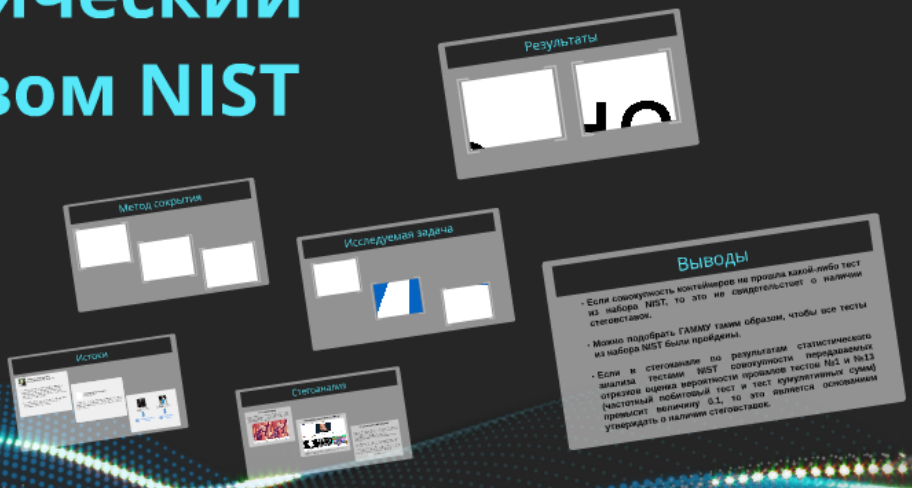
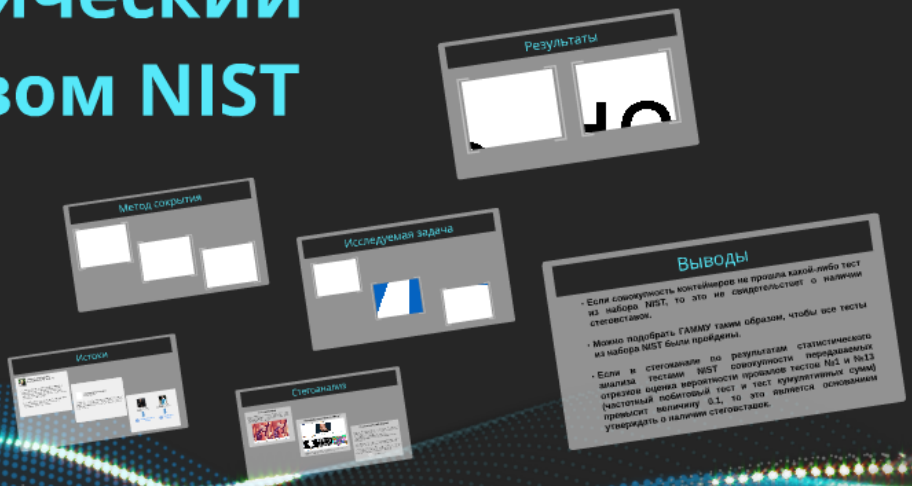
Стеганографический метод ассоциативной защиты. Статистический анализ посредством NIST

Р.Ф. Гибадуллин,
И.С. Вершинин



Стеганографический метод ассоциативной защиты. Статистический анализ посредством NIST

Р.Ф. Гибадуллин,
И.С. Вершинин



Истоки

Райхлин Вадим Абрамович,
доктор физ.-мат. наук, профессор

- Райхлин В.А. Об использовании аппарата двумерного ассоциативного поиска в процессе распознавания // Проблемно-ориентированные средства повышения эффективности вычислительных систем. – Казань: КАИ им. А.Н. Туполева. – 1991. – С.38-54.
- Райхлин В.А. Анализ производительности процессорных матриц при распознавании двоичных образов // Автометрия. – 1996. – № 5. – С.97-103.

Вершинин Игорь Сергеевич,
к.т.н., доцент

Вершинин И. С. Моделирование двумерно-ассоциативных механизмов маскирования стилизованных бинарных изображений //Диссертация. – Казань: КГТУ им. А.Н. Туполева. – 2004. – С.10-18.

Гибадуллин Руслан Фаршатович, ***к.т.н.***

Пыстогов Сергей Васильевич, ***соискатель***

1. Security Map-Point Cluster
2. Security Map Cluster

Райхлин Вадим Абрамович,
доктор физ.-мат. наук, профессор

- Райхлин В.А. Об использовании аппарата двумерного ассоциативного поиска в процессе распознавания // Проблемно-ориентированные средства повышения эффективности вычислительных систем. – Казань: КАИ им. А.Н. Туполева. – 1991. – С.38-54.

- Райхлин В.А. Анализ производительности процессорных матриц при распознавании двоичных образов // Автометрия. – 1996. – № 5. – С.97-103.

Вершинин Игорь Сергеевич,
к.т.н., доцент

Вершинин И. С. Моделирование двумерно-ассоциативных механизмов маскирования стилизованных бинарных изображений //Диссертация. – Казань: КГТУ им. А.Н. Туполева. – 2004. – С.10-18.

Гибадуллин Руслан Фаршатович, *к.т.н.*

Пыстогов Сергей Васильевич, *соискатель*

1 **Security Map-Point Cluster**

2 **Security Map Cluster**

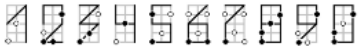
Метод сокрытия

Алгоритм маскирования (*Вершинин И.С.*)

1. $i := 0$.
2. Генерировать случайную перестановку эталонного множества D_0 с натуральными числами $\Theta_1, \Theta_2, \ldots, \Theta_r$, образуя тем самым список $C_0 = (\Theta_i)_{i=1,2,\ldots,r}$. Дополнить этот список нулевым элементом Θ_{r+1}. Ни один элемент списка C_0 изначально не отмечен.
3. $i := i + 1$.
4. $j := 1$, $k := 1$. Считать Θ_1 первым элементом множества D_0.
5. Пока не встретился неотмеченный элемент списка C_0:
   $j := j + 1$. Если Θ_j не пуст, идти к п. 6. Иначе - к п. 10.
6. $\Theta_j \oplus \Theta_i$ (побитно) $\to A_0$ (булева матрица).
7. Пока не встретился неотмеченный элемент списка Γ_i:
   $j := j + 1$. Если Θ_j не пуст, идти к п. 8. Иначе - к п. 13.
8. $\Theta_j \oplus \Theta_i$ (побитно) $\to A_2$ (булева матрица).
9. $A_2 = A_1$.
10. $A_1 \& A_2$ (побитно) $\to A_1$. Если $A_1 \neq 0$, идти к п. 7. Иначе - к п. 11.
11. $k := k - 1$. Считать Θ_k k- элементом множества D_0.
12. $A_1 = A_0$. Перейти к п. 7.
13. Случайным образом набрать один из единичных элементов матрицы A_1. Его координаты (p, q) определяют номер единичной матрицы $\overline{\Gamma}_{pq}$ инверсной матрицы списка $\overline{\Gamma}$, для всех единичных элементов списка C_1.
14. Отметить элементы списка C_1, включенные во множество D_{j+1}.
15. $l := l + 1$; $D_l := D_{j+1}$; $j := k$. Перейти к п. 2.
16. Сформировать инверсной маски для последнего единичного элемента списка C_l считать заполненным. Отметить этот элемент, аннулировав его номер списке C_l и множестве D_{j+1} (сделав их пустыми).
17. $l := l - 1$. Если $l \geq 0$, идти к п. 18. Иначе - к п. 19.
18. Пока не встретился неотмеченный элемент списка Γ_i:
    $i := i + 1$. Перейти к п. 4.
19. КОНЕЦ.

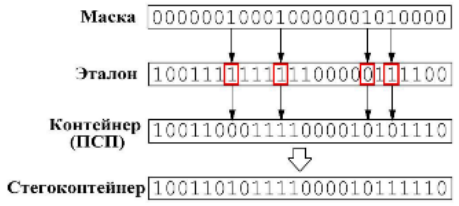
Пример

001	111	111	101	111	001	111	111	111	111
011	101	010	101	100	010	010	101	101	101
101	101	111	111	111	111	100	111	111	101
001	010	010	001	001	101	100	101	010	101
001	111	100	001	111	111	100	111	100	111



000	000	000	000	000	000	000	000	000	000
000	001	001	000	101	101	101	001	001	001
010	000	000	010	000	000	000	010	000	010
000	100	010	000	010	010	010	100	100	100
100	110	100	100	100	101	101	100	110	100
---	---	---	---	---	---	---	---	---	---
---	--1	--0	---	1-0	0-0	0-0	--1	--1	--1
-0-	---	---	-1-	---	---	---	-1-	---	-0-
---	0--	-1-	---	-0-	-0-	-0-	1--	0--	1--
0--	11-	1--	0--	1--	1-1	1-0	1--	10-	1--

Формирование стегоконтейнера



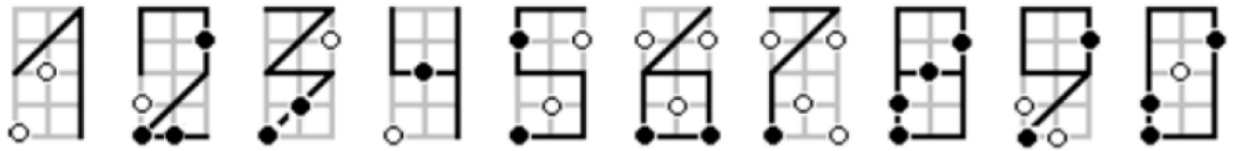
Prezi

Алгоритм маскирования (*Вершинин И.С.*)

1. $l := 0$.
2. Занумеровать случайную перестановку эталонов множества D_l в натуральном порядке $Э_1, Э_2, \dots, Э_\gamma$, образовав тем самым список $C_l = (Э_j)$, $j = 1, 2, \dots \gamma$. Дополнить этот список пустым элементом $Э_{\gamma+1}$. Ни один элемент списка C_l изначально не отмечен.
3. $i := 1$.
4. $j := i$, $k := 1$. Считать $Э_i$ первым элементом множества D_{l+1}.
5. Пока не встретится неотмеченный элемент списка C_l
   $j := j + 1$. Если $Э_j$ не пуст, идти к п. 6. Иначе - к п. 16.
6. $Э_i \oplus Э_j$ (побитно) $\rightarrow A_1$ (булева матрица).
7. Пока не встретится неотмеченный элемент списка C_l
   $j := j + 1$. Если $Э_j$ не пуст, идти к п. 8. Иначе - к п. 13.
8. $Э_i \oplus Э_j$ (побитно) $\rightarrow A_2$ (булева матрица).
9. $A_3 := A_1$.
10. A_1 & A_2 (побитно) $\rightarrow A_1$. Если $A_1 \neq |0|$, идти к п. 7. Иначе - к п. 11.
11. $k := k + 1$. Считать $Э_j$ k– элементом множества D_{l+1}.
12. $A_1 := A_3$. Переход к п. 7.
13. Случайным образом выбрать один из единичных элементов матрицы A_1. Его координаты (p, q) определяют новый единичный элемент $\overline{m}_{pq}$ инверсной матрицы масок $\overline{M}$ для всех неотмеченных эталонов списка C_l.
14. Отметить элементы списка C_l, включенные во множество D_{l+1}.
15. $l := l + 1$; $D_l := D_{l+1}$, $\gamma := k$. Переход к п. 2.
16. Формирование инверсной маски для последнего неотмеченного элемента списка C_l считать законченным. Отметить этот элемент, аннулировав тем самым список C_l и множество D_{l+1} (сделав их пустыми).
17. $l := l - 1$. Если $l \geq 0$, идти к п. 18. Иначе - к п. 19.
18. Пока не встретится неотмеченный элемент списка C_l
    $i := i + 1$. Переход к п. 4.
19. КОНЕЦ.

Пример

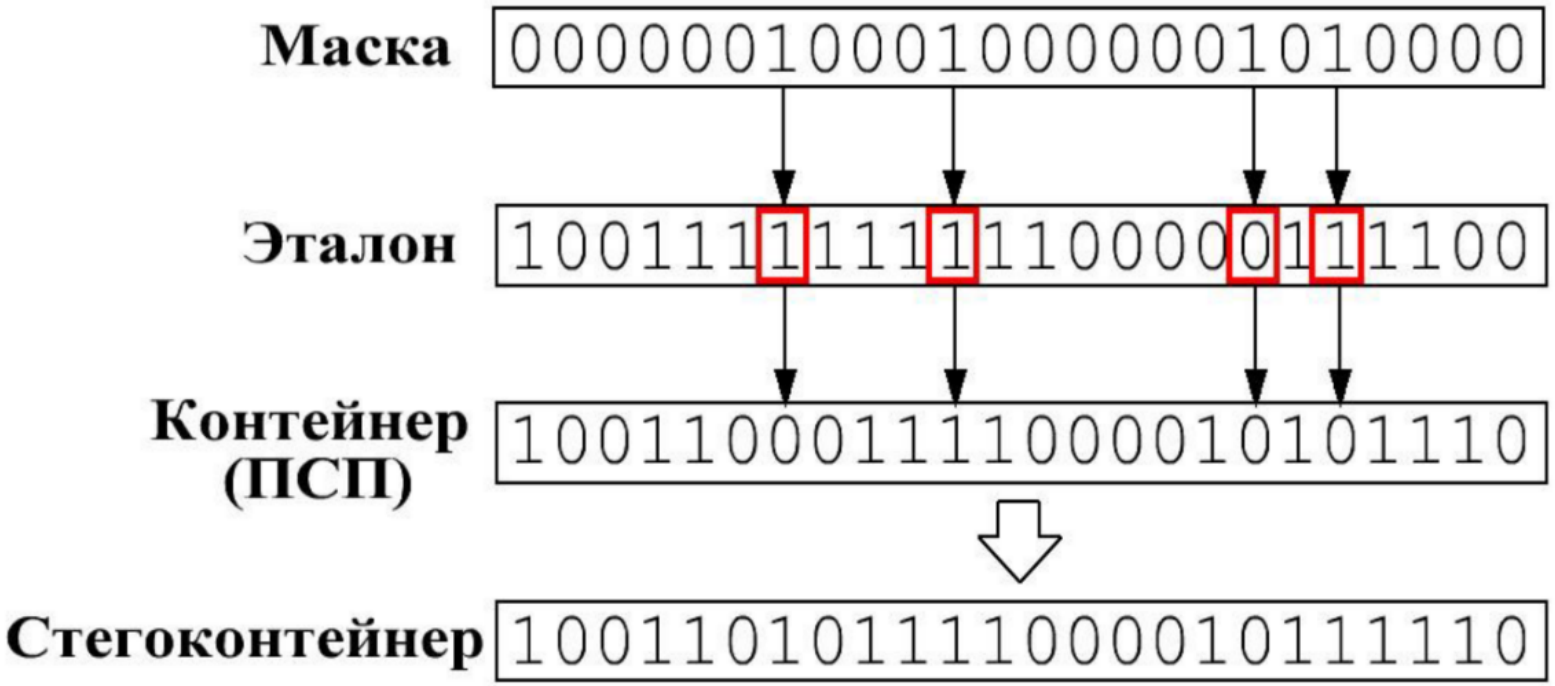
001	111	111	101	111	001	111	111	111	111
011	101	010	101	100	010	010	101	101	101
101	101	111	111	111	111	100	111	111	101
001	010	010	001	001	101	100	101	010	101
001	111	100	001	111	111	100	111	100	111



000	000	000	000	000	000	000	000	000	000
000	001	001	000	101	101	101	001	001	001
010	000	000	010	000	000	000	010	000	010
000	100	010	000	010	010	010	100	100	100
100	110	100	100	100	101	101	100	110	100

– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –	– – –
– – –	– – 1	– – 0	– – –	1 – 0	0 – 0	0 – 0	– – 1	– – 1	– – 1
– 0 –	– – –	– – –	– 1 –	– – –	– – –	– – –	– 1 –	– – –	– 0 –
– – –	0 – –	– 1 –	– – –	– 0 –	– 0 –	– 0 –	1 – –	0 – –	1 – –
0 – –	1 1 –	1 – –	0 – –	1– –	1 – 1	1 – 0	1 – –	1 0 –	1 – –

Формирование стегоконтейнера



Стегоанализ

Стеганография

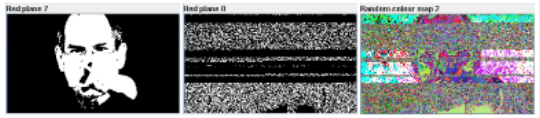
Стеганография (от греч. στεγανός — скрытый + γράφω — пишу; буквально «тайнопись») — наука, позволяющая спрятать передаваемые данные в некотором контейнере, таким образом скрыв сам факт передачи информации.

Оба изображения «весят» 786 486 байт, но правое содержит сообщения 10 первых глав «Лолиты» Набокова.

Контейнер или стегоконтейнер?

Рассмотрим на примере изображения с фотографией Стива Джобса:

Применим к этому изображению визуальную атаку, построим новые изображения из отдельных значащих бит соответствующих разрядов:



На втором и третьем изображении заметны области с высокой энтропией (высокой плотностью данных) — это и есть внедренное сообщение.

[1] Алексей Шульмин, Евгения Крылова Стеганография в современных кибератаках (Август, 2017): https://securelist.ru/steganography-in-contemporary-cyberattacks/79090/

Статистический анализ

Статистические методы анализа для выявления стегоконтейнера имеют определенный успех [1]. Например, для изображений применяются такие методы, как
- «хи-квадрат»-метод (А. Вестфелд, А. Пфитцман; 2000 г.),
- «регулярный-сингулярный» или RS-метод (Д. Фридрих, М. Гольян, А. Пфитцман; 2001 г.)

Предполагается, что в нашем случае для выявления факта наличия стеговставок подходит статистический анализ посредством NIST [2].

[2] Статистические тесты NIST - пакет статистических тестов, разработанный Лабораторией информационных технологий Национального института стандартов и технологий (NIST):

https://www.nist.gov

Стеганография

Стеганография (от греч. στεγανός — скрытый + γράφω — пишу; буквально «тайнопись») — наука, позволяющая спрятать передаваемые данные в некотором контейнере, таким образом скрыв сам факт передачи информации.

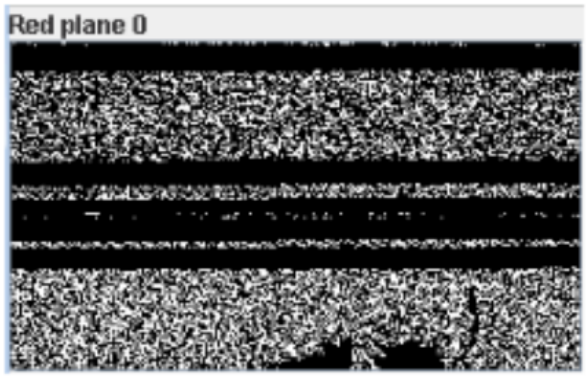
Lenna.bmp Lenna_stego.bmp

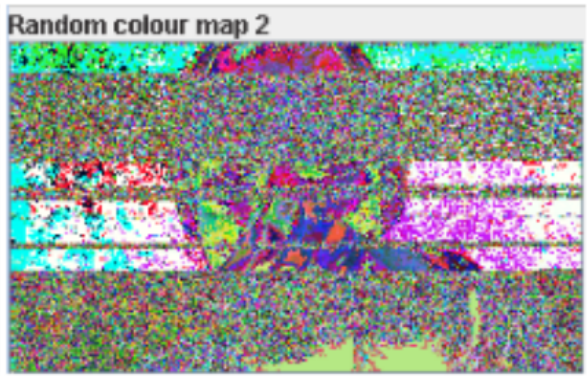
Оба изображения «весят» 786 486 байт, но правое содержит сообщения 10 первых глав «Лолиты» Набокова.

Контейнер или стегоконтейнер?

Рассмотрим на примере изображения с фотографией Стива Джобса:

Применим к этому изображению визуальную атаку, построим новые изображения из отдельных значащих бит соответствующих разрядов:





На втором и третьем изображении заметны области с высокой энтропией (высокой плотностью данных) — это и есть внедренное сообщение.

[1] Алексей Шульмин, Евгения Крылова Стеганография в современных кибератаках (Август, 2017): https://securelist.ru/steganography-in-contemporary-cyberattacks/79090/

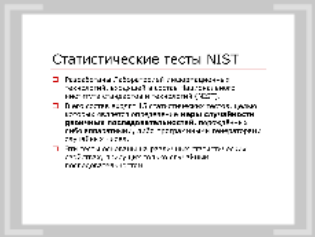
Статистический анализ

Статистические методы анализа для выявления стегоконтейнера имеют определенный успех [1]. Например, для изображений применяются такие методы, как
- «хи-квадрат»-метод (А. Вестфелд, А. Пфитцман; 2000 г.),
- «регулярный-сингулярный» или RS-метод (Д. Фридрих, М. Гольян, А. Пфитцман; 2001 г.)

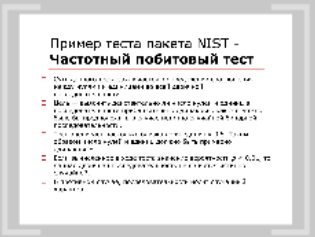
Предполагается, что в нашем случае для выявления факта наличия стеговставок подходит статистический анализ посредством NIST [2].

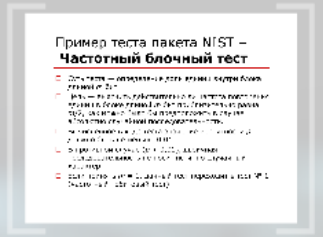
[2] Статистические тесты NIST - пакет статистических тестов, разработанный Лабораторией информационных технологий Национального института стандартов и технологий (NIST):

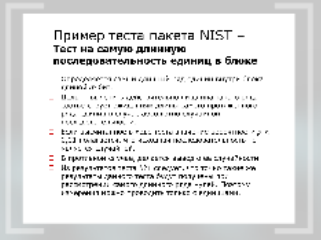
https://www.nist.gov


Статистические тесты NIST


Состав пакета NIST


Пример теста пакета NIST -
Частотный побитовый тест


Пример теста пакета NIST –
Частотный блочный тест


Пример теста пакета NIST –
Тест на самую длинную последовательность единиц в блоке

Статистические тесты NIST

- Разработаны Лабораторией информационных технологий, входящей в состав Национального института стандартов и технологий (NIST).
- В его состав входят 15 статистических тестов, целью которых является определение **меры случайности двоичных последовательностей**, порождённых либо аппаратными, либо программными генераторами случайных чисел.
- Эти тесты основаны на различных статистических свойствах, присущих только случайным последовательностям.

Состав пакета NIST

1.1 Частотный побитовый тест
1.2 Частотный блочный тест
1.3 Тест на последовательность одинаковых битов
1.4 Тест на самую длинную последовательность единиц в блоке
1.5 Тест рангов бинарных матриц
1.6 Спектральный тест
1.7 Тест на совпадение неперекрывающихся шаблонов
1.8 Тест на совпадение перекрывающихся шаблонов
1.9 Универсальный статистический тест Маурера
1.10 Тест на линейную сложность
1.11 Тест на периодичность
1.12 Тест приблизительной энтропии
1.13 Тест кумулятивных сумм
1.14 Тест на произвольные отклонения
1.15 Другой тест на произвольные отклонения

Пример теста пакета NIST - **Частотный побитовый тест**

- Суть данного теста заключается в определении соотношения между нулями и единицами во всей двоичной последовательности.
- Цель — выяснить действительно ли число нулей и единиц в последовательности приблизительно одинаковы, как это можно было бы предположить в случае истинно случайной бинарной последовательности.
- Тест оценивает насколько близка доля единиц к 0,5. Таким образом число нулей и единиц должно быть примерно одинаковым.
- Если вычисленное в ходе теста значение вероятности $p < 0{,}01$, то данная двоичная последовательность не является истинно случайной.
- В противном случае, последовательность носит случайный характер.

Пример теста пакета NIST – **Частотный блочный тест**

- Суть теста — определение доли единиц внутри блока длиной m бит.
- Цель — выяснить действительно ли частота повторения единиц в блоке длиной m бит приблизительно равна $m/2$, как можно было бы предположить в случае абсолютно случайной последовательности.
- Вычисленное в ходе теста значение вероятности p должно быть не меньше 0,01.
- В противном случае ($p < 0{,}01$), двоичная последовательность не носит истинно случайный характер.
- Если принять $m = 1$, данный тест переходит в тест № 1 (частотный побитовый тест).

Пример теста пакета NIST – **Тест на самую длинную последовательность единиц в блоке**

- Определяется самый длинный ряд единиц внутри блока длиной *m* бит.
- Цель — выяснить действительно ли длина такого ряда соответствует ожиданиям длины самого протяжённого ряда единиц в случае абсолютно случайной последовательности.
- Если высчитанное в ходе теста значение вероятности $p < 0{,}01$ полагается, что исходная последовательность не является случайной.
- В противном случае, делается вывод о ее случайности.
- Из результатов теста №1 следует, что точно такие же результаты данного теста будут получены при рассмотрении самого длинного ряда нулей. Поэтому измерения можно проводить только с единицами.

Исследуемая задача

Цель, задача и принятые ограничения

Цель:

Попытаться установить статистическую неразличимость совокупностей пустых и стегоконтейнеров на основании статистического анализа посредством NIST.

Задача и принятые ограничения:

Провести статистический анализ посредством набора тестов NIST двух бинарных отрезков длиной ***660 КБ***, где один из отрезков состоит из контейнеров, полученных последовательной генерацией участков ПСП генератором Вихрь Мерсенна [3], а второй отрезок – из стегоконтейнеров, полученных из первого отрезка путем стегозатравки. При этом размер каждого контейнера (стегоконтейнера) равен ***528 бит*** (***m = 60***), число опытов по анализу каждого отрезка ***N = 200***.

[3] Mersenne Twister: math.sci.hiroshima-u.ac.jp/~m-mat/MT/VERSIONS/C-LANG/mt19937ar.cs

Для генерации контейнеров (стегоконтейнеров) реализована программа на языке C#:

- https://bitbucket.org/landwatersun/forum/downloads/MCipher-20199114.zip

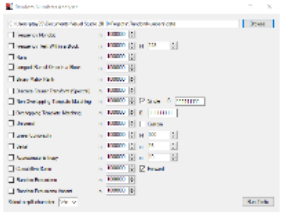
Возможности программы-генератора:

- ✓ Генерация стегоключа, контейнера, стегоконтейнера;
- ✓ Выбор размера контейнера (стегоконтейнера);
- ✓ Выбор количества контейнеров (стегоконтейнеров);
- ✓ Открытый исходный код (*проект реализован на базе .NET Framework 4.5.2 и C# program for MT19937 Copyright (C) 2006, Mitil URL:* http://meisui.psk.jp/).

Для тестирования контейнеров (стегоконтейнеров) было уделено внимание двум эквивалентным между собой проектам:

1) https://github.com/Steppenwolfe65/NIST-Statistical-Test-Suite
2) https://sourceforge.net/projects/randomanalysis/

В работе был использован второй проект



Цель, задача и принятые ограничения

Цель:

Попытаться установить статистическую неразличимость совокупностей пустых и стегоконтейнеров на основании статистического анализа посредством NIST.

Задача и принятые ограничения:

Провести статистический анализ посредством набора тестов NIST двух бинарных отрезков длиной ***660 КБ***, где один из отрезков состоит из контейнеров, полученных последовательной генерацией участков ПСП генератором Вихрь Мерсенна [3], а второй отрезок – из стегоконтейнеров, полученных из первого отрезка путем стеговставок. При этом размер каждого контейнера (стегоконтейнера) равен ***528 бит*** (***m = 60***), число опытов по анализу каждого отрезка ***N = 200***.

[3] Mersenne Twister: math.sci.hiroshima-u.ac.jp/~m-mat/MT/VERSIONS/C-LANG/mt19937ar.cs

Для генерации контейнеров (стегоконтейнеров) реализована программа на языке C#:

- https://bitbucket.org/landwatersun/forum/downloads/MCipher-20199114.zip

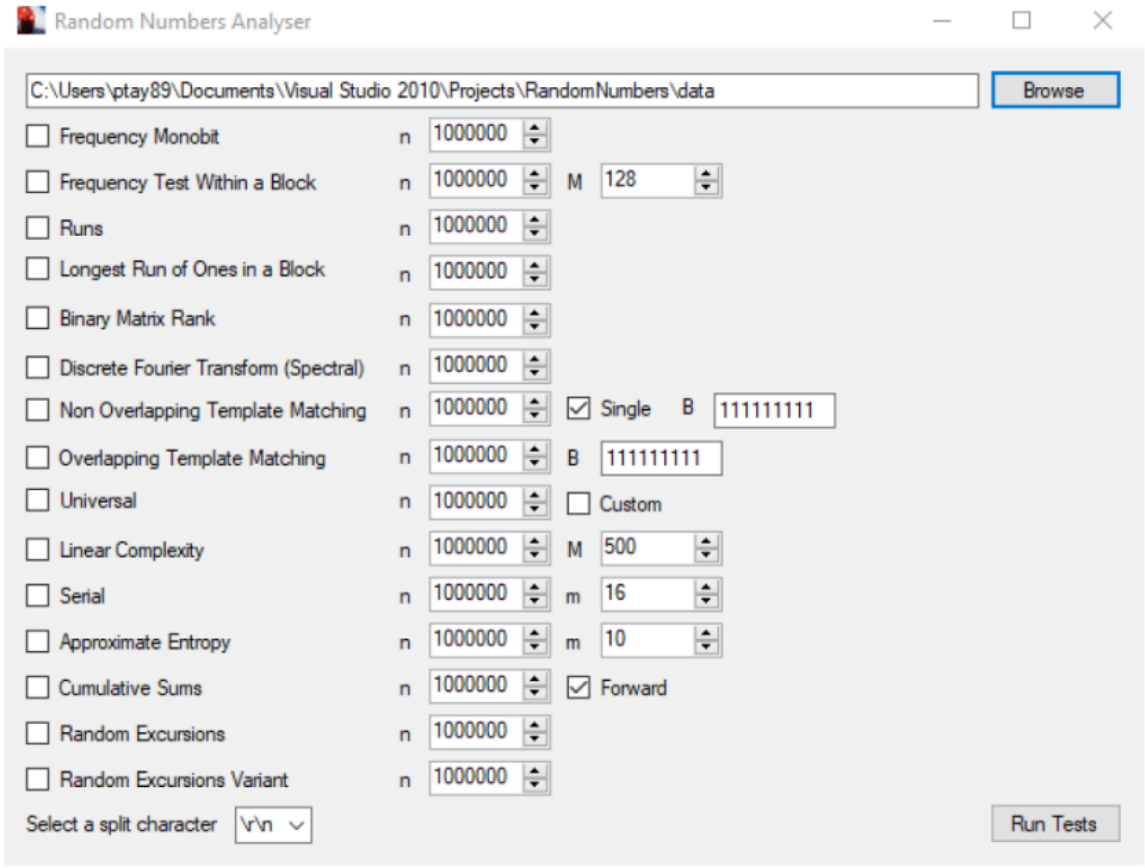
Возможности программы-генератора:

- ✓ Генерация стегоключа, контейнера, стегоконтейнера;
- ✓ Выбор размера контейнера (стегоконтейнера);
- ✓ Выбор количества контейнеров (стегоконтейнеров);
- ✓ Открытый исходный код (*проект реализован на базе .NET Framework 4.5.2 и C#-program for MT19937 Copyright (C) 2006, Mitil URL:* http://meisui.psk.jp/).

Для тестирования контейнеров (стегоконтейнеров) было уделено внимание двум эквивалентным между собой проектам:

1) https://github.com/Steppenwolfe65/NIST-Statistical-Test-Suite
2) https://sourceforge.net/projects/randomanalysis/

В работе был использован второй проект:



Результаты

Число провалов теста NIST из 200 опытов составляет:

- Для совокупности пустых контейнеров (*s1*) – 53.
- Для совокупности стегоконтейнеров (*s2*) – 101.

Таблица ниже указывает, сколько всего провалов было по каждому тесту из набора NIST в ходе данного анализа. Где P_{s1} (P_{s2}) это оценка вероятностей того, что *s1* (*s2*) не пройдет тест.

№ теста	s1	s2	s1/s2	P_{s1}	P_{s2}
1	**8**	**78**	**0,1**	**0,04**	**0,39**
2	4	3	1,33	0,02	0,015
3	6	6	1	0,03	0,03
4	3	3	1	0,015	0,015
5	3	5	0,6	0,015	0,025
6	1	5	0,2	0,005	0,025
7	0	1	0	0	0,005
8	2	3	0,67	0,01	0,015
9	1	1	1	0,005	0,005
10	4	1	4	0,02	0,005
11	4	3	1,33	0,02	0,015
12	2	4	0,5	0,01	0,02
13	**9**	**73**	**0,12**	**0,045**	**0,365**
14	9	9	1	0,045	0,045
15	14	7	2	0,07	0,035

Дополнительно получены результаты анализа стегоконтейнеров с избыточным маскированием (с числом масок $k_m = 5$)
[см. Вершинин И.С., Информация и безопасность. 2016. Т. 19. № 4]

Число провалов теста NIST из 200 опытов составляет:

- Для совокупности пустых контейнеров (*s1*) – 35.
- Для совокупности стегоконтейнеров (*s2**) – 193.

№ теста	s1	s2*	s1/s2*	P_{s1}	P_{s2^*}
1	**1**	**182**	**0,01**	**0,005**	**0,91**
2	3	4	0,75	0,015	0,02
3	1	1	1	0,005	0,005
4	3	4	0,75	0,015	0,02
5	3	3	1	0,015	0,015
6	**3**	**155**	**0,02**	**0,015**	**0,775**
7	1	56	0,02	0,005	0,28
8	0	4	0	0	0,02
9	3	1	3	0,015	0,005
10	1	4	0,25	0,005	0,02
11	1	1	1	0,005	0,005
12	2	67	0,03	0,01	0,335
13	**1**	**183**	**0,01**	**0,005**	**0,915**
14	14	1	14	0,07	0,005
15	7	1	7	0,035	0,005

При этом установлено, что {оценка мат. ожидания (M), оценка среднеквадратичного отклонения (σ)} числа бит в *s2* и *s2** отличных от *s1* из 200 опытов составляет {M = 18 623, σ = 1137} и {M = 88 300, σ = 3204} соответственно.

Число провалов теста NIST из 200 опытов составляет:

- Для совокупности пустых контейнеров (*s1*) – 53.
- Для совокупности стегоконтейнеров (*s2*) – 101.

Таблица ниже указывает, сколько всего провалов было по каждому тесту из набора NIST в ходе данного анализа. Где P_{s1} (P_{s2}) это оценка вероятностей того, что *s1* (*s2*) не пройдет тест.

№ теста	s1	s2	s1/s2	P_{s1}	P_{s2}
1	**8**	**78**	**0,1**	**0,04**	**0,39**
2	4	3	1,33	0,02	0,015
3	6	6	1	0,03	0,03
4	3	3	1	0,015	0,015
5	3	5	0,6	0,015	0,025
6	1	5	0,2	0,005	0,025
7	0	1	0	0	0,005
8	2	3	0,67	0,01	0,015
9	1	1	1	0,005	0,005
10	4	1	4	0,02	0,005
11	4	3	1,33	0,02	0,015
12	2	4	0,5	0,01	0,02
13	**9**	**73**	**0,12**	**0,045**	**0,365**
14	9	9	1	0,045	0,045
15	14	7	2	0,07	0,035

Дополнительно получены результаты анализа стегоконтейнеров с избыточным маскированием (с числом масок $k_m = 5$)
[см. Вершинин И.С., Информация и безопасность. 2016. Т. 19. № 4]

Число провалов теста NIST из 200 опытов составляет:

- Для совокупности пустых контейнеров (*s1*) – 35.
- Для совокупности стегоконтейнеров (*s2**) – 193.

№ теста	s1	s2*	s1/s2*	Ps1	Ps2*
1	**1**	**182**	**0,01**	**0,005**	**0,91**
2	3	4	0,75	0,015	0,02
3	1	1	1	0,005	0,005
4	3	4	0,75	0,015	0,02
5	3	3	1	0,015	0,015
6	**3**	**155**	**0,02**	**0,015**	**0,775**
7	1	56	0,02	0,005	0,28
8	0	4	0	0	0,02
9	3	1	3	0,015	0,005
10	1	4	0,25	0,005	0,02
11	1	1	1	0,005	0,005
12	2	67	0,03	0,01	0,335
13	**1**	**183**	**0,01**	**0,005**	**0,915**
14	14	1	14	0,07	0,005
15	7	1	7	0,035	0,005

При этом установлено, что {оценка мат. ожидания (M), оценка среднеквадратичного отклонения (σ)} числа бит в *s2* и *s2** отличных от *s1* из 200 опытов составляет {M = 18 623, σ = 1137} и { M = 88 300, σ = 3204} соответственно.

Выводы

- Если совокупность контейнеров не прошла какой-либо тест из набора NIST, то это не свидетельствет о наличии стеговставок.
- Можно подобрать ГАММУ таким образом, чтобы все тесты из набора NIST были пройдены.
- Если в стегоканале по результатам статистического анализа тестами NIST совокупности передаваемых отрезков оценка вероятности провалов тестов №1 и №13 (частотный побитовый тест и тест кумулятивных сумм) превысит величину 0.1, то это является основанием утверждать о наличии стеговставок.

Стеганографический метод ассоциативной защиты. Статистический анализ посредством NIST

Р.Ф. Гибадуллин,
И.С. Вершинин