# МЕТОД СТЕГАНОАНАЛИЗА СТАТИЧЕСКИХ ИЗОБРАЖЕНИЙ ФОРМАТА JPEG НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

*Шниперов А.Н.[1], Прокофьева А.В.[2]*

***Целью** настоящего исследования является разработка метода стеганоанализа статических изображений формата JPEG, основанного на применении искусственных иммунных систем.*

***Метод исследования:** эвристический метод с использованием эволюционных алгоритмов и элементов методов обучения с подкреплением.*

***Полученный результат.***

*Спроектирована и разработана модель искусственной иммунной системы для задачи обнаружения скрытой информации в изображениях формата JPEG, а именно: определены базовые требования и рассмотрены основные элементы искусственной иммунной системы, введены операции мутации и клонирования антител, а также приведено формальное описание на псевдоязыке реализации основных узлов искусственной иммунной системы с последующей реализацией алгоритмов. Кроме того, в статье приводится краткий обзор и анализ состояния проблематики стеганоанализа, а также анализ полученных экспериментальных результатов и оценка эффективности разработанного метода. Предложенный метод позволяет детектировать наличие скрытой информации, внедренной различными популярными инструментами стеганографии в статические изображения формата JPEG с достаточно высокой точностью. Теоретическая значимость данной работы состоит в развитии достаточно перспективного подхода эвристического стеганоанализа с использованием искусственных иммунных систем. Практическая значимость заключается в разработанном программном продукте, а также в экспериментальных данных, подтверждающих эффективность метода стеганоанализа в отношении детектирования скрытой информации в изображениях формата JPEG.*

***Ключевые слова:** Стеганография, Steghide, OutGuess, F5, бинарная классификация, вейвлет-преобразование Хаара, алгоритм клонального отбора, алгоритм отрицательного отбора.*

**DOI:** 10.21681/2311-3456-2020-2-22-31

## Введение

Цифровые изображения формата JPEG наиболее широко распространены в сети Интернет, а их повседневный оборот представляет весьма значительную долю интернет-трафика, включая социальные сети, мессенджеры, порталы по обмену изображениями и другие ресурсы. Высокая популярность данного формата изображений стала одной из причин довольно быстрого появления новых методов скрытой передачи информации, где в качестве стеганоконтейнера выступает само изображение. Так, согласно статистике, предоставленной компанией Wetstone Technologies, по состоянию на март 2014 года в общей сложности существует более 300 приложений, которые позволяют скрывать данные в изображениях формата JPEG [1].

Весьма серьезной проблемой является использование таких средств скрытой передачи информации в противоправных целях (в том числе и террористических), а также для обхода мониторинга систем предотвращения утечек конфиденциальной информации (DLP-систем). В последнее время разработчики данных систем начали активно уделять внимание данной проблеме и внедрять соответствующие инструменты. Однако задача стеганоанализа является весьма сложной, а ее решение требует еще многих исследований в данной области. Вследствие чего можно говорить о высокой актуальности задачи обнаружения скрытых каналов передачи информации и разработке методов обнаружения их.

Данная работа посвящена применению концепции искусственной иммунологии в решении задачи стеганоанализа, т.к. искусственные иммунные системы вобрали в себя лучшие особенности биоинспирированных методов, такие как динамическое расположение элементов из эволюционных алгоритмов и принципы обучения из искусственных нейронных сетей. В статье предлагается новый метод стеганоанализа статических изображений формата JPEG, позволяющий детектировать наличие скрытой информации в них с достаточно высокой точностью, внедренной различными популярными инструментами стеганографии.

1 Шниперов Алексей Николаевич, кандидат технических наук, доцент научно-учебной лаборатории «Информационная безопасность» кафедры прикладной математики и компьютерной безопасности Сибирского Федерального Университета, г. Красноярск, Россия. E-mail: ashniperov@sfu-kras.ru

2 Прокофьева Александра Владимировна, инженер отдела аппаратных комплексов защиты информации ООО ИК «СИБИНТЕК», г. Красноярск, Россия. E-mail: prokofe-aleksandra@yandex.ru

**Обзор методов стеганоанализа jpeg изображений**

Существует множество методов стеганоанализа, которые различаются по используемым характеристикам изображения и методам встраивания, которым они противодействуют. В зависимости от используемых исходных данных методы стеганоанализа традиционно разделяют на сигнатурные, статистические и эвристические.

Сигнатурные методы стеганоанализа предназначены для работы с форматными методами сокрытия информации, которые в процессе сокрытия оставляют специфические маркеры (сигнатуры), по которым и удаётся детектировать скрытое вложение.

Статистические методы стеганоанализа основываются на анализе статистических характеристик исследуемого изображения с целью установления, как они коррелируют с характеристиками пустых стеганоконтейнеров такого же типа. Наиболее известными статистическими методами являются *RS*-стеганоанализ и *WS*-стеганоанализ [2], гистограммный [3], *SPAM* (*subtractive pixel adjacency matrix*) стеганоанализ [4] и другие подходы. Данные методы могут показывать очень высокую чувствительность к обнаружению заполненного стеганоконтейнера и даже определению объёма скрываемой в нём информации, однако их точность в значительной степени зависит от алгоритма встраивания.

Эвристические методы стеганоанализа представляют большой интерес для исследователей, т.к. они более универсальны, поскольку не привязаны к какому-то алгоритму внедрения скрытой информации, хоть и несколько менее точны в целом. В основном, данные методы базируются на решении задачи бинарной классификации с применением методов машинного обучения, например, методы, предложенные в работах [5-10]. Рассмотрим некоторые их них подробнее, так, в работе [6] приводится метод стеганоанализа, основанный на анализе гистограмм, построенных на основе таблицы кодов Хаффмана, использующихся для кодирования значений дискретного косинусного преобразования (ДКП) кодами переменной длины. Для анализа гистограмм применяется машинное обучение с использованием искусственной нейронной сети. По выводам авторов, данный метод позволяет детектировать заполненные стеганоконтейнеры, полученные с помощью двух алгоритмов: Steghide и OutGuess, с точностью от 95.4% до 98.8%. Метод достигает большей точности на изображениях большого размера (4200 × 2358 пикселей).

В работе [9] предлагается алгоритм стеганоанализа, также основанный на сегментации изображений, но формируемые фрагменты образуются в соответствии со сложностью текстуры. В качестве вектора характеристик изображений используется набор PEV-274, предложенный в работе [11] и в настоящее время является распространенным в системах стеганоанализа. Задача классификации решается посредством применения метода опорных векторов. Точность метода, по оценкам авторов, составляет от 85 до 97% для алогитма JPHide, от 67 до 77% – для алгоритма F5 и всего 57-62% для алгоритма PQ.

Отдельным, достаточно интересным и перспективным, направлением развития эвристических методов стеганоанализа можно выделить искусственные иммунные системы (ИИС), биологическим прототипом которых является иммунная система живых организмов. Основной функцией иммунной системы является обнаружение и нейтрализация чужеродных объектов (антигенов) к которым относятся, например, бактерии и вирусы. Антигены провоцируют иммунный ответ организма, который начинает вырабатывать защитные клетки иммунной системы – антитела различных видов, пока не найдётся такое антитело, которое специфично связывается с антигеном и нейтрализует его, обеспечивая тем самым естественную защиту организма. Совокупность сформированных в течение жизни антител формируют иммунитет организма. В свою очередь, ИИС является неким функциональным аналогом иммунной системы, способной к обучению и являющейся децентрализованной распределенной системой обработки и анализа информации [12]. Применение ИИС для решения задач стеганоанализа является сравнительно новым, однако за последние несколько лет уже опубликован ряд работ в этой области.

Так, в работе [13] авторы строят ИИС, которая оперирует векторами характеристик изображений, формирующимися с помощью итеративного применения вейвлет-преобразования Хаара, результирующая матрица которого имеет следующий вид:

$$\begin{pmatrix} AC & HC \\ VC & DC \end{pmatrix} \quad (1)$$

где *AC* – субдиапазон приближения, а *HC, VC*; *DC* – горизонтальный, вертикальный и диагональный субдиапазоны изображения. Для составления векторов характеристик авторы используют только группы коэффициентов *HC*, *VC* и *DC* для каждого цветового канала RGB-модели, в результате чего получается вектор, состоящий из 36 значений. Такие векторы характеристик изображений из обучающей выборки, к каждому из которых добавляются два значения евклидова расстояния от него до внутренней и внешней среды, образуют множество антител. Авторы выделяют три основных концепции, на которых строится искусственная иммунная система:

1. Представление компонентов в системе: внутренняя среда соответствует пустым стеганоконтейнерам; внешняя среда – заполненным стеганоконтейнерам; антителами являются векторы характеристик, используемые для детектирования заполненных стеганоконтейнеров.
2. Механизм обнаружения антигенов (заполненных контейнеров): основывается на нахождении евклидова расстояния между каждым антителом и внутренней средой, и между каждым антителом и внешней средой.

DOI:10.21681/2311-3456-2020-2-22-31

$$A(ab, env) = \min\left(\sqrt{\sum_{j=1}^{n}(ab_j - env_{i_j})^2}\right), \quad (2)$$

где n = 36, $ab_j$ – представляет собой элемент вектора антитела, $env_i$ – элемент, принадлежащий внутренней или внешней среде (т. е. пустой стеганоконтейнер для внутренней среды и заполненный – для внешней среды).

3. Набор процедур обучения, которые позволяют сгенерировать наиболее эффективные антитела. Авторами предлагается использовать алгоритм отрицательного отбора: если антитело реагирует на пустой стеганоконтейнер как на заполненный (ошибка 1-го рода), то оно уничтожается и вместо него случайным образом генерируется другое антитело.

Точность обнаружения, заявленная авторами, составила от 67 до 90%. Заявляемое время проведения одного теста (то есть обработки 6 тысяч изображений) составляет всего – 3-5 минут.

Данный подход к созданию ИИС был выбран нами для дальнейшего исследования и развития в качестве базового, поскольку он является достаточно универсальным (по отношению к алгоритмам внедрения скрытой информации), по оценкам авторов показывает достаточно неплохие результаты по времени обработки изображения в сравнении с другими методами, а также позволяет достичь хороших результатов точности классификации.

### Постановка задачи и описание предлагаемого метода

Практическая реализация искусственой имунной системы, описанной в работе [13], а также серия экспериментов показала, что показатели точности решения задачи классификации находятся в крайне высокой зависимости от объёма обучающей выборки. Таким образом, для реального применения данного подхода к построению ИИС на практике требуется создать очень большую выборку, содержащую как можно больше изображений, т.к. в случаях, когда анализируемое изображение отсутствует в обучающей выборке, точность детектирования скрытой информации в нём (для заполненного стеганоконтейнера) составляет приблизительно 50% и задача бинарной классификации не может быть решена. При этом точность значительно повышается, если изображение добавить в обучающую выборку. Кроме того, необходимо отметить, что данный подход работоспособен только в случае работы с квадратными изображениями.

Таким образом, сформировалась научно-техническая задача – разработать подход к построению обучаемой ИИС, способной к детектированию скрытой информации в изображениях формата JPEG, отсутствующих в обучающей выборке.

Сформулируем общую задачу разрабатываемой ИИС. Пусть $I = C \cup S$ – множество объектов заданного типа (изображений формата JPEG), *S* – множество заполненных стеганоконтейнеров, каждый из которых содержит скрытую информацию, *C* – множество пустых стеганоконтейнеров, не содержащих скрытой информации, полагаем $S \cap C = \varnothing$. Каждый из объектов $img \in I$ представлен вектором *D* его характеристик. Общая постановка задачи стеганоанализа изображения $img \in I$ заключается в решении задачи бинарной классификации $def : img \to S$ искусственной иммунной системой, т.е. детектированию скрытой вложенной информации в изображении. При этом ИИС рассматривается как некоторая система, способная распознать «свой» объект, которым является пустой контейнер *C*, от «чужеродного», которым является заполненный контейнер *S*.

### Получение вектора характеристик изображения

Поскольку одним из недостатков базового метода [9] является высокая зависимость от обучающей выборки, было принято решение ввести дополнительные преобразования изображения (калибровку) для получения его вектора характеристик. Необходимо отметить, что калибровка изображения довольно часто используется в различных методах стеганоанализа, например, в работах [9] и [11]. Общая схема этапа получения вектора характеристик изображения представлена на (рис. 1).

На первом шаге функции калибровки анализируемое изображение *img*, размерностью $n \times m$ пикселей формата JPEG переводится в пространственную область с использованием функции *IDCT* – обратного ДКП-преобразования. Далее изображение обрезается на четыре пикселя с двух сторон и повторно сжимается с использованием матрицы квантования исходного изображения *img*. Калибровка позволяет достаточно эффективно подавить влияние JPEG-сжатия исходного изображения *img*, а также потенциально возможного внедрённого в него скрытого вложения, на ДКП-

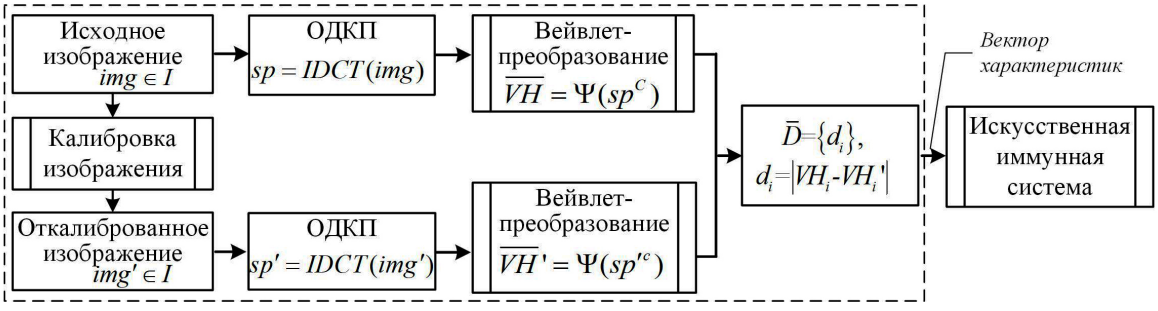


*Рис. 1. Общая схема получения вектора характеристик изображения*

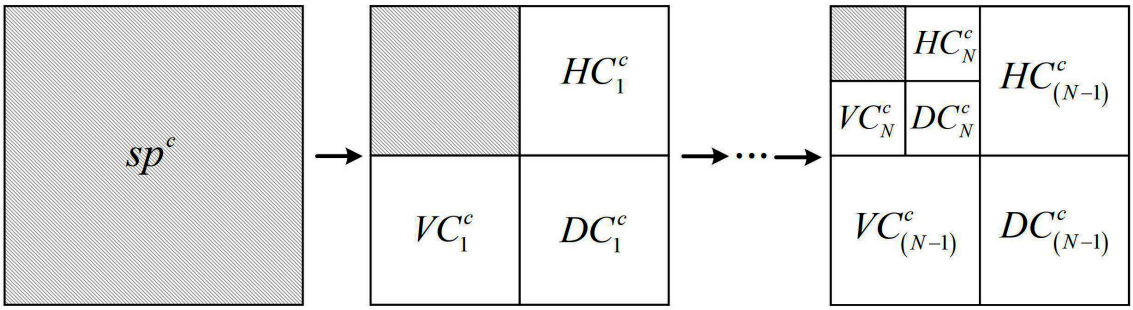


*Рис. 2. Схема итеративного применения ДВП*

коэффициенты откалиброванного изображения $img'$. Таким образом, вектор характеристик откалиброванного изображения $img'$ является некоторым отражением статистических свойств пустого стеганоконтейнера для анализируемого изображения *img*.

Далее исходное и откалиброванное изображения переводятся в пространственную область: $sp = IDCT(img)$, $sp' = IDCT(img')$ где *IDCT* – обратное дискретное косинусное преобразование. Для каждого цветового канала итеративно применяется вейвлет-преобразование Хаара:

$$\overline{VH} = \Psi(sp^c) = \left\langle HC^c, VC^c, DC^c \right\rangle$$

и $\overline{VH}' = \Psi(sp^c{}') = \left\langle HC^c{}', HV^c{}', DC^c{}' \right\rangle$,

где $c \in \{r, g, b\}$ – соответствующий цветовой канал. Общий принцип итеративного применения вейвлет-преобразования для каждого из цветовых каналов показан на (рис. 2).

На первой итерации на вход дискретного вейвлет-преобразования (ДВП) Хаара подается цветовой канал исходного изображения, результатом преобразования являются четыре результата разложения вдвое меньших размеров (т. е. приближение, горизонтальная, вертикальная и диагональная детальные части), которые расположены в соответствии с (рис. 2). Аналогично производятся и последующие итерации, но на вход ДВП подается субдиапазон приближения, полученный на предыдущей итерации и имеющий вдвое меньший масштаб. В итоге для квадратных изображений на последней итерации получаем матрицу ДВП-коэффициентов размером 2×2, для прямоугольных изображений с горизонтальной ориентацией – 2×3, а с вертикальной ориентацией – 3×2. При этом общее количество итераций вейвлет-преобразования N зависит от размера изображения и определяется:

$$N = \left\lfloor \log_2 \left( \min\{n, m\} \right) \right\rfloor, \qquad (3)$$

где *n* – ширина изображения (в пикселях), а *m* – высота изображения.

На последней итерации вейвлет-преобразования получаем 4 группы коэффициентов для каждого цветового канала изображения модели RGB, из которых значимыми являются группы $HC^c$, $VC^c$ и $DC^c$. Далее необходимо сгруппировать результаты вейвлет-преобразований в вектора $\overline{VH}$ и $\overline{VH}'$. Число компонент в векторе коэффициентов ДВП будет следующим:

$$k = \left| \left\langle HC^C, VC^C, DC^C \right\rangle \right| \times \left| \left\langle R, G, B \right\rangle \right| \times u \times v, \qquad (4)$$

где $\left| \left\langle HC^C, VC^C, DC^C \right\rangle \right|$ – мощность множества, состоящего из групп коэффициентов, получаемых в результате вейвлет-преобразования, где $HC^c$ – горизонтальный субдиапазон; $VC^c$ – вертикальный; $DC^c$ – диагональный; R, G, B – красный, зеленый и синий цветовые каналы RGB-модели, $u \times v$ – размер матрицы на последней итерации вейвлет-преобразования.

Таким образом, квадратное изображение будет представлено вектором, состоящим из 36 значений, а прямоугольное изображение – вектором из 54 значений:

$$\overline{VH} = \left\langle HC^C, VC^C, DC^C \right\rangle, \text{ где}$$

$$\begin{aligned}
HC^C &= \left\langle HC_R^C, HC_G^C, HC_B^C \mid HC_R^C = \left\{ hc_{R_i}^C \right\}, \right. \\
HC_G^C &= \left\{ hc_{G_i}^C \right\}, HC_B^C = \left\{ h_{B_i}^C \right\} \Big\rangle \\
VC^C &= \left\langle VC_R^C, VC_G^C, VC_B^C \mid VC_R^C = \left\{ vc_{R_i}^C \right\}, \right. \\
VC_G^C &= \left\{ vc_{G_i}^C \right\}, VC_B^C = \left\{ v_{B_i}^C \right\} \Big\rangle \\
DC^C &= \left\langle DC_R^C, DC_G^C, DC_B^C \mid DC_R^C = \left\{ dc_{R_i}^C \right\}, \right. \\
DC_G^C &= \left\{ dc_{G_i}^C \right\}, DC_B^C = \left\{ dc_{B_i}^C \right\} \Big\rangle
\end{aligned} \qquad (5)$$

$i = [0, \ldots, 3]$, если *img* – квадратное изображение, $i = [0, \ldots, 5]$, если *img* – прямоугольное.

Итоговый вектор характеристик изображения находим следующим образом:

$$\bar{D} = \left\{ d_i \mid \forall d_i = \left| VH_i - VH'_i \right| \right\}, \qquad (6)$$

$$i \in (0, \ldots, k),\ d_i \in R$$

где *k* находится по формуле (4), $VH_i$ – *i*-ая компонента вектора коэффициентов ДВП исходного изображения img, $VH'_i$ – i-ая компонента вектора коэффициентов ДВП откалиброванного изображения $img'$.

**Инициализация искусственной иммунной системы**

Пусть собственными клетками ИИС будет считать пустые стеганоконтейнеры, чужеродными (антигенами) – заполненные. Первоначальный набор антител будем получать не случайным образом, как предлагали авторы в работе [13], а использовать в качестве них набор антигенов (заполненных стеганоконтейнеров) из обучающей выборки, как было предложено группой исследователей во главе с Т. Лу в статье [14]. Данный подход напоминает процесс вакцинации, позволяющий научить иммунную систему сопротивляться антигенам без развития болезни. Для того чтобы разрабатываемая иммунная система могла работать в реальных условиях, необходимо создать два набора антител: $\{A_{36}\}$ и $\{A_{54}\}$, отдельно для квадратных изображений, отдельно для прямоугольных, поскольку приведение к пространству одной размерности неизбежно приведет к появлению дополнительного шума в изображениях, а следовательно, и к снижению точности. Каждое антитело в первоначальном наборе представляет собой набор векторов характеристик изображений $\{D_1,...,D_N\}$ из обучающей выборки, являющихся заполненными контейнерами.

Для построения ИИС необходимо определить способ представления компонентов системы, механизмы адаптации ИИС к изменениям системы во времени, а также механизм, позволяющий оценить степень схожести генетических наборов антигенов и антител, то есть оценить взаимодействие компонентов системы.

Взаимодействие между антителами и антигенами можно описать геометрически, с помощью *k*-мерного пространства, в котором измерения соответствуют набору характеристик изображения, используемого для оценки взаимодействия антител и антигенов, где *k* определяется по формуле (4). Таким образом, векторы характеристик изображений могут рассматриваться как точки в *k*-мерном пространстве.

Антитела в ИИС представляются векторами из (*k*+1) значений: к вектору характеристик изображения $\overline{D}$ добавляться значение r, которое представляет собой радиус окрестности точки (т.е. антитела) в *k*-мерном пространстве:

$$\overline{Ig} = \left\langle \overline{D}, r \right\rangle \qquad (7)$$

Концепция ИИС основывается на том, что если вектор характеристик анализируемого изображения попадает в окрестность хотя бы одного антитела, то такое изображение будет отнесено к множеству заполненных стеганоконтейнеров. В противном случае, к множеству пустых стеганоконтейнеров. Поскольку компоненты векторов характеристик изображений представляют собой вещественные числа, взаимодействие компонентов системы оценивается с помощью евклидова расстояния. Так как первоначальный набор антител получаем из набора заполненных стеганоконтейнеров S из обучающей выборки, *r* определим как минимальное евклидово расстояние от вектора характеристик заполненного контейнера из обучающей выборки $\boldsymbol{s} \in S$ до векторов характеристик пустых стеганоконтейнеров, принадлежащих обучающей выборке $lc_i \in LC$:

$$r = \min\left\{d(\boldsymbol{s}, lc_i) \mid i \in (0,...,N_{LC})\right\}, \qquad (8)$$

где $N_{LC}$ – число пустых стеганоконтейнеров в обучающей выборке; *d* – евклидово расстояние.

Заметим, что у собственных клеток ИИС также есть радиус окрестности $-0 \le r_c < 0{,}1$, значение которого задается статически и одинаково для всех собственных клеток ИИС. Данный параметр необходим для того, чтобы учесть появление незначительных изменений в пустых стеганоконтейнерах, появившихся, например, в результате преобразования изображения из одного цветового пространства в другое, или в результате искажений помехами, возникающими в каналах связи. Другими словами, параметр позволяет уменьшить ошибку первого рода.

Для обучения ИСС используются алгоритмы, основанные как на теории клонального отбора, так и на теории отрицательного отбора, которые применяются последовательно: отрицательный отбор применяется при инициализации ИИС, а клональный – при обучении.

Алгоритм инициализации, предлагаемой в работе ИИС, основывается на теории отрицательного отбора и в общем виде выглядит следующим образом.

[Шаг 1]: Пока не получено требуемое начальное множества антител $A_0$ для каждого вектора характеристик заполненного стеганоконтейнера из обучающей выборки $\boldsymbol{s}_i \in S$ выполнить шаги 2 – 6.

Шаг 2]: Изначально r принимаем бесконечно большим: $r \leftarrow \infty$.

[Шаг 3]: Для каждого вектора характеристик пустого стеганоконтейнера из обучающей выборки $\boldsymbol{lc}_j \in LC$ выполнить шаги 4 – 5.

[Шаг 4]: Вычислить евклидово расстояние d между $s_i$ и $lc_j$.

[Шаг 5]: Если $\left(d - r_c\right) \le r$, то $r = d - r_c$.

[Шаг 6]: Если $r > r_c$, то $A_0 \leftarrow A_0 \cup \left\langle \boldsymbol{s}_i, r \right\rangle$. В противном случае, данное антитело уничтожается, т.к. такие антитела будут вызывать аутоиммунную реакцию и приведут к значительному увеличению величины ошибки первого рода. Вместо уничтоженного антитела генерируется случайный вектор из 36-ти или 54-х значений, после чего переходим на шаг 2.

**Обучение иммунной системы**

Для обучения и тестирования ИИС была подготовлена база, состоящая из 7,5 тысяч цветных изображений формата JPEG различного размера: от $1024 \times 512$ до

$4800 \times 4888$ пикселей. Источниками изображений послужили:

- выборки изображений с платформы «Kaggle»[3], посвященной анализу данных и машинному обучению. Для создания базы для обучения и тестирования искусственной иммунной системы использовались выборки: «Natural Images», «Fruits 360 dataset», «Cats», «Test2015».
- различные фотографии, полученные с помощью цифровых камер (Canon, Kodak). Выборка включает в себя фотографии естественных ландшафтов, зданий, животных и растений.

Далее эти изображения были поделены на две группы: обучающую и тестовую выборки, мощностью по 3,75 тысяч изображений в каждой. Первая половина каждой из групп оставалась неизменной и образовывала собой множество пустых стеганоконтейнеров. Над второй половиной производилось встраивание скрытого сообщения с помощью инструментов стеганографии Steghide, OutGuess и F5, которые являются наиболее популярными, достаточно стойкими и используют неформатные методы сокрытия информации в графических изображениях.

Обучение в ИСС обеспечивается увеличением относительного размера популяции тех антител, которые доказали свою ценность при распознавании, посредством механизмов клонального отбора, которая предполагает мутации антител. В ходе экспериментов были апробированы несколько вариантов мутаций, в том числе случайные незначительные (в пределах от 0,001 до 0,01) изменения компонент векторов антител, перестановки элементов в векторах в пределах только одного из блоков $HC^c$, $VC^c$, $DC^c$ при неизменности остальных. Этот вариант подобен тому, как происходят мутации антител в иммунной системе живых организмов. В ходе экспериментов лучшие результаты показал вариант мутаций, основанный на перестановках компонент векторов антител в пределах одного блока, поскольку в данном случае достигается наименьшее число ложных срабатываний.

Клональный отбор следует производить итерационно, поскольку одиночные изменения антигенов незначительно влияют на свойства системы. Следовательно, с увеличением числа поколений мутации (итераций цикла клональной селекции) повышается точность определения неизвестных антигенов. Общая схема одного поколения мутации приведена на (рис. 3).

Обучение ИИС также производится дважды: отдельно для наборов антител, полученных на этапе инициализации иммунной системы и получении начального набора антител, для квадратных изображений и прямоугольных изображений.

Таким образом, каждый этап обучения представим следующим алгоритмом. На вход алгоритма подается: начальный набор антител $A_0$, $N$ – число антител в наборе. В качестве результата работы алгоритма получим рабочий набор антител $A_w$.

Шаг [1]. Пока не пройдено требуемое количество поколений мутации $P$ повторять шаги 2–8. Заметим, что значение $P=10$ поколений мутации уже позволяет достичь неплохих показателей точности.

Шаг [2]. Распознавание антигена. Иммунная система с помощью начального набора антител $A_0$ решает задачу классификации $def : img \to S$ для изображения из обучающей выборки.

Шаг [3]. Вычисление аффинности. На данном шаге происходит отбор наиболее эффективных антител на основе величины аффинности $Af_i$, определяемой отдельно для каждого антитела $Ig_i \in A_0$ и равной числу таких заполненных стеганоконтейнеров из обучающей выборки, которые будут отнесены к множеству заполненных, благодаря этому антителу:

$$Af_i = \sum_{j=0}^{N} \begin{cases} 1, \text{ если } Ig_i \text{ относит } img_j \\ \text{к заполненным контейнерам} \\ 0, \text{ иначе} \end{cases} \quad (9)$$

Шаг [4]. Клонирование (поэлементное копирование в памяти) антител с наибольшей аффинностью. Причем число клонов антитела прямо пропорционально величине его аффинности.

Шаг [5]. Мутация антител. Небольшие случайные изменения векторов антител позволяют достичь более высокого соответствия к распознаваемому антигену. Степень мутации обратно пропорциональна величине аффинности (чем выше аффинность родительской клетки, тем в меньшей степени они подвергаются мутации, и наоборот).

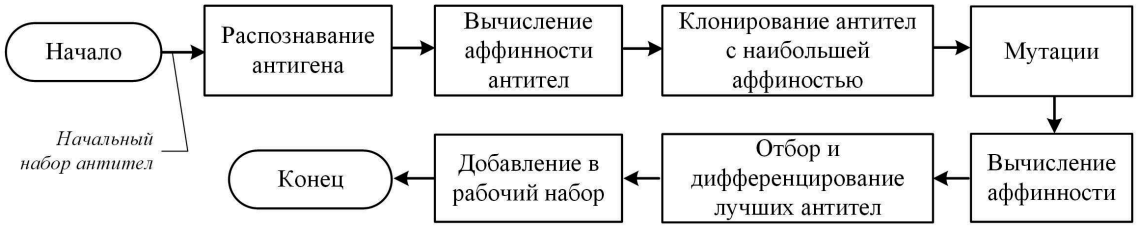


*Рис. 3. Общая схема одной итерации алгоритма клональной селекции*

3 Goldbloom A., Hamner B. Datasets | Kaggle [Электронный ресурс]. 2019. URL: https://www.kaggle.com/datasets.

Шаг [6]. Для полученных на предыдущем шаге антител согласно выражению (9) вычисляется их аффинность.

Шаг [7]. Отбор и дифференцирование лучших антител на основе значения их аффинности антител. Если аффинность модифицированного антитела больше, чем у его прообраза, то оно проходит отбор, в противном случае – уничтожается. Уничтожение антител необходимо производить для уменьшения ошибки второго рода.

Шаг [8]. Добавление антител, прошедших отбор на предыдущем шаге, в рабочий набор антител $A_W \leftarrow A_W \cup Ig_{new}$.

Таким образом, во время обучения ИИС происходит итеративный процесс воспроизведения новых антител из лучших представителей предыдущего поколения и в итоговый набор добавляются только те антитела, которые максимально подходят под найденный антиген [15].

### Классификации изображения искусственной иммунной системой

Основа работы ИИС строится на том, что если вектор характеристик анализируемого изображения попадает в окрестность хотя бы одного антитела, то такое изображение будет отнесено к множеству заполненных стеганоконтейнеров. В противном случае, к множеству пустых.

На данном этапе в ИИС подается некоторое, нерассмотренное ранее, изображение $img \in I$, к которому применяются преобразования, рассмотренные в п.п. 3.1, результатом которых является вектор характеристик изображения $\overline{D}_{img}$. В зависимости от формы изображения выбирается рабочий набор антител $A_W$ и все последующие действия уже будут производиться с ним в пространстве соответствующей размерности.

Далее для вектора характеристик $\overline{D}_{img}$ и каждого антитела $Ig_i \in A_k$, где $k$ = 36 или 54 (в зависимости от формата изображения) выполняются следующие действия.

Шаг [1]. Рассчитывается евклидово расстояние $d$ между $\overline{D}_{img}$ и первыми k элементами вектора $Ig_i$:

$$d = \sqrt{\sum_{j=0}^{k} (D_j - Ig_{i_j})^2} \qquad (10)$$

Шаг [2]. Если расстояние $d$ меньше, чем ($k$+1)-ый элемент вектора $Ig_i$ (последним элементом антитела $Ig_i$ элементом является $r$), то считаем, что вектор характеристик $\overline{D}_{img}$ рассматриваемого изображения $img$ попадает в окрестность антитела $Ig_i$ с радиусом $r$. Соответственно, решена задача бинарной классификации и изображение $img$ относится к классу заполненных стеганоконтейнеров: $img \in S$.

Шаг [3]. Если ни для одного антитела $Ig_i \in A_k$ не выполнилось условие шага 2, оно классифицируется как пустой стеганоконтейнер: $img \notin S \Rightarrow img \in C$.

В общем виде структурно-функциональную схему предлагаемой искусственной имунной сети для решения задачи стеганоанализа изображений можно представить на (рис. 4).

### Анализ полученных результатов и оценка эффективности предлагаемого метода

Алгоритмы предлагаемой ИИС были реализованы в виде программного продукта. Обучение и тестирование ИИС проводилось на основе базы изображений с раз-

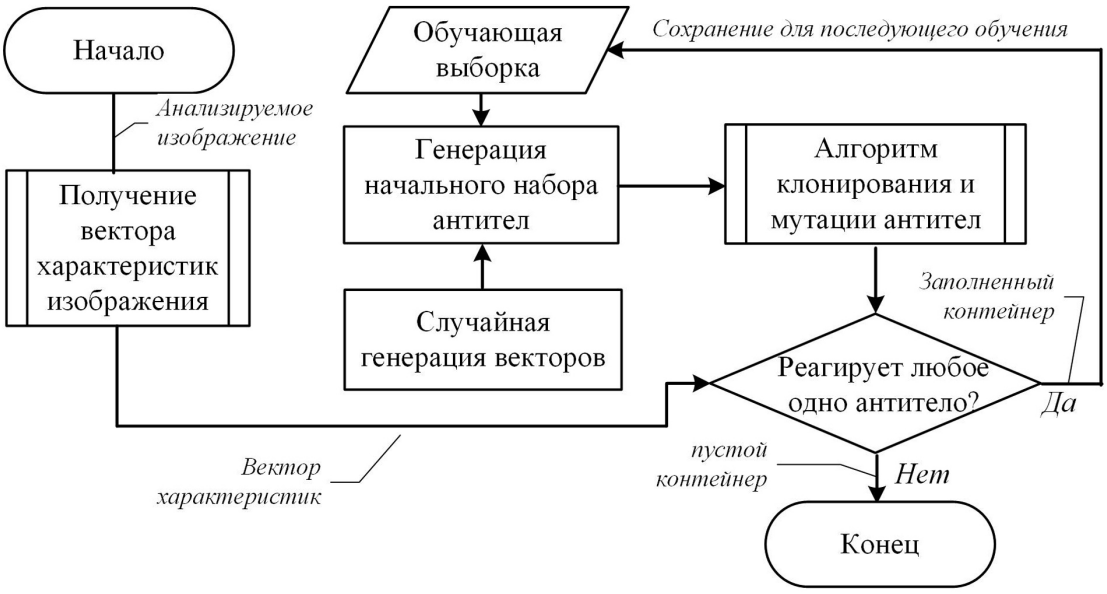


*Рис. 4. Структурно-функциональная схема искусственной имунной сети*

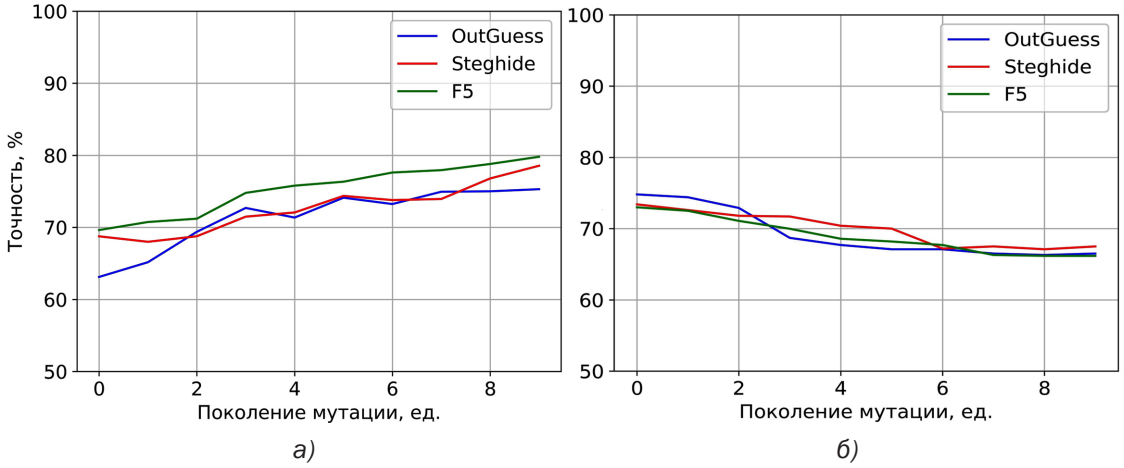


*Рис. 5. Зависимость точности классификации от числа поколений мутации для а) – заполненных стеганоконтейнеров; б) – пустых стеганоконтейнеров для алгоритмов OutGuess, Steghide и F5*

личными статистическими характеристиками (с различными размерами, коэффициентами сжатия JPEG) для того, чтобы приблизить систему к реальным условиям эксплуатации.

На рис. 5 (а, б) представлены графики зависимости точности классификации заполненных и пустых стеганоконтейнеров от числа поколений мутации антител для алгоритмов стеганографии OutGuess, Steghide и F5. Можно отметить, что с большим количеством поколений мутаций, точность обнаружения заполненных стеганоконтейнеров возрастает. Но одновременно с этим, увеличивается величина ошибки второго рода. Это связано с тем, что при мутации антитела затруднительно предугадать, какие пустые контейнеры, неизвестные на данный момент нашей системе, попадут в окрестность антител.

Средняя продолжительность фазы обучения (с десятью поколениями мутаций) составляет в текущей реализации порядка 11 часов. Среднее время решения задачи бинарной классификации одного изображения составляет 0,3 – 0,5 секунды в зависимости от размера изображения. Эксперименты проводились на компьютере со следующими характеристиками: 8 Гб RAM, процессор Intel Core i5 с тактовой частотой 2.5 ГГц.

## ВЫВОДЫ

Данная статья содержит развитие достаточно перспективного подхода эвристического стеганоанализа с использованием искусственных имунных систем. В работе приводится краткий анализ состояния проблематики стеганоанализа изображений и отмечается, что эвристические подходы к стеганоанализу изображений являются на данный момент наиболее перспективными.

Приводится описание разработанного метода стеганоанализа статических изображений формата JPEG, базирующегося на принципах работы искуственных имунных систем, включая формальные описания его ключевых алгоритмов. В целом, можно сделать вывод о достаточной эффективности предлагаемого метода для выявления факта скрытой передачи информации посредством изображений формата JPEG. Точность обнаружения заполненных стеганоконтейнеров составляет около 75-80%, а точность пустых стеганоконтейнеров близка к 70%.

Справедливо отметить, что на данный момент весьма значимым недостатком предлагаемого метода является продолжительное время обучения искусственной иммунной системы. Данную проблему можно решить с использованием гибридной вычислительной системы, включающую графические процессоры (GPU) современных видеокарт, путем распаралелливания необходимых вычислений в используемых алгоритмах.

**Рецензент:** *Марков Алексей Сергеевич, доктор технических наук, старший научный сотрудник, МГТУ им. Н.Э. Баумана, г. Москва, Россия, E-mail: a.markov@bmstu.ru*

**Литература**

1. Holub V., Fridrich J. Low-complexity features for JPEG steganalysis using undecimated DCT // IEEE Trans. Inf. Forensics Secur. 2015. T. 10, № 2. pp. 219–228. DOI: 10.1109/TIFS.2014.2364918.
2. Gulášová M., Jókay M. Steganalysis of stegostorage library // Tatra Mountains Mathematical Publications. 2016. T. 67, № 1. pp. 99–116. DOI: 67. 10.1515/tmmp-2016-0034.
3. Fridrich J.J., Goljan M., Hogea D. Steganalysis of JPEG Images: Breaking the F5 Algorithm // Information Hiding. Lecture Notes in Computer Science. 2002. pp. 310-323. DOI: 10.1007/3-540-36415-3_20.
4. Pevny T., Bas P., Fridrich J. Steganalysis by subtractive pixel adjacency matrix. // IEEE Trans. Inf. Forensics Secur. 2010. № 5 (2). pp. 215–224. DOI: 10.1109/TIFS.2010.2045842.
5. Евсютин О.О., Шумская О.О. Сравнение линейного дискриминанта Фишера и наивного байесовского классификатора в задаче стегоанализа JPEG- изображений / ЭЛЕКТРОННЫЕ СРЕДСТВА И СИСТЕМЫ УПРАВЛЕНИЯ. Томский государственный университет систем управления и радиоэлектроники, Томск. 2017. №1-2. Стр. 79-82.
6. Hendrych J., Kunčický R., Ličev L. New Approach to Steganography Detection via Steganalysis Framework. // Proceedings of the Second International Scientific Conference “Intelligent Information Technologies for Industry” (IITI’17). 2017. Advances in Intelligent Systems and Computing, vol 679. Springer, Cham. DOI: 10.1007/978-3-319-68321-8_51.
7. Ziou D., Jafari R. Efficient steganalysis of images: Learning is good for anticipation // Pattern Analysis and Applications. 2014. Vol. 17, № 2. pp. 279–289. DOI: 10.1007/s10044-012-0303-9.
8. Watanabe S., Murakami K., Furukawa T. and Zhao Q. Steganalysis of JPEG image-based steganography with support vector machine // 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Shanghai. 2016. pp. 631–636. DOI: 10.1109/SNPD.2016.7515970
9. Wang R., Xu M., Ping X., Zhang T. Steganalysis of JPEG images by block texture based segmentation // Multimedia Tools and Applications. 2015. T. 74, № 15. pp. 5725–5746. DOI: 10.1007/s11042-014-1880-y.
10. Kodovský J., Fridrich J. Steganalysis of JPEG images using rich models // Proceedings of SPIE – The International Society for Optical Engineering. 2012. Vol. 8303. pp. 1. DOI: 10.1117/12.907495.
11. Pevny T., Fridrich J. Merging Markov and DCT features for multi-class JPEG steganalysis // Proceedings of SPIE – The International Society for Optical Engineering. 6505. 2007. DOI: 10.1117/12.696774.
12. Дасгупта Д. Искусственные иммунные системы и их применение / под ред. Романюха А.. ФИЗМАТЛИТ, 2006. 344 с.
13. Pérez J.D.J.S., Rosales M.S., Cruz-Cortés N. Universal steganography detector based on an artificial immune system for JPEG images // Proc. – 15th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 10th IEEE Int. Conf. Big Data Sci. Eng. 14th IEEE Int. Symp. Parallel Distrib. Proce. 2017. pp. 1896–1903. DOI: 10.1109/TrustCom.2016.0290.
14. Lu T., Zhang L., Wang S., Gong, Q. Ransomware detection based on V-detector negative selection algorithm // 2017 International Conference on Security, Pattern Analysis, and Cybernetics, SPAC 2017. pp. 531-536. DOI: 10.1109/SPAC.2017.8304335
15. Кушнир Н.В., Кушнир А.В., Анацкая Е.В., Катышева П.А., Устинов К.Г. Искусственные иммунные системы: обзор и современное состояние. // Электронный сетевой политематический журнал “Научные труды КУБГТУ”. Краснодар. 2015. №12. Стр. 382-391.

# STEGANALYSIS METHOD OF STATIC JPEG IMAGES BASED ON ARTIFICIAL IMMUNE SYSTEM

**Shniperov A.N.[4], Prokofieva A.V.[5]**

***Abstract.*** *The purpose of this work is to develop the method for steganalysis of static JPEG images, based on the usage of artificial immune systems.*

*In this paper, a model of an artificial immune system was developed for the task of detecting hidden information in JPEG images. Basic requirements were determined and the basic elements of an artificial immune system were considered, mutation and antibody cloning operations were introduced. Also, formal description of main nodes of the artificial immune system is given. In addition, a brief overview and analysis of the state of the problem of steganalysis are provided in the paper. Also analysis of the obtained experimental results and an assessment of the effectiveness of the developed method is made.*

*The proposed method allows to detect the presence of hidden information, embedded by various popular steganography tools (like OutGuess, Steghide and F5) in static JPEG images with a sufficiently high accuracy. The theoretical significance of this work consists in the development of a fairly promising approach of heuristic steganalysis*

4 Alexey Shniperov, Ph.D., Assistant Professor at laboratory of Information Security of the Department of Applied Mathematics and Computer Security of Siberian Federal University, Krasnoyarsk, Russia. E-mail: ashniperov@sfu-kras.ru

5 Aleksandra Prokofieva, security engineer at the Department of Information Security Hardware Complexes at “Siberian Internet Company” (“SIBINTEK” LLC), Krasnoyarsk, Russia. E-mail: prokofe-aleksandra@yandex.ru

*using artificial immune systems. The practical significance lies in the developed software product, as well as in experimental data confirming the effectiveness of the method of steganalysis in point of the detection of hidden information in JPEG images.*

***Keywords**: Steganography, Steghide, OutGuess, F5, binary classification, Haar wavelet-transform, Clonal selection, Negative selection*

**References**

1. Holub V., Fridrich J. Low-complexity features for JPEG steganalysis using undecimated DCT // IEEE Trans. Inf. Forensics Secur. 2015. T. 10, № 2. pp. 219–228. DOI: 10.1109/TIFS.2014.2364918.
2. Gulášová M., Jókay M. Steganalysis of stegostorage library // Tatra Mountains Mathematical Publications. 2016. T. 67, № 1. pp. 99–116. DOI: 67. 10.1515/tmmp-2016-0034.
3. Fridrich J.J., Goljan M., Hogea D. Steganalysis of JPEG Images: Breaking the F5 Algorithm // Information Hiding. Lecture Notes in Computer Science. 2002. pp. 310-323. DOI: 10.1007/3-540-36415-3_20.
4. Pevny T., Bas P., Fridrich J. Steganalysis by subtractive pixel adjacency matrix. // IEEE Trans. Inf. Forensics Secur. 2010. № 5 (2). pp. 215–224. DOI: 10.1109/TIFS.2010.2045842.
5. Evsyutin O.O., Shumskaya O.O. Sravnenie linejnogo diskriminanta Fishera i naivnogo bajesovskogo klassifikatora v zadache stegoanaliza JPEG- izobrazhenij // ELEKTRONNYE SREDSTVA I SISTEMY UPRAVLENIYA. Tomskij gosudarstvennyj universitet sistem upravleniya i radioelektroniki, Tomsk. 2017.№1-2. pp. 79-82.
6. Hendrych J., Kunčický R., Ličev L. New Approach to Steganography Detection via Steganalysis Framework. // Proceedings of the Second International Scientific Conference “Intelligent Information Technologies for Industry” (IITI’17). 2017. Advances in Intelligent Systems and Computing, vol 679. Springer, Cham. DOI: 10.1007/978-3-319-68321-8_51.
7. Ziou D., Jafari R. Efficient steganalysis of images: Learning is good for anticipation // Pattern Analysis and Applications. 2014. Vol. 17, № 2. pp. 279–289. DOI: 10.1007/s10044-012-0303-9.
8. Watanabe S., Murakami K., Furukawa T. and Zhao Q. Steganalysis of JPEG image-based steganography with support vector machine // 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Shanghai. 2016. pp. 631–636. DOI: 10.1109/SNPD.2016.7515970
9. Wang R., Xu M., Ping X., Zhang T. Steganalysis of JPEG images by block texture based segmentation // Multimedia Tools and Applications. 2015. T. 74, № 15. pp. 5725–5746. DOI: 10.1007/s11042-014-1880-y.
10. Kodovský J., Fridrich J. Steganalysis of JPEG images using rich models // Proceedings of SPIE – The International Society for Optical Engineering. 2012. Vol. 8303. pp. 1–DOI: 10.1117/12.907495.
11. Pevny T., Fridrich J. Merging Markov and DCT features for multi-class JPEG steganalysis. // Proceedings of SPIE – The International Society for Optical Engineering. 6505. 2007. DOI: 10.1117/12.696774.
12. Dasgupta D. Iskusstvennye immunnye sistemy i ih primenenie. / edited by Romanyuha A.. FIZMATLIT, 2006. 344 p.
13. Pérez J.D.J.S., Rosales M.S., Cruz-Cortés N. Universal steganography detector based on an artificial immune system for JPEG images // Proc. – 15th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 10th IEEE Int. Conf. Big Data Sci. Eng. 14th IEEE Int. Symp. Parallel Distrib. Proce. 2017. pp. 1896–1903. DOI: 10.1109/TrustCom.2016.0290.
14. Lu T., Zhang L., Wang S., Gong, Q. Ransomware detection based on V-detector negative selection algorithm // 2017 International Conference on Security, Pattern Analysis, and Cybernetics, SPAC 2017. pp. 531-536. DOI: 10.1109/SPAC.2017.830433515.
15. Kushnir N.V., Kushnir A.V., Anackaya E.V., Katysheva P.A., Ustinov K.G. Iskusstvennye immunnye sistemy: obzor i sovremennoe sostoyanie. // Elektronnyj setevoj politematicheskij zhurnal “Nauchnye trudy KUBGTU”. Kubanskij gosudarstvennyj tekhnologicheskij universitet, Krasnodar. 2015. №12. pp. 382-391.