УЧЕНЫЕ ЗАПИСКИ КАЗАНСКОГО УНИВЕРСИТЕТА.
СЕРИЯ ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

2024, Т. 166, кн. 3
С. 320–330

ISSN 2541–7746 (Print)
ISSN 2500–2198 (Online)

ОРИГИНАЛЬНАЯ СТАТЬЯ

УДК 004.056

doi: 10.26907/2541-7746.2024.3.320-330

ЭЛЕМЕНТЫ ТЕОРИИ АССОЦИАТИВНОЙ ЗАЩИТЫ ИНФОРМАЦИИ

И.С. Вершинин

Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, г. Казань, 420111, Россия

Аннотация

Рассмотрены элементы теории нового научного направления – ассоциативной защиты информации при ее хранении и передаче. Показано, что применение данного подхода ведет к повышению уровня защиты и помехоустойчивости анализа сцен.

Ключевые слова: ассоциативная защита, анализ сцен, помехоустойчивость, криптостойкость, стегостойкость

Введение

Интенсивное развитие компьютерной стеганографии началось с 1990-х годов, когда возникли новые задачи сокрытия информации, которые не могла решить классическая криптография. Известные в настоящее время стеганографические методы предполагают внедрение информации в изображение, видео- или аудиофайлы. Было развито множество методов с применением серьезного математического аппарата.

Так, для погружения скрытой информации в неподвижные изображения большое распространение получили алгоритмы Jsteg, Outguess, F5.

Jsteg вкладывает информацию в наименьшие значащие биты (НЗБ) частотных коэффициентов цветных изображений в формате JPEG. Outguess также вкладывает сообщение в наименее значимые биты (НЗБ) частотных коэффициентов цветных изображений, однако для повышения стойкости к стегоанализу выполняется повторное вложение с целью приближения гистограммы покрывающих сообщений к гистограмме стегосистемы.

F5 отличается от Jsetg и Outguess тем, что при вложении минимизируется количество изменяемых коэффициентов, что затрудняет использование простейших методов стегоанализа.

Известны более сложные стегосистемы, такие как Model based или Perturbed Stegosystem (PS), которые оказываются трудно обнаруживаемыми. Идея построения первой состоит в том, что статистика НЗБ частотных коэффициентов изображения после вложения «подгоняется» под статистику этих же НЗБ покрывающих сообщений (без вложения) по некоторой аналитической модели. В случае же PS используются факт двойного квантования с ухудшением качества (который имеет место в реальных алгоритмах сжатия изображения) и вложение в определенные

коэффициенты. Такой метод почти не искажает статистику по сравнению с исходным покрывающим сообщением. В настоящее время известно понятие совершенных стегосистем. В них исходные контейнеры и соответствующие им стегоконтейнеры подчинены одному закону распределения вероятностей. Однако применение таких стегосистем на практике является затруднительным.

Как известно, для большинства современных стегосистем, скрывающих информацию в различного рода файлах, надежность этих систем зависит от объема встраиваемой информации. Чем он больше, тем ниже надежность, и наоборот. Отсюда получаем критерий выбора оптимального решения в зависимости от указанного соотношения. Варьируя эти параметры, можно добиться либо высокого качества стегопоследовательности, либо большого объема скрываемых в стегоконтейнере данных. Увеличение одновременно обоих параметров недостижимо. Таким образом, можно сделать вывод о том, что применение существующих стегосистем к защите сцен с большим объемом данных является затруднительным, а в ряде случаев – невозможным, т. к. при большом объеме скрываемых данных неизбежна существенная потеря стойкости. Рассматриваемый далее ассоциативный подход более подходит для целей защиты данных анализируемых сцен. В рамках представленного научного направления ассоциативной защиты информации ведется двумерно-ассоциативный анализ сцен, которые могут быть представлены в виде картографической и текстовой информации. Проводится распознавание сцен, итогом которого является описание сцены в терминах «объекты – координаты» [1]. Количество типов объектов и их координат полагаются заранее известными.

Каждому объекту/координате ставится в соответствие k-разрядный десятичный код. Местоположение объекта определяется двумя координатами X и Y (рис. 1). Каждый разряд десятичного кода есть бинарная матрица A^t в алфавите почтовых индексов (десятичных цифр), размер которой $m \times n$, $m = 2n - 1$. Рис. 2 – пример отображения цифры 9 при $n = 5$.

Код объекта	Код координаты X	Код координаты Y

Рис. 1. Структура записи в таблице данных сцены

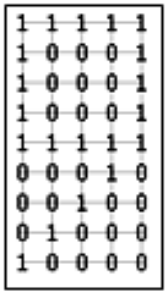

Рис. 2. Цифра 9

Далее проводится процедура маскирования. Для этого по разработанному АЛГОРИТМУ (см. далее) каждой бинарной матрице ставится в соответствие инверсная матрица маски с теми же размерами. Единичные биты матриц масок указывают на биты эталона, которые должны сохраняться и будут использоваться

в дальнейшем для проведения процедуры распознавания. Все остальные биты эталона (т. н. замаскированные биты) заменяются случайными значениями в процессе проведения процедуры рандомизации. В итоге исходные бинарные матрицы будут представлены (заменены) троичными матрицами, элементы которых принадлежат множеству $\{0, 1, -\}$.

Маскирование осуществляется из условия минимизации сохраняемых бит в матрицах-эталонах. Тем самым повышается стойкость защиты. Распознавание букв кодового слова осуществляется сопоставлением сохраненных бит на множестве троичных эталонов, т. е. проводится ассоциативная обработка.

Рассмотрим наиболее существенные отличия ассоциативной стегозащиты от существующих стего- и криптометодов защиты информации. Основой ассоциативной защиты является процедура маскирования. При этом маски могут быть использованы в двух случаях: для нейтрализации противодействия распознаванию санкционированным пользователем и для противодействия распознаванию несанкционированным пользователем. В первом случае речь идет об искажениях в изображениях, которые могут быть как случайными, так и преднамеренными. Это могут быть либо помехи, возникающие в процессе хранения или передачи информации, либо внедряемые злоумышленником преднамеренно путем искажения бит замаскированного изображения.

Во втором случае процедура рандомизации приводит к искажению части изображения, однако знание маски позволяет провести правильную идентификацию объектов при распознавании, т. е. процедура маскирования в этом случае является определяющей. Рандомизация же призвана скрыть сообщение, оставив истинными лишь незамаскированные биты.

Распознавание проводится по маскам. Для этого для каждого объекта/координаты проводится сопоставление со всеми троичными эталонами с использованием масок, указывающих местоположение незамаскированных (существенных) бит.

Таким образом, здесь можно говорить о симбиозе понятий крипто- и стеганографии. Оба понятия дополняют друг друга. Как известно, криптография путем применения сложных математических алгоритмов приводит к получению шифртекста той же длины, что и исходное сообщение. В случае ассоциативной защиты полезное сообщение – это небольшое количество сохраненных по маске бит, которые погружаются в пустой контейнер достаточно большого объема, что характерно для стеганографии. Вместе с тем, представление десятичных цифр бинарными матрицами (определяющими размер носителя) с последующим маскированием и рандомизацией – своеобразное шифрование, хотя оно и отлично от принятого криптографического.

Процедуру маскирования можно отнести к трафаретному способу стеганографии (в качестве частного случая). Известно, что трафарет «накладывается» на исходное (передаваемое) сообщение. Далее «поверх» трафарета наносится текст, имеющий некоторый смысл. В случае ассоциативной стегозащиты основное отличие в том, что исходные (незамаскированные) биты сообщения погружаются в шумовую картину. Другими словами, сообщение при передаче маскируется под шум канала. Противник в этом случае должен будет определить, присутствуют ли в системе связи только наложение шума канала или совокупность шума канала и стегосигнала [2]. В результате проведенных процедур маскирования и рандомизации из исходных представлений сцен формируются (по аналогии со стеганографией) стегоконтейнеры. Исходный (пустой) контейнер имеет длину $L = k(9n-12)$. Количество $(9n - 12)$ определяется совокупным контуром существенных бит би-

нарных матриц. В алфавите почтовых индексов существенные биты располагаются по внешнему контуру и внутреннему «зигзагу» соответствующих бинарных матриц. Рис. 3, а – пример представления совокупного контура для $n = 3$, рис. 3, б – для $n = 7$. В сформированные контейнеры заносится отрезок ПСП, далее в них по маске внедряются сохраняемые истинные биты. Их число определяется примерным значением $q = 5k$, что значительно меньше значения L. Это характерно для стеганографии.

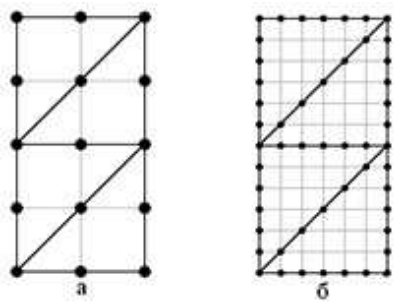

Рис. 3. Существенные биты

Метод ассоциативной стегозащиты, в сравнении с известными методами стеганографии, позволяет потенциально обеспечить безусловную стегостойкость. При этом рассматриваемый подход обладает свойством более высокой помехоустойчивости хранимой или передаваемой информации по сравнению с известными шифрами.

1. Базовый алгоритм маскирования

Далее бинарные матрицы эталонов A^t будем обозначать как эталоны символов $Э_t$. Разработанный базовый алгоритм маскирования или просто АЛГОРИТМ [3] осуществляет случайный поиск сохраняемых (немаскируемых) бит эталонов заданного множества. На рис. 4 и 5 представлена блок-схема АЛГОРИТМА. Суть работы АЛГОРИТМА заключается в выборе одного из вариантов маскирования на каждом этапе работы АЛГОРИТМА для рассматриваемой на этом этапе совокупности эталонов. АЛГОРИТМ построен таким образом, что при отсутствии информации об истинном ключе, генерируемом с его помощью, вероятность правильного распознавания должна быть невелика.

Пусть D_l – множество бинарных матриц эталонов, рассматриваемых на l-этапе работы алгоритма. Множество D_0 содержит полное множество типов эталонов.

По результатам проведенных исследований установлено, что для проведения правильного распознавания объектов в стегоконтейнерах необходимо и достаточно сохранить значения лишь нескольких бит. АЛГОРИТМ выполняет поиск таких бит. Сохраняемые биты выбираются из условия взаимной непокрываемости, т. е. на каждом этапе работы АЛГОРИТМА идет противопоставление рассматриваемых на этом этапе эталонов по значению одного бита, который должен сохраняться истинным для всех эталонов подмножества.

На рис. 6 показан пример одного из результатов работы АЛГОРИТМА. Биты, которые не маскируются (т. е. сохраняются), представлены точками.

В теореме 1 сформулировано важнейшее свойство АЛГОРИТМА, служащее основой для всего ассоциативного подхода.

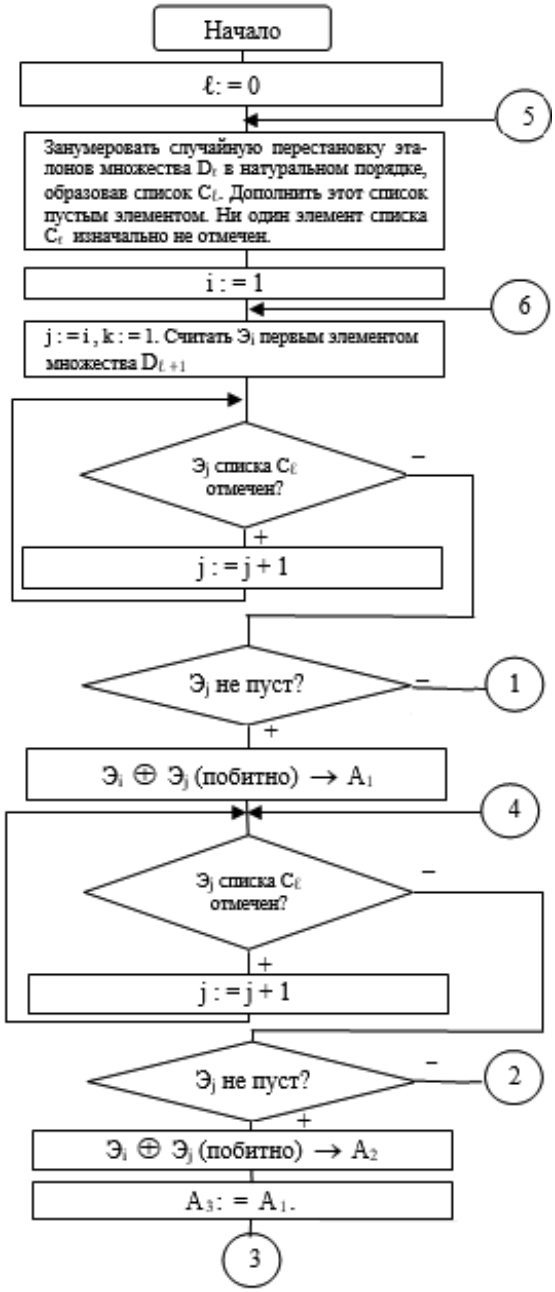

Рис. 4. Блок-схема АЛГОРИТМА

Теорема 1. *Для произвольной бинарной матрицы размером $m \times n$ 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.*

В частности, данная теорема используется для определения стегостойкости и помехоустойчивости ассоциативной защиты.

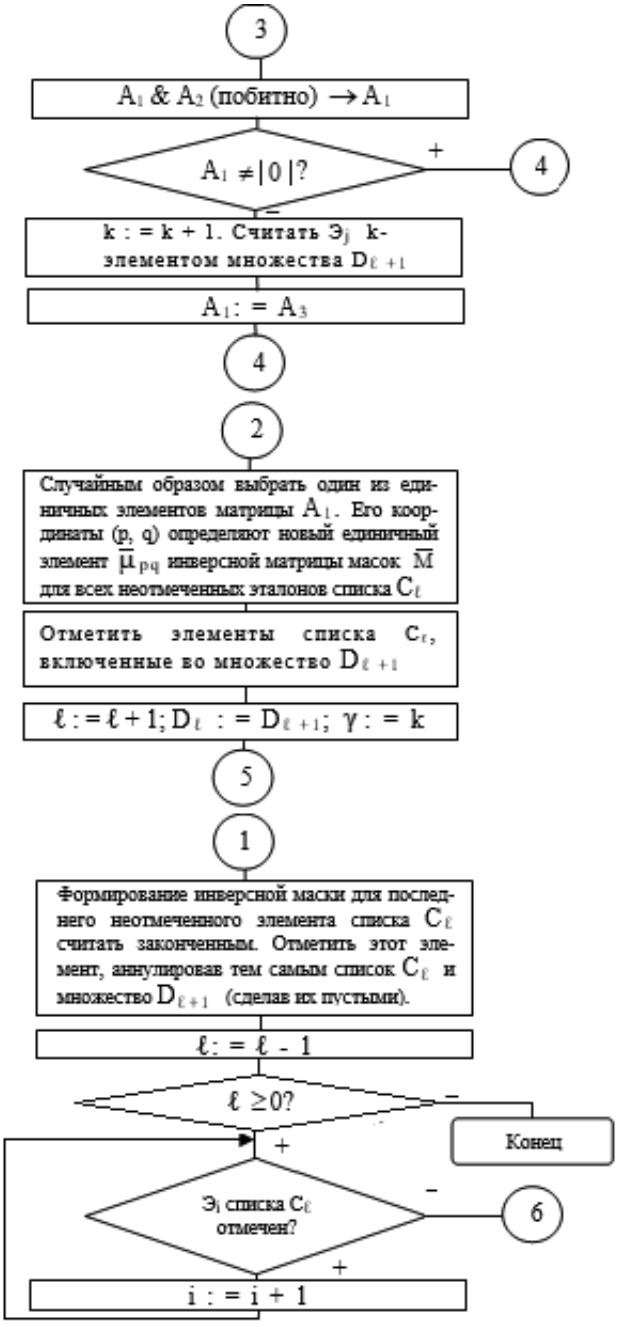

Рис. 5. Блок-схема АЛГОРИТМА (продолжение)

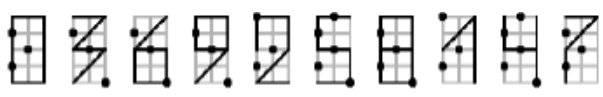

Рис. 6. Результат работы АЛГОРИТМА

2. Помехоустойчивость

При хранении и/или передаче информации возможно воздействие помех, как случайных, так и преднамеренных. Целью проведенных исследований было рассмотрение потенциальных возможностей рассматриваемого метода по противодействию указанным помехам, т. е. определение помехоустойчивости метода. Исследования проводились для трехразрядных стегоконтейнеров при $n = 60$ путем сравнения с (255, 223)-кодом Рида – Соломона [4].

Рассматривалось два случая. В первом помехам в стегоконтейнере было подвержено не более 16 байт (указанный код Рида – Соломона в этом случае позволяет скорректировать имеющиеся ошибки), во втором – более 16 байт. В результате проведенных исследований была установлена неудовлетворительная способность рассматриваемого подхода противостоять действию помех. Для повышения помехоустойчивости метода необходимо введение избыточности. При этом было предложено вводить избыточность не в передаваемое сообщение, а увеличивать число используемых ключей (наборов масок). Это означает следующее.

При погружении информации в стегоконтейнеры необходимо использовать не один, а несколько наборов масок Q (их нечетное количество). Эти же наборы масок используются и при распознавании. В случае воздействия помех результаты распознавания на этих ключах могут отличаться. За результат распознавания принимается код символа, полученный при соблюдении т. н. мажоритарного принципа, т. е. для каждого разряда стегоконтейнера за результат принимается та кодовая буква, которая была распознана не менее $(Q + 1)/2$. При несоблюдении условия устанавливается отказ от распознавания (фиксируется обнаружение ошибки).

Установлено, что на практике целесообразно использовать $Q = 7$. Указанное количество ключей позволяет не допустить прием зашумленного сообщения как истинного при искажении более 16 байт в сообщении (в отличие от рассматриваемого варианта кода Рида – Соломона) при действии как случайных, так и преднамеренных помех. Следует отметить, что для случайных помех достаточно выбора $Q = 5$. Однако, поскольку в реальных системах неизвестно, какая именно помеха будет действовать (возможно также воздействие и обеих помех), целесообразно использование семи наборов масок.

3. Стойкость ассоциативной защиты

Рассматриваются два аспекта стойкости – стего- и криптостойкость. Это обусловлено следующим. В работе [5] утверждается, что стегостойкость безусловна, если псевдослучайная последовательность (ПСП) непрерывно генерируется на множестве контейнеров. При этом увеличение совокупной длины ПСП возрастает при увеличении n, но среднее количество вкраплений остается неизменным. Однако для объемных сцен это условие не выполняется. В случае увеличения числа контейнеров значение n остается неизменным, т. е. увеличивается длина ПСП и среднее число вкраплений (они возрастают линейно). Это приводит к нарушению условия неразличимости исходного и заполненного контейнеров, т. е. необходимо проведение стегоанализа. Если по его результатам неразличимость не будет установлена, требуется проведение криптоанализа.

3.1. Стегостойкость. Стегостойкость определялась экспериментально путем применения к ПСП набора статистических тестов NIST. При успешном прохождении тестирования по NIST (по всем 15 тестам, входящим в набор), ПСП считается истинно случайной или «белой». Если хотя бы один тест не пройден,

то ПСП признается «черной». Получены оценки вероятности генерации случайных («белых») ГАММ для формирования пустых контейнеров и получения на их основе «белых» стегоконтейнеров путем вставки по маске истинной информации. По результатам проведенных экспериментов установлена безусловная стегостойкость ассоциативной защиты при произвольном выборе ГАММЫ.

Если предположить, что при передаче информации пересылки пустых и стегоконтейнеров имеют равную вероятность, то получаем равенство априорных и апостериорных вероятностей (как при передаче «белых» и «черных» ГАММ, так и при передаче соответствующих стегоконтейнеров), т. е. выполняется критерий совершенной секретности К. Шеннона. Однако его выполнение не является обязательным, т. к. выбор «белой» ГАММЫ при формировании стегоконтейнера позволяет получить безусловную стегостойкость.

Дополнительно был проведен стегоанализ при действии следующих атак на стегоконтейнер (в скобках приведены соответствующие им криптоатаки):

A. Атака на основе известного заполненного контейнера (атака с использованием только шифртекста);

B. Атака на основе известного встроенного сообщения (атака с использованием выбранного открытого текста);

C. Атака разрушения или подмены скрытого сообщения.

В первом случае определяющую роль сыграло число всевозможных ключей для используемого метода. Вычислительная стойкость современных систем сокрытия информации определяется невозможностью полного перебора ключей за приемлемое время для выявления истинного ключа даже с применением суперкомпьютерных технологий. Согласно проведенной оценке числа ключей вероятность случайного подбора истинного ключа при $n = 60$ составляет 10^{-29}. Она ничтожна.

Во втором случае было проведено исследование, в котором открытый текст был предстален 10-разрядным кодом (1 2 3 4 5 6 7 8 9 0) в формате почтовых индексов. При проведении исследования считалось, что противнику известно соответствующее открытому тексту стегосообщение. При заданных условиях оценивалось количество ключей (наборов масок), которые потенциально могли быть использованы при генерации данного стегосообщения. Для этого осуществлялся поиск бит, которые могли быть использованы в качестве дихотомальных при работе АЛГОРИТМА. Эксперимент проводился при $n = 60$.

Количество найденных наборов масок (ключей) составило 10^{14}. При этом полное множество ключей – 10^{29}. Если принять время испытания одного ключа (с учетом выявления семантики кодируемых сущностей) равным 1 мкс, то полный перебор в этом случае займет более 6 лет. Тем не менее потенциально существует возможность успешного проведения данной атаки. Однако использование 10-разрядного кода маловероятно, т. к. на практике применяется существенно меньшее число разрядов. В третьем случае подмена или разрушение противником стегосообщения эквивалентны воздействию помех. Тогда использование избыточного числа ключей приведет к отказу у распознавания. Таким образом, будет детектирована попытка подмены или разрушения сообщения.

3.2. Проведенный криптоанализ. Он был ограничен случаем воздействия трех характерных криптоатак: «лобовой» (со знанием только шифртекста) и со знанием открытого текста, которые уже были рассмотрены в п. 3.1, а также на ГАММУ. В рамках атаки на ГАММУ были рассмотрены 2 вопроса: 1) можно ли при

знании отрезка ГАММЫ, примененной при создании стегоконтейнеров, сформировать ключевой набор масок; 2) какова временная сложность экспериментального определения истинной ГАММЫ при использовании ГПСП «Вихрь Мерсенна»?

Идея рассмотрения первого вопроса состояла в выявлении истинных бит и формировании матриц масок на основе найденной совокупности координат этих бит. Каждому конкретному эталону ставилась в соответствие та или иная маска из найденного набора. Установлено, что эксперимент закончится успешно при числе стегоконтейнеров, равном 33.

Получение ответа на второй вопрос связано с выявлением принципиальной возможности нахождения истинного отрезка ГАММЫ, использованного при рандомизации. Для стегоконтейнера длины L случайным образом выбиралось начальное состояние ГПСП. В сгенерированной ПСП последовательно выделялись окна длиной L со сдвигом на 1 бит. Далее выполнялась проверка истинности ГАММЫ путем ее поразрядного суммирования по модулю 2 с контейнером. Если число единиц в результате больше 24, процедура с окнами продолжалась. Иначе найденная ГАММА полагалась истинной. По алгоритму и программе, разработанным в соответствии с этой методикой, в течение суток удалось провести всего лишь $2,88 \times 10^9$ сдвигов окна. Истинной ГАММЫ выявлено не было. Поскольку длина периода использованного ГПСП равна $2^{19937} - 1$, то просмотр всех указанных окон за приемлемое время невозможен.

Таким образом, проведенные исследования позволяют установить доказуемую стойкость ассоциативного подхода. Данное свойство ранее было установлено только для шифрования методом гаммирования.

Заключение

В статье дано понятие ассоциативной защиты, рассмотрен АЛГОРИТМ маскирования, сформулирована теорема, положенная в основу всего ассоциативного подхода. Приведены результаты исследований по помехоустойчивости к действию случайных и преднамеренных помех, а также по стего- и криптостойкости.

Практическое применение ассоциативного подхода возможно для защиты данных картографии и текстовых сцен [6, 7]. Мыслимы и иные сферы применения.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Список литературы

1. *Дуда Р., Харт П.* Распознавание образов и анализ сцен. М.: Мир, 1976. 511 с.
2. *Коржик В.И., Небаева К.А., Алексеев М.* Использование модели канала с шумом для построения стегосистемы // Телекоммуникации. 2013. № S7. С. 33–36.
3. *Райхлин В.А., Вершинин И.С., Глебов Е.Е.* К решению задачи маскирования стилизованных двоичных изображений // Вестн. КГТУ им. А.Н. Туполева. 2001. № 1. С. 42–47.
4. *Морелос-Сарагоса Р.* Искусство помехоустойчивого кодирования. М.: Техносфера, 2006. 320 с.
5. *Ker D.A.* A capacity result for batch steganography // IEEE Signal Process. Lett. 2007. V. 14, No 8. P. 525–528. https://doi.org/10.1109/LSP.2006.891319.
6. *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 J. Math. 2013. V. 34, No 4. P. 319–325. https://doi.org/10.1134/S1995080213040112.

7. *Vershinin I.S., Gibadullin R.F., Pystogov S.V., Raikhlin V.A.* Associative steganography of text messages // Moscow Univ. Comput. Math. Cybern. 2021. V. 45, No 1. P. 1–11. https://doi.org/10.3103/S0278641921010076.

Поступила в редакцию 23.06.2024
Принята к публикации 15.08.2024

Вершинин Игорь Сергеевич, кандидат технических наук, доцент, заведующий кафедрой компьютерных систем
Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ
ул. К. Маркса, д. 10, г. Казань, 420111, Россия
E-mail: *ISVershinin@kai.ru*

ISSN 2541–7746 (Print)
ISSN 2500–2198 (Online)

**UCHENYE ZAPISKI KAZANSKOGO UNIVERSITETA.
SERIYA FIZIKO-MATEMATICHESKIE NAUKI
(Proceedings of Kazan University. Physics and Mathematics Series)**

2024, vol. 166, no. 3, pp. 320–330

ORIGINAL ARTICLE
doi: 10.26907/2541-7746.2024.3.320-330

Elements of the Theory of Associative Information Protection

I.S. Vershinin

Kazan National Research Technical University named after A.N. Tupolev – KAI, Kazan, 420111 Russia

E-mail: *ISVershinin@kai.ru*

Received June 23, 2024; Accepted August 15, 2024

Abstract

This article explores elements of the theory of a new scientific field – associative protection of information during storage and transmission. The approach under study is shown to enhance the level of protection and noise immunity in scene analysis.

Keywords: associative protection, scene analysis, noise immunity, cryptoresistance, stegoresistance

Conflicts of Interest. The authors declare no conflicts of interest.

Figure Captions

Fig. 1. Entry structure in the scene data table.
Fig. 2. The digit 9.
Fig. 3. Significant bits.
Fig. 4. Block diagram of the ALGORITHM.
Fig. 5. Block diagram of the ALGORITHM (continued from Fig. 4).
Fig. 6. Result of the ALGORITHM.

References

1. Duda R., Hart P. *Raspoznavanie obrazov i stsen* [Pattern Classification and Scene Analysis]. Moscow, Mir, 1976. 511 p. (In Russian)
2. Korzhik V.I., Nebaeva K.A., Alekseev M. Using the noisy channel model for stegosystem design. *Telekommunikatsii*, 2013, no. S7, pp. 33–36. (In Russian)
3. Raikhlin V.A., Vershinin I.S., Glebov E.E. On solving the problem of masking of conventionalized dual images. *Vestn. KGTU im. A.N. Tupoleva*, 2001, no. 1, pp. 42–47. (In Russian)
4. Morelos-Zaragoza R. *Iskusstvo pomekhoustoichivogo kodirovaniya* [The Art of Error Correcting Coding]. Moscow, Tekhnosfera, 2006. 320 p. (In Russian)
5. Ker D.A. A capacity result for batch steganography. *IEEE Signal Process. Lett.*, 2007, vol. 14, no. 8, pp. 525–528. https://doi.org/10.1109/LSP.2006.891319.
6. 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 J. Math.*, 2013, vol. 34, no. 4, pp. 319–325. https://doi.org/10.1134/S1995080213040112.
7. Vershinin I.S., Gibadullin R.F., Pystogov S.V., Raikhlin V.A. Associative steganography of text messages. *Moscow Univ. Comput. Math. Cybern.*, 2021, vol. 45, no. 1, pp. 1–11. https://doi.org/10.3103/S0278641921010076.

⟨ ***Для цитирования*:** *Вершинин И.С.* Элементы теории ассоциативной защиты информации // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. 2024. Т. 166, кн. 3. С. 320–330. https://doi.org/10.26907/2541-7746.2024.3.320-330. ⟩

⟨ ***For citation*:** Vershinin I.S. Elements of the theory of associative information protection. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2024, vol. 166, no. 3, pp. 320–330. https://doi.org/10.26907/2541-7746.2024.3.320-330. (In Russian) ⟩