DOI 10.15276/imms.v15.no1.39
УДК 004.056.5

Informatics and Mathematical Methods in Simulation
Vol.15 (2025), No.1, pp. 39-47

ТЕОРІЯ ГРАФІВ ЯК ОСНОВА МЕТОДІВ ВБУДОВУВАННЯ ІНФОРМАЦІЇ

І.І. Борисенко, І.С. Вінковська

Національний університет «Одеська політехніка»
1 Шевченка пр., Одеса, 65044, Україна
Email: boris_enko63@ukr.net

У сучасному цифровому середовищі, де витоки даних і кібератаки стають все більш поширеними, задача створення надійних засобів захисту інформації є надзвичайно актуальною. У епоху цифрових технологій захист веб – сайтів, організація надійних способів передачі та зберігання даних, особливо конфіденційної інформації, стали важливими аспектами будь-якого бізнесу чи організації і активно розвиваються. Однак, із зростанням залежності від Інтернету ризик кіберзагроз також зріс експоненціально. В комплексних системах захисту інформації широке застосування знайшли стеганографічні методи, принцип роботи яких полягає в створенні скритих каналів зв'язку у вже існуючих потоках даних в інформаційно-телекомунікаційних системах. Задача вибору стеганографічного методу дозволяє вирішити деякі вимоги, що висуваються перед стеганосистемою при її побудові. Однією із найважливіших вимог є забезпечення стійкості стеганосистеми до атак стеганоаналізу. Статистичні методи стеганоаналізу намагаються виявити найменші зміни у статистичній поведінці файла, викликані стеганографічним перетворенням. Суть статистичних методів полягає в оцінюванні ймовірності існування стеганографічного приховування з невідомою стеганосистемою на основі критерію оцінки наближення досліджуваного контейнера до "природнього". Метою роботи є підвищення ефективності стеганосистеми шляхом модифікації методу приховування інформації, розробленого на основі теорії графів. Поставлена мета була досягнута шляхом вводу в зазначений метод блоку знаходження максимального паросполучення. Результатом роботи є розробка модифікованого методу приховування інформації, який можна застосовувати для побудови стеганографічних систем. Ефективність запропонованого методу досягається за рахунок збереження статистик першого порядку контейнера і має кращі показники ніж у метода - прототипа. Значущість результату полягає у підвищені загальної стійкості стеганосистеми до статистичних атак.

Ключові слова: стеганосистема, стеганографічний метод, стійкість стеганосистеми, статистичні методи стеганоаналізу, стеганографічне перетворення

Вступ. Одним з найбільш перспективних підходів до виявлення існування прихованого каналу передачі інформації є підхід, що представляє собою введення прихованої інформації в файл як порушення статистичних законів природних контейнерів. При такому підході аналізуються статистичні характеристики досліджуваного контейнера і встановлюється, чи схожі вони (характеристики) з характеристиками природних контейнерів (якщо так, то прихованої передачі інформації немає), або вони схожі з характеристиками контейнерів з вбудованою інформацією (якщо так, то виявляється факт існування прихованого каналу передачі даних). Цей клас стегоатак є ймовірносним, тобто вони не дають однозначної відповіді, а формують оцінки типу «ця досліджувана послідовність містить приховане повідомлення з ймовірністю 90%». Ймовірнісний характер статистичних методів стегоаналізу не є істотним недоліком, оскільки на практиці ці методи часто дають оцінки ймовірності існування стегоканала, що відрізняються від одиниці або нуля на нескінченно малі значення [1].

У класі статистичних методів стегоаналізу використовується безліч статистичних характеристик, таких як оцінка ентропії, коефіцієнти кореляції, залежності між елементами послідовностей, умовні розподіли, відмінність розподілів за

критерієм хі-квадрат, оцінка числа переходів значень молодших бітів у сусідніх елементах контейнера і багато інших [2,3].

Більшість статистичних атак, як правило, не потребують спеціальних технічних та програмних засобів і високої кваліфікації атакуючого і можуть проводитись без специфічного програмного або технічного забезпечення, без спеціальної підготовки та кваліфікації зловмисника.

Так, наприклад, найпростіші тести оцінюють кореляцію елементів контейнера, в які можуть бути вбудовані приховані повідомлення. Це робить такий і подібний вид атак надзвичайно простим і поширеним, і саме це диктує високу потребу у забезпеченні стійкості прихованого повідомлення до збурних дій викликаних вбудовуванням повідомлення в контейнер.

Програмні інструменти приховування інформації такі як Steganos, Outguess, Jsteg, Jphs, S-Tools та інші прості в використанні і здатні створити стеганографічний канал з великою пропускною здатністю. Ці інструменти, як правило, використовують метод найменшого значущого біта (LSB) та його модифікації, але стеганоконтейнери, які створені за допомогою LSB, успішно виявляються методами стеганоаналізу.

Цей факт призвів до появи цілого ряду робіт, присвячених методам вбудовування в молодший біт без суттєвого порушення закону розподілення бітів.

Наприклад, тривіальною модифікацією методу LSB-replacement є LSB-matching, який випадковим чином змінює піксельні значення на ± 1 так, що молодші біти пікселів відповідають бітам повідомлення, що вбудовується. Завдяки такій модифікації LSB-*matching* стеганоконтейнери набагато важче розпізнаються методами стеганоаналізу.

Вбудовування інформації в контейнер призводить до зміни його характеристик. Ці зміни використовуються методами стеганоаналізу для розпізнавання стеганоконтейнерів.

Чим менше збурень зазнає контейнер під час вбудовування інформації, тим важче стеганоаналітичним методам забезпечити низький рівень похибки при розпізнаванні.

В останній час активно ведуться роботи по створенню стеганографічних методів та алгоритмів, розробники яких намагаються забезпечити найменш можливий вплив на контейнер вбудованої інформації як за рахунок вибору елементів контейнера для вбудовування так і специфіки самого алгоритму [4-7].

Група методів, так звані методи мінімального стеганографічного збурення, значно підвищує стійкість стеганографічних систем.

Отже, якщо забезпечити обмін елементів контейнера, а не їх модифікацію, то тим самим будуть максимально збережені статистичні характеристики контейнера. Саме такий принцип положено в метод представлений в [8].

Постановка задачи. Метою роботи є підвищення ефективності стеганосистеми шляхом модифікації методу приховування інформації, розробленого на основі теорії графів запропонованого в [8].

Для досягнення мети вирішуються задачі.

1. Модифікувати алгоритм знаходження паросполучень, що є основою для модифікації методу [8].
2. Провести оцінку ефективності, зокрема порівняльну, модифікованого і базового методу.

Контейнер, який буде використовуватися для вбудовування інформації позначимо літерою С, а його елементи s_i, тоді $C = \{s_1, ..., s_i, ..., s_n\}$, де n – кількість елементів контейнера, $s_i \in S$ – множина значень.

Визначимо функцію $f:S\to\{0,...,p-1\}$, яка кожному s_i ставить у відповідність залишок від ділення s_i на p, позначимо цю операцію $s_i \oplus p$. Для вбудовування одного елемента повідомлення будемо використовувати не один елемент контейнера, а групу з c елементів так, як це запропоновано в [8]. Такий підхід дає більшу свободу вибору одного елемента з c для модифікації. Таким чином, після розбивки елементів контейнера на $k=\left\lfloor \frac{n}{c} \right\rfloor$ неперетинаючихся груп одержимо структуру $C'=\{c_1,...,c_i,...,c_k\}$, де $c_i=(s_{i1},...,s_{ic})$. Будь-яке $c_i^{'}=f(f(s_{i1})+...+f(s_{ic}))$ визначає деяке значення, яке вже вбудоване в контейнер.

Повідомлення, позначимо його $M=\{m_1,...,m_i,...,m_k\}$ кодуємо за тим же принципом, що й елементи контейнера, тобто $m_i\in\{0,...,p-1\}$. Наприклад, якщо $p=2$, то значеннями m_i є біти.

Вузол графа представляє собою структуру $v_i(X_i,Y_i)$, де $Y_i=(y_{i1},...,y_{ic})$ – i-та група цільових значень, а $X_i=(x_{i1},...,x_{ic})$ – позиції елементів контейнера, які складають Y_i. Вузли створюються тільки у випадку, коли $m_i\neq c_i'$. Використання парасполучень побудованого графа дасть можливість зменшити кількість корегованих елементів контейнера і вбудовування інформації буде в більшій мірі відбуватися за рахунок їх обміну.

Як вже відмічалося, що вузли графа створюються тільки у випадку, коли $m_i\neq c_i$, але не зважаючи на це їх кількість є значною.

Так, наприклад, при вбудовуванні повідомлення розміром 1КВ в середньому маємо три тисячі вузлів і до десяти тисяч ребер, а при збільшенні розміру вкладень до 4КВ кількість вузлів відповідно збільшується в чотири рази, а кількість ребер сягає півтора мільйони. Тому автори [8] відмовилися від використання списків суміжності, а запропонували дві структури, які також є громіздкими.

В даній роботі пропонується контейнер розбити на блоки і для кожного блоку будувати граф, розмір якого дозволить використовувати його матрицю суміжності, яка однозначно представляє граф.

Основна частина. Важливим етапом роботи нового стеганографічного алгоритму є знаходження найбільшого парасполучення в графі, що є самостійною комбінаторною задачею, яка відноситься до NP-повних і має експоненціальну часову складність. У цьому зв'язку розробляються різні евристики з поліноміальною часовою складністю, які в обмін на зниження часу роботи алгоритму не дають точного розв'язку, тому розробка нових алгоритмів залишається актуальним завданням.

Парасполученням (або незалежною множиною ребер) графа G називають множину ребер у якій ніякі два ребра не суміжні.

Парасполучення графа G називають максимальним, якщо воно не міститься в жодному парасполученні з більшою кількістю ребер, і найбільшим, якщо кількість ребер у ньому найбільша серед усіх парасполучень графа G [9].

В роботі пропонується алгоритм знаходження найбільшого парасполучення, назвемо його z_blok, на базі матриці суміжності графа, двоїстого до даного.

Розглянемо вихідний граф (рис.1), та побудуємо до нього двоїстий (рис.2) за правилом: вузлами будуть ребра вихідного графа, які позначені номерами від одного до десяти, і два вузли суміжні, якщо ребра, які їм відповідають, також суміжні в вихідному графі, тобто інциндентні одному й тому ж вузлу.

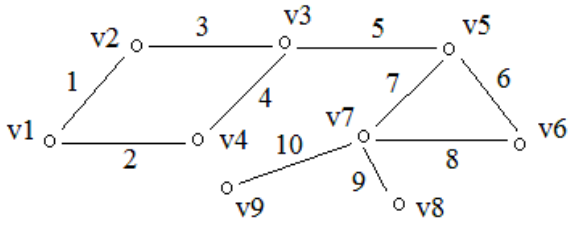

Рис.1. Вихідний граф

Найбільше паросполучення відповідає найбільшій незалежній множині вершин [9].

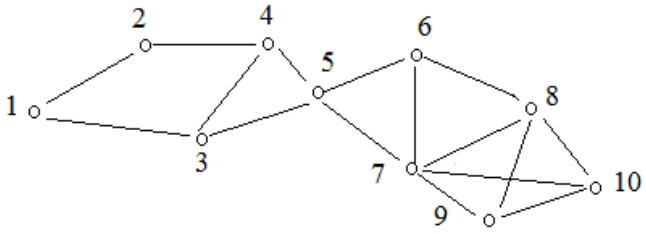

Рис.2. Граф, двоїстий до вихідного

Побудуємо матрицю суміжності A (рис.3) для двоїстого графа.

номер вузла	1	2	3	4	5	6	7	8	9	10
1	X	1	1							
2	1	X		1						
3	1		X	1	1					
4		1	1	X	1					
5			1	1	X	1	1			
6					1	X	1	1		
7					1	1	X	1	1	1
8						1	1	X	1	1
9							1	1	X	1
10							1	1	1	X

Рис.3. Матриця суміжності двоїстого графа

Аналіз матриці суміжності A показує, що якщо, наприклад, строки та стовпці з номерами 1, 4, 6 та 9 (вузли з такими номерами утворюють незалежну множину) переставити так, щоб вони стояли поряд, то на діагоналі матриці утворюється нульовий блок. Отже, матрицю суміжності треба перебудувати так, щоб вона мала блочно-діагональний вигляд. Нульові квадратні блоки будуть відповідати максимальним незалежним множинам вершин.

Алгоритм z_blok :

а) побудувати граф, двоїстий до вихідного з матрицею суміжності $A=(a_{ij})$ розміром $n\times n$;

б) аналізувати пари сусідніх стрічок $(i,i+1)$, i - непарне:

1) $\sum_{j=1}^{i-1} a_{ij} > \sum_{j=1}^{i-1} a_{i+1,j}$,

2) $\sum_{j=1}^{i-1} a_{ij} = \sum_{j=1}^{i-1} a_{i+1,j}$ і $\sum_{j=i+2}^{n} a_{ij} > \sum_{j=i+2}^{n} a_{i+1,j}$,

3) $\sum_{j=1}^{i-1} a_{ij} = \sum_{j=1}^{i-1} a_{i+1,j}$ і $\sum_{j=i+2}^{n} a_{ij} = \sum_{j=i+2}^{n} a_{i+1,j}$ але одиниця в стрічці i до головної діагоналі зустрілася раніше, ніж в стрічці $i+1$;

при виконанні будь-якої умови виконати перестановку стрічок i та $i+1$ та відповідних стовпців;

в) аналізувати пари сусідніх стрічок $(i, i+1)$, i - парне;

виконати перестановку стрічок i та $i+1$ та відповідних стовпців при виконанні однієї з умов представлених у пункті б).

Алгоритм закінчує свою роботу, якщо немає пар $(i, i+1)$, для яких виконуються умови 1) – 3). З побудованих блоків вибрати блок найбільшого розміру, номери стрічок (стовпців), що йому відповідають будуть складати найбільшу незалежну множину двоїстого графа і найбільше паросполучення вихідного.

П'ята ітерація для непарного i являється останньою ітерацією роботи алгоритму z_blok.

Алгоритмом виділено чотири нульові блоки: перший блок розміром 4x4 відповідає вузлам з номерами 1, 6, 9 та 4, другий блок розміром 3x3 відповідає вузлам 10, 2, 3, третій блок розміром 2x2 відповідає вузлам 8, 5, останній блок відповідає вузлу 7 (рис.4). Отже, найбільший блок є шуканим, таким чином, найбільша незалежна множина представлена вузлами 1, 6, 9 ,4 в двоїстому графі і ребра саме з такими номерами є найбільшим паросполученням у вихідному графі.

номер вузла	1	6	9	4	10	2	3	8	5	7
1	X					1	1			
6		X						1	1	1
9			X		1			1		1
4				X		1	1		1	
10			1		X			1		1
2	1			1		X				
3	1			1			X		1	
8		1	1		1			X		1
5		1		1			1		X	1
7		1	1		1			1	1	X

Рис. 4. Матриця суміжності A , i - непарне, п'ята ітерація

Алгоритм вбудовування повідомлення.

Процес вбудовування повідомлення представимо схемою (рис.5), взявши конкретні значення параметрів графової моделі c=3 та p=4 [8].

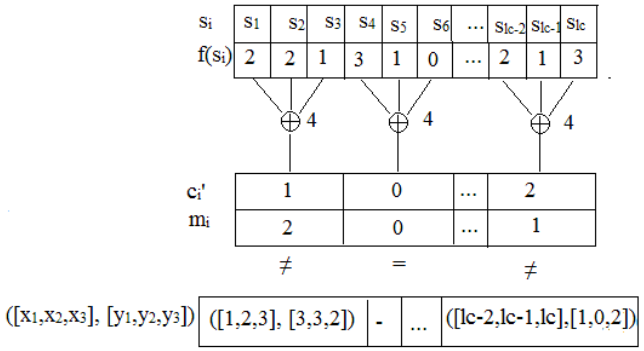

Рис.5. Схема вбудовування повідомлення

Перша стрічка містить елементи контейнера. Друга – визначені значення функцією f. Три (в загальному випадку c) елемента $f(s_i)$ модифікуються завдяки застосуванню операції додавання по модулю 4 (в загальному випадку по модулю p) щоб одержати значення c_i', які порівнюються з елементами повідомлення m_i. Якщо значення не співпали, як в першому та останньому випадку, то створюється вузол, як показано в останній стрічці. Цільові значення вузла [y_1,y_2,y_3] обчислюються додаванням різниці $m_i - c_i'$ до кожного значення $f(s_i)$. Заміна одного з $f(s_i)$ на його цільове значення приведе у відповідність c_i' та m_i.

Розглянемо як створюються ребра [8]. Вище було зазначено, що алгоритм вбудовування переслідує мету в більшій мірі елементи контейнера обмінювати, ніж модифікувати. Розглянемо вузли перший і останній схеми, представленої на рисунку 5. Цільове значення елемента s_1 дорівнює трьом при значенні $f(s_1) = 2$, а елемента s_{lc} дорівнює двом при $f(s_{lc}) = 3$. Якщо різниця їх значень $d = | s_1 - s_{lc} |$ дозволяє обміняти ці елементи місцями, тобто обмін не визве видимого спотворення контейнера, то створюється ребро. Легко помітити, що ребер між двома вузлами може бути декілька, наприклад, також можна створити ребро (s_3, s_{lc-2}) за умови виконання вимоги до різниці їх значень. Це дає більше ступенів свободи для вибору з цих ребер одного, виходячи, наприклад, з мінімальності значення d.

Стеганографічний алгоритм, назвемо його graf_matching, представимо наступними кроками.

Крок 1. Розбиваємо матрицю контейнера – зображення на блоки заданого розміру.

Крок 2. Для кожного блоку будуємо граф G_i, виконуючи дії представлені схемою, зображеною на рисунку 5.

Крок 3. Для побудованого графа G_i будуємо граф, двоїстий до даного G_{di}, який представляється матрицею суміжності.

Крок 4. Знаходимо незалежні множини вузлів на графі G_{di}, використовуючи алгоритм z_blok.

Крок 5. На графі G_i визначаємо найбільше паросполучення.

Крок 6. Для вузлів, які належать найбільшому паросполученню виконуємо обмін елементів контейнера.

Крок 7. Для вузлів, які не належать найбільшому паросполученню виконуємо модифікацію елементів контейнера.

Щоб декодувати повідомлення, вбудоване алгоритмом graf_matching треба розбити матрицю стеганоконтейнера на блоки того самого розміру, що і при вбудовуванні. Використовуючи ті самі значення для параметрів c і p, які використовувалися при вбудовуванні, обчислити $f(s_i)$ та c_i'. Значення c_i' є елементами повідомлення.

Оцінка ефективності модифікованого методу. Для оцінки ефективності модифікованого методу вбудовування повідомлення було проведено обчислювальний експеримент в ході якого була зроблена оцінка алгоритму знаходження паросполучень z_blok шляхом порівняння його з відомими алгоритмами karp-sipser та Куна (табл.1) та порівняльний аналіз модифікованого методу з методом [8] (табл.2).

Таблиця 1.

Визначення частки об'єму контейнера, що підлягає корегуванню P(%)

Контейнер	Алгоритм	Блок 1	P	Блок 2	P	Блок 3	P	Блок 4	P
Autumn.tif									
Кількість елементів		47 - 53		49-57		45-59		50-58	
Кількість паросплучень	karp-sipser	22 - 20	13-3	24-18	21-1	24-19	11-7	28-21	8-2
	куна	21 - 20	13-5	24-17	22-1	24-19	11-7	27-20	10-4
	z_blok	22 - 19	13-3	25-17	20-3	23-20	11-5	28-20	9-2
Canoe.tif									
Кількість елементів	karp-sipser	44-50		43-51		45-60		46-61	
Кількість паросп-чень	karp-sipser	18-19	12-3	18-19	12-2	19-20	12-1	20-21	12-1
	куна	17-19	13-3	18-19	12-2	19-20	10-1	21-22	11-2
	z_blok	18-20	13-3	19-20	11-1	20-21	9-1	22-23	9-1

Як видно з табл.1, z_blok не тільки не гірший за методи karp-sipser і Куна, але і показав кращі результати в деяких випадках. Так, наприклад, блок 2 і блок 4 для контейнера Autumn.tif і блоки 2,3,4 для контейнера **Canoe.tif** вказують на те, що алгоритм z_blok знаходить більше паросполучень і частка об'єму інформації для корегування менша, а отже робота z_blok більш якісна.

Таблиця 2.

Статистичні показники викривлення контейнера

Назва показника	Оригінал	Алгоритм [8]	Graf_matching
Средня абсолютна різниця $\mathrm{AD} = \frac{1}{XY}\sum_{x,y}\left\|C_{x,y} - S_{x,y}\right\|$	0	0,0823	0,0228
Нормована середня абсолютна різниця $\mathrm{NAD} = \frac{\sum_{x,y}\left\|C_{x,y} - S_{x,y}\right\|}{\sum_{x,y}\left\|C_{x,y}\right\|}$	0	3,7577e-04	1,8760e-04
Середньоквадратична помилка $\mathrm{MSE} = \frac{1}{XY}\sum_{x,y}\left(C_{x,y} - S_{x,y}\right)^2$	0	0,0823	0,0228
Відношення «сигнал/шум» $\mathrm{SNR} = \frac{\sum_{x,y}\left(C_{x,y}\right)^2}{\sum_{x,y}\left(C_{x,y} - S_{x,y}\right)^2}$	∞	6,4724e+05	9,3503e+05
Максимальне відношення «сигнал/шум» $\mathrm{PSNR} = \mathrm{XY} \cdot \frac{\max_{x,y}\left(C_{x,y}\right)^2}{\sum_{x,y}\left(C_{x,y} - S_{x,y}\right)^2}$	∞	7,1958e+05	2,6918+06

В таблиці 2 наведені деякі статистичні показники, які свідчать про те, що в результаті застосування розробленого методу викривлень контейнера, при вбудовуванні повідомлення, відбувається менше, ніж коли застосовується метод прототип, а отже статистики контейнера зберігаються краще. Це відбувається за рахунок модифікації алгоритму знаходження парасполучень в результаті чого парасполучення знаходяться більш якісно і кількість пікселів, які треба модифікувати, зменшується.

Висновок. В роботі вирішено задачу підвищення ефективності стеганосистеми шляхом розробки модифікації методу вбудовування повідомлення, запропонованого в [8].

В ході модифікації запропоновано поліпшення знаходження максимального парасполучення, як основи алгоритму вбудовування повідомлень на основі теорії графів. Проведено експериментальні дослідження на реальних зображеннях, які використовувалися в якості контейнера та на різних повідомленнях, які кодувалися різноманітним чином. Експериментально доведена більша ефективність розробленого метода в порівнянні з [8]. При застосуванні розробленого методу краще зберігаються статистики першого порядку, а отже можна стверджувати, що статистичному стегааналізу виявити вкладення буде важче, ніж при застосуванні [8].

Результати роботи можуть бути використані для побудови нових стеганографічних систем, стійких до статистичних атак та для розв’язку задач, які зводяться до пошуку парасполучень у графі.

Список літератури

1. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография/ М.: Солон-Пресс, 2002. 272 с.
2. Швидченко И.В. Методы стеганоанализа для графических файлов. *Штучний інтелект.* 2010. №4. С. 696-705.
3. BÖhme *R.* Advanced statistical steganalysis. Berlin, Heidelberg: Springer-Verlag, 2010.
4. Filler T., Judas J., Fridrich J. Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes. *Forensics and Security*. 2011. V. 6(1). P. 920–935.
5. Kodovský J., Fridrich J., Holub V. On Dangers of Overtraining Steganography to an Incomplete Cover Model. *Proc. ACM Multimedia & Security Workshop, Niagara Falls, New York*. 2011. P. 69-76.
6. Filler T., Fridrich J. Gibbs construction in Steganography. *Forensics and Security*. 2010. V.5(4). P. 705-720.
7. Fridrich J., Filler T. Practical methods for minimizing embedding impact in steganography. *Proceedings SPIE, Electronic Imaging, Steganography, and Watermarking of Multimedia Contents IX*. 2007. 6505. P. 2-3.
8. Hetzl S., Mutzel P. A graph-theoretic approach to steganography. *Proc. Communication and Multimedia security*. 2005. P.119-128.
9. Харари Ф. Теория графов, М.:Мир, 1993. С.203.

GRAPH THEORY AS THE BASIS OF METHODS FOR EMBEDDING INFORMATION

I.I. Borysenko, I.S. Vinkovska

National Odesa Polytechnic University
1, Shevchenko Ave, Odesa, 65044, Ukraine
Email: boris_enko63@ukr.net

In today's digital environment, where data leaks and cyberattacks are becoming more common, the task of creation reliable means of information protection is extremely urgent. In the digital age, the protection of websites, the organization of reliable ways to transfer and store data, especially confidential information, have become important aspects of any business or organization and are actively developing. However, with the increase in dependence on the Internet, the risk of cyber threats has also increased exponentially. In complex information security systems, steganographic methods are widely used, the principle of which is to create hidden communication channels in existing data flows in information and telecommunication systems. The problem of choosing a steganographic method allows you to solve some of the requirements put forward for the steganographic system when it is built. One of the most important requirements is to ensure the resistance of the stegosystem to stegoanalysis attacks. Statistical methods of steganoanalysis attempt to detect the slightest changes in the statistical behavior of a file caused by steganographic transformation. The essence of statistical methods is to assess the probability of the existence of the steganographic of hiding of an unknown stegano system based on the criterion for assessing the approh of the study container to the "natural" one. The purpose of the work is to increase the efficiency of the stegan system by modifying the method of hiding information developed on the basis of graph theory. The goal was achieved by entering the maximum steam combination block into the specified method. The result of the work is the development of a modified method of hiding information, which can be used to build steganographic systems. The effectiveness of the proposed method is achieved by preserving the first-order statistics of the container and has better performance than that of the prototype method. The significance of the result lies in increasing the overall resistance of the hip system to statistical attacks.

Keywords: steganosystem, steganographic method, steganosystem stability, statistical methods of steganoanalysis, steganographic transformation