УДК 003.26:004.056.55 **DOI:** 10.18372/2310-5461.39.13092

Р. С. Одарченко, канд. техн. наук, доц.
Національний авіаційний університет
orcid.org/0000-0002-7130-1375
e-mail: odarchenko.r.s@ukr.net

Є. О. Самойлик
Національний авіаційний університет
orcid.org/0000-0003-4090-8681
e-mail: sea110913@gmail.com

А. О. Абакумова, аспірант
Національний авіаційний університет
orcid.org/0000-0002-2844-096X
e-mail: nastia.abakumova@gmail.com

МЕТОД ПОБУДОВИ СЕМАНТИЧНОГО СЛОВНИКА У СКЛАДІ ДОСКОНАЛО СТІЙКОЇ КРИПТОСИСТЕМИ ЗАХИСТУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Вступ

Розвиток нових інформаційних технологій і впровадження комп'ютерних систем в усі сфери людської діяльності стали причиною різкого зростання зацікавленості широкого кола користувачів до проблеми інформаційного захисту. Провідна роль у забезпеченні інформаційної безпеки в інформаційно-телекомунікаційних системах відводиться криптографії, одним із головних завдань якої є: забезпечення конфіденційності, цілісність та автентичність даних, що передаються. Зважаючи на це в рамках сьогодення питання використання, удосконалення та розроблення досконало стійких криптографічних системи (криптосистем), які забезпечують різноманітний рівень криптостійкості, є досить актуальним.

Аналіз останніх досліджень і публікацій

На сьогодні багатьма вченими 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]. Тим не менш, ці криптосистеми не гарантують формальну, теоретично доведену неможливість їхнього злому [4–7]. Згідно з цим існує проблема недовіри до надійності цих систем в задачах передавання інформації, що характеризуються високими рівнями секретності.

Водночас численні дослідження багатьох авторів указують на відносно невеликі значення відстані єдиності при шифруванні повідомлень, складених із символів алфавіту будь-якої із природних мов [8; 9]. Відтак це призводить до необхідності частої зміни ключової інформації, що є проблемою для багатьох застосувань.

Постановка завдання

З урахуванням вище зазначеного можна стверджувати, що створення нових методів побудови досконало стійких криптосистем, які 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.

За таких умов для вирішення даного завдання в першу чергу необхідно розробити метод побудови семантичного тезаурусу, придатного для використання у складі досконало стійкої криптосистеми захисту текстової інформації. А саме: увести показники семантичних зв'язків між смисловими конструкціями мови відображення прикладної області, насамперед показники правдоподібності, і на цій основі здійснити синтез структури тезаурусу.

Структура тезаурусу смислових образів

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.

В першу чергу 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, … , 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 (рис. 1).

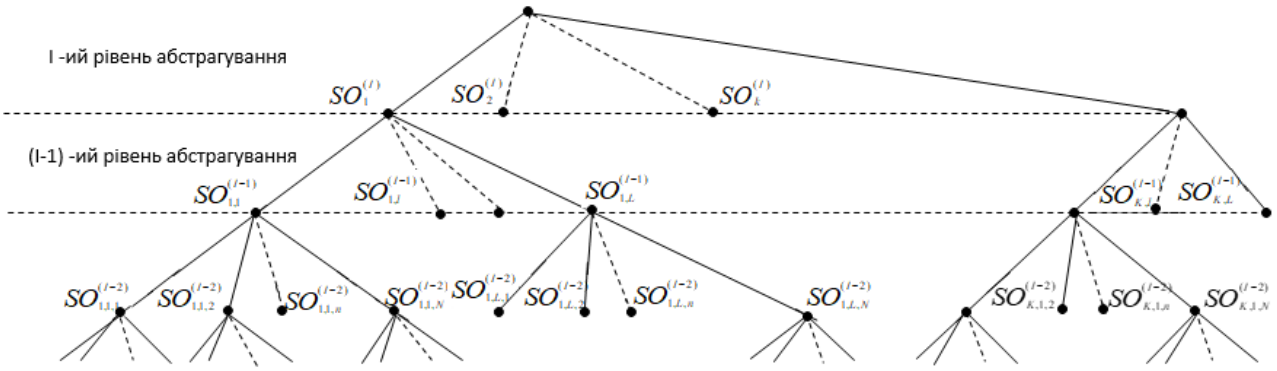

Рис. 1. Прошаркова коренево-подібна структура простору смислових образів, що складають тезаурус

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, …, I. При цьому підтезауруси у складі TZ розташовані прошарками у вигляді гілкоподібної кореневої системи, що розростаються зверху вниз. В основі кореня лежить семантичний образ $SO^{(I+1)}$ з максимальним (I+1)-м рівнем абстрагування відображення SO, що є доступним інтелекту у конкретній області розумової діяльності.

У загальному випадку структуру тезаурусу TZ можливо представити у вигляді рекурентної коренево-подібної схеми як

$$TZ \in \{TZ^{(I)}_k\}, \quad k = \overline{1, K},$$

$$\text{де } TZ^{(I)}_k \in \{TZ^{(I-1)}_{k,l}\}, \quad l = \overline{1, L}, \text{ де } TZ^{(I-1)}_{k,l} \in \{TZ^{(I-2)}_{k,l,n}\}, \quad n = \overline{1, N},$$

$$\text{де } TZ^{(2)}_{k,l,n,\dots,p} \in \{SO^{(1)}_{k,l,n,\dots,j}\}, \quad j = \overline{1, J}. \qquad (1)$$

У виразі (1) прийняті такі 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; $k \in \{1, 2, \dots, K\}$ — порядковий номер елемента $TZ^{(I)}_k$ у множині тезаурусів, що у сукупності визначають простір смислових образів I-го рівня абстрагування; 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 $l \in \{1, 2, \dots, L\}$ — порядковий номер тезаурусу (I-1)-го рівня абстрагування у множині тезаурусів, які у сукупності визначають простір смислових образів (I-1)-го рівню 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 $n \in \{1, 2, \dots, N\}$ — 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 $TZ^{(I-1)}_{k,l}$; $TZ^{(2)}_{k,l,n,\dots,p}$ — тезаурус другого рівня абстрагування представлення SO, що конкретизує смислові образи $TZ^{(3)}_{k,l,n,\dots,s}$, де $p \in \{1, 2, \dots, P\}$ — порядковий номер тезаурусу другого рівня абстрагування у складі тезаурусу третього рівня $TZ^{(3)}_{k,l,n,\dots,s}$, де $s \in \{1, 2, \dots, S\}$ — порядковий номер тезаурусу третього рівня абстрагування; P — кількість тезаурусів другого рівня абстрагування, що входять до складу $TZ^{(3)}_{k,l,n,\dots,s}$; S — кількість тезаурусів третього рівня абстрагування, що входять до складу відповідного тезаурусу четвертого рівня абстрактності і т. д. уздовж ланцюгу тезаурусів із зростанням значення індексу i; $SO^{(1)}_{k,l,n,\dots,j}$ — семантичний словник, що відображає тезаурус $TZ^{(2)}_{k,l,n,\dots,p}$; J — кількість слів у тезаурусі $TZ^{(2)}_{k,l,n,\dots,j}$.

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

Якщо розглядати структуру тезаурусів мови відображення будь-якої прикладної області відповідно до розтину за рівнями абстрагування відображення смислових образів (рис. 1), то у загальному випадку доцільно задати таку ієрархію її семантичних одиниць:

прикладна область (напрямок знань)/ тема/сценарій/ситуація/ фраза/слово/символ алфавіту/ код символу алфавіту (2)

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

сценарій/ситуація/фраза/слово (3)

У цьому випадку, у виразі (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 (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 повною мірою відображав визначену область активності. Однак створення такого простору через його велику розмірність потребує занадто великих ресурсних витрат. На практиці область активності обмежується визначеною сферою прикладних застосувань.

Формальне відображення семантики текстової інформації

Відобразимо зразок текстової послідовності на рівні семантичних одиниць i-го рівню абстрагування так:

$$SO^{(i)}{}_{k(1)}, SO^{(i)}{}_{k(2)},\ldots, SO^{(i)}{}_{k(n)},\ldots, SO^{(i)}{}_{k(Ni)}, \quad (4)$$

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 підчас 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 (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 $\equiv$ означає відношення смислової ідентичності.

У свою чергу, кожен елемент випадкової послідовності (4) може бути представлений послідовністю більш конкретного змісту — на рівні семантичних одиниць (i-1)-го рівня абстрагування:

$$SO^{(i)}{}_k \rightarrow SO^{(i-1)}{}_{k,\,l(1)}, SO^{(i-1)}{}_{k,\,l(2)}, \ldots, SO^{(i-1)}{}_{k,\,l(Nl)}, \quad (5)$$

а кожен елемент випадкової послідовності (5) має бути представлений послідовністю ще більш конкретного змісту — на рівні семантичних одиниць (i-2)-го рівня абстрагування:

$$SO^{(i-1)}{}_{k,l} \rightarrow SO^{(i-2)}{}_{k,l,m(1)}, SO^{(i-2)}{}_{k,l,m(2)},\ldots, SO^{(i-2)}{}_{k,l,m(Nm)}. \quad (6)$$

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

$$SO^{(2)}{}_{k,l,n,\ldots,p} \rightarrow SO^{(1)}{}_{k,l,n,\ldots,j\,(1)}, SO^{(1)}{}_{k,l,n,\ldots,j\,(2)}, \ldots, SO^{(1)}{}_{k,l,n,\ldots,j\,(Nj)}. \quad (7)$$

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.

Смислові відношення між семантичними одиницями

Метод побудови досконало стійкої криптосистеми базується на використанні лексикографічних ефектів, що пов'язані з 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 крипто-аналітиком 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 криптосемантичну систему захисту, необхідно надати формальні визначення таким смисловим відно-

шенням між семантичними одиницями як *смислова ідентичність*, *смислова відмінність*, сумнівна смислова ідентичність (або *правдоподібність*), сумнівна смислова відмінність (або *фальшиво-подібність*).

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 самому тезаурусу мови, що відображає певним чином обрану область активності;

2) смислові відношення визначаються між семантичними одиницями будь-якого, але однакового рівню 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

$$TZ_a^{(i)}{}_k \equiv TZ_b^{(i)}{}_k, \quad (8)$$

якщо є ідентичними за сутністю змісту.

Позначка → означає приналежність будь-якого тезаурусу $TZ^{(i)}{}_k$ із множини тезаурусів *i*-го рівня абстрагування до тезаурусу обраної мови *TZ*, а у виразі (8) позначка ≡ означає відношення смислової ідентичності.

Зрозуміло, якщо

$$TZ_a^{(i)}{}_k \equiv TZ_b^{(i)}{}_k, \text{ а } TZ_b^{(i)}{}_k \equiv TZ_c^{(i)}{}_k, \text{ то } TZ_a^{(i)}{}_k \equiv TZ_c^{(i)}{}_k. \quad (9)$$

Зрозуміло також, що

$$SO^{(i)}{}_k \equiv TZ^{(i)}{}_k \text{ для будь-яких значень } i \text{ та } k \quad (10)$$

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 рівню 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

$$TZ_a^{(i)}{}_k \text{ ⌂ } TZ_b^{(i)}{}_l, \quad (11)$$

якщо не є ідентичними за сутністю змісту. Позначка ⌂ означає смислову відмінність.

Зрозуміло, що різні гілки будь-якого, але одного рівня абстрагування представлення *SO* у корені, що представляє структуру *TZ* (рис. 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

$$TZ_a^{(i)}{}_k \text{ ^ } TZ_b^{(i))}{}_k, \quad (12)$$

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

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

$$SO^{(I)}{}_1 \text{ ^ } SO^{(I)}{}_2 \text{ ^ } \dots \text{ ^ } SO^{(I)}{}_k \text{ ^ } \dots \text{ ^ } SO^{(I)}{}_K,$$

$$\text{якщо } \{TZ^{(I)}{}_k\} \rightarrow TZ, \quad (13)$$

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

(14)

У виразі (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 (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

$$TZ_a^{(i)}{}_k \text{ ¦} TZ_b^{(i)}{}_l, \quad (15)$$

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

Структура тезаурусу бази захисту

Структура тезаурусу прикладної системи TZ_{PS} відображена на рис. 1 і є рекурентною схемою (1). Як бачимо, тезаурус TZ_{PS} містить усі можливі семантичні образи на всіх заданих рівнях абстрагування, що в сукупності складають мову відображення цієї прикладної області. 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} (рис. 2).

Для захисту смислу текстових повідомлень пропонується такий методологічний підхід, 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 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, можливо формалізувати цю структуру таким чином, щоб параметри локалізації були представлені в явному вигляді, наприклад:

$$SO^{(i)}_{k} \in \left[\begin{array}{l} SO^{(i-1)}_{k,1,1};\ SO^{(i-1)}_{k,1,2};\ \ldots;\ SO^{(i-1)}_{k,1,n};\ \ldots;\ SO^{(i-1)}_{k,1,N(l=1)} \\ SO^{(i-1)}_{k,2,1};\ SO^{(i-1)}_{k,2,2};\ \ldots;\ SO^{(i-1)}_{k,2,n};\ \ldots;\ SO^{(i-1)}_{k,2,N(l=2)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ SO^{(i-1)}_{k,l,1};\ SO^{(i-1)}_{k,l,2};\ \ldots;\ SO^{(i-1)}_{k,l,n};\ \ldots;\ SO^{(i-1)}_{k,l,N(l)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ SO^{(i-1)}_{k,L,1};\ SO^{(i-1)}_{k,L,2};\ \ldots;\ SO^{(i-1)}_{k,L,n};\ \ldots;\ SO^{(i-1)}_{k,L,N(L)} \end{array} \right. \qquad (16)$$

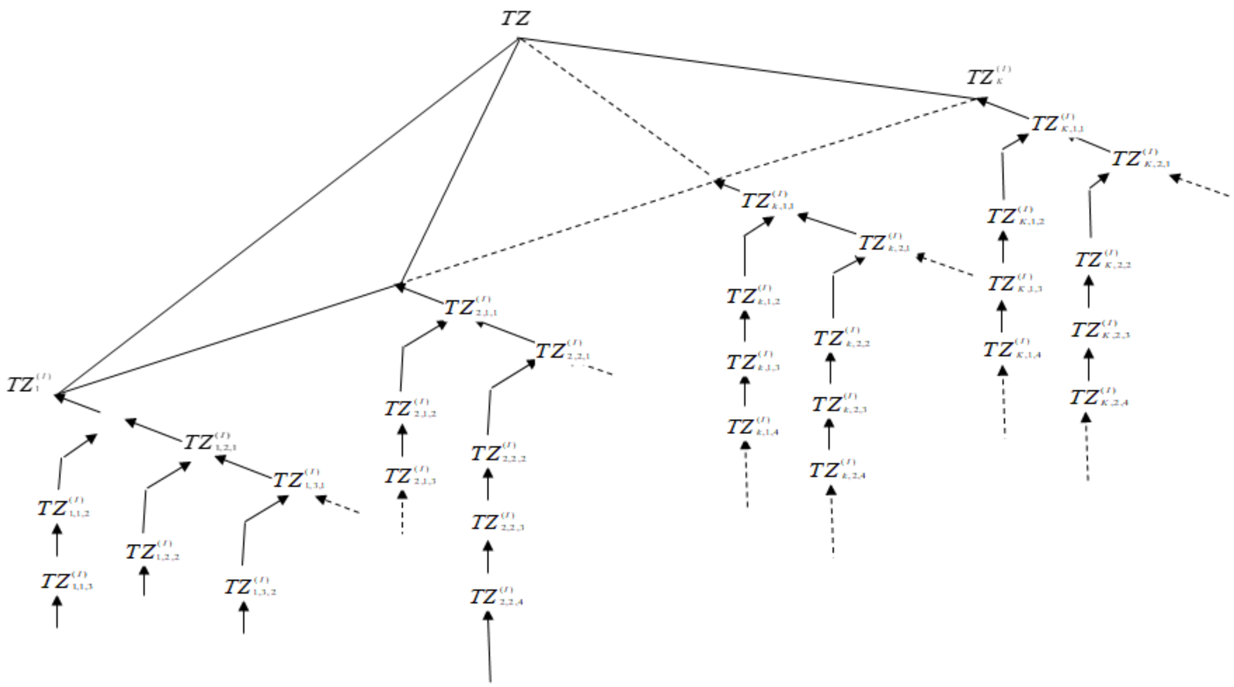

Рис. 2. Структура тезаурусу бази захисту текстової інформації TZ_{BZ} у прикладній системі, що відображається тезаурусом TZ_{PS}

Поточні значення нижніх індексів у позначеннях семантичних одиниць у виразі (16) однозначно визначають місця розташування цих семантичних одиниць у структурі TZ_{BZ}.

Отже, будь-яка семантична одиниця $TZ^{(I)}{}_k$ із множини семантичних одиниць I-го рівню абстрагування, що входять до складу тезаурусу мови обраної області активності TZ_{PS}, під час крипто аналізу сприймається суб'єктом як можливий кандидат на смислову ідентичність із вихідною семантичною одиницею відкритого тексту $TZ^{(I)}{}_l$.

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

$$SO^{(I-1)}{}_{k,1} \wedge SO^{(I-1)}{}_{k,2} \wedge \ldots \wedge SO^{(I-1)}{}_{k,l} \wedge \ldots \wedge SO^{(I-1)}{}_{k,L},$$
$$\text{якщо } \{TZ^{(I-1)}{}_{k,l}\} \xrightarrow[L]{} TZ^{(I)}{}_k, \qquad (17)$$

за умови $SO^{(I-1)}{}_k \equiv TZ^{(I-1)}{}_k$, для будь-яких значень (I-1), k та l. (18)

Аналогічно за рекурентною схемою можна записати вирази щодо відношень між смисловими одиницями для будь-яких інших більш деталізованих рівнів абстрагування представлення смислових образів. Із виразів (16)–(18) випливає, що семантичні одиниці, що входять до складу будь-якої однієї гілки у структурі тезаурусу TZ_{BZ}, знаходяться між собою у відношенні семантичної правдоподібності.

Висновки

Таким чином розроблено метод побудови семантичного словника, який за рахунок уведення в прикладну лексикографічну систему показників семантичних зв'язків між смисловими конструкціями мови відображення прикладної області застосування задає семантичну структуру словника прикладної області, що дає можливість використати його у складі досконало стійкої криптосистеми захисту текстової інформації.

Водночас 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, правдоподібність, фальшиво-подібність.

Нарешті, на закінчення запропоновано методологічний підхід до захисту смислу текстових повідомлень, що заснований на використанні відношень смислової правдоподібності.

Перспективи подальших досліджень

Перспективи дослідження полягають у технічній реалізації крипто-семантичного метода захисту текстових даних, що буде заснований на використанні прикладного тезаурусу смислових образів, який дозволить забезпечити режим досконалої стійкості у рамках конкретно визначених прикладних систем.

Технічне рішення дасть можливість забезпечити досконалу стійкість захисту для умов, коли обсяги зашифрованих семантичних одиниць обраної мови представлення смислових образів не обмежуються відстанню єдиності та довжиною ключової інформації, що сприяє усуненню проблеми розповсюдження ключів шифрів.

ЛІТЕРАТУРА

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. **Тилборг Ван Х. К. А**. Основы криптологии. Профессиональное руководство и интерактивный ученик. — М. : Мир, 2006. — 471 с.

5. **Гроздов Г. Г.** Современные методы защиты информации; под ред. чл.-корр. НАН Украины В. В. Васильева. — К., 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.

8. **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.

9. **Заєць В. В., Чуприн В. М.** Розрахунок ефективності криптосемантичної системи захисту інформації // Збірник наукових праць «Управління розвитком». — Харків : ХНЕУ, 2008. — №6 — С. 68–70.

Одарченко Р. С., Самойлик Є. О., Абакумова А. О.

МЕТОД ПОБУДОВИ СЕМАНТИЧНОГО СЛОВНИКА У СКЛАДІ ДОСКОНАЛО СТІЙКОЇ КРИПТОСИСТЕМИ ЗАХИСТУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

У статті розроблено формальну структуру тезаурусу смислових образів для будь-якої мови взагалі або для будь-якого суб'єкту окремо. 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. Одночасно зазначалося, що для побудови будь-якої досконало стійкої лексикографічної крипто-семантичної 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). Відповідно з цим, запропоновано методологічний підхід до захисту смислу текстових повідомлень, що заснований на використанні відношень смислової правдоподібності. На закінчення представлено формальні відображення семантики текстової інформації, що дозволяє автоматизувати процес створення семантичних словників.

Ключові слова: криптосистема; захист; тезаурус; семантичний словник; текстова інформація.

Odarchenko R. S., Samoilyk Ye. O., Abakumova A. O.

METHOD OF CONSTRUCTING A SEMANTIC DICTIONARY CONSISTING OF A PERFECTLY STABLE CRYPTOSYSTEM OF TEXT INFORMATION PROTECTION

The article describes the formal structure of semantic images thesaurus for any language in general or for any subject separately. It is shown that in many applied spheres it is advisable to assign the following hierarchy of semantic units that reflect logically completed thoughts of a certain level of abstraction, where the degree of abstraction is reduced from left to right: script/situation/phrase word. At the same time, it was noted that for the construction of any perfectly stable lexicographic crypto-semantic protection system of text information it is necessary to have formal definitions of such semantic relations between semantic units as semantic identity, semantic difference, questionable semantic identity (or plausibility), questionable semantic difference (or false similarity). Accordingly, a methodological approach is proposed to protect the meaning of text messages, which is based on the use of semantic plausibility relations. In conclusion, the formal representation of textual information semantics is presented, which allows automating the process of creating semantic dictionaries.

Keywords: cryptosystem; protection; thesaurus; semantic dictionary; text information.

Одарченко Р. С., Самойлик Е. О., Абакумова А. А.

МЕТОД ПОСТРОЕНИЯ СЕМАНТИЧЕСКОГО СЛОВАРЯ В СОСТАВЕ СОВЕРШЕННО УСТОЙЧИВОЙ КРИПТОСИСТЕМЫ ЗАЩИТЫ ТЕКСТОВОЙ ИНФОРМАЦИИ

В статье разработано формальную структуру тезауруса смысловых образов для любого языка вообще или для любого субъекта в отдельности. Показано, что во многих прикладных сферах целесообразно назначить следующую иерархию смысловых единиц, отражающих логически завершенные мысли определенного уровня абстрагирования, где убывания степени абстрагирования принято слева направо: сценарий/ситуация/фраза/слово. Одновременно отмечалось, что для построения любой совершенно устойчивой лексикографической крипто-семантической системы защиты текстовой информации необходимо иметь формальные определения таких смысловых отношений между семантическими единицами как смысловая идентичность, смысловое отличие, сомнительная смысловая идентичность (или правдоподобие), сомнительное смысловое отличие (или фальшиво-сходство). Согласно с этим, предложено методологический подход к защите смысла текстовых сообщений, основанный на использовании отношений смыслового правдоподобия. В заключение представлены формальные отображения семантики текстовой информации, что позволяет автоматизировать процесс создания семантических словарей.

Ключевые слова: криптосистема; защита; тезаурус; семантический словарь; текстовая информация.

Стаття надійшла до редакції 26.06.2018 р.
Прийнято до друку 28.08.2018 р.
Рецензент – д-р техн. наук, проф. Конахович Г. Ф.