

Краткие сообщения



УДК 621.391

И.Л. ЧВАРКОВА, В.С. САДОВ, А.Ф. ЧЕРНЯВСКИЙ

ОЦЕНКА ПРИМЕНИМОСТИ КРИТЕРИЯ ХИ-КВАДРАТ ДЛЯ ОБНАРУЖЕНИЯ СТЕГАНОГРАФИЧЕСКОГО КАНАЛА В АУДИОДАННЫХ

In the article consideration is given to the problems of the chi-square criterion adaptability for disclosure of secret messages in steganographic audio-information transmission channels with classification into different classes according to the scatter in meanings of the signal samples.

Цифровая стеганография - наука о незаметном и устойчивом к атакам сокрытии одних данных в других. Методы стеганографии позволяют передавать конфиденциальную информацию таким образом, что неизвестным остается сам факт передачи сообщения. Одним из распространенных методов внедрения скрываемых сообщений в мультимедийные данные, использующим модуляцию контейнера во временной области, является метод замены младших значащих бит в оцифрованных отсчетах сигналов. Значения этих бит практически не несут каких-либо сведений об аудиосигнале, поэтому человек не способен на слух определить их изменения [1]. Однако, несмотря на относительную простоту, этот метод имеет серьезные недостатки, в частности неустойчивость.

Наилучший способ оценки наличия или отсутствия скрытого канала передачи информации - применение статистических атак, исследующих такие статистические характеристики последовательностей данных, как оценка энтропии, коэффициенты корреляции, вероятностные характеристики и зависимости между элементами последовательностей, условные распределения и их различимость по критерию Хи-квадрат.

Считается, что критерий Хи-квадрат - один из перспективных подходов для выявления факта наличия скрытого сообщения в статических изображениях [2]. Оценим применимость этого критерия для стегосистем с аудиоданными, полученными методом замены младшего значащего бита.

Пусть исходный аудиосигнал, или контейнер, до встраивания состоит из N отсчетов $x(i)$, где $i=1...256$. Если $f(x(i), s(i))$, где $s(i)$ - отсчет скрываемого сообщения, является функцией, описывающей процесс встраивания сообщения в контейнер [3], то с учетом этого процесс стегокодирования и извлечения сообщения можно представить как на рис. 1.

Рассмотрим два соседних двоичных отсчета $x(i)$ 8-битной последовательности, равных 00000000 и 00000001. При встраивании сообщения в младший значащий бит значение отсчета 00000000 либо изменится на 00000001, либо останется без изменения. Следовательно, вероятности появления отсчетов $y(i)$ звука, равных 00000000 и 00000001, при стегокодировании одинаковы, что состав-

ляет полусумму вероятностей появления соседних отсчетов незаполненного контейнера.

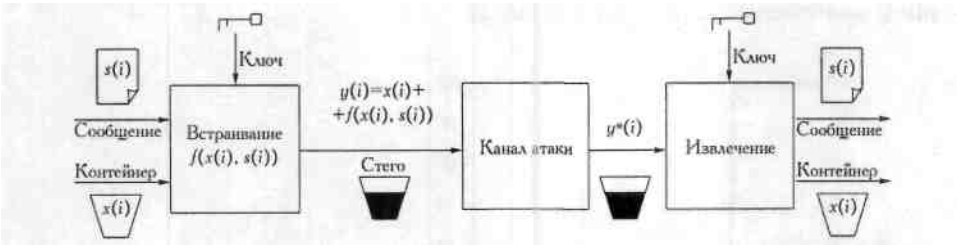


Рис. 1. Структура процесса стегокодирования

Критерий Хи-квадрат основывается на предположении о равных вероятностях появления соседних уровней громкости в стегофайле [4], т. е. при стегокодировании аудиоданных вероятности появления соседних уровней громкости усредняются.

Величина Хи-квадрат вычисляется по формуле [5]:

$$\chi^2 = \sum_{j=1}^v \frac{(n_j - np_j^*)^2}{np_j^*}, \quad (1)$$

где v - число степеней свободы, n_j - количество появлений j -го уровня громкости ($j=0...255$), n - общее число появлений всех уровней громкости, $p_j^* = \frac{n_j}{n}$ - ожидаемая вероятность появления j -го уровня громкости. Ожидаемое число появлений соседних уровней громкости составляет: $n_j^* = \frac{n_j + n_{j+1}}{2}$.

Вероятность обнаружения скрытого сообщения p определяется как:

$$p = 1 - \int_0^{\chi^2} \frac{t^{(v-2)/2} e^{-t/2}}{2^{v/2} \Gamma(v/2)} dt. \quad (2)$$

Для решения поставленной задачи были подобраны 50 аудиоконтейнеров, которые условно разбиты на три класса:

- 1 - речь (стихи, голос диктора, речь с фоновой музыкой и т. д.);
- 2 - «естественная» музыка (классическая, джазовая, поп-музыка и т. д.);
- 3 - «синтезированная» музыка (электронная и прошедшая обработку на компьютере, синтезированные звуки природы и т. д.).

Для каждого блока аудиофайла, составляющего сотую часть его размера, определялись n_j , т. е. количества появлений всех 256 возможных значений отсчетов звукового сигнала. Затем с учетом общего числа появлений всех отсчетов вычислялись значение χ^2 по формуле (1) и вероятность обнаружения скрытого сообщения p по формуле (2).

На рис. 2 представлены зависимости вероятности обнаружения скрытого сообщения $p(l)$ в заполненных и незаполненных наиболее характерных аудио-контейнерах различных классов, где l - размер файла. Объем встраиваемого сообщения составлял 27 % объема контейнера. Встраивание производилось от начала контейнера во все младшие биты отсчетов звука.

Как видно из рис. 2, для аудиофайлов первого класса вероятность обнаружения сообщения в заполненной части контейнера равна 100 %, в незаполненной - 0 %. При анализе вероятностей обнаружения скрываемой информации в заполненном и незаполненном контейнерах второго класса можно выделить области, в которых наиболее велика вероятность найти сообщение. Однако критерий Хи-квадрат для таких аудиофайлов показывает 100 % вероятность

обнаружения сообщения даже в некоторых незаполненных частях контейнера. Для аудиофайлов третьего класса такая вероятность в заполненном и незаполненном контейнерах стремится к единице.

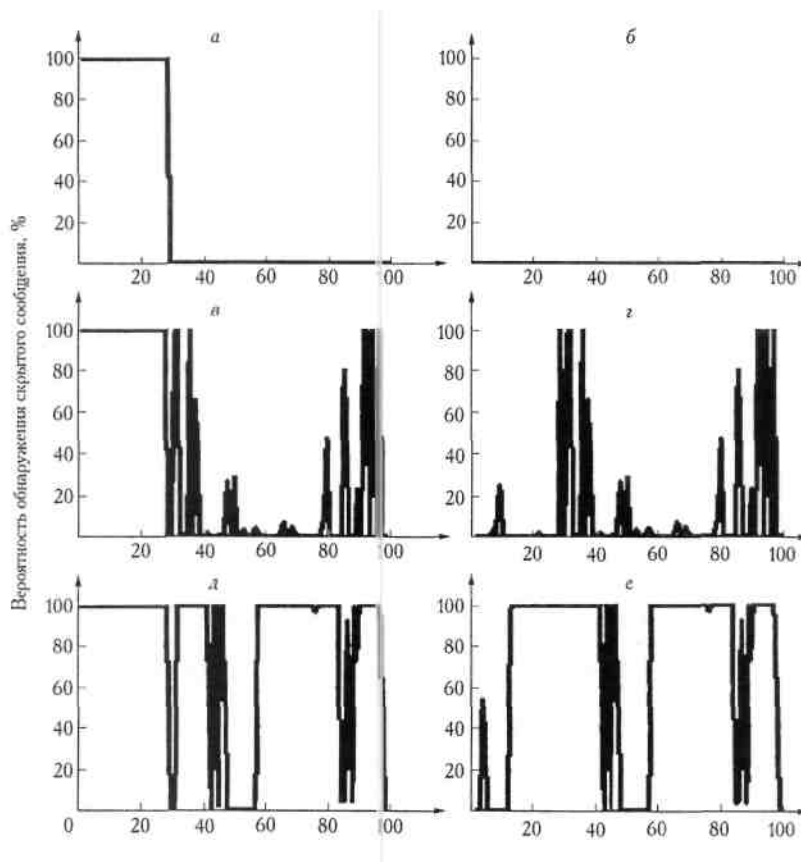


Рис. 2. Вероятности обнаружения скрытого сообщения в аудиофайлах первого (а, б), второго (в, г), третьего (д, е) классов.
а, в, д - заполненный контейнер; б, г, е - пустой контейнер

Таким образом, отнесение звуковых файлов к предложенному условному второму или третьему классу не позволяет определить применимость для них данного критерия с целью оценить наличие скрытого канала передачи сообщений.

Так как критерий Хи-квадрат основывается на оценке равномерности распределения соседних значений отсчетов сигналов, то вероятность обнаружения сообщения тем больше, чем меньше разброс значений этих отсчетов и чем больше крутизна распределения заполнения отсчетов, характеризующая их наибольшее различие. Соответственно критерий выявляет наличие скрываемой информации тем хуже, чем распределение ближе к равномерному. Степень равномерности можно оценить по величине доверительного интервала. Процесс принятия решения о применимости критерия Хи-квадрат для оценки стего-контейнеров путем вычисления доверительного интервала представлен на рис. 3.

На рис. 4 приведены примеры усредненных гистограмм, отображающих частоту появления значений отсчетов различных сигналов.

Результаты вычисления величины доверительного интервала наиболее характерных аудиофайлов одинаковой длины, относящихся к различным классам, представлены в таблице.

Статистика исследования большого числа контейнеров подтвердила, что наибольшая величина доверительного интервала соответствует аудиосигналам третьего класса, а наименьшая - первого.



Рис. 3. Алгоритм принятия решения о применимости критерия Хи-квадрат для оценки наличия скрытого канала передачи сообщений

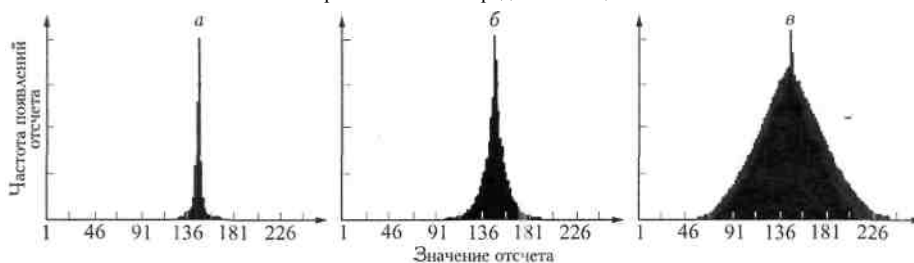


Рис. 4. Гистограммы значений отсчетов аудиоконтейнеров: а - первого класса, б - второго класса, в - третьего класса

Величина доверительного интервала для аудиосигналов различных классов									
Характеристика	Класс								
	речь			«естественная» музыка			«синтезированная» музыка		
	1	2	3	1	2	3	1	2	3
Max	1980514	4150814	1713721	1612275	1330277	1415539	414860	962761	607614
Threshold	39610	83016	34274	32245	26605	28310	8297	19255	12152
Scattering	21	30	36	59	93	92	148	118	132
Basic scattering	0-40			40-100			100-255		

Примечание . 1-3 - характерные представители класса. Max - максимальное число появления значения отсчета в контейнере; threshold - величина порога, по которому вычисляется доверительный интервал ($threshold = \max 0,02 = n/0,02$); scattering - доверительный интервал для каждого представителя класса; basic scattering - доверительный интервал для класса в целом.

При величине доверительного интервала больше 60 критерий Хи-квадрат не позволяет выявить наличие встроенного сообщения. Для аудиоконтейнеров со значением интервала меньше 60 критерий обнаруживает скрываемое сообщение, и вероятность его обнаружения стего стремится к 100 % лишь в заполненной части контейнера.

1. Чваркова И.Л., Тихоненко С.Г., Садов В.С. // Изв. Бел. инж. акад. 2004. № 1(17)/2. С. 171.

2. Moskowitz I.S., Longdon G.E., Chang L. // Proceedings of Workshop «New Security Paradigms», 2000. P. 41.

3. Bassia P., Pitas I. // IX European Signal Processing Conference (EUSIPCO'98), 8-11 September 1998. Rhodes, 1998. Vol. 1. P. 25.

Вестник БГУ. Сер. 1. 2005. № 2

4. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография. М., 2002. С. 137.

5. Гмурман В.С. Теория вероятностей и математическая статистика: Учеб. пособие для студентов вузов. М., 2003.

Поступила в редакцию 28.12.04.

Ирина Леонидовна Чваркова - студентка 5-го курса факультета радиофизики и электроники.

Василий Сергеевич Садов - кандидат технических наук, доцент кафедры интеллектуальных систем.

Александр Федорович Чернявский - доктор технических наук, профессор, академик НАН Беларуси, директор НИИПФП им. А.Н. Севченко БГУ.