алгоритми знайшли застосування в різних областях, включаючи фінанси, охорону здоров'я та державні установи. Поєднуючи криптографію та семантику, ці алгоритми забезпечують потужне рішення для захисту конфіденційної текстової інформації, підтримки цілісності даних і забезпечення безпечного пошуку інформації.

Криптосемантичний метод базується на математичних моделях, які описують процес шифрування та розшифрування даних. Один з найвідоміших підходів до криптосемантичного шифрування - це використання блочних шифрів. Вони дозволяють розбити дані на блоки та застосовувати до них операції шифрування та розшифрування. При цьому використовуються різні алгоритми та ключі для забезпечення безпеки даних.

Криптосемантичний метод відрізняється від інших методів захисту інформації своїми особливостями. Він забезпечує не тільки конфіденційність даних, але і зберігання семантики, тобто смислу та значення даних. Це дозволяє зберігати цілісність та коректність інформації навіть під час її шифрування та передачі.

РОЗДІЛ 3
КРИПТОСЕМАНТИЧНИХ МЕТОД ЗАХИСТУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

3.1. Схема реалізації криптосемантичного методу

Метод конфіденційності інформації, який розглядається, належить до криптографічних методів симетричного шифрування текстових даних, представлених у комп'ютеризованому вигляді будь-якою з мов спілкування (природною людською чи формально визначеною штучною мовою), для якої існує відповідна лексикографічна система (зокрема, створено лінгвістичний корпус). Цей метод називається криптосемантичним методом (CSM). Передумовою використання CSM є наступне: семантичний тезаурус (тобто належним чином структурована семантична система лексиконів), що охоплює предметну сферу використання даного методу захисту, повинен бути включений до лексикографічної системи. Лінгвістичний корпус визначається як програмний інструмент, який автоматично розділяє електронний текст, що підлягає захисту, на «мікроконтексти» – фрагменти тексту, згруповані навколо мовних одиниць (зокрема, слів, фраз, шрифтів тощо), які є об'єктами інтерпретації [7,8]. Мовний тезаурус — це лексикон мови з визначеними семантичними відношеннями між мовними одиницями. У даному випадку тезаурус – це ієрархічна структура семантичних лексиконів, що визначають семантичні відношення між мовними одиницями мови, що використовуються для відображення текстових повідомлень вибраної сфери прикладного використання.

В третьому розділі демонструються результати створення криптосемантичного методу захисту текстових даних, що характеризується принаймні двома суттєвими перевагами: CSM за певних умов забезпечує необмежену криптостійкість, але, водночас, не погіршує якість функціонування прикладні комп'ютерні системи, інформаційні ресурси яких підлягають захисту. Практична реалізація цього методу цілком можлива, якщо для його побудови використовується теорія лексикографічних схем і сфера використання обмежується захистом даних, що містяться у наперед

визначених табличних формах.

В кваліфікаційній роботі розглядається метод крипто-семантичного захисту, який використовується, коли текстова інформація, яка підлягає шифруванню, представлена в таблиці довільного типу.

На рис. 1 наведена одна з можливих моделей системи захисту даних CSM, принцип функціонування якої базується на синхронізації генераторів псевдовипадкової послідовності (ГПВЧ), розташованих на передавальній та приймальній сторонах захищеного каналу обміну за допомогою відомий ключ шифрування.

Модель на рисунку 1 містить усі основні елементи криптографічної системи [5,9]. Однак, крім цього, в блок-схему включений додатковий елемент – тезаурус предметної області, в рамках якої планується використовувати систему захисту CSM. Тезаурус створюється на основі результатів статистичного та семантичного аналізу предметної області та має включати всі лінгвістичні одиниці, які потенційно можуть бути включені до відкритого тексту, вставленого у попередньо визначену форму таблиці. Маючи під рукою тезаурус предметної області, відомий пароль і попередньо визначену форму таблиці, адміністратор прикладної системи на кінці передачі формує оригінальний текст і надсилає його механізму шифрування (або програмному кодеру). За допомогою ключової інформації, тобто пароля, ГПВЧ встановлюється в певний початковий стан і надсилає згенеровану послідовність на приймальну сторону механізму шифрування. Зашифрований текст, отриманий в результаті роботи шифрувальника, зберігається на комп’ютері та (або) передається на приймаючу сторону по відкритому каналу зв’язку. Адміністратор прикладної системи на приймальній стороні відправляє отриманий зашифрований текст на машину декодування та за допомогою відомого пароля встановлює початковий стан ГПВЧ. Машина для декодування створює оригінальний текст, сформований на стороні передачі.

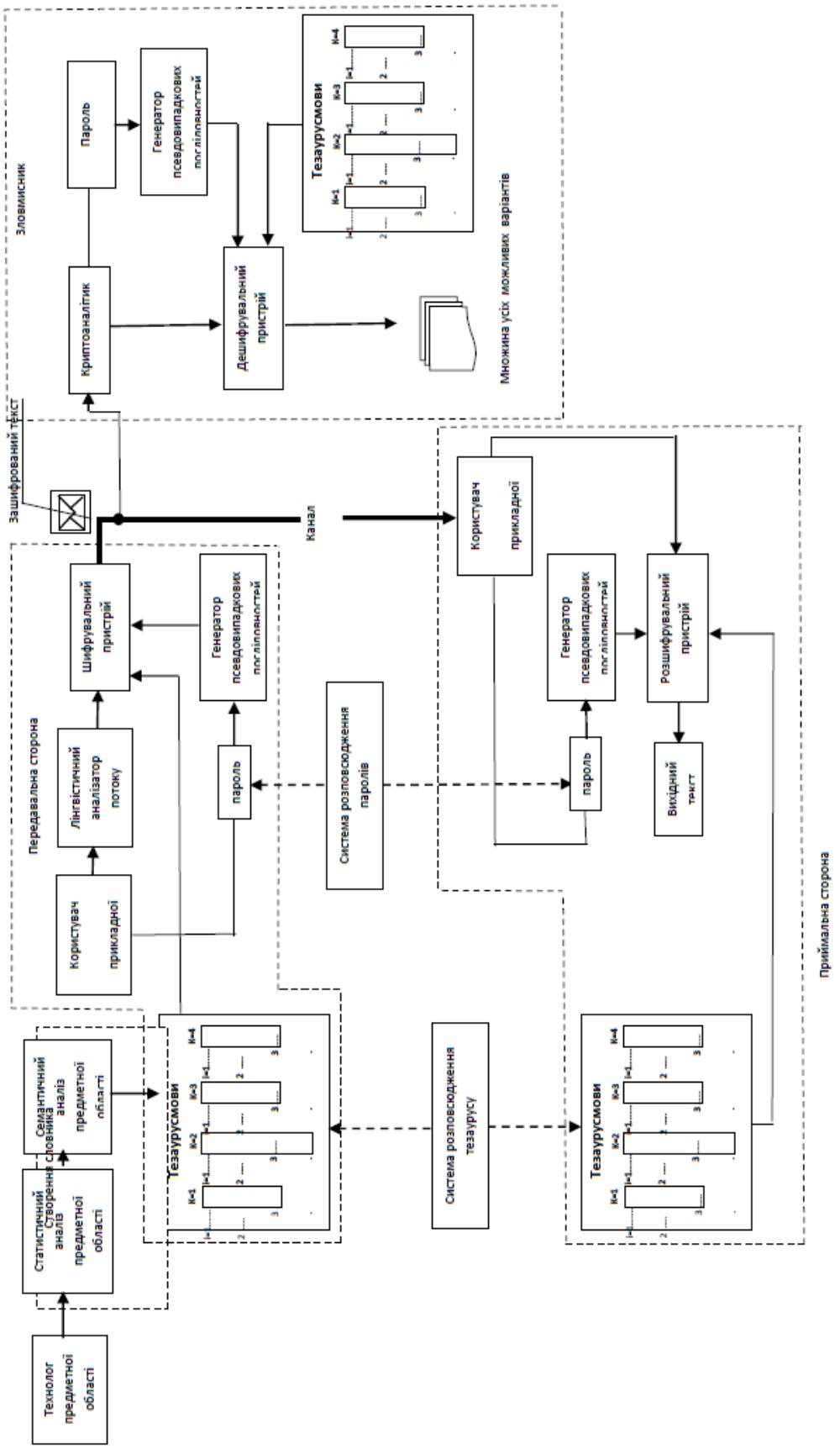

Рис. 3.1. Базовий варіант реалізації методу побудови лексикографічної криптосистеми

Криптосемантичний метод захисту інформації передбачає обов'язкове використання особливо структурованих лексиконів мов, на яких відображається інформація, що захищається. Такі лексикони називають тезаурусами цих мов. Структура зв’язків між лінгвістичними елементами тезаурусів повинна забезпечувати можливість приховати справжній семантичний зміст інформації, що захищається, у ствреному потоці так званого правдоподібного «семантичного шуму», роль якого відіграє семантичний зміст зашифрованого тексту.

Щоб криптоаналітики не могли оцінити рівень успішності реалізованих ними атак на шифртекст, необхідно надати шифртексту ознак правдоподібності. Як ознак правдоподібності зашифрованого тексту доцільно обрати ознаки логічної несуперечності та/або формальної правдоподібності його змісту.

Умова формальної правдоподібності передбачає створення ситуації, коли:

1) передбачений результат контент-аналізу зашифрованого тексту ідентифікує відмінну від нуля ймовірність подій або ненульову ймовірність достовірності інформації, що відображається в цьому зашифрованому тексті;

2) в умовах відсутності апріорних знань можна припустити, що з точки зору порушника інформація, що міститься в зашифрованому тексті, має приблизно такий же рівень достовірності, як і багато іншої інформації з сукупності можливих. види інформації в межах даної прикладної області. Зрозуміло, що чим вищий рівень формальної правдоподібності досягається в процесі побудови тезауруса мови додатка, тим менше у порушника можливостей успішно розшифрувати перехоплений текст.

На відміну від стенографічних методів, спрямованих на приховування самого факту передачі інформації, що підлягає захисту, механізм захисту CSM може використовуватися в умовах, коли порушник усвідомлює, що об’єктом його аналізу є зашифрований текст, а не відкритий текст. Однак аналіз зашифрованого тексту, отриманого за допомогою перетворень CSM, позбавляє порушника можливості відрізнити справжній зміст захищеної інформації від правдоподібного вмісту зашифрованих текстів у розпорядженні порушника, оскільки відсутній критерій для виявлення такої різниці. Через це порушник не може дійти обґрунтованих висновків

щодо рівня правдивості інформації шифротексту за умови відсутності апріорної інформації про зміст інформації, що захищається, і, природно, інформація про використані секретні ключі шифрування.

3.2. Синтез тезаурусу та схема структурування

Перший крок у процедурі створення механізму захисту CSM, застосування якого обмежене певною сферою застосування, зводиться до виконання лексикографічного завдання складання та структурування лексикону. Цей лексикон, з одного боку, має відображати повний тезаурус (тобто семантичну основу) обраної області застосування, а з іншого боку, має бути структурований таким чином, щоб набір елементів тезауруса був розділений на на певні підмножини формально правдоподібних логічно правильних конструкцій (тобто конструкцій FSP) обраної мови відображення інформації.

Запропоновано наступну узагальнену схему синтезу та структурування тезауруса для випадку, коли необхідно організувати захищений обмін інформацією в певно визначеній прикладній сфері застосування.

1) Вирішується лексикографічне завдання складання лексикону мови прикладної області, який комплексно відображає тезаурус цієї прикладної області. Записи цього лексикону є мовними одиницями та/або конструкціями мови відображення інформації, що підлягає захисту. Набір мовних одиниць і/або конструкцій у скінченному вигляді. За цих умов набір записів лексикону мови області застосування відображається у вигляді підмножини в наборі записів лексикону мови, що використовується для людського спілкування.

2) Набір записів лексикону мови прикладної області розділено (на основі семантичного аналізу мови прикладної області) на n підмножин, кожна з яких містить елементи, які формально семантично пов’язані (тобто кожна підмножина в лексиконі повинна містять елементи FSP лише одного свого виду). За умови такого поділу лексики кожна з отриманих підмножин включає лише ті мовні одиниці (або мовні конструкції), між якими існує однозначно визначений семантичний взаємозв’язок.

Такий взаємозв’язок можна визначити (але не обов’язково) за допомогою більш абстрактної лінгвістичної одиниці та/або конструкції, взятої зі словника визначень тієї ж мови. Кожна підмножина елементів FSP у вибраній області застосування має бути повною, тобто містити всі мовні одиниці або мовні конструкції даної лексики, які перебувають у зв’язках FSP з елементами цієї підмножини. Коли лексикон розділено таким чином, під позначенням n розуміється кількість різних типів FSP у тезаурусі вибраної області застосування.

3) Кожному встановленому виду формального семантичного зв’язку між елементами лексики надається однозначне (унікальне) визначення у вигляді мовної одиниці або мовної конструкції обраної мови, якою відображається інформація, що підлягає захисту. Кількість таких визначень має дорівнювати n. По суті, кожне з визначень є ідентифікатором відповідної підмножини елементів FSP (або ідентифікатором виду формального семантичного зв'язку між елементами лексикону).

4) Тезаурус обраної мови прикладної області синтезується у вигляді двох однозначно взаємопов’язаних лексиконів: один лексикон містить ідентифікатори видів FSP, а інший лексикон складається з N сублексиконів, кожен з яких є структурованою повною підмножиною FSP. елементів. Таким чином, маючи тезаурус для обраної предметної області, структурований вищеописаним чином, можна організувати безпечний обмін інформацією на основі методу захисту CSM. Для цього всі учасники такого обміну повинні використовувати однаковий тезаурус.

Процедура шифрування заснована на заміні істинного елемента простої послідовності лінгвістичних одиниць на маскувальний елемент захищеної послідовності. Припустимо, що тезаурус складається з n сублексиконів, де n — кількість стовпців у визначеній формі таблиці. Кожен сублексикон використовується для обробки даних з одного стовпця, тому кількість сублексикій тезауруса, які використовуються для шифрування/дешифрування, дорівнює кількості стовпців визначеної форми таблиці. Виберіть сублексикон, елементи якого типу FSP збігаються з видом FSP першого елемента вхідних простих послідовностей мовних одиниць. Припустимо, що це k-й сублексикон з n сублексиконів, які складають

тезаурус. Усі елементи цієї підлексики пронумеровані. Припустимо також, що розташування першого справжнього елемента вхідної послідовності має порядковий номер у структурі вибраного сублексикона Nor, а перше випадкове число, надане ГПВЧ, дорівнює Nran, де 0 < Nran < pk , pk є розмір (тобто загальна кількість слів) k-го сублексикону. Тоді порядковий номер розташування в структурі k-го сублексикону «несправжнього» (тобто маскуючого) елемента мови, розміщеного як перший елемент у вихідній зашифрованій послідовності замість справжнього першого елемента мови, N_{enc} дорівнюватиме:

$$N^i_{k\,enc} = mod_{p_k}\left(N^i_{k\,or} + mod_{p_k}\left(N^i_{k\,ran}\right) + p_k\right) \tag{3.1}$$

Вираз (1) вірний, коли розмірність використаного сублексикону збігається з розкидом випадкових послідовностей, згенерованих ГПВЧ (розкид є різницею між максимальним і мінімальним значенням псевдовипадкових чисел у послідовності). У загальному випадку, коли поширення ГПВЧ дорівнює розмірності найбільшого сублексикону, рівняння шифрування, яке є формулою для визначення порядкового номера розташування маскуючого слова маскувального твору в k-му сублексиконі тезауруса, має наступне: форма:

$$N_{enc} = mod_{p_k}\left(N_{or} + mod_{p_k}(N_{ran})\right) \tag{3.2}$$

де N_{or} є порядковим номером розташування в k-му сублексиконі i-го справжнього мовного елемента, взятого з оригінальної простої послідовності лінгвістичних одиниць, що підлягають шифруванню, де i є порядковим номером розташування цього елемента в простій послідовності;

N_{enc} — порядковий номер розташування в k-му сублексиконі маскуючого елемента, поставленого на i-ту позицію (замість справжнього елемента) вихідної послідовності зашифрованої мовної одиниці;

N_{ran} — псевдовипадкове ціле число, згенероване ГПВЧ на і-му етапі генерації для шифрування і-го істинного елемента з порядковим номером N_{or}; pk – розмірність k-го сублексикону;

k — порядковий номер сублексикону в тезаурусі мови області застосування, елементи якого з'єднані загальним FSP з і-м мовним елементом вихідної простої послідовності.

Рівняння для розшифровки інформації, зашифрованої за допомогою будь-якого механізму CSM, є формулою для визначення порядкового номера розташування у вибраному сублексиконі (сублексикон вибирається з n набору сублексиконів використовуваного тезауруса на основі FSP їх елементів). з елементами оригінальної зашифрованої послідовності) і-ої справжньої мовної одиниці, розміщеної у вихідній розшифрованій послідовності замість маскуючої лінгвістичної одиниці, взятої з вихідної зашифрованої послідовності.

Рівняння розшифровки має такий вигляд:

$$N_{or} = mod_{p_k}\left(N_{enc} + mod_{p_k}(N_{ran}) + p_k\right) \tag{3.3}$$

де N_{or} — порядковий номер мовної одиниці (відповідно до нумерації в k-му сублексиконі), ідентичної справжній одиниці, відображеній у вихідній простій послідовності мовних одиниць; N_{enc} – порядковий номер зашифрованої мовної одиниці (відповідно до нумерації в k-му сублексиконі), взятий із зашифрованої послідовності, що обробляється, на поточному етапі дешифрування; N_{ran} — це псевдовипадкове число, згенероване ГПВЧ для розшифровки зашифрованої одиниці, що має порядковий номер N_{enc};

p_k — розмірність (тобто загальна кількість слів) k-го сублексикону;

k - порядковий номер сублексикону, обраного на поточному етапі процедури дешифрування. Формула (3) вірна, якщо розповсюдження ГПВЧ дорівнює розмірності сублексикону, що містить найбільшу кількість мовних одиниць серед набору всіх сублексиконів тезауруса.

3.3. Забезпечення випадковості процедури заміни

Зрозуміло, що протягом одного терміну дії пароля за стандартних умов описаний вище механізм заміни не забезпечує випадковості змін, оскільки, ізолюючи елементи зашифрованого тексту, розташовані один від одного точно на один період генерації ГПВЧ, можна виділити групи елементів зашифрованого тексту. в якому ці елементи замінюються на ті самі елементи тезауруса [10]. Це, у свою чергу, дозволяє розраховувати на успішність криптоатак на основі статистичного аналізу появи елементів у перехоплених зашифрованих текстах. Тому умова правдоподібності мовних структур на передавальному кінці кодера недостатня. У таких випадках необхідно забезпечити випадковість процедури заміни істинних мовних структур структурами, взятими з лексиконів тезауруса предметної області. Випадковість підстановки може бути забезпечена, якщо в схему ГПВЧ вставити таблицю для зберігання даних про її поточний стан на момент завершення сеансу зв'язку. У цьому випадку наступний сеанс зв'язку починається з того, що ГПВЧ генерує послідовність не з початкового стану, а з того місця, де генерація псевдовипадкової послідовності була завершена в попередньому сеансі. І якщо протягом одного періоду генерації ГПВЧ забезпечується його випадковість (що цілком можливо на практиці) і за умов, коли кількість рядків у будь-якій формі таблиці значно менше періоду генерації ГПВЧ, схема злому системи шифрування на основі ймовірнісний аналіз частоти появи мовних структур на передавальному кінці кодера не буде ефективним.

Область застосування. Механізми захисту CSM можуть бути реалізовані в класі на будь-яких шифрах заміни. Зокрема, для шифрування текстової інформації можуть бути використані будь-які різновиди поліалфавітних схем підстановки (наприклад, Віженера, Бофора тощо), які не маскують частоту появи символів алфавіту на передавальному кінці кодера і, таким чином, є чутливий до злому, коли використовується ймовірнісний аналіз зашифрованого тексту. Також можливе використання різних варіантів, заснованих на використанні генераторів псевдовипадкової послідовності (ГПВЧ). За стандартних умов такі системи захисту також чутливі до злому, оскільки не забезпечують випадковість заміни символів

вихідної текстової послідовності, коли аналізовані обсяги зашифрованого тексту значно перевищують період згенерованих ГПВЧ-послідовностей. Проте ГПВЧ, побудований з належним урахуванням вищесказаного, забезпечує випадковість заміни символів звичайного тексту символами, взятими з тезауруса предметної області, і таким чином забезпечує стійкість проти атак на основі ймовірнісного аналізу криптограм.

ВИСНОВКИ ДО РОЗДІЛУ 3

В третьому розділі запропоновано метод захисту текстових даних під назвою «криптосемантичний метод». Метод базується на використанні лексикографічних систем у формі прикладних лінгвістичних корпусів у формально визначених обмеженнях вибраних сфер прикладного використання. CSM забезпечує абсолютну гарантію захисту текстових даних від компрометації конфіденційності навіть за умов, коли криптоаналітику доступна досить велика кількість зразків зашифрованої інформації (очевидно більше, ніж обсяг інформації про пароль). Не визначено умов та обмежень, за яких доцільне використання CSM. Відповідних формальних підстав і доказів також не надано. Ця стаття спрямована на усунення цього недоліку. У цій статті визначити рівні криптографічної стійкості, які здатний забезпечити криптосемантичний метод захисту текстових даних, з урахуванням ресурсних можливостей криптоаналітика. Визначити умови, за яких CSM забезпечує абсолютну гарантію захисту текстових даних від порушення конфіденційності.

Механізми захисту CSM можуть бути реалізовані в класі на будь-яких шифрах заміни. Зокрема, для шифрування текстової інформації можуть бути використані будь-які різновиди поліалфавітних схем підстановки (наприклад, Віженера, Бофора тощо), які не маскують частоту появи символів алфавіту на передавальному кінці кодера і, таким чином, є чутливий до злому, коли використовується ймовірнісний аналіз зашифрованого тексту. Також можливе використання різних варіантів, заснованих на використанні генераторів псевдовипадкової послідовності (ГПВЧ). За стандартних умов такі системи захисту також чутливі до злому, оскільки не забезпечують випадковість заміни символів вихідної текстової послідовності, коли аналізовані обсяги зашифрованого тексту значно перевищують період згенерованих ГПВЧ-послідовностей.

ВИСНОВКИ

Запропоновано метод захисту текстових даних, який називається криптосемантичним методом (CSM). Для реалізації методу в рамках формально визначених обмежень вибраної сфери прикладного використання необхідно розробити відповідну лексикографічну систему у вигляді прикладного лінгвістичного корпусу та семантично структурувати інформацію за допомогою сконструйованого лінгвістичного корпусу таким чином, щоб зашифрована зразки текстових повідомлень представляють семантично правдоподібні фрагменти тексту. За певних умов метод забезпечує абсолютну гарантію захисту текстових даних від порушення конфіденційності.

Розглянуто особливості реалізації методу крипто-семантичного захисту для випадку, коли текстова інформація, що шифрується, представлена в таблиці довільного типу. Форма таблиці заздалегідь визначена. Жодної іншої інформації, окрім тієї, що введена в таблицю, немає. Представлено функціональну модель системи захисту табличних даних CSM. Наведено відповідні рівняння шифрування та дешифрування інформації.

Передумовою використання CSM є включення в лексикографічну систему семантичного тезауруса (тобто особливо структурованої системи семантичних лексиконів), що охоплює предметну сферу використання цього методу захисту. Тезаурус є необхідним елементом у механізмі перетворення вихідного простого табличного тексту в зашифрований табличний текст, що має ознаки формальної семантичної вірогідності (FSP).

Слід зазначити, що індекс міри схожості з оригінальним (істинним) текстом має бути випадковою величиною, конкретне значення якої невідомо криптоаналітику. Тезаурус повинен містити всі можливі види лінгвістичних конструкцій, які використовуються для побудови (синтезу) оригінальних відкритих текстів, введених у визначену табличну форму. Вимог до секретності тезаурусу не встановлено. Цей тезаурус має бути доступним для всіх потенційних сторін безпечного обміну та

використовуватися в процесах шифрування на стороні передачі (той, що генерує інформацію) і в процесах дешифрування на стороні прийому (читання).

Розроблено схему синтезу та структурування тезауруса на випадок, коли необхідно організувати захищений обмін інформацією в конкретно визначеній прикладній області застосування. Ця розробка уможливила організацію безпечного обміну інформацією на основі запропонованого механізму CSM.

Основна характеристика криптосемантичних методів шифрування інформації пов'язана з особливостями структурування тезауруса. По-перше, тезаурус повинен адекватно відображати сферу застосування системи захисту CSM, оскільки будь-яка інформація, яка в принципі може потребувати захисту, повинна бути повністю відображена в тезаурусі побудованої системи CSM. По-друге, тезаурус має бути спеціально структурований таким чином, щоб забезпечувати можливість приховати зміст інформації, яка підлягає захисту, у згенерованому потоці правдоподібного семантичного шуму.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Shalini Subramani, Selvi M, Kannan A & Santhosh Kumar Svn (2023) Review of Security Methods Based on Classical Cryptography and Quantum Cryptography, Cybernetics and Systems
2. Sushil Jajodia, Pierangela Samarati, Moti Yung (2019) Encyclopedia of Cryptography, Security and Privacy, 1142 pp.
3. Gupta, H.; van Oorschot, P.C. Onboarding and Software Update Architecture for IoT Devices. In Proceedings of the 17th International Conference on Privacy, Security and Trust, PST 2019, Fredericton, NB, Canada, 26–28 August 2019; pp. 1–11.
4. Basin, D., Cremers, C., Miyazaki, K., Radomirovic, S., & Watanabe, D. (2020, May/June). Improving of the Cryptographic Protocol Standards.
5. Schneier B., Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2nd ed. New York // John Wiley and Sons, 1996.
6. Zandberg, K.; Schleiser, K.; Acosta, F.; Tschofenig, H.; Baccelli, E. Secure Firmware Updates for Constrained IoT Devices Using Open Standards: A Reality Check. IEEE Access 2019.
7. Шеннон К.Е. "Теорія зв'язку в секретних системах". У кн.: Шеннон К.Е. Роботи з теорії інформації та кібернетиці. 1963, с. 332-402, -829 с.
8. Naccache, D.; Stern, J. A New Public-Key Cryptosystem; Lecture Notes in Computer Science; Springer: Germany, 2019
9. IEEE CSP--EI, Scopus 2022: 2022 IEEE 6th International Conference on Cryptography, Security and Privacy (CSP 2022)
10. Munjal, K.; Bhatia, R. A Systematic Review of Homomorphic Encryption and its Contributions in Healthcare Industry. Complex Intell. Syst. 2022, 1–28.
11. A statistical Test Suite for the Validation of Random Number Generators and Pseudo Random Number Generators for Cryptographic Applications [Text]: NIST Special Publication 800-22 Rev1. – Gaithersburg, Maryland: NIST, 2019. – 153 p.