МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

DOI 10.54398/20741707_2022_2_84
УДК 681.3

АССОЦИАТИВНАЯ СТЕГАНОГРАФИЯ: СОСТОЯНИЕ И ПЕРСПЕКТИВЫ

Статья поступила в редакцию 14.04.2022, в окончательном варианте – 02.05.2022.

Вершинин Игорь Сергеевич, Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, 420111, Российская Федерация, г. Казань, ул. К. Маркса, 10,
кандидат технических наук, зав. кафедрой компьютерных систем, ORCID: 0000-0001-5166-2862, e-mail: ISVershinin@kai.ru

Дается определение понятия ассоциативной стеганографии (АС). Рассматривается базовый для ассоциативной стеганографии алгоритм маскирования заданного конечного множества бинарных матриц-эталонов десятичных цифр одинаковых размеров. Проведено исследование свойств базового алгоритма, помехоустойчивости ассоциативной защиты данных, предложен метод ее повышения. Дается оценка стегостойкости и криптостойкости (если стегостойкость не безусловна) ассоциативной защиты данных. Рассматриваются вопросы применения ассоциативной стеганографии для защиты картографических сцен и текстов, организации систем управления базами данных с такой защитой.

Ключевые слова: ассоциативная стеганография, помехоустойчивость, стегостойкость, криптостойкость, ассоциативная защита картографических и текстовых сцен

ASSOCIATIVE STEGANOGRAPHY: STATE AND PROSPECTS

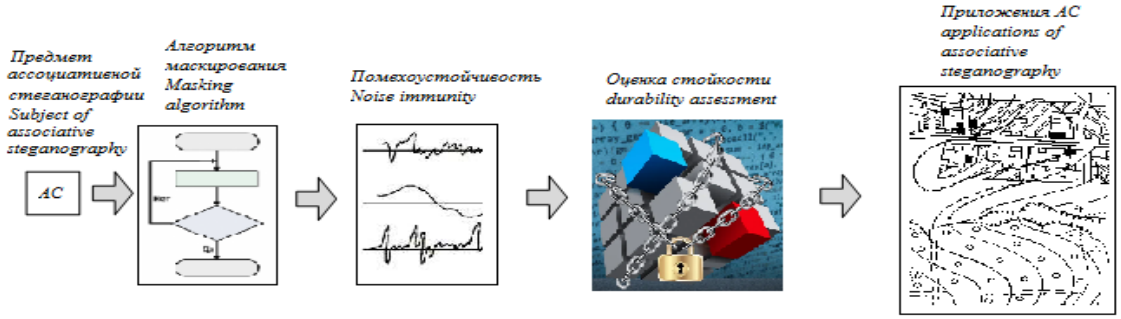
The article was received by the editorial board on 14.04.2022, in the final version – 02.05.2022.

Vershinin Igor S., Kazan National Research Technical University named after A.N. Tupolev-KAI, 10 K. Marx St., Kazan, 420111, Russian Federation,
Cand. Sci. (Engineering), Head of the Computer Systems Department, ORCID: 0000-0001-5166-2862, e-mail: ISVershinin@kai.ru

The definition of the concept of associative steganography is given. The basic algorithm for associative steganography for masking a given finite set of binary matrices-standards of decimal digits of the same size is considered. A study of the properties of the basic algorithm, noise immunity of associative data protection, and a method for improving it is proposed. The assessment of the stegodurability and cryptodurability (if the stegodurability is not unconditional) of associative data protection is given. The issues of the use of associative steganography for the protection of cartographic scenes and texts, the organization of database management systems with such protection are considered.

Keywords: associative steganography, noise immunity, stegodurability, cryptodurability, associative protection of cartographic and text scenes

Graphical annotation (Графическая аннотация)



Введение. Предметом ассоциативной стеганографии является анализ защищенных двумерных сцен. Под сценой понимают картину (изображение) с множеством объектов. Задача анализа сцен – одна из задач распознавания образов, когда не интересуются «тонкой структурой» изображения, а всего лишь укрупненным описанием того, что на нем представлено, в терминах «объекты – координаты» [1]. Число имен объектов и градаций их координат полагается конечным и заведомо известным.

Используется k-разрядное десятичное кодирование координат и имен объектов. Исходная информация по сцене структурируется как таблица с множеством записей, представленная на рисунке 1. Каждая десятичная цифра представлена своей двоичной матрицей-эталоном размерами m×n, m = 2n – 1 (рисунок 2 – пример представления символа 9 для n = 5).

Код объекта	Координата X	Координата Y

Рисунок 1 – Вид структурированной таблицы

Защита организуется трансформацией этих матриц в троичные путем их маскирования. Под маскированием понимается генерация матрицы масок для каждой матрицы-эталона.

Размеры матриц-эталонов и масок – одинаковы. Маска определяет позиции тех битов в матрице-эталоне, которые не должны подвергаться изменению и используются для дальнейшей идентификации эталона. Генерация масок осуществляется случайным образом с использованием АЛГОРИТМА маскирования. Полученный в результате набор масок представляет собой ключ распознавания, которое выполняется путем двумерно-ассоциативного различения каждой цифры кода на множестве троичных эталонов, элементы которого принадлежат {0, 1, –}.

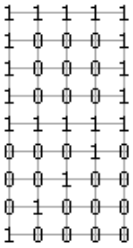


Рисунок 2 – Представление символа 9

Троичное представление любого объекта (координаты) помещается с использованием масок в стегоконтейнер. Стегоконтейнер изначально заполнен отрезком псевдослучайной последовательности (ПСП). Его длина, определяемая разрядностью кода и размерами матриц, всегда много больше объема полезной информации, определяемого числом сохраняемых (незамаскированных) бит. Принцип формирования стегоконтейнеров показан на рисунке 3.

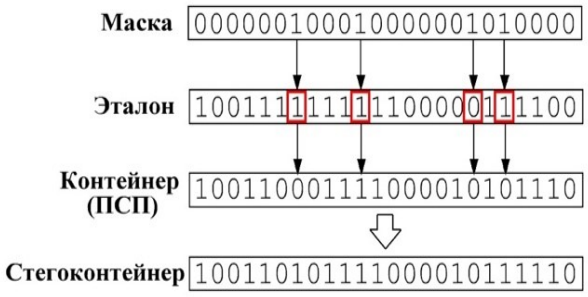


Рисунок 3 – Принцип формирования стегоконтейнеров

В качестве основы предлагаемого метода защиты используется противопоставление сообщений неединичного множества. Для заданного множества эталонов формируется ключ, представляющий собой набор масок. Он определяет ограниченное количество бит в бинарных матрицах-эталонах, которые остаются истинными. Эти биты случайно распределены по битовой сетке эталона. Существенные биты символов располагаются по внешнему контуру и внутреннему «зигзагу» матриц (рис. 4; а – n = 3, б – n = 7) суммарной длиной (9n – 12). Размер ключа определяется только количеством эталонов и размерами матриц-эталонов вне зависимости от размера сообщения.

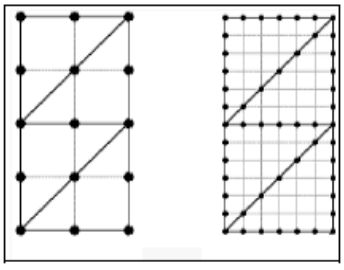
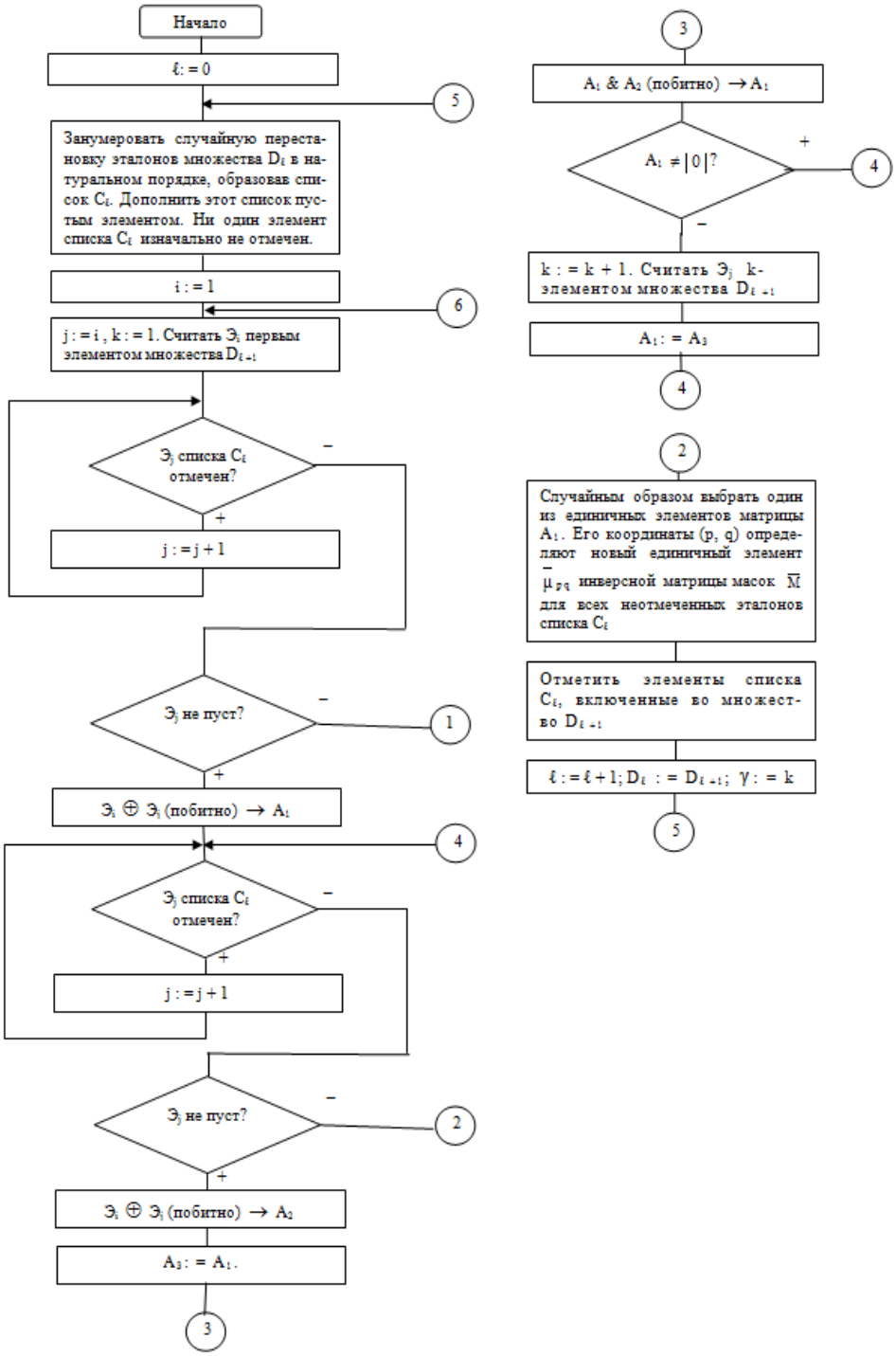


Рисунок 4 – Внешний контур и внутренний «зигзаг»: а – n = 3; б –n = 7

Внедренная шумовая картина (ГАММА) не оказывает влияния в случае знания ключа (т.е. для санкционированного пользователя), однако создает существенную и практически непреодолимую преграду для несанкционированного. Организация процедур ассоциативной защиты сцен и дальнейшее управление процессом поиска требуют больших вычислительных ресурсов в случае выполнения этих процедур в «реальном времени». Это определяет необходимость использования высокопроизводительных вычислительных кластеров [2].

Развиваемый подход, в отличие от известных методов стеганографического преобразования, в состоянии обеспечить практически *безусловную стегостойкость*, а в сравнении с известными шифрами – более высокую помехоустойчивость при хранении и передаче информации по открытым каналам связи.

Базовый алгоритм маскирования. На рисунке 5 в виде блок-схемы представлен алгоритм формирования масок путем случайного определения битов, подлежащих сохранению в бинарных матрицах заданного множества эталонов $Э_t$, D_ℓ обозначает совокупность бинарных матриц-эталонов, которые рассматриваются на ℓ-этапе работы АЛГОРИТМА. Для $\ell = 0$ соответствующее множество D_0 включает все типы эталонов.



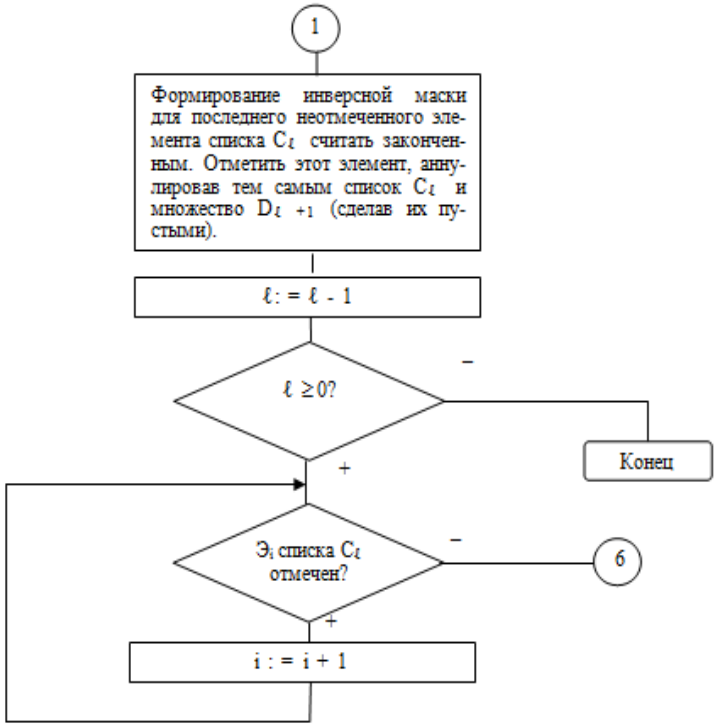


Рисунок 5 – Блок-схема АЛГОРИТМА

Один из возможных результатов работы АЛГОРИТМА представлен на рисунке 6. Точками показаны сохраняемые биты, размер эталонов – 5×3.

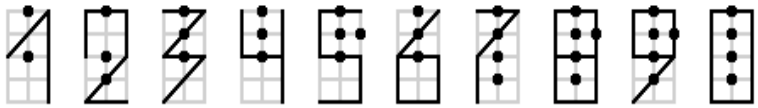


Рисунок 6 – Результат работы АЛГОРИТМА

Основополагающее свойство АЛГОРИТМА для всей ассоциативной стеганографии устанавливает следующее.

Теорема 1. Для произвольной бинарной матрицы размером m x n проведение процедуры распознавания на множестве эталонов тех же размеров по маскам, сгенерированным с использованием АЛГОРИТМА, приведет к распознаванию в этой матрице одного и только одного эталона из указанного множества.

Указанное свойство используется при определении помехоустойчивости и стойкости предлагаемого метода.

Помехоустойчивость. Проведены исследования по определению помехоустойчивости предлагаемого метода к действию случайных и преднамеренных помех. В качестве исходной информации, подлежащей хранению или передаче, используется набор стегоконтейнеров (кодовых слов), состоящих из трех букв (почтовых индексов), погруженных по маске в ПСП. Размер каждого стегоконтейнера составляет 198 байт при n = 60.

Для случайных помех рассматривалось два случая:

1. Для каждого кодового слова случайные помехи локализованы в пределах 16 байт
2. В каждом кодовом слове искажено более 16 байт.

В обоих случаях проводилось сравнение с (255, 223)-кодом Рида – Соломона.

Также проводилось исследование помехоустойчивости и к действию преднамеренных помех. Установлена необходимость введения избыточности для повышения помехоустойчивости. При этом происходит внедрение избыточности не в передаваемом сообщении, а на уровне ключей. Суть введения избыточности на уровне ключей состоит в следующем. Процедура маскирования и формирование стегконтейнеров происходит с использованием множества наборов масок Q. Количество наборов масок всегда нечетное (3, 5, 7, …). Распознавание также происходит с использованием всех наборов масок на основе принципа большинства, т.е. результатом распознавания является символ, число распознаваний которого больше или равно (Q+1)/2. Если данное условие не выполняется, фиксируется факт отказа от распознавания.

Но даже при отсутствии избыточности применение метода ассоциативной стегозащиты позволяет увеличить степень стойкости к обоим типам помех в сравнении с иными существующими методами защиты информации: допускается искажение до 3 % хранимых и передаваемых бит вместо 1,5 % – для ГОСТ 28147-89 [3, 4]. Использование избыточного маскирования повышает эту оценку до 6 %.

Самым значимым результатом при введении избыточности на уровне ключей является отсутствие неверных распознаваний по сравнению с рассмотренным (255, 223)-кодом Рида – Соломона в случае действия случайных помех при искажении более 16 байт. Также установлена принципиальная возможность противодействия наличию многократной аддитивной помехи при выборе Q = 5. Обобщая оба результата (для борьбы как со случайными, так и с преднамеренными помехами), на практике следует использовать Q = 7. Это можно обосновать тем, что в реальных системах приемопередачи сообщений практически невозможно определить тип действующей помехи (либо обеих помех), а также ту часть сообщения, на которые эти помехи воздействуют. Исходя из этого, целесообразно смириться с различного рода потерями временного и технического характера при использовании семи наборов масок в угоду сохранения целостности данных, так как их потеря является гораздо более критическим событием.

Стойкость ассоциативной защиты. Особенностью стегосистем являются затруднения в установлении самого факта передачи сообщений, связываемые с понятием стегостойкости. Она безусловна, если затруднения трансформируются в практическую невозможность. Это имеет место, когда ПСП непрерывно генерируется на множестве контейнеров и рост суммарной длины ПСП растет с увеличением n при неизменном среднем числе вкраплений. Но при постоянстве n и росте числа контейнеров в стегосообщении объем вкраплений и длина ПСП нарастают линейно, и условие неразличимости «пустого» и стегоконтейнеров не выполняется. Тогда необходимо провести специальный стегоанализ по выявлению требований неразличимости для рассматриваемых систем. Если и ее не будет, необходим дополнительный криптоанализ. Он полезен и при возникновении у противника каких-то сомнений в передаче «пустых» контейнеров. Проведенный криптоанализ ограничен случаем воздействия всего лишь трех характерных криптоатак – «лобовая», на ГАММУ и со знанием открытого текста, в том числе и для избыточного маскирования, когда для сокрытия сообщений одновременно используют несколько наборов масок с целью повышения помехоустойчивости при хранении и передаче данных.

По результатам проведенных исследований [5, 6] установлено, что следование принципам ассоциативной стеганографии позволяет обеспечить доказуемую стойкость предлагаемого метода. Это – достаточно «сильное» свойство, которое строго установлено только для шифров с применением гаммирования [4].

Стратегия ассоциативной стегозащиты сцен картографии. В основе принятой стратегии защиты картографических сцен лежит представление сцены в виде множества таблиц-отношений, атрибутами которых являются десятичные коды объектов и их координат. Замаскированные тройки <ИМЯ ОБЪЕКТА><КООРДИНАТА X><КООРДИНАТА Y> размещаются в трех стегоконтейнерах. Изначально каждый стегоконтейнер заполняется отрезком псевдослучайной последовательности (ПСП) той же длины. Затем выполняется «вкрапление» в него по маске значимых бит матриц A^t.

Для пояснения принципов формирования искомого множества таблиц-отношений для случая точечных объектов рассмотрим пример картографической сцены с нанесенной на ней глобальной координатной сеткой (рис. 6, Y и X – максимальные значения координат y и x).

Кластер отображается на сцене прямоугольником определенных размеров. Внутри кластера наносится локальная координатная сетка. Принимая в качестве погрешности величину ε, получаем шаг локальной координатной сетки 2ε. Понятие ассоциативной стеганографии предполагает полноту множества задействованных кодов объектов и градаций их координат для данной разрядности кода k. Иными словами, мощность этого множества $\Gamma = 10^k$. Значение k выбирается из условия равного числа градаций локальной и глобальной сеток.

Принцип формирования кластеров:

1. Случайным образом выбираем некоторую строку исходной таблицы. Отмечаем эту строку. Определяем *глобальные координаты* выделяемого кластера (рис. 6). Сама же запись преобразуется к виду:

Код объекта	(x, y)

Здесь (x, y) – локальные координаты выделенного объекта в данном кластере.

2. Пункт 1 повторяется на множестве неотмеченных строк. При этом каждый раз устанавливается, принадлежит ли вновь выделенная строка к одному из ранее введенных кластеров. Далее происходит преобразование координат из глобальных в тексте в локальные в кластере. Если такого кластера нет, создается новый кластер. И так до тех пор, пока не будут рассмотрены все записи исходной таблицы.

3. Обеспечивается равное количество записей во всех кластерах для повышения стойкости защиты. Для этого вводятся так называемые «пустые» (несуществующие) объекты и координаты. Они подлежат сокрытию вместе с реально существующими объектами и координатами.

4. Чтобы скрыть положение в кластере его «родителя», все записи в каждом кластере подлежат перемешиванию.

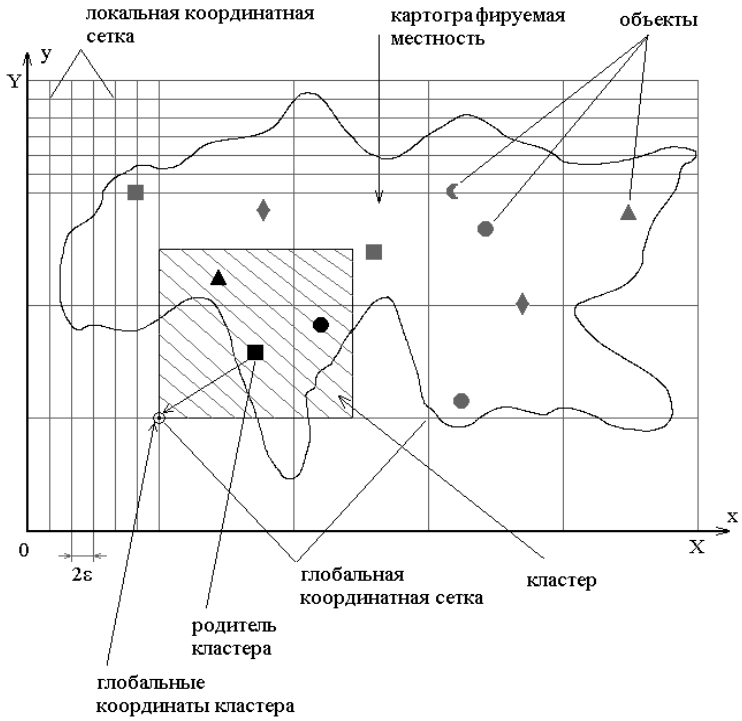


Рисунок 6 – Пример картографической сцены

Организация систем БД картографии с ассоциативной защитой. За основу построения специализированной СУБД с ассоциативной защитой (АЗ) берется «интеллектуальная» файл-серверная организация взаимодействия (рис. 7). Предлагаемая инфологическая схема БД полнообъектных картографических сцен (ПКС) с ассоциативной защитой показана на рисунке 8. Принят единообразный формат хранения данных для всех (точечных, линейных и площадных) объектов [7].

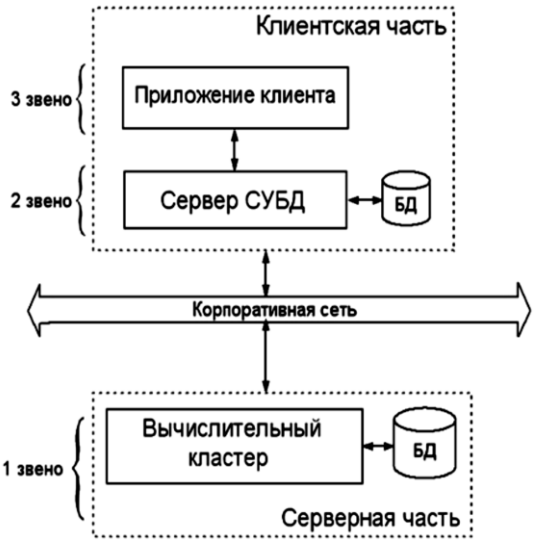


Рисунок 7 – Организация СУБД с ассоциативной защитой

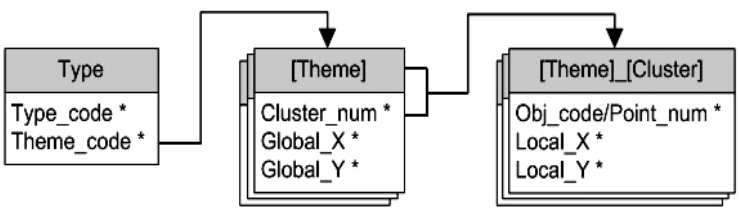


Рисунок 8 – Инфологическая схема баз данных полнообъектных картографических сцен с ассоциативной защитой

В рамках предлагаемой схемы выделяются наборы таблиц-отношений. Для каждого тематического слоя создается отдельная таблица, которая описывает кластеры внутри этого слоя. Также отдельная таблица создается для каждого кластера, в которой описываются объекты внутри этого кластера.

Принципы взаимодействия между отношениями этих таблиц определяются процедурами серверной программы во время обработки запроса. Связи на рисунке 9 указывают на формирование SQL-поискового запроса в таблице, имя которой определяется по окончании формирования раскрытого стегокода из предыдущей таблицы.

Описание отношений:

- Type – отношение, содержащее информацию о всех тематических слоях сцены;
- [Theme] – набор отношений, каждое из которых описывает отдельный тематический слой;
- [Theme]_[Cluster] – набор отношений, каждое из которых описывает содержимое (объекты) одного кластера какого-либо тематического слоя;
- Type_code* – определяет тип слоя: слой точечных, слой линейных или слой площадных объектов;
- Theme_Code* – уникальный код тематического слоя;
- Cluster_num* – код номера кластера в тематическом слое;
- Global_X*, Global_Y*– глобальные координаты нижнего левого угла данного кластера;
- Obj_code/Point_num* – код точечного объекта (код номера узловой точки линейного или площадного объекта);
- Local_X*, Local_Y* – локальные координаты точки/узла внутри кластера.

Звездочкой отмечены данные об атрибутах, которые должны храниться в замаскированном виде.

Создание кластеров производится в соответствии с принципами формирования кластеров, рассмотренными выше. Для каждого создаваемого кластера осуществляется присвоение ему очередного (по порядку) номера. Координаты кластера выбираются согласно рисунку 6. Номер и координаты записываются в рабочее отношение [Наименование слоя], которое описывает кластеры этого слоя. Порядок строк в табличных отношениях-кластерах, которые включают как значимые, так и пустые записи, выбирается произвольно.

Для пустых элементов данного тематического слоя (т.е. для пустых точечных объектов либо узловых точек линейных или площадных объектов) произвольно осуществляется выбор кода из множества неиспользуемых в этом слое кодов объектов /узловых точек. Для этого определяется количество точек N^{points}_{max} в кластере данного тематического слоя с наибольшим количеством записей. Затем для каждого кластера-отношения данного слоя осуществляется дополнение «пустыми точками» до значения N^{points}_{max}.

С помощью дополнительного отношения со следующим форматом замаскированных записей клиент выполняет фильтрацию пустых записей:

theme_type	theme_code	N^{points}_{max}

Следует отметить, что никаких дополнительных ограничений из-за наличия нумерации узловых точек на линейные размеры линейных/площадных объектов и размер кластеров, получаемых в процессе формирования сцены, не накладывается.

После получения ответа на запрос пользователя от сервера клиентская сторона должна выполнить процедуру раскрытия сокрытых данных. Помимо этого, необходимо восстановить порядок следования узловых точек для линейных/площадных объектов. Это достигается проведением процедуры сортировки.

Ассоциативно-стеганографический подход к защите текстовых сообщений. Ассоциативная стеганография вполне приемлема и для защиты текстовых характеристик различных объектов [8, 9]. Предлагаемая инфологическая схема БД замаскированной текстовой информации показана на рисунке 9.

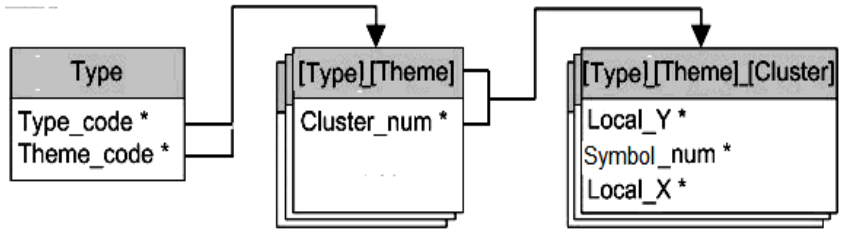


Рисунок 9 – Инфологическая схема текстовой базы данных с ассоциативной защитой

Описание отношений:

- Type – отношение, содержащее информацию о всех тематических слоях, представленных парой сокрытых кодов: Код типа – Код слоя;
- [Type]_[Theme] – набор отношений, каждое из которых описывает отдельный тематический слой;
- [Type]_[Theme]_[Cluster] – набор отношений, каждое из которых представляет одну из характеристик данного объекта;
- Cluster_num* – может отражать номер страницы. Один кластер (фрагмент) на одну страницу;
- Local_X / Local_Y * – номер строки и позиция в строке;
- Symbol_num * – трехразрядный код символа.

Как и ранее, звездочкой отмечены данные об атрибутах, которые должны храниться в замаскированном виде. Формат записи в кластере:

Код номера строки (Local _Y*)	Код символа (Symbol_num*)	Код позиции в строке (Local _X*)

Перспективы. Дальнейшее развитие ассоциативной стеганографии связывается с:

1) анализом целесообразности непрерывной матричной бинаризации и маскирования символов без указания их координат при ассоциативной защите текстов с целью дополнительного снижения объема передач в три раза;

2) доведением разработанных исследовательских версий соответствующих СУБД до готовых к широкому практическому применению программных продуктов;

3) развитием *новых приложений* ассоциативной стеганографии.

Заключение. По результатам проведенного рассмотрения определено понятие ассоциативной стеганографии, разработан базовый алгоритм маскирования кодовых представлений данных схемы (объектов и их координат), сформулирована основополагающая теорема, важная для всей ассоциативной стеганографии. Разработан метод повышения помехоустойчивости ассоциативной стеганографии к действию случайных и преднамеренных помех, основанный на использовании избыточного маскирования. Предложена стратегия защиты картографических и текстовых сцен, основанная на принципах ассоциативной стеганографии. Развиты подходы к управлению базами данных картографических и текстовых сцен с ассоциативной защитой, основанные на построении инфологических схем БД, что позволило выделить принципиальные особенности специализированных СУБД [10] с ассоциативной защитой.

Библиографический список

1. Дуда, Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – Москва : Мир, 1976. – 511 с.

2. Вершинин, И. С. Распределенное управление защищенными картографическими базами данных / И. С. Вершинин, Р. Ф. Гибадуллин, А. Е. Прохоров // Высокопроизводительные параллельные вычисления на кластерных системах : материалы 8-й Междунар. конф. HPC-2008. – Казань : КГТУ, 2008. – С. 216–221.

3. ГОСТ 28147-89. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования. – Москва : Госстандарт СССР, 1989.

4. Schneier, B. Applied Cryptography / B. Schneier. – 2nd ed. – John Wiley&Sons, 1996.

5. Raikhlin, V. A. The Elements of Associative Steganography Theory / V. A. Raikhlin, R. F. Gibadullin, I. S. Vershinin // Moscow University Computational Mathematics and Cybernetics. – 2019. – Vol. 43, iss. 1. – P. 40–46.

6. Vershinin, I. S. Associative Steganography. Durability of Associative Protection of Information / I. S. Vershinin, R. F. Gibadullin, S. V. Pystogov, V. A. Raikhlin // Lobachevskii Journal of Mathematics. – 2020. – № 3. – P. 439–449.

7. Raikhlin, V. A. Reliable Recognition of Masked Binary Matrices. Connection to Information Security in Map Systems / V. A. Raikhlin, I. S. Vershinin, R. F. Gibadullin, S. V. Pystogov // Lobachevskii Journal of Mathematics. – 2013. – Vol. 34, № 4. – P. 319–325.

8. Vershinin, I. S. Associative Steganography of Text Messages / I. S. Vershinin, R. F. Gibadullin, S. V. Pystogov, V. A. Raikhlin // Moscow University Computational Mathematics and Cybernetic. – 2021. – Vol. 45, № 1. – P. 1–11.

9. Raikhlin, V. A. Is it possible to reduce the sizes of stegomessages in associative steganography? / V. A. Raikhlin, R. F. Gibadullin, I. S. Vershinin // Lobachevskii Journal of Mathematics. – 2022. – Vol. 43, № 2. – P. 455–462.

10. Карасев, Д. С. Исследование проблем работы и создания специализированной СУБД / Д. С. Карасев // Научный аспект. – 2012. – № 2. – С. 148–149.

References

1. Duda, R., Hart, P. *Raspoznavanie obrazov i analiz stsen* [Pattern recognition and image analyze]. Moscow, Mir Publ., 1976. 511 p.

2. Vershinin, I. S., Gibadullin, R. F., Prokhorov, A. E. Raspredelennoe upravlenie zashishennymi kartograficheskimi bazami dannykh [Distributed management of secure cartographic databases]. *Vysokoproizvoditelnye parallelnye vychisleniya na klasternykh sistemakh : materialy 8-y Mezhdunarodnoy conferentsii NRS-2008* [High-performace parallel computing on cluster systems: Proceedings of the 8th International Conference NRS-2008]. Kazan, Kazan State Technical University, 2008, pp. 216–221.

3. *GOST 28147-89. Sistemy obrabotki informacii. Zashita kriptograficheskaya. Algoritm kriptograficheskogo preobrazovaniya* [Information processing systems. Cryptographic protection. Cryptographic Conversion Algorithm]. Moscow, Gosstandart SSSR Publ., 1989.

4. Schneier, B. *Applied Cryptography*. 2nd ed. John Wiley&Sons, 1996.

5. Raikhlin, V. A., Vershinin, I. S., Gibadullin, R. F. The Elements of Associative Steganography Theory. *Moscow University Computational Mathematics and Cybernetics*, 2019, vol. 43, iss.1, pp. 40–46.

6. Vershinin, I. S., Gibadullin, R. F., Pystogov, S. V., Raikhlin, V. A. Associative Steganography. Durability of Associative Protection of Information. *Lobachevskii Journal of Mathematics*, 2020, no. 3, pp. 439–449.

7. Raikhlin, V. A., Vershinin, I. S., Gibadullin, R. F., Pystogov, S. V. Reliable Recognition of Masked Binary Matrices. Connection to Information Security in Map Systems. *Lobachevskii Journal of Mathematics*, 2013, vol. 34, no. 4, pp. 319–325.

8. Vershinin, I. S., Gibadullin, R. F, Pystogov, S. V., Raikhlin, V. A. Associative Steganography of Text Messages. *Moscow University Computational Mathematics and Cybernetic*, 2021, vol. 45, no. 1, pp. 1–11.

9. Raikhlin, V. A., Gibadullin, R. F., Vershinin, I. S. Is it possible to reduce the sizes of stegomessages in associative steganography? *Lobachevskii Journal of Mathematics*, 2022, vol. 43, no. 2, pp. 455–462.

10. Karasev, D. S. Issledovanie problem raboty i sozdaniia spetsializirovannoi SUBD [Research of problems of work and creation of a specialized DBMS]. *Nauchnyi aspect* [Scientific Aspect], 2012, no. 2, pp. 148–149.