МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

На правах рукопису

Самойлик Є.О.

УДК 004.7

ЛЕКСИКОГРАФІЧНІ МЕТОДИ ЗАХИСТУ МОВНОЇ ІНФОРМАЦІЇ

Спеціальність 05.13.21 – системи захисту інформації

Дисертація на здобуття наукового ступеня кандидата технічних наук

Науковий керівник
Одарченко Р.С.,
д-р техн. наук, доцент

Київ – 2020

АНОТАЦІЯ

Самойлик Є.О. Лексикографічні методи захисту мовної інформації. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.21 «Системи захисту інформації». – Національний авіаційний університет, Київ, 2020.

Наразі розроблено різноманітні практично стійкі криптографічні системи, що знайшли застосування для вирішення широкого спектру прикладних задач, де необхідно забезпечити надійний захист від порушень конфіденційності інформації, що передається відкритими каналами зв'язку. Проте ці криптосистеми не гарантують формальну, теоретично доведену неможливість їхнього злому. Отже, існує проблема недовіри до надійності цих систем в задачах передавання інформації, що характеризуються високими рівнями секретності. Зокрема, неможливо в повній мірі довіряти імпортованим засобам «сильної криптографії» та прикладним системам, що мають у своєму складі такі засоби, захист критично важливої державної інформації, наприклад з грифом „таємно" або „цілком таємно", від атак з боку розвідувальних служб потужних організаційних структур.

Дисертаційна робота присвячена створенню симетричних криптосистем, що не пред'являють жорстких вимог до системи розповсюдження ключової інформації.

Було показано, що методи і моделі, що використовують у задачах теорії інформації і зв'язку, не повною мірою придатні для опису систем семантичної криптографії – систем, призначених для захисту смислового змісту інформації від перехоплення суб'єктами, що не мають відповідних прав легального доступу до неї. Для адекватного опису цих систем у даній роботі уточнені і (або) перевизначені відомі поняття, такі як середовище, суб'єкт, інтелект, семантичний тезаурус, інформація, форма представлення і

сенс (зміст) інформації, а також представлена модель осмислення суб'єктом змісту інформації стосовно завдань семантичної криптографії. В роботі була представлена узагальнена інформаційна модель суб'єкта, що придатна для вирішення завдань семантичної криптографії. Модель відбиває основну життєву функцію суб'єкта – адекватно реагувати на інформацію, що поступає із середовища його існування. Згідно цієї моделі потоки інформації, що сприймаються чутливими елементами сенсорних систем суб'єкта, подаються в реальному часі на обробку засобами інтелекту. Інтелект із сприйнятого інформаційного потоку виділяє потік смислових образів, обробляє цей потік за допомогою наявних інтелектуальних засобів і, в результаті, формує потік управлінських рішень у вигляді потоку сигналів впливу на елементи психофізіологічних систем організму і опорно-рухового апарату суб'єкта.

При цьому, уперше представлена інформаційна модель інтелекту, що визначає його структуру і деталізує його функції в аспектах, що розглядають під час вирішення завдань семантичної криптографії. Інтелект суб'єкта у рамках даної моделі розглядається як програмно-керована машина, що реалізовує в реальному часі комплекс програм підтримки впорядкованої сукупності фізіологічних процесів в організмі суб'єкта. В рамках роботи було пояснено функції шести основних інформаційних модулів цього комплексу, керованих супервайзером – основною системо-утворюючою програмою, що організує функціонування інтелекту по цілеспрямованому аналізу, обробці і зберіганню потоку смислових образів, витягуваних з потоку інформації, що надходить. Функціональність цих модулів передбачає можливість: 1) сприйняття і оцінювання потоків інформації, що надходять від рецепторів сенсорних систем організму, а також формування (за результатами оцінки) і запам'ятовування потоків смислових образів; 2) підтримки і оновлення тезауруса персональних знань суб'єкта; 3) підтримки і оновлення профілю персональної культури суб'єкта (системи персональних моральних критеріїв, переваг і обмежень); 4) підтримки і оновлення профілю персональних компетенцій; 5) формувань управлінських рішень, що визначають фізичну

поведінку суб'єкта, а також фізіологічний і емоційний стан його організму; 6) самонавчань інтелекту, в процесі якого реалізуються процеси вдосконалення тезауруса знань про середовище, профілю персональної культури і профілю персональних компетенцій суб'єкта.

Крім того, було вперше розроблено формальну модель розуміння мовної інформації, відповідно до якої у процесі розумової діяльності суб’єкт генерує детерміновані скінчені дискретні часові ряди смислових образів, що вибираються ним (зокрема, за допомогою асоціативних механізмів) із доступного йому тезаурусу. Вперше тезаурус інтелекту суб’єкта визначено як структурований у вигляді прошаркової коренево-подібної ієрархії набір смислових образів, що упорядковані за рівнями абстрактності цих смислових образів у напрямі усе більш конкретного (більш деталізованого) їхнього сприйняття суб’єктом. Показано, що якість тезаурусу визначається як розмірністю бази смислових образів, якими здатен оперувати суб’єкт у процесі розумової діяльності, так і кількістю рівнів абстрактності відображення смислових образів у структурі тезаурусу.

Таким чином, в роботі було визначено характеристики лексикографічних систем, що є необхідними для створення досконало стійких криптосистем захисту текстової інформації. Показано, що досягнення поставленої в дисертаційній роботі мети слід шукати на шляху розробки нових методів побудови стійких криптосистем, які здатні збільшити відстань єдиності за ключем або, в ідеалі, забезпечити стійкий захист при відсутності будь-яких обмежень на обсяг текстової інформації, що підлягає шифруванню. Якщо збільшення відстані єдиності може бути досягнуто шляхом укрупнення алфавіту джерела відкритих текстових повідомлень, то уникнення виконання шеннонівської умови досконалої стійкості захисту є можливим лише за рахунок використання особливостей семантичних характеристик текстової інформації на основі поєднання методів симетричної криптографії і семантичної обробки змісту текстових повідомлень, зокрема за допомогою певним чином побудованих лексикографічних систем,

реалізованих у вигляді проблемно-орієнтованих лінгвістичних корпусів. У цьому випадку має здійснюватися шифрування не форми відображення структури співвідношень між характеристиками об’єктів, що відображають текстові повідомлення, а безпосередньо сутність (смисл) цієї структури.

Таким чином, основними науковими результатами дисертаційної роботи є: вперше розроблена модель крипто-семантичного словника, яка за рахунок уведення в прикладну лексикографічну систему показників семантичних зв’язків між смисловими конструкціями мови відображення області застосування дозволяє визначити тезаурус бази захисту інформації у прикладній системі та семантичну структуру словників прикладної області; вперше розроблений метод побудови лексикографічної системи захисту мовної інформації, який за рахунок розширення базового алфавіту лінгвістичної формальної системи джерела мовних повідомлень забезпечує збільшення відстані єдиності шифру за ключем та дозволяє збільшити довжину шифрованих повідомлень відносно довжини ключової інформації та відповідно зменшити частоту зміни ключів шифру у порівнянні із схемою Вернама; вперше запропонований лексикографічний метод захисту текстової інформації, що за рахунок випадкової заміни первинного смислового образу повідомлення на інший правдоподібний елемент, узятий із семантичного тезаурусу бази захисту прикладної області, дозволяє забезпечити підвищену стійкість захисту при відсутності будь-яких обмежень на обсяг мовної інформації, що підлягає шифруванню, і, тим самим, усуває необхідність у періодичній зміні ключової інформації.

Теоретичні результати, отримані в дисертаційному дослідженні, відкривають можливість виявити і запропонувати нові практичні шляхи підвищення ефективності шифрування мовної інформації в різноманітних каналах передачі інформації.

При цьому отримані результати дозволяють суттєво спростити систему розповсюдження ключової інформації.

Практична цінність дисертаційної роботи полягає в такому: розроблено методику автоматизації розробки тезаурусу бази захисту інформації у прикладній області під час побудови лексикографічної криптосистеми; запропоновано схему технічної реалізації методу побудови криптосистеми із збільшеною відстанню єдиності за ключем шифру для стійкої системи передавання текстових даних, представлених у вигляді табличних форм; розроблено алгоритмічне та програмне забезпечення крипто-семантичного захисту текстових даних, що засноване на використанні прикладного тезаурусу смислових образів, який здатний забезпечити режим підвищеної стійкості у рамках конкретно визначених прикладних систем.

На базі розроблених в роботі методів та моделей також було запропоновано алгоритмічне та програмне забезпечення, яке може бути використано для шифрування/розшифрування мовної інформації. Даний додаток був протестований для радіообміну «диспетчер-пілот».

Матеріали дисертаційної роботи упроваджено у діяльність ДП «Антонов», навчальний процес Національного авіаційного університету. Використання результатів дисертаційної роботи підтверджено відповідними актами впровадження.

Також представлені в роботі результати можна розглядати як основні базові відомості, необхідні для розуміння роботи крипто-семантических систем захисту смислового змісту інформації, застосування яких на практиці у ряді областей є досить доцільним.

Ключові слова: технічний захист інформації, лексикографічна система, відношення правдоподібності, семантичний тезаурус прикладних застосувань, захист текстової інформації, відстань єдиності.

ABSTRACT

Yevgen Samoilik. Lexicographic methods for the protection of linguistic information. - Qualifying scientific work on the rights of the manuscript.

Thesis for a Candidate Degree in Engineering in the specialty 05.13.21 "Information Security Systems". - National Aviation University, Kyiv, 2020.

Currently, a variety of virtually robust cryptographic systems have been developed that have been deployed to address a wide range of applications that require reliable protection against the privacy of information transmitted through open communication channels. However, these cryptosystems do not guarantee the formal, theoretically proven inability to break them. Therefore, there is a problem of mistrust of the reliability of these systems in information transmission tasks characterized by high levels of secrecy. In particular, it is impossible to fully trust the import of "strong cryptography" tools and applications that have such tools, the protection of critical state information, such as "secret" or "top secret", from attacks by powerful intelligence services organizational structures.

The dissertation is devoted to the creation of symmetric cryptosystems that do not make strict demands on the system of dissemination of key information.

It has been shown that the methods and models used in the problems of information and communication theory are not fully suitable for describing systems of semantic cryptography - systems designed to protect the semantic content of information from interception by entities that do not have appropriate legal access rights. For an adequate description of these systems in this work clarified and (or) redefined known concepts, such as environment, subject, intelligence, semantic thesaurus, information, form of presentation and meaning (content) of information, as well as a model of understanding the subject of content information on the tasks of semantic cryptography. The thesis presents a generalized information model of the subject, which is suitable for solving problems of semantic cryptography. The model reflects the main vital function of

the subject - to adequately respond to information coming from its environment. According to this model, the streams of information perceived by the sensitive elements of the sensory systems of the subject, are fed in real time for processing by intelligence. Intelligence from the perceived information flow allocates the flow of semantic images, processes this flow with the help of available intellectual means and, as a result, forms a flow of management decisions in the form of a flow of signals influencing elements of psychophysiological systems.

Thus, for the first time the information model of intelligence defining its structure and detailing its functions in the aspects considered at the decision of problems of semantic cryptography is presented. The intelligence of the subject in this model is considered as a software-controlled machine that implements in real time a set of programs to support an ordered set of physiological processes in the body of the subject. The work explained the functions of the six main information modules of this complex, controlled by the supervisor - the main system-forming program that organizes the functioning of intelligence for targeted analysis, processing and storage of semantic images extracted from the flow of incoming information. The functionality of these modules provides the ability to: 1) perception and evaluation of information flows coming from the receptors of sensory systems of the body, as well as the formation (according to the evaluation) and memorization of flows of semantic images; 2) maintenance and updating of the thesaurus of personal knowledge of the subject; 3) support and updating of the profile of personal culture of the subject (system of personal moral criteria, advantages and limitations); 4) maintaining and updating the profile of personal competencies; 5) the formation of management decisions that determine the physical behavior of the subject, as well as the physiological and emotional state of his body; 6) self-learning of the intellect, in the process of which the processes of improving the thesaurus of knowledge about the environment, the profile of personal culture and the profile of personal competencies of the subject are realized. In addition, a formal model of understanding linguistic information was developed for the first time, according to which in the process of mental activity

the subject generates deterministic finite discrete time series of semantic images selected by him (in particular, by associative mechanisms) from the available thesaurus. For the first time, the subject's thesaurus of intelligence is defined as a set of semantic images structured in the form of a layered root-like hierarchy, arranged according to the levels of abstractness of these semantic images in the direction of their more specific (more detailed) perception. It is shown that the quality of the thesaurus is determined both by the dimension of the base of semantic images, which the subject is able to operate in the process of mental activity, and the number of levels of abstract reflection of semantic images in the structure of the thesaurus.

Thus, the thesis identifies the characteristics of lexicographic systems that are necessary for the creation of perfectly stable cryptosystems for the protection of textual information. It is shown that the achievement of the goal set in the dissertation should be sought by developing new methods of building stable cryptosystems that can increase the distance of the unity of the key or, ideally, provide stable protection in the absence of any restrictions on the amount of textual information to be encrypted. If the increase in the distance of unity can be achieved by enlarging the alphabet of the source of open text messages, then avoiding the Shannon condition of perfect stability of protection is possible only through the use of semantic characteristics of text information based on a combination of symmetric cryptography and semantic processing of text messages. in some way constructed lexicographic systems, implemented in the form of problem-oriented linguistic corpora. In this case, it is not the form of displaying the structure of the relationship between the characteristics of objects that display text messages that is encrypted, but the essence (meaning) of this structure.

The characteristics of lexicographic systems, which are necessary for creation of perfectly stable cryptosystems of protection of text information, are determined. It is shown that the achievement of the goal set in the dissertation should be sought in the way of development of new methods of building perfectly stable cryptosystems, which are able to increase the distance of unity by key or, ideally,

to provide stable protection in the absence of any restrictions on the amount of textual information subject to scam . If increasing the distance of uniqueness can be achieved by enlarging the alphabet of the source of open text messages, avoiding the fulfillment of the Shannon condition of perfect stability of protection is possible only through the use of features of semantic characteristics of textual information based on a combination of methods of symmetric cryptography and semantic processing of changes in some way constructed lexicographic systems, implemented in the form of problem-oriented linguistic orpusiv. In this case, the encryption not of the structure of the relations between the characteristics of the objects displaying the text messages should be encrypted, but of the essence (meaning) of the structure. In order to adequately reflect the methods of using lexicographic systems in information security problems, definitions of well-known concepts, such as environment, subject, intelligence, semantic thesaurus, information, form of presentation and semantic content of information, as well as a model of understanding of the subject are given semantic content of information.

Thus, the main scientific results of the dissertation are: for the first time a crypto-semantic dictionary model has been developed, which, by introducing into the applied lexicographic system of indicators of semantic connections between semantic constructions of the language of display of the field of application, allows to determine the thesaurus of the information protection base in the application system and semantic structure application area dictionaries; for the first time developed a method of constructing a lexicographic system for the protection of speech information, which, by expanding the basic alphabet of the linguistic formal system of the source of voice messages, provides an increase in the distance of the cipher uniqueness by key and allows to increase the length of encrypted messages relative to the length of the key information and, accordingly, to reduce the frequency of change of the cipher keys Vernama; the first proposed lexicographic method of protecting textual information, which by accidentally replacing the primary semantic image of the message with another plausible element, taken from the semantic thesaurus of the base of protection of the applied area, allows to

provide increased stability of protection in the absence of any restrictions on the volume of linguistic information , and thus eliminates the need for periodic changes to key information.

The theoretical results obtained in the dissertation research open the opportunity to identify and propose new practical ways to improve the efficiency of encryption of linguistic information in various channels of information transmission. At the same time, the results obtained can significantly simplify the system of dissemination of key information.

The practical value of the dissertation is as follows: the method of automation of development of thesaurus of the information protection base in the applied area during the construction of the lexicographic cryptosystem is developed; a scheme of technical implementation of the method of construction of a cryptosystem with increased distance of unity on a key of a cipher for the stable system of transfer of the text data presented in the form of tabular forms is offered; Algorithmic and software of crypto-semantic protection of text data is developed, based on the use of applied thesaurus of semantic images, which is able to provide the mode of increased stability within the specifically defined application systems. Algorithmic and software that can be used to encrypt / decrypt speech information was also proposed on the basis of the methods and models developed in the work. This application has been tested for radio exchange "pilot-dispatcher".

The materials of the dissertation were introduced into the activities of the State Enterprise "Antonov", the educational process of the National Aviation University. The use of the results of the thesis is confirmed by the relevant acts of implementation.

Also the results presented in the work can be considered as the basic basic information necessary for understanding of work of crypto-semantic systems of protection of semantic content of the information.

Key words: technical protection of information, lexicographic system, likelihood ratio, semantic application thesaurus, protection of text information, distance of unity.

Список публікацій здобувача

1. Klimchuk, V., Samoylik, E., Gnatyuk, V., Prysiazhnyy, D., Buryachok, V. Synthesis of quite proof cryptosystem with increased unicity distance for cloud computing CEUR Workshop Proceedings 2104, pp. 596-607.

2. Самойлик Є.О. Ефективність досконало стійкої криптосистеми із збільшеною відстаню єдиності. Захист інформації. - Том 19, № 2 (2017). – С. 184-192.

3. Одарченко Р.С., Самойлик Є.О., Сімахін В.М., Боровик В.О., Тимчишин Р.М. Криптосемантична система захисту текстової інформації. Control Systems and Computers, N1, 2020. С. 35-46.

4. Р.С. Одарченко, Є.О. Самойлик, А.О. Абакумова Метод побудови семантичного словника у складі досконало стійкої криптосистеми захисту текстової інформації. Наукоємні технології. - № 3 (39), 2018. – С. 355-361.

5. Самойлик Є.О., Одарченко Р.С. Визначення характеристик лексикографічних систем, придатних для створення досконало стійких систем захисту текстової інформації. Вісник інженерної академії наук. - № 2, 2018. – С. 96-102.

6. Одарченко Р.С., Гнатюк В.О., Самойлик Є.О. Удосконалена архітектура системи безпеки стільникових мереж нового покоління. Наукоємні технології в інфокомунікаціях: обробка, захист та передача інформації [Текст] : монографія / під заг. ред.: В. М. Безрука, В. В. Баранніка ; Харків. нац. ун-т радіоелектроніки. – Харків : Бровін О. В., 2018. – 327 с.

7. Самойлик Є.О., Одарченко Р.С. Визначення характеристик лексикографічних систем з позицій технічного захисту текстової інформації. Збірник тез всеукраїнської науково-практичної конференції молодих учених і студентів «Проблеми навігації та управління рухом розвитку глобальної системи зв’язку, навігації, спостереження та організації повітряного руху CNS/ATM»: 21-23 листопада 2018 р. - С. 28.

8. Самойлик Є.О., Одарченко Р.С., Жмурко Т.О., Лукашенко В.В. Структура семантичного тезаурусу для лексикографічних систем. Збірник тез

III міжнародної науково-технічної конференції «Інформаційна безпека у сучасному суспільстві»: 29-30 листопада 2018 р., С. 34-35.

9. Hrystak A., Kinzeryavyy V., Prysiazhnyi D., Burmak Y., Samoylik Y. High-speed and hash function for blockchain security mechanisms. Scientific and practical cyber security journal (SPCSJ) 4(1): 65-70 ISSN 2587-4667 Scientific Cyber Security Association (SCSA).

10. Самойлик Е.А.. Одарченко Р.С. Крипто-семантическая система защиты текстовой информации. Тезисы докладов I Международной научно-практической конференции «Современные технологии кибербезопасности», 14 червня 2019 р., м. Алмати, Казахстан. - С. 102-111.

11. Самойлик Є.О. Структура системи криптосемантичного захисту інформації // Збірник тез міжнародної науково-практичної конференції молодих учених і студентів «Політ-2019.Сучасні проблеми науки»: 3–4 квітня 2019 р.: тези доп. – К., 2019. – С. 114.

12. Самойлик Є.О. Практичні рекомендації по організації криптосемантичного каналу захисту текстової інформації // Збірник тез науково-практичної конференціїї «Проблеми експлуатації та захисту інформаційно-комунікаційних систем»: 7 – 9 червня 2019 р.: тези доп. – К., 2019. – С. 67-68.

13. Самойлик Є.О., Одарченко Р.С. Визначення характеристик лексикографічних систем з позицій технічного захисту текстової інформації. Збірник тез всеукраїнської науково-практичної конференції молодих учених і студентів «Проблеми навігації та управління рухом розвитку глобальної системи зв’язку, навігації, спостереження та організації повітряного руху CNS/ATM»: 21-23 листопада 2018 р. - С. 28.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ

НСД – несанкціонований доступ;

ОС – операційна система;

ТЗІ – технічний захист інформації;

ПБ – послуга безпеки;

ФПЗ – функціональний профіль захисту;

L – довжина (обсяг) тексту, що підлягає захисту;

КЗВЄ – криптосистема із збільшеною відстанню єдиності;

S – словник, що відображає певну множину ***M*** можливих у тексті повідомлень $M_{j,i}$;

M – повідомлення як елемент словника ***S***;

N – кількість смислових варіантів повідомлень (тобто, розмір словника ***S***);

H(M) – ентропія повідомлення, узятого із словника ***S***;

n – довжина повідомлення;

r – ентропія мови, за допомогою якої відображаються повідомлення із ***M*** (тобто, це середня кількість інформації, що міститься в одній літері цієї мови і залежить від *n*);

R – максимальна ентропія мови (тобто, максимальне число бітів, яке може бути передано кожним символом алфавіту цієї мови за умови рівноймовірності виникнення усіх послідовностей символів);

D – надлишковість (інформаційна) мови ($D = R - r$);

H(K) – ентропія системи захисту як розмір простору ключів шифру, що залежить від кількості можливих для використання ключів шифру;

K – кількість можливих для використання ключів шифру у системі захисту (*H(K)*= log(*K*));

U – відстань єдиності (що також називають точкою єдиності) – такий приблизний розмір шифрованого тексту, для якого сума ентропії відкритого

тексту та ентропії ключа шифру дорівнює числу бітів, що міститься у цьому шифрованому тексті (*U= H(K)/D*);

E – множина можливих шифрограм, що утворюються шляхом застосування оператора перетворення $\boldsymbol{T_i}$ до відкритих повідомлень *M* із множини ***M***, а індекс *i* відповідає конкретному ключу, що був при цьому застосований;

$\boldsymbol{T_i}$ – оператор перетворення ***M*** у ***E***;

n_1 – довжина повідомлення [*літер*], що записане мовою людського спілкування (українська, російська, англійська і т.д.), що має алфавіт B_1 [*літер*];

SO – смисловий образ;

{**S**} – простір смислових образів, елементи котрого використовуються під час формування вихідних (рос., - исходных) текстових повідомлень;

TZ – семантичний тезаурус;

TZ_M – тезаурус мови відображення людської діяльності;

TZ_S – тезаурус суб'єкта;

TZ_{PS} – тезаурус мови відображення прикладної області;

TZ_{BZ} – тезаурус бази захисту текстової інформації;

$TZ^{(i)}$ – підтезаурус семантичних образів *i*-го рівня абстрагування;

{**F**} – простір відображень засобами обраної мови смислових образів, узятих із простору {**S**};

{$\mathbf{F_Ш}$} – простір зашифрованих відображень смислових образів із простору {**S**}; $\mathbf{P_{SF}}$ – оператор перетворення смислових образів у їхні відображення у рамках обраної мови;

$\mathbf{P_Ш(x)}$ – оператор шифрування, що залежить від вибору ключа шифру **x**;

$\mathbf{P_{рш}(x)}$ – оператор розшифрування, що залежить від вибору ключа шифру **x;**

$\mathbf{P_{FS}}$ – оператор перетворення відображень смислових образів у вихідні смислові образи;

$\mathbf{P_{шS}(x)}$ – оператор зворотного перетворення зашифрованих текстових повідомлень безпосередньо у зміст вихідних відкритих повідомлень;

Φ_i – функціональна структура і-ої інформаційної системи, що складається із елементів множини Φ_i припустимих для виконання функцій $\Phi_{i,к}$, де к=1, 2, …., N;

Z_i – функціональна структура і-ої інформаційної системи, що складається із елементів множини Z_i обмежувальних умов Z_i, де k = 1, 2,…,M, за яких виконання функцій і-ої інформаційної системи є припустимим;

$\Phi_{i,k}$ – k-й функціональний елемент із заданої множини припустимих для виконання функцій і-ої інформаційної системи Φ_i, де k = 1, 2, …, N;

$Z_{i,k}$ – k-та обмежувальна умова із заданої множини обмежувальних умов Z_i, де k = 1, 2, …, M;

N, M – кількість елементів відповідно у множинах Φ_i та Z_i.

ВСТУП

Актуальність теми

Наразі розроблено різноманітні практично стійкі криптографічні системи, що знайшли застосування для вирішення широкого спектру прикладних задач, де необхідно забезпечити надійний захист від порушень конфіденційності інформації, що передається відкритими каналами зв'язку [1-5]. Проте ці криптосистеми не гарантують формальну, теоретично доведену неможливість їхнього злому [4-7]. Отже, існує проблема недовіри до надійності цих систем в задачах передавання інформації, що характеризуються високими рівнями секретності. Тому тема цієї дисертаційної роботи спрямована на створення засобів, що дозволяють відносно мало затратним шляхом забезпечити ефективний захист такої інформації.

З урахуванням вище зазначеного у даній роботі розглянуто стійкі криптосистеми захисту текстової інформації з теоретично доведеною ідеальною теоретико-інформаційною стійкістю [8], що гарантують неможливість однозначного відновлення відкритих текстових повідомлень навіть за умов, коли у розпорядженні криптоаналітика знаходяться зразки зашифрованих повідомлень скільки завгодно великої сумарної довжини, а криптоаналітик має необмежений час та необмежені обчислювальні ресурси для дешифрування перехоплених криптограм [8].

Проте, сучасним досконало стійким криптосистемам притаманні недоліки, що суттєво звужують області їхнього використання. Необхідно слідкувати, щоб поточна сумарна довжина зашифрованих текстових повідомлень у процесі шифрування не перевищувала довжину ключа шифру. Інакше виникає теоретична можливість розкриття ключа шифру.

Численні дослідження багатьох авторів вказують на відносно невеликі значення відстані єдиності при шифруванні повідомлень, складених із символів алфавіту будь-якої із природних мов [10,11]. Це призводить до

необхідності частої зміни ключової інформації, що є проблемою для багатьох застосувань. Окрім того, необхідно забезпечувати випадковість й однакову ймовірність вибору варіантів реалізації ключа. Так що створення нових методів побудови стійких криптосистем, що забезпечують більш великі значення відстані єдиності або, взагалі, забезпечують можливість шифрування скільки завгодно великих обсягів текстових, зокрема голосових, повідомлень незалежно від значень відстані єдиності являє актуальне завдання.

Аналіз потокових шифрів з рівно ймовірними ключами показує, що існує можливість збільшення відстані єдиності за рахунок синтезу штучної мови відображення прикладної області з алфавітом великої розмірності. У цьому випадку відкривається можливість передавання у режимі секретності більшого обсягу зашифрованої інформації без необхідності зміни ключа шифру. До того ж такий підхід не потребує визначення семантичних співвідношень між лінгвістичними конструкціями синтезованої мови і, отже, створення семантичного тезаурусу прикладної області. Тим не менш, неперевищення відстані єдиності є обов’язковою умовою дотримання режиму досконалої секретності у створених таким чином криптосистемах, що не усуває жорстких вимог до системи розповсюдження ключової інформації.

З іншого боку, можливо використати крипто-семантичний підхід до створення стійкої криптографічної системи. У цьому випадку відпадає необхідність дотримання умови неперевищення відстані єдиності і відкривається можливість вибору довжини ключів шифру незалежно від обсягів інформації, що потребують шифрування. Проте на цьому шляху виникає необхідність дослідження семантичних зв’язків між лінгвістичними конструкціями синтезованої мови відображення прикладної області і створення відповідного семантичного тезаурусу. Вищезгадана задача, яка вирішувалась в даній дисертаційній роботі, обумовлює її **актуальність.**

Зв'язок роботи з науковими програмами, планами, темами. Тематика

дисертаційної роботи та одержані результати безпосередньо пов’язані з «Основними науковими напрямами та найважливішими проблемами фундаментальних досліджень у галузі природничих, технічних і гуманітарних наук НАН України на 2014-2018 роки», Стратегією національної безпеки України від 26 травня 2015 року № 287/2015, Стратегією кібербезпеки України від 15 березня 2016 року №96/2016 і Рамковою програмою ЄС з досліджень та інновацій «Горизонт 2020».

Висвітлені в дисертації наукові результати отримано, здебільшого, в рамках науково-дослідних робіт, які були виконані в Національному авіаційного університеті (НАУ), а також на ДП «Антонов»: шифр 874-ДБ13, тема «Створення та дослідження нових систем захищеного авіаційного радіозв’язку в рамках Концепції CNS/ATM ICAO» (НДР 0110U000225). Отримані результати використовуються у навчальному процесі НАУ. Використання результатів дисертаційної роботи підтверджено відповідними актами впровадження.

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

Задачі дослідження:

1. Проаналізувати ефективність сучасних методів захисту мовної інформації та можливість її підвищення за рахунок використання семантичних характеристик текстової інформації.

2. Розробити модель крипто-семантичного словника за рахунок уведення показників семантичних зв’язків між смисловими конструкціями мови відображення прикладної області і на цій основі здійснити синтез структури тезаурусу смислових образів.

3. Розробити метод побудови лексикографічної криптосистеми, заснованої на використанні механізму укрупнення алфавіту джерела текстових повідомлень.

4. Розробити лексикографічний метод захисту текстової інформації.

5. Розробити схему реалізації, програмне забезпечення крипто-семантичного захисту мовної інформації.

Об’єктом дослідження є процеси захисту мовної інформації.

Предметом дослідження є методи підвищення ефективності шифрування текстової інформації, що спрямовані на зменшення необхідності у періодичній зміні використання ключа шифру.

Методи дослідження

Розробка методів, що забезпечують підвищення ефективності шифрування мовної інформації при обміні інформацією через відкриті канали зв’язку, виконана на основі результатів теорії випадкових процесів, теорії телетрафіка, теорії криптографічних систем та теорії лексикографічних систем. В основу роботи покладена теорія секретного зв’язку К. Шеннона. Широко використано результати теорії лінгвістичних корпусів, що відображені, зокрема, у праці Широкова «Корпусна лінгвістика». Під час натурного моделювання запропонованих криптосистем використано сучасні методи комп’ютерного моделювання.

Наукова новизна одержаних результатів

Автором одержані наступні нові наукові результати:

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

2. *Вперше розроблено метод побудови лексикографічної криптосистеми*, який за рахунок розширення базового алфавіту лінгвістичної формальної системи джерела мовних повідомлень забезпечує збільшення відстані єдиності шифру за ключем та дозволяє збільшити довжину шифрованих повідомлень відносно довжини ключової інформації та відповідно зменшити частоту зміни ключів шифру.

3. В*перше запропоновано лексикографічний метод захисту текстової інформації*, що за рахунок випадкової заміни первинного смислового образу повідомлення на інший правдоподібний елемент, узятий із семантичного тезаурусу бази захисту прикладної області, дозволяє забезпечити підвищену стійкість захисту при відсутності будь-яких обмежень на обсяг мовної інформації, що підлягає шифруванню, і, тим самим, усуває необхідність у періодичній зміні ключової інформації.

Практична значення одержаних результатів

Автором були отримані нові результати для впровадження на практиці:

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

2. Запропоновано схему технічної реалізації методу побудови криптосистеми із збільшеною відстанню єдиності за ключем шифру для стійкої системи передавання текстових даних, представлених у вигляді табличних форм.

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

Особистий внесок здобувача

Основні положення й результати дисертаційної роботи отримані автором самостійно. Роботи, виконані разом із співавторами, наведені в переліку публікацій. З робіт, що опубліковані у співавторстві, використовуються результати, отримані особисто здобувачем. У роботах, опублікованих у співавторстві, автору дисертації належить: розробка методу створення крипто-семантичної системи захисту інформації для використання в хмарних технологіях [1]; розробка криптосемантичної системи захисту текстової інформації [3]; розробка методу побудови семантичного словника у складі стійкої криптосистеми захисту текстової інформації [4]; визначення основних

характеристик лексикографічних систем, придатних для створення стійких систем захисту текстової інформації [5]; удосконалення алгоритму шифрування даних в стільникових мережах із використанням лексикографічних підходів [6]; визначення характеристик лексикографічних систем для шифрування даних [7]; розробка методу створення семантичного тезаурусу для лексикографічних систем [8]; розробка структури крипто-семантичної системи захисту інформації [9]; визначення характеристик лексикографічних систем з позицій технічного захисту текстової інформації [12].

Апробація результатів дисертації

Основні теоретичні та практичні результати дисертаційної роботи доповідались і обговорювались на таких конференціях і семінарах: Міжнародна науково-практична конференція молодих учених та студентів «Політ. Сучасні проблеми науки» (Київ, НАУ, 2017-2019 рр.); Міжнародний круглий стіл «Про національну і інформаційну безпеку РК» (Казахстан, Алмати, 2016 р.); IEEE International Scientific-Practical Conference «Problems of Infocommunications Science and Technology (PIC S&T)» (Харків, ХНУРЕ, 2018 р.); VIII Міжнародна науково-технічна конференція «Комп'ютерні системи і мережні технології» (Київ, НАУ, 2015 р.); Міжнародна науково-технічна конференція «ITSEC» (Київ, НАУ, 2017 – 2018 рр.); Міжнародна науково-технічна конференція «ABIA-2015» (Київ, НАУ, 2015 р.); Міжнародна науково-практична конференція «Інформаційна безпека та комп'ютерні технології» (Кропивницький, КНТУ, 2016 р.); Науково-технічна конференція «Проблеми розвитку глобальної системи зв'язку, навігації, спостереження та організації повітряного руху CNS/ATM» (Київ, НАУ, 2018 р.); XXI Міжнародна науково-технічна конференція «Сучасні засоби зв'язку» (Мінськ, 2018 р.); Науково-технічна конференція «Проблеми експлуатації та захисту інформаційно-комунікаційних систем» (Київ, НАУ, 2017 – 2019 рр.).

Публікації. За матеріалами дисертаційної роботи опубліковано 13 наукових праць, у тому числі 1 розділ колективної монографії, 4 статті у

фахових виданнях, які входять в перелік наукових видань, затверджений МОН України, 1 працю, яка включена до науково-метричної бази Scopus, матеріали доповідей на науково-технічних конференціях – 6.

Структура і зміст роботи. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг роботи складає 157 с. друкованого тексту, у тому числі містить 17 рисунків та 2 таблиці. Список використаних джерел на 10 с. містить 122 найменування.

РОЗДІЛ 1
ВИЗНАЧЕННЯ ХАРАКТЕРИСТИК ЛЕКСИКОГРАФІЧНИХ СИСТЕМ, ПРИДАТНИХ ДЛЯ СТВОРЕННЯ СИСТЕМ ЗАХИСТУ МОВНОЇ ІНФОРМАЦІЇ

1.1. Аналіз шляхів забезпечення стійкості криптосистем

У [12,13] надана алгебраїчна модель шифру. Але така модель лише відображає функціональні властивості шифрування без урахування особливостей мови повідомлень та їх статистичних ознак. Тому К. Шенноном була запропонована імовірнісна модель шифру [14-17]. Згідно цієї моделі за відомими розподілами ймовірностей відкритих текстів і ключів та перехопленою криптограмою криптоаналітик може обчислити безумовний розподіл імовірностей криптограм, умовні апостеріорні ймовірності відкритих текстів і, таким чином, отримати інформацію для визначення вихідного тексту та ключа.

Основною характеристикою шифру є його криптографічна стійкість, під якою розуміють його здатність протистояти атакам криптоаналітика [18-21]. Дана робота присвячена конструюванню досконало стійких шифрів з абсолютною теоретико-інформаційною стійкостю, щодо яких дано наступне визначення [22-23].

Шифр із множиною відкритих текстів X, множиною криптограм Y, множиною ключів K і криптографічним перетворенням $y = E_k(x,k)$ з імовірнісними розподілами $P_{відкр.}(X)$ та $P_{кл.}(K)$ зветься досконало стійким (з атакою на x при перехопленні y), якщо для всіх $x \in X$ і $y \in Y$ має місце рівність:

$$P_{відкр./кр.}(x/y) = P_{відкр.}(x) \qquad (1.1)$$

(іншими словами, відкритий та зашифрований тексти є статистично незалежними).

Для досконало стійкого шифру має місце нерівність:

$$|K| \geq |Y| \geq |X|, \qquad (1.2)$$

де $|K|$, $|Y|$, $|X|$ – кількість можливих ключів, криптограм та відкритих текстів відповідно.

Шифр зветься ідеально стійким, якщо неможливо однозначно відновити відкритий текст за відомим шифрованим текстом скільки завгодно великої довжини. Тож, досконало стійкий шифр є ідеально стійким.

Важливим класом шифрів є ендоморфні шифри, для яких множина відкритих текстів X збігається з множиною криптограм Y [21]. Мінімально можлива кількість ключів ендоморфних шифрів K дорівнює кількості можливих криптограм, тобто $|K| = |Y|$ [24]. Щоб ендоморфний шифр, для якого виконується умова $|X|=|Y|=|K|$, був досконало стійким, необхідно і достатньо, аби: 1) для будь-якого відкритого тексту $x \in X$ і криптограм $y \in Y$ існував лише один ключ k з умовою $y = E_k(x,k)$; 2) розподіл імовірностей $P_{кл.}(K)$ був рівномірним, тобто щоб імовірність вибору будь-якого ключа $k \in K$ залишалась однаковою і дорівнювала $1/|K|$ [23].

Прикладом ендоморфного досконало стійкого шифру, що задовольняє вищенаведеним умовам, є так званий латинський квадрат [25,26] – квадратна матриця, рядки і стовпці якої є результатом перестановки елементів деякої множини. Рядки цього квадрата задають множину ключів шифру $|K|$, стовпці – множину відкритих текстів $|X|$, а клітини – множину відповідних ним криптограм $|Y|$.

Іншим прикладом [27] досконало стійких шифрів можуть бути шифри простої заміни з множиною ключів $K= SYM(Z_m)$, де $SYM(Z_m)$ – симетрична група підстановок з рівно ймовірним вибором ключа та множиною відкритих текстів із слів довжини L без однакових букв.

Проте не тільки ендоморфні шифри є досконало стійкими. Відносно широке розповсюдження отримали шифри, що реалізуються з використанням джерела випадкових послідовностей та операції *XOR* [23].

Нехай знаки відкритого тексту, криптограми і ключа набувають значення з кільця лишків Z_m, а довжини ключа і криптограми збігаються з

довжиною n відкритого тексту. Тоді рівняння шифрування досконало стійкого шифру може бути задано наступним рівнянням:

$$y \equiv (x_i + k_i) \bmod m,\ i=1,2,\ldots,n, \qquad (1.3)$$

що визначає операцію шифрування n - грами відкритого тексту $(x_1,x_2,\ldots,x_n)$ на ключі $k=(k_1,k_2,\ldots,k_n)$, у результаті якої утворюється n - грама шифрованого тексту $(y_1,y_2,\ldots,y_n)$, m – основа алфавіта символів текста.

Існує немало модифікацій схеми (1.3), що отримала назву «схема одноразових блокнотів». Найбільш відома з них має назву «шифр Вернама». Доведено [23], що шифр Вернама за модулем m відповідає умові (1.1) і, тому, є досконало стійким. Випадкова ключова послідовність $k=(k_1,k_2,\ldots,k_n)$ додається до невипадкового відкритого тексту, генерується довільний шифрований текст, зламати котрий не допоможуть ніякі обчислювальні потужності, тому що сформовані криптограми є рівно ймовірними. Так що увесь криптоаналіз зводиться до аналізу методу генерації символів ключа. Основний недолік шифру Вернама пов'язаний з необхідністю частої зміни ключа, оскільки для підтримки режиму досконалої стійкості потрібен ключ такої ж довжини, що й повідомлення. Тобто, доводиться вирішувати проблему розподілу ключів.

Розглянемо особливість шифрування осмислених текстових повідомлень потоковим шифром з рівно ймовірними ключами. Під час криптоаналізу при спробі визначити істинний ключ шифру шляхом прямого перебору варіантів ключа криптоаналітик може отримати декілька осмислених відкритих текстів для різних ключів. При цьому тільки один з цих ключів буде істинним ключем, а решта – фальшивими. Натуральне число L_0, для якого очікувана кількість фальшивих ключів дорівнює нулю, визначає відстань єдиності за ключем [23]. Відстань єдиності – це мінімальна довжина шифрованого тексту, необхідного для однозначного встановлення істинного ключа шифру, що визначається через нерівність [23]

$$L \geq \log_2|K|\backslash(D\log_2 m), \qquad (1.4)$$

де D – надлишковість мови відображення текстових повідомлень.

Найменший цілий розв'язок L_0 цієї нерівності і приймається за відстань єдиності за ключем.

Відстань єдиності – це не міра довжини криптограми, що є необхідною для успішності криптоаналізу, а міра довжини криптограми, необхідної для єдиності результату криптоаналізу.

Бажано, щоб $L_0 \rightarrow \infty$, але на практиці при шифруванні символами природних мов, зокрема українською [26], відстань єдиності виявляється небезпечно малою.

За результатами аналізу відомих досконало стійких методів захисту текстової інформації зробимо наступні висновки:

1) існуючі досконало стійкі шифри при шифруванні повідомлень, складених із символів алфавіту будь-якої із природних мов, мають відносно невеликі значення відстані єдиності за ключем, що призводить до необхідності частої зміни ключової інформації;

2) існують проблеми із синхронізацією генераторів дійсно випадкових послідовностей (ГВП) у випадках, коли вони входять до складу досконало стійких криптосистем, що ставить під сумнів доцільність їхнього застосування на практиці;

3) існують труднощі з гарантуванням стійкості генераторів псевдовипадкових послідовностей (ГПВП), що не завжди робить доцільним їхнє використання у складі досконало стійких криптосистем.

Враховуючи вищенаведене, досягнення мети даного дослідження – уникнення необхідності частої зміни ключів шифрів – слід шукати на шляху розробки нових методів побудови досконало стійких криптосистем, які здатні збільшити відстань єдиності за ключем або, в ідеалі, забезпечить досконало стійкий захист при відсутності будь-яких обмежень на обсяг текстової інформації, що підлягає шифруванню.

Якщо збільшення відстані єдиності може бути досягнуто шляхом укрупнення алфавіту джерела відкритих текстових повідомлень, то уникнення виконання шенноновської умови досконалої стійкості захисту є

можливим лише за рахунок використання особливостей семантичних характеристик текстової інформації на основі поєднання відомих методів симетричної криптографії і методів семантичної обробки змісту цих повідомлень, зокрема за допомогою певним чином побудованих лексикографічних систем, реалізованих у вигляді проблемно-орієнтованих лінгвістичних корпусів. У цьому випадку буде здійснюватися шифрування не форми відображення структури співвідношень між характеристиками об'єктів, що відображають текстові повідомлення, а безпосередньо сутність (смисл) цієї структури. (У даній роботі поняття інформації визначається з використанням співвідношення філософських категорій «форма – сутність», що надане, наприклад, у [28]).

1.2. Визначення характеристик лексикографічних систем з позицій технічного захисту текстової інформації

Центральна ідея, що спрямована на досягнення мети даного дослідження, базується на використанні крипто-семантичного підходу до створення досконало стійкої системи захисту текстової інформації, зокрема результатів теорії лексикографічних систем та методів побудови відповідних лінгвістичних корпусів [28,29]. Якщо здійснювати захист смислового змісту текстової інформації, а не відображення цього змісту у певним чином обраній формі (наприклад, кодових послідовностей), то слід сподіватися, що при побудові досконало стійких криптосистем відпаде необхідність дотримання умови неперевищення відстані єдиності шифру за ключем і, як результат, відкриється можливість вибору довжини ключів шифру незалежно від обсягів інформації, що потребують шифрування.

На цьому шляху, перш за все, слід отримати формальні докази щодо можливості використання семантичних характеристик текстової інформації для забезпечення стовідсоткових гарантій забезпечення її конфіденційності. Зокрема, слід визначити показники цих характеристик, що можуть бути використані при побудові досконало стійких криптосистем захисту текстової

інформації. У цьому випадку виникає необхідність дослідження семантичних зв’язків між лінгвістичними конструкціями синтезованої мови відображення прикладної області і створення відповідного семантичного тезаурусу.

Примітка 1.1. *Підкреслимо, що мова йде не про захист кодових відображень інформації, що циркулює у каналах зв’язку або оброблюється на комп’ютерах чи фіксується на фізичних носіях, а про захист смислу інформації, що виявляється у процесі інтелектуальної діяльності людини.*

Визначення та моделі, що використовуються у задачах теорії інформації та зв’язку, не у повній мірі придатні для опису лексикографічних систем [30-34]. Тому для адекватного відображення методів їхнього застосування в задачах ТЗІ слід уточнити або перевизначити відомі поняття, такі як середовище, суб’єкт, інтелект, семантичний тезаурус, інформація, форма представлення та смисловий зміст інформації, а також надати модель розуміння суб’єктом смислового змісту інформації.

Примітка 1.2. *Автор даної роботи не претендує на виключне авторство наведених нижче визначень понять. Більшість наданих визначень слід розглядати як відображення результатів теорії лексикографічних систем на проблематику захисту інформації. Проте існують й інші визначення цих же понять. Включення наведених нижче визначень у даний текст спрямовано на усунення неоднозначності їхнього сприйняття.*

1.2.1. Загальні визначення

Об’єкт – те, що можливо розглядати окремо як єдине ціле.

Середовище – це множина об'єктів, що характеризується унікальною структурою взаємозв'язків і характеристик властивостей компонентів цього середовища, зокрема різних мікро і макро елементів, що входять до складу об'єктів, об'єктів в цілому, фрагментів середовища і середовища в цілому. (Тобто, допускається, залежно від поставлених завдань, представляти середовище у різних аспектах і з різним рівнем її деталізації).

Суб'єкт – елемент підмножини об'єктів середовища, що відрізняються від усіх інших об'єктів наявністю інтелекту. Це, передусім, люди і інші живі об'єкти, але і, можливо, штучно створені об'єкти у середовищі існування людини. Суб'єкт завдяки наявності інтелекту має здатність виявлення, фіксації і оцінювання характеристик суб'єктів і об'єктів середовища існування, навчатися, формувати раціональні стратегії поведінки у середовищі існування, емоційно реагувати на події, що відбуваються, і здійснювати цілеспрямовані дії зі зміни характеристик середовища з метою підвищення комфортності свого існування або існування соціуму, до котрого він належить.

Інтелект – жива (тобто, що має свідомість) програмно керована машина, що інтегрована в організм суб'єкта для забезпечення безпеки і комфортності його існування або соціуму, до якого суб'єкт належить. Свідомість – поточний комплекс відчуттів, що формуються інтелектом, котрі відбивають у реальному часі стан усіх систем організму в пам'яті суб'єкта. Завдяки інтелекту суб'єкт здатний відчувати (тобто, виявляти, фіксувати і оцінювати), створювати, а також використовувати інформацію в корисних для себе цілях.

Інтелект реалізує в реальному часі взаємопов'язану сукупність процесів з тим, щоб підвищити (чи, щонайменше, зберегти) рівень комфортності і гарантій існування в умовах середовища існування, що змінюються. З усього різноманіття процесів, що реалізовуються інтелектом, у цій роботі розглядаються наступні процеси: синтез тезауруса семантичних образів, необхідного для осмислення суб'єктом характеристик середовища; виділення з інформації, що поступає через сенсорну систему суб'єкта, смислових образів, необхідних, зокрема, для формування програми конкретних дій у середовищі існування, що змінюється; синтез програми конкретних дій (раціональній стратегії поведінки); самонавчання в частині синтезу нових знань, модифікації профілю моральної поведінки (персональної культури) і модифікації профілю персональних компетенцій (умінь) суб'єкта.

Примітка 1.3. *Такі важливі інтелектуальні процеси як управління фізіологічними процесами в організмі, емоційними станами суб'єкта, органами впливу суб'єкта на об'єкти середовища існування та ін. у даній роботі не розглядаються.*

Примітка 1.4. *Механізми реалізації управлінських рішень, вироблених інтелектом, не є функцією інтелекту і також тут не розглядаються.*

Інформація (узагальнене визначення) – це закодоване відображення у свідомості суб'єкта структури взаємозв'язків і характеристик властивостей компонентів середовища (чи середовища в цілому), аналізоване інтелектом у філософських категоріях "форма - суть" [8,29]. Інформація включає відомості (дані) про форму структури, аналізуючи які можна отримати відомості про суть структури, зокрема виявити смислові образи, що відображені у формі аналізованої структури. Так що, інформація містить відомості як про форму, так і про суть структури.

Форма структури – результат відображення структури об'єкту, до якого суб'єкт виявив цікавість, в пам'яті суб'єкта. Відображення, що сприймається через рецепторну систему суб'єкта з використанням його органів чуття. Форму структури не слід ототожнювати з формою (кодом) відображення інформації про форму структури, тобто з алфавітом і форматами одиниць мови представлення інформації про відображення структури.

Суть структури – результат обробки сприйнятого інтелектом відображення структури об'єкту, тобто результат інтелектуальної обробки форми структури. Суть виявляється шляхом аналізу отриманого відображення (тобто, отриманої форми) за допомогою інтелекту суб'єкта. Суть структури не слід ототожнювати зі змістом інформації про форму структури, а слід ототожнювати зі змістом інформації про сенс структури. Суть структури може мати сенс (тобто, бути змістовною), а може не мати сенсу (тобто, бути беззмістовною). Якщо в результаті аналізу відомостей про форму структури, інтелект виявив які-небудь смислові образи, то отримана інформація класифікується інтелектом як що має сенс. Інакше інформація

вважається позбавленою сенсу. Інформація доносить до інтелекту суб'єкта відображення (форму) суті (змісту) структури, яке притягнуло увагу суб'єкта, в кодах представлення цього відображення, сприйманих органами чуттів суб'єкта (зір, слух, нюх та дотик). А суть інформації виявляється шляхом аналізу отриманого відображення за допомогою інтелекту суб'єкта. Так що, інформація – це відомості, дані як про відображення суті (у кодах представлення цього відображення), так і про саму суть (у кодах представлення смислових образах). По відображенню суті інтелект суб'єкта намагається виявити саму суть. Так що, інформація може вміщувати в себе одночасно і форму, і суть структури взаємозв'язків і характеристик, що притягнула увагу суб'єкта.

Примітка 1.5. *Помітимо, що інформація - це не відомості про суть змісту, а відомості про відображення суті. Зокрема, якщо ми маємо справу з людиною і його відображенням в дзеркалі, то інформація - це не відомості про саму людину (тобто, про його суть), а відомості про його відображення в дзеркалі. Чи можна по відображенню в дзеркалі скласти уявлення про суть людини? У якійсь мірі - можливо, а в якійсь - не можна. Інтелект якраз і намагається по відображенню виявити суть. Відображення в дзеркалі однієї і тієї ж людини різними суб'єктами в якихось деталях може усвідомлюватися по-різному, оскільки суб'єкти мають різний інтелект. Суть інформації може бути змістовною (тобто, мати сенс), а може бути беззмістовною (що не має сенсу). Інформація може відображати форму без змісту. Інтелект суб'єкта може усвідомлювати смислові образи без яких-небудь форм їх відображення (наприклад, сни суб'єктів).*

Примітка 1.6. *Якщо ставляться завдання передачі інформації (коли джерело і приймач інформації віддалені один від одного), або якщо виникає необхідність в запам'ятовуванні інформації поза свідомістю суб'єкта (наприклад, в книгах або в комп'ютерній пам'яті), то в цих випадках йдеться про інформацію, яка відображає форму суті структури в кодах представлення цього відображення. Суттю інформації в цих випадках не*

цікавляться. В процесі передачі або запам'ятовування може спотворюватися відображення (форма) суті передаваної інформації, що може утрудняти виявлення цієї суті інтелектом суб'єкта.

1.2.2. Характеристики інтелекту

Інтелект має здатність сприймати таку унікальну властивість простору і часу як спрямованість, що дозволяє суб'єктові усвідомлювати своє існування, існування інших об'єктів, оцінювати структуру і властивості компонентів середовища у прив'язці до координат простору і часу. Інтелект, зокрема, здатний накладати на семантичний образ, що відбиває розташування компонент середовища у просторі, придуману ним координатну сітку і, в результаті, на кількісному рівні визначати місце розташування, форму і розміри цих компонент.Так само інтелект здатний накладати на семантичний образ, що відбиває часове розташування подій, що відбуваються, деяку часову шкалу і, таким чином, на кількісному рівні визначати моменти виникнень і тривалості різних подій, змін характеристик і фаз (стадій в змінах) в станах цих компонент на осі часу.

Інтелект забезпечує можливість людині усвідомлювати структуру і властивості компонент реально існуючого середовища, закодоване відображення яких завдяки органам чуття він здатний відчувати реально. В той же час, завдяки інтелекту, суб'єкт має здатність абстрактно мислити, тобто створювати віртуальне (уявне, реально не відчутне) середовище, заповнене придуманими ним об'єктами. Можливості сенсорної системи суб'єкта обмежують відчуття об'єктів в реальному просторі трьома мірами і однією мірою реально поточного часу. Але завдяки здатності абстрактно мислити, суб'єкт може сконструювати багатовимірне віртуальне середовище з уявною структурою взаємозв'язків і властивостей її віртуальних компонент.

Мета функціонування інтелекту – підвищити (чи, щонайменше, зберегти) рівень комфортності існування у змінюваних умовах середовища мешкання. Якщо суб'єкт відчуває дискомфорт, то його метою може бути зниження рівня

персонального дискомфорту або дискомфорту соціуму (у випадках, коли забезпечення комфортності існування соціуму є моральною потребою суб'єкта). При цьому під місцем існування розуміється множина об'єктів і суб'єктів, характеристики яких відчуваються суб'єктом в координатах простору і часу. Відчуття комфортності пов'язане з позитивними фізіологічними відчуттями, почуттями і емоціями. Наприклад, з відчуттями безпеки, свободи дій або задоволеності, з почуттями задоволення, гордості або оптимізму, з емоціями радості або доброзичливості. Відчуття дискомфорту пов'язане з негативними відчуттями, почуттями і емоціями. Наприклад, з відчуттями небезпеки, тривоги, страху, обмежень свободи дій або незадоволення, а також з почуттями стурбованості, пригніченості, депресії, песимізму, приниження або болю, а також з емоціями обурення або злості. Суб'єкт може в один і той же час переживати і позитивні, і негативні відчуття. Так що, можна говорити про певний профіль відчуттів.

Суть роботи інтелекту полягає в реалізації взаємопов'язаної сукупності процесів – синтезу смислових образів, управління органами дії суб'єкта на об'єкти і суб'єкти середовища існування, управління елементами психофізіологічної системи суб'єкта, самонавчання (в т.ч., поповнення тезауруса знань, розширення профілів культури і компетенцій суб'єкта).

Ефективність роботи інтелекту визначається рівнем якості тезауруса знань, культури і персональних компетенцій суб'єкта.

Як правило, розглядають, в тій чи іншій мірі, шість механізмів, що реалізовуються інтелектом:

1) механізм витягання (синтезу) смислових образів з інформації, що входить, з використанням наявного персонального тезауруса знань;

2) механізм ухвалення осмислених рішень по управлінню органами дії суб'єкта на об'єкти і суб'єкти середовища існування (в т.ч., на себе) з метою підвищення рівня комфортності існування (чи зниження рівня дискомфорту);

3) механізм ухвалення підсвідомих рішень по управлінню елементами психофізіологічної системи суб'єкта (в т.ч., емоціями);

4) механізм оновлення структури і поповнення знаннями наявного тезауруса з використанням результатів відповідної обробки витягнутих з інформації, що входить, смислових образів;

5) механізм вдосконалення профілю персональної культури суб'єкта з використанням результатів навчання на витягнутих вибірках смислових образів;

6) механізм вдосконалення профілю персональних компетенцій суб'єкта з використанням результатів навчання на витягнутих вибірках смислових образів.

Інформація використовується інтелектом як сировина, необхідна для витягання з неї смислового змісту – смислових образів. Іншими словами, інтелект витягає з інформації її сенс. Витягнутий сенс використовується інтелектом, передусім, для підвищення рівня комфортності (чи пониження рівня дискомфорту) існування суб'єкта, що має цей інтелект, шляхом цілеспрямованої дії на характеристики самого суб'єкта і його середовища існування з використанням відповідних психофізіологічних систем і опорно-рухового апарату суб'єкта, які управляються інтелектом. Крім того, витягнутий сенс використовується для підвищення рівня якості самого інтелекту за рахунок самонавчання, в результаті якого поповнюються (оновлюються) знання про характеристики середовища існування, що дозволяє цілеспрямовано модифікувати (оновити) структуру тезауруса знань, підвищити персональну культуру і наростити персональні компетенції суб'єкта.

1.2.3. Характеристики інформації

З урахуванням викладених вище відомостей представляється можливим визначити поняття "інформація" таким чином.

Інформація (уточнене визначення) – це закодоване природою відображення структури реально існуючих логічних і кількісних співвідношень між характеристиками (показниками) властивостей

матеріальних об'єктів середовища або їх елементів, а також закодоване інтелектом відображення придуманої структури співвідношень між придуманими характеристиками віртуальних (придуманих) об'єктів або їх елементів. Співвідношень, що незмінні або змінюються в координатах досліджуваного простору і часу. Співвідношень, реально існуючих в середовирщі існування суб'єкта і, отже, реально відчутних за допомогою органів почуттів суб'єкта, або реально не існуючих, уявних, існуючих тільки в уяві суб'єкта.

Інформація – це відомості як про форму відображення в пам'яті суб'єкта, так і про виявлену інтелектом суть структури співвідношень між характеристиками властивостей об'єктів середовища. Відомості про форму структури сприймаються інтелектом через рецептори сенсорних систем суб'єкта. Відомості про суть структури виявляються інтелектом на основі аналізу відомостей, що поступили, про форму структури. Це різні види відомостей. У завданнях криптосемантики їх слід розрізняти.

Відомості про форму структури називатимемо первинною інформацією (у теорії зв'язку або при описі комп'ютерних технологій цей вид відомостей називають інформацією (чи даними) без якого-небудь прикметника, оскільки її смислові характеристики знаходяться поза межами цих сфер використання поняття "інформація"). Відомості про суть структури називатимемо вторинною інформацією, яка може бути беззмістовною (позбавленою сенсу) або представляти відомості про смислові образи, що називаються знаннями.

Первинна інформація – абстрактна категорія, але цей вид відомостей може сприйматися сенсорними системами суб'єкта у реальному часі або фіксуватися на матеріальних носіях. Завдяки чому первинну інформацію можна запам'ятовувати і зберігати у часі, передавати через канали зв'язку і обробляти за допомогою комп'ютерних систем не лише у реальному часі, але і в зручному для суб'єкта темпі і в слушний для нього час.

Вторинна інформація – також абстрактна категорія, але цей вид відомостей може відображатися в нервових клітинах організму суб'єкта.

Отримати вторинну інформацію, використовуючи як початкові дані первинну інформацію, а потім витягнути з вторинної інформації знання може тільки інтелект суб'єкта. При цьому іноді можна витягнути знання, отримавши відомості від неживих, але реально існуючих об'єктів. Наприклад, при спостереженні за контуром гір на тлі неба можуть виникнути цілком певні смислові образи (типу образ кішки при спостереженні за контуром гори під назвою Кішка або образ ведмедя при спостереженні за контуром гори під назвою Ведмідь-гора в Криму).

Примітка 1.7. *Смислові віртуальні образи можуть усвідомлюватися інтелектом у багатовимірному віртуальному просторі, а також реальні образи на реальній площині без прив'язки до координати часу (наприклад, зображення картинок в дитячій книжці).*

У криптосемантиці слід відрізняти форми представлення структури взаємозв'язків характеристик об'єктів середовища від форм представлення первинної або вторинної інформації, яку обробляє інтелект. Особливий інтерес виявляють форми представлення первинної інформації, оскільки характеристики цих форм можуть піддаватися оптимізації в процесі створення криптосемантических систем захисту знань.

1.2.4. Основні види первинної інформації

Основні види первинної інформації можуть бути сформульовані наступним чином:

1. Інформація про характеристики об'єктів, що сприймаються як неживі, що фіксується і оцінюється за допомогою відповідних технічних засобів – сенсорів, перетворювачів і вимірювачів. В основному, йдеться про необроблені дані, представлені у формах неврегульованих послідовностей символів з алфавіту використовуваної мови представлення цих даних - шуму, шереху, даних телеметричних систем, метеорологічних даних, даних спостережень за природними космічними об'єктами, даних спостереження за неживими об'єктами. Ці дані, перш ніж їх використати, як правило,

потребують обробки, наприклад у впорядкуванні, згладжуванні, проріджуванні, сортуванні, класифікації або стискуванні. На кожному з етапів обробки цих даних використовуються відповідні форми їх представлення. Якщо дані використовуються далеко від місця їх генерації, то для їх транспортування в місця споживання можливе використання відповідних каналів передачі даних і відповідного телекомунікаційного устаткування, призначеного для узгодження характеристик каналів і сигналів, що переносять первинну інформацію. Залежно від цілей і умов застосування на різних етапах процесу передачі виникає необхідність в різних перетвореннях первинної інформації і сигналів, що її переносять (скремблювання, кодування, шифрування, модуляція, синхронізація і так далі). У цих випадках також використовуються різні форми представлення інформації і сигналів.

2. Інформація про характеристики об'єктів, живих або не живих, реальних або уявних, яка створюється, оцінюється, обробляється і фіксується безпосередньо людиною або під її контролем. В даному випадку йдеться про оброблену первинну інформацію, представлену у формах впорядкованих послідовностей символів з алфавіту використовуваної мови представлення цієї інформації, які зручні для фіксації на фізичних носіях. Це, передусім, текстова інформація, нотні записи, малюнки, креслення, слайди, мультимедіа, комп'ютерні програми і деякі інші артефакти, що мають місце у середовищі існування. Ці форми представлення інформації припускають наявність загальних тезаурусів відомостей як на стороні виникнення інформації, так і на стороні споживання інформації, і призначені для використання як в режимі реального часу (тобто, коли інформація споживається відразу ж після моменту її виникнення), так і у бажаний для споживача час. Цю інформацію можна зберігати на фізичних носіях та (або) передавати через відповідні канали зв’язку.

3. Інформація, що фіксується і оцінюється безпосередньо сенсорною системою суб'єкта, тобто інформація, яка безпосередньо представлена у

свідомості суб'єкта у вигляді впорядкованих або неврегульованих послідовностей значень показників звукових, візуальних, дотикових і інших чутливих елементів сенсорної системи цього суб'єкта. Саме ця форма представлення інформації використовується як вихідні дані в процесі функціонування інтелекту.

1.2.5. Основні форми представлення первинної інформації

Основні форми представлення первинної інформаці наступні:

1) неврегульовані послідовності символів з алфавіту використовуваної мови представлення даних, зафіксованих і оцінених без участі суб'єктів за допомогою відповідних датчиків, перетворювачів і вимірювачів;

2) впорядковані суб'єктом послідовності символів з алфавіту використовуваної мови представлення інформації;

3) впорядковані послідовності значень показників звукових, візуальних, дотикових і інших чутливих елементів сенсорної системи суб'єкта.

Як правило, форми представлення неврегульованих послідовностей символів перетворюються у форми представлення впорядкованих послідовностей символів, що зручні для подальшого осмислення перетвореної первинної інформації засобами інтелекту. У свою чергу, впорядковані послідовності символів з мови, доступної для розуміння суб'єктом, в процесі споживання інтелектом первинної інформації в реальному часі перетворюються у впорядковані послідовності значень показників звукових, візуальних, дотикових і інших чутливих елементів сенсорної системи суб'єкта. Якщо суб'єкти знаходяться усередині зон чутливості їх сенсорів, то інформаційна взаємодія між ними може здійснюватися безпосередньо шляхом генерації потоків впорядкованих змін в показниках відповідних матеріальних середовищ взаємодії цих суб'єктів.

1.2.6. Інформаційна модель суб'єкта, призначена для використання в завданнях криптосемантики

Узагальнена інформаційна модель суб'єкта, що відбиває його основну життєву функцію, – адекватно реагувати на інформацію, що потрапляє із середовища його існування (в т.ч., і від психофізіологічних систем його власного організму), показана на рис.1.1.

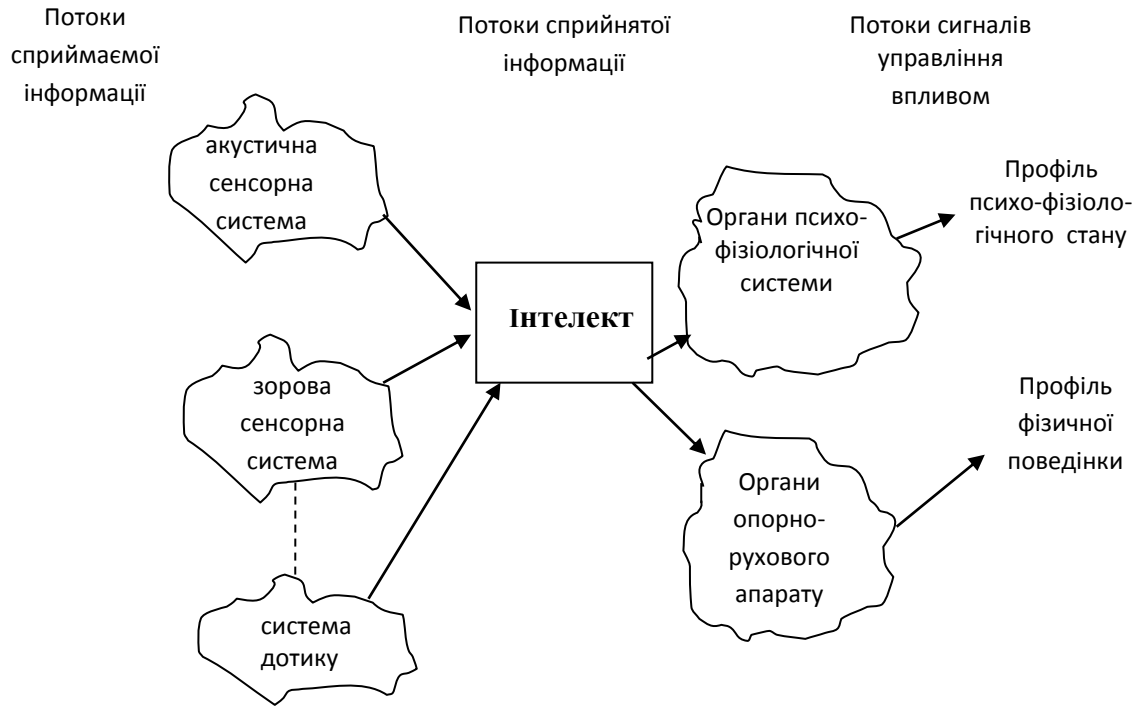

Рис.1.1. Інформаційна модель суб'єкта

На рис.1.1 відбиті наступні інформаційні потоки:

1 – потік сприймаємої первинної інформації (відображення структури середовища);

2 – потік сприйнятої первинної інформації (на виході сенсорної системи);

3 – потік сигналів управління впливом на психофізіологічну систему організму;

4 – потік сигналів управління впливом на опорно-руховий апарат суб'єкта.

Потоки інформації, що сприймаються чутливими елементами сенсорних систем суб'єкта, подаються в реальному часі на обробку засобами інтелекту. Інтелект із сприйнятого інформаційного потоку виділяє потік смислових образів, обробляє цей потік за допомогою наявних інтелектуальних засобів і, в результаті, формує потік управлінських рішень у вигляді потоку сигналів впливу на елементи психофізіологічних систем організму і опорно-рухового апарату суб'єкта. В результаті формується фізична поведінка суб'єкта, а також фізіологічний і емоційний стан його організму, які згідно з синтезованими критеріями є адекватними змінам у середовищі існування і відповідають поставленим цілям.

1.2.7. Інформаційна модель інтелекту

Оскільки інтелект в завданнях криптосемантики розглядається як програмно-керована машина, то інформаційну модель інтелекту зручно представляти у вигляді впорядкованої сукупності розгорнутих в реальному часі фізіологічних процесів живого організму, що реалізовують комплекс програм, що складається із семи компонент:

1) програма - супервайзер, що запускається на виконання в реальному часі у момент виникнення життя, яка організовує сумісне функціонування усіх процесів, що відбуваються в організмі суб'єкта. Ця програма знаходиться в активному стані упродовж усього життя суб'єкта і виконує функції, багато в чому аналогічні функціям супервайзера комп'ютерних операційних систем. У цій моделі супервайзер розглядається як основна системно-утворююча програма, яка організовує цілеспрямований аналіз, обробку і зберігання смислової компоненти інформації;

2) програма сприйняття і оцінювання інформації, що надійшла від рецепторів сенсорних систем організму, а також формування (за результатами оцінки) і запам'ятовування смислових образів;

3) комплекс програм управління персональною базою знань, основною компонентою якої є смисловий (семантичний) тезаурус (словник)

персональних знань про середовище існування, що постійно оновлюється і особливим чином структурується в реальному масштабі часу, що служить для зіставлення змісту отримуваної від середовища інформації із вже відомими знаннями про неї. (Структура тезауруса знань про середовище існування – основний об'єкт досліджень в області криптосемантики);

4) комплекс програм управління персональною базою знань, що містить інформацію про унікальну, властиву тільки цьому суб'єктові, систему персональних моральних критеріїв, пріоритетів (переваг), обмежень ("червоних ліній", табу) і звичних профілів моральної поведінки, яку ідеалісти називають "душею" суб'єкта, а матеріалісти – профілем його персональної культури (культура розуміється в широкому значенні цього слова);

5) комплекс програм управління персональною базою знань, що містить інформацію про унікальну, властиву тільки цьому суб'єктові, сукупність персональних компетенцій (умінь) суб'єкта, що забезпечують можливості активної поведінки суб'єкта;

6) програма формування управлінських рішень, що визначають фізичну поведінку суб'єкта, а також фізіологічний і емоційний стан його організму. Якість прийнятих рішень залежить не лише від результатів оцінювання інформації, що надходить, але і від профілю культури суб'єкта, а також від профілю його персональних компетенцій. Ця програма генерує потоки сигналів впливу на психофізіологічні системи організму і на опорно-руховий апарат суб'єкта;

7) комплекс програм самонавчання інтелекту, що реалізовують процеси вдосконалення тезауруса знань про середовище, профілю персональної культури і профілю персональних компетенцій суб'єкта.

Примітка 1.8. *Тезаурус знань інтелекту слід відрізняти від тезауруса відомостей, необхідного для забезпечення роботи систем передачі або зберігання інформації. Якщо в тезаурусі знань інтелекту зберігаються знання про смислові образи, то в тезаурусі відомостей системи передачі*

зберігаються відомості, які необхідно знати на передавальній і приймальній стороні перед тим, як починати сеанси передачі. Це, зокрема: відомості про алфавіти, формати, коди і словники мови представлення сигналів і (чи) передаваних повідомлень, які можуть передаватися у відкритому і (чи) захищеному від зловмисників виді; про характеристики прийнятої системи модуляції/демодуляції сигналів, використовуваних кодів з виявленням або корекцією помилок, синхронізуючої і фазуючої інформації; відомості про характеристики систем скремблювання, мультиплексування, комутації, маршрутизації і інкапсуляції і тому подібне. Зрозуміло, що тезаурус відомостей (явних і (чи) прийнятих за умовчанням) має бути однаковим на передавальній і приймальній сторонах системи передачі впродовж усього періоду її експлуатації.

Якщо використати процесний підхід до опису моделі, то інформаційна модель інтелекту відображається у вигляді рис.1.2.

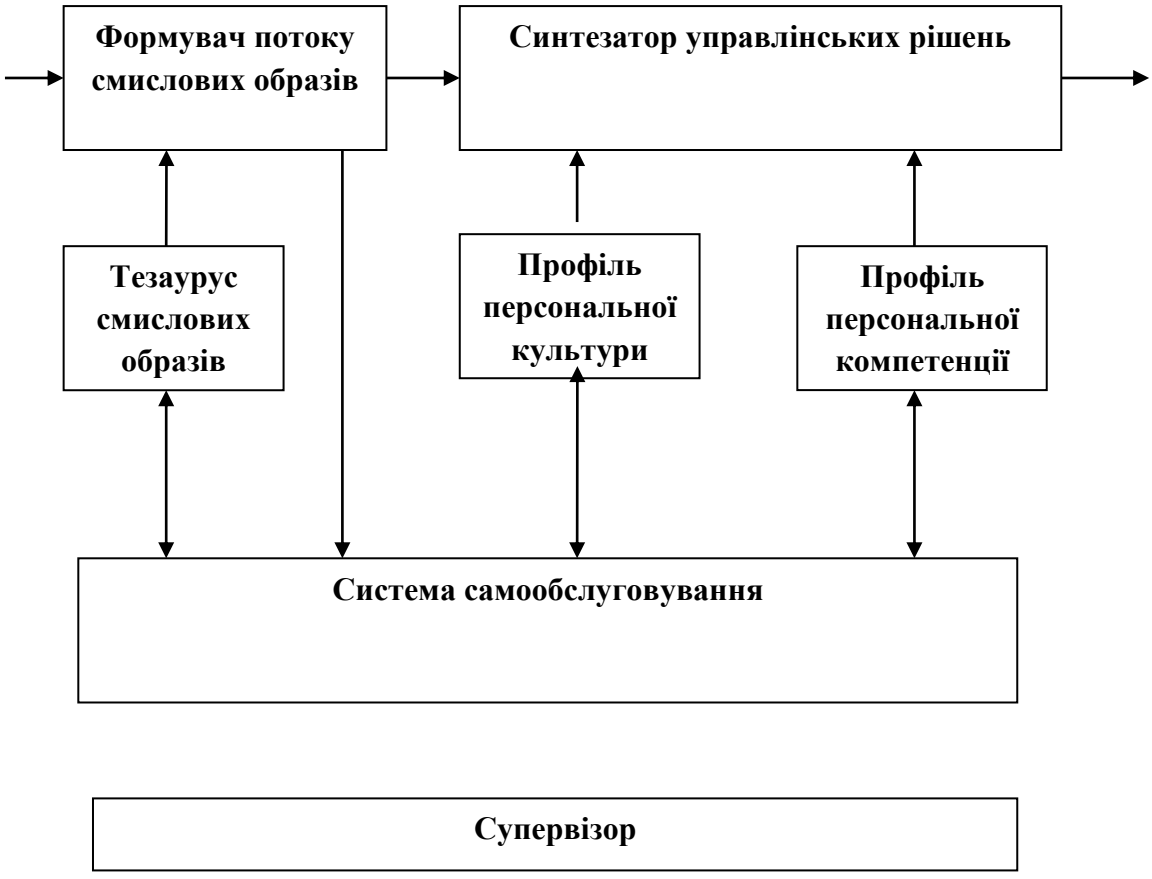

Рис.1.2. Інформаційна модель інтелекту

На рис.1.2 відображені наступні інформаційні потоки:

1 – потік сприйнятої первинної інформації (на виході сенсорної системи);

2 – потік поточних смислових образів, синтезованих формувачем;

3 – потік сигналів впливу на психофізіологічну і опорно-рухову системи організму;

4 – потік відомих зразків з тезауруса смислових образів;

5 – потік відібраних профілів поведінки, використовуваних при синтезі сигналів впливу на психофізіологічну і опорно-рухову системи організму;

6 – потік відібраних компетенцій, використовуваних при синтезі сигналів впливу на психофізіологічну і опорно-рухову системи організму;

7 – потік оновлень тезауруса смислових образів;

8 – потік оновлень профілю культури;

9 – потік оновлень профілю компетенцій.

Примітка 1.9. *При розгляді представленої моделі виникає філософське питання про існування інтегрального інтелекту, тобто про реальність існування в просторі деякого суб'єкта, інтелект якого включає інтелекти усіх раніше існуючих, нині існуючих і створених в майбутньому суб'єктів матеріального світу.*

1.3. Процес розуміння мови з позицій прикладної лінгвістики та теорії лексикографічних систем

Розумова діяльність суб’єкта заснована на використанні первинної інформації, що містить відомості про форму відображення у пам’яті суб’єкта структури співвідношень між характеристиками властивостей об’єктів середовища існування цього суб’єкту. Відповідні визначення та пояснення надано вище. Із трьох основних видів форм представлення первинної інформації у крипто-семантиці розглядається, головним чином, третя форма (див. розділ 1.2.5), а саме, потоки упорядкованих послідовностей значень показників мовних, візуальних та інших чутливих елементів сенсорних систем суб’єкту. У даній роботі досліджуються найбільш важливі з точки

зору сфер застосування два види текстової інформації – мовна усна (голосова) (рос., – речевая, устная) та мовна письмова. Перший вид відповідає голосовій інформації, що сприймається інтелектом через слуховий апарат, а другий вид, тобто письмова мова, сприймається через зір. В обох випадках маємо справу із зразками первинної інформації, що закодовані з використанням правил побудови будь-якої із відомих природних або штучних мов, що зрозумілі суб'єкту. Передбачається, що у результаті інтелектуальної обробки цих зразків може виявитися їхній смисл (тобто, утворитися смислові образи).

Семантична криптографія заснована на використанні такої властивості мовної інформації як правдоподібність. Для формального визначення показників цієї властивості необхідно, перш за все, мати модель розуміння мови, яка була б придатна для синтезу крипто-семантичних методів захисту смислу текстової інформації. Визначення цієї моделі – необхідний етап створення семантичної криптосистеми.

1.3.1. Основні визначення

Синтез моделі розуміння мови слід починати із визначень основних понять та явищ, що відносяться до прикладної лінгвістики та теорії лексикографічних систем, з урахуванням подальшої сфери застосування цих понять та явищ у задачах крипто-семантики.

Визначимо **розумову (рос. – мыслительную) діяльність** людини як психофізіологічний процес створення (генерування) віртуальних смислових потоків, що являють собою кінцево-мірні детерміновані (не випадкові) послідовності дискретних смислових образів (*SO*), яким притаманна властивість осмисленості. У процесі розумової діяльності суб'єкт може створювати як статистично, так і семантично непов'язані між собою потоки. Однак внаслідок лексикографічних ефектів, що притаманні будь-якій мові відображення смислових образів [8], окремо визначений смисловий потік можливо і доцільно розглядати як дискретну невипадкову послідовність *SO*,

що відображають хід думок людини. При цьому будемо вважати, що **смисловий образ *SO*** є дискретним елементом смислового потоку, що сприймається суб'єктом як віртуальна логічно несуперечлива смислова конструкція, що має ознаки смислової завершеності (тобто, самодостатності змісту, який не потребує обов'язкових додаткових пояснень). Оскільки окремо визначений смисловий потік (*SP*) являє детерміновану послідовність *SO*, то цьому *SP* також притаманна ознака осмисленості.

Якщо вважати, що основною формальною характеристикою як *SO*, так і *SP* є **рівень абстрагування (агрегованості, узагальненості)** їхнього представлення, то потік смислових образів (*SO*) у загальному випадку можливо представити наступним чином:

$$SP^{(i)} = SO_1^{(i-1)}, SO_2^{(i-1)}, \ldots, SO_k^{(i-1)}, \ldots, SO_N^{(i-1)}, \qquad (1.5)$$

де $SP^{(i)}$ – смисловий потік *i*-го рівня абстрактності; *i* – показник рівня абстрактності; $SO_k^{(i-1)}$ – *к*-й елемент $SP^{(i)}$, що має більш низький (тобто, більш деталізований) рівень абстрактності відображення смислової одиниці; *к* – порядковий номер *SO* у послідовності смислових образів, що відображають *SP*; *N* – довжина послідовності *SO*, що складають *SP*.

Приймемо наступну **узагальнену модель процесу розумової діяльності** суб'єкта, що наділений інтелектом. Згідно цієї моделі у пам'яті суб'єкта знаходяться його власні бази різноманітних смислових образів, що структуруються ним на основі його індивідуальних, можливо, підсвідомих уявлень (асоціацій) на доступних йому рівнях абстрактності відображення інформації. Умисно чи (та) неумисно спостерігаючи за ходом подій у середовищі перебування, суб'єкт постійно поповнює свої бази смислових образів новими для нього смисловими образами (а також постійно забуває ті смислові образи, які згідно критеріїв, що формуються на підсвідомому рівні на основі набутого досвіду, не можуть бути використані у майбутньому). Поповнення баз смислових образів може здійснюватися цілеспрямовано,

зокрема шляхом навчання та (або) несвідомо (у фоновому режимі). Проте відображення смислових образів у цих базах не завжди є достатньо конкретними, щоб користуватися ними у повсякденному житті. Тому з метою наділення набутої інформації ознаками корисності суб'єкт деталізує та певним чином структурує цю інформацію, перетворюючи її у знання. Іншими словами, на основі інформації, що міститься у базах смислових образів, суб'єкт створює свої власні бази знань у вигляді ієрархічно структурованих щодо рівнів абстрагування детермінованих кінцевих дискретних послідовностей смислових образів. Сукупність множин дискретних послідовностей смислових образів, упорядкованих за рівнями абстрагування відображення смислу, що є доступна суб'єкту для оперування у повсякденному житті, визначає тезаурус його інтелекту.

Визначимо **простір смислових образів суб'єкта** як кінцеву множину смислових образів, яка доступна цьому суб'єкту в процесі здійснення розумової діяльності, а також **простір смислових образів області (сфери) прикладних застосувань** як кінцеву множину смислових образів, що використовують різні суб'єкти в процесі вирішення завдань у рамках визначеної області прикладних застосувань.

Текстова інформація - це відображення (у рамках обраної мови) розумової діяльності людини (або штучного інтелекту) у вигляді упорядкованих за смисловими ознаками послідовностей текстових одиниць (фраз, речень, слів, букв, фонем тощо): послідовності відображень текстових одиниць у вигляді звуків, що генеруються людиною, якщо мається на увазі усна мова; послідовності символьних позначень текстових одиниць, що відображаються відповідними технічними пристроями у комп'ютерах, на паперових, електронних та інших фізичних носіях інформації, якщо йдеться про письмову мову.

З урахуванням вищезазначеного визначимо **тезаурус** *TZ* як структурований у вигляді прошаркової (рос. – слоистой) коренево-подібної ієрархії простір смислових образів *SO*, що упорядковані за рівнями

абстрагування цих смислових образів у напрямі усе більш конкретного (більш деталізованого) їхнього сприйняття суб'єктами. За таким визначенням можливо говорити про **тезаурус суб'єкта** TZ_S як про упорядкований набір *SO*, що зафіксовані у його пам'яті. Тезаурус суб'єкта – це його база знань. Можливо також говорити про **тезаурус прикладної області** TZ_M як про упорядкований набір тлумачних (сценарних, фразеологічних, семантичних тощо) словників, що упорядковані за рівнями абстрагування смислових образів. Словник, що визначає сукупність *SO* *i*-го рівня абстрагування, представляє тезаурус семантичних одиниць $TZ^{(i)}$ *i*-го рівня абстрагування. Диз'юнкція тезаурусів *TZ* усіх можливих прикладних областей, що охоплює усю множину певним чином визначених послідовностей смислових образів, доступних певній спільноті людей, складає **тезаурус** TZ_O цієї **спільноти** людей:

$$TZ_O = TZ_1 \; or \; TZ_2 \; or, \; \ldots \; or \; TZ_T \,, \qquad (1.6)$$

де *T* – кількість врахованих у TZ_O прикладних областей використання тезаурусу.

1.4. Формальна модель розуміння інформації

Найбільш абстрактна модель розуміння мовної інформації суб'єктом розумової діяльності (тобто, людиною або комп'ютерною програмою), що запропонована, зокрема, у [29], показана на рис.1.3, де прийняті наступні позначення:

S – суб'єкт розумової діяльності (людина, комп'ютерна програма);

D – зразок текстового повідомлення, що осмислюється суб'єктом;

F_D – результат неосмисленого сприйняття відображення форми досліджуваного зразка текстового повідомлення (форми символів алфавіту та спеціальних символів, з використанням котрих складено зразок D, місць розташування символів у дискретній послідовності цих символів, структури

текстового повідомлення як одного із можливих варіантів побудови форми повідомлення відповідно до прийнятих правил граматики використаної мови тощо);

C_D – результат сприйняття смислу досліджуваного зразка текстового повідомлення;

S_F – оператор відображення форми досліджуваного зразка текстового повідомлення на основі його безпосереднього сприйняття суб'єктом;

S_C – оператор безпосереднього відображення смислу досліджуваного зразка текстового повідомлення без аналізу його форми;

H – оператор відображення смислу досліджуваного зразка текстового повідомлення на основі аналізу його форми.

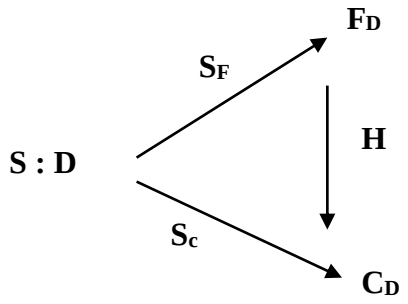

Рис.1.3. Абстрактна модель розуміння текстової інформації

Дамо наступну інтерпретацію вищезазначеної моделі. У процесі вирішення прикладних завдань суб'єкт формує упорядковані щодо смислу послідовності мовних одиниць шляхом осмисленого вибору цих одиниць із доступного йому тезаурусу. Так що, **думка (рос. – мысль)** будь-якого суб'єкту являє собою дискретний часовий ряд смислових образів, що вибираються ним (зокрема, за допомогою асоціативних механізмів) із доступного тезаурусу. У процесі навчання або творчої діяльності суб'єкт намагається розширити та (або) підвищити якість тезаурусу TZ_S шляхом уведення у його склад додаткових мовних одиниць та (або) уточнення смислового змісту вже відомих мовних одиниць. Іноді суб'єкт намагається

модернізувати структуру тезаурусу шляхом збільшення кількості рівнів абстрагування відображення смислових образів.

Суб’єкти (групи суб’єктів) різняться за характеристиками своїх тезаурусів – як за складом і кількістю доступних смислових образів, так і за складом і кількістю доступних рівнів абстрагування їхнього представлення.

Індивіду може бути доступним кілька різнорідних просторів смислових образів – музичних, візуальних, поетичних, смакових, нюхових тощо. Однак більшість індивідів для вирішення більшості життєво важливих завдань, що пов’язані із застосуванням інтелекту та накопичених знань, використовує механізми природної мови. Простір смислових образів природної мови – це простір текстових одиниць цієї мови, оскільки, зазвичай, людина мислить мовними образами [28,29]. У подальшому будемо розглядати лише простори мовних одиниць і, отже, під назвою «тезаурус» будемо розуміти лише простори мовних одиниць.

Висновки до першого розділу

1. Існуючі визначення і моделі, що використовують у задачах теорії інформації і зв'язку, не повною мірою придатні для опису систем семантичної криптографії – систем, призначених для захисту смислового змісту інформації від перехоплення суб'єктами, що не мають відповідних прав легального доступу до неї. Для адекватного опису цих систем у даній роботі уточнені і (або) перевизначені відомі поняття, такі як середовище, суб'єкт, інтелект, семантичний тезаурус, інформація, форма представлення і сенс (зміст) інформації, а також представлена модель осмислення суб'єктом змісту інформації стосовно завдань семантичної криптографії. Інформація в цій роботі визначена як закодоване відображення у свідомості суб'єкта структури взаємозв'язків і характеристик властивостей компонентів середовища (чи середовища в цілому), аналізоване інтелектом в категоріях "форма - суть". Інформація включає відомості (дані) про форму структури, аналізуючи які можна отримати відомості про суть структури, зокрема

виявити смислові образи, що містяться у формі аналізованої структури. Так що, згідно з цим визначенням інформація містить відомості як про форму, так і про суть структури, що дозволяє відокремити процеси, пов'язані з обробкою відомостей про форму структури (методи обробки саме цих відомостей розглядаються в теорії зв'язку), від процесів, пов'язаних з обробкою відомостей про суть структури (методи захисту цих відомостей розглядаються в семантичній криптографії).

2. Представлена узагальнена інформаційна модель суб'єкта, що придатна для вирішення завдань семантичної криптографії. Модель відбиває основну життєву функцію суб'єкта – адекватно реагувати на інформацію, що поступає із середовища його існування. Згідно цієї моделі потоки інформації, що сприймаються чутливими елементами сенсорних систем суб'єкта, подаються в реальному часі на обробку засобами інтелекту. Інтелект із сприйнятого інформаційного потоку виділяє потік смислових образів, обробляє цей потік за допомогою наявних інтелектуальних засобів і, в результаті, формує потік управлінських рішень у вигляді потоку сигналів впливу на елементи психофізіологічних систем організму і опорно-рухового апарату суб'єкта.

3. Уперше представлена інформаційна модель інтелекту, що визначає його структуру і деталізує його функції в аспектах, що розглядають під час вирішення завдань семантичної криптографії. Інтелект суб'єкта у рамках даної моделі розглядається як програмно-керована машина, що реалізовує в реальному часі комплекс програм підтримки впорядкованої сукупності фізіологічних процесів в організмі суб'єкта. Пояснено функції шести основних інформаційних модулів цього комплексу, керованих супервайзером – основною системо-утворюючою програмою, що організує функціонування інтелекту по цілеспрямованому аналізу, обробці і зберіганню потоку смислових образів, витягуваних з потоку інформації, що надходить. Функціональність цих модулів передбачає можливість: 1) сприйняття і оцінювання потоків інформації, що надходять від рецепторів сенсорних систем організму, а також формування (за результатами оцінки) і

запам'ятовування потоків смислових образів; 2) підтримки і оновлення тезауруса персональних знань суб'єкта; 3) підтримки і оновлення профілю персональної культури суб'єкта (системи персональних моральних критеріїв, переваг і обмежень); 4) підтримки і оновлення профілю персональних компетенцій; 5) формувань управлінських рішень, що визначають фізичну поведінку суб'єкта, а також фізіологічний і емоційний стан його організму; 6) самонавчань інтелекту, в процесі якого реалізуються процеси вдосконалення тезауруса знань про середовище, профілю персональної культури і профілю персональних компетенцій суб'єкта.

4. Вперше розроблено формальну модель розуміння мовної інформації, відповідно до якої у процесі розумової діяльності суб’єкт генерує детерміновані скінчені дискретні часові ряди смислових образів, що вибираються ним (зокрема, за допомогою асоціативних механізмів) із доступного йому тезаурусу. Вперше тезаурус інтелекту суб’єкта визначено як структурований у вигляді прошаркової (рос. – слоистой) кореневоподібної ієрархії набір смислових образів, що упорядковані за рівнями абстрактності цих смислових образів у напрямі усе більш конкретного (більш деталізованого) їхнього сприйняття суб’єктом. Показано, що якість тезаурусу визначається як розмірністю бази смислових образів, якими здатен оперувати суб’єкт у процесі розумової діяльності, так і кількістю рівнів абстрактності відображення смислових образів у структурі тезаурусу.

5. Представлені в роботі результати слід розглядати як основні базові відомості, необхідні для розуміння роботи крипто-семантических систем захисту смислового змісту інформації, застосування яких на практиці у ряді областей, як показано в роботах [10,11], є доцільним.

РОЗДІЛ 2
РОЗРОБКА МЕТОДУ ПОБУДОВИ СЕМАНТИЧНИХ ТЕЗАУРУСІВ ДЛЯ ЛЕКСИКОГРАФІЧНИХ КРИПТОСИСТЕМ

Неодмінним елементом будь-якої досконало стійкої криптосистеми, що заснована на застосуванні певним чином побудованої лексикографічної системи, є тезаурус бази захисту інформації у прикладній системі, де ця криптосистема використовується. У даному випадку тезаурус – це семантичний словник, структура якого відображає структуру семантичних зв'язків між смисловими конструкціями мови відображення прикладної області його застосування. У досконало стійких криптосистемах структура семантичних зв'язків між елементами тезауруса має бути відображена на формальному рівні. Визначення цієї структури є основним завданням даного розділу.

2.1. Структура тезаурусу смислових образів

2.1.1. Формальна структура тезаурусів

Наповнимо описану у розділі 1 абстрактну модель розуміння текстової інформації конкретним змістом, зокрема визначимо на формальному рівні структуру тезаурусу смислових образів мови відображення області розумової діяльності суб'єкта. Будемо вважати, що тезаурус будь-якої мови взагалі TZ_M або будь-якого суб'єкту окремо TZ_S має ієрархічну прошаркову (рос. – слоистую) структуру і за ступенем абстрагування відображення смислових образів розподіляються на i рівнів, де $i = 1, 2, \ldots, I$ – кількість рівнів абстрагування відображення смислових образів, якими оперує колективний інтелект носіїв цієї мови взагалі або індивідуальний інтелект суб'єкту розумової діяльності окремо, а I_{max} – максимальна кількість рівнів абстрагування відображення SO, що є доступною інтелекту. Так що, простір смислових образів, доступний суб'єкту (або групі суб'єктів), тобто його тезаурус TZ_S, є дискретним, кінцево-мірним, який щодо рівнів абстрагування

представлення образів має прошаркову коренево-подібну структуру (див. рис. 2.1).

За таким визначенням будь-який тезаурус *TZ* (зокрема, TZ_M або TZ_S) складається із сукупності підтезаурусів семантичних одиниць усіх доступних для розуміння рівнів абстрагування $TZ^{(i)}$, де $i=1,2, \dots, I$. При цьому підтезауруси у складі *TZ* розташовані прошарками у вигляді гілко-подібної кореневої системи, що розростаються зверху униз. В основі кореня лежить семантичний образ $SO^{(I+1)}$ з максимальним (I+1)-им рівнем абстрагування відображення *SO*, що є доступним інтелекту у конкретній області розумової діяльності. Перший знизу від основи кореня прошарок підтезаурусів, що складають множину $\{TZ^{(I)}_k\}$, визначають простір $\{SO^{(I)}_k\}$, що конкретизують $SO^{(I+1)}$ із I-им рівнем абстрагування відображення смислових образів. Другий знизу від основи кореня прошарок підтезаурусів, що складають множину $\{TZ^{(I-1)}_{k,l}\}$, визначають простір $\{SO^{(I-1)}_{k,l}\}$, що конкретизують $SO^{(I)}$ із більш детальним (I-1)-им рівнем абстрагування відображення смислових образів. І т.д. – від більш узагальненого до більш конкретного відображення смислових образів.

Так що, структуру тезаурусу *TZ* у загальному випадку можливо представити у вигляді рекурентної коренево-подібної схеми як

$$TZ \in \{TZ^{(I)}_k\}, \text{ де } TZ^{(I)}_k \in \{TZ^{(I-1)}_{k,l}\}, \text{ де } TZ^{(I-1)}_{k,l} \in \{TZ^{(I-2)}_{k,l,n}\}, \dots$$
$$K \qquad\qquad L \qquad\qquad N$$
$$\dots\dots\dots, \text{ де } TZ^{(2)}_{k,l,n,\dots,p} \in \{SO^{(1)}_{k,l,n,\dots,j}\}. \qquad (2.1)$$
$$J$$

У виразі (2.1) прийнято наступні позначення:

$TZ^{(I)}_k$ – тезаурус I-го рівня абстрагування представлення *SO*;

$I \in \{1, 2, \dots\ i, \dots, \dots, I_{max}\}$ – кількість рівнів абстрагування відображення *SO*, що доступна інтелектові; I_{max} – теоретично будь-яке велике, але скінчене ціле; $k \in \{1, 2, \dots, K\}$ – порядковий номер елемента $TZ^{(I)}_k$ у множині

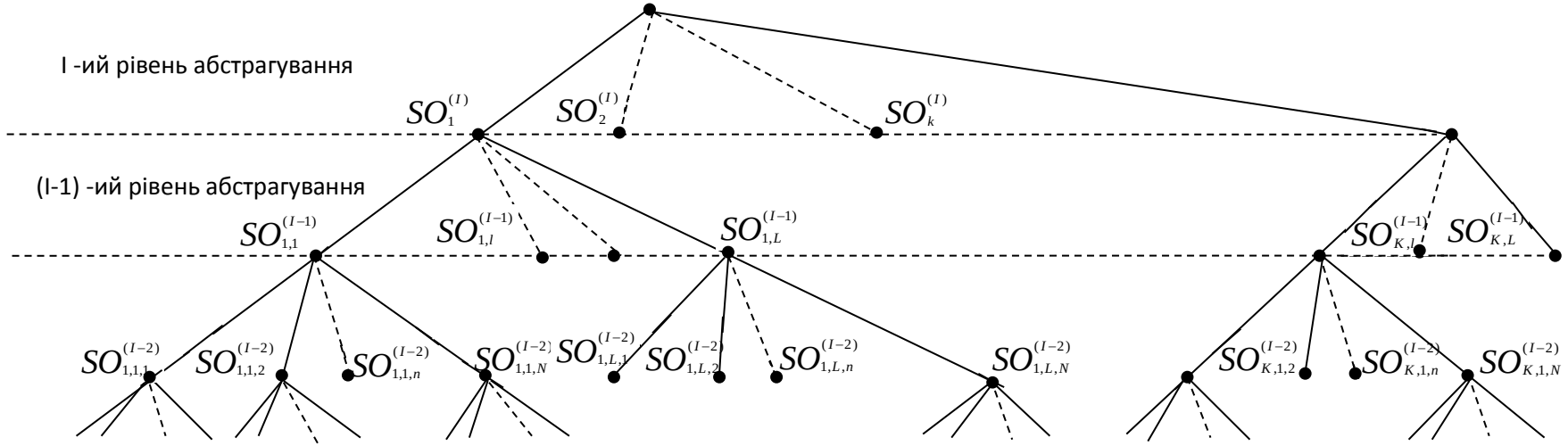

Рис. 2.1. Прошаркова коренево-подібна структура простору смислових образів, що складають тезаурус

тезаурусів, що у сукупності визначають простір смислових образів *I*-го рівня абстрагування, тобто $\{SO^{(I)}\}$; *K* – кількість тезаурусів *I*-го рівня абстрагування відображення *SO* , що входять до складу *TZ*.

$TZ^{(I-1)}_{k,\,l}$ – тезаурус (*I-1*)-го рівня абстрагування представлення *SO*, що конкретизує смислові образи $TZ^{(I)}_k$, де $l \in \{1,2, \ldots, L\}$ – порядковий номер тезаурусу (*I-1*)-го рівня абстрагування у множині тезаурусів, які у сукупності визначають простір смислових образів (*I-1*)-го рівня абстрагування у рамках тезаурусу $TZ^{(I)}_k$; *L* – кількість тезаурусів (*I-1*)-го рівня абстрагування відображення *SO*, що входять до складу $TZ^{(I)}_k$;

$TZ^{(I-2)}_{k,\,l,\,n}$ – тезаурус (*I*-2)-го рівня абстрагування представлення *SO*, що конкретизує смислові образи $TZ^{(I-1)}_{k,\,l}$, де $n \in \{1,2, \ldots, N\}$ – порядковий номер тезаурусу (*I-2*)-го рівня абстрагування у тезаурусі $TZ^{(I-1)}_{k,,l}$; *N* – кількість тезаурусів (*I-2*)-го рівня абстрагування відображення *SO*, що входять до складу $TZ^{(I-1)}_{k,\,l}$;

$TZ^{(2)}_{k,\,l,\,n,\,\ldots,\,p}$ – тезаурус другого рівня абстрагування представлення *SO*, що конкретизує смислові образи $TZ^{(3)}_{k,\,l,n,\,\ldots,s}$, де $p \in \{1,2, \ldots, P\}$ – порядковий номер тезаурусу другого рівня абстрагування у складі тезаурусу третього рівня $TZ^{(3)}_{k,\,l,n,\,\ldots,s}$, де $s \in \{1,2, \ldots, S\}$ – порядковий номер тезаурусу третього рівня абстрагування; *P* – кількість тезаурусів другого рівня абстрагування, що входять до складу $TZ^{(3)}_{k,\,l,n,\,\ldots,s}$; *S* – кількість тезаурусів третього рівня абстрагування, що входять до складу відповідного тезаурусу четвертого рівня абстрактності і т.д. уздовж ланцюга тезаурусів із зростанням значення індексу *i*;

$SO^{(1)}_{k,\,l,\,n,\,\ldots,\,j}$ – семантичний словник, що відображає тезаурус $TZ^{(2)}_{k,l,\,n,\,\ldots,\,p}$; *J* – кількість слів у тезаурусі $TZ^{(2)}_{k,l,\,n,\,\ldots,\,j}$.

Отже, структура тезаурусу *TZ* представляється у вигляді розгалуженого кореня підтезаурусів $TZ^{(i)}$, де $i \in \{1, 2, \ldots, \ldots, I_{max}\}$.

Примітка 2.1. *У подальшому з метою спрощення викладу підтезауруси будемо називати тезаурусами. Тобто, будемо розглядати* TZ_M *або* TZ_S *як сукупність тезаурусів, що упорядковані за рівнями абстрагування представлення семантичних одиниць.*

2.1.2. Структура тезаурусів мови відображення прикладної області

Якщо розглядати структуру тезаурусів мови відображення будь-якої прикладної області відповідно до розтину за рівнями абстрагування відображення смислових образів (див. рис. 2.1), то у загальному випадку доцільно задати наступну ієрархію її семантичних одиниць:

прикладна область (напрямок знань)/ тема/ сценарій/ ситуація/

(2.2)

фраза/ слово/ символ алфавіту/ код символу алфавіту.

Зокрема, у багатьох сферах прикладних застосувань, які у подальшому назвемо областями активності, доцільно призначити наступну ієрархію смислових одиниць, що відображають логічно завершені думки певного рівня абстрагування, де убування ступеня абстрагування прийнято зліва направо:

сценарій/ ситуація / фраза / слово. (2.3)

У цьому випадку, у виразі (2.1) $I = 4$, тобто будемо мати чотирьох ступеневу ієрархічну структуру у вигляді розгалуженого кореня тезаурусів з різним рівнем абстрагування представлення смислових образів. Значення параметрів K, L, M, N, S та P у виразі (2.1) необхідно визначати, виходячи із феноменології заданої області активності.

Під сценарієм (або темою у рамках визначеної області активності) розуміється упорядкована щодо смислу послідовність ситуацій, під ситуацією – упорядкована щодо смислу послідовність фраз, під фразою – упорядкована щодо смислу послідовність слів. А під словом (письмової мови) – упорядкована щодо смислу послідовність символів обраного алфавіту.

Область активності – це область розумової діяльності, яка відображається заданим простором смислових образів (які, у свою чергу, відображаються відповідним простором мовних одиниць усіх рівнів абстрагування у рамках обраної мови). Бажано, щоб цей простір у повній мірі відображав визначену

область активності. Зокрема, наприклад, з теоретичної точки зору національний лінгвістичний корпус, що розроблений та підтримується Українським мовно-інформаційним фондом НАН України [28], може бути використано для синтезу універсального простору мовних одиниць, що охоплює максимально можливу кількість областей розумової діяльності. Однак створення такого простору через його велику розмірність потребує занадто великих ресурсних витрат. На практиці область активності обмежується визначеною сферою прикладних застосувань.

Наприклад, якщо хочемо побудувати тезаурус TZ_M мови відображення такої прикладної області як фразеологія радіообміну голосовими повідомленнями між авіадиспетчером та пілотом цивільного літака, що має здійснюватися у відповідності з регламентом [42], то простір смислових образів *I*-го рівня абстрагування має складатися із *K* можливих сценаріїв. А простір смислових образів (*I-1*)-го рівня абстрагування має складатися із *L* можливих ситуацій. А простір смислових образів (*I-2*)-го рівня абстрагування має складатися із *M* можливих фраз. На кінець, простір смислових образів (*I-3*)-го рівня абстрагування має складатися із *N* можливих слів. Якийсь *к*-ий сценарій відображається кінцево-мірною послідовністю ситуацій, що обираються суб'єктом із множини смислових образів ситуацій і які у сукупності являють смисловий образ сценарію. У свою чергу, якась *l* –та ситуація відображається кінцево-мірною послідовністю фраз, що обираються суб'єктом із множини смислових образів фраз і які у сукупності являють смисловий образ ситуації. І т.д. – до більш детального відображення смислу текстового повідомлення.

2.2. Формальне відображення семантики текстової інформації

Відобразимо зразок текстової послідовності на рівні семантичних одиниць *i*-го рівня абстрагування наступним чином:

$$SO^{(i)}_{k(1)}, SO^{(i)}_{k(2)}, \dots, SO^{(i)}_{k(n)}, \dots \; SO^{(i)}_{k(Ni)}, \qquad (2.4)$$

де k – порядковий номер семантичної одиниці $TZ^{(i)}$ у тезаурусі $TZ^{(i+1)}$, тобто $k = 1, 2, \ldots, K$. Значення індексу k генеруються інтелектом суб'єкта під час формування смислового образу тексту і залежить від порядкового номеру n семантичної одиниці у тексті, $n=1,2, \ldots, Ni$. Параметр Ni визначає розмірність зразка текстової послідовності.

Зрозуміло, що запис (2.4) буде справедливим, якщо для будь-якого елементу зразка $SO^{(i)}$ знайдеться відповідний тезаурус $TZ^{(i)}$ у тезаурусі $TZ^{(i+1)}$, тобто якщо $SO^{(i)} \equiv TZ^{(i)}$, де символ ≡ означає відношення смислової ідентичності.

У свою чергу, кожен елемент випадкової послідовності (2.4) може бути представлений послідовністю більш конкретного змісту – на рівні семантичних одиниць (i-1)-го рівня абстрагування:

$$SO^{(i)}_k \rightarrow SO^{(i-1)}_{k,\,l(1)}, SO^{(i-1)}_{k,\,l(2)}, \ldots, SO^{(i-1)}_{k,\,l(Nl)}, \qquad (2.5)$$

а кожен елемент випадкової послідовності (2.5) має бути представлений послідовністю іще більш конкретного змісту – на рівні семантичних одиниць (i-2)-го рівня абстрагування:

$$SO^{(i-1)}_{k,l} \rightarrow SO^{(i-2)}_{k,l,m(1)}, SO^{(i-2)}_{k,l,m(2)}, \ldots, SO^{(i-2)}_{k,l,m(Nm)}. \qquad (2.6)$$

На кінець, найбільш конкретний рівень представлення смислу текстових повідомлень здійснюється тоді, коли кожен елемент другого рівня абстрагування буде являти собою послідовність слів, тобто послідовність семантичних елементів першого рівня абстрагування:

$$SO^{(2)}_{k,l,n,\ldots,p} \rightarrow SO^{(1)}_{k,l,n,\ldots,j(1)}, SO^{(1)}_{k,l,n,\ldots,j(2)}, \ldots, SO^{(1)}_{k,l,n,\ldots,j(Nj)}. \qquad (2.7)$$

Надані вище семантичні відношення можливо визначити між смисловими одиницями, що відображають логічно завершені думки певного рівня абстрагування, зокрема на рівні ситуацій, на рівні фраз або на рівні слів.

2.3. Смислові відношення між семантичними одиницями

Метод побудови досконало стійкої криптосистеми, що пропонується у даній роботі, базується на використанні лексикографічних ефектів, що пов'язані із такою властивістю текстової (голосової чи письмової) інформації як можливість її неоднозначного сприйняття суб'єктом. Механізми такої криптосистеми мають забезпечити створення умов, за яких зашифровані зразки текстової інформації сприймаються криптоаналітиком як правдоподібні з невизначеною ймовірністю їхньої появи у шифрограмі, що не дає йому змогу відрізнити вихідні відкриті зразки повідомлень, що були зашифровані, від інших правдоподібних відображень цих повідомлень, що утворились у результаті шифрування. Отже, щоб побудувати будь-яку крипто-семантичну систему захисту, необхідно надати формальні визначення таким смисловим відношенням між семантичними одиницями як **смислова ідентичність, смислова відмінність**, сумнівна смислова ідентичність (або **правдоподібність**), сумнівна смислова відмінність (або **фальшиво-подібність**). Тобто, щодо смислових відношень між смисловими образами необхідно визначити показники смислової ідентичності, смислової відмінності, правдоподібності та фальшиво-подібності.

Відмінність між правдоподібністю та фальшиво-подібністю полягає в особливостях суб'єктивного сприйняття смислових відношень. Якщо смисловий образ сприймається суб'єктом як ймовірний щодо появи у відкритому тексті, але визначити ймовірність його появи не виявляється можливим, а у суб'єкта існують небезпідставні сумніви щодо істинності такого сприймання, то у цьому випадку доцільно визначити смисловий образ як правдоподібний. Однак, якщо смисловий образ сприймається суб'єктом як неймовірний щодо появи у відкритому тексті, але у нього існують небезпідставні сумніви щодо істинності такого сприймання, то у цьому випадку

доцільно визначити смисловий образ як фальшиво-подібний. Наприклад, якщо криптоаналітик бачить перед собою уривок тексту, що являє беззмістовний набір символів алфавіту, який утворився, на його думку, не внаслідок випадкового збігу зовнішніх обставин, а внаслідок цілеспрямованої діяльності шифрувальника, то, з точки зору криптоаналітика, смисловий зв'язок між отриманим уривком тексту та можливим вихідним відкритим зразком текстової інформації має ознаки фальшиво-подібності.

Побудована вище модель розуміння мови та створений на основі цієї моделі формалізм структури тезаурусів, яким оперує суб'єкт під час розумової діяльності, дозволяють визначити вищезазначені показники наступним чином.

Обмеження щодо області визначення смислових відношень

Смислові відношення у даній роботі визначаються за таких умов:

1) семантичні одиниці, між котрими визначаються смислові відношення, мають належати одному і тому ж тезаурусу мови, що відображає певним чином обрану область активності;

2) смислові відношення визначаються між семантичними одиницями будь-якого, але однакового рівня абстрагування.

Відношення смислової ідентичності

Якщо серед множини семантичних одиниць, що представляють тезаурус *TZ*, існує кілька одиниць, що є ідентичними за сутністю змісту, то ці одиниці щодо смислу знаходяться у відношенні смислової ідентичності (синоніми). Тобто, будь-які дві семантичні одиниці будь-якого, але однакового рівня абстрагування $TZ_a{}^{(i)}{}_k$ та $TZ_b{}^{(i)}{}_k$ із множини $\{TZ^{(I)}{}_k\} \rightarrow TZ$ знаходяться у відношенні смислової ідентичності

$$TZ_a{}^{(i)}{}_k \equiv TZ_b{}^{(i)}{}_k, \qquad (2.8)$$

якщо є ідентичними за сутністю змісту. Позначка $\rightarrow$ означає приналежність будь-якого тезаурусу $TZ^{(i)}{}_k$ із множини тезаурусів *i*–го рівня абстрагування до тезаурусу обраної мови *TZ*, а у виразі (2.8) позначка ≡ означає відношення смислової ідентичності.

Зрозуміло, якщо

$$TZ_a{}^{(i)}{}_k \equiv TZ_b{}^{(i)}{}_k, \text{ а } TZ_b{}^{(i)}{}_k \equiv TZ_c{}^{(i)}{}_k, \text{ то } TZ_a{}^{(i)}{}_k \equiv TZc^{(i)}{}_k. \quad (2.9)$$

Зрозуміло також, що

$$SO^{(i)}{}_k \equiv TZ^{(i)}{}_k \text{ для будь-яких значень } i \text{ та } k, \quad (2.10)$$

тобто, будь-який елемент будь-якого тезаурусу, що входить до складу тезаурусу будь-якої мови будь-якої області активності відображає відповідний смисловий образ.

Відношення смислової відмінності

Якщо серед множини семантичних одиниць, що представляють тезаурус *TZ*, існує кілька одиниць, що не є ідентичними за сутністю змісту, то ці одиниці щодо смислу знаходяться у відношенні смислової відмінності. Тобто, семантична одиниця *i*-го рівня абстрактності $TZ_a{}^{(i)}{}_k$ із множини $\{TZ^{(I)}{}_k\} \rightarrow TZ$ та семантична одиниця $TZ_b{}^{(i)}{}_l$ із множини $\{TZ^{(I)}{}_l\} \rightarrow TZ$ знаходяться у відношенні смислової відмінності

$$TZ_a{}^{(i)}{}_k \triangle TZ_b{}^{(i)}{}_l, \quad (2.11)$$

якщо не є ідентичними за сутністю змісту. Позначка $\triangle$ означає смислову відмінність.

Зрозуміло, що різні гілки будь-якого, але одного рівня абстрагування представлення *SO* у корені, що представляє структуру *TZ* (див. рис. 2.1), знаходяться у стані смислової відмінності.

Відношення правдоподібності

Якщо порівнювальні смислові образи сприймаються суб'єктом як ідентичні щодо смислу, але у нього існують небезпідставні сумніви щодо істинності такого сприймання, то у цьому випадку доцільно визначити

смислове відношення як правдоподібне. Тобто, якщо в якості позначки відношення правдоподібності обрати символ ^ та врахувати, що $SO^{(i)}{}_k \equiv TZ^{(i)}{}_k$ для будь-яких значень *i* та *k*, то будь-які дві семантичні одиниці будь-якого, але однакового рівня абстрагування $TZ_a{}^{(i)}{}_k$ та $TZ_b{}^{(i)}{}_k$ із множини $\{TZ^{(I)}{}_k\} \rightarrow TZ$ знаходяться у відношенні смислової правдоподібності

$$TZ_a{}^{(i)}{}_k \wedge TZ_b{}^{(i)}{}_k, \qquad (2.12)$$

якщо ймовірність того, що $TZ_a{}^{(i)}{}_k \equiv TZ_b{}^{(i)}{}_k$ є меншою, ніж 1.

Підкреслимо, що відношення правдоподібності визначаються між семантичними одиницями будь-якого, але однакового рівня абстрагування.

Відносно семантичних одиниць *I*-го (найвищого у даній мові даної області активності) рівня абстрагування буде справедливим наступне твердження:

$$SO^{(I)}{}_1 \wedge SO^{(I)}{}_2 \wedge \ \ldots \wedge SO^{(I)}{}_k \wedge \ldots \wedge SO^{(I)}{}_K, \text{ якщо } \{TZ^{(I)}{}_k\} \underset{K}{\rightarrow} TZ \ , \qquad (2.13)$$

за умови $SO^{(I)}{}_k \equiv TZ^{(I)}{}_k$, для будь-яких значень *I* та *k*. (2.14)

У виразі (2.13) позначка → означає приналежність будь-якого тезаурусу $TZ^{(I)}{}_k$ із множини тезаурусів *I*-го рівня абстрагування до тезаурусу обраної мови *TZ*, а у виразі (2.14) позначка ≡ означає відношення смислової ідентичності.

Відношення фальшиво-подібності

Якщо порівнювальні смислові образи сприймаються суб'єктом як відмінні щодо смислу, але у нього існують небезпідставні сумніви щодо істинності такого сприймання, то у цьому випадку доцільно визначити смислове відношення як фальшиво-подібне. Тобто, якщо в якості позначки відношення фальшиво-подібності обрати символ ¦ та врахувати, що $SO^{(i)}{}_k \equiv TZ^{(i)}{}_k$ для будь-яких значень i та k, то будь-які дві семантичні одиниці будь-якого, але однакового рівня абстрагування $TZ_a{}^{(i)}{}_k$ із множини $\{TZ^{(I)}{}_k\} \rightarrow TZ$ та $TZ_b{}^{(i)}{}_l$ із

множини $\{TZ^{(I)}{}_l\} \rightarrow TZ$ знаходяться у відношенні смислової фальшиво-подібності

$$TZ_a{}^{(i)}{}_k \,\dot{|}\, TZ_b{}^{(i)}{}_l, \qquad (2.15)$$

якщо ймовірність того, що $TZ_a{}^{(i)}{}_k \triangle TZ_b{}^{(i)}$ є меншою, ніж 1.

2.4. Структура тезаурусу бази захисту

Структура тезаурусу прикладної системи TZ_{PS} відображена на рис. 2.1 і представляється рекурентною схемою (2.1). Як бачимо, тезаурус TZ_{PS} містить усі можливі семантичні образі на усіх заданих рівнях абстрагування, що у сукупності складають мову відображення цієї прикладної області. Так що, семантичні відношення між семантичними образами будь-якого одного рівня абстрагування можуть бути семантично відмінні, правдоподібні або фальшиво-подібні. Оскільки будь-який лексикографічний метод захисту текстової інформації базується на випадковій заміні істинного смислового образу повідомлення на інший правдоподібний смисловий образ, то користування тезаурусом TZ_{PS} для здійснення таких замін не уявляється можливим. Лексикографічна система у складі досконало стійкої криптосистеми має спиратися на тезаурус, усі елементи котрого пов’язані між собою відношенням правдоподібності, коли будь-яка заміна одного семантичного образу на інший не порушує відношення правдоподібності. Такий тезаурус назвемо тезаурусом бази захисту інформації TZ_{BZ} у прикладній системі, що відображається тезаурусом TZ_{PS}.

Структура TZ_{BZ} показана на рис. 2.2.

Для захисту смислу текстових повідомлень пропонується наступний методологічний підхід, що заснований на використанні відношень смислової правдоподібності. Для того, щоб криптоаналітик ні за яких умов не мав

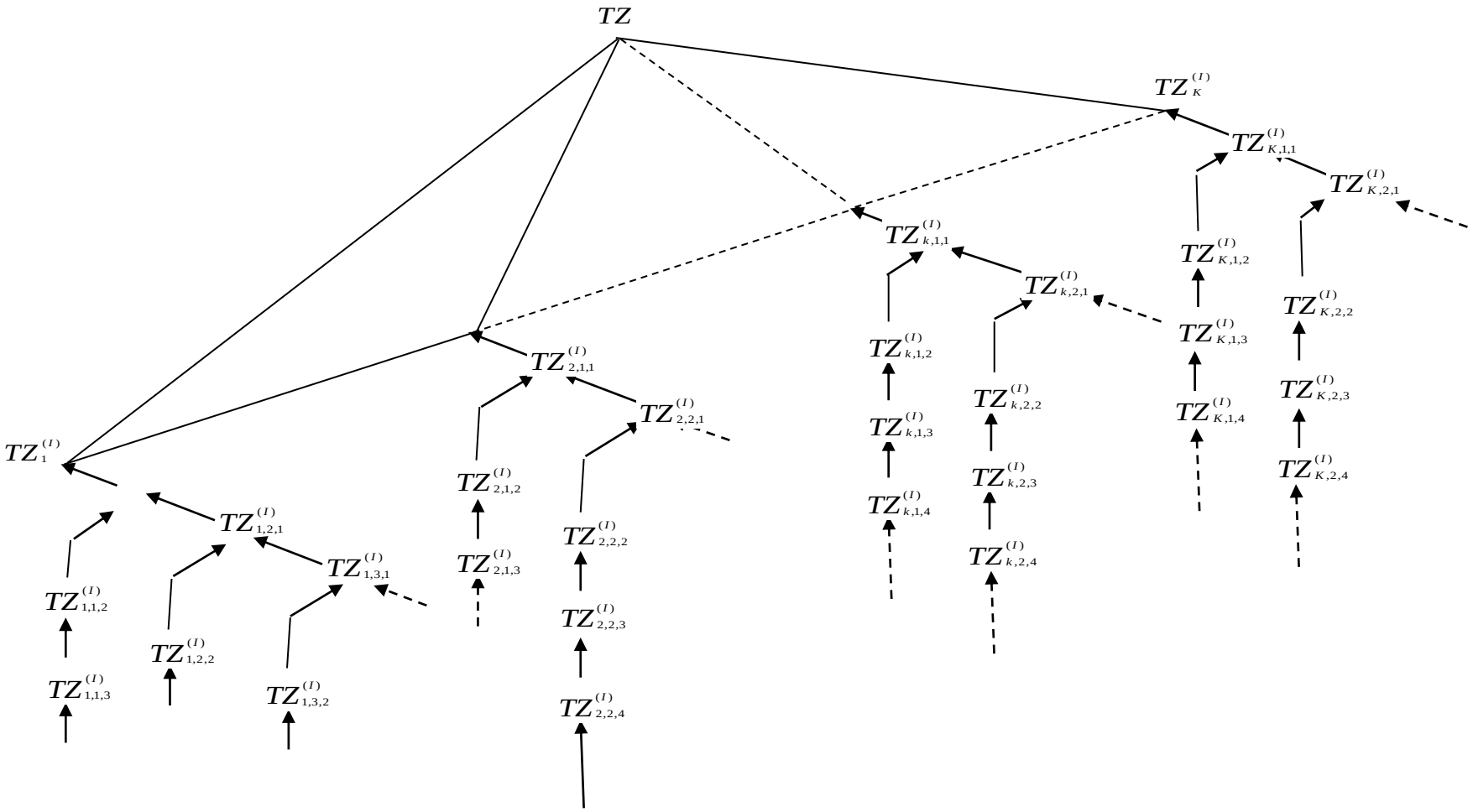

Рис. 2.2. Структура тезаурусу бази захисту текстової інформації TZ_{BZ} у прикладній системі, що відображається тезаурусом TZ_{PS}

можливостей скласти будь-яке уявлення щодо істинності смислу перехоплених текстових повідомлень, необхідно і достатньо замінити смисл SO, що входять до складу вихідних відкритих текстових повідомлень, на правдоподібні їхні відображення, що беруться із тезаурусу бази захисту інформації TZ_{BZ} обраної області активності з тезаурусом TZ_{PS}.

Для здійснення такої заміни в автоматичному режимі необхідно мати формальні позначення місць розташування відображень SO у структурі TZ_{BZ}, тобто необхідно локалізувати мовні одиниці у структурі тезауруса TZ_{BZ}.

Аналізуючи структуру TZ_{BZ} на рис. 2.2, можливо формалізувати цю структуру таким чином, щоб параметри локалізації були представлені у явному вигляді, наприклад у наступному вигляді:

$$SO^{(i)}_k \in \left\{ \begin{array}{l} SO^{(i-1)}_{k,1,1};\ SO^{(i-1)}_{k,1,2};\ \ldots;\ SO^{(i-1)}_{k,1,n};\ \ldots;\ SO^{(i-1)}_{k,1,N(l=1)} \\ SO^{(i-1)}_{k,2,1};\ SO^{(i-1)}_{k,2,2};\ \ldots;\ SO^{(i-1)}_{k,2,n};\ \ldots;\ SO^{(i-1)}_{k,2,N(l=2)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ SO^{(i-1)}_{k,l,1};\ SO^{(i-1)}_{k,l,2};\ \ldots;\ SO^{(i-1)}_{k,l,n};\ \ldots;\ SO^{(i-1)}_{k,l,N(l)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ SO^{(i-1)}_{k,L,1};\ SO^{(i-1)}_{k,L,2};\ \ldots;\ SO^{(i-1)}_{k,L,n};\ \ldots;\ SO^{(i-1)}_{k,L,N(L)} \end{array} \right. \quad (2.16)$$

Поточні значення нижніх індексів у позначеннях семантичних одиниць у виразі (2.16) однозначно визначають місця розташування цих семантичних одиниць у структурі TZ_{BZ}.

Визначені вище відношення можливо визначити між смисловими одиницями, що відображають логічно завершені думки певного рівня

абстрагування, зокрема на рівні ситуацій, на рівні фраз або на рівні слів (на кшталт: правдоподібна ситуація, правдоподібна фраза, правдоподібне слово). Тому інтерес являють упорядковані за темами (сценаріями) ситуації, упорядковані за ситуаціями фразеологічні словники та упорядковані за фразами тлумачні словники. Так що, у рамках структури на рис. 2.2 правдоподібний сценарій має складатися із правдоподібних ситуацій. У свою чергу, правдоподібна ситуація має складатися із правдоподібних фраз, а правдоподібна фраза – із правдоподібних слів. Тобто, якщо тезаурус мови обраної області активності позначити як TZ_{PS}, то у потоці SO, що відображає зашифрований зразок семантичної одиниці, усі смислові образи найвищого I-го рівня абстрагування мають бути вибрані під час шифрування із множини сценаріїв $\{TZ^{(I-1)}{}_k\}$, що визначає простір SO (I-1)-го рівня абстрагування у мові обраної області активності. У свою чергу, у рамках кожного сценарію $\{TZ^{(I-1)}{}_k\}$ усі смислові образи (I-2)-го рівня абстрактності мають бути вибрані із множини фраз$\{TZ^{(I-2)}{}_{k,l}\}$, а усі смислові образи (I-3)-го рівня абстрактності мають бути вибрані із множини слів $\{SO^{(1)}{}_{k,l,n,\dots,j}\}$.

Отже, будь-яка семантична одиниця $TZ^{(I)}{}_k$ із множини семантичних одиниць I-го рівня абстрагування, що входять до складу тезаурусу мови обраної області активності TZ_{PS}, під час криптоаналізу сприймається суб'єктом як можливий кандидат на смислову ідентичність із вихідною семантичною одиницею відкритого тексту $TZ^{(I)}{}_l$. Проте суб'єкт, якщо він має інформацію щодо прийнятої структури тезаурусів TZ_{PS} та TZ_{BZ}, якими він керується, розуміє, що $TZ^{(I)}{}_k \triangle TZ^{(I)}{}_l$ для будь-яких значень $k \neq l$, де $\triangle$ – позначка смислової відмінності. За цих умов суб'єкт усвідомлює, що вищезазначені семантичні одиниці, якщо вони відображені у TZ_{BZ}, знаходяться між собою у відношенні смислової правдоподібності. І знайти серед усіх правдоподібних смислових образів, що відображені у TZ_{BZ}, такий, що знаходиться із зашифрованим смисловим образом у відношенні смислової ідентичності, не уявляється можливим.

Щодо смислової правдоподібності між семантичними одиницями (I–1)-го рівня абстрагування за аналогією буде справедливим наступне твердження:

$$SO^{(I-1)}_{k,1} \wedge SO^{(I-1)}_{k,2} \wedge \ldots \wedge SO^{(I-1)}_{k,l} \wedge \ldots \wedge SO^{(I-1)}_{k,L}, \text{ якщо } \{TZ^{(I-1)}_{k,l}\} \underset{L}{\rightarrow} TZ^{(I)}_{k}, \quad (2.17)$$

за умови $SO^{(I-1)}_{k} \equiv TZ^{(I-1)}_{k}$, для будь-яких значень (I -1), k та l. (2.18)

Аналогічно за рекурентною схемою можливо записати вирази щодо відношень між смисловими одиницями для будь-яких інших більш деталізованих рівнів абстрагування представлення смислових образів. Із виразів (2.16) – (2.18) витікає, що семантичні одиниці, що входять до складу будь-якої однієї гілки у структурі тезаурусу TZ_{BZ}, знаходяться між собою у відношенні семантичної правдоподібності.

Висновки до другого розділу

2.1. Розроблено формальну структуру тезаурусу смислових образів для будь-якої мови взагалі або для будь-якого суб'єкту окремо. Показано, що такий тезаурус має ієрархічну прошаркову (рос., – слоистую) структуру і за ступенем абстрагування відображення смислових образів розподіляється на i рівнів, де $i = 1, 2, \ldots, I$ – кількість рівнів абстрагування відображення смислових образів, якими оперує колективний інтелект носіїв мови взагалі або індивідуальний інтелект суб'єкту розумової діяльності окремо, а I – максимальна кількість рівнів абстрагування відображення SO, що є доступною інтелекту. Так що, простір смислових образів, доступний суб'єкту (або групі суб'єктів), є дискретним, кінцево-мірним, який щодо рівнів абстрагування представлення образів має прошаркову коренево-подібну структуру.

Показано також, що структуру тезаурусу у загальному випадку можливо представити у вигляді рекурентної схеми (див. вираз (2.1)).

2.2. Показано, що у багатьох прикладних сферах доцільно призначити наступну ієрархію смислових одиниць, що відображають логічно завершені думки певного рівня абстрагування, де убування ступеню абстрагування прийнято зліва направо: сценарій / ситуація / фраза / слово. У цьому випадку для досконало стійкого захисту текстових повідомлень будемо мати ієрархічну структуру у вигляді розгалуженого кореня тезаурусів із чотирма рівннями абстрагування представлення смислових образів.

2.3. Показано, що для побудови будь-якої досконало стійкої лексикографічної криптосистеми системи захисту текстової інформації необхідно мати формальні визначення таких смислових відношень між семантичними одиницями як **смислова ідентичність, смислова відмінність,** сумнівна смислова ідентичність (або **правдоподібність**), сумнівна смислова відмінність (або **фальшиво-подібність**). Тобто, щодо смислових відношень між смисловими образами необхідно мати показники смислової ідентичності, смислової відмінності, правдоподібності та фальшиво-подібності. Надано формальні визначення цих показників (див вирази (2.8) – (2.15).

2.4. Запропоновано наступний методологічний підхід до захисту смислу текстових повідомлень, що заснований на використанні відношень смислової правдоподібності. Для того, щоб криптоаналітик ні за яких умов не мав можливостей скласти будь-яке уявлення щодо істинності смислу перехоплених текстових повідомлень, необхідно і достатньо замінити смисл *SO*, що містяться у вихідних текстових повідомленнях, на правдоподібні їхні відображення, що беруться із тезаурусу бази захисту текстової інформації TZ_{BZ} у прикладній системі, що відображається тезаурусом TZ_{PS} мови обраної області активності. Ці тезауруси повинні мати структуру виду рис. 2.2 та рис. 2.1 відповідно і у повній мірі наповнені відповідними смисловими образами на усіх рівнях абстрагування, що необхідні для повного відображення заданої прикладної області.

2.5. Для здійснення заміни істинних смислових образів SO на правдоподібні необхідно мати формальні позначення місць розташування відображень SO у структурі бази захисту текстової інформації TZ_{BZ}, тобто необхідно локалізувати семантичні одиниці у структурі TZ_{BZ}. Отримано вираз (2.16), що представляє структуру тезауруса TZ_{BZ} з параметрами локалізації, що надані у явному вигляді.

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

РОЗДІЛ 3
РОЗРОБКА МЕТОДУ ПОБУДОВИ ЛЕКСИКОГРАФІЧНОЇ КРИПТОСИСТЕМИ

Численні дослідження властивостей досконало стійких криптосистем вказують на відносно невеликі значення відстані єдиності при шифруванні повідомлень, складених із символів алфавіту будь-якої із природних мов [26,28,29]. Це призводить до необхідності частої зміни ключової інформації, що є проблемою для багатьох застосувань. Окрім того, необхідно забезпечувати випадковість й однакову ймовірність вибору варіантів реалізації ключа. Так що створення методу побудови криптосистем, що забезпечує можливість роботи у режимі досконалої секретності за суттєво більших значень відстані єдиності у порівнянні, зокрема, із методом одноразових блокнотів (схеми Вернама), являє актуальне завдання.

3.1. Постановка завдання

Розглянемо досконало стійку криптосистему захисту текстової інформації, що поміщена у задану табличну форму за умови, що ця текстова інформація береться із семантичного словника наперед визначеної прикладної області. Тобто, перш ніж здійснювати шифрування, використовуючи будь-який відомий досконало стійкий шифр, необхідно створити семантичний словник, лінгвістичні одиниці котрого відображають змістовний простір цієї прикладної області.

Розробимо механізм збільшення відстані єдиності у рамках цієї досконало стійкої криптографічної системи шляхом синтезу штучної мови відображення прикладної області, де ця система використовується.

Текстова інформація, що підлягає захисту, розглядається як потік текстових повідомлень (зокрема, слів або словосполучень) $M_{j,v}$, де $j=1,2,\dots, L$ – індекс, що означає порядковий номер повідомлення у потоці, $v= 1, 2,\dots, N$

– індекс, що означає смисловий варіант повідомлення із наперед визначеного словника повідомлень ***S*.**

За цих умов використаємо наступні позначення.

L – довжина (обсяг) тексту, що підлягає захисту;

S – словник, що відображає певну множину ***M*** можливих у тексті повідомлень $M_{j,i}$;

M – повідомлення як елемент словника ***S***;

N – кількість смислових варіантів повідомлень (тобто, розмір словника ***S***);

H(M) – ентропія повідомлення, узятого із словника ***S***;

n – довжина повідомлення;

r – ентропія мови, за допомогою якої відображаються повідомлення із ***M*** (тобто, це середня кількість інформації, що міститься в одній літері цієї мови і залежить від *n*);

R – максимальна ентропія мови (тобто, максимальне число бітів, яке може бути передано кожним символом алфавіту цієї мови за умови рівноймовірності виникнення усіх послідовностей символів);

D – надлишковість (інформаційна) мови ($D = R - r$);

H(K) – ентропія системи захисту як розмір простору ключів шифру, що залежить від кількості можливих для використання ключів шифру;

K – кількість можливих для використання ключів шифру у системі захисту (*H(K)*= log(*K*));

U – відстань єдиності (що також називають точкою єдиності) – такий приблизний розмір шифрованого тексту, для якого сума ентропії відкритого тексту та ентропії ключа шифру дорівнює числу бітів, що міститься у цьому шифрованому тексті (*U= H(K)/D*);

E – множина можливих шифрограм, що утворюються шляхом застосування оператора перетворення $\boldsymbol{T_i}$ до відкритих повідомлень *M* із множини ***M***, а індекс *i* відповідає конкретному ключу, що був при цьому застосований;

$\boldsymbol{T_i}$ – оператор перетворення $\boldsymbol{M}$ у $\boldsymbol{E}$.

Припустимо, що критично важлива текстова інформація поміщена у табличну форму розміром *s*x*L2*, що задана у вигляді таблиці 3.1, де *s* – кількість стовпців таблиці, а *L2* – кількість рядків цієї ж таблиці.

Вміст кожної клітини таблиці будемо розглядати як окреме повідомлення (слово або словосполучення) *M* із множини $\boldsymbol{M}$. Отже, дані таблиці – це текстова інформація, яка підлягає захисту, і при передаванні через канал зв'язку розглядається як потік текстових повідомлень $M_{j,v}$, де *j= 1, 2, ... , L* – індекс, що означає порядковий номер повідомлення у потоці, ν = *1, 2, ... , N* – індекс, що означає смисловий варіант повідомлення із наперед визначеного словника повідомлень $\boldsymbol{S}$.

Таблиця 3.1

Задана таблична форма

	Найменування стовпця №*1*	Найменування стовпця №*2*	---	---	Найменування стовпця № *s*
Найменування рядка №*1*					
Найменування рядка №*2*					

Найменування рядка №*L2*					

Дані таблиці необхідно передати через відкритий канал зв'язку у зашифрованому вигляді таким чином, щоб забезпечити досконало стійкий захист від порушень конфіденційності шляхом синтезу криптосистеми, що реалізує розроблюваний метод збільшення відстані єдиності.

Ураховуючи критичну важливість текстової інформації, що поміщена у таблицю, досконалість теоретико-інформаційної стійкості синтезованої криптосистеми захисту має бути теоретично доведеною. Що означає необхідність формального визначення обмежень, за яких забезпечується стовідсоткова гарантія неможливості однозначного відновлення відкритої інформації, що поміщена у табл. 3.1, навіть за умов, коли у розпорядженні криптоаналітика знаходяться зразки зашифрованих даних скільки завгодно великої сумарної довжини, а криптоаналітик має необмежений час та необмежені обчислювальні ресурси для дешифрування перехоплених криптограм.

3.2. Синтез методу

3.2.1. Укрупнення алфавіту як механізм збільшення відстані єдиності

Уведемо поняття *семантичний словник таблиці* як множину $\boldsymbol{M}$ усіх можливих семантичних варіантів повідомлень $M_{j,i}$, що можуть знайти своє відображення у таблиці заданої форми [10,11]. Розмір словника $\boldsymbol{S}$ це є розмір множини $\boldsymbol{M}$, що дорівнює N.

Будемо вважати, що на практиці у більшості випадків семантичні зв’язки між інформацією різних стовпців таблиці не спостерігаються. Тому доцільно припустити, що семантичний словник даної табличної форми складається із s семантичних підсловників відповідно до кількості стовпців у таблиці. Текстові повідомлення, що вносяться у перший стовпець, являють собою семантичні одиниці першого підсловника, у другий стовпець вносяться семантичні одиниці другого підсловника і т.д.

Синтезуємо мову відображення текстової інформації, що поміщається у задану табличну форму. Для цього визначимо в якості літери алфавіту цієї форми один рядок табл. 3.1. Іншими словами, уявімо, що кожен рядок табл. 3.1 є літерою певним чином визначеного алфавіту цієї табличної форми. У

цьому разі кількість літер у такому алфавіті B визначиться за наступною формулою:

$$B = \prod_{i=1}^{s} S_i \,, \tag{3.1}$$

де s – кількість підсловників у словнику табличної форми, S_i – кількість семантичних елементів (тобто, елементарних повідомлень M) в і-ому підсловнику.

Припустимо, що повідомлення, які поміщаються у клітини заданої табличної форми, є однакової довжини n.

Визначимо кількість можливих смислових варіантів повідомлень N (тобто, розмір словника $\boldsymbol{S}$) довжиною n, що можуть бути внесені у клітини цієї таблиці. Розглянемо випадок, коли усі клітини таблиці є заповненими елементарними повідомленнями, узятими із підсловників. У цьому випадку будемо мати максимально можливу довжину послідовності повідомлень, що відображають інформацію у табл. 3.1 і мають бути передані протягом одного сеансу зв'язку. Розмір цієї послідовності дорівнює кількості літер у визначеному алфавіті табличної форми. Якщо припустити, що літери із алфавіту даної табличної форми під час формування даних таблиці будуть не (не будуть) повторюватися, тобто кожен рядок у таблиці буде зустрічатися тільки один раз, то кількість можливих варіантів повідомлень N довжиною n розраховується за наступною формулою:

$$N = (S_1 S_2 \ldots S_s)\,(S_1 S_2 \ldots S_s - 1)\,\ldots\,(S_1 S_2 \ldots S_s - (s-1)), \tag{3.2}$$

де S_i – розмір і-го підсловника, i = 1, 2, 3, …, s.

Знаючи N, використаємо ланцюг широко відомих математичних виразів щодо:

- ентропії повідомлення $H(M) = \log_2 N$, що вимірюється у бітах; (3.3)
- ентропії мови $r = H(M)/n$ заданої табличної форми; (3.4)
- абсолютної ентропії $R = log_2 B$ мови заданої табличної форми; (3.5)
- надлишковості мови $D = R - r$ заданої табличної форми; (3.6)

- ентропії ключової системи $H(K)=\log_2 K$, (3.7)

де K – кількість ключів у СЗІ.

І, як результат, визначимо відстань єдиності для розроблюваної ключової системи за формулою [23]:

$$U= H(K)/D. \quad (3.8)$$

Аналізуючи результати розрахунків за наведеними вище виразами, легко побачити, що надлишковість штучної мови відображення інформації D, що поміщена у задану табличну форму, є надзвичайно малою, що, у свою чергу, згідно (3.8) визначає суттєво великі значення відстані єдиності U.

Приклад

Припустимо, що маємо таблицю із чотирма рядками та чотирма стовпцями. Припустимо також, що відповідно для обраної табличної форми маємо $S_1 = 8$, $S_2 = 100$, $S_3 = 8$, $S_4 = 8$.

За формулою (3.1) кількість літер у алфавіті даної табличної форми визначається як $B = 8 \cdot 100 \cdot 8 \cdot 8 = 51200$. Так що розмір словника табличної форми визначається як $B = 51200$ [літер].

Розрахуємо кількість можливих повідомлень N довжиною 4 згідно з виразом (3.2). Для нашого прикладу:

$$N = (S_1 \cdot S_2 \cdot S_3 \cdot S_4) \cdot (S_1 \cdot S_2 \cdot S_3 \cdot S_4 - 1) \cdot (S_1 \cdot S_2 \cdot S_3 \cdot S_4 - 2) \cdot (S_1 \cdot S_2 \cdot S_3 \cdot S_4 - 3).$$

Отже, у цьому випадку $N= 6871142396067532800$.

За формулою (3.3) визначимо ентропію повідомлення $H(M)$:

$$\log_2 6871142396067532800 = 62.5752556894213,$$

тобто, $H(M)= 62.5752556894213$ [біт].

Визначимо ентропію мови r заданої табличної форми за формулою (3.4):

$$\frac{62.5752556894213}{4} = 15.6438139223553,$$

тобто, $r = 15,6438139223553$ [біт/літера].

Визначимо абсолютну ентропію *R* мови заданої табличної форми за формулою (3.5):

$$\log_2 51200 = 15.6438561897747,$$

тобто, *R = 15,6438561897747* [біт/літера].

Надлишковість мови *D* заданої табличної форми визначимо за формулою (3.6):

$$15.6438561897747 - 15.6438139223553 = 0.000042267419412,$$

тобто *D = 0,000042267419412* [біт/літера].

Визначимо ентропію розроблюваної ключової системи *H(K)* за формулою (3.7). Слід зазначити, що при розрахунку ентропії кількість ключів обираємо, виходячи із міркувань необхідності побудови системи із досконалою стійкістю. Отже, нехай кількість ключів *K* дорівнює *N* – кількості можливих повідомлень довжиною *n*, тоді

$$\log_2 687114239\mathbf{6}067532800 = 62.575255\mathbf{6}894213,$$

тобто, *H(K) = 62.5752556894213* [біт].

І, на кінець, визначимо відстань єдиності для розроблюваної ключової системи за формулою (3.8):

$$\frac{62.5752556894213}{0.000042267419412} = 1480460.75583801,$$

тобто, *U = 1480460,75583801* [літер].

Виберемо для нашого прикладу довжину ключового слова за формулою

$$k = \log_x H(M), \qquad (3.9)$$

де *x* – кількість літер в алфавіті, що використовується для складання ключового слова, *H(M)* – ентропія повідомлення.

Якщо для ключового слова використовувати лише літери української мови (33 літери), тоді довжина ключового слова *k* буде дорівнювати:

$\log_{33} 687114239\mathbf{6}067532800 = 12.405$, тобто $k \approx 13$ [літер].

Якщо ж, для побудови ключового слова використовувати ще й цифри (0–9), тоді довжина ключового слова повинна буди не меншою за 12 літер.

$\log_{43} 68711423960675328000 = 11.53$, тобто $k \approx 12$ [літер].

Порівнюючи отримані значення відстані єдиності та довжини ключового слова, бачимо, що на відміну від відомих методів забезпечення режиму досконалої стійкості, обсяги текстової інформації, що потребують досконало стійкого захисту, у даному випадку можуть суттєво перевищувати довжину ключового слова.

3.2.2. Залежність відстані єдиності та ентропії ключа шифру від довжини повідомлення

Підставляючи вирази (3.4) і (3.5) у вираз (3.8), отримаємо наступне:

$$U = \frac{H(K)}{\log_2(B) - \frac{H(M)}{n}}. \tag{3.10}$$

Виходячи з властивостей досконало стійкої криптосистеми системи [23], кількість ключів шифру K повинна дорівнювати N – кількості смислових варіантів повідомлень довжиною n. Таким чином, за умови

$$H(K) = H(M) = \log_2 N \tag{3.11}$$

підставляючи (3.11) у (3.10), вираз для визначення відстані єдиності можна записати наступним чином:

$$U = \frac{\log_2 N}{\log_2(B) - \frac{\log_2 N}{n}}. \tag{3.12}$$

Тепер знайдемо залежність N – кількості можливих варіантів повідомлень від n – довжини повідомлення. При визначені N слід мати на увазі те, що кожен рядок у таблиці (тобто, кожна літера у повідомленні, якщо використати розглянутий вище спосіб укрупнення алфавіту) зустрічається тільки один раз (тобто, літери не повторюються). У цьому випадку максимально можлива довжина повідомлення (тобто, максимальна кількість символів у текстовій інформації, що відображають зміст таблиці і передаються через канал зв'язку) дорівнює кількості літер в алфавіті мови відображення інформації, що синтезована для даної табличної форми. Отже,

з урахуванням (3.1) та (3.2) відобразимо вираз для визначення N – кількості можливих значень повідомлення при різних n:

$$N = \prod_{s=1}^{n}[B-(s-1)]. \tag{3.13}$$

Таким чином, при B = *const* і при $H(K)=H(M)$ (умова забезпечення режиму досконало стійкої криптосистеми) можна записати наступне (підставивши вираз (3.13) у (3.12)):

$$U(n) = \frac{\log_2 \prod_{s=1}^{n}[B-(s-1)]}{\log_2(B) - \frac{\log_2 \prod_{s=1}^{n}[B-(s-1)]}{n}}. \tag{3.14}$$

Вираз (3.14) визначає залежність значень відстані єдиностіU від довжини повідомлення n, що представлена на рис. 3.1.

Вираз, що визначає залежність ентропії ключа шифру $H(K)$ від довжини повідомлення n, з урахуванням (3.13) можна записати наступним чином:

$$H(K) = \log_2 \prod_{s=1}^{n}[B-(s-1)]. \tag{3.15}$$

Графіки залежності $U = f(n)$, $H(K) = f(n)$ показано на рис. 3.1 та рис. 3.2 відповідно.

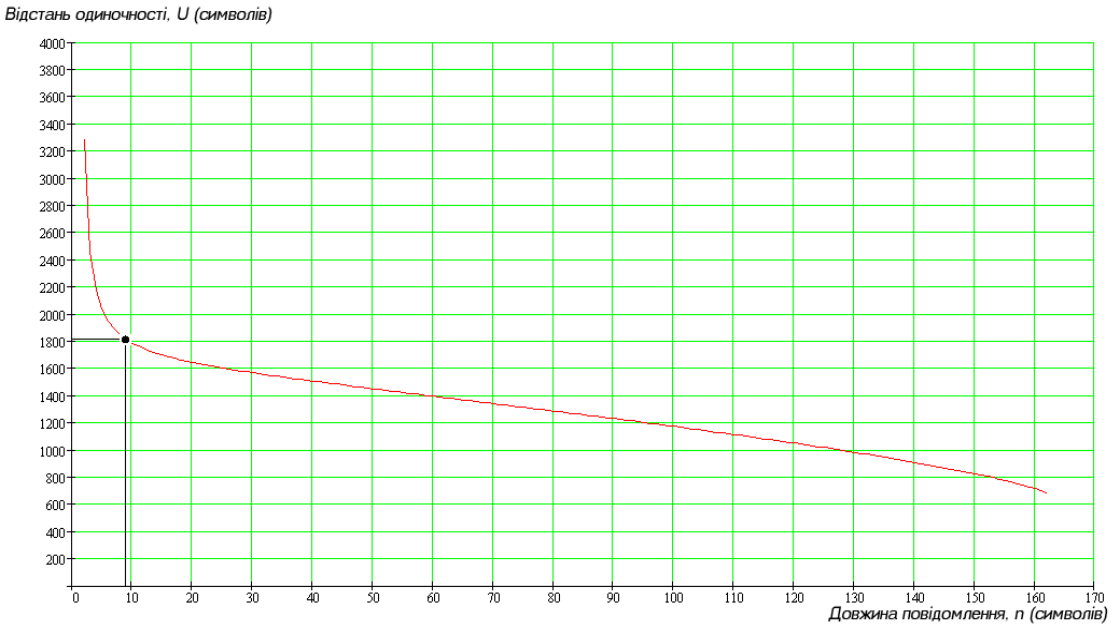

Рис. 3.1. Графік залежності відстані єдиностіU від довжини повідомлення n

Приклад

Нехай маємо семантичний словник заданої табличної форми, що складається із чотирьох підсловників, тобто $s=4$. Ці підсловники мають певну кількість лінгвістичних одиниць, а саме $S_1=3$, $S_2=6$, $S_3=3$, $S_4=3$.

Визначимо кількість літер в алфавіті мови відображення інформації у даній табличній формі відповідно до формули (3.1). Отримаємо у даному прикладі $B=162$. У той же час величина B визначає максимальну довжину обсягу текстової інформації, що підлягає захисту повідомлення протягом одного сеансу зв'язку, оскільки кожна літера у повідомленні зустрічається тільки один раз.

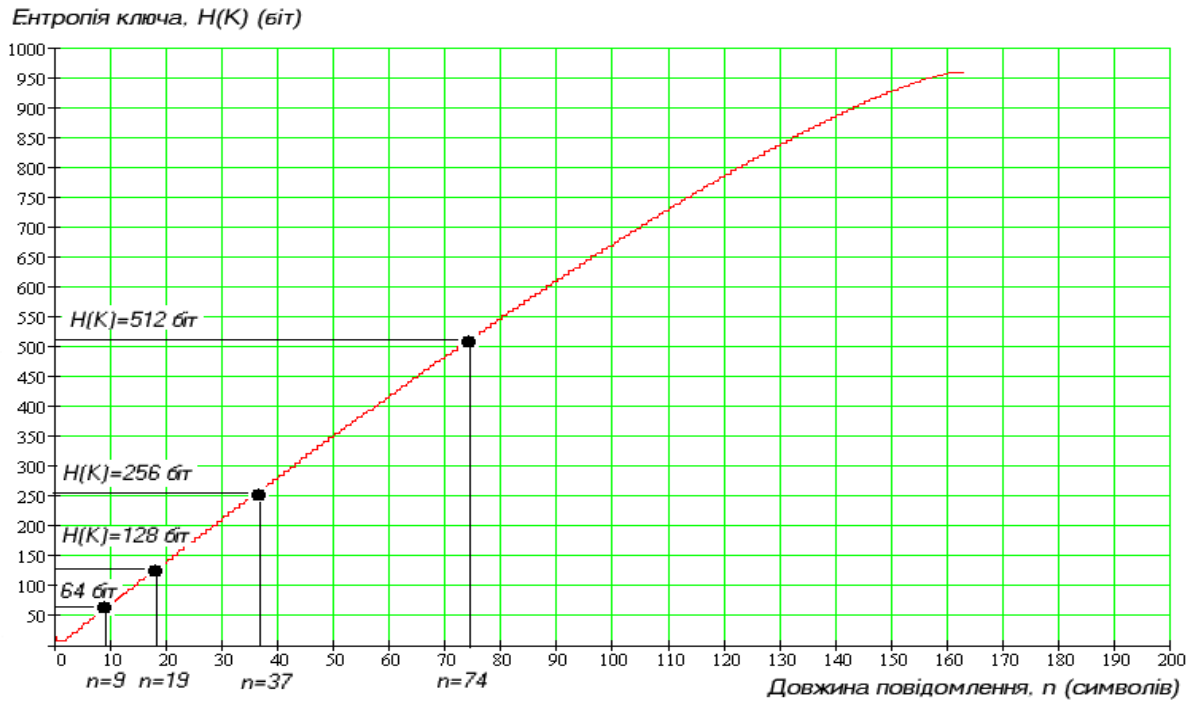

Рис. 3.2. Графік залежності ентропії ключа шифру $H(K)$ від довжини повідомлення n

З графіку на рис. 3.1 видно, що навіть при максимальній за цих умов довжині повідомлення $n=162$, відстань єдиності дорівнює $U \approx 680$ символів.

Отже, шляхом укрупнення алфавіту мови відображення інформації у таблиці заданої форми можемо після відповідних перетворень мати

шифротексти, які значно коротші за відстань єдиності. Так що, згідно теорії К.Шеннона [17], такі шифротексти можна дешифрувати декількома способами, причому кожен з них може бути коректним, і таким чином забезпечувати досконалий захист текстової інформації, поставивши аналітика перед вибором правильного повідомлення із множини кількох можливих правильних повідомлень.

Графіки, що зображені на рис. 3.1 та рис. 3.2, доцільно використовувати для обчислення максимально можливої кількості сеансів зв'язку без зміни ключа шифру при дотриманні умов, коли не порушується ознака досконало стійкої криптосистеми. Зокрема, користуватися цими графіками пропонується наступним чином.

Користувач криптосистемою задається сталим розміром ключа шифру, наприклад 64 біти. З графіку на рис. 3.2 видно, що така довжина ключа шифру відповідає довжині повідомлення у 9 символів (тобто, має бути 9 рядків у таблиці заданої форми). З графіку, що зображений на рис. 3.1, видно, що відстань єдиності, що відповідає цій довжині повідомлення дорівнює $U=1812$ символів. Таким чином, для того, щоб криптосистема зберігала властивості досконало стійкої, при використанні ключа шифру довжиною 64 біти, користувач (передавальна сторона) повинен передавати повідомлення довжиною 9 символів (рядків таблиці) не більше ніж $1812/9 \approx 200$ разів. Іншими словами, користувач повинен змінювати ключ шифру після передавання кожних 200 повідомлень.

3.2.3. Залежність показника ефективності захисту від довжини повідомлення та кількості символів алфавіту

Ефективність методу захисту текстової інформації з укрупненням алфавіту мови відображення цієї інформації доцільно порівнювати із ефективністю методу одноразових блокнотів, оскільки обидва вищезазначені методи захисту можуть бути реалізованими у режимі досконало стійких криптосистем.

У разі застосування методу одноразових блокнотів для захисту інформації у режимі досконалої секретності довжина ключа шифру має дорівнювати довжині повідомлення [23]. Для визначення ентропії такого ключа використовують наступний вираз:

$$H_1(K) = \log_2(B_1^{n_1}), \tag{3.16}$$

де n_1 – довжина повідомлення [*літер*], що записане мовою людського спілкування (українська, російська, англійська і т.д.), що має алфавіт B_1 [*літер*].

У той час, як ентропія ключа, що застосовується для захисту інформації при використанні укрупненого алфавіту мови відображення цієї інформації обчислюється за виразом (3.15).

Отже, в загальному випадку для визначення показника виграшу у довжині ключа (за інших рівних умов) у разі застосування у криптосистемі укрупненого алфавіту у порівнянні із методом одноразових блокнотів слід використати наступний вираз:

$$Z = \frac{H_1(K)}{H(K)}\ . \tag{3.17}$$

Підставивши вираз (3.16) у чисельник, а вираз (3.15) у знаменник виразу (3.17) отримаємо наступне:

$$Z = \frac{\log_2(B_1^{n_1})}{\log_2 \prod_{s=1}^{n}\left[B-(s-1)\right]}. \tag{3.18}$$

При цьому слід пам’ятати, що B – алфавіт табличної форми, тобто кількість можливих комбінацій табличних рядків семантичного словника, що штучно створюється в результаті статистичного та семантичного аналізу предметної області, а n – кількість рядків табличної форми.

Приклад

Нехай вихідна таблиця з відкритою інформацією, що підлягає шифруванню, має форму у вигляді три рядки на чотири стовпця. Так що семантичний словник складається з чотирьох підсловників, тобто $s=4$. Ці

підсловники мають наступну кількість лінгвістичних одиниць: $S_1=3$, $S_2=6$, $S_3=3$, $S_4=3$. Отже маємо: $n=3$, $B=162$, $n_1=180$, $B_1=50$ (літери української мови, цифри *0 - 9*, такі знаки як *“-”, “,”, “.”, “*пробі*л”, “»”, “«”*).

За таких вихідних даних відповідно до виразу (3.18) значення показника виграшу у довжині ключа $Z = 46{,}2$. Тобто, при застосуванні досконало стійкої криптосистеми з укрупненим алфавітом довжина ключа шифру може бути у *46,2* рази менша за довжину ключа шифру при застосуванні методу одноразових блокнотів.

Побудуємо графіки залежності виграшу у довжині ключа від довжини повідомлення $Z = f(n_1)$ (див. рис. 3.3). Порівнюється ефективність застосування криптосистеми з укрупненим алфавітом (щодо довжини ключа шифру) у порівнянні із ефективністю застосування методу одноразових блокнотів.

Із графіків на рис. 3.3 видно, що виграш у ефективності застосування криптосистеми з укрупненим алфавітом у порівнянні з криптосистемою, що реалізує метод одноразових блокнотів, лінійно залежить від довжини повідомлення.

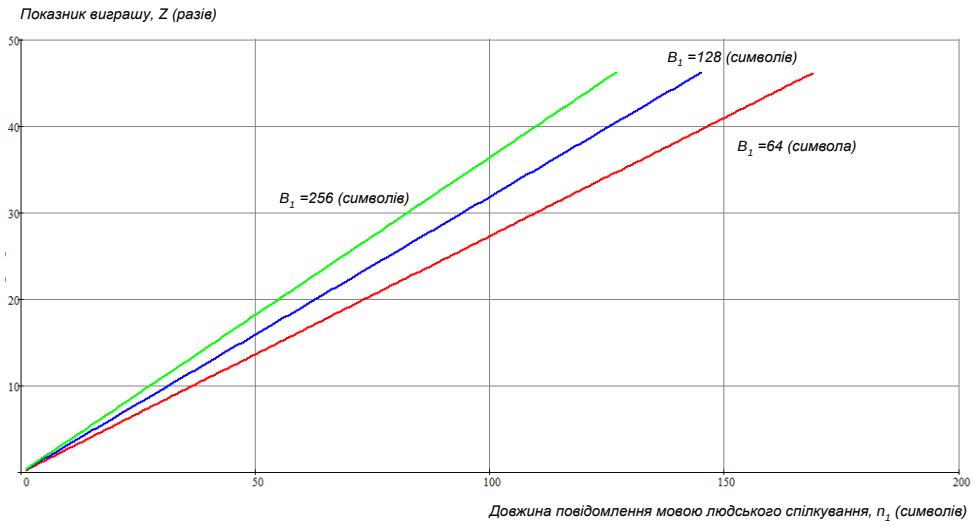

Рис. 3.3. Графік залежності виграшу в ефективності застосування криптосистеми з укрупненим алфавітом (щодо довжини ключа шифру) від довжини повідомлення $Z = f(n_1)$ у порівнянні із ефективністю застосування методу одноразових блокнотів

Розглянемо графіки залежності виграшу в ефективності застосування криптосистеми з укрупненим алфавітом (у порівнянні із ефективністю методу одноразових блокнотів) від кількості символів алфавіту звичайної мови людського спілкування, тобто визначимо залежність $Z = f(B_1)$ (див. рис. 3.4).

Із графіків на рис. 3.4 видно, що виграш в ефективності застосування криптосистеми з укрупненим алфавітом у порівнянні з методом одноразових блокнотів практично не залежить від обсягу алфавіту використаної мови людського спілкування. Цей виграш суттєво залежить від довжини повідомлень, що представлені на цій мові (див. рис. 3.3).

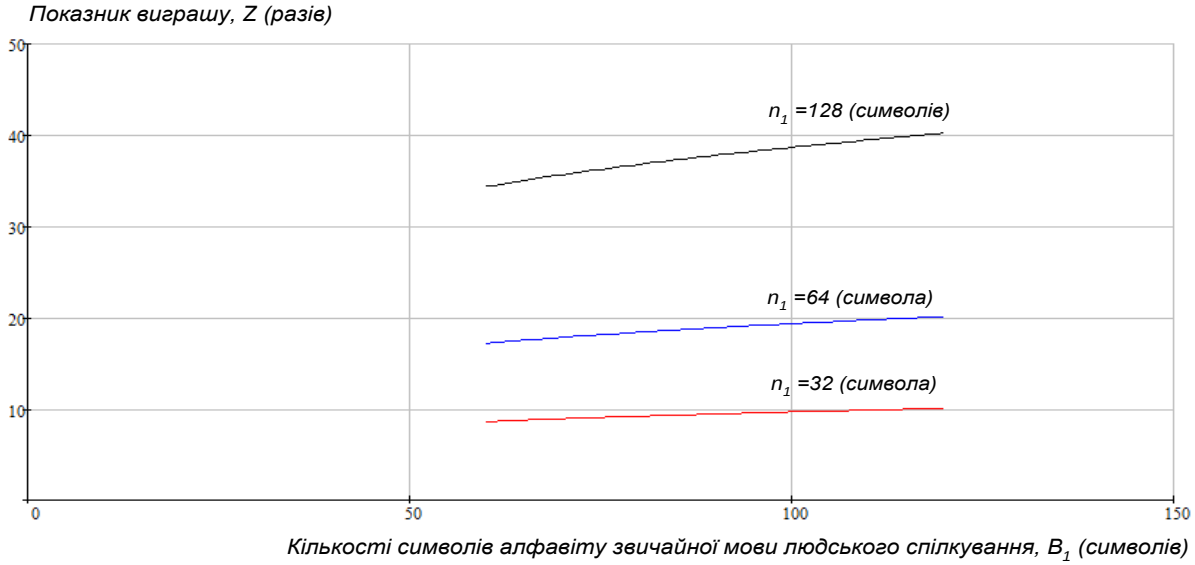

Рис. 3.4. Графік залежності виграшу Z від кількості символів алфавіту звичайної мови людського спілкування $Z = f(B_1)$

3.2.4. Варіант побудови криптосистеми із збільшеною відстанню єдиності

Розглянемо особливості побудови досконало стійкої криптосистеми із збільшеною відстанню єдиності для захисту текстових повідомлень, що поміщені у задану табличну форму (надалі – КЗВЄ).

Вихідні умови

1. При шифрування текстові повідомлення беруться виключно із семантичного словника наперед визначеної прикладної області.

2. Дозволено застосування різноманітних варіантів, що засновані на використанні генераторів псевдовипадкових послідовностей (ГПВП). Слід використовувати ГПВП, що здатні забезпечити стійкість проти атак, що базуються на імовірнісному аналізі шифрограм.

Примітка 3.1. *Для шифрування також можуть застосовуватися будь-які різновиди поліалфавітних схем підстанов (типу Віженера, Бофора тощо) [23], що не маскують частоту появ символів алфавіту на виході шифратора і, отже, є чутливими до злому при застосуванні імовірнісного аналізу шифрограм.*

3. Сторони секретного обміну текстовими повідомленнями через незахищений канал зв’язку повинні мати ідентичне обладнання КЗВЄ.

Порядок дій

1. Створюється семантичний словник, лінгвістичні одиниці котрого у повній мірі відображають мовний простір прикладної області. Цей словник містить лінгвістичні одиниці, котрі потенційно можуть бути включені до складу вихідних відкритих елементарних текстових повідомлень, що поміщаються адміністратором у задану табличну форму під час підготовки таблиці, що має бути передана через канал зв’язку. Оскільки згідно даного методу при використанні КЗВЄ у режимі досконалої стійкості передбачається дотримання умови щодо відстані єдиності, то формальний синтез тезаурусу може не здійснюватися, що значно спрощує процес його створення.

2. Сторони секретного обміну текстовими повідомленнями перед тим, як скористатися даним методом захисту, узгоджують ключ шифру.

3. При наявності словника, відомого ключа шифру та заданої табличної форми адміністратор прикладної системи на передавальній стороні формує вихідний відкритий текст у вигляді таблиці заданої форми.

4. Маючи таблицю із сформованим відкритим текстом, адміністратор запускає на виконання програмний засіб реалізації даного методу (шифрувальний пристрій або програмний шифратор) і, як результат, отримує зашифрований зразок тексту, котрий може зберігати на своєму комп’ютері або передати через відкритий канал зв’язку.

Функціональна схема КЗВЄ

Принцип роботи КЗВЄ базується на синхронізації ГПВП, що розташовані на передавальній та приймальній сторонах каналу секретного обміну інформацією, за допомогою відомого ключа шифру (див. рис. 3.5).

Схема на рис. 3.5 містить у своєму складі усі основні елементи симетричної криптографічної системи захисту текстової інформації, яка за певних умов здатна функціонувати у режимі досконалої стійкості [23]. Однак, окрім цього, використано додатковий елемент – словник предметної області, у рамках якої планується застосовувати КЗВЄ. Шифратор реалізує механізм укрупнення алфавіту мови відображення інформації, що поміщається у задану табличну форму, так, як це показано у попередніх підрозділах. Зокрема, кожен окремий рядок таблиці представляється як окрема літера укрупненого алфавіту мови відображення табличної інформації. За допомогою ключової інформації ГПВП ставиться у певний початковий стан, з якого і починає подавати згенеровані псевдовипадкові числа на вхід шифрувального пристрою. Генерується одне псевдовипадкове число на кожен рядок таблиці.

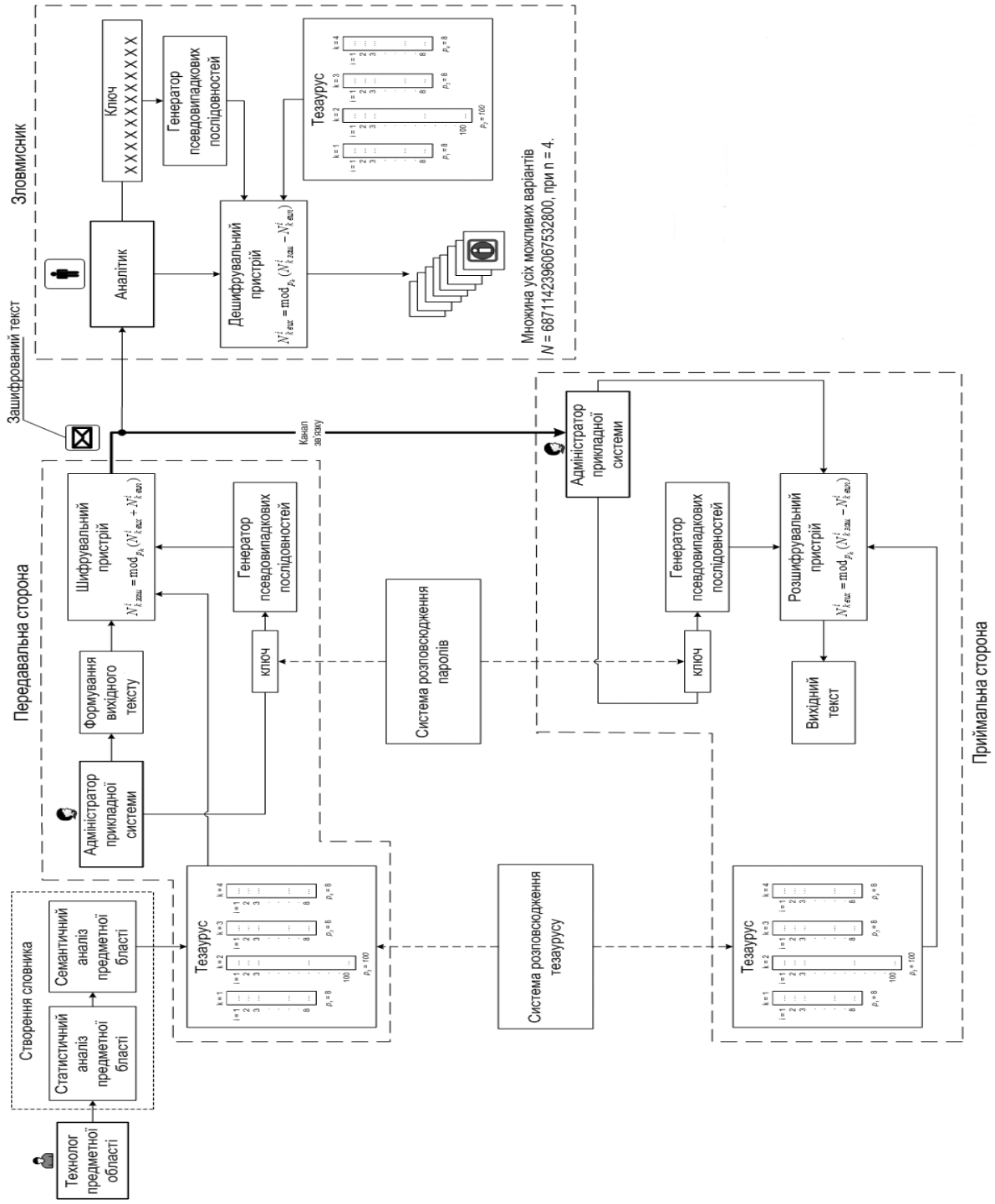

Рис. 3.5. Варіант блок-схеми КЗВЄ

На третій вхід шифратора подаються лінгвістичні одиниці від словника предметної області, а також дані щодо порядкових номерів розташування цих одиниць у структурі словника. Шифратор здійснює заміну істинного укрупненого елементу таблиці, у якості якого розглядається поточний рядок таблиці, на маскуючий рядок лінгвістичних елементів, узятих із словника.

Для кожного поточного рядка сформованої адміністратором вихідної таблиці знаходиться його відображення у словнику та визначається порядковий номер розташування цього відображення у структурі словника. Потім визначається порядковий номер розташування маскуючого рядка таблиці у структурі словника, який використовується як замінник істинного рядка таблиці. Цей порядковий номер визначається шляхом складання за модулем 2 порядкового номеру істинного елемента та псевдовипадкового числа, одержаного від ГПВП. Результатом роботи шифрувального пристрою є зашифрований текст таблиці, котрий передається на приймальну сторону через відкритий канал зв'язку. Адміністратор прикладної системи на приймальній стороні подає отриманий зашифрований текст на розшифрувальний пристрій, а за допомогою відомого ключа задає той самий початковий стан ГПВП, що і початковий стан генератора передавальної сторони. У розшифраторі здійснюються ті ж самі перетворення, що і у шифраторі. Результатом роботи розшифрувального пристрою є вихідна таблиця, що була сформована на передавальній стороні.

Якщо характеристики синтезованої таким чином КЗВЄ визначено так, як це раніш показано у даній роботі, то буде забезпечено функціонування цієї криптосистеми у режимі досконалої стійкості.

Як бачимо, для збільшення відстані єдиності у склад даної КЗВЄ включено словник мови відображення інформації, лінгвістичні одиниці із котрого поміщаються у задану табличну форму [8]. Структура словника узгоджується із структурою табличної форми, у яку поміщаються текстові повідомлення, що потребують захисту. Якщо, наприклад, у стовпцях таблиці відображаються семантично зв'язані лінгвістичні одиниці, то логічно представити даний словник у вигляді сукупності із n підсловників, де n – кількість стовпців заданої табличної форми. Кожен підсловник використовується для обробки даних одного стовпця. Тому кількість підсловників тезаурусу, що використовуються для

шифрування/розшифрування, дорівнює кількості стовпців заданої табличної форми.

З метою пояснення механізму заміни істинного елементу відкритої текстової послідовності на маскуючий елемент закритої послідовності у процедурі шифрування розглянемо наступний приклад. Припустимо, що словник предметної області складається із n підсловників. Усі елементи кожного підсловника є пронумерованими. Припустимо також, що розташування першого істинного елементу увідної послідовності у структурі визначеного шифратором k-ого підсловника має порядковий номер $N_{вих}$, а перше випадкове число, надане генератором ГПВП, дорівнює $N_{вип}$, де $0 \le N_{вип} \le p_к$, $p_к$ – розмірність (тобто, загальна кількість слів) k-ого підсловника.

Тоді порядковий номер розташування у структурі k-ого підсловника „не істинного" (тобто, маскуючого) мовного елементу, що поміщається у якості першого елементу у вивідну шифровану послідовність замість істинного першого мовного елементу, $N_{заш}$ буде дорівнювати

$$N^i_{k\,заш} = \mathrm{mod}_{p_k}(N^i_{k\,вих} + \mathrm{mod}_{p_k}(N^i_{k\,вип}) + p_k). \qquad (3.19)$$

Вираз (3.19) є справедливим за умов співпадіння розмірності застосованого підсловника з розмахом випадкових послідовностей, що генеруються ГПВП. (Розмах – це різниця між максимальним та мінімальним значеннями псевдовипадкових чисел у послідовності). У загальному випадку при розмаху ГПВП, що дорівнює розмірності найбільшого підсловника, рівняння шифрування, яке являє собою формулу для визначення порядкового номеру розташування маскуючого слова у $к$-ому підсловнику тезаурусу, має наступний вигляд:

$$N_{заш} = \mathrm{mod}_{p_k}(N_{вих} + N_{вип}), \qquad (3.20)$$

де $N_{вих}$ – порядковий номер розташування у $к$-ому підсловнику i-го істинного мовного елемента, узятого із вихідної, що підлягає шифруванню, відкритої послідовності лінгвістичних одиниць, де i – порядковий номер розташування цього елементу у відкритій послідовності;

$N_{зав.}$ – порядковий номер розташування у *к*-ому підсловнику маскуючого елементу, що поміщається в *і*-ту позицію (замість істинного елементу) вивідної зашифрованої послідовності лінгвістичних одиниць;

$N_{вип}$ – псевдовипадкове ціле число, що згенероване ГПВП на *і*-ому кроці генерування для шифрування *і*-го істинного елементу із порядковим номером $N_{вих}$;

$p_к$ - розмірність *к*–го підсловника;

к – порядковий номер підсловникау словнику мови прикладної області.

Рівняння розшифрування інформації у даному випадку являє собою формулу для визначення порядкового номеру розташування в обраному підсловнику *і*-ої істинної лінгвістичної одиниці, що поміщається у вивідну розшифровану послідовність замість маскуючої лінгвістичної одиниці, узятої із вихідної зашифрованої послідовності.

Рівняння розшифрування має наступний вигляд:

$$N_{вих} = \mathrm{mod}_{p_k}(N_{заш} - N_{вип}), \tag{3.21}$$

де $N_{вих}$ – порядковий номер лінгвістичної одиниці (згідно нумерації у *к*-ому підсловнику), що є ідентичним істинній одиниці, відображеній у вихідній відкритій послідовності лінгвістичних одиниць;

$N_{заш}$ – порядковий номер зашифрованої лінгвістичної одиниці (згідно нумерації у *к*-ом підсловнику), узятої із зашифрованої послідовності, що надана на обробку, на поточному кроці процедури розшифрування;

$N_{вип}$ – псевдовипадкове ціле число, згенероване ГПВП, для розшифрування зашифрованої одиниці із порядковим номером $N_{заш}$;

$p_к$ – розмірність (тобто, загальна кількість слів) *к*-го підсловника;

к – порядковий номер підсловника, вибраного на поточному кроці процедури розшифрування.

Вираз (3.21) є справедливим, якщо розмах ГПВП дорівнює розмірності підсловника, котрий має найбільшу кількість мовних елементів серед множини усіх підсловників використаного словника предметної області.

3.3. Аналіз методу

3.3.1. Особливості побудови генератора псевдовипадкових послідовностей для криптосистем із збільшеною відстанню єдиності

Слід зазначити, що ГПВП, які використовуються в системах захисту інформації, повинні відповідати наступним вимогам [35-39]:

- висока криптографічна стійкість;
- хороші статистичні властивості, тобто псевдовипадкові послідовності, що генеруються, за своїми статистичними властивостями мають не відрізнятися за однозначно визначених умов від істинно випадкових послідовностей;
- великий період випадкової послідовності, що генерується;
- ефективна апаратна та програмна реалізація.

На рис. 3.6 зображена схема роботи ГПВП у складі КЗВЄ. З рисунку видно, що знаючи величину періоду випадкової послідовності *T*, аналітик може зробити фільтр і „виловлювати” символи з періодом *T*.

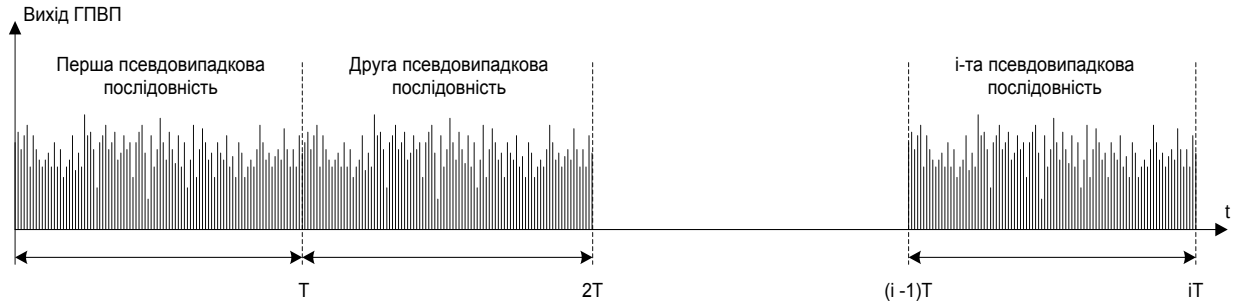

Рис. 3.6. Робота ГПВП у складі КЗВЄ без забезпечення випадкового характеру підмін символів

І, якщо застосувати процедуру шифрування, за якою кожна наступна таблиця шифрується із першого значення псевдовипадкової послідовності чисел, то у такому разі не забезпечується випадковий характер підмін символів.

На рис. 3.7 показано, яким чином ГПВП, що функціонує у складі КЗВЄ, при шифруванні таблиць з текстовою інформацією забезпечує випадковий

характер підмін. З рис. 3.7 видно, що кожна наступна таблиця шифрується послідовністю псевдовипадкових чисел без скидання ГПВП у початковий стан. Після кожного сеансу шифрування фіксується кількість псевдовипадкових чисел, що були використанні під час шифрування. І кожна наступна таблиця шифрується не із початкового стану ГПВП, а з тої точки послідовності, де вона була зафіксована на попередньому акті шифрування.

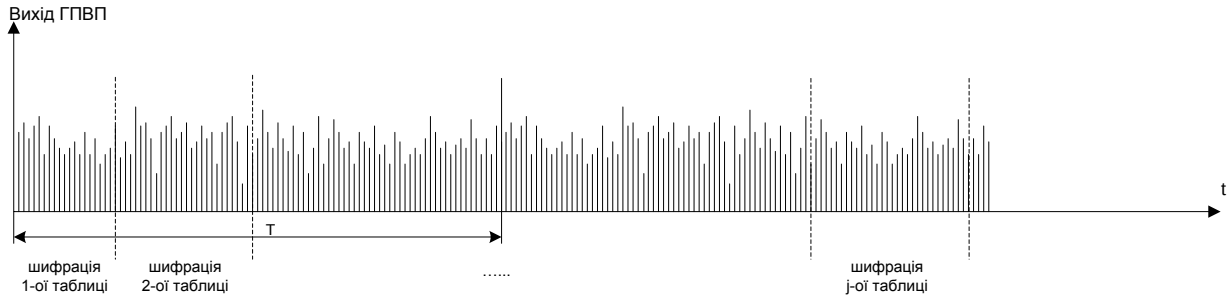

Рис. 3.7. Робота ГПВП у складі КЗВЄ із забезпеченням випадкового характеру підмін

Тим самим забезпечується випадковий характер підмін і, отже, нейтралізуються атаки типу імовірнісний аналіз частоти повторень символів вихідного тексту. Для забезпечення випадковості характеру підмін необхідно, щоб період ГПВП *T* був значно більшим за довжину таблиці *n*.

3.3.2. Визначення стійкості методу

1. Аналіз стійкості КЗВЄ за умов, коли хоча б один зразок вихідного (відкритого) текстового повідомлення та зразок кореспондованої з ним шифрограми є відомими

Початкові умови та ресурсні можливості зловмисника:

1) аналітику відомий хоча б один зразок вихідної відкритої і кореспондованої з нею зашифрованої таблиці;

2) учасники секретного обміну інформацією не змінюють ключ шифру на проміжку часу, коли відбувається атака;

3) аналітик має у розпорядженні засоби шифрування/розшифрування КЗВЄ, знає структуру ключа шифру, структуру заданої табличної форми та словник прикладної області, структура якого є узгодженою із структурою табличної форми;

4) аналіз стійкості задіяного алгоритму шифроперетворень на основі перехоплених шифрограм не проводився або не дав позитивного результату.

Склад обладнання, що використовується аналітиком для здійснення атаки на шифр за вищенаведених умов, показано на рис. 3.5.

Дії аналітика. Аналітик здійснює дешифрування відомого зашифрованого зразка відомої вихідної таблиці та робить спроби визначення ключа шифру шляхом прямого перебору усіх можливих значень цього ключа до тих пір, коли буде отримано відповідне відоме відкрите вихідне повідомлення.

Показник стійкості: K_1 – максимально можлива кількість переборів ключа шифру (дорівнює кількості можливих значень ключа шифру).

$$K_1 = x^k, \tag{3.22}$$

де x – основа алфавіту мови ключа шифру, k – розрядність ключа шифру.

Висновок щодо вищерозглянутої моделі атаки:

1) аналітик має можливість визначити факт успішного завершення атаки шляхом порівняння розшифрованого зразка відомої шифрограми з відповідним їй відкритим зразком вихідного повідомлення;

2) у цьому випадку дотримання чи недотримання обмежень щодо відстані єдиності на стійкість КЗВЄ не впливає, а стійкість КЗВЄ співпадає із стійкістю алгоритму шифроперетворень, що використовується згідно схеми, що зображена на рис. 3.5.

2. Аналіз стійкості КЗВЄ за умов відсутності кореспондованих зразків вихідних та зашифрованих текстових повідомлень та недотримання обмежень щодо відстані єдиності

Початкові умови та ресурсні можливості зловмисника:

1) у розпорядження аналітика надано достатній обсяг перехопленого шифротексту (отриманого на проміжку часу, коли ключ шифру не був змінений), щоб робити обґрунтовані статистичні висновки щодо імовірності появ окремих його елементів;

2) відсутність будь-яких кореспондованих пар зразків вихідної та зашифрованої інформації, тобто аналіз стійкості може виконуватися тільки на основі перехоплених шифротекстів;

3) аналітик має змогу визначити, що умова дотримання відстані єдиності не виконується, і, отже, атака на шифр не втрачає сенс;

4) схема генератора псевдовипадкових послідовностей (ГПВП) забезпечує випадковий характер підмін;

5) можливість отримати відомості про статистичні властивості словника прикладної області, узгодженого із структурою заданої табличної форми;

6) аналітик має у розпорядженні засоби шифрування/розшифрування КЗВЄ та словник прикладної області, знає структуру ключа шифру та структуру табличної форми.

Блок-схема обладнання, що використовується для здійснення атаки на шифр за вищенаведених умов аналогічна тій, що показана на рис. 3.5, але обсяг робіт, що має виконувати аналітик, є значно ширший.

Дії аналітика.

Підготовча стадія

1) Попереднє отримання інформації про статистичні властивості інформації секретного обміну, а саме:

а) отримання статистично повної вибірки шифрограм по проблемі у рамках заданого тезаурусу прикладної області;

б) обробка вибірки з метою отримання еталонної статистичної функції розподілу семантичних одиниць, що складають інформацію секретного обміну по проблемі.

Стадія атаки

1) дешифрування перехоплених зразків зашифрованого тексту методом прямого перебору усіх значень ключів шифру до тих пір, поки не будуть отримані усі зразки функції розподілу ймовірностей появи мовних елементів на виході дешифратору;

2) порівняння отриманих зразків функції розподілу з еталонною функцією, отриманою на підготовчій стадії. Прийняття рішення про найбільш імовірний варіант ключа.

Показник стійкості – статистичний:

$$K_2 = K_1 \times V \,, \tag{3.23}$$

де K_1 – показник стійкості шифроалгоритму, V – об’єм статистичної вибірки інформації секретного обміну.

Висновок щодо вищерозглянутої моделі атаки: стійкість КЗВЄ у V разів вища за стійкість методу шифроперетворень, що використовується, та є статистичною величиною.

Розглянемо графік залежності показника стійкості K від розрядності ключа шифру k щодо двох вищерозглянутих моделей атак (див. рис. 3.8).

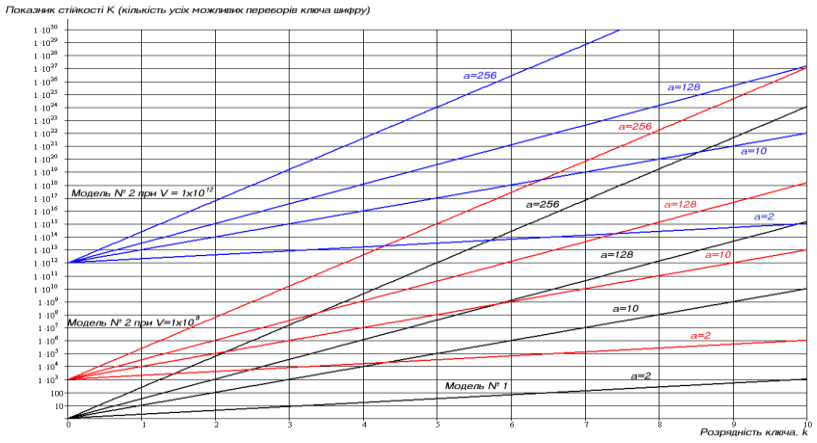

Рис. 3.8. Графік залежності показника стійкості КЗВЄ від розрядності ключа шифру

По-перше, тривіальний результат: із ростом довжини ключа шифру завжди стійкість системи захисту зростає. По-друге, у будь-яких випадках стійкість КЗВЄ щодо атак типу №1 суттєво нижча за стійкість цієї системи відносно атак типу №2. По-третє, з ростом обсягу вибірки V стійкість КЗВЄ зростає.

Розглянемо залежність показника стійкості K від параметру a, де a – основа алфавіту мови ключової інформації.

З графіку на рис. 3.8 видно, що при збільшенні значення параметру a значно збільшується показник стійкості K при сталому (постійному, незмінному) значенні розрядності ключа шифру k. Наприклад, якщо розглянути на графіку сімейство кривих, при моделі атак №1, то видно, що при розрядності ключа $k = 8$, показник стійкості приймає наступні значення: $K = 256$ при $a = 2$, $K = 1 \cdot 10^8$ при $a = 10$, $K = 1{,}4064 \cdot 10^{12}$ при $a = 128$ і $K = 1{,}8447 \cdot 10^{19}$ при $a = 256$. Отже, бажано збільшувати основу алфавіту мови ключової інформації.

3.4. Умови забезпечення режиму досконалої стійкості у криптосистемах із збільшеною відстанню єдиності

Початкові умови та ресурсні можливості зловмисника (випадок, коли параметри КЗВЄ забезпечують умови підтримки досконало секретної системи):

1) аналітик має у розпорядженні засоби шифрування/розшифрування КЗВЄ, у т.ч. словник прикладної області, але відсутня можливість перехоплення обсягу шифрограм, що перевищує відстань єдиності, оскільки система захисту задовольняє умовам досконалої стійкості;

2) відсутність будь-яких апріорних відомостей про вихідну текстову інформацію та кореспондований з нею зашифрований текст;

3) відсутність необхідності отримання будь-яких апріорних відомостей про статистичні властивості інформації секретного обміну (оскільки забезпечується умова підтримки досконало секретної системи).

Модель атаки (якщо параметри КЗВЄ відповідають умовам досконало секретної системи) - не потребує визначення.

Показник стійкості – не визначений, оскільки відсутні умови для визначення моменту успішного завершення атаки. В якості показника ефективності КЗВЄ у цьому випадку використовується відстань єдиності *U*.

Критерій коректного функціонування побудованої КЗВЄ – дотримання відстані єдиності (умова того, що обсяг зашифрованого одним ключем шифру не перевищить відстань єдиності), що визначається як [23]

$$U = \frac{\log_2(K)}{D}, \tag{3.24}$$

де U – відстань єдиності, K – максимально можлива кількість переборів ключа шифру, D – надлишковість прийнятої для відображення повідомлень мови.

Для виконання умови дотримання відстані єдиності необхідно коректно визначити розрядність (довжину) ключа k у кореляції з довжиною повідомлення n, маючи на увазі, що [23]

$$k = \log_2 N, \tag{3.25}$$

де k – розрядність ключа, N – кількість можливих значень повідомлення довжиною n. Отже, залежність ентропії (тобто, довжини у бітах) ключа шифру від довжини повідомлення можна записати наступним чином (вивід формули наданий у [8]):

$$H(K) = \log_2 \prod_{s=1}^{n} [B - (s-1)], \tag{3.26}$$

де B – кількість символів алфавіту мови, на якій представлене повідомлення.

Висновок щодо вищерозглянутої моделі атаки:

КЗВЄ за умов дотримання відстані єдиності має властивості досконало секретної системи.

3.5. Обмеження у застосуванні методу

Збільшення відстані єдиності є можливим лише для криптосистем, що використовуються у прикладних областях, для котрих створені відповідні штучні мови відображення цих прикладних областей, що характеризуються алфавітами великої розмірності. У цьому випадку відкривається можливість суттєвого спрощення системи розповсюдження ключової інформації, оскільки режим досконалої секретності може бути забезпечено за більших обсягів зашифрованої інформації без необхідності зміни ключа шифру. Окрім того, такі криптосистеми не потребують визначення семантичних співвідношень між лінгвістичними конструкціями синтезованих мов і, отже, створення семантичних тезаурусів прикладних областей. Проте, неперевищення відстані єдиності є обов'язковою умовою дотримання режиму досконалої секретності у створених таким чином криптосистемах, що не усуває жорстких вимог до системи розповсюдження ключової інформації [23].

Висновки до третього розділу

1. Запропоновано метод побудови досконало стійкої криптосистеми захисту від порушень конфіденційності текстової інформації, що береться із семантичного словника наперед визначеної прикладної області і поміщається у задану табличну форму. Метод базується на застосуванні механізму укрупнення алфавіту мови відображення текстової інформації, що використовує словник прикладної області і враховує структуру табличної форми. Внаслідок укрупнення алфавіту збільшується так звана відстань єдиності, що є основним пороговим показником приналежності криптосистеми до класу досконало стійких систем захисту з теоретично доведеною ідеальною теоретико-інформаційною стійкістю. Невеликі значення відстані єдиності при шифруванні повідомлень, складених із символів алфавітів природних мов, обумовлюють необхідність частої зміни ключової інформації, що є проблемою для багатьох прикладних застосувань.

Даний метод певною мірою усуває цю проблему і дозволяє розширити області використання досконало стійких криптосистем.

2. Отримано математичні вирази та побудовано відповідні графіки, що визначають залежності відстані єдиності та ентропії ключа шифру від довжини повідомлення. Графіки показують, що шляхом укрупнення алфавіту мови відображення інформації у таблиці заданої форми можемо після відповідних перетворень мати шифротексти, які значно коротші за відстань єдиності. Отримані результати доцільно використовувати для обчислення максимально можливої кількості сеансів зв'язку без зміни ключа шифру при дотриманні умов, коли не порушується ознака досконало стійкої криптосистеми.

3. Порівняно ефективність методу захисту з укрупненням алфавіту із ефективністю методу одноразових блокнотів, оскільки обидва вищезазначені методи захисту можуть бути реалізованими у рамках досконало стійких криптосистем. Ефективність у даному випадку розуміється як виграш у довжині ключа (за інших рівних умов) у разі застосування у криптосистемі укрупненого алфавіту у порівнянні із методом одноразових блокнотів. Отримано математичний вираз для визначеного показника виграшу. Показано, що показник виграшу лінійно залежить від довжини повідомлень, що представлені на синтезованій мові, і практично не залежить від обсягу алфавіту використаної мови людського спілкування.

Ефективність запропонованого методу побудови досконало стійкої криптосистеми, якщо в якості критерію ефективності обрана відстань єдиності, характеризується суттєво вищими рівнями у порівнянні, із ефективністю інших методів (зокрема, методу одноразових блокнотів) забезпечення режиму досконалої секретності. У наведеному прикладі маємо майже 50-ти кратний виграш за даним критерієм ефективності.

4. Отримано математичні вирази, що відображають процедури шифрування/розшифрування, і є справедливими за умов співпадіння розмірності застосованого підсловника з розмахом випадкових

послідовностей, що генеруються ГПВП. Рівняння шифрування являє формулу для визначення порядкового номеру розташування маскуючого слова у *к*-ому підсловнику тезаурусу. Рівняння розшифрування являє формулу для визначення порядкового номеру розташування в обраному підсловнику *і*-ої істинної лінгвістичної одиниці, що поміщається у вивідну розшифровану послідовність замість маскуючої лінгвістичної одиниці, узятої із вихідної зашифрованої послідовності.

5. Розглянуто особливості побудови ГПВП. Кожна наступна таблиця шифрується послідовністю псевдовипадкових чисел без скидання ГПВП у початковий стан. Тим самим забезпечується випадковий характер підмін і, отже, нейтралізуються атаки типу імовірнісний аналіз частоти повторень символів вихідного тексту.

6. Визначено стійкість криптосистеми із збільшеною відстанню єдиності за різних ресурсних можливостей порушника. За умов, коли хоча б один зразок вихідного (відкритого) текстового повідомлення та зразок кореспондованої з ним шифрограми є відомими, стійкість синтезованої криптосистеми співпадає із стійкістю алгоритму шифроперетворень, що використовується. За умов відсутності кореспондованих зразків вихідних та зашифрованих текстових повідомлень та недотримання обмежень щодо відстані єдиності стійкість криптосистеми є статистичною величиною і у *v* разів вища за стійкість методу шифроперетворень, що використовується, де *v* – об’єм статистичної вибірки інформації секретного обміну. З ростом обсягу вибірки *v* стійкість криптосистеми зростає.

Отримано математичні вирази та надано графік залежності показника стійкості від розрядності ключа шифру. З ростом довжини ключа шифру стійкість системи захисту зростає. При збільшенні основи алфавіту мови ключової інформації значно збільшується показник стійкості криптосистеми при сталому значенні розрядності ключа шифру.

7. Визначено формальні умови коректного функціонування побудованої криптосистеми та розрядність ключа шифру у кореляції із довжиною повідомлень, що закриваються.

8. Застосування методу є можливим лише для криптосистем, що використовуються у прикладних областях, для котрих створені відповідні штучні мови відображення цих прикладних областей, що характеризуються алфавітами великої розмірності.

РОЗДІЛ 4
РОЗРОБКА МЕТОДУ ПОБУДОВИ КРИПТО-СЕМАНТИЧНОЇ СИСТЕМИ ЗАХИСТУ МОВНОЇ ІНФОРМАЦІЇ

Метод побудови досконало стійкої криптосистеми, розробка котрого відображена у третьому розділі даної роботи, за рахунок укрупнення алфавіту джерела текстових повідомлень забезпечив збільшення відстані єдиності шифру за ключем, що надало змогу суттєво збільшити довжину шифрованих повідомлень відносно довжини ключової інформації [22-23], не порушуючи при цьому шеннонівської умови досконалої стійкості захисту. У результаті проміжок часу між зміною ключів шифру у порівнянні із схемою Вернама може бути суттєво збільшеним (у наведеному прикладі – майже у 50 разів). Але з формальної точки зору цей метод не усуває необхідність у періодичній зміні ключової інформації.

У цьому розділі поставлено завдання з розробки методу побудови досконало стійкої криптосистеми, заснованої на використанні прикладних лексикографічних систем захисту текстової інформації, зокрема тезаурусів смислових образів (синтез структури котрих виконано у другому розділі), що забезпечує незалежність обсягу текстової інформації, що підлягає шифруванню, від значень відстані єдиності та довжини ключів шифрів. Тобто, у цьому випадку відпадає необхідність дотримання умови неперевищення відстані єдиності і відкривається можливість вибору довжини ключів шифру незалежно від обсягів інформації, що потребують шифрування.

4.1. Формальне обґрунтування можливості створення досконало стійких незалежних від довжини ключу шифру криптосистем

Процес шифрування текстових повідомлень, що надані у письмовій або голосовій формі, представимо у наступному вигляді:

$$\{\mathbf{S}\} \xrightarrow{\mathbf{P_{SF}}} \{\mathbf{F}\} \xrightarrow{\mathbf{P_{Ш}(x)}} \{\mathbf{F_{Ш}}\}, \qquad (4.1)$$

де {**S**} – простір смислових образів, елементи котрого використовуються під час формування вихідних (рос., – исходных) текстових повідомлень; {**F**} – простір відображень засобами обраної мови смислових образів, узятих із простору {**S**}; {$\mathbf{F_{Ш}}$} – простір зашифрованих відображень смислових образів із простору {**S**}; $\mathbf{P_{SF}}$ – оператор перетворення смислових образів у їхні відображення у рамках обраної мови, $\mathbf{P_{Ш}(x)}$ – оператор шифрування, що залежить від вибору ключа шифру **x**.

Процес розшифрування текстових повідомлень представимо у наступному вигляді:

$$\{\mathbf{F_{Ш}}\} \xrightarrow{\mathbf{P_{рш}(x)}} \{\mathbf{F}\} \xrightarrow{\mathbf{P_{FS}}} \{\mathbf{S}\}, \qquad (4.2)$$

де $\mathbf{P_{рш}(x)}$ – оператор розшифрування, що залежить від вибору ключа шифру **x**, а $\mathbf{P_{FS}}$ – оператор перетворення відображень смислових образів у вихідні смислові образи.

Для відтворення операторів $\mathbf{P_{SF}}$ та $\mathbf{P_{FS}}$ використаємо крипто-семантичний підхід до забезпечення досконало стійкого захисту інформації, а оператори $\mathbf{P_{Ш}(x)}$ та $\mathbf{P_{рш}(x)}$ реалізуємо будь-яким із відомих методів симетричної криптографії.

Підкреслимо, що у виразах (4.1) та (4.2) оператори $\mathbf{P_{Ш}(x)}$ та $\mathbf{P_{рш}(x)}$ не залежать від семантичних характеристик елементів простору {**S**}, тобто шифрування здійснюється без урахування смислу текстових повідомлень (саме тому методи класичної криптографії мають універсальний щодо смислу інформації характер).

Підкреслимо також, що у цьому випадку оператор перетворення форми відображення будь-якого смислового образа засобами будь-якої природної мови у його смисловий зміст P_{FS} має детермінований, однозначно визначений (правилами граматики) характер.

В основі відомих методів криптоаналізу зашифрованих текстових повідомлень лежить **аксіома** щодо однозначності оператора зворотного перетворення «перехоплений зразок шифрограми» в «істинний зміст зразка вихідного відкритого повідомлення». Тобто, якщо

$$\{\mathbf{F_Ш}\} \xrightarrow{\mathbf{P_{шS}(x)}} \{\mathbf{F}\} \qquad (4.3)$$

(де $\mathbf{P_{шS}(x)}$ – оператор зворотного перетворення зашифрованих текстових повідомлень безпосередньо у зміст вихідних відкритих повідомлень), то $\mathbf{P_{шS}(x)}$ має детермінований та однозначно визначений характер [22,23].

Примітка 4.1. *Якщо враховувати визначення поняття «інформація», що надано у розділі 1, під словосполученням «зміст повідомлення» у поясненнях до виразу (4.3) розуміється відображення смислових образів в кодах первинної інформації, а не їхній смисл (тобто, безпосередньо смислові образи, що є елементами смислового тезаурусу).*

Тому яку б теорію чи гіпотезу не висував криптоаналітик щодо істинності визначеного ним варіанту ключа шифру, у нього завжди існує можливість скористатися ефективним інструментом перевірки коректності результатів своїх крипто-аналітичних досліджень, а саме, якщо визначений ним варіант ключу забезпечує перетворення беззмістовного зразка шифрограми у текст, що має конкретний зміст, то результати криптоаналізу згідно з вищенаведеною аксіомою про однозначність перетворення вважаються успішними. І, отже, будь-який відомий шифр (за виключенням шифрів типу схеми Моборна/Вернама, які за певних умов не забезпечують однозначність перетворення) з теоретичної точки зору може бути скомпрометований.

Основна ціль дослідження, представленого у цьому розділі, – створити метод захисту інформації від порушень конфіденційності, що забезпечує з формальної точки зору абсолютну конфіденційність повідомлень, що

підлягають захисту, за умови, коли сумарні обсяги зашифрованої інформації у формально необмеженій мірі перевищують довжину паролів, що використовуються. (Реалізація такого методу дозволить, з одного боку, отримати стовідсоткову упевненість в забезпеченні конфіденційності повідомлень, а з другого боку, позбавитись проблеми розповсюдження ключів шифрів, оскільки у цьому разі при оптимальному виборі довжини ключа шифру відпадає необхідність додержання відстані єдиності).

Для досягнення цієї цілі використаємо крипто-семантичний підхід до забезпечення конфіденційності, що передбачає залежність операторів $\mathbf{P_{SF}}$ та $\mathbf{P_{FS}}$ від семантичних характеристик елементів простору смислових образів$\{\mathbf{S}\}$, тобто оператори шифрування/розшифрування крипто-семантичної системи захисту $\mathbf{P_{Ш}(x,\ S)}$ та $\mathbf{P_{рш}(x,S)}$ мають бути залежними як від вибору ключа шифру, так і від вибору елементу із простору смислових образів. Тобто, процес шифрування будемо здійснювати у наступному вигляді:

$$\{\mathbf{S}\} \xrightarrow{\mathbf{P_{ш}(x,S)}} \{\mathbf{F_{Ш}}\}, \qquad (4.4)$$

де $\mathbf{P_{ш}(x,S)}$ – оператор шифрування крипто-семантичної системи, що є залежним як від вибору ключа шифру **x**, так і від вибору елементу із простору смислових образів $\{\mathbf{S}\}$.

У цьому випадку шифрування буде здійснюватися із урахуванням смислу текстових повідомлень, а вищенаведена аксіома щодо однозначності перетворення «перехоплений зразок шифрограми» в її істинний зміст згідно виразу (4.3) не буде відповідати дійсності.

У даній роботі для отримання залежності оператора шифрування від $\{\mathbf{S}\}$, пропонується здійснювати заміну істинного смислового образу вихідного повідомлення (тобто, одного із елементів простору $\{\mathbf{S}\}$) на інший елемент семантичного тезаурусу даної прикладної області, смисловий образ

котрого знаходиться у відношенні смислової правдоподібності із істинним смисловим образом вихідного повідомлення, тобто

$$SO_k^{(1)} \wedge \{ SO_1^{(1)} \wedge SO_2^{(1)} \wedge \ldots \ldots \wedge SO_M^{(1)} \}, \qquad (4.5)$$

де $SO_k^{(1)}$ – істинний, наприклад, *к*-й смисловий образ *i*-го рівня абстрактності у смисловому потоці$SP^{(i+1)}$ (i+1) рівня абстрактності, тобто послідовності смислових образів і-го рівня абстрактності, що представляють $SP^{(i+1)}$ (див. вираз (1.5));

$\{ SO_1^{(1)} \wedge SO_2^{(1)} \wedge \ldots \ldots \wedge SO_M^{(1)} \}$ – множина із *M* правдоподібних щодо істинного смислових образів вихідного повідомлення, що входять до складу семантичного тезаурусу;

$к \neq \{1, 2, \ldots\ldots, M\}$.

Примітка 4.2. *Більш детально характеристики семантичних тезаурусів надано у розділі 2.*

За цих умов робота криптоаналітика втрачає сенс, оскільки він буде позбавлений можливості відрізнити істинний смисл зашифрованих повідомлень від багатьох інших фальшивих, але правдоподібних смислових образів.

4.2. Синтез крипто-семантичного методу побудови системи захисту

4.2.1. Загальна постановка завдання

Припустимо, що задано функціональний простір прикладної області застосування певної і-ої інформаційної системи $\mathbf{П_i}$, цільове використання котрої пов’язано з передаванням, обробкою та зберіганням текстових повідомлень, у вигляді

$$\mathbf{П_i}(\mathbf{Ф_i},\mathbf{Z_i}) \in \{\mathbf{Ф_i},\mathbf{Z_i}\} = \{\Phi_{i,1}, \Phi_{i,2}, \ldots \Phi_{i,N}; Z_{i,1}, Z_{i,2}, \ldots Z_{i,M}\}, \qquad (4.6)$$

де $\mathbf{\Phi_i}$ – функціональна структура і-ої інформаційної системи, що складається із елементів множини Φ_i припустимих для виконання функцій $\Phi_{i,к}$, де к=1, 2, …. , N;

$\mathbf{Z_i}$ – функціональна структура і-ої інформаційної системи, що складається із елементів множини Z_i обмежувальних умов Z_i, де k = 1, 2,…,M, за яких виконання функцій і-ої інформаційної системи є припустимим;

$\Phi_{i,k}$ – k-й функціональний елемент із заданої множини припустимих для виконання функцій і-ої інформаційної системи Φ_i, де k = 1, 2, …, N;

$Z_{i,k}$ – k-та обмежувальна умова із заданої множини обмежувальних умов Z_i, де k = 1, 2, …, M;

N, M – кількість елементів відповідно у множинах Φ_i та Z_i.

Вираз (4.6) задає функціональність простору $\mathbf{\Phi_i}$ прикладної області застосування і-ої інформаційної системи та простір умов $\mathbf{Z_i}$, за яких дана інформаційна система може бути використана за цільовим призначенням.

Необхідно забезпечити захист смислу інформації, що передається, оброблюється та зберігається засобами і-ої інформаційної системи, від порушень конфіденційності на рівні надання абсолютних гарантій неможливості злому системи захисту як з теоретичної (формальної), так і з практичної точок зору. При цьому обсяг інформації, що підлягає захисту, за будь-якої довжини ключа шифру, що використовується, має не обмежуватися довжиною цього ключа, а довжина ключа обиратися в залежності від припустимого значення ймовірності прийняття безпомилкових рішень в процесі дешифрування смислового потоку.

4.2.2. Формальний спосіб виконання завдання

Для виконання поставленого завдання використаємо крипто-семантичний підхід у рамках представлених вище уявлень та інформаційних моделей щодо функціонування інтелекту (див. розділ 1). Для того, щоб оператори шифрування/розшифрування $\mathbf{P_{SF}}$ та $\mathbf{P_{FS}}$ у виразах (4.1) та (4.2) могли бути реалізовані методами семантичної криптографії, необхідно

простір заданої прикладної області застосування і-ої інформаційної системи $\mathbf{П_i}$ відобразити на просторі смислових образів цієї системи $\mathbf{S_i}$ з урахуванням семантичних співвідношень між ними, що визначаються заданим простором обмежувальних умов $\mathbf{Z_i}$, тобто $\mathbf{P_{ФS(Zi)}}$

$$\mathbf{П_i(Ф_i,Z_i) \rightarrow \{S_{i,},Z_i\}}, \qquad (4.7)$$

де $\mathbf{P_{ФS(Zi)}}$ – оператор відображення $\mathbf{П_i}$ на $\mathbf{S_i}$ з урахуванням $\mathbf{Z_i}$.

Фактично $\{\mathbf{S_{i,}Z_i}\}$ задає структуру семантичного тезаурусу заданої прикладної області, формальний синтез котрої виконано у розділі 2.

Примітка 4.3. *У задачах криптосемантики основу лінгвістичних корпусів складають семантичні словники (інша назва - тезауруси знань), що містять у своєму складі усю множину певним чином пов'язаних між собою семантичних елементів (семантичних образів), які у сукупності являють повне мовне відображення простору тих прикладних сфер, щодо яких відповідні семантичні словники були створені.*

Ця структура може бути використана при створенні системи захисту, що передбачає (згідно висновків, зроблених у попередньому підрозділі) залежність операторів P_{SF} та P_{FS} від семантичних характеристик елементів простору смислових образів $\{S\}$.

Пропонується крипто-семантичний метод захисту смислу текстових повідомлень шляхом створення симетричної криптосистеми, у рамках якої оператори шифрування/розшифрування $\mathbf{P_{Ш}(x, S)}$ та $\mathbf{P_{рш}(x,S)}$ є залежними не тільки від вибору ключа шифру із простору **x**, але і від вибору елементу із простору смислових образів $\mathbf{S_i}$.

Згідно даного методу у структурі тезаурусу кожному смисловому образу із $\mathbf{S_i}$ ставиться у відповідність певним чином визначена множина інших смислових образів із цього ж $\mathbf{S_i}$, що знаходяться із цим смисловим образом у відношенні смислової правдоподібності (див. підрозділ 2.2, вираз (2.8)). Така структура тезаурусу надає можливість за певних умов під час шифрування

замінювати повідомлення, що підлягають шифруванню, на інші правдоподібні повідомлення, що не відображають істинний смисл вихідних повідомлень. А під час розшифрування, якщо пароль є відомим, здійснювати зворотні заміни правдоподібних зашифрованих повідомлень на істинні щодо смислу. Зрозуміло, що для зловмисника пароль є невідомим. Отже, можливостей щодо здійснення зворотних замін, як показує подальший аналіз цього способу захисту, він не має.

Запропонований у даній роботі метод захисту передбачає створення системи захисту у два етапи: спочатку створюється семантичний словник (тезаурус) $\mathbf{T_Z}$ прикладної області $\mathbf{П_i}$, а потім розроблюються програмно-технічні засоби реалізації операторів шифрування/розшифрування $\mathbf{P_Ш(x, S)}$ та $\mathbf{P_{рш}(x,S)}$ (див. вирази (4.1)-(4.2)), у складі яких використовується створений тезаурус.

Оператори шифрування/розшифрування крипто-семантичної системи захисту $\mathbf{P_Ш(x, S)}$ та $\mathbf{P_{рш}(x,S)}$, згідно даного методу, є залежними як від вибору ключа шифру із **x**, так і від вибору елементу із простору смислових образів **S.** Процес шифрування передбачає генерування поряд із зашифрованим відображенням істинного смислового образу вихідного повідомлення (тобто, одного із елементів простору {**F**}), ще і представницької вибірки – множини зашифрованих відображень фальшивих, але правдоподібних смислових образів (тобто, певним чином визначену кількість елементів із простору $\{\mathbf{F_Ш}\}$, вихідні смислові образи котрих пов'язані відношенням правдоподібності із зашифрованим відображенням смислового образу істинного зразка мовного повідомлення). Як показує подальший аналіз (див. підрозділ 4.3), робота криптоаналітика за цих умов, як правило, втрачає сенс, оскільки він позбавлений можливості відрізнити істинний смисл зашифрованих повідомлень від багатьох інших фальшивих, але правдоподібних смислових образів.

В деяких випадках може виявитися корисною для криптоаналітика інформація про кількість L елементів у підмножині смислових образів $\{F^k_Ш\}$,

що пов’язані відношенням правдоподібності із зашифрованим відображенням смислового образу істинного зразка текстового повідомлення S_k. Ця інформація вважається апріорі відомою і, якщо L має невеликі відносно 1 значення, то значення ймовірності прийняття безпомилкового рішення щодо S_k в процесі дешифрування може виявитися достатньо значною, щоб вважати результат дешифрування корисним для зловмисників. Тому з точки зору легальних користувачів системи захисту в залежності від конкретних умов прикладної області $\mathbf{П_i}$ доцільно збільшувати значення L до певного припустимого рівня L_0, при перевищенні котрого робота криптоаналітика втрачає сенс. За цих умов у загальному випадку (якщо не здійснювати ймовірнісний аналіз потоку вихідних смислових образів, що зроблено у наступному підрозділі) ймовірність прийняття безпомилкового рішення в процесі дешифрування щодо істинності S_k визначиться із виразу

$$p(S_k) \approx 1/ L_0. \quad (4.8)$$

Тому мінімально можливе значення довжини ключа шифру n_{min} доцільно обирати із нижче наведеного співвідношення:

$$n_{min} \geq | \log_2 L_0 |. \quad (4.9)$$

Тобто, довжина ключа шифру для крипто-семантичного шифрування має обиратися в залежності від припустимого значення ймовірності прийняття безпомилкових рішень в процесі дешифрування смислового потоку.

4.2.3. Алгоритм реалізації методу побудови лексикографічної криптосистеми

Нехай будуть прийняті наступні позначення:

S – суб’єкт, що синтезує зразок вихідного розмовного повідомлення D у процесі вирішення прикладних завдань заданої області застосувань;

D – зразок необробленого вихідного відкритого (незашифрованого) текстового повідомлення, смисловий зміст якого потребує захисту (S:D, де «:» –знак візуального або розмовного сприйняття);

S_F – оператор обробки вихідного зразка за правилами граматики мови (природної або штучної), що прийнята 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 щодо структури цього повідомлення (прийнята 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 семантичному тезаурусу прикладної області та визначення параметрів локалізації даного повідомлення у структурі тезаурусу;

C_D – зразок відкритого текстового повідомлення, що безпосередньо відображає смисл вихідного текстового повідомлення і відповідає прийнятим граматичним правилам та семантичним обмеженням;

S_Z – оператор шифрування, що перетворює вихідний зразок відкритого текстового повідомлення на зразок зашифрованого текстового повідомлення, що має правдоподібний, але, найбільш ймовірно, інший смисловий зміст (чим забезпечується неоднозначність у сприйнятті смислу зашифрованого повідомлення);

Z_D – зразок зашифрованого текстового повідомлення (синтезований із семантичних одиниць використаного тезаурусу), смисловий зміст котрого, найбільш ймовірно, відрізняється від істинного смислу відкритого вихідного текстового повідомлення;

S^0_Z – оператор розшифрування, що забезпечує перетворення неоднозначного щодо смислу зашифрованого текстового повідомлення на зразок текстового повідомлення, що однозначно відображає істинний смисл відкритого вихідного текстового повідомлення;

Z^0_D – зразок розшифрованого текстового повідомлення, що повністю співпадає із C_D, тобто $Z^0_D \leftrightarrow C_D$, де символ $\leftrightarrow$ означає збіжність форми та смислу повідомлень Z^0_D та C_D.

Тоді алгоритм реалізації методу побудови лексикографічної криптосистеми має відтворювати послідовність операцій з обробки текстових повідомлень, що показана на рис. 4.1 у вигляді діаграми форм представлення оброблюваного зразка текстового повідомлення та операторів перетворення цих форм засобами системи захисту.

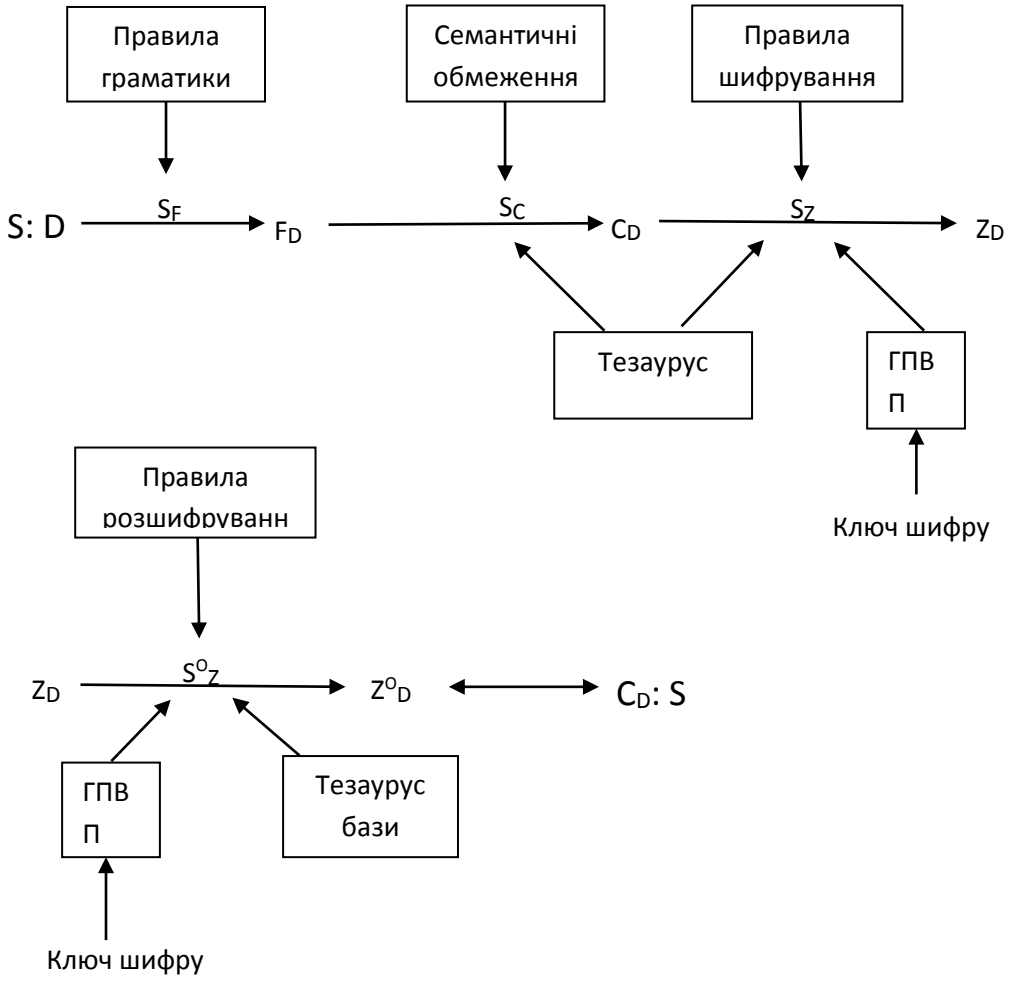

Рис. 4.1. Алгоритм реалізації лексикографічної криптосистеми

Конфіденційність текстових повідомлень згідно даного методу забезпечують наступним чином. Зразок незашифрованого вихідного

текстового повідомлення D за допомогою засобів спеціально розробленого лінгвістичного корпусу оброблюють за правилами S_F граматики мови (природної або штучної), що прийнята для відображення мовних повідомлень заданої області прикладних застосувань.

Як результат, отримують коректно сформований за правилами граматики зразок F_D відкритого вихідного текстового повідомлення та додаткову інформацію щодо структури цього повідомлення (прийняту систему мовних одиниць, локалізацію мовних одиниць у дискретній послідовності цих одиниць, результати маркування (розмітки) вихідного повідомлення за лінгвістичними характеристиками тощо).

Далі виконують аналіз S_C зразка F_D на відповідність елементам семантичному тезаурусу, спеціально розробленого для відображення смислу мовних повідомлень та семантичних зв'язків між ними у вигляді відповідних семантичних співвідношень для заданої прикладної області. Як результат, отримують зразок вихідного відкритого повідомлення C_D, що без спотворень відображає істинний смисл вихідного текстового повідомлення і відповідає прийнятим граматичним правилам та семантичним обмеженням. Далі безпосередньо виконують шифрування з використанням будь-якого відомого симетричного шифру S_Z, що заснований на застосуванні генератора псевдо випадкових послідовностей (ГПВП). А саме, згідно з паролем встановлюють ГПВП у початковий стан і далі шляхом гамування (тобто, виконання операції *XOR* - додавання за модулем 2) елементи повідомлення C_D заміняють на правдоподібні елементи із тезауруса. (Вибір методів заміни істинних щодо смислу повідомлень на правдоподібні залежить від синтезованої структури тезаурусу, зокрема від кількості врахованих рівнів абстрагування його семантичних одиниць.) Як результат, отримують зразок зашифрованого текстового повідомлення Z_D, що має правдоподібний, але, найбільш ймовірно, інший смисл (чим забезпечується неоднозначність у сприйнятті смислу зашифрованого повідомлення). Для розшифрування повідомлення Z_D застосовують оператор зворотного перетворення S^0_Z. А

саме, за допомогою відомого паролю встановлюють ГПВП у той самий початковий стан, що був під час шифрування, і далі використовують операцію *XOR* повторно, тобто елементи неоднозначного щодо смислу зашифрованого повідомлення Z_D шляхом відповідного вибору із тезаурусу заміняють на елементи повідомлення Z^0_D, що однозначно відображає істинний смисл вихідного мовного повідомлення C_D і відповідає прийнятим граматичним правилам та семантичним обмеженням заданої прикладної області.

Стисло суть методу можна визначити як включення до симетричної системи криптографічного захисту інформації засобів лексикографічної системи таким чином, щоб у процесі шифрування забезпечувалась семантична неоднозначність зашифрованих зразків повідомлень. При цьому засоби лексикографічної системи, що використовуються, мають бути спроможними здійснювати граматичний та семантичний аналіз повідомлень у рамках заданої прикладної області.

4.3. Аналіз методу побудови лексикографічної криптосистеми

Даний метод захисту смислу текстових повідомлень базується на забезпеченні неоднозначності зворотного переходу від відображення зашифрованої сутності (тобто, смислу або, іншим словом, суті змісту) вихідного розмовного повідомлення, представленого у вигляді однієї із можливих форм цього відображення, до відображення істинної суті змісту вихідного розмовного повідомлення. В основу даного методу покладена фундаментальна особливість філософської категорії «співвідношення форма - суть», що розглянута у розділі 1.2. А саме: будь-яка сутність може бути відображена різноманітними формами. Теорія крипто-семантики базується на гіпотезі, що, у загальному випадку, однозначного зв’язку між сутністю розмовних повідомлень та формою їхнього відображення не існує. Отже, якщо перехоплені криптоаналітиком зразки зашифрованих мовних повідомлень (у рамках формально визначених обмежень обраної області

прикладних застосувань) являють собою семантично правдоподібні повідомлення, а будь-який результат застосування будь-якого із можливих способів криптоаналізу приводить до отримання нехай іншого щодо смислу, але теж семантично правдоподібного повідомлення, то істинність вихідного повідомлення не є можливим виявити, оскільки криптоаналітик не має можливостей визначити факт успішного завершення своєї роботи.

Примітка 4.6. *З теоретичної точки зору за певних умов є можливим здійснювати апостеріорні ймовірнісні оцінки появи тих чи інших смислових одиниць у потоці смислових образів, що надходять на увід приймальної частини прикладної інформаційної системи. Проте здійснювати таке оцінювання під час криптоаналізу не має сенсу, оскільки отримання функції розподілу апостеріорних оцінок ймовірностей появи смислових одиниць не надає додаткових можливостей для криптоаналізу відносно знань функції розподілу апріорних оцінок ймовірностей появи цих смислових одиниць.*

Вищенаведене твердження поширюється на симетричні крипто-семантичні системи за умов, коли:

1) усі семантичні одиниці вихідних текстових повідомлень є елементами семантичного тезаурусу інформаційної системи, що захищається:

$$\{SO\} \in \{TZ_M\}, \tag{4.10}$$

де $\{SO\}$ – кінцева множина смислових образів, що використовують різні суб'єкти в процесі виконання завдань у рамках визначеної області прикладних застосувань;

$\{TZ_M\}$ – тезаурус прикладної області, тобто упорядкований набір тлумачних (сценарних, фразеологічних, семантичних тощо) словників, що упорядковані за рівнями абстрактності смислових образів.

2) структура тезаурусу (профіль функціональності тезаурусу) відображає структуру співвідношень між семантичними одиницями

прикладної області (профіль функціональності прикладної області), в рамках якої використовується інформаційна система, що захищається:

$$\text{🏛}\{SO\} \equiv \text{🏛}\{TZ_M\}, \quad (4.11)$$

де 🏛 – знак профілю функціональності;

3) тезаурус побудовано таким чином, що кожній семантичній одиниці, що входить до складу тезаурусу $\{TZ_M\}$, відповідає кілька інших семантичних одиниць, що знаходяться між собою у відношенні семантичної правдоподібності:

$$SO_k^{(1)} \wedge \{SO_1^{(1)} \wedge SO_2^{(1)} \wedge \ \dots \ \dots \wedge SO_M^{(1)}\}, \quad (4.12)$$

де $SO_k^{(1)}$ – $к$-та семантична одиниця i-го рівня абстрактності;

$\{ SO_1^{(1)} \wedge SO_2^{(1)} \wedge \ \dots \ \dots \ \wedge SO_M^{(1)} \}$ – множина із M правдоподібних щодо $SO_k^{(1)}$ семантичних одиниць, що входять до складу семантичного тезаурусу;

$$к = \{1, 2, \dots , M\};$$

$$i = \{1, 2, \dots , I\}.$$

Один із можливих шляхів реалізації даного методу захисту полягає у наступному. У рамках формально визначених обмежень щодо множини смислових образів, що відображають тезаурус обраної сфери прикладних застосувань, розроблюють лексикографічну систему, яку реалізують у вигляді прикладного лінгвістичного корпусу, і далі безпосередньо перед шифруванням текстових повідомлень згідно з будь-яким обраним способом симетричної криптографії здійснюють семантичну структуризацію цієї інформації з використанням засобів побудованого лінгвістичного корпусу таким чином, щоб зразки зашифрованих текстових повідомлень представлялися як семантично правдоподібні повідомлення, а будь-який результат застосування будь-якого із можливих способів криптоаналізу

приводив до отримання нехай інших щодо смислу, але теж семантично правдоподібних текстових повідомлень.

Запропонований метод забезпечення конфіденційності текстових повідомлень, згідно якого здійснюють шифрування/розшифрування зразків текстових повідомлень із застосуванням симетричного криптографічного алгоритму, базується на використанні властивостей генератора дискретних (псевдо)випадкових послідовностей, ***який відрізняється тим***, що перед шифруванням зразки вихідних повідомлень за допомогою засобів лінгвістичного корпусу структуруються за лінгвістичними характеристиками таким чином, щоб шифрування будь-якого зразка вихідного повідомлення, що складається із елементів семантичного тезаурусу мови відображення заданої кінцевої множини текстових повідомлень, приводило до отримання правдоподібного зразка текстового повідомлення, смисл котрого випадковим чином відрізняється від смислу зразка вихідного відкритого повідомлення.

Запропонований набір операцій з обробки текстових повідомлень та послідовність їхнього виконання забезпечують можливість отримання абсолютної гарантії захисту смислу текстових повідомлень від порушень конфіденційності навіть в умовах, коли обсяги текстових повідомлень, що підлягають шифруванню, у багато разів перевищують обсяги парольної (ключової) інформації, що містяться у ключах задіяних шифрів. Оскільки семантична структура тезаурусу будується таким чином, щоб будь-яких статистичних та (або) семантичних зв’язків між відображенням смислу правдоподібних, але фальшивих текстових повідомлень та відображенням істинного смислу вихідних повідомлень не має існувати. Іншими словами, крипто-семантичний метод забезпечує абсолютне теоретично стійке шифрування.

Сукупність методів захисту текстових повідомлень, алгоритми реалізації котрих передбачають необхідність використання перед шифруванням методів лінгвістичної обробки текстових повідомлень за їхніми семантичними характеристиками з метою визначення відношень

правдоподібності, доцільно виділити в окрему групу криптографічних методів, яку назвати як семантична криптографія, оскільки об’єктами захисту у рамках цієї групи є смисловий зміст текстових повідомлень, що захищаються.

Для реалізації крипто-семантичного методу захисту смислу текстових повідомлень необхідно оброблювати вихідні зразки текстових повідомлень за правилами граматики мови (природної або штучної), що прийнята для формування мовних повідомлень заданої сфери застосувань. Для цього доцільно використати засоби спеціально розробленого лінгвістичного корпусу [23], зокрема використати, окрім вихідних зразків текстових повідомлень, дані щодо структури потоків текстових повідомлень, їхніх лінгвістичних характеристик, правил граматики мови, що прийнята для відображення повідомлень заданої сфери застосувань. Засоби відтворення цих даних складають так званий лінгвістичний корпус, з використанням котрого має здійснюватися маркування (розмітка) потоків текстових повідомлень за їхніми лінгвістичними характеристиками. Має забезпечуватися коректність граматичної структури вихідних повідомлень, що створюються людиною або комп’ютерною програмою, відповідно до правил та обмежень, що задаються засобами лінгвістичного корпусу. У випадках невідповідності цим правилам мовні повідомлення мають повертатися їхнім створювачам на доопрацювання.

Засоби лексикографічної системи, що використовуються, мають бути спроможними здійснювати граматичний та семантичний аналіз текстів у рамках заданої прикладної області.

В якості вихідних даних, що є необхідними для реалізації методу, окрім граматично коректних вихідних мовних повідомлень, мають використовуватися семантичні конструкції задіяного тезаурусу мови.

Область використання крипто-семантичного методу обмежується умовами (4.10) – (4.12), тобто елементи усіх можливих зразків лінгвістичних конструкцій, що відображають смисл текстових повідомлень, мають

міститися у тезаурусі, спеціально розробленого для відображення смислу мовних повідомлень заданої сфери застосувань.

Ймовірнісний аналіз

Для оцінки ефективності систем захисту мовної інформації знайшов широке застосування ймовірнісний (частотний) метод аналізу, оскільки у природних мовах завжди існують ймовірнісні зв'язки між лінгвістичними одиницями цих мов [9,16,26,40]. Покажемо, що крипто-семантичний спосіб захисту мовної інформації є не чутливим до результатів ймовірнісного криптоаналізу.

Припустимо, що дискретна функція розподілу апріорних ймовірностей семантичних одиниць і-го рівня абстрагування із тезаурусу даної прикладної області є відомою і має наступний вигляд:

$$p\,(SO^{(i)}{}_1)=p_1; p\,(SO^{(i)}{}_2)=p_2; \ldots\ldots\ldots, p\,(SO^{(i)}{}_L)=p_L, \qquad (4.13)$$

де L – кількість видів семантичних одиниць і-го рівня абстрагування в структурі тезаурусу.

Отримання цієї функції розподілу уявляється вельми утрудненим завданням, оскільки необхідно мати доступ до статистичних даних щодо частостей використання усіх функцій, що складають профіль функціональності даної прикладної області (див. вираз (4.6)). Тим не менш, з теоретичної точки зору це є можливим.

Припустимо також, що у розпорядженні криптоаналітика знаходяться крипто-семантичні засоби захисту інформації ідентичні тим, якими користуються легальні користувачі, а також представницькій набір перехоплених шифрограм, що за будь-якими статистичними критеріями дає йому змогу отримати оцінку дискретної функції розподілу апостеріорних ймовірностей семантичних одиниць і-го рівня абстрагування на множині перехоплених шифрограм:

$$p^*SO^{(i)}{}_1)=p^*{}_1;\ p^*(SO^{(i)}{}_2)=p^*{}_2;\ \ldots\ldots\ldots\ldots..,\ p^*(SO^{(i)}{}_L)=p^*{}_L.. \quad (4.14)$$

Припустимо, що криптоаналітик виконав частотний аналіз ГПВП і у результаті побудував дискретну функцію розподілу ймовірностей появи конкретних значень псевдовипадкових послідовностей в межах заданого розмаху ГПВП (величина розмаху *K* дорівнює кількості семантично правдоподібних одиниць і-го рівня абстрагування у заданому тезаурусі прикладної області) у вигляді:

$$P(VP_1)=P_1;\ P(VP_2)=P_2;\ \ldots\ldots\ldots\ \ldots,\ P(VP_K)=P_K. \quad (4.15)$$

З урахуванням формули Байєса [40, 41] можливо записати наступну систему рівнянь:

$$\begin{aligned}
&p^*SO^{(i)}{}_1)= p\,(SO^{(i)}{}_1)\,P(VP_1)+ p\,(SO^{(i)}{}_1)\,P(VP_2)+ \ldots. + p\,(SO^{(i)}{}_1)\,P(VP_K);\\
&p^*(SO^{(i)}{}_2)= p\,(SO^{(i)}{}_2)P(VP_1)+ p\,(SO^{(i)}{}_2)\,P(VP_2)+ \ldots. + p\,(SO^{(i)}{}_2)\,P(VP_K);\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (4.16)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&p^*(SO^{(i)}{}_L)= p\,(SO^{(i)}{}_L)\,P(VP_1)+ p\,(SO^{(i)}{}_L)\,P(VP_2)+ \ldots.+ p\,(SO^{(i)}{}_L)\,P(VP_K).
\end{aligned}$$

Аналіз виразу (4.16) показує, що у даному випадку частотний аналіз міг би мати успіх, якщо б існували умови, за яких права частина хоча б одного із рівнянь (4.16) дорівнювала 1. Таке могло статися тільки у разі одночасної втрати ознак випадковості як у послідовностях смислових образів, що генерує джерело інформації, так і в послідовностях, що генерує ГПВП. Однак за таких умов ймовірнісний аналіз втрачає сенс.

4.4. Базова схема реалізації крипто-семантичного методу побудови системи захисту

Розглянемо основний (базовий) варіант технічної реалізації крипто-семантичного методу забезпечення абсолютної конфіденційності смислу

текстових повідомлень. Можливі й інші варіанти реалізації цього методу захисту.

Вважається, що текстові повідомлення можуть представлятися на будь-якій із мов спілкування (природній людській або формально визначеній штучній), для якої створена відповідна лексикографічна система [29] (зокрема, так званий лінгвістичний корпус [28]), у складі котрого розроблено прикладний семантичний тезаурус (тобто, відповідним чином структурована система семантичних словників), що повністю охоплює предметну область використання даного методу захисту.

Найбільш поширеного застосування базовий варіант даного методу матиме у системах зберігання та передавання текстових повідомлень через відкриті канали зв'язку, що незахищені від перехоплення інформації, у випадках, коли порушення конфіденційності смислу цих повідомлень може призвести до неприйнятних негативних наслідків для її власників. Використання цього варіанту реалізації крипто-семантичного методу може виявитися безальтернативним технічним рішенням у прикладних задачах, де необхідно забезпечити абсолютну гарантованість захисту текстових повідомлень в умовах, коли відсутня довіра до будь-яких структур та суб'єктів.

У процесі розробки даного варіанту реалізації крипто-семантичного методу було поставлено завдання: за рахунок попередньої лінгвістичної обробки зразків текстових повідомлень, смисловий зміст котрих підлягає шифруванню, забезпечити можливість функціонування симетричної криптографічної системи захисту з будь-яким наперед визначеним формально обґрунтованим рівнем стійкості до крипто-аналітичних атак, а за умов (4.10) – (4.12) забезпечити абсолютну гарантію захисту смислу текстових повідомлень від порушень конфіденційності як під час їхнього зберігання у комп'ютерах, так і під час їхнього передавання через незахищені середовища транспортування інформації. При цьому абсолютна гарантованість захисту має забезпечуватися як з теоретичної, так і з

практичної точок зору, і не залежати від умови неперевищення відстані єдиності.

Поставлене завдання вирішується тим, що у рамках формально визначених обмежень обраної сфери прикладних застосувань розроблюють лексикографічну систему, яку реалізують у вигляді прикладного лінгвістичного корпусу, і далі безпосередньо перед шифруванням текстових повідомлень згідно з будь-яким обраним способом симетричної криптографії здійснюють семантичну структуризацію цієї інформації з використанням засобів побудованого лінгвістичного корпусу таким чином, щоб зразки зашифрованих текстових повідомлень представляли семантично правдоподібні семантичні образи у контексті обраної прикладної області, а будь-який результат застосування будь-якого із можливих способів криптоаналізу приводив до отримання нехай інших щодо смислу, але теж семантично правдоподібних текстових повідомлень.

Лінгвістичний корпус у даній роботі визначається як програмний засіб автоматичного розбиття потоку текстових повідомлень, смисловий зміст котрих підлягає захисту, на «мікроконтексти» – фрагменти потоку, які «групуються» навколо лінгвістичних одиниць (зокрема, слів, фраз, сценаріїв тощо), що є об’єктами тлумачення [28-30]. Під тезаурусом мови розуміється лексика мови із визначеними семантичними відношеннями між лінгвістичними одиницями. У даному випадку тезаурус – ієрархічна структура семантичних словників, що визначає семантичні відношення між лінгвістичними одиницями мови, що прийнята для відображення смислу текстових повідомлень заданої сфери застосувань.

Примітка 4.1. *Слова «сутність», «смисл», «смисловий зміст» та «суть змісту» вважаються синонімами.*

Стисло суть даного варіанту технічної реалізації крипто-семантичного методу захисту можна визначити як включення до симетричної криптографічної системи засобів лексикографічної системи таким чином, щоб у процесі шифрування забезпечувалась семантична неоднозначність

зашифрованих зразків текстових повідомлень. При цьому засоби лексикографічної системи, що використовуються, мають бути спроможними здійснювати граматичний та семантичний аналіз вихідного потоку текстових повідомлень у рамках заданої прикладної області.

Послідовність операцій з обробки текстових повідомлень, умови та результати такої обробки на кожному із технологічних етапів реалізації даного методу захисту показана на рис. 4.1 у вигляді діаграми форм представлення оброблюваного зразка текстового повідомлення та операторів перетворення цих форм засобами системи захисту.

В якості вихідних даних, що є необхідними для реалізації оператора S_F, окрім вихідних текстових повідомлень, використовують дані щодо структури потоку текстових повідомлень, їхніх лінгвістичних характеристик, зокрема правила граматики мови, що прийнята для відображення текстових повідомлень заданої сфери застосувань. Засоби відтворення цих даних складають так званий лінгвістичний корпус, з використанням котрого здійснюється маркування (розмітка) потоку текстових повідомлень за його лінгвістичними характеристиками [29]. Оператор S_F забезпечує коректність граматичної структури вихідних повідомлень, що створюються людиною або комп'ютерною програмою, відповідно до правил та обмежень, що задаються засобами лінгвістичного корпусу. У випадках невідповідності цим правилам голосові повідомлення повертаються їхнім створювачам на доопрацювання.

В якості вихідних даних, що є необхідними для реалізації оператора S_C, окрім граматично коректних вихідних текстових повідомлень, використовують семантичні конструкції задіяного тезаурусу мови. Оператор S_C забезпечує відповідність текстових повідомлень елементам тезаурусу, що гарантує потенціальну можливість шифрування смислу цих повідомлень.

Оператор шифрування S_Z здійснює заміну вихідного (скоригованого операторами S_F та S_C) зразка відкритого текстового повідомлення на лінгвістичну конструкцію, елементи котрої обрані випадковим чином із тезаурусу. Характер випадковості визначається властивостями генератора

псевдо випадкових чисел (ГПВП), що має функціонувати у складі задіяної схеми шифрування.

На рис. 4.2 відображена блок-схема базової моделі системи, принцип дії котрої відтворює крипто-семантичний метод захисту текстових повідомлень.

Дана система містить у своєму складі усі основні елементи симетричної криптографічної системи і забезпечує захист смислу текстових повідомлень за умови синхронізації генераторів псевдовипадкових послідовностей (ГПВП), що розташовані на передавальній та приймальній сторонах каналу секретного обміну інформацією. Розшифрування здійснюється з використанням звісного ключа шифру (паролю). Однак, окрім цього, у блок-схему моделі додатково включено засоби лінгвістичного корпусу у складі лінгвістичного аналізатора текстових повідомлень, тезаурусу смислових образів предметної області, у рамках якої планується використовувати систему захисту, та програмного комплексу маркування (розмітки) вихідних потоків текстових повідомлень. Лінгвістичний аналізатор контролює відповідність вихідних зразків текстових повідомлень граматичним та семантичним правилам та обмеженням, що прийняті у рамках заданої предметної області.

Тезаурус створюється за результатами статистичного та семантичного аналізів предметної області і повинен містити усі лінгвістичні одиниці, котрі потенційно можуть бути включені до складу будь-якого зразка (фрагмента, повідомлення) потоку текстових повідомлень, що має бути захищений від втрати конфіденційності.

Система функціонує наступним чином. Користувач прикладної системи формує вихідний потік відкритих голосових повідомлень, смисл котрих потребує захисту, та подає його на обробку засобами лінгвістичного корпусу. Засоби лінгвістичного корпусу здійснюють лексикографічну обробку потоку вихідних зразків голосових повідомлень у два етапи.

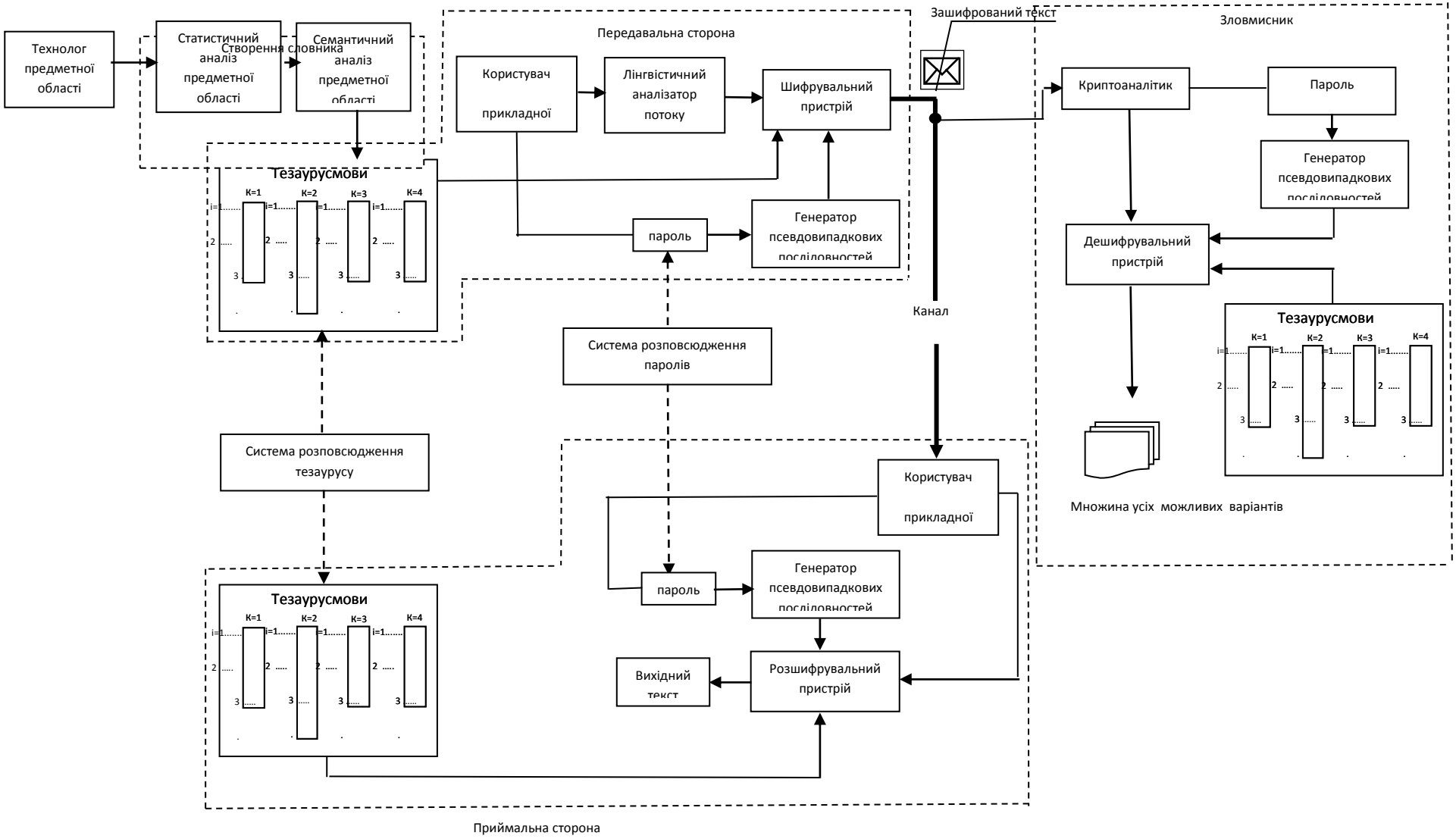

Рис. 4.2. Базовий варіант реалізації методу побудови лексикографічної криптосистеми

Спочатку, на першому етапі, здійснюється розмітка структури потоку голосових повідомлень за граматичними характеристиками та перевіряється його відповідність правилам граматики мови, що використовується. У випадках невідповідності цим правилам зразки голосових повідомлень повертаються на доопрацювання. Далі, на другому етапі, сформований відповідно до прийнятих граматичних правил потік голосових повідомлень подається на семантичний аналізатор, що перевіряє відповідність лінгвістичних конструкцій цього потоку елементам тезаурусу, що використовується. (Тобто, здійснює розмітку потоку щодо семантичних характеристик. Для маркування зразків лінгвістичних конструкцій у потоці голосових повідомлень може бути використана одна із відомих уніфікованих систем розмітки документів – SGML, HTML, XML тощо.). Якщо лінгвістична конструкція містить елементи, що не відображені у структурі тезаурусу, то вона повертається на доопрацювання. У протилежному випадку робиться висновок, що оброблена лінгвістична конструкція повністю відображає смисл вихідного голосового повідомлення, відповідає прийнятим граматичним правилам та семантичним обмеженням і є придатна для шифрування її смислового змісту.

Для захисту обробленої лінгвістичної конструкції від порушень конфіденційності її смислу користувач (або відповідна комп’ютерна програма) має подати символьну послідовність, що відображає цю конструкцію, на вхід шифрувального пристрою (або програмного шифратора) і увести ключову інформацію (пароль). За допомогою ключової інформації генератор псевдовипадкових послідовностей (ГПВП) ставиться у певний початковий стан. Шифратор опрацьовує символьні послідовності шляхом реалізації будь-якого відомого алгоритму шифрування, робота котрого 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 тезаурусу. Результатом роботи шифрувального пристрою є потік

зашифрованих відображень голосових повідомлень, якому притаманна властивість семантичної неоднозначності. У разі його нелегального перехоплення криптоаналітик не буде мати можливість визначити істинний смисл зразків вихідних голосових повідомлень, оскільки навіть при прямому переборі ключів шифру в умовах його повної обізнаності щодо задіяного лінгвістичного корпусу та прийнятої системи захисту він буде отримувати щоразу правдоподібні зразки голосових повідомлень (що складені із елементів тезаурусу), які із наперед заданою ймовірністю (див. вираз (4.8)) не відображають істинний смисл повідомлень. І 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, що був зашифрований, від інших правдоподібних зразків голосових повідомлень.

Зашифрований потік відображень голосових повідомлень може зберігатися у запам’ятовуючих пристроях або бути переданий через будь-яке незахищене фізичне середовище, наприклад, через радіоканал зв’язку. У будь-якому випадку зашифрований потік відображень голосових повідомлень може бути розшифрованим на будь-якому комп’ютері, на якому інстальовані відповідні засоби лінгвістичного корпусу та криптографічної системи захисту інформації. Щоб здійснити розшифрування, необхідно подати символьну послідовність, що відображає зашифрований потік, на розшифрувальний пристрій (або відповідну комп’ютерну програму розшифрування) та за допомогою звісного ключу шифру задати початковий стан ГПВП, що має бути ідентичним початковому стану ГПВП – того, що був під час шифрування вихідного потоку голосових повідомлень. Розшифратор опрацьовує символьні послідовності шляхом реалізації будь-якого відомого алгоритму розшифрування, робота котрого базується на використанні ГПВП. Суть розшифрування полягає у заміні лінгвістичних елементів зашифрованого потоку на елементи тезаурусу, що є ідентичними елементам вихідного потоку. Результатом роботи розшифрувального пристрою є потік

голосових повідомлень, що за смислом і формою його відображення є ідентичним потоку вихідних голосових повідомлень.

Область застосування цієї схеми обмежується умовами (4.10) – (4.12).

Дана схема не надає можливостей щодо схову самого факту використання крипто-семантичного шифрування.

4.5. Розробка програмного забезпечення для шифрування мовної інформації

Розроблене програмне забезпечення повинно здійснювати шифрування та відповідно розшифрування мовної інформації, що передається. Для реалізації даного прикладу було обрано використовувати прикладну область «Радіообмін диспетчер-пілот». Для реалізації поставленого завдання було вирішено використовувати мову програмування C# та вбудовані бібліотеки операційної системи Windows для розпізнавання мовної інформації.

В програмі було застосовано наступну відносно просту структуру тезаурусу радіообміну (табл. 4.1).

Таблиця 4.1

Тестовий тезаурус

Negative	Zero	Turn left	Request progressive taxi
Ready	One	Turn right	Straight ahead, second right.
	Two	Line up runway	Continue straight ahead.
	Three	Line up /and wait/	Turn left now
	Four	Cleared for take-off /report airborne/	Turn first right and report marshaller in sight
	Five	Cleared for take-off runway	First right. Marshaller in sight
	Six	Hold position, cancel i say again cancel take-off	Hold south of stand
	Seven	Stop immediately	Hold short of runway
	Eight	Cleared for take-off from	Request cross runway

	Nine		Hold position, traffic 1 kilometer final
	Ten		Expedite crossing runway
	Decimal		Are you ready for immediate departure
	Hundred		Climb straight ahead
	Thousand		Climb to reach flight level
			Descend to reach flight level

Принцип роботи програмного забезпечення детально описаний в попередньому підрозділі.

Одразу після запуску програми на екрані з’являються три вікна (рис. 4.3): вікно шифрування, вікно вибору режиму (шифрування/дешифрування) та вікно із зашифрованим/розшифрованим текстом.

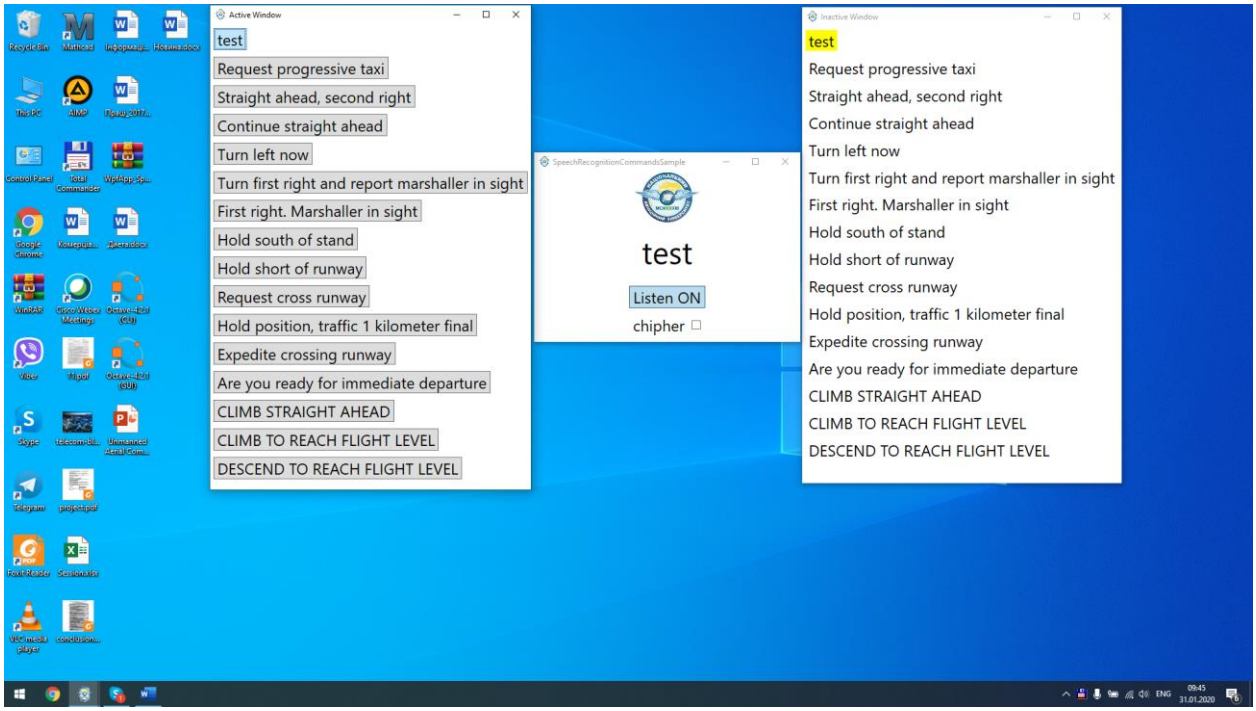

Рис. 4.3. Інтерфейс розробленого програмного забезпечення

Програмне забезпечення може використовувати для шифрування як текст, що вводиться через мікрофон, так і фрази, які обираються вручну із тезаурусу (рис. 4.4).

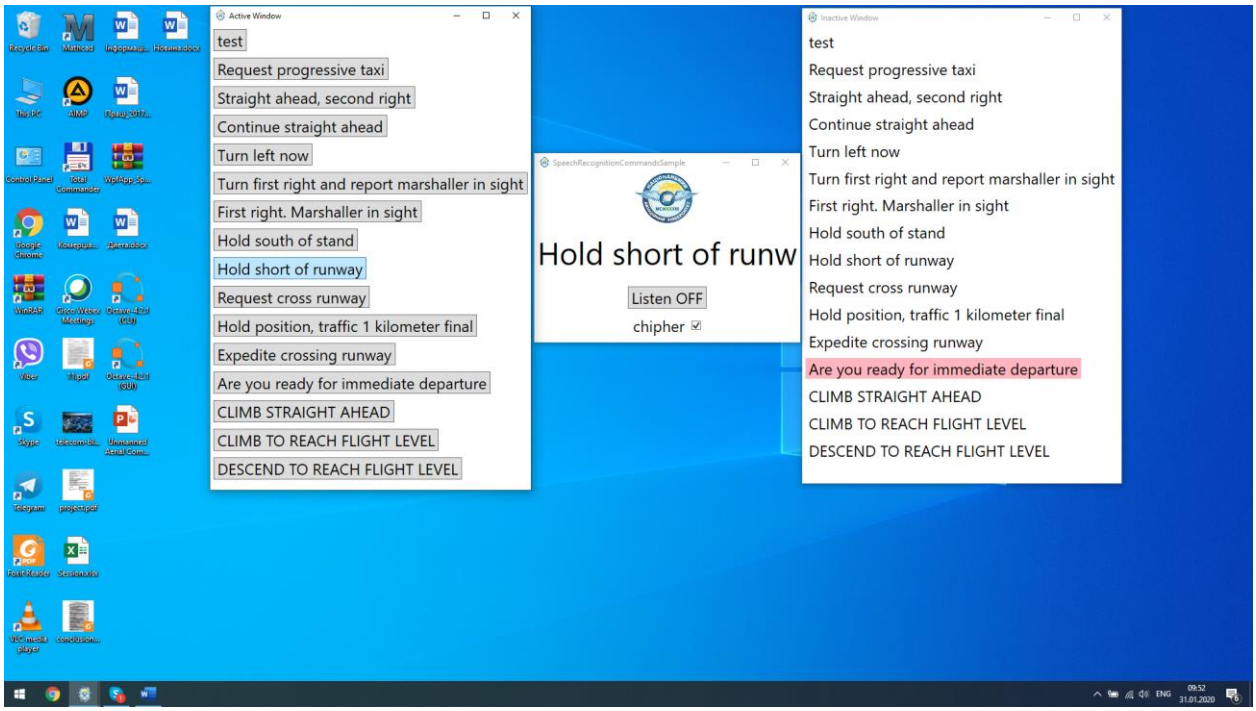

Рис. 4.4. Робота програмного забезпечення в ручному режимі

Після тривалого тестування розробленого додатку стало зрозуміло, що його можна використовувати на практиці. Покращити ефективність його роботи можна шляхом використання більш досконалих бібліотек для розпізнавання мовної інформації.

Висновки до четвертого розділу

1. Вперше на формальному рівні обґрунтовано можливість створення досконало стійких систем захисту смислового змісту текстових повідомлень, рівні стійкості котрих не залежать від довжини ключа шифру. Показано, що для забезпечення абсолютної конфіденційності повідомлень за умови, коли сумарні обсяги зашифрованої інформації перевищують довжину паролів, необхідно, щоб оператори шифрування/розшифрування системи захисту були залежними не тільки від вибору ключа шифру, але і від вибору елементу із простору смислових образів прикладної області. Названа умова може бути виконана, якщо задіяти результати теорії лексикографічних систем до забезпечення конфіденційності. Для отримання залежності

оператора шифрування від елементів простору смислових образів пропонується здійснювати заміну істинного смислового образу вихідного повідомлення (тобто, одного із елементів цього простору) на елемент семантичного тезаурусу даної прикладної області, смисловий образ котрого знаходиться у відношенні смислової правдоподібності із істинним смисловим образом вихідного повідомлення. За цих умов робота криптоаналітика втрачає сенс, оскільки він буде позбавлений можливості відрізнити істинний смисл зашифрованих повідомлень від багатьох інших фальшивих, але правдоподібних смислових образів.

2. Вперше на формальному рівні здійснено синтез методу побудови лексикографічної криптосистеми. Згідно даного методу простір заданої прикладної області застосування *i*-ої інформаційної системи відображається на простір смислових образів цієї системи з урахуванням семантичних співвідношень між ними, що визначаються заданим простором обмежувальних умов. Тобто, синтезується структура семантичного тезаурусу заданої прикладної області. У структурі тезаурусу кожному смисловому образу із простору смислових образів ставиться у відповідність множина інших смислових образів із цього ж простору, що знаходяться із цим смисловим образом у відношенні смислової правдоподібності. Така структура тезаурусу надає можливість під час шифрування замінювати повідомлення, що підлягають шифруванню, на інші правдоподібні повідомлення, що не відображають істинний смисл вихідних повідомлень. А під час розшифрування, якщо пароль є відомим, здійснювати зворотні заміни правдоподібних фальшивих повідомлень на істинні щодо смислу. Даний метод передбачає створення системи захисту у два етапи: спочатку створюється семантичний словник (тезаурус) прикладної області, а потім розроблюються програмно-технічні засоби реалізації операторів шифрування/розшифрування, у складі яких використовується створений тезаурус.

Показано також, що в деяких випадках довжина ключа шифру для крипто-семантичного шифрування має обиратися в залежності від припустимого значення ймовірності прийняття безпомилкових рішень в процесі дешифрування смислового потоку. Надано відповідний вираз для вибору мінімально можливого значення довжини ключа шифру.

Вперше синтезовано алгоритм реалізації запропонованого методу захисту смислу текстової інформації.

3. Вперше здійснено формальний, у т.ч. ймовірнісний, аналіз стійкості захисту смислу текстових повідомлень, що забезпечує лексикографічна криптосистема, яка побудована згідно запропонованого методу. Показано, що цей метод забезпечує абсолютно досконалу криптостійкість системи захисту за умов виконання обмежень (4.10) – (4.12), тобто елементи усіх можливих зразків текстових повідомлень мають знайти своє відображення у тезаурусі бази захисту текстової інформації, спеціально розробленого для відображення смислу текстових повідомлень заданої сфери застосувань.

Показано, що частотний аналіз міг би мати успіх тільки у разі одночасної втрати ознак випадковості як у послідовностях смислових образів, що генерує джерело інформації, так і в послідовностях, що генерує ГПВП. Однак за таких умов ймовірнісний аналіз втрачає сенс.

4. Вперше розроблено базовий крипто-семантичний метод, що реалізує запропонований спосіб забезпечення абсолютної стійкості системи захисту смислового змісту текстових повідомлень від порушень його конфіденційності.

Стисло суть даного методу захисту можна визначити як включення до симетричної системи криптографічного захисту інформації засобів лексикографічної системи таким чином, щоб у процесі шифрування забезпечувалась семантична неоднозначність зашифрованих зразків текстових повідомлень. При цьому засоби лексикографічної системи, що використовуються, мають бути спроможними здійснювати граматичний та семантичний аналіз вихідного потоку текстових повідомлень у рамках

заданої прикладної області. Область застосування обмежується умовами (4.10) – (4.12). Даний метод не надає можливостей щодо схову самого факту використання крипто-семантичного шифрування.

5. На базі розроблених в роботі методів та моделей було запропоновано алгоритмічне та програмне забезпечення, яке може бути використано для шифрування/розшифрування мовної інформації. Також було протестовано відповідний додаток.

ВИСНОВКИ

Сукупність наукових положень, сформульованих та обґрунтованих в дисертаційній роботі, складає вирішення науково-технічної задачі, яка полягала в необхідності підвищення ефективності шифрування текстових даних. У дисертаційній роботі отримані такі теоретичні та практичні результати.

1. В результаті проведеного аналізу було встановлено, що існуючі визначення і моделі, що використовують у задачах теорії інформації і зв'язку, не повною мірою придатні для опису систем семантичної криптографії. Для адекватного опису цих систем представлена модель осмислення суб'єктом змісту інформації стосовно завдань семантичної криптографії.

2. Розроблено модель семантичних тезаурусів, придатних для використання у складі стійкої криптосистеми захисту текстової інформації. А саме: уведено показники семантичних зв’язків між смисловими конструкціями мови відображення прикладної області, насамперед показник правдоподібності, і на цій основі здійснити синтез структури тезаурусів. Надано структуру тезауруса у розтині за координатою, що характеризує рівень абстрагування відображення смислових образів. Ця структура однозначно визначає місця розташування семантичних одиниць у тезаурусі, що дозволяє автоматизувати процес його створення.

3. Вперше розроблено метод побудови лексикографічної криптосистеми, який за рахунок розширення базового алфавіту лінгвістичної формальної системи джерела мовних повідомлень забезпечує збільшення відстані єдиності шифру за ключем та дозволяє збільшити довжину шифрованих повідомлень відносно довжини ключової інформації та відповідно зменшити частоту зміни ключів шифру у порівнянні із схемою Вернама. У наведеному прикладі маємо майже 50-ти кратний виграш за даним критерієм ефективності.

4. Вперше запропоновано лексикографічний метод захисту текстової інформації, що за рахунок випадкової заміни первинного смислового образу

повідомлення на інший правдоподібний елемент, узятий із семантичного тезаурусу бази захисту прикладної області, дозволяє забезпечити підвищену стійкість захисту при відсутності будь-яких обмежень на обсяг мовної інформації, що підлягає шифруванню, і, тим самим, усуває необхідність у періодичній зміні ключової інформації.

5. На базі розроблених методів та моделей було розроблене відповідно програмне забезпечення, яке надало змогу шифрувати/розшифровувати смислові образи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Смарт Н. Криптография / С.А. Кулешова (пер.с англ.). — М. : Техносфера, 2006. — 519 с.

2. Месси Дж. Л. Введение в современную криптологию. Малый тематический выпуск «Защита информации». ТИИЭР, 1988, № 5.

3. Encyclopedia of cryptography and security / ed.-in-chief Henk C. A. van Tilborg. New York : Springer , cop. 2005. - X 684.

4. Тил борг Ван Х.К.А. Основы криптологии. Профессиональное руководство и интерактивный ученик. – М.: Мир, 2006. – 471 с.

5. Грездов Г.Г. Современные методы защиты информации. Под ред. чл. – корр. НАН Украины Васильева В.В.- Киев, 2002. -32с. (Препринт / НАН Украины. Отделение гибридных моделирующих и управляющих систем в энергетике ИПМЭ; №01/2001).

6. Соломаа А. Криптография с открытым ключем: Пер. с англ. – М.: Мир, 1995. – 318 с.

7. Щербаков Л. Ю., Домашев А.В. Прикладная криптография. Использование и синтез криптографических интерфейсов. – М.: Издательско-торговый дом „Русская редакция”, 2003. – 416 с.: ил.

8. Самойлик Э.О. Ефективність досконало стійкої криптосистеми із збільшеною відстанню єдиності. Захист інформації, Том 19,№2, Квітень-червень, 2017. –С.184-192.

9. Сушко С.О., Кузнецов Г.В., Фомичова Л.Я., Корабльов А.В. Математичні основи криптоаналізу. – Д.: Національний гірничий університет, 2010. -465 с.

10. Заєць В.В., Чуприн В.М. Визначення стійкості криптостеганографічних методів захисту інформації // Збірник наукових праць Інституту проблем моделювання в енергетиці ім. Г.Є.Пухова. – Київ: ІПМЕ ім. Г.Є.Пухова НАН України, 2007. - № 44. – С. 9-19.

11. Заєць В.В., Чуприн В.М. Розрахунок ефективності криптосемантичної системи захисту інформації // Збірник наукових праць „Управління розвитком”. – Харків: ХНЕУ, 2008. – №6 – С. 68-70.

12. Хорошко В.А., Чекатков А.А. Методы и средства защиты информации / под общ. ред. Ковтанюка Ю.С. – К.: Юниор, 2003. -504 с.

13. Горбенко І.Д., Гріненко Т.О. Захист інформації в інформаційно-телекомунікаційних системах. –Х.: ХНУРЕ, 2004. -368 с.

14. *Shannon C.E.* A Mathematical Theory of Communication // Bell System Technical Journal, v. 27, n. 4, 1948, pp. 379-423, 623-656.

15. *Shannon C.E.* Communication Theory of Secrecy Systems // Bell System Technical Journal, v. 28, n. 4, 1949, pp. 379-423, 656-715.

16. *Shannon C.E.* Predication and Entropy in Printed English // Bell System Technical Journal, v. 30, n. 1, 1951, pp. 50-64.

17. Шеннон К.Э. «Теория связи в секретных системах». В кн.: Шеннон К.Э. Работы по теории информации и кибернетике. М.: Иностранная литература. 1963, с. 332-402, -829 с.

18. Бабенко Л.К., Мишустина Е.А. Методическое пособие по изучению современных методов криптоанализа. –Таганрог:ТРТУ, 2003. -66с.

19. Menezes A.J., Oorschot P.K., Vanstone S.A. The Handbook of Applied Cryptography. -New York: CRC Press, 1997. -816 p.

20. Шнайер Б. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си. –М.: Триумф, 2002. -916 с.

21. Schneier B., Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2nd ed. New York // John Wiley and Sons, 1996.

22. Математичні основи криптографії: навч. посібник / Кузнецов Г.В., Фомичов В.В., Сушко С.О., Фомичова Л.Я. – Д.: Національний гірничий університет, 2004. – 391 с.

23. Математичні основи крипто аналізу: навч. посібник / Сушко С.О.,Кузнецов Г.В., Фомичова Л.Я., Корабльов А.В. – Д.: Національний гірничий університет, 2010. -465 с.

24. Бабаш А.В., Шанкин Г.П. Криптография. –М.: СОЛОН-Р, 2002. -511с.

25. Харин Е.С., Берник В.И., Матвеев Г.В. Математические основы криптологии. – Минск:БГУ, 1999. -319 с.

26. Кригін М.Ю., Широков В.А. Дослідження інформаційно-статистичних властивостей українського тексту // Математичні машини і системи. -2000. -№1. –С.120-127.

27. Гундарь К.Ю. и др. Защита информации в компьютерных системах. Киев.: Корнейчук, 2000.-152 с.

28. Корпусна лінгвістика / В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна та ін.. – К.: Довіра, 2005. – 471 с.

29. Широков В,А. Феноменологія лексикографічних систем. – К.: Наукова думка, 2004. – 326 с.

30. Белявская Е.Г. Семантическая структура слова в номинативном и коммуникативном аспектах (когнитивные основания формирования и функционирования семантической структуры слова). М., 1992.

31. Кобрин Р. Ю. Лингвистическое описание терминологии как база концептуального моделирования в информационных системах. АДД. – Л., 1989.

32. Баранов А. Н. Введение в прикладную лингвистику: Учебное пособие. Изд. 2-е, испр. – М.: Едиториал УРСС, 2003. – 360 с.

33. Филлмор Ч. Об организации семантической информации в словаре // Новое в зарубежной лингвистике. – Вып. XIV: Проблемы и методы лексикографии. – М.: Прогресс, 1983.

34. Комарова З.И. Семантическая структура слова и ее лексикографическое описание. - Свердловск: Изд-во Уральск. ун-та, 1991.-156 с.

35. Иванов М.А., Чугунков И.В., Теория, применение и оценка качества генераторов псевдослучайных последовательностей. – М.: КУДНИЦ-ОБРАЗ, 2003. – 240 с.

36. D. Eastlake, S. Crocker, and J. Schiller, Randomness Recommendations for Security, RFC 1750, December 1994.

37. Barker E., Kelsey J. NIS Special Publication 800-90 / Recommendation for RandomNumber Generation Using Deterministic Random Bit Generators / Computer Security Division, Inforamtion Technology Laboratory.

38. C. Plumb Truly Random Numbers, Dr.Dobbs Journal, November 1994, p.113.

39. M.Jacobsson, E.Shriver, B.Hillyer and A.Juels. A Practical Secure Physical Random Bit Generator, 5th ACM Conference on Computer and Communications Security, 1998, p.103.

40. Вентецль Е.С. Теория вероятностей: Учеб. для вузов. – 5-е изд. сгер. – М.: высш. шк., 1998. – 576 с.: ил.

41. *Cover T.M., King R.C.* A Convergent Gambling Estimate of the Entropy of English, IEEE Transactions on Information Theory, v. IT-24, n. 4, Jul 1978, pp. 413-421.

42. Правила ведення радіотелефонного зв'язку та фразеологія радіообміну в повітряному просторі України. Наказ Державіаслужби №1328 від 02.092002 р.

43. Системы авиационной радио связи: Учеб. пособие / Под ред. В. А. Силякова; СПбГУАП. СПб., 2004. 160 с. ISBN 5-8088-0136-2.

44. Заєць В.В. Узагальнений крипто-семантичний алгоритм шифрування/розшифрування даних // Науково-Технічний журнал „Захист Інформації". – Київ: ДУІКТ, 2008. – № 2(38) – С. 39-42.

45. Заєць В.В., Чуприн В.М. Визначення стійкості криптостеганографічних методів захисту інформації // Збірник наукових праць Інституту проблем моделювання в енергетиці ім. Г.Є.Пухова. – Київ: ІПМЕ ім. Г.Є.Пухова НАН України, 2007. - № 44 – С. 9-19.

46. Заєць В.В., Чуприн В.М. Розрахунок ефективності криптосемантичної системи захисту інформації // Збірник наукових праць „Управління розвитком". – Харків: ХНЕУ, 2008. – №6 – С. 68-70.

47. Герасименко В.А., Размахнин М.К. Криптографические методы защиты информации. Зарубежная электроника. 1982, №5, с. 97-123.

48. Герасименко В.А. Размахни М.К. Программные средства защиты информации в вычислительных, информационных и управляющих системах и сетях. Зарубежная радиоэлектроника. 1986, №5, с. 73-91.

49. Люцарев В.С. и др. Безопасность компьютерных сетей на основе Windows NT. М.: Русская редакция, 1998. – 279 с.

50. Хоффман Л.Дж. Современные методы защиты информации. М.: Советское радио, 1980. – 264 с.

51. Use of taxonomy of security faults. COAST Laboratory, Purdue University, Technical report TR-96-051.

52. Landwehr C., Bull A., McDermott J. A taxonometry of computer security flaws, with examples. Information Technology Division, code 5542, Naval research laboratory, Washington D.C., 20375-5337.

53. Leveson N., Turner C.S. An investigation of the Therac-25 accidents. UCI TR92-108, Inf. And Comp. Sci. Dept., of Cal-Irvine, Irvine, CA.

54. Спесивцев, Вегнер и др. Защита информации в ПЭВМ. М.: Радио и связь, 1992, -192 с.

55. Ахо А., Хопрокфт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. М.: Мир, 1979. – 536 с.

56. Lempel A. Cryptography in transition. Computing surveys, vol. 11, pp. 285-303, Dec 1979.

57. Диффи У. Первые десять лет криптографии с открытым ключом. Малый тематический выпуск «Защита информации». ТИИЭР, 1988, № 5.

58. Merkle R.C. Secure communication over insecure channels. Comm ACM, pp. 294-299, Apr. 1978.

59. Герасименко В. А. Защита информации в автоматизированных системах обработки данных. Книги 1, 2. М.: Энергоатомиздат, 1994.-400 с. и 176 с.

60. Хорошко В.О., Азаров О.Д., Шелест М.Є., Яремчук Ю.Є. Основи комп’ютерної стеганографії: Навчальний посібник для студентів і аспірантів. – Вінниця: ВДТУ, 2003. – 143 с.

61. W. Bender, D. Gruhl, N. Morimoto, A. Lu, Technicues for Data Hiding. IBM Systems Journal, 35(3&4): pp. 313-336, 1996.

62. P.Bassia, I. Pitas, Robust Audio Watermarking In The Time Domain // Department of Informatics, University of Thessalonica (http://poseidon.csd.auth.gr/voyatis/creus.zip)

63. Husrev T. Sencar, Muhalingam Ramkumar, Ali N. Akansu, Data Hiding Fundamentals and Applications. Content Security In Digital Multimedia. ELSEVIER science and technology books, 2004. 364p.

64. Neil F. Johnson, Zoran Duric, Sushil Jajodia, Information Hiding: Steganography and Watermarking. – Attaks and Countermeasures. Kluwer Academic Puplishers. 2001. 160p.

65. S. Katzenbeisser, Fabien A. P. Petitcoals (Editors), Information Hiding Techniques for Steganography And Digital Watermark. Artech House Publishers. 1999. 220p.

66. Романовский И.В. Дискретный анализ. – 3-е изд., перераб. и доп. – СПб.: Невский Диалект; БХВ-Петербург, 2003. – 320 с.: ил.

67. Конахович Г.Ф., Пузиренко О.Ю. Комп'ютерна стеганографія. Теорія і практика. – К.: „МК-Пресс", 2006. – 288 с.: іл.

68. Закон України "Про інформацію" / Верховна Рада України. — Офіц. вид. — К. : Парламентське вид-во, 1997. — 32с.

69. Закон України «Про захист інформації в інформаційно-телекомунікаційних системах» № 80/94-ВР від 05.07.1994.

70. Закон України «Про науково-технічну інформацію» № 3322-XII від 25.06.93.

71. Закон України «Про державну таємницю» № 3855-XII від 21.01.1994.

72. ДСТУ 3396.0-96 – Захист інформації. Технічний захист інформації. Основні положення.

73. ДСТУ 3396.1-96 – Захист інформації. Технічний захист інформації. Порядок виконання робіт.

74. ДСТУ 3396.2-97 – Захист інформації. Технічний захист інформації. Терміни і визначення.

75. ДСТУ 4145-2002 Інформаційні технології. Криптографічний захист інформації. Цифровий підпис, що грунтується на еліптичних кривих. Формування та перевіряння.

76. НД ТЗІ 1.1–003–99 Термінологія в галузі захисту інформації у комп’ютерних системах від несанкціонованого доступу.

77. НД ТЗІ 1.1–002–99 Загальні положення щодо захисту інформації в комп’ютерних системах від несанкціонованого доступу.

78. НД ТЗІ 2.5–004–99 Критерії оцінки захищеності інформації в комп’ютерних системах від несанкціонованого доступу.

79. НД ТЗІ 2.5–005–99. Захист інформації. Технічний захист інформації. Класифікація автоматизованих систем і стандартні функціональні профілі захищеності оброблюваної інформації від несанкціонованого доступу.

80. НД ТЗІ 3.6-001-2000 Технічний захист інформації. Комп'ютерні системи. Порядок створення, впровадження, супроводження та модернізації засобів технічного захисту інформації від несанкціонованого доступу.

81. НД ТЗІ 1.4-001-2000 Типове положення про службу захисту інформації в автоматизованій системі.

82. НД ТЗІ 2.1-001-2001 Створення комплексів технічного захисту інформації. Атестація комплексів. Основні положення.

83. Наказ ДСТСЗІ СБУ № 76 «Про затвердження Порядку захисту державних інформаційних ресурсів в інформаційно-телекомунікаційних системах» від 24 грудня 2001 року.

84. Архангельский А.Я. Программирование в C++Builder 5. – М.: ЗАО «Издательство БИНОМ», 2000 г . – 1152 с.: ил.

85. Архангельский А.Я. Разработка прикладных программ для Windows в C++Builder 5. – М.: ЗАО «Издательство БИНОМ», 2000

86. Архангельский А.Я. Функции C++, C++Builder 5 и API Windows. – М.: ЗАО «Издательство БИНОМ», 2000

87. Архангельский А.Я. Работа с локальными базами данных в C++Builder 5 – М.: ЗАО «Издательство БИНОМ», 2000

88. Холингвэрт Д., Батерфилд Д., Сворт Б. и др. C++Builder 5. Руководство разработчика, том. 1. Основы : Пер. с англ.: Уч. пос. – М.: Издательский дом «Вильямс», 2001. – 880 с.: ил. – Парал. тит. англ.

89. Холингвэрт Д., Батерфилд Д., Сворт Б. и др. C++Builder 5. Руководство разработчика, том. 2. Сложные вопросы программирования: Пер. с англ. – М.: Издательский дом «Вильямс», 2001. – 832 с.: ил. – Парал. тит. англ.

90. Труб И.И. Объектно-ориентированное моделирование на C++: Учебный курс. – СПб.: Питер, 2006. – 411 с.: ил.

91. ANSI/ISO C++ Professional Programmer's Handbook. – Indianapolis: Macmillan Computer Publishing, 1999.

92. Хогланд Г., Мак-Гроу Г. Взлом программного обеспечения: анализ и использование кода.: Пер. с англ. – М.: Издательский дом «Вильямс», 2005. – 400 с.

93. Коблик Н. Курс теории чисел и криптографии – М.: Научное изд-во ТВП, 20001. – 254 с.

94. Коробейников А.Г. Математические основы криптографии. Учебное пособие. СПб: СПб ГИТМО (ТУ). 2002. – 41 с.

95. Гатчин Ю.А., Коробейников А.Г. Основы криптографических алгоритмов. Учебное пособие. СПб: СПб ГИТМО (ТУ). 2002. – 29 с.

96. Черемушки А.В. Лекции по арифметическим алгоритмам в криптографии. – М.: МЦНМО, 2002. – 104 с.

97. Словарь криптографических терминов / Под ред. Б.А. Погорелова и В.Н. Скачкова. – М.: МЦНМО, 2006. – 94 с.

98. Петров А.А. Компьютерная безопасность. Криптографические методы защиты. – М.: ДМК, 2000. – 448 с.: ил.

99. Макнамара Д. Секреты компьютерного шпионажа: Тактика и контрмеры / Д. Макнамара; Пер. с англ.; Под ред. С.М. Малявко. – М.: БИНОМ. Лаборатория знаний, 2004. – 536 с., ил.

100. Мак-Клар С., Шах С., Шах Ш. Хакинг в Web: атаки и защита: Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 384 с.: ил. – Парал. тит. англ.

101. Кузнецов И. Н. Информация: поиск, анализ, защита. – Минск: Амалфея, 2002. - 319 с.

102. Домарев В.В. Безопасность информационных технологий. Методология создания систем защиты М.: DiaSoft, 2002. - 671 с., ил.

103. Скляров, Д. В. Искусство защиты и взлома информации СПб.: БХВ-Петербург, 2004. - 276 с., ил.

104. Хорев П. Б. Методы и средства защиты информации в компьютерных системах. – М.: Академия, 2007. – 254 с., ил.

105. Белов Е.Б., Лось В.П., Мещеряков Р.В. и др. Основы информационной безопасности. Учебное пособие для вузов. М.: Горячая линия – Телеком, 2006. – 544 с.: ил.

106. Василенко О.Н. Теоретико-числовые алгоритмы в криптографии. – М.: МЦНМО, 2003. – 328 с.

107. Домарев В.В. Безопасность информационных технологий. Системный поход: - К.: ООО „ТИД ”ДС”, 2004. – 992 с.

108. Загородний В.И. Комплексная защита информациив компьютерных системах: Учебное пособие. – М.: Логос, 2001. – 264 с.: ил.

109. Малюк А.А. Информационная безопасность: концептуальне и методологические основы защиты информации. Учеб. пособие для вузов. – М.: Горячая линия-Телеком, 2004. – 280 с.

110. Иванов М.А., Чугунков И.В. Теория, применение и оценка качества генераторов псевдослучайных последовательностей. – М.: КУДНИЦ-ОБРАЗ, 2003. – 240 с.

111. Gutmann P. Software Generation of Practically Strong Random Numbers Department of Computer Science University of Auckland

112. D. Eastlake, S. Crocker, and J. Schiller, Randomness Recommendations for Security, RFC 1750, December 1994.

113. Barker E., Kelsey J. NIST Special Publication 800-90 / Recommendation for RandomNumber Generation Using Deterministic Random Bit Generators / Computer Security Division, Inforamtion Technology Laboratory

114. C. Plumb Truly Random Numbers, Dr.Dobbs Journal, November 1994, p.113.

115. M.Jacobsson, E.Shriver, B.Hillyer and A.Juels A Practical Secure Physical Random Bit Generator, 5th ACM Conference on Computer and Communications Security, 1998, p.103.

116. Филлмор Ч. Об организации семантической информации в словаре // Новое в зарубежной лингвистике. – Вып. XIV: Проблемы и методы лексикографии. – М.: Прогресс, 1983.

117. Комарова З.И. Семантическая структура специального слова и ее лексикографическое описание. - Свердловск: Изд-во Уральск. ун-та,1991.-156 с.

118. Белявская Е.Г. Семантическая структура слова в номинативном и коммуникативном аспектах (когнитивные основания формирования и функционирования семантической структуры слова). М., 1992.

119. Кобрин Р. Ю. Лингвистическое описание терминологии как база концептуального моделирования в информационных системах. АДД. – Л., 1989.

120. Баранов А. Н. Введение в прикладную лингвистику: Учебное пособие. Изд. 2-е, испр. – М.: Едиториал УРСС, 2003. – 360 с.

121. Зубов А.Ю. Совершенные шифры: Вступительное слово чл.-корр. РАН Б.А. Севастьянова. – М.: Гелиос АРВ, 2003. – 160 с.

122. Алферов А.П., Зубов А.Ю., Кузьмин А.С., Черемушкин А.В. Основы криптографии. – М.: «Гелиос АРВ», 2001. – 479 с.

ДОДАТОК А
ЛІСТИНГ ПРОГРАМНОГО КОДУ ДО РОЗРОБЛЕНОГО ДОДАТКУ

Код головного вікна на мові XAML:

```
<Window x:Class="WpfApp_SpeechRecognitionTextCommands.MainWindow"
      xmlns="http://schemas.microsoft.com/winfx/2006/xaml/presentation"
      xmlns:x="http://schemas.microsoft.com/winfx/2006/xaml"
      xmlns:d="http://schemas.microsoft.com/expression/blend/2008"
      xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006"
      xmlns:local="clr-namespace:WpfApp_SpeechRecognitionTextCommands"
      mc:Ignorable="d"
      Title="SpeechRecognitionCommandsSample" Height="300" Width="425"
      Closing="Window_Closing" FontSize="24"
WindowStartupLocation="CenterScreen" Icon="Res/nau_logo.ico">
   <DockPanel>
      <WrapPanel DockPanel.Dock="Bottom" HorizontalAlignment="Center"
Margin="0,0,0,10">
         <TextBlock  VerticalAlignment="Center">chipher</TextBlock>
         <CheckBox x:Name="cb_chipher" IsChecked="True"
VerticalAlignment="Center" Margin="10, 0, 0, 0"></CheckBox>
      </WrapPanel>
      <WrapPanel DockPanel.Dock="Bottom" HorizontalAlignment="Center"
Margin="0,10">
         <ToggleButton Name="btnToggleListening"
Click="btnToggleListening_Click" Padding="5,0">Listen</ToggleButton>
         <Button Visibility="Collapsed" x:Name="bt_speek" Content="Speaking"
Margin="5,0" Click="bt_speek_Click">
            <Button.ContextMenu>
               <ContextMenu>
                  <MenuItem Header="Menu item 1" />
               </ContextMenu>
            </Button.ContextMenu>
         </Button>
      </WrapPanel>

      <Image DockPanel.Dock="Top" Width="100"
Source="Res/nau_logo.ico"></Image>
      <Label  Name="lblDemo" HorizontalAlignment="Stretch"
HorizontalContentAlignment="Center" VerticalAlignment="Center"
FontSize="48" >
         <TextBlock TextWrapping="WrapWithOverflow">
            Hello, board!
```

```
            </TextBlock>
        </Label>
    </DockPanel>
</Window>
```

Код вікна Speaking на мові XAML:

```
<Window
x:Class="WpfApp_SpeechRecognitionTextCommands.Window_speaking"
        xmlns="http://schemas.microsoft.com/winfx/2006/xaml/presentation"
        xmlns:x="http://schemas.microsoft.com/winfx/2006/xaml"
        xmlns:d="http://schemas.microsoft.com/expression/blend/2008"
        xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006"
        xmlns:local="clr-namespace:WpfApp_SpeechRecognitionTextCommands"
        mc:Ignorable="d"
        Title="Window_speaking" Height="450" Width="800"
Activated="Window_Activated" Deactivated="Window_Deactivated"
FontSize="24" SizeToContent="WidthAndHeight" Loaded="Window_Loaded">
    <Window.Resources>
        <ContextMenu x:Key="cmButton">
            <MenuItem Header="Menu item 1" />
            <MenuItem Header="Menu item 2" />
            <Separator />
            <MenuItem Header="Menu item 3" />
        </ContextMenu>
    </Window.Resources>
    <StackPanel x:Name="myStackPanel" >

    </StackPanel>
</Window>
```

Код головного вікна на мові C#:

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows;
using System.Windows.Controls;
using System.Windows.Data;
using System.Windows.Documents;
using System.Windows.Input;
using System.Windows.Media;
```

```
using System.Windows.Media.Imaging;
using System.Windows.Navigation;
using System.Windows.Shapes;

using System.Speech.Recognition;
using System.Speech.Synthesis;

using System.Globalization;

namespace WpfApp_SpeechRecognitionTextCommands
{
   /// <summary>
   /// Interaction logic for MainWindow.xaml
   /// </summary>
   public partial class MainWindow : Window
   {
      System.Globalization.CultureInfo ci = new
System.Globalization.CultureInfo("en-US");
      private SpeechRecognitionEngine speechRecognizer = new
SpeechRecognitionEngine();
      private SpeechSynthesizer sythesizer = new SpeechSynthesizer();

      Window_speaking window2;
         Window_Text window3;

      public delegate void AccountHandler();
      public event AccountHandler Notify;

      public MainWindow()
      {
         InitializeComponent();
         //System.Globalization.CultureInfo ci = new
System.Globalization.CultureInfo("En-en");
         speechRecognizer = new SpeechRecognitionEngine(ci);



         speechRecognizer.SpeechRecognized +=
speechRecognizer_SpeechRecognized;

         GrammarBuilder grammarBuilder = new GrammarBuilder();
         //Choices commandChoices = new Choices(
         //    "Request progressive taxi",
         //    "Straight ahead, second right",
```

```
    //    "Continue straight ahead",
    //    "Turn left now"
    //    );
    Choices commandChoices = new Choices(myWords.arrWords);
    //Choices commandChoices = new Choices(
    //    "test"
    //    );
    grammarBuilder.Append(commandChoices);

    //Choices valueChoices = new Choices();
    //valueChoices.Add("left", "right");
    //valueChoices.Add("runway", "wait");
    //valueChoices.Add("foward", "on", "for");
    //valueChoices.Add("position", "cancel", "traffic");
    //grammarBuilder.Append(valueChoices);

    speechRecognizer.LoadGrammar(new Grammar(grammarBuilder));
    speechRecognizer.SetInputToDefaultAudioDevice();

    //Window_speaking window2 = new Window_speaking();
    window2 = new Window_speaking(ref lblDemo, this);
    window2.Show();
    //Window_Text window3 = new Window_Text();
    window3 = new Window_Text(ref lblDemo, this);
    Notify += window3.DisplayMessage;
    window3.Show();
    Application.Current.MainWindow.Activate();
}

//public string DisplayMessage()
//{
//    string str = lblDemo.Content.ToString();
//    return str;
//}

private void btnToggleListening_Click(object sender, RoutedEventArgs e)
{
    //lblDemo.Content = "1";
    Notify?.Invoke(); //$"Событие"
    if (btnToggleListening.IsChecked == true)
    {
        speechRecognizer.RecognizeAsync(RecognizeMode.Multiple);
        btnToggleListening.Content = "Listen ON";
    }
```

```
            else
            {
                speechRecognizer.RecognizeAsyncStop();
                btnToggleListening.Content = "Listen OFF";
            }
        }

        public void speechRecognizer_SpeechRecognized(object sender,
SpeechRecognizedEventArgs e)
        {
            if (e != null) lblDemo.Content = e.Result.Text;
            Notify?.Invoke(); //$"Событие"
            #region if (e.Result.Words.Count == 2)
            //if (e.Result.Words.Count == 2)
            //{
            //    string command = e.Result.Words[0].Text.ToLower();
            //    string value = e.Result.Words[1].Text.ToLower();
            //lblDemo.Content = command + " " + value;

            //switch (command)
            //{
            //    case "turn":
            //        switch (value)
            //        {
            //            case "left":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //            case "right":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //        }
            //        break;
            //    case "line up":
            //        switch (value)
            //        {
            //            case "runway":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //            case "and wait":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //        }
            //        break;
            //    case "cleared for take-off":
```

```
            //        switch (value)
            //        {
            //            case "report":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //            case "runway":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //            case "from":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //        }
            //        break;
            //    case "hold":
            //        switch (value)
            //        {
            //            case "position":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //            case "cancel":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //            case "traffic":
            //                lblDemo.Content = command + " " + value;
            //                break;
            //        }
            //        break;
            //}

            //SpeechSynthesizer sythesizer = new SpeechSynthesizer();
            //sythesizer.Speak(command + " " + value);
            //}
            #endregion
        }

        private void Window_Closing(object sender,
System.ComponentModel.CancelEventArgs e)
        {
            speechRecognizer.Dispose();
            foreach (Window item in Application.Current.Windows)
            {
                if (item != this)
                item.Close();
            }
```

```
        }

        private void bt_speek_Click(object sender, RoutedEventArgs e)
        {
            Button myBut = (Button)sender;
            string str = myBut.Content.ToString();
            //SpeechSynthesizer sythesizer = new SpeechSynthesizer();
            sythesizer.Speak(str);
        }


    }
}
```

ДОДАТОК Б
АКТИ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ

Висвітлені в дисертації наукові результати отримано, здебільшого, в рамках науково-дослідних робіт, які були виконані в Національному авіаційного університеті (НАУ), а також на ДП «Антонов»: шифр 874-ДБ13, тема «Створення та дослідження нових систем захищеного авіаційного радіозв'язку в рамках Концепції CNS/ATM ICAO» (НДР 0110U000225). Отримані результати використовуються у навчальному процесі НАУ. Використання результатів дисертаційної роботи підтверджено відповідними актами впровадження.

Надалі в даному додатку наведені результати впровадження напрацювань, отриманих в рамках даної дисертаційної роботи.

ЗАТВЕРДЖУЮ:

Головний конструктор
кандидат технічних наук

_____ О.З. Двейрін

«___»_____ 2020 г.

АКТ

Цим актом підтверджується, що результати дисертаційної роботи працівника ДП «АНТОНОВ» Самойлика Євгена Олександровича «Лексикографічні методи захисту мовної інформації» можуть бути використані при створенні обладнання нових типів літаків та модернізації літаків для виконання спеціальних завдань в інтересах МО України.

В роботі запропонована розробка методів, що забезпечують підвищення ефективності шифрування мовної інформації при обміні інформацією через відкриті канали зв'язку, виконана на основі результатів теорії випадкових процесів, теорії телетрафіка, теорії криптографічних систем та теорії лексикографічних систем. В основу роботи покладена теорія секретного зв'язку К. Шеннона. Використано результати теорії лінгвістичних корпусів, що відображені, зокрема, у праці Широкова В.А. «Корпусна лінгвістика». При час натурного моделювання запропонованих криптосистем використано сучасні методи комп'ютерного моделювання.

Методика дає підвищення ефективності шифрування мовної інформації за рахунок використання особливостей семантичних характеристик мовної інформації.

Робота виконувалась на ДП «Антонов» та Національному авіаційному університеті.

Головний конструктор
кандидат технічних наук Г.І. Рудюк

Заступник Головного конструктора О.Ю. Курганський

П О Г О Д Ж У Ю

Проректор Національного авіаційного університету з навчальної роботи

_____ А.Г. Гудманян

«10» 06 2020 р.

З А Т В Е Р Д Ж У Ю

Проректор Національного авіаційного університету з наукової роботи

_____ В.П. Харченко

«___» _____ 2020 р.

АКТ ВПРОВАДЖЕННЯ

результатів дисертаційної роботи здобувача наукового ступеня кандидата технічних наук Самойлика Євгена Олександровича у навчальний процес Національного авіаційного університету

Ми, що нижче підписалися, *завідувач кафедри телекомунікаційних та радіоелектронних систем Факультету аеронавігації, електроніки та телекомунікацій, д.т.н., проф. Г.Ф. Конахович, декан Факультету аеронавігації, електроніки та телекомунікацій Національного авіаційного університету д.т.н., проф. І.О. Мачалін*
склали цей акт про те, що результати наукових досліджень *за темою дисертаційної роботи на здобуття наукового ступеня кандидата технічних наук Самойлика Євгена Олександровича*
використовуються у навчальному процесі *на кафедрі телекомунікаційних та радіоелектронних систем Факультету аеронавігації, електроніки та телекомунікацій Національного авіаційного університету.*

Найменування впровадженого результату	Форма впровадження і досягнутий фактичний ефект
Алгоритмічне та програмне забезпечення крипто-семантичного захисту текстових даних	*У навчальний процес як методичні рекомендації для проведення лабораторних занять з дисциплін «Безпека інформаційно-комунікаційних систем». Завдяки своїй наочності впроваджені методичні рекомендації підвищили ефективність засвоєння студентами матеріалу зазначеної дисципліни, зокрема під час вивчення новітніх методів захисту інформації*
Модель та метод побудови крипто-семантичного словника для прикладної галузі застосування	*У навчальний процес як методичні матеріали для проведення курсового проектування з дисципліни «Захист інформації в ТКС». Впроваджений метод дозволив підвищили якість підготовки студентів, що навчаються за спеціальністю 172 «Телекомунікації та радіотехніка», освітньою програмою «Телекомунікаційні системи та мережі» шляхом набуття ними знань про методи лексикографічного захисту інформації.*

Завідувач кафедри телекомунікаційних та радіоелектроннихсистем, д.т.н., проф. _____ Г.Ф. Конахович

Декан Факультету аеронавігації, електроніки та телекомунікацій д.т.н., проф. _____ І.О. Мачалін