УДК 003.26

А.С. Елисеев, А.Р. Тикиджи-Хамбурьян

СТАТИСТИЧЕСКИЙ СТЕГАНОГРАФИЧЕСКИЙ АНАЛИЗ ИСТОЧНИКОВ КОНТЕЙНЕРОВ ОДИНАКОВОГО ТИПА С ИСПОЛЬЗОВАНИЕМ БАЗОВОГО МЕТОДА АНАЛИЗА ОТДЕЛЬНЫХ КОНТЕЙНЕРОВ НЕИЗВЕСТНОЙ СТРУКТУРЫ

Исследуется вопрос повышения точности классификации пустых и заполненных контейнеров в случае малых относительных объемов скрываемого в контейнер сообщения. Для решения данного вопроса предлагается рассматривать не единичные информационные объекты, а их совокупности, полученные от одного источника. При этом источник фактически отождествляется с набором контейнеров, ассоциированных с ним. Например, источником может быть тот адрес электронной почты, с которого осуществлялась передача всех контейнеров совокупности, или тот аккаунт на фотохостинге, от имени которого были выложены все изображения. Предлагается метод стеганографического анализа источников контейнеров, основанный на объединении результатов анализа отдельных контейнеров совокупности, полученной от источника, при помощи базового метода, представленного как черный ящик. Целью анализа источников является выявление среди них тех, которые используют средства сокрытия информации. Показано, что в ряде случаев предложенный метод позволяет добиться значительного увеличения точности классификации источников относительно точности классификации отдельных контейнеров базовым методом.

Стеганография; стеганоанализ; наборы контейнеров; классификация источников контейнеров; нормальное распределение.

A.S. Eliseev, A.R. Tikidzhi-Khamburian

STATISTICAL STEGANALYSIS OF IMAGES SOURCES, BASED ON OPAQUE SEPARATE IMAGES STEGANALYSIS TECHNIQUE

Paper is devoted to steganalysis accuracy in case of low relative embedding rates. Its main idea is to consider not single images, but sets of related images, e.g. images, received from the single source. Images source is identified with the appropriate set of images. For instance, E-Mail address of sender of all images in the set, can be thought of as a source. Similarly, account in photo sharing network, in behalf of which all the images in the set were shared, is another example of images source. Steganalysis technique for images sources is proposed. The technique is based on uniting results of analyses of single images from source's set. Analyses of single images is considered to be held by some unknown opaque algorithm, with only statistical accuracy characteristics known. The proposed sources steganalysis aims to classify appropriate images sets into sets with and without steganographically processed images. It is shown, that proposed technique's classification accuracy is in some cases much higher then the accuracy of classification of single images by the base opaque algorithm.

Steganography; steganalysis; images sets; images sources classification; Gaussian distribution.

Введение. Современная методология математического исследования стойкости методов сокрытия информации базируется на предположении о том, что аналитику доступен для изучения один стеганоконтейнер или, во всяком случае, интерес для него представляет факт наличия или отсутствия скрытой информации в одном контейнере. Практика же применения стеганографических систем такова, что, во-первых, одним пользователем один метод сокрытия используется многократно, во-вторых, зачастую, для передачи стеганоконтейнеров используется один и тот же открытый канал связи (возможно доступный для изучения одному и тому же аналитику на протяжении многих сеансов передачи скрытых данных) и,

в-третьих, нежелательным для пользователя стеганосистемы является не обнаружение факта наличия скрытой информации в одном из передаваемых им контейнеров, а обнаружение факта наличия скрытой информации в каком-либо из множества передаваемых контейнеров.

Ясная практическая значимость данного нюанса делает актуальной разработку методов выявления факта сокрытия данных в наборе контейнеров, полученных от одного и того же источника. В области сокрытия данных с учетом их возможного распределения по нескольким контейнерам наиболее успешными являются работы [1, 2], в которых при помощи теоретико-информационного подхода дан характер изменения скрытности при изменении числа контейнеров и плотности сокрытия одновременно в нескольких контейнерах. Ответная область стеганографического анализа наборов контейнеров исследована хуже. В [3], сформулирована общая идея повышения точности анализа за счет рассмотрения множества контейнеров. В [4], данные идеи получили развитие и были доведены до практически значимых результатов. В настоящей работе обобщена часть результатов, полученных в [4]. Кроме того получена явная оценка числа контейнеров в наборе, необходимого для достижения желаемой точности анализа. Описываемый в работе метод, предполагает, что все контейнеры от исследуемого источника, которые аналитику удалось собрать, имеют один и тот же тип (или, как минимум, к ним всем применим один и тот же взятый за основу метод анализа отдельных контейнеров).

Стеганографический анализ наборов контейнеров. Предположим, что аналитику удалось получить в свое распоряжение n контейнеров от, интересующего его источника, причем все эти контейнеры принадлежат некоторому классу T. В его распоряжении имеется некоторый метод анализа отдельных контейнеров, который можно применить ко всем контейнерам класса T $U : T \to \{0,1\}$ (т.е. считается, что базовый метод выдает только ответы "да" или "нет"). Никаких ограничений на внутреннее устройство данного метода не накладывается. Будем лишь считать известными некоторые характеристики качества данного метода, а именно частоты его ошибок первого и второго рода, которые обозначим через α и β соответственно (здесь и далее, руководствуясь принципами презумпции невиновности за основную гипотезу, принимаем предположение об отсутствии в объекте дополнительной встроенной информации, а за альтернативную – о наличии в объекте стегановложения). Все, что аналитик может сделать с имеющимся у него набором объектов – применить метод U к каждому из имеющихся у него контейнеров в отдельности. Далее предлагается подсчитать количество контейнеров в исследуемой совокупности, которые оказались подозрительными на наличие сокрытия с точки зрения метода U. Для того, чтобы классифицировать набор, как содержащий или не содержащий заполненные контейнеры (и, следовательно, источник данного набора – как использующий стеганографию или как не использующий ее), следует задаться некоторым порогом числа срабатываний базового метода анализа отдельных контейнеров. Данный порог может зависеть от средних частот ошибок базового метода, а также от желаемой чувствительности анализа. Разберемся, каким образом следует выбирать данный порог.

Рассмотрим альтернативные гипотезы H_r о совокупности контейнеров, состоящие в том, что доля заполненных контейнеров в совокупности составляет r, т.е. среди n контейнеров имеется nr заполненных и $(1-r)n$ пустых контейнеров. Гипотеза H_0 соответствует полному отсутствию заполненных контейнеров, т.е. ситуации, когда источник исследуемого набора вообще не использует средства стеганографического сокрытия данных. Зададимся некоторой фиксированной долей контейнеров r – минимальной долей стеганоконтейнеров, присутствие которой в наборе объектов будем стараться отличить от полного отсутствия заполненных объектов в исследуемой совокупности (из общих соображений интуитивно ясно,

что чем больше зафиксированная волевым образом доля *r*, тем легче отличить, содержащую данную долю совокупность от совокупности, соответствующей H_0; данная интуитивная догадка будет позже обоснована). Итак, будем проверять основную гипотезу H_0 при альтернативе H_r.

Обозначим через ξ_0 случайное число срабатываний базового метода на совокупности контейнеров $c_1,\ldots,c_n$, которая заведомо соответствует основной гипотезе:

$$\xi_0 = \sum_{i=1}^{n} U(c_i). \quad (1)$$

Случайная величина ξ_0, являясь суммой из *n* случайных величин, имеющих распределение Бернулли с одним и тем же значением параметра, и независимых в силу независимости анализа отдельных контейнеров, которые также предполагаются независимыми, удовлетворяет биномиальному распределению с числом опытов *n* и вероятностью успеха в каждом опыте α: $\xi_0 \sim Bi(n,\alpha)$. Через ξ_r обозначим число срабатываний базового метода при условиях, определяемых гипотезой H_r. Аналогично: $\xi_r \sim Bi(n(1-r),\alpha) + Bi(nr,1-\beta)$.

Считая общее число контейнеров в рассматриваемой совокупности *n* достаточно большим, а вероятности (частоты) ошибок α и β достаточно далекими от граничных значений 0 и 1, перейдем к Гауссовой аппроксимации биномиальных распределений: $Bi(n,p) \approx N(np, np(1-p))$, где через $N(a,\sigma^2)$ обозначено нормальное распределение с математическим ожиданием a и дисперсией σ^2.

Таким образом, имеем:

$$\xi_0 \sim N(n\alpha, n\alpha(1-\alpha)), \quad (2)$$

$$\xi_r \sim N(n(1-r)\alpha, n(1-r)\alpha(1-\alpha)) + N(nr(1-\beta), nr\beta(1-\beta)),$$

или, по свойству линейной замкнутости нормального распределения:

$$\xi_r \sim N(n(1-r)\alpha + nr(1-\beta), n(1-r)\alpha(1-\alpha) + nr\beta(1-\beta)). \quad (3)$$

Дополнительно обозначим:

$$\mu_0 = n\alpha, \sigma_0^2 = n\alpha(1-\alpha), \quad (4)$$

$$\mu_r = n(1-r)\alpha + nr(1-\beta),\ \sigma_r^2 = n(1-r)\alpha(1-\alpha) + nr\beta(1-\beta). \quad (5)$$

Тогда $\xi_0 \sim N(\mu_0, \sigma_0^2)$, $\xi_r \sim N(\mu_r, \sigma_r^2)$.

Имеем

$$\mu_r = n(1-r)\alpha + nr(1-\beta) = n\alpha - nr\alpha + nr - nr\beta = \\ = n\alpha + nr(1-a-\beta) > n\alpha,$$

при $\alpha + \beta < 1$, т.е. если базовый метод анализа отдельных контейнеров дает хотя бы какое-то преимущество перед случайным гаданием (в противном случае нет никаких причин его использовать).

Итак, математическое ожидание ξ_r больше математического ожидания ξ_0. Будем в дальнейшем в таких случаях говорить, что распределение ξ_r "лежит правее" распределения ξ_0. Данный факт подтверждает интуитивную догадку о том, что чем больше срабатываний дал базовый метод, тем вероятнее наличие заполненных контейнеров в исследуемом наборе (и тем вероятнее использование стега-

нографии источником данного набора). После выбора порогового значения нашей статистики следует при превышении данного порогового значения принимать альтернативную гипотезу H_r, и основную гипотезу H_0 – в противном случае.

Пороговое значение числа срабатываний базового метода определяет число ошибок первого и второго рода при принятии решения по всему набору контейнеров. Частота ошибок первого рода (ложных срабатываний) соответствует части распределения величины ξ_0, лежащей правее выбранного порога, а частота ошибок второго рода (пропусков наборов, содержащих достаточное число заполненных контейнеров) – частью распределения ξ_r, лежащей левее порога. В дальнейшем для краткости частоту (вероятность) ошибок (первого или второго рода) метода классификации будем называть просто ошибкой метода (первого или второго рода).

Теперь становится понятно, каким образом следует выбирать пороговое значение. Пусть, например, необходимо выбрать порог таким образом, чтобы при заданном объеме *n* совокупности анализируемых контейнеров и заданной максимальной ошибке первого рода A при классификации совокупности, минимизировать ошибку второго рода. Тогда следует расположить порог так, чтобы справа от него располагалось доля распределения величины ξ_0, равная A, т.е. порог должен совпадать с $(1-\mathrm{A})$-квантилем распределения ξ_0, который обозначим через $p_{1-\mathrm{A}}$.

γ-квантиль p_γ нормального распределения с параметрами μ_0 и σ_0^2 определяется из соотношения: $\gamma = \Phi\left(\frac{p_\gamma - \mu_0}{\sigma_0}\right)$, где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$ – функция распределения стандартного распределения (т.е. нормированного и центрированного нормального распределения $N(0,1)$), которую иногда называют также функцией Лапласа.

Имеется очевидная связь функции $\Phi(x)$ со стандартной функцией ошибок $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$, присутствующей в библиотеках всех современных языков программирования, а также в программных пакетах для компьютерных вычислений и в приложениях электронных таблиц:

$$\Phi(x) = \frac{1}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right).$$

Однако для вычисления γ-квантиля p_γ необходима обратная функция $\Phi^{-1}(x)$, иногда называемая также пробит-функцией, либо обратная функция ошибок. Действительно:

$$p_\gamma = \Phi^{-1}(\gamma)\cdot\sigma_0 + \mu_0 = erf^{-1}(2\gamma - 1)\cdot\sigma_0\sqrt{2} + \mu_0.$$

Реализации пробит-функции или обратной функции ошибок также присутствуют во многих статистических библиотеках и пакетах, а также в приложениях электронных таблиц. Поэтому искомое пороговое значение достаточно легко вычислить на практике:

$$p_{1-\mathrm{A}} = \Phi^{-1}(1-\mathrm{A})\cdot\sigma_0 + \mu_0 = erf^{-1}(1-2\mathrm{A})\cdot\sigma_0\sqrt{2} + \mu_0. \qquad (6)$$

Аналогично, порог для случая, когда необходимо при заданной максимальной ошибке второго рода B минимизировать ошибку первого рода A, следует выбрать следующим образом:

$$p_{\text{В}} = \Phi^{-1}(\text{В})\cdot\sigma_r + \mu_r = erf^{-1}(2\text{В}-1)\cdot\sigma_r\sqrt{2} + \mu_r \cdot \quad (7)$$

Сравнение точности анализа наборов с точностью анализа отдельных контейнеров. Рассмотрим теперь, как изменяется точность классификации наборов контейнеров (и, следовательно, источников данных наборов) относительно точности классификации отдельных контейнеров. Исчерпывающей характеристикой точности параметрической бинарной классификации является соответствующая ROC-кривая, т.е. зависимость числа правильно выявляемых объектов от ошибки первого рода при различных значениях параметра классификации. В данном случае параметром является порог принятия решения по числу срабатываний базового метода. Вычислим ошибку второго рода В при классификации наборов контейнеров в зависимости от порога p_γ. Поскольку величина ξ_r также распределена нормально, то, рассуждая как и раньше, приходим к тому, что:

$$\text{В} = \Phi\left(\frac{p_\gamma - \mu_r}{\sigma_r}\right) = \frac{1}{2}\left(1 + erf\left(\frac{p_\gamma - \mu_r}{\sigma_r\sqrt{2}}\right)\right)\cdot$$

Теперь легко вычислить и саму ROC-кривую, т.е. зависимость величины $1-\text{В}$ от ошибки первого рода А:

$$1-\text{В} = 1-\Phi\left(\frac{\Phi^{-1}(1-\text{А})\cdot\sigma_0 + \mu_0 - \mu_r}{\sigma_r}\right), \text{ или} \quad (8)$$

или $1-\text{В} = \frac{1}{2}\left(1 - erf\left(\frac{erf^{-1}(1-2\text{А})\cdot\sigma_0\sqrt{2} + \mu_0 - \mu_r}{\sigma_r\sqrt{2}}\right)\right)$.

Параметрами данной кривой являются:

- объем исследуемой выборки контейнеров *n*;
- частота ошибок первого рода α базового метода;
- частота ошибок второго рода β базового метода;
- выбранная выявляемая частота r использования средств сокрытия.

На рис. 1 слева представлены ROC-кривые, соответствующие $\alpha = 0.05$, $\beta = 0.3$, $r = 0.1$ при различных объемах выборки контейнеров. Видно, что точность классификации малых наборов контейнеров меньше точности классификации отдельных контейнеров (соответствующие ROC-кривые проходят ниже и правее точки, соответствующей базовому методу). Это объясняется тем, что за счет разбавления заполненных контейнеров пустыми (доля *r*) плотность использования сокрытия данных снижается в 10 раз. Однако, когда число контейнеров в наборе достигает $n = 58$, точность классификации набора начинает превосходить точность базового метода за счет дополнительных данных, которые подвергаются анализу. При $n = 200$ получаем кривую, площадь под которой приближенно равна 0.997. Данная кривая соответствует весьма качественному методу классификации источников наборов контейнеров.

Попробуем варьировать долю *r*, обнаружения которой хотим добиться, при неизменном числе контейнеров $n = 20$. Соответствующие кривые изображены на рис. 1 справа. Из рисунка также видно, что когда анализу подвергается набор лишь из 20 контейнеров, точность классификации данного набора становится выше точности классификации отдельных контейнеров исходным методом лишь когда заполненные контейнеры составляют как минимум долю $r = 0.18$ от общего числа контейнеров. Причем качество предложенного метода анализа быстро растет с ростом доли *r*. Так при $r = 0.05$ площадь под ROC-кривой метода составляет все-

го около 0.674, а при $r = 0.2$ – уже 0.948. Дальнейшее повышение долевого содержания заполненных контейнеров в исследуемой совокупности быстро приводит к тому, что метод начинает показывать очень хорошее качество классификации: при $r = 0.3$ площадь под ROC-кривой составляет 0.989, а при $r = 0.5$ – уже более 0.999.

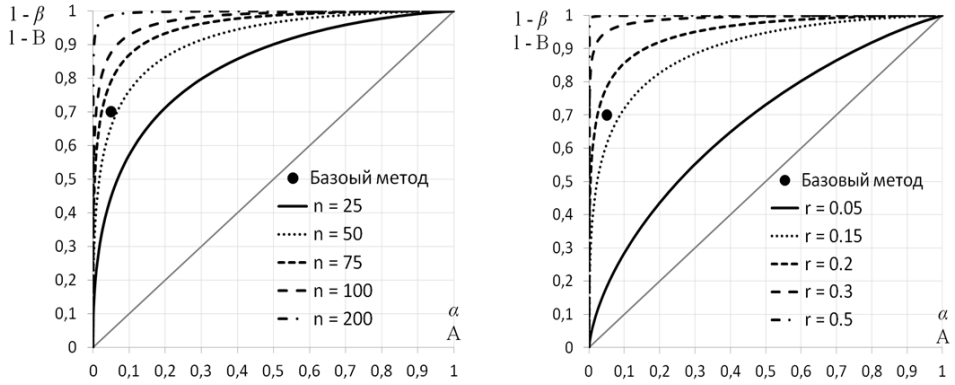


Рис. 1. ROC-кривые классификации наборов слева – при $\alpha = 0.05, \beta = 0.3, r = 0.1$*;*
справа – при $\alpha = 0.05, \beta = 0.3, n = 20$

Рассмотрим, что происходит при использовании менее качественного базового метода. Например, возьмем $\alpha = 0.3, \beta = 0.5$. Соответствующие ROC-кривые при $n = 50$ изображены на рис. 2 слева. Точность базового метода достигается при $r = 0.17$, однако сама эта точность не велика: площадь под соответствующей ROC-кривой составляет 0.643. При данном объеме набора контейнеров хорошо отличать на практике удастся только наборы, у которых $r \geq 0.4$ (площадь под кривой при $r = 0.4$ составляет 0.804, при $r = 0.6$ – 0.898, а при $r = 0.9$ – 0.97). С повышением объема выборки (и при фиксированной доле *r*) точность растет не так быстро, как в случае более качественного базового метода анализа.

Рассмотрим еще более плохой базовый метод, у которого $\alpha = 0.5, \beta = 0.4$. Например, такие показатели могут быть у различных достаточно хороших методов, в случае низкой плотности сокрытия. При $n = 100$ соответствующие ему ROC-кривые изображены на рис. 2 справа. В данных условиях на практике вполне можно отличать наборы, в которых заполненные контейнеры (возможно, с низкой плотностью) встречаются в $r \geq 0.6$ случаях. Площадь под ROC-кривой $r = 0.6$ составляет 0.803. А точность базового метода за счет общего числа контейнеров превышается уже при $r = 0.13$.

Заметим, что как следует из (8), для достижения заданных уровней ошибок первого и второго рода классификации источника A и B (при также заданных α, β, r) достаточно выполнения условия $\mathrm{B} \geq \Phi\left(\dfrac{\Phi^{-1}(1-\mathrm{A})\cdot\sigma_0 + \mu_0 - \mu_r}{\sigma_r}\right)$, которое равносильно $\Phi^{-1}(\mathrm{B})\cdot\sigma_r + \mu_r \geq \Phi^{-1}(1-\mathrm{A})\cdot\sigma_0 + \mu_0$, или, с учетом (4), (5):

$$n \geq \frac{\left(\sqrt{\alpha(1-\alpha)}\Phi^{-1}(1-\mathrm{A}) - \sqrt{(1-r)\alpha(1-\alpha) + r\beta(1-\beta)}\Phi^{-1}(\mathrm{B})\right)^2}{r^2\left(1-(\alpha+\beta)\right)^2}. \qquad (9)$$

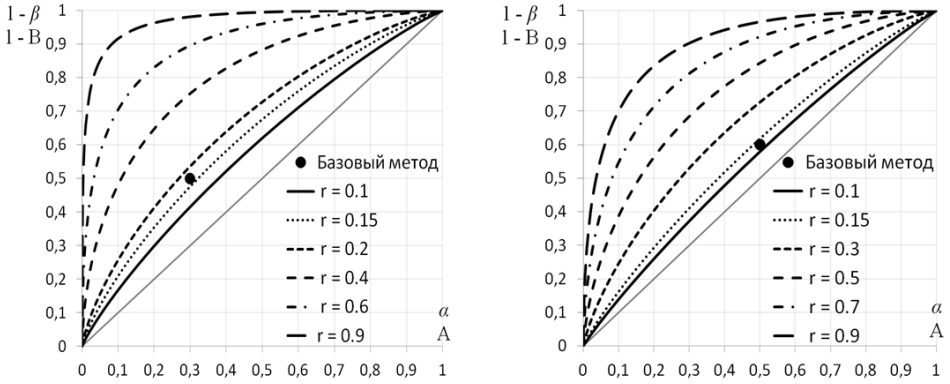


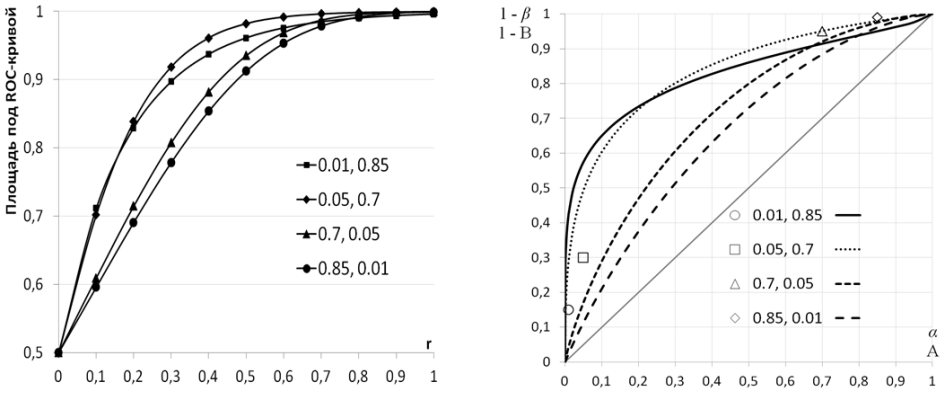
Рис. 2. ROC-кривые классификации наборов слева – при $\alpha = 0.3, \beta = 0.5, n = 50$*;*
справа – при $\alpha = 0.5, \beta = 0.4, n = 100$

Формула (9) дает оценку числа контейнеров от источника, которые необходимо собрать, чтобы достичь заданных уровней ошибок первого и второго рода при классификации данного источника.

Выбор базового метода. Итак, выше было продемонстрировано, как на основе произвольного базового метода анализа отдельных контейнеров можно построить метод анализа источников наборов таких контейнеров, который по точности классификации превосходит базовый. Как лучше поступить, если в наличии имеется несколько методов классификации отдельных контейнеров, каждый из которых можно использовать в качестве базового? В настоящей работе предполагается, что о базовом методе известны только частоты допускаемых им ошибок первого и второго рода. Поэтому и при сравнении базовых методов будем учитывать лишь частоты ошибок. Выше был показан тот очевидный факт, что чем точнее базовый метод анализа (чем меньше частоты его ошибок), тем точнее получается анализ наборов контейнеров, проводимый на его основе. Разберемся теперь, какая чувствительность базового метода (т.е. соотношение его ошибок первого и второго рода) приводит к лучшим результатам при анализе контейнеров. Для этого рассмотрим четыре базовых метода с характеристиками (в порядке возрастания их чувствительности): $\alpha = 0.01, \beta = 0.85$; $\alpha = 0.05, \beta = 0.7$; $\alpha = 0.7, \beta = 0.05$; $\alpha = 0.85, \beta = 0.01$.

На рис. 3,а изображена зависимость площади под ROC-кривой, характеризующей точность классификации набора, от доли *r* заполненных контейнеров в наборе, если из $n = 50$ контейнеров. Из рисунка видно, что при небольших долях *r* значительно выгоднее использовать один из специфичных базовых методов, нежели один из чувствительных. Справа изображены сами соответствующие ROC-кривые при $n = 50, r = 0.2$. Данная асимметрия легко объяснима тем, что как вытекает из формул (2) и (3), ошибка первого рода базового метода α полностью (без участия ошибки второго рода β) определяет распределение ξ_0, а также влияет на распределение ξ_r, причем влияет тем больше, чем меньше *r*. Поэтому при малых *r* вклад β в определение ROC-кривой значительно меньше вклада α. При $r = 1$ их вклады уравниваются, поэтому в данной точке ординаты всех графиков на рисунке 3А практически совпадают (это, однако, совсем не означает, что сами ROC-кривые также совпадают). Если бы было необходимо, наоборот, отличать наборы с частично заполненными контейнера-

ми, от наборов полностью заполненных контейнеров, наблюдалась бы асимметрия в противоположную сторону. Поэтому для данной двойственной задачи больше подходит чувствительный базовый метод.



а б

Рис. 3. а – зависимость площади под ROC-кривой от r при $n = 50$*;*
б – ROC-кривые при $n = 50, r = 0.2$

Кроме того, при небольших долях *r* скорость роста точности классификации с ростом объема выборки *n* также зависит в основном от ошибки первого рода базового метода, что хорошо видно из рис. 4, на котором изображена зависимость площади под ROC-кривой от объема выборки *n* при фиксированном $r = 0.2$ для рассматриваемых базовых методов. Т.е. чем более специфичным является базовый метод, тем быстрее точность классификации набора увеличивается до "хороших" значений с ростом объема набора контейнеров от данного источника.

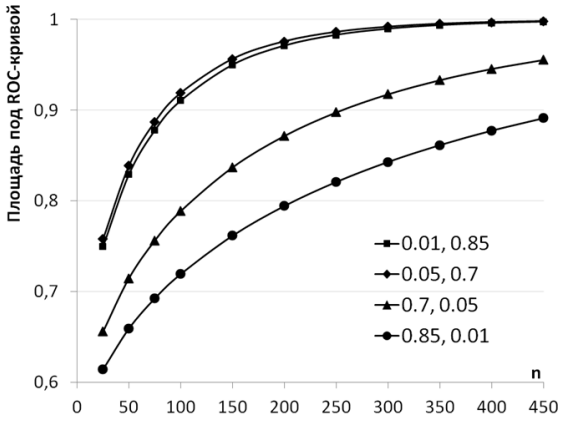


Рис. 4. Зависимость площади под ROC-кривой от n при $r = 0.2$

Однако, заметим, что из рис. 3,б видно, что хотя общие площади под кривыми, основанными на специфичных методах, заметно больше, однако при больших ошибках первого рода ошибки второго рода меньше, когда за основу берется чувствительный метод. Поэтому для построения чувствительного метода может быть выгоднее и за основу брать чувствительный.

Выводы. Таким образом, представленный метод классификации использующих и не использующих стеганографию источников наборов контейнеров состоит в следующем.

1. Задаться частотой *r* использования средств стеганографического сокрытия данных, которую необходимо выявить.

2. Собрать как можно большее количество контейнеров, переданных исследуемым источником либо по крайней мере столько контейнеров, сколько необходимо для достижения необходимой точности классификации в соответствии с формулой (9).

3. Исходя из числа собранных контейнеров *n*, выбранной частоты *r, и* частот ошибок первого и второго рода базового метода α и β вычислить порог для числа срабатываний базового метода на собранной совокупности контейнеров, например, по формуле (6) или (7), задавшись максимально допустимой вероятностью ложной тревоги A или пропуска B соответственно.

4. Подсчитать число срабатываний базового метода на собранной совокупности контейнеров по формуле (1).

5. Если число срабатываний превысило порог, то сделать вывод об использовании источником средств сокрытия данных. В противном случае сделать вывод о том, что метод сокрытия, против которого спроектирован базовый метод анализа не применялся достаточно часто (или с достаточной плотностью).

В работе показано, что за счет анализа всего набора контейнеров, ассоциированных с данным источником, представленный метод позволяет значительно повысить точность классификации по сравнению с базовым методом анализа отдельных контейнеров. При этом точность повышается тем больше, чем чаще источник использует средства сокрытия данных и чем больше контейнеров, с ним ассоциированных, удалось собрать и подвергнуть анализу. Однако, в случае малого числа контейнеров для анализа или малой частоты встречаемости заполненных контейнеров среди исследуемого набора контейнеров, точность классификации источников таких наборов может быть даже меньше точности классификации отдельных контейнеров базовым методом. Кроме того, точность классификации источников увеличивается с увеличением точности базового метода и при уменьшении чувствительности (с сохранением точности).

Представленный в работе метод объединения результатов применения базового метода к отдельным, независимым контейнерам предназначен для использования в случае, когда базовый метод рассматривается как черный ящик, выдающий лишь бинарный ответ на вопрос о наличии стегановложения в анализируемом контейнере. Таким образом, условия на базовый метод являются весьма необременительными, что позволяет использовать в качестве базового практически любой метод анализа отдельных контейнеров и делает представленный метод достаточно универсальным.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ker A.D.* A capacity result for batch steganography // IEEE Signal Processing Letters. – 2007. – № 14 (8). – P. 525-528.
2. *Ker A.D.* Batch steganography and pooled steganalysis // IH-8, LNCS 4437. – 2007. – P. 265-281.
3. *Балакин А.В., Гуфан А.Ю.* Обнаружение стегановложений во множестве однотипных объектов // T-Comm – Телекоммуникации и транспорт. Спецвыпуск апрель 2009 (Цифровая обработка сигналов). – М.: ООО "Издательский дом Медиа Паблишер", 2009. – С. 39-41.
4. *Елисеев А.С.* Исследование и разработка методов и алгоритмов стеганографического анализа отдельных контейнеров и их связанных наборов: Дис. … канд. техн. наук. – Ростов-на-Дону: Изд-во ЮФУ, 2013. – 169 с.

Статью рекомендовал к опубликованию д.ф.-м.н., профессор М.Ф. Куприянов.

Елисеев Алексей Сергеевич – ФГАНУ НИИ "Спецвузавтоматика", г. Ростов-на-Дону; e-mail: alexey.meteorite@gmail.com, alexeliseev@bk.ru; 344065, г. Ростов-на-Дону, пер. Днепровский, 124/6, кв. 99; тел.: +79043418984; к.т.н.; старший научный сотрудник.

Тикиджи-Хамбурьян Анна Рубеновна – e-mail: azick9@gmail.com; 344000, г. Ростов-на-Дону, ул. Соколова, 64/89, кв. 7; тел.: +79888990248; инженер-программист.

Eliseev Alexey Sergeevich – FSASE SRI "Specvuzavtomatika", Rostov-on-Don; e-mail: alexey.meteorite@gmail.com, alexeliseev@bk.ru; 124/6, Dneprovskiy pr., app. 99, Rostov-on-Don, 344065, Russia; phone: +79043418984; cand. of eng. sc.; senior scientist.

Tikidzhi-Khamburian Anna Rubenovna – e-mail: azick9@gmail.com; 64/89, Sokolova street, app. 7, Rostov-on-Don, 344000, Russia; phone: +79888990248; engineer-programmer.

УДК 004.056

Ю.Е. Рябинин, О.А. Финько

УСТОЙЧИВАЯ К АТАКАМ СТЕГАНОГРАФИЧЕСКАЯ СИСТЕМА В РАСШИРЕННОМ МОДУЛЯРНОМ КОДЕ

Не имея возможности вскрыть передаваемые по стеганографическому каналу легальными пользователями сообщения, злоумышленник применяет атаки, разрушающие встроенные в стеганографический контейнер данные. Одним из возможных решений описанной проблемы является применение распределенной многоканальной системы обмена стеганографическими сообщениями в расширенном модулярном коде. Суть предлагаемой стеганографической системы заключается в представлении встраиваемых в стеганографические контейнеры сообщений (предварительно зашифрованных по известному алгоритму и ключу) наименьшими неотрицательными вычетами модулярного кода. Вычисляются избыточные символы модулярного кода. Это обеспечивает устойчивую передачу информации по стеганографическим каналам связи при деструктивных воздействиях злоумышленника в сетях общего пользования. Представлена оценка достоверности передаваемых данных, получены расчетные формулы для вычисления вероятностей возникновения ошибок. Достигнутые результаты позволяют проектировать стеганографические системы, устойчивые к возможным атакам, характерным для указанных систем и обеспечивающие высокую степень достоверности передаваемых шифрованных данных в сети общего пользования.

Модулярный код; стеганография; Китайская теорема об остатках; остаток; основание; достоверность.

Ju.E. Ryabinin, O.A. Finko

STEGANOGRAPHIC SYSTEM RESISTANT TO ATTACK IN THE EXTENDE MODULAR ARITHMETIC

Not being to open transmitted legitimate users by steganography channel from of the message, the attacker uses the attacks, destroying built-in steganographic the container data. One possible solution for the above problems is the use of distributed multi-channel system of exchange of steganographic messages in the extended modular code. The essence of the proposed steganographic system is to provide embedded in containers steganographic messages (pre- encrypted by a known algorithm and key) least nonnegative residue modular code. Calculated excess characters modular code. It provides a robust data transmission over steganography communication channels with the destructive effects attacker in public networks. Provides an assessment of the reliability of the transmitted data , obtained formulas for calculating the probability of errors. The achieved results allow for the design of steganographic systems, resistant to possible attacks, characteristic for these systems and ensure high reliability of the transmitted encrypted data within a public network.

Modular code; steganography; Chinese Remainder Theorem; remainder; basis; reliability.