

Академия Федеральной службы охраны Российской Федерации

Подход к внедрению робастного водяного знака в текстовые данные

кандидат технических наук Козачок Александр Васильевич Копылов Сергей Александрович



22 марта 2018 г.



- 1 Характеристика предметной области исследования
- 2 Модель защиты текстовой информации на основе внедрения робастного водяного знака
- 3 Алгоритмы внедрения и извлечения робастного водяного знака
- 4 Оценка робастности разработанного подхода
- 5 Оценка точности извлечения встроенной информации



- 1 Характеристика предметной области исследования
 - Модель защиты текстовой информации на основе внедрения робастного водяного знака
- 3 Алгоритмы внедрения и извлечения робастного водяного знака
- 4 Оценка робастности разработанного подхода
- 5 Оценка точности извлечения встроенной информации

Распределение инцидентов информационной безопасности





Внутренние штатные администраторы

Прочие внутренние пользователи

Устройства прямого доступа в интернет
 б)

• Съемные носители

О Печать

Рис. 1. Распределение инцидентов информационной безопасности*: a) по типу, б) по источнику внутренних инцидентов, в) по каналу утечки информации по внутренним инцидентам

ВНЕШНИЕ

a)



- 1 Характеристика предметной области исследования
- 2 Модель защиты текстовой информации на основе внедрения робастного водяного знака
- 3 Алгоритмы внедрения и извлечения робастного водяного знака
- 4 Оценка робастности разработанного подхода
- 5 Оценка точности извлечения встроенной информации

Модель защиты текстовой информации



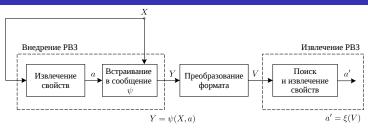


Рис. 2. Модель защиты текстовой информации от НСД за счет внедрения робастного водяного знака (РВЗ) при преобразовании формата

- 1 Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi.
- 2 Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobor
- 3 vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan biben
 - 4 erat ligula aliquet magna, vitae ornare odio metus a mi. $h+\Delta$
- 5 tesque a nulla. Cum sociis natoque penatibus et magnis dis parturi
- $_{6}$ montes, nascetur ridiculus mus. Aliquam tincidunt urna. $h-\Delta$
- 7 per vestibulum turpis. Pellentesque cursus luctus mauris.

Оценка емкости встраивания информации



Гарнитура используемого шрифта:

- шрифт с засечками (serif) Computer Modern Roman (аналог Times New Roman);
- 2 шрифт без засечек (sans serif) Computer Modern Sans Serif (аналог Arial);
- 3 моноширинный шрифт (monospace) Computer Modern Typewriter (аналог Courier New).

Таблица 1. Зависимость емкости встраивания от изменения межстрочного интервала и кегля шрифта

Кегль	Межстрочный	Предельно допустимая емкость
шрифта	интервал	встраивания (бит)
10	1	60
10	1,5	40
10	2	30
12	1	49
12	1,5	33
12	2	24
14	1	42
14	1,5	28
14	2	21



- 1 Характеристика предметной области исследования
 - Модель защиты текстовой информации на основе внедрения робастного водяного знака
- 3 Алгоритмы внедрения и извлечения робастного водяного знака
- 4 Оценка робастности разработанного подхода
- 5 Оценка точности извлечения встроенной информации

Алгоритмы внедрения и извлечения робастного водяного знака



```
Data: Текстовый документ TD_0,
           встраиваемая информация І
   Result: Подписанный текстовый документ
            TD_{\varsigma}
1 Len \leftarrow GetLength (I)
2 N \leftarrow \text{CountLines } (TD_0)
3 if N > (Len + 1) then
         for i \leftarrow 0 to (N-2) do
               i \leftarrow i \mod Len
                if I_j = 1 then
                       TD_0 \leftarrow \text{Embed } (TD_0, i)
         TD_s \leftarrow TD_0
9 return TD_s
```

5

7

Рис. 4. Алгоритм внедрения РВЗ в текстовый документ

```
Data: Изображение, содержащее текст Im_t
    Result: Встроенная информация I_e
 1 Im_{arey} \leftarrow ConvertToGray (Im_t)
 2 Im_{filt} \leftarrow Filtration (Im_{grey})
 3 sinogram \leftarrow RadonTransform (Im_{filt})
 4 for l \leftarrow 0 to 180 do
    R[l] \leftarrow \mathsf{RmsCalculation} \ (sinogram)
 6 rot ← Argmax(R)
 7 row \leftarrow sinoaram[rot]
   M \leftarrow \mathsf{FindPicks}(row)
 9 D ← CorrectErrors(M)
10 min, max \leftarrow FindModes(D)
    if |D| > 3 then
          if Std(D) > 2 then
13
                  GaussianMixture(D, min, max)
14
15
16
17
      |I_e \leftarrow \{\}
19 return I_e
```



- 1 Характеристика предметной области исследования
 - Модель защиты текстовой информации на основе внедрения робастного водяного знака
- 3 Алгоритмы внедрения и извлечения робастного водяного знака
- 4 Оценка робастности разработанного подхода
- 5 Оценка точности извлечения встроенной информации

Оценка свойств робастности разработанного водяного знака I



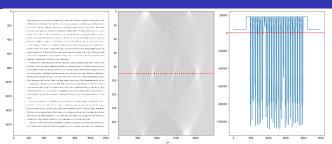


Рис. 6. Пример извлечения линий текста из изображения, содержащего текст

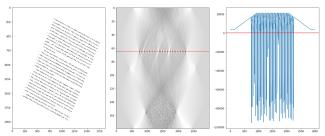


Рис. 7. Извлечение данных из изображения, повернутого на 25°

Оценка свойств робастности разработанного водяного знака II



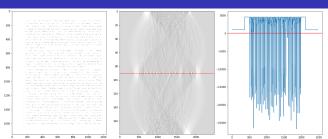


Рис. 8. Извлечение данных из изображения с примененным медианным фильтром

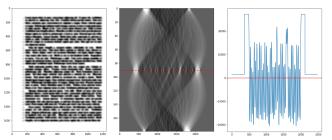


Рис. 9. Извлечение данных из изображения с примененным гауссовским фильтром

Оценка свойств робастности разработанного водяного знака III



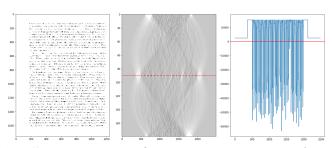


Рис. 10. Извлечение данных из изображения с примененным медианным фильтром

Тип преобразования	Устойчивость разработанного алгоритма к преобразованию		
Поворот	поворот на любой угол		
Масштабирование	коэффициент масштабирования до 150%		
Преобразование формата	в любой формат растровых изображений		
Медианная фильтрация	с пределом ядра свертки 9 пикселей		
Гауссовская фильтрация	с пределом радиуса размытия 8 пикселей		
Усредненная фильтрация	с пределом ядра свертки 5 пикселей		
Уменьшение DPI	до 25 пикселей на дюйм		

Таблица 2. Стойкость разработанного алгоритма к преобразованиям применяемым к изображениям



- 1 Характеристика предметной области исследования
 - Модель защиты текстовой информации на основе внедрения робастного водяного знака
- 3 Алгоритмы внедрения и извлечения робастного водяного знака
- 4 Оценка робастности разработанного подхода
- 5 Оценка точности извлечения встроенной информации

Оценка точности извлечения количества строк



Таблица 3. Массив извлеч	енных величин межстрочных интервалов
--------------------------	--------------------------------------

Файл	Число	Вычис-	Массив извлеченных величин межстрочных
	строк в	ленное	интервалов
	исходном	число	
	документе	строк	
10_1	61	57	15, 15, 16, 15, 15, 15, <u>39</u> , 15, 15, 15, 15, 15, 15, <u>40</u> , 15,
			15, 14, 15, <u>40</u> , 15, 15, 15, 15, 15, 16, 15, 14, 15, 15, 15,
			15, 15, 15, 15, <u>40</u> , 15, 14, 15, 15, 15, 15, 15, 15, 15, 15,
			16, 14, 15, 15, 15, 15, 15, 15, 15, 15
10_2	31	34	40, 40, 40, 39, 3 , 36, 40, 40, 3 , <u>86</u> , 39, 40, 40, 40, 40,
			39, <u>3</u> , 36, 40, 40, 40, 40, 39, 40, 40, <u>3</u> , 36, 40, 39, 40,
			40, 40, 40

Таблица 4. Скорректированные массивы величин межстрочных интервалов

Файл	Число	Исправ-	Массив исправленных величин межстрочных			
	строк в	ленное	интервалов			
	исходном	число				
	документе	строк				
10_1	61	61	15, 15, 16, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15			
			15, 15, 15, 14, 15, 15, 15, 15, 15, 15, 15, 16, 15, 14,			
			15, 15, 15, 15, 15, 15, 15, 15, 15, 15,			
			15, 15, 15, 15, 16, 14, 15, 15, 15, 15, 15, 15, 15, 15			
10 2	31	31	40, 40, 40, 39, 36, 40, 40, 40, 40, 39, 40, 40, 40, 40, 39,			
			36, 40, 40, 40, 40, 39, 40, 40, 36, 40, 39, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40			
10_2	31	31				

Оценка точности извлечения встроенной информации I



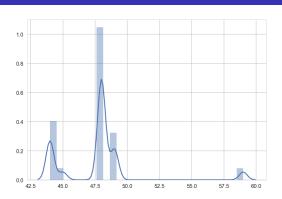


Рис. 11. Гистограмма с ядерной оценкой плотности распределения

Таблица 5. Результат извлечения информации из изображения со встроенным РВЗ

таблица 3. г свультат извисчения информации из изображения со встроенным г во				
Файл	In	Двоичный массив межстрочных интервалов	Out	
14_1	0110	0, 1, 1, 0, <u>1</u> , <u>0</u> , 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, <u>1</u> , 0, 1, 1, 0, 0, 1, 1, 0,	0110	
		0, 1, 1, 0, 0, 1, 1, 0, 0, 1		
14_1	0111	0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1,	0111	
		0, 1, 1, 1, 0, 1, 1, 1, 0		

Оценка точности извлечения встроенной информации II



Таблица 6. Результат извлечения информации при встраивании ее в документ с параметрами шрифта: 14 пт, межстрочный интервал – 1, величина изменения интервала – 1,1 интервала = 1,4 пт \approx 0,49 мм

Разрешение изображения (DPI)	Время обработки, сек	Число строк в исходном документе	Извлеченное число строк	Точность	Вероятность ложных срабатываний	Вероятность пропуска цели
25	0,9	41	40	0,667	0,033	0,30
50	4	41	41	0,886	0,033	0,081
100	15	41	41	0,984	0,016	0
150	34	41	41	0,992	0,008	0
200	62	41	41	0,992	0,008	0
250	105	41	41	0,992	0,008	0
300	166	41	41	0,992	0,008	0

Таблица 7. Результат извлечения информации при встраивании ее в документ с параметрами шрифта: 14 пт, межстрочный интервал -1,5, величина изменения интервала -1,1 интервала =1,4 пт ≈ 0.49 мм

Разрешение изображения (DPI)	Время обработки, сек	Число строк в исходном документе	Извлеченное число строк	Точность	Вероятность ложных срабатываний	Вероятность пропуска цели
25	0,8	28	28	0,62	0,083	0,297
50	4	28	28	0,75	0,024	0,226
100	16	28	28	0,988	0,012	0
150	35	28	28	1	0	0
200	70	28	28	1	0	0
250	101	28	28	1	0	0
300	143	28	28	1	0	0 17/20

Оценка точности извлечения встроенной информации III



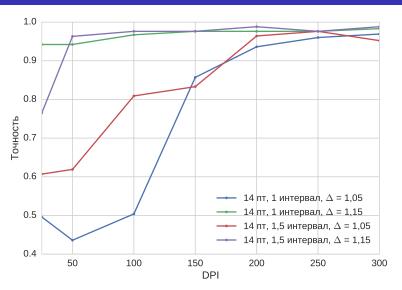


Рис. 12. Зависимость точности извлечения информации от DPI

Оценка точности извлечения встроенной информации IV



10 processors a serial correspond processes, sandpostations STREET, CONTRACTOR STREET, PURSUITABLE STREET, SCHOOL SCHOOL STREET, приводиливаеть из подворожний, положный и транторовае фодуть: Ю учинавления, в головичения г закондачивачного Российной Форрации прадва планевичноми, правими, учеть в венежности CHARGE CONTRACT AND ADDRESS OF THE PARTY OF прина, портина во глитентичности учета, и также преблазава в принятывая A REPORT OF THE PARTY OF THE PA П) година, в учиналивно поряде проприяти, учрежники в применация, исобирання для выполнения обисимостой, исопаления на примы готорутичный примы и обстатива доставления такивания 10000 10 приложен в учинивания перера для варабота ревозий ис ворого, отвечным в обруг довужения примя порартичной reporte, my time: a most riprostructor, y terrain a recognisation, a true factor m. proposed reserv

Рис. 13. Фрагмент изображения с DPI = 20, увеличенного в 6 раз

Таблица 8. Результат извлечения информации из изображения со встроенным PB3 (DPI=20)

Размер	Извлечен-	Двоичный массив межстрочных интервалов	Точность
изобра-	ное число		извлечения
жения	строк		информации
166×234	39	0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0,	0,487
		0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,	
		0, 0, 0	
996 ×	40	1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0,	0,872
1404		1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,	
		0, 0, 1	19 / 20

Подход к внедрению робастного водяного знака в текстовые данные



Спасибо за внимание! Вопросы?





a.kozachok@academ.msk.rsnet.ru